Walden University
ScholarWorks

Walden Dissertations and Doctoral Studies

Walden Dissertations and Doctoral Studies Collection

2018

Analytic Extensions to the Data Model for Management Analytics and Decision Support in the Big Data Environment

Nsikak Etim Akpakpan
Walden University

Follow this and additional works at: https://scholarworks.waldenu.edu/dissertations

Part of the Business Administration, Management, and Operations Commons, Databases and Information Systems Commons, Library and Information Science Commons, and the Management Sciences and Quantitative Methods Commons

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact ScholarWorks@waldenu.edu.

Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Nsikak E. Akpakpan

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Aridaman Jain, Committee Chairperson,
Applied Management and Decision Sciences Faculty

Dr. Robert Levasseur, Committee Member,
Applied Management and Decision Sciences Faculty

Dr. Raghu Korrapati, University Reviewer
Applied Management and Decision Sciences Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2018

Abstract

Analytic Extensions to the Data Model for Management Analytics

and Decision Support in the Big Data Environment

by

Nsikak E. Akpakpan

MBA, University of Illinois at Chicago, 1996

MPH, University of Illinois at Chicago, 1996

DA, University of Calabar Teaching Hospital, 1989

MB, BS, University of Nigeria, 1985

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Sciences

Walden University

August 2018

Abstract

From 2006 to 2016, an estimated average of 50% of big data analytics and decision support projects failed to deliver acceptable and actionable outputs to business users. The resulting management inefficiency came with high cost, and wasted investments estimated at $2.7 trillion in 2016 for companies in the United States. The purpose of this quantitative descriptive study was to examine the data model of a typical data analytics project in a big data environment for opportunities to improve the information created for management problem-solving. The research questions focused on finding artifacts within enterprise data to model key business scenarios for management action. The foundations of the study were information and decision sciences theories, especially information entropy and high-dimensional utility theories. The design-based research in a nonexperimental format was used to examine the data model for the functional forms that mapped the available data to the conceptual formulation of the management problem by combining ontology learning, data engineering, and analytic formulation methodologies. Semantic, symbolic, and dimensional extensions emerged as key functional forms of analytic extension of the data model. The data-modeling approach was applied to 15-terabyte secondary data set from a multinational medical product distribution company with profit growth problem. The extended data model simplified the composition of acceptable analytic insights, the derivation of business solutions, and the design of programs to address the ill-defined management problem. The implication for positive social change was the potential for overall improvement in management efficiency and increasing participation in advocacy and sponsorship of social initiatives.

Analytic Extensions to the Data Model for Management Analytics

and Decision Support in the Big Data Environment

by

Nsikak E. Akpakpan

MBA, University of Illinois at Chicago, 1996

MPH, University of Illinois at Chicago, 1996

DA, University of Calabar Teaching Hospital, 1989

MB, BS, University of Nigeria, 1985

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Sciences

Walden University

August 2018

Dedication

I dedicate this dissertation to the incredible work of pioneers in Analytical and Design Sciences who provided the framework for the development of Applied Management and Decision Sciences. They are Herbert Simon, Allen Newell, Daniel Kahneman, Chester Bernard, John McCarthy, Amos Tversky, Marvin Minsky, Edgar F. Codd, Edwin Diday, Claude Shannon, Philip B. Crosby, W. Edwards Deming, Armand V. Feigenbaum, Kaoru Ishikawa, Joseph M. Juran, John Nash, John Von Neumann, Oskar Morgenstein, Henri Fayol, and Henry Mintzberg, John Willard Milnor, Jonathan Borwein, Michael Jordan, to name a few. Their work heralded a fresh approach to management, driven by information, numeracy, data, and analytics. Special dedications go to the Golden Jubilee of the Edgar F. Codd’s pioneering work on relational data-modeling, which provided the core ideas for this dissertation; and John Willard Milnor whose core ideas of Surgery Theory in analytical geometry helped me connect my passions for Medicine, Management, and Mathematics. Finally, I dedicate this dissertation to my family: my wife, Christiana Udoh; my children, Itorobong, Inemesit, and Edidiong who, playfully, would not let the turmoil of life and health derail this dream.

Acknowledgments

I would like to acknowledge my committee members, Dr. Aridaman Jain, Dr. Robert Levasseur, and Dr. Raghu Korrapati for their help with this dissertation. I am grateful to every Walden faculty member and colleague I learned from, especially, my former Committee Chair, Dr. Reza Hamzaee. To my friends, far and near, I say, this would not have been possible without all the encouragement and ideas.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

This study explored the use of applied data-modeling concepts to refine the data model for management analytics and decision support in a big data environment. The study sought to address the challenges facing data analytics projects, which included, the overwhelming availability of big data, the growing complexity of business domains, the demands of operational accountability, and the explosion of analytic techniques (De Smedt, 2013; McAfee & Brynjolfsson, 2012; Storey & Song, 2017). The issue was that insights and solutions from these projects lost alignment to well-known data and the intuitive cognitive models required for the management problem-solving.

Chapter 1 covers the following topics: background of the study, the purpose of the study, the research questions, the nature of the study, the theoretical foundation, the definition of critical terms; the scope, delimitations, and limitations of the study; and the significance of the study to management theory, business practice, and social change.

Background of the Study

I was motivated to study this topic due to a combination of research and personal experience showing that companies' efforts in the areas of data analytics and decision support were often neither effective nor efficient. Most management decisions and actions of business analysts and executives used intuitions and cognitive models, and not insights or solutions from data analytics and decision support systems (Yeoh & Popovič, 2015). The difficulty was the framing of specific management problems or opportunities with available data. For this reason, management questioned the value proposition of investments in data analytics and decision support systems. Strategic decision failures, such as, the 2008 global economic collapse and many other such occurrences in history were examples that made formal analysis and decision support systems suspicious as viable management problem-solving tools (Bosch, Nguyen, & Buckle-Henning, 2014).

With the advent of business big data, the data analytics projects faced three challenges: (a) taming the information chaos caused by the exponential growth of information assets, (b) relieving the mounting pressure to use these information assets to advance efficiency and predictability of decision-making, and (c) addressing acute problems of information deluge on decision-making, including analysis paralysis, escalating indecision, reification, and strategic ambiguity (Block, 2012; Tien, 2013). From these challenges the following two problem scenarios arose. The more important was the difficulty in the discovery of underlying structures and associations about subjects of interest. The other issue was transforming these structures and associations into actionable business insights and organizing them into scenarios to improve management programs for predictable and positive outcomes (Resmini, 2012).

These challenges were reduced when the data were in models that connected underlying elements and their associations (Hand, 2012; Thompson, 2011). A well-constructed data model captured the structure, content, and context of the underlying elements. Such models also captured mechanisms and situations responsible for the outcome observations of the domain of interest (Burch, 2018). The data model provided the stable and accurate representation of subjects within the enterprise (Johnson, 2014). Furthermore, the data model provided the foundation for the continuous discovery of the attributes of dominant subjects for management problem-solving (Beroggi, 2010; Kwakkel, Walker, & Marchau, 2010). Additionally, the data model carefully rationalized and integrated the attributes from all relevant data sources without compromising the integrity of the data generation processes, therefore, provided the most comprehensive collection of the subjects responsible for the performance of the enterprise.

With the advent of big data, input data structures came in many different forms. It was not uncommon for data structures like online transaction processing (OLTP), online analytic processing (OLAP), relational, object-relational, hierarchical (or graph), network, document, and flat data structures to be part of a single data analytics project scope. Because of the size and dimensionality of these data sources, it was typical for contemporary analytic processes to partition and sample the data to limit complexity. Partitioning led to the deluge of partial analytic solutions and data silos, for example, static reports, dynamic reports (dashboards, scorecards), and analytic algorithms (Kalou & Koutsomitropoulos, 2014). Sampling raised issues of representativeness, bias, and the requirement of statistical validation. New and advanced data analytics methodologies arose to overcome these concerns.

The advanced data analytics techniques included semantic data analysis, statistical data analysis, symbolic data analysis, functional data analysis, topographical data analysis, projection pursuit analysis, exploratory system dynamics and modeling, data mining, and deep learning to name a few. These advanced data analytics techniques were responses to the growing availability of data and the demand to use them to guide knowledge and learning (Paganoni & Secchi, 2014). A unique challenge of these advanced analytic methods was in the pre-processing of the data for the analytic technique selected (Kaisler, Espinosa, Armour, & Money, 2014). This pre-processing step required the selection of attributes, sampling of the data, and transformations of the data in ways that caused loss of business interpretability and value (Kalou & Koutsomitropoulos, 2014; Ma et al., 2014). For this reason, I chose the approach of analyzing the data to determine how to extend the model to accommodate the unique challenges posed by big data in data analytics projects for management problem solving without the constraints imposed by analytic methodologies.

Problem Statement

Most business analysts and executives found the outputs from big data projects inadequate for management analytics and decision support (Bendre & Thool, 2016). From 2006 to 2016, an estimated average of 50% of these projects failed to deliver acceptable and actionable outputs to business users (Gartner Inc, 2016). Also, the percentage of failed data analytics projects continued to rise with the exponential growth of data within organizations (Khan et al., 2014). The general management problem was that the outputs did not reconcile the intuitive cognitive model of the problem situation of business analysts and executives and the accustomed available data (Zicari et al., 2016). In many cases, these outputs were incomplete, difficult to understand, and difficult to translate into management actions because of their black-box nature (Günther, Mehrizi, Huysman, & Feldberg, 2017). The outputs were also disconnected from available data and from dominant cognitive conceptualization of the management problems and solutions by analysts and executives (Flath & Stein, 2018; Ransbotham, Kiron, & Prentice, 2017). The specific management problem was the inappropriate representation of information by big data projects for management analytics and decision support (Sivarajah, Kamal, Irani, & Weerakkody, 2017; Storey & Song, 2017).

Purpose of the Study

The purpose of this quantitative descriptive study was to examine the data model of a typical data analytics project in a big data environment for opportunities to improve the representation of information. I identified the data model as the primary focus of the study because the expression of information in the data model was known to improve understanding and application of the data (Burch, 2018). I adopted nonexperimental design-based research (DBR) to study the data model for artifacts that would improve the expression of the underlying management situations.

The research questions of this study focused on extracting expressions in the available data to improve the discovery, identification, specification, and resolution of management problems. Using 15-terabyte secondary data sets from U. S.-based multinational medical product distribution company on orders, payments, products, customers, sales channels, and marketing activities, I applied ontology learning to identify operational concepts within the business domain. I used data engineering to connect the concepts to available data through direct transformations, and analytic formulations to abstract functional forms from the available data. This approach ensured that any resulting analytic insights and solutions maintained the connection to the available data.

I assessed the performance of the data model artifacts on empirical measures of analytic importance such as information gain, intelligence density, decision yield, cognitive gain, empirical lift, Bayesian yield, the weight of evidence, and strength of association measures, as necessary. The results of this study could increase (a) the acceptance of big data analytics outputs by business analysts and executives, (b) the return on investment for big data analytics projects, and (c) the overall efficiency of data-driven management analytics and decision support. The social change implication was an increase in management engagement in social programs to sustain good corporate citizenship within stakeholder communities, including sponsorship of community events and social programs.

Research Questions

The research questions focused on finding artifacts within enterprise data to model key business scenarios for management problem-solving as follows:

Research Question 1: Can data model extensions improve the discovery of management scenarios from big data?

Research Question 2: Can data model extensions improve insights about the management scenarios?

Research Question 3: Can data model extensions express the complex constraints and rules needed to compose the acceptable and actionable solutions for analysts and executives?

Research Question 1

I relied on the relational model as the primary approach to modeling enterprise data. This modeling approach and subsequent enhancements solved significant problems in the use of databases to deliver information systems. The initial relational data model proposal unified data representation and addressed issues of data integrity. The proposal also added enhancements, for example, the relationship and the data catalog (or dictionary extensions) to improve the capture of the meaning of data and the use of the database for analysis (Werro, 2015). However, the capture of meaning was limited to low-order predicate logic, based on the quantities of attributes. Advanced analytics and decision support required higher-order logic, ontological argument assertions, and association reification to address complex analytic needs of management (Fried, Jansen, Hahn-Powell, Surdeanu, & Clark, 2015). The premise of this research question was that the manifestation of this higher-order logic, ontological argument assertions, and association reification at the data level had the potential to improve the analytics and decision support for management problem-solving.

Research Question 2

The challenge of representing business insights and solutions derived from big data was the consequence of the increase in complexity of enterprise business processes which manifested in applications, systems, and data environments. Addressing this complexity in the use of databases for analysis led to data warehousing and business intelligence applications. Data warehouses consolidated the data into single logical or physical repositories, while business intelligence applications automated the exploration of the data. From these systems and other sources, the creation of specialized datasets for advanced analysis, for example, statistical analysis, mathematical programming, system dynamics modeling, data mining, symbolic data analysis, functional data analysis, deep learning, to name a few, became a necessity. This practice resulted in analytic silos which constrained general expression of the enterprise within analytic solutions. The need to segment analytic processing arose when the business analysis was limited to simple aggregations in the presence of lots of data. The need also arose due to inadequate computational power for all attributes and instances of the data in analytic processing. Fortunately, these situations have changed in the modern enterprise, so high-dimensionality analysis can be taken advantage of in creating insights for management analytics and decision support (Liu, Liu, & Li, 2017). This new perspective allows information about randomness, uncertainty, and dynamism to be expressed within available data. It also allows the supporting data processing to adopt a distributed and parallel approach, co-opting the resources needed for the computational task at hand.

Research Question 3

The success of algorithms in analytic processing was an essential contribution of the last decade. In advanced analytic processing, extensions to properties of infinitely differentiable functions are used to specify real complex lines and planes (Veech, 2014). These extensions established analytic continuity and discontinuity (or breaks) in analytic scenarios. The extensions contributed to analytic solutions such as complex response surface topology, convoluted neural networks, restricted Boltzmann machines, and many others that are capable of expressing difficult conditions and constraints as chains, trees or forests of logic within the analytic space (Paganoni & Secchi, 2015). The implication was that these techniques could be incorporated into the formulation of analytic characteristics and associations to enhance data for management analytics and decision support.

To address these research questions, I investigated methods of analytic data representation. The investigation involved exploration of metadata, the underlying ontology of the available data, and the intuitive cognitive conceptualization of the management problem scenario. Since the business environment was not static, it was critical to integrate continuous adaptation of the representation and annotation of the characteristics and facts in the business domain. The implication was that contemporary approaches to analytic data-modeling, which were mostly static, needed innovation to capture changes in the attribution of concepts within the domain. The innovation was the application of analytic formulation techniques to derive additional data from the source data inputs while preserving the links between the input and derived data. Preserving the links improved the explainability of the insights generated, when the derived data were multi-valued, non-decomposable attributes, statistical moments, weighted scores (for example, propensity scores, rank scores, linear weights or variates), domain markers, patterns, profiles, perceptrons, coefficients, and so on. These derived data expressed concepts and constructs not directly captured by the available data to broaden the scope of the data for management problem-solving. Addressing the complexity of the derived data in the data model was critical. I used partitioning, classification, segmentation, grouping, and so on, to control the extant complexity under consideration, much the same way as randomization and blocking during experimentation.

Theoretical Foundation

In this study, I integrated theories of information science and theories in applied management and decision science. The key theories from information sciences were relational, dimension, and information theories. I used these theories from Information sciences to extend the theories from applied management and decision sciences in the design of the data model for the management analytics and decision problem representation. Specifically, information entropy and high-dimensional utility theories were critical in the deconstruction of data for management problem-solving. A brief discussion of these theories follows.

The relational theory provided the grounding for representing data as relations and specializing these relations as facts and dimensions in the multidimensional data model for analytic processing (Gosain & Singh, 2015). The multidimensional data model fact relation types were the numerical attributes and dimension relations were categorical attributes or derivations thereof which lacked formalized analytic space. With large and complex business scenarios, classical multidimensional designs lost flexibility due to high dimensionality and complex interdependencies (Al-Aqrabi, Liu, Hill, & Antonopoulos, 2015).

Dimension theory addresses complex attribution and interdependencies through the synthesis of the invariant properties required to specify the metric or vector space expressed by available data (Rasetti & Merelli, 2016). The theory guided quantitative expression of the dimensionality of the abstract space (Shen, Davis, Lin, & Nachtsheim, 2013). Its application resulted in the projection of classical multidimensional space into a metric space for analytical processing. The techniques depended on the assumptions of the nature of the space under consideration as follows. Programmatic methods (for example, linear, stochastic, integer programming; time series) mapped well-defined input-output spaces. Statistical (for example, analysis of variances, regression) and probability (for example, bayesian, frequency) methods defined linear smooth metric spaces. Numerical methods (for example, neural networks, decision trees, evolutionary algorithms) defined nonlinear smooth metric spaces. Finally, algorithmic heuristics (for example, data mining, deep learning, artificial intelligence algorithms) applied to unknown metric spaces. However, the specification of the metric space required standard measurements, which was lacking in management (Diamantini, Potena, & Storti, 2013). Therefore, it was critical to use the available data to formulate the ontology to enhance the representation and interpretation of expressions of underlying subjects of interest, as proposed by information theory (Schutz, Neumayr, & Schrefl, 2013).

Information theory supports the recoding of available data to improve the representation of a subject (Budhathoki & Vreeken, 2017). This application of information theory abstracts available data into specific elements for the analytic requirements. The application of information theory to data analysis created a number of methods, including classical data analysis, semantic data analysis, symbolic data analysis, functional data analysis, topological data analysis, projection pursuit analysis, symbolic dynamics, complexity analysis to name a few. These methods contributed to data abstraction as follows.

Classical data analysis described the standard data table which contains raw information while semantic data analysis re-described the data using atomic and molecular predicate logic in specific and intended decision-support problem-solving scenarios (Kaytoue, Kuznetsov, Napoli, & Polaillon, 2011). Functional data analysis represented information as mathematical and logical functions of underlying elements. Symbolic dynamics captured multilevel, multiphase information for complex dynamic analysis and decision-support problem-solving, with a well-developed construct of symbolic extension which organized each level or phase of a subject into differentiated zero-dimensional arrays. (Downarowicz, Travisany, Montecino, & Maass, 2014). In complex analysis, analytic extensions are used to generalize the solution for infinitely differentiable functions and variables without setting the thresholds beyond which variables had no business analytic or decision significance and lost management problem-solving value. The theories of applied management and decision science established the significant threshold of analytic and decision value for management.

The integration of these analytic formulation constraints imposed by theories of applied management and decision science transformed available data into the ontology for management analytics and problem-solving. Rasch theory adds to this through the construction of the measurements (or mereology and metrology) for management tasks using latent variables (Bond & Fox, 2013; Sofroniou, 2011). Shafer-Dempster theory generalized the Bayesian belief by integrating uncertainty reasoning into evidence derived from available data (Beynon, 2011). Analytical hierarchical process theory proposed steps for aligning the order of the contributing factors and influences exerted by ontological and epistemological elements (Deng, 2017). The Blackwell theorem expanded the application of information filters to isolate signals that were most critical to decision making (Roy & Rao, 2017)

The organizational theory proposed that the factors and influences exerted by the business elements occurred in the transactions it conducted. The opportunity to control the behavior of organizations was in administering their transactions efficiently and effectively (Powell & DiMaggio, 2012). Organizational theories evolved through task specialization (or division of labor), behavioral, contingency, information processing, and computational organization propositions. Each of these propositions established the decision as the most critical cognitive activity of the organization. Therefore, the decision theory was a framework for problem identification, specification, and resolution. The role of data processing was central to decision theory formulations, which determined the prevailing operational decision theory as rational, cognitive, behavioral, naturalistic, garbage can, computational, or combinations thereof (Cegielski, Allison Jones-Farmer, Wu, & Hazen, 2012; Pourshahid, Richards, & Amyot, 2011).

Nature of the Study

In this study, I examined the data model of a typical data analytics project in a big data environment. I used design-based research (DBR) because of the focus on the design of artifacts to support the research (Chakrabarti & Blessing, 2014). The focus was on the design of data model for a typical data analytics project in a big data environment for management problem-solving. The DBR approach had gained popularity in design science disciplines like Information systems, Computer sciences, Engineering, Cybernetics, Artificial intelligence, and others (Cronholm & Göbel, 2015). The research approach focused on the scheme of the items within a subject under study to highlight relationships and the impact of changes in the scheme on the overall expression of the subject (Chakrabarti & Blessing, 2014). With DBR methodology, I was able to evaluate and compare designs of the situation under consideration (Cronholm & Göbel, 2015).

I used a nonexperimental descriptive format. This format supported the discussion of the methodology used in the progressive transformation of the data into the concepts of the management problem. I applied ontology learning, data engineering, and analytic formulation techniques to extend the data model. The ontology learning identified the concepts and the cognitive map of the business problem domain. Data engineering transformed the concepts to the available data. Analytic formulation fostered the discovery and quantification of the associations and dependencies embedded in the data. This approach ensured that representation of analytic insights and solutions retained the connections to the available data.

To illustrate the data-modeling approach, I used secondary data from a U. S.-based medical products manufacturer and distributor. This analysis scenario required an integrated corporate action sequence of six different management areas of responsibility within the enterprise: customer service, marketing, pricing, product development, sales, and distribution. The case illustration reflected a typical data analytics project situation in modern organizations where there were lots of data but no clarity on management problems or the strategies to resolve them.

Definitions

This section includes definitions of key terms used throughout this study.

Analytic extension: The result of the process of expanding or continuing complex function(s) or variable(s) into simpler function(s) or variables to derive solutions (Segura, & Sepulcre, 2015).

Bayesian yield: The degree to which the data model facilitates the generation and evaluation of alternatives, derived from the conditional entropy of Bayes (Deng et al., 2014).

Classical data attribute: An attribute defined by the values captured at the lowest level of granularity possible for item or individual of interest (Diday, 2012).

Classical multidimensional data model or data cube: A subject-based arrangement of measures by categorical attributes to support online analytic processing (OLAP) operations including slice, dice, roll-up, drill-down, and pivot (Kuznetsov & Kudryavtsev, 2009).

Classical dimension attributes: A set of categorical attributes organized in a hierarchy for the partitioning of measures during OLAP operations (Kuznetsov & Kudryavtsev, 2009).

Classical measure or fact attributes: A set of numerical attributes which are quantitative expressions of the subject(s) of interest (Kuznetsov & Kudryavtsev, 2009).

Cognitive gain: The degree to which data improved the understanding, reasoning, and inference within the domain of interest (Curşeu, Jensen, & Chappin, 2013).

Data model extension: an appendage of a data model used to express specific characteristics of underlying subjects to improve the depth of information representation, for example, relationship, semantic, temporal, spatial, graphic, provenance, and others (Smirnov & Kovalchuk, 2014).

Decision yield: The estimate of the likelihood of the use of the data in the resolution of the decision problem because of the added precision, consistency, simplicity, cost efficiency, and agility (Fish, 2012; Wu et al., 2012).

Empirical lift: The degree of expression of the critical empirical factors in the data model, derived from information entropy concept of Claude Shannon (Deng et al., 2014).

Enterprise data model: Rationalized integrated third normal form data model of application and systems used to capture activities of the enterprise (Metz, 2014).

Enterprise data warehouse: Physical implementation of an enterprise data model in as a database management system for analytic uses (Metz, 2014).

Intelligence density: The ratio of conceptually recognizable attributes to a total number of data elements in the model (Bai, White, & Sundaram, 2011).

Symbolic data attribute: An attribute defined by values transformed from classical data to express the characteristic of an attribute for specific analytic intentions (Diday, 2012).

Symbolic dimension attributes: A set of attributes that form an axis of analysis used to qualify a subject of interest in specific terms for specific analytic objectives (Noirhomme-Fraiture & Brito, 2011).

Symbolic measures or facts attributes: Measures of a domain of interest used to express numerical characteristics of a domain for specific analytic objectives (Noirhomme-Fraiture & Brito, 2011).

Symbolic extension: Specialized encoding of attributes that uniquely represents the distinct state of existence of a subject of interest (Downarowicz et al., 2014).

Symbolic primitives: Functions automatically generated by data mining algorithms, for example, symbolic regression, classification or time series, which include fit functions, formulae, control commands, and so on, used in expressing the mathematical relationship between attributes (Zelinka, Davendra, Senkerik, Kasek, & Oplatkova, 2011).

Assumptions

Assumptions are conditions that a researcher holds as true with no demonstrable proof. The first assumption in this study was that the available data for the data analytics project in the big data environment were comprehensive and reflected the real world of the enterprise and its management decision problems. The complexity of the enterprise reflected its management problems, such that data model would offer the analysts and decision makers the ability to establish the importance of operational concepts within the management domain. This use of data-modeling preserved the lineage between the raw data input and enhanced data generated for problem-solving (Caron, 2013).

The second assumption was that the abstraction of data preserved the validity of the derived insights. The application of analytic formulation techniques to transform attributes emphasized associations and influences that were specific to the analysis situation under consideration. For example, analytic transformations such as class assignments, use of nth order statistical moments, frequency estimates, probability distribution functions, the coefficient of determination, correlation coefficient, and so on, expressed association between the indicator and response attributes under consideration. For example, joint probability estimates applied to situations where independence was verifiable. Conditional probabilities were the preferred method of quantifying association when independence was not verifiable.

The third assumption was that it was possible to extract insight from available data. This perspective was different from contemporary research studies, in which the data were from a controlled data generation process, an experiment. In this study, the focus was on the data model of the available data for data analytics project. The data combined information generated in the day-to-day operations of business integrated with information captured by other sources external to the organization. In this scenario, the data analyst had no control of the data generation process and was unable to manipulate the situation directly. Data analysis and modeling required inferring influences of attributes on one another to determine their consequences on management decisions and business programs.

Scope and Delimitations

This study focused on the enhancements to the data model of the available data for data analytics project in a big data environment. I did not construct a separate OLAP multidimensional model or create an alternative analytical model building outside the context of the data model. The former was the case with OLAP application system, while the latter was the case with statistical and mathematical programming, system dynamics, decision analytic processing, data mining, deep learning techniques, and algorithmic heuristics applications and systems. This focus on the data model of the available data for data analytics project in a big data environment was adopted because it offered the most elegant solution to analytics in management compared to the alternative approach of contrived subject-oriented OLAP models or constrained analytic algorithms. The OLAP data model limited associations between the subject areas of the enterprise. Analytic algorithms further limited data participation as required to control dimensionality of the input data for computational and methodological purposes.

This data model research defined data structures that advanced data-driven problem-solving in management. The tasks included decision discovery, scenario generation, prediction, inference, evaluation, and choice tasks. The approach focused on the abstraction of data elements from their raw form into structures, referred to as analytic extensions, and their application to the creation of solutions to support these tasks. This approach was different from the classical research approach in which empirical study drove data generation and analysis. Instead, this work aligned the objectives of the data model to structural, formal, and resolution expectations of the area of interest. Through the data model, established relationships between the data objects and analytic methods fulfilled the requirements of composing evidence and determining effects and influences on entities.

I did not provide the detailed treatment of any of the analytical techniques used, or their mathematical proofs because all the techniques were mainstream and did not require justification as part of this study. I focused on the applied aspects of these concepts and constructs, and their integration into the data model for management analysts and decision makers.

Limitations

I used secondary data to illustrate the enhancements of the data model of the available big data for analytics and decision-support in management. The source of data for the study was proprietary, so the data was de-identified as required by the data owners to protect the sources. The validation of the results, presented for the study, may not account for all the situations of anomalies in the data or with the analytic formulation techniques applied.

The interviews of business analysts and executives conducted established the conceptual scope and the prevailing hypothesis of the analytic problem. The evaluation of the resulting data model depended on management acceptance and actionability criteria established by the business analysts and executives through the interviews. I also used empirical measures of business and analytic significance, for example, information gain, Bayesian yield, intelligence density, and other similar measures. This business result orientation was different from traditional research where the statistical evaluation was preferred.

I drew from my experiences as a management analyst and researcher for Fortune 10, 50 and 500 companies and government agencies in the United States, seeking assistance with measurement, estimation, inference, and forecasting solutions to address transactional, operational, or strategic problems. In this role, I needed to advance capabilities in existing business intelligence and decision support systems to integrate inferential capabilities (programmatic, diagnostic, predictive, intelligence) into their decision-support environment and analytic processing workflow. Expectations included the creation of a measurement and metrics framework for shared performance management across diverse management domains. I designed and implemented application systems to support management effectiveness and efficiency. The driving force was to generate value from data assets and monetize them through the creation of value-added information solutions and services, for both internal and external use. Since these situations were specific subsets of data, analytics, and decision-support scenarios faced in business management, the perspectives driving this work were from these business settings. Therefore, the application of the study outside the business management context would be limited. Extension of the data model may not be necessary for data gathered through a controlled experiment or in situations where measurements of underlying elements are well established as in science and engineering contexts.

Significance of the Study

Significance to Theory

In this study, I addressed gaps in data-modeling of extensive secondary data for analytics and decision-support in management research. I advanced the use of ontology learning, data engineering, and analytic formulation techniques to transform available data from the classical data format in the form of scalar data types, through matrices and arrays, to functionals with specific ontological commitments. This approach closed the conceptual gap between analytical insights and cognitive concepts of domains of interest which, Beroggi (2010) argued, lagged behind advances in information and computer technology.

This study systematized an adaptive and progressive stepwise process of designing data models that connect meanings and signals embedded in the data. This framework generated derived data elements in specific analytic contexts. Many conventional approaches to decision modeling such as the analytical hierarchical process (AHP) of Saaty, generalized utility models, generalized risk models, and others which required heuristic approximations by experts. A meticulous transformation of existing data through ontology learning, data engineering and analytic formulation of the metric space of the subject of interest replaced the rates and weights approach of decision analysis methodologies.

The adaptive approach relaxed the controls and assumptions of traditional data-modeling and allowed relationships captured within data to drive the formulation of empirically rigorous, pragmatic data models that incorporated hierarchies not purely based on cardinality and linear functional dependencies. This integrative approach to data analytics in management maximized utilization of information and knowledge assets for decision processing. It aligned the processing of available data to the ontology of the subject under consideration specified by business analysts and executives.

Significance to Practice

The significance of this study was in the construction of a data model for data analytics projects in a big data environment for management problem-solving. The focus was the use of the data model to deconstruct complexity within available data and the management problem-solving situation. The goal of this data model research was to make transparent the discovery, evaluation, and resolution of management opportunities within the domain.

The deconstruction of complexity was crucial to the creation of a useful data model. Complexity is the state of lack of transparency between inputs (causes) and outputs (effects) of nondeterministic systems. Complexity manifests as the interaction of the inputs, the input output process, and the outputs themselves. Analytic deconstruction of complexity is critical to decision processing, through programmatic (if known inputs, outputs), diagnostic (if unknown input, known output), predictive (if known input, unknown output), and intelligent (if unknown input, unknown output) means. The construction of data models that accounted for the complexity of underlying data elements and their interactions improved analytics and decision-support in management. Making analysis more concrete and quantitative furthered Busemeyer and Townsend's (1993) decision field theory proposal. The decision field theory reflected some universal propositions for resolution of choice problems through systematic perceptions of the environment based on the information. It also included the utility of numeracy in the decision makers' coping to determine the need for decision-support by information systems and technology (Peters, 2012).

The final area of professional application of this study was in the creation of management support applications. The typical input to decision processes was a set of rates and weights compiled from experts and surveys. Decision analytic techniques such as hierarchical analytic processing, network analytic processing, info-gap decision processing, and many others, required in-depth knowledge of the domain of interest and the ability to reduce the knowledge into weights and rates of the decision problem. The weights and rates formulated analytically from the available data were more accurate than those defined by experts (Dezert & Tchamova, 2014).

Significance to Social Change

Business enterprises are essential instruments of societal prosperity because they provide employment, support the needs of the population by providing goods and services, and contribute to social efforts within many communities through donations and volunteerism. The influence of business enterprises have increased due to globalization, the information age, the convergence of business and politics (for example, the U.S. Supreme Court Citizens United decision), and the adoption of free-market economics around the world. These developments have added complexity to the working environment for executives and managers of enterprises. The modern business enterprise is not just expected to be solvent; it is also expected to contribute to the social aspects of the communities in which it is doing business by improving the quality of life of customers and community. The evidence needed to guide decisions and actions to maximize benefits of the business enterprise to all its stakeholders and the public at large was made possible by extending analytics to account for these considerations (Burns & Jindra, 2013). Broadening the characterization of the influences of the enterprise highlighted opportunities for management engagement in social issues. In many cases, the issues that impact the marketplace also influenced the performance of the organization. An example of social change that could be realized through the case illustration includes social programs to improve daily activities of patients and residents of health care institutions served by the company, especially incentives for sales representatives to volunteer their time at facilities they covered.

Summary and Transition

This chapter introduced the study of a model of the available data for data analytics project in a big data environment. The goal of the study was to search for data model extensions to address the issues of representation of insights about problems and solutions. This approach required organizing all available data into structures for that mapped the available data to the cognitive conceptualization of the management problem situation. This study is expected to contribute to reducing the high degree of failure in management analytics and decision-support, which accounted for an estimated $2.7 trillion in wasteful spending in 2016. The link between available data and solutions of management decision problems established a favorable relationship between investments in data asset development, quality of decision-making, and the business value achieved.

In Chapter 2, I review the literature on online analytic processing (multidimensional) data-modeling, the use of dimensional analytic techniques to achieve functional form expression of available data, issues of big data analytic scenarios, and challenges with computational and algorithmic analytic processing. In Chapter 3, I describe the research methodology, including a justification for a DBR methodology, the use of the descriptive, nonexperimental, quantitative format, the choice of the secondary data, and the data abstraction methodology that integrated ontology learning, data engineering, and analytic formulation techniques. The results of the study are in presented in Chapter 4, and the discussion of the results are presented in Chapter 5.

Chapter 2: Literature Review

Introduction

The purpose of this descriptive nonexperimental quantitative, DBR study was to examine the data model in a typical data analytics project scenario to address difficulties encountered with the acceptance of big data projects outputs. The literature on data models with data analytics revealed a very strong favorable association (Zohuri & Moghaddam, 2017). The conceptual data model was the primary tool for communicating the structure, content, and context of available data in organizations, yet big data analytics projects favored an approach that bypassed this critical artifact. The result was that business analysts and executives found the outputs from data analytics projects inadequate for management analytics and decision-support (Bendre & Thool, 2016).

In this chapter, I describe the literature search strategy on big data analytics process. I preview the state of analytic data-modeling, the role of functional form expression in data models, the problem of representing large scenarios for analytic processing, and the challenges with computational/algorithmic solutions in data analytics. I conclude with a discussion of the issue of the dissociation of the data from resulting analytic solutions which was my motivation for this data model approach to the challenges of big data project outputs (MacLeod & Nersessian, 2018).

Literature Search Strategy

The primary source of material for the study was Academic Search Complete, an EBSCO academic research database, available through Walden University. Searches included Google Scholar, Elvsier, Association for Computing Machinery (ACM), and the Institute of Electrical and Electronic Engineers (IEEE) digital databases.

The keywords used in the search included *empirical model building, analytic model building, multidimensional modeling, online analytic processing), exploratory model building, exploratory system dynamics modeling, statistical database, business intelligence, knowledge discovery from databases, data mining, data modeling, decision models, domain models, big data modeling, business intelligence, expert systems, symbolic data analysis, dimensional analysis, symbolic dynamics, artificial intelligence, reasoning systems, artificial intelligence modeling, deep learning modeling,* and *symbolic computation.* The search was conducted from the year of study until about 600 articles were retrieved and reviewed. Changes in popularity of these keywords over time complicated the task of limiting materials included in the study to publications in the last five years, as required by Walden dissertation guidelines. Some of the most relevant materials cited publication dates as early as 1990, which indicates that the problem of making sense of data emerged with Information / Systems era of this decade. Despite the age of these materials, the concepts expressed aligned with contemporary usage and understanding.

Of the roughly 600 articles I retrieved and reviewed. I cited 259 articles in this document. Of these, 87% were peer-reviewed and published between 2013 and 2018 based on Walden library databases designations. Ten percent of these citations were either books or conference materials.

Theoretical Foundation

This study integrated theories in information science and applied management and decision sciences to extend the data model for management analytics and decision problem representation. Relevant theories in information science included relational, dimension, and information theories. The applied management and decision sciences theories were organization and decision theories. A brief discussion of these theories follows.

The relational theory provided the grounding for representing data as relations of attributes such that every record within them was an instance of occurrence or members of the relation. The relational theory also provided the constructs for specializing these relations as fact relations and dimension relations in the multidimensional data model for analytic processing (Gosain & Singh, 2015). The multidimensional data model facts relations were the numerical attributes and dimension relations were categorical attributes or derivations thereof, without any attempt to formalize the space defined. With large and complex business scenarios, the classical multidimensional designs resulted in problems of large dimension sizes and complex interdependencies (Al-Aqrabi, Liu, Hill, & Antonopoulos, 2015).

Dimension theory addresses complex attribution and interdependencies through the synthesis of the invariant properties required to specify the metric or vector space expressed by available data. The theory guided quantitative expression of the dimensionality of the abstract space (Shen, Davis, Lin, & Nachtsheim, 2013). Its application resulted in the projection of classical multidimensional space into a metric space for analytical processing. The techniques used depending on the assumptions of the nature of the space under consideration as follows. Programmatic methods (for example, linear, stochastic, integer programming; time series) projected well-defined input-output spaces. Statistical (for example, analysis of variances, regression) and probability (for example, Bayesian, Frequency) methods projected linear smooth metric spaces. Numerical methods (for example, neural networks, decision trees, evolutionary algorithms, etc.) applied to nonlinear smooth metric spaces. While algorithmic heuristics (for example, data mining, deep learning, artificial intelligence algorithms) came in useful in projecting unknown metric spaces. However, the specification of the metric space required standard measurements, which was lacking in the field of management (Diamantini, Potena, & Storti, 2013). Therefore, it was critical to use the available data and subsequent derivations to formulate the ontology for the representation and interpretation of expressions of underlying subjects of interest, using Information theory proposals (Schutz, Neumayr, & Schrefl, 2013).

Information theory supported the re-coding of available data to improve the representation of a subject. This application of information theory abstracted available data into elements specific for analytic requirements. Many methods of data analysis resulted from the application of information theory. Examples were classical data analysis, semantic data analysis, symbolic data analysis, functional data analysis, topological data analysis, projection pursuit analysis, symbolic dynamics, complexity analysis to name a few. Essentially, these were methods of data abstraction that can be integrated into the data analytics framework to drive extensions of the data model for insight generation discussed below.

Classical data analysis described the data available in the classical data table, while semantic data analysis extended the data analysis to the underlying atomic and molecular predicate logic (Nalepa, 2017). Symbolic data analysis further abstracted classical or semantic data for intended analysis and decision-support problem-solving (Kaytoue, Kuznetsov, Napoli, & Polaillon, 2011). Functional data analysis provided the framework for the representation of information as mathematical and logical functions of underlying elements. Symbolic dynamics provided the framework for representing multi-level, multi-phase information for complex dynamic analysis and decision-support problem-solving (Downarowicz, Travisany, Montecino, & Maass, 2014). Furthermore, SD has a well-developed construct of symbolic extension which organizes the data at each level or phase of a subject into a zero-dimensioned array for differentiation. In complex analysis, analytic extensions were used to generalize the solution for infinitely differentiable functions and variables. However, the boundaries of analytic or decision significance and management problem-solving value were not considerations of these methods. The theories of applied management and decision science established the significance and value threshold of analytic outputs in management.

The translation of available data into an ontology for managerial tasks required the application of theories of applied management and decision science. Applied theories like Rasch, Shafer-Dempster, analytical hierarchical process theories advanced the integration of analysis into management and decisions sciences. Rasch theory was useful in the construction of the measurements (or mereology and metrology) within the management domain using latent variables (Bond & Fox, 2013; Sofroniou, 2011). Shafer-Dempster theory generalized the Bayesian belief by integrating uncertainty reasoning into evidence derived from available data (Beynon, 2011). Analytical hierarchical process theory proposed steps for aligning the order of the contributing factors and influences exerted by ontological elements (Deng, 2017).

The organizational theory proposed that the factors and influences, exerted by the business elements, occurred in the transactions it conducted. The opportunity to control the behavior of organizations lay in administering these transactions efficiently and effectively (Powell & DiMaggio, 2012). For this reason, the organizational theories evolved through task specialization (or division of labor), behavioral, contingency, information processing, and computational organization propositions. Each of these propositions held as its central theme that the decision was the most critical cognitive activity of the organization. Decision theory provided the framework for problem identification, specification, and resolution. This placed analytic processing at the center of decision theory proposition. The degree of analytic processing was responsible for the prevailing operational decision theory as rational, cognitive, behavioral, naturalistic, garbage can, computational, or combinations thereof (Cegielski, Allison Jones-Farmer, Wu, & Hazen, 2012; Pourshahid, Richards, & Amyot, 2011).

The integration of these theories converged on the utility of analytic processing in the disambiguation of the business environment for analysts and executives. The essential contribution of analytic processing compared to other analytic techniques (i.e., reporting, modeling, algorithms, and computation) was complete automation of the data analytics process from input to the generation of actionable insights and recommendations for all levels of the enterprise. The requirement to integrate data and technology assets, i.e., database management systems, and computer application programs into seamless processing were critical. Equally important was ensuring the outputs of the analytic processing exercise was transparent in management decision making. The transparency of analytic processing remained the primary challenge of applied management and decision science practitioners and researchers, hence the primary motivation for this study.

Literature Review

As noted above, analytical techniques provided frameworks for systematizing analytic processing (Kwakkel at al., 2010). They helped determine the nature of associations between attributes in the data to answer business and research questions about underlying subjects (Chen, Chiang, & Storey, 2012). The structure, content, context, unit of analysis and granularity of the data dictated their breadth, depth, and application to management analytics and decision-support. In recent years, users have challenged the utility of analytical techniques in addressing complex business questions facing management (Gomes, 2014). The response to this challenge was online analytic processing (OLAP). OLAP has two aspects: the multidimensional data model and algebraic operations. The OLAP data model provided the framework for organizing data the multidimensional structure. The multidimensional structure is an n-relational structure or data cube. The OLAP algebraic operations specified exploration and navigation procedures for the data cube (for example, slice, dice, drill, pivot). Currently, OLAP remains state of the art in the analytic processing despite challenges of limited analytic capabilities. To gain perspective on solutions to the challenges and issues with OLAP, I review the literature on the synthesis of a logical representation of complex subjects and large business analytic scenarios that advances high dimensional analytic processing. I provide a discussion of multilevel ensemble formulation through algorithmic/computational analytic processing. I highlight the absence of data models to support these higher forms of analytic processing, which is the gap I am seeking to address with this study.

Online Analytic Processing

Edgar F. Codd was the central figure in data-modeling literature for proposing both relational and online analytic processing (OLAP) data-modeling techniques (Wade & Chamberlin, 2012). Relational data-modeling drove advances in database technology, including the principle of data definition and manipulation using declarative language such as the structured query language (SQL). The framework of the relational data model was the theory, algebra, and calculus of relations which were stable and closed. At the core was the representation of data items as related sets, to which rules of normalization were applied to ensure efficiency and accuracy of data capture and storage. The OLAP proposal generalized the relational data-modeling approach from few to large relational structures. The OLAP model was responsible for the rapid adoption of data-driven DSS of the last decade, including a change in the role of the data warehouse from a passive repository for static enterprise reporting to an active platform for dynamic real-time analytics and decision-support.

At the core of the OLAP proposal was the multidimensional data-modeling technique. This data-modeling technique organized numerical data as facts (or measures) and categorical data as dimensions to form a multidimensional array (Gosain & Singh, 2015). This scheme enabled sophisticated navigation of large data sets and high-performance data retrieval operations.

The original multidimensional data-modeling proposal by Codd was rather strict about the designation of data attributes as measures or dimensions, and about the relationship between fact and dimension relations. Intense research into multidimensional data-modeling led to revisions. Gosain & Singh (2015) presented the most comprehensive survey of such revisions, which identified 23 characteristics of the 16 most complete multidimensional models. Table 1 shows the characteristics of the revisions.

Table 1

Characteristics of OLAP Multidimensional Designs

Aspect	Characteristic	Rationale
General	Atomic and non-atomic measures	Capture of measures at whatever level of granularity available
	Derived measure	Deriving new measures from existing ones, as needed
	Derived dimension attributes	Deriving new dimension attributes, as needed
	Flexible additivity	Support for full additivity, semi-additivity, and non-additivity
	Non-hierarchical dimension	A single level dimension attribute
	Cross dimension attributes	Dimension attributes that reference multiple dimensions
	Degenerate facts	Measures that may not be accurate all the time
	Degenerate dimensions	Dimension attribute with no content except its primary key
	Sharing dimensions	Dimension shared by multiple fact relations
	Sharing dimension levels	Dimension level sharing by multiple fact relations
	Parallel hierarchies	Creation of more than one hierarchy in a dimension
	Different roles of dimensions	Dimensions that serve different roles depending on the context
Fact-dimension relationship	Incompleteness association	Allowing the occurrence of missing associations
	Non-strictness association	Dynamic associations
Fact-dimension relationship	Incompleteness association	Allowing the occurrence of missing associations
	Non-strictness association	Dynamic associations
Inter-dimension relationship	Generalization	Generalization/ Specialization relationship between levels of dimension
	Association	Functional dependencies between dimension attributes
	Fact constellation	More than one fact in a dimensional model
Implementation	Technique	Modeling technique include ad-hoc, E-R, UML
	Mathematical/analytical constructs	Inclusion of mathematical/analytic operations
	Transformation of hierarchy	Mapping for transforming hierarchies
	Guidelines	Availability of an implementation guideline

According to Gosain & Singh (2015), the state-of-the-art analytic data-modeling retained the basic n-dimensional schema of fact relations and corresponding dimension relations. Representation of fact and dimension as relations allowed fact elements with the same dimensional architecture to connect to dimension elements at the same group level. This representation created the classical multidimensional structure commonly referred to as snowflake schema design, providing significant flexibility over the earlier proposal, the star schema design (Sharma & Sood, 2013).

Each dimension of the multidimensional schema represents sets of categorical data elements with a partial order from top to bottom, such that one categorical data element is greater than another if the members of the former are subsumed by the latter. The topmost element of the dimension corresponds to the largest possible dimension element size because it logically subsumes all the other elements in the dimension. The partial order of the categories forms the hierarchy of the dimension. The hierarchy of the dimensions was the navigational paths or graphs. Essential characteristics of these paths or graphs are: (a) that they are acyclic paths or graphs which means no re-entry loop and (b) that their direction reflects the cardinality of the relationship between the sets of dimension elements based on their occurrence (Pedersen, 2013).

The practice was to apply Codd’s rules of normalization to the structuring of the dimension elements to create homogeneous dimension levels. This practice allowed multiple hierarchies for different navigation and aggregation paths on the data. It also allowed specialization of the relationship between dimension levels into six types: (a) covered relationship in which the lower dimension level subsumes all the elements of the higher dimension level; (b) onto relationship in which there is a one-to-one correspondence between the dimension levels, typically modeled implicitly within the relation defined for the dimension level; (c) non-covering relationship which implies the dimension level is in a path parallel to a considered dimension level, with skipped levels; (d) non-onto relationships which are the absence of a parallel relationship at the one-to-one cardinality; (e) self-into relationship in which there is a self-referential requirement at the one-to-one level of cardinality creating an implicit hierarchy in the dimension level; and (f) self-onto relationship, a situation where a self-reference returns an empty set, which was the condition of a fully normalized dimension design (Pedersen, 2013).

The nature of the dimension is also an essential consideration in modeling. A dimension can be universal or domain. Universal dimensions include time and location, which can be modeled on their own or used to qualify other dimensions, as is the case in the spatiotemporal data model. Domain dimensions are those that have a specific significance in the subject under consideration; for instance, in the business domain, examples of dimensions were Store, Product, Customer, and so on. It is also essential to determine whether the dimension is static or dynamic and, if dynamic, whether it has a cycle and whether the cycle is or is not stationary (Pedersen, 2013). Managing dynamism in the design of dimensions creates the concept of slowly changing dimensions, which have defined types as follows: (a) Type 0 – insert only; (b) Type 1 – update in place; (c)

Type 2 – dimension versioning; (d) Type 3 – use of dimension effective and expiry date; (e) Type 4 – use dimension change or history relation to capture changes; and (f) Type 6 (hybrid of 1, 2, 3) with the current value, old value, start date, end date and current status flag (Kimball & Ross, 2011; Leonard, Mitchell, Masson, Moss, & Ufford, 2014). The assumption was that rapidly changing dimensions should not exist, but they did. For example, the customer was a very popular dimension in the business domain model which grew with changes in essential characteristics. The characteristics of the customer were not part of the classical fact-dimension scheme of the multidimensional model. The model did not explicitly reflect the change in state of the customer related to its activities and did not establish a connection with related concepts like party, prospect, and so on (so-called polymorphism).

The modeling of the dimensions was critical as it defined the axis of analysis or navigation for the user and provided the analysis flow process the user could adapt to formulate explanations to situations of interest progressively. However, some problems emerged with this design of dimensions including (a) that the relationship between the dimensions was primary key-foreign key reference; (b) that the dimensions are independent of each other; (c) within each dimension the different dimension hierarchies partition the dimension space equally or carry the same weight in terms of impact; (d) at each dimension level the effect of the dimension values were equally weighted; and (e) when there were elements in the dimensions that had a numerical value, they should be treated as categorical attributes during analytic processing (Caron, 2013).

The measure or facts of the multidimensional data model are typically the numerical attributes in the available data set. Fact or measures are assumed to be the numerical translation of the results of the interaction of the dimensions at the appropriate levels of details (Schutz, Neumayr, & Schrefl, 2013). For example, sales facts or measures such as sale amount, sale quantity, sale price, or sale discounts are a numerical representation of the interaction of customer and product dimensions within the business domain.

Different approaches were used to derive the facts or measures. One approach is the use of the concept of key performance indicators (KPI), which identifies measures that were significant contributors to the performance of the domain of interest (Diamantini et al., 2013). Another approach is the concept of the balanced scorecard (BSC), proposed by Norton and Kaplan, as a measure of organizational growth and learning that integrates operational and financial perspectives of organizations (Morard, Stancu, & Jeannette, 2012). KPIs and BSCs were part of visual displays commonly known as dashboards, which are constructed at different levels of an organization to provide a point-in-time (cross-sectional) or progression-over-time (longitudinal) view of performance. Current challenges with the definition of measures, related to the question of constructing an appropriate measurement model for items and activities that were not directly measurable. Morard et al. (2012) determined that the measurement model derived deductively from available data differed from the BSCs and KPIs expressed by management.

In a classical OLAP conceptual data model, there is no assumption of independence in the facts or measures, so they should not be combined. Also, contemporary designs advocate annotation of facts or measures such that there is information on whether they are natural or derived. When they are derived, it is also necessary to specify what operations (statistical, mathematical, or logical, for instance) were applied. Because the classical OLAP data model design constrains the implementation of hierarchies between measures, navigating the facts or measures in the same way as dimensions were not allowed. The design became an important issue when data gathered was at multiple levels of granularity and association between facts could not be derived through the navigation of the dimensions. Contemporary OLAP designs also assume that the value of the fact or measure is immutable and that the significance of the value of the fact or measure is stable over time (Diamantini et al., 2013). For this reason, there is no formal concept of changing facts and measures or adjustments to facts or measures to ensure that change in the significance of the value is in the classical multidimensional model.

Another important aspect of a multidimensional model is the relationships between fact and dimension relations. The contemporary approach advocates relating the fact to the dimension at the right level of granularity. The nature of this relationship is essential to the accurate functioning of OLAP operations, especially aggregation operations. An important reason for this is that aggregation operations navigate lattice expressed by the intersection of the dimension and fact elements in three different ways:

additively (sum), non-additively (average, min, max), and by counts (cardinality). Other considerations handled in contemporary design include (a) ranges by using value-equivalent tuples with annotations to specialize OLAP operations on the ranges as slice operations – time slice or space slice operations (Pedersen, 2013); (b) handling of uncertainty in the data value and relationships as probability or conditional probability using the probability operations applied, within and between fact relations and/or within and between dimension relations (Cuzzocrea, 2011; Moole, 2005); and (c) heuristic mapping of fuzzy attributes to actual dimensions and measures based on specified rules (Fasel, 2014).

The process of determining the attributes in an OLAP data model was not straightforward, mainly because the availability of data from multiple sources was overwhelming (Romero & Abello, 2011). Data modeling methods took on two main frameworks: demand-driven based on user requirements or supply-driven based on the available metadata and data. A hybrid which integrates both frameworks was gaining popularity (Romero & Abello, 2011).

According to Romero and Abello (2011), the demand-driven approach followed the classical Information System (IS) engineering process which depended on the end-users to provide input to inform the data-modeling process, while the supply-driven approach depended on the available metadata. However, in real-world scenarios, end users may not be aware of all the potential analysis opportunities and may overlook critical requirements that could dramatically improve decision-making. Also, metadata may not be comprehensive to allow the data modeler to infer all the attributes required for analysis of the data elements. Also, available data was the noisy and unsupervised discovery of features within the data was overwhelming and useless to decision-making, and sometimes downright misleading. Most recent proposals called for the use of ontologies to model data for analytics and decision-support (Padillo & Mazon, 2011).

Ontology, in the context of the data model, was the formalized conceptualization of a subject within the domain of interest through its available data. Ontology was, therefore, the most differentiated version of the “data about the data” or metadata (Jareevongpiboon & Janecek, 2013). Contemporary documentation of data was in the form of the data dictionary, which was limited to the name, description and the syntactic attributes of the data including data type, uniqueness, nullable, and so forth. Data glossaries expanded the number of semantic attributes that were captured to include examples, concepts, constructs, to name a few. Thesaurus and vocabularies extended the symbolic attributes further to include lateral relationships like types, similarity, dissimilarity. Taxonomies captured dimensional attributes, including hierarchical relationships within a set of concepts allowing partial ordering of these concepts. Ontology brought all these characteristics together to achieve an ultimate conceptualization of a subject of interest capturing all relevant concepts, constructs, constraints, controls, and constants (Martinez-Cruz, Blanco, & Vila, 2012).

According to Pardillo and Mazon (2011), there were ten shortcomings of the multidimensional model design, use of ontologies solved. Table 2 summarizes these shortcomings and related solutions.

Table 2

Ontology proposals for OLAP Data Models

No	Situation	Current state	Rationale	Solution
1	Multidimensional design requirements	Limited to classical attributes defined by the users	Needs for specific concepts that provide meaning to the analytics and decision-support situation	• Use of Foundation ontology with representation and interpretation mappings
2	Requirement for data source reconciliation	Direct or syntactic reconciliation only	Semantic reconciliation is more appropriate	• Dimension and measure discovery through matching and subsumption, • Selecting measure and dimensions for defining facts and classes, • Establishing bases for searching and pruning (grouping rules etc.), • Defining aggregation hierarchies through part-whole relationships, • Use of heuristics on structural aspects, for example, instance counts, • Inserting frequencies, cardinality, etc.

(table continues)

No	Situation	Current state	Rationale	Solution
3	Data model completeness	Syntactic completeness	Semantic completeness	• Use of published ontologies and taxonomies to ensure representation of metonymy/homonymy or hyponymy/hyponymy relationship and other issues of polysemy
4	Data types of measures	Measures of analysis declared as numeric without any sense of unit or scale	Measures have units and scales normalized	• Use of levels of measurement: nominal, ordinal, interval and ratio, which improved the implementation of aggregation semantics for the different measurement levels, for example, mode and chi-squared aggregation for nominal measures, mean, standard deviation, correlation, etc. aggregation for interval data
5	Summarizability	Additivity constraint	Semantic summarizability	• Classification of measures: additive, semi-additive and non-additive; • Classification of summary attributes as flow (rate), stock (level) or value per unit; • Classification of non-additive measures into ratios, percentages, measures of intensity, average, minimum, maximum, etc.
6	Conformed dimensions	Hub-and-Spoke vs. Bus approach to design	Design consistency that allows complete specification of subject of interest	• Use of annotation and links that map results to the input data used

(table continues)

No	Situation	Current state	Rationale	Solution
7	Traceability	Loss of traceability between the source and model semantics	Semantically traceable data models	• Integration of transformation logic into the data model
8	Reasoning support	OLAP algebra and calculus	Reasoning requires logic for proof	• Integrate logical propositions provided by ontology into the data model •
9	Visualization	Issues of visualization of high dimensional hierarchical data	Layered visualization with appropriate visual gallery	• Semantic annotation of measure and dimensions for visualization
10	Security	Ad-hoc		• Inferred from ontology about credentials, permissions, and rights

Table 2 refined the approach to the determination of the content of the multidimensional data model. The ontology approach emphasized the explicit specification of knowledge available about the domain, either from internal sources or public sources. The approach required inference of any domain-specific attribution not available in the data. Hoang, Jung, and Tran (2014) advocated the creation of this enterprise ontology, independent of the information systems development projects to ensure that there was a systematic approach to qualification and quantification of the elements relevant to knowledge of the domain of interest.

While ontologies captured comprehensive conceptualization of the domain of interest, it provided no guidance on their essential and relative influence on the events and activities of the domain of interest. It also did not provide a framework to reduce a complex domain or concept into its components for examination. The dimensional analysis technique provided such a framework by enabling functional form expression as discussed below.

Functional Form Expression

The primary reason multidimensional models was so useful in analytics and decision-support was their structural alignment to dimensional analysis and reasoning than contemporary relational models(Savinov, 2013). Dimensional analysis generalized linear algebra, reducing complex problems into simple forms for solutions (Shen et al., 2013). The principal use of dimensional analysis was to deduce from data the final form of quantities of dependent and independent attributes of the subject of interest devoid of scale or units, according to Buckingham's π-theorem. This dependence on normalized standard quantities for expressing relationships preserved the concept of similarity and prevented coincidence of equivalence and differences caused by measurement units and scales. Using the similarity principle, it was possible to formalize the problem mathematically and simplify the solution by reducing the space of the data matrix to achieve a better functional form for underlying relationships.

The dimensional analysis required the manipulation of three classical constructs: properties, quantities, and units to derive attributes whose units canceled out when multiplied or divided, such that their absolute significance was maintained despite the change in numerical magnitude (Bridgman's principle) (Shen et al., 2013). The formula that satisfied this principle of absolute significance of relative magnitude was the power law form expression:

$$Q = \alpha A^{a} B^{b} C^{c} \ldots \tag{1}$$

where

- Q is the derived attribute
- A, B, C. are numerical values of base quantities
- a, b, c are real numbers whose values distinguish one type of base quantity from another
- α invariant scale that guarantees similarity of Q and base quantities (similarity coefficient)

These derived power form attributes were the dimensions. A dimension of the first kind was from the base units of the numerical value of base quantities, and dimensions of a subsequent kind from dimensions of the first kind, and so on. In this context, the dimensions may not represent a tangible characteristic of the subject of interest. Each base quantity, by definition, was its dimension. The dimension was, therefore, a formulaic expression of how the value of the quantities transformed when the size of the base units changed. For example, the dimension of a base quantity, Q,

$$[Q]=W \tag{2}$$

Where

- [Q] represents a dimension of property Q
- W represents the concept of the measurement unit, in this case, the concept of width

If the width unit size, W, increases by a factor of f, the numerical value of Q will increase by a factor of f^{-1}. Also, the dimension of a dimension conferred the same information about the general form. A dimension, Q, defined by:

$$Q=\alpha L_1^{l_1} L_2^{l_2} \ldots M_1^{m_1} M_2^{m_2} \ldots t_1^{\tau_1} t_2^{\tau_2} \ldots \quad (3)$$

Where

- L_i, numerical values of certain lengths
- M_i, numerical value of mass
- t_i, values of certain times
- α, exponents of real numbers

If the length unit changes by a factor, *l*, mass unit changes by *m* and time unit changes by *t*, the value of Q changes to:

$$Q^1=n^{-1}Q \quad (4)$$

where

$$n=(n_L)^{\Sigma l_i}(n_m)^{\Sigma m_i}(n_t)^{\Sigma t_i}$$

Q transformed like the numerical value of the base quantities with a unit whose size was proportional to the sizes of the underlying units. When the numerical value did not change with its base unit value, then the dimension was considered stable or dimensionless.

In analytics and decision-support, one seeks functional relationships between numerical values of quantities that describe, estimate, infer, or forecast the situation of interest, devoid of coincidence of choice of units - dimension homogeneity. Dimensional homogeneity implied both sides of the quantitative expression should have the same dimension, and dimensionless, the quantities and the terms must be of the same dimension or dimensionless, and any arguments of any exponential, logarithm, trigonometric or other special functions that appear in the equation must be dimensionless. Dimensional analysis demanded formulation of equations to capture the functional relationships between sets of independent and dependent quantities expressed in equation form as follows.

$$Q_0 = f(Q_1, Q_2, \ldots, Q_n) \tag{5}$$

Where

Q_0 is the dependent quantity

$Q_1, Q_2, \ldots, Q_n$ are independent quantities

f is the conversion factor that confers similarity to the expression

The relationships expressed in (5) above was the result of laws or policies governing the occurrence of the quantities of the property of the subject of interest. This relationship should hold despite the sizes of the base units of the quantities included, per Bridgman’s principle. The system of units that defined the quantities determined its dimension along with exponents that were dimensionless numbers following from this definition. Assuming that

1. $Q_{1,}$ $Q_{2,}$…, Q_k were dimensionally independent subset of quantities, where none of the members had a dimension that expressed the dimensions of the remaining members
2. $Q_{k+1,}$ $Q_{k+2,\ldots}Q_n$ were the rest of remaining independent attributes expressed regarding the dimensions of the subset $Q_{1,}$ $Q_{2,}$…, Q_k
3. Q_0 remained the product of powers of $Q_{1,}$ $Q_{2,}$…, Q_k and $Q_{k+1,}$ $Q_{k+2,\ldots}Q_n$ to achieve dimensionally homogeneous expression
4. $k < n$

Then

$$\pi_i = \frac{Q_{k+i}}{Q_1^{N(k+i)1} Q_2^{N(k+i)2} \ldots Q_k^{N(k+i)k}} \tag{6}$$

where

i=1, 2, …, n-k were dimensionless form of the dependent variable Q_0

and,

$$\pi_0 = \frac{Q_0}{Q_1^{N_{01}} Q_2^{N_{02}} \ldots Q_k^{N_{0k}}} \tag{7}$$

where

1 .. k was the dimensionally independent form of the dependent variable, Q_0

Then,

$$\pi_0 = f(Q_1, Q_2, \ldots, Q_k; \pi_1, \pi_2, \ldots, \pi_{n-k}) \tag{8}$$

According to Bridgman's principle and following the Buckingham's π-theorem, the reduced form of the expression of the expression should be:

$$\pi_0 = f(\pi_1, \pi_2, \ldots, \pi_{n-k}) \tag{9}$$

This final form satisfied, the Buckingham's π theorem which stated that when a complete relationship between dimensional quantities was in the dimensionless form, the number of independent quantities that appear reduced from the original n to n-k where k was the maximum number of the original n that are dimensionally independent. This theorem facilitated the discovery of the dimensions of dependent attributes, but not the form of the dimension. The form had to be discovered deductively from both exploration of the properties and the values of the data set, guided by existing knowledge of the subject of interest, available data, theories, propositions, and experimentation (Shen et al., 2013).

Dimensional transformation of data in a pre-determined fashion ensured that the underlying relationships remained intact and enhanced as needed for the analysis under consideration (Shen et al., 2013). This analytic process eliminated coincidences of similarity that may occur. Dimensional independence conferred statistical and mathematical independence which made the analysis much more valuable and informative. The reduction in the number of attributes eliminated redundancies encountered with large data sets, (for example, redundant non-distinguishing dimension attributes and records; identification of dimensions with similar effects of interest). Also, dimensional transformation demanded numerical expression for dimensions, which is different from the concept of dimension in a classical multidimensional model.

The requirement of numerical expression of attributes can be problematic with non-numeric properties or attributes. Multivariate algebra, the grounding for multivariate statistics, solved this problem through the coding of attributes, using functions. Examples were enumeration, dummy coding (or identity coding), threshold-based coding, target-based coding, the weight of evidence coding, cluster coding, smoothed weight of evidence, etc. (Wickens, 2014, pp. 5-15). Other methods of categorical data transformations include Rasch model of measurement based on tabulation of expected frequencies and Shafer-Dempster model of evidence-based on the tabulation of the log-odds of probabilities (Bond & Fox, 2013, pp. 15 – 28; Cuzzolin, 2012). The typical dimensional analysis focused on extents of objects or subjects under consideration, as the generalization of their linear algebraic expression. Extending this concept from defined measurable objects or subjects to undefined abstract space covering the interaction of objects and subjects, required specification and integration of subspaces. The specification of large complex scenarios became the primary challenge of management analytics and decision-support.

Expression of Large and Complex Scenarios

Data warehouses and OLAP applications evolved as a response to growing complexity of information technology and data environments supporting business functions and management activities. A typical enterprise data warehouse was made up of many records with a large number of attributes. A simple mathematical estimate of candidate models in an enterprise model design space can be calculated using the formula, L^A , where A is the number of attribute and L is the average number of levels (or values) of the attributes. For a simple modeling problem with one hundred attributes at two levels each, the number of solutions would be about 10^{30} (Michalewicz, Schmidt, Michalewicz, & Chiriac, 2011, p. 25). Technically, the number of empirical model candidates within a model design space was huge, but there were a limited number of these models that would satisfy the design requirements of the analysis exercise.

Therefore, the characterization of the enterprise model design space required a careful examination of the underlying analytics opportunities. Model spaces were the factors and functions that drove the transactions to express states of existence (of entities, domains, systems) responsible for the outcome variations, which made up the utility and preference relations for the management decision maker (Hsu, Ito, Schweikert, Matsuda, & Shimojo, 2011). Considering the potentially large number of solutions within an enterprise model design space and the constraints imposed by subject based multidimensional modeling approaches, the consensus in the literature converged on multi-tier ensemble analytical architecture. Hsu et al. (2011) presented three-tier architecture paradigm based on computational informatics perspectives to include: (1) structure layer models for structural components of the domain, (2) function layer models for functional components of the domain, and (3) application layer models for application components of the domain. The application of this architectural approach to the analysis of the brain system resulted in a computational fusion method for the assessment of gender variation in facial attractiveness is shown in Figure 1 below.

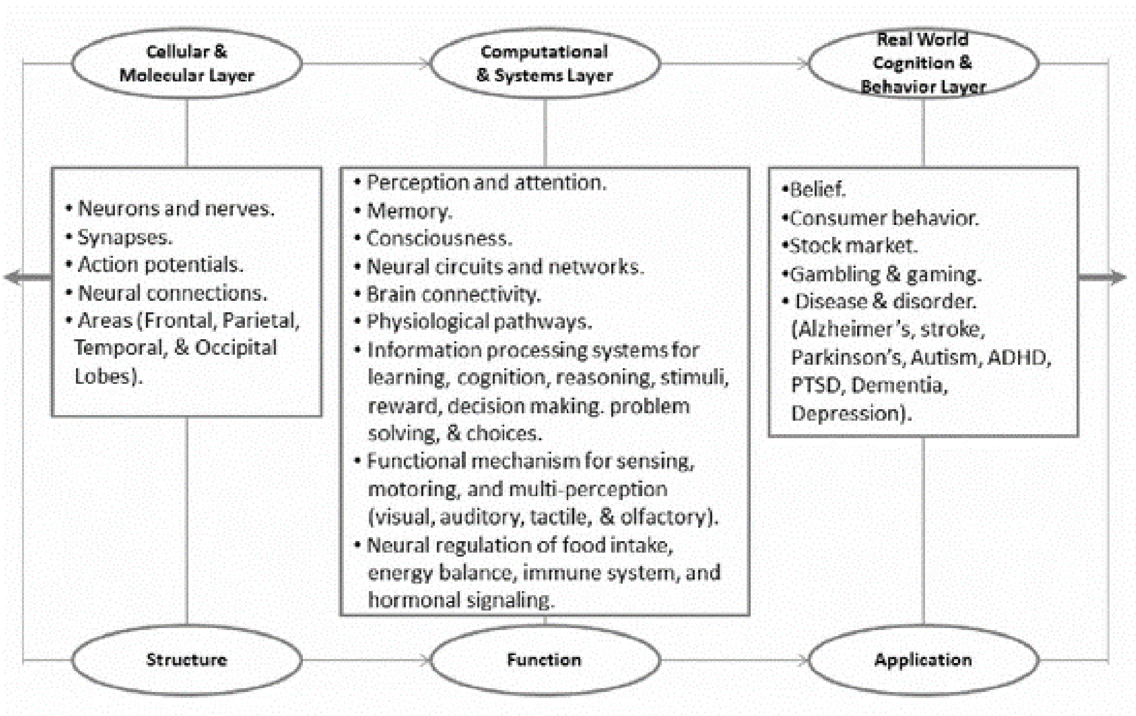

Figure 1. Multilevel modeling applied to the brain system. From "Combinatorial Fusion Analysis in Brain Informatics: Gender Variation in Facial Attractiveness Judgment," by D. F. Hsu, T. Ito, C. Schweikert, T. Matsuda & S. Shimojo, 2011, *Active media technology,* p.9. Copyright 2011 by Springer-Verlag Berlin Heidelberg. Adapted with permission of the author.

Beroggi (2010, p. 12) discussed a three-step analytical formulation process: structural, formal, and resolution steps, across three common modeling paradigms: data (observation), domain (or system), and decision to create a 3X3 analytical model architecture matrix. The first step of the analytical formulation was the structural level, a graphical portrayal of the relations and dependencies which may be causal (non-symmetric), correlational (symmetric), conditional (probabilistic) or informational (definitional) allowing the subject of interest to reflect the underlying data structures. The second level was the formal level where the relations and dependencies transformed into attributes to calibrate or define the subject of interest. The third level was the resolution level in which procedures were applied to generate solutions about the subject of interest. At each level of the analytic formulation, the level of analysis determined the format, content, and context of expressed relations. Appendices F and G were compilations of the details of the approaches. Further, Hendry (2009, pp. 16-19) identified four practical knowledge levels: measurement, estimation, modeling and forecasting levels, based on the nature of probability distribution and data generation processes. Table 3 below combined these proposals on representing complex subjects of interest from its available data.

Table 3

Multi-Level Model Ensemble for Complex Subjects

Knowledge Level: Formulation Steps	Dominant approach	Modeling paradigm		
		Observation / Record	Domain / System	Decision / Application
Level A: Structural	Measurement, probability, and statistical theory	Underlying transactions organized to expose effect and coefficients	Underlying processes that connect the transactions, establishing input and output attributes	Underlying mechanisms to which the knowledge of transactions and processes can be applied
Level B: Formal	Estimation	Measurement and latent models that underlie the coefficients of interaction captured in the transaction	Functions and equations for each of the processes within the domain and system	Estimates (heuristic, probabilistic, statistical, mathematical) that underlie the mechanisms
Level C: Inferential	Reduction, Likelihood, Expectation maximization	Given specific conditions, determine what inferences made about the data	Starting with initial values, determine new values for the levels, flows, and converters	Evaluate the formal model and determine elimination, integration and recursive approach
Level D: Forecasting	Numerical optimization, simulation, heuristics	Extend conditions beyond representation in the data to include potential innovations and emergence	Extend values within the model to new futures for levels, flows, and converters	Evaluate the new futures for applicability and implementation

In Table 3, an integration of the dominant approaches and modeling paradigms in the literature cuts across disciplines from statistics to cybernetics. This matrix charts the paths for data from the left upper corner through insights to foresight on the right lower corner of the matrix. The modeling paradigms had modeling standards or patterns. The data model paradigm represented the entity objects as the primary subject of interest, while the domain (or system) models represented a collection of entity objects that interact to achieve congruent outcomes. The decision (or application) model paradigm represented the expression of relations to achieve alternative futures of existence for specific goals and objectives.

The discussion so far established the multi-level design architecture as the effective analytic representation in the presence of complexity. This architecture was achievable through progressive reduction of the available data, and the exploration of the results for candidate representations of the subject of interest. This approach included a feedback loop for incremental updating of the model to improve its performance over time. Analysis of the data, domain, and decision situations in complex areas of endeavor required incremental construction and manipulation of models, such that one set of models replaced another set of models. The ability to integrate and compare multiple models in a domain of interest was critical to this type of empirical model building. Analytic model building in complex domains demanded the construction of an ensemble of models of various forms and specifications, for comparative analysis and integration (Windschitl, Thompson, & Braaten, 2007). An epistemic comparison of the classical approach and this alternate approach, multi-level empirical model building, was synthesized in Table 4 below.

Table 4

Classical versus Layered Empirical Modeling

Model characteristic	Classical	Layered
Goal	Find patterns in natural phenomena	Find patterns and defensible explanations for the way the natural world works
Construction	Hypothesis often stated as predictions about isolated aspects of the phenomena	Adaptation often represented as interaction of the phenomena within the domain of interest
Validation and Revisability	Evaluation of predictions result in acceptance or rejection of the prediction, with limited opportunity to revise them	Evaluation of hypothesis occurs in the context of design, revisions to the design allow further validation of the predictions
Explainability	In the form of conclusions, summarizing the trends and patterns in the data	Uses patterns in the data to build evidence for explaining the design. The design serves as a tool for explaining the phenomena
Extensibility	Insights on how the phenomena could beyond the scope of the data are not possible	Provides insights into the phenomena beyond the scope of the data, through analysis of alternative designs
Generative	New hypothesis or theories are end-product	Alternative designs are the product

A necessary implication of the multi-level empirical model development in the large enterprise space was the need to automate the generation and comparison of a large number of models as part of the analytic processing. This created demand for computational and algorithmic analytic processing techniques discussed below.

Computational/Algorithmic Analytic Processing

An emerging approach to handling large analysis scenarios was the adoption of a computational or algorithmic approach. With this approach, computer scientists or algorithm designers look at exploring large complex datasets as a computation or algorithmic problem. The objective was the discovery of the models for the data using statistical, machine learning (numerical heuristics), summarization (likelihood, similarity), and feature extraction (frequent itemsets, similar items) methods. A large number of algorithms had been developed to support computational extraction of models from data. The most common algorithms being C4.5 and higher, k-means, support vector machines, apriori algorithm, Expectation-Maximization algorithm, PageRank, AdaBoost, naïve Bayes, Classification And Regression Trees (CART) (Wu et al., 2008). Other algorithms included autometrics (Hendry & Mizon, 2011), neural networks, genetic programming, grammatical evolution, multi-expression programming, evolutionary algorithms, self- organizing migrating algorithm, differential evolution, simulated annealing, analytical programming, Pareto genetic programming (Zelinka et al., 2011). These algorithms integrated attribute manipulation, numerical simulation, numerical optimization, bootstrapping, and other techniques in the evaluation of the data to discover models that can be used to measure, estimate, infer or forecast a subject of interest.

A class of algorithms referred to as symbolic algorithms had become particularly popular because they modify the underlying data model to improve the efficiency of discovery of models within the data. Symbolic programming discovered symbolic data from the available data (Syme, Granicz, & Cisternino, 2012). This approach constructed symbolic structures from available data and used this structure in the discovery of models. The structures in Figure 2 represent the result of implementing symbolic regression or classification to produce fit functions, formulae, control commands examples from Zelinka et al. (2011).

$$x\left(K_1 + \frac{(x^2K_3)}{K_4\,(K_5 + K_6)}\right) * (-1 + K_2 + 2x\,(-x - K_7)) \tag{1}$$

$$\sqrt{t}\left(\frac{1}{\log(t)}\right)^{\sec^{-1}(1.28)} \log^{\sec^{-1}(1.28)}(\sinh(\sec(\cos(1)))) \tag{2}$$

$$\begin{aligned}&Nor[(Nand[Nand[B||B, B\&\&A], B])\&\&C\&\&A\&\&B,\\&Nor[(!C\&\&B\&\&A||!A\&\&C\&\&B||!C\&\&!B\&\&!A)\&\&\\&(!C\&\&B\&\&A||!A\&\&C\&\&B||!C\&\&!B\&\&!A)||\\&A\&\&(!C\&\&B\&\&A||!A\&\&C\&\&B||!C\&\&!B\&\&!A),\\&(C||!C\&\&B\&\&A||!A\&\&C\&\&B||!C\&\&!B\&\&!A)\&\&A]]\end{aligned} \tag{3}$$

$$\begin{aligned}&\text{Prog2[Prog3[Move, Right, IfFoodAhead[Left, Right]],}\\&\text{IfFoodAhead[IfFoodAhead[Left, Right], Prog2[IfFoodAhead[}\\&\text{IfFoodAhead[IfFoodAhead[Left, Right], Right], Right],}\\&\text{IfFoodAhead[Prog2[Move, Move], Right]]]]}\end{aligned} \tag{4}$$

Figure 2. Examples of structures derived from symbolic algorithms.

This approach to analytic processing had been equally enabled and challenged by data size and dimension explosion. Algorithms implement a static set of procedures with established decision thresholds. The black box nature of their implementation required that the user acquire a significant mathematical skill. Also, in many practical situations, the nature of the available data was overwhelming, noisy, localized, inconsistent, and incomplete. As such, the extraction of valuable insight through generalization of every numerical relationship had limits. Input from expert to differentiate the raw input became necessary to simplify the computational complexity of the algorithms. However, input from expert had limits and introduced bias into the generation of useful generalizations of relations within the subjects of interest. A much more comprehensive approach was direct manipulation of the data model to discover the attributes that support analytic continuation beyond the classical multidimensional space into formalized metric spaces and subspaces.

Summary and Conclusions

In this chapter, I discussed different analytic processing methods. OLAP was the only method with a well formalized for data-modeling process. I highlighted the issues with OLAP data models, including, heterogeneous and irregular dimensions, handling of different types of aggregation operations, handling time and uncertainty, symmetrical treatment for dimension and fact elements, and support for different levels of granularity in the facts and dimensions. Also, contemporary analysis scenarios handled with OLAP were small and narrowly defined data cubes. As the size of data and complexity of the analytic scenarios increased, specialized designs were introduced to improve their analytics and decision-support capabilities, for example, planning, prediction, inference, probabilistic, Gaussian, OLAM, programmatic OLAP (prolap), and many other OLAP formats. Another critical issue was the determination of the proper attribution of fact and dimension relations of the data cubes to ensure alignment with cognitive models of underlying subjects. The classical multidimensional models and data cubes did not address secondary issues associated with independence and parsimony of attributes in the data model design. The result was sparse data cubes whose application was limited to descriptive analytics.

Improving data models for advanced analytics and decision-support required three key changes to the data-modeling process. It was important to learn the ontology of the subject of interest from the available data, not the other way around as it has been the case in contemporary ontology engineering. An important part of ensuring that the data model has the right content for analytic processing was leveraging data engineering and analytic formulation techniques to evolve the data to the right unit and level for analytic processing, such that similarity principle critical to data analysis and decision-support problem-solving would be applicable. Finally, refining the normed metric space through rigorous specification of proper functional forms of relationships in the data ensured reduction of the metric space to orthogonal expressions of underlying data to limit interdependence of indicator or characteristic attributes.

The solution opportunities in the use of ontologies included representation and interpretation mappings, and the use of measures and encoded indicator attributes in fact relations. Ontologies would also establish the base for searching and pruning of categorical and measure attributes in fact relations. Also, ontologies would use heuristics to express structural aspects of the data. Other opportunities with ontologies were the use of summary attributes and transformation logic and many others. Other solution options were the use of data engineering and analytic formulation techniques, especially semantic data analysis, symbolic data analysis and dimensional data analysis to establish the ontology of the subject of interest within the analytic data models. Additionally, using constructs from complexity analysis and symbolic dynamics, the analytic data model should be transformed from a static artifact to an active one by expressing dynamism, uncertainty, and fuzziness within the data model.

The mathematical constructs of Bridgman's principle, Buckingham's π-theorem, and Blackwell theorem were helpful in the construction of extensions to express the complex features embedded in the data. This integration of critical concepts from dimension and complexity analysis into the relational data model to derive solutions in management analytics and decision-support provided the basis for the analytic extensions on relations adopted by this study. In the next chapter, I describe the research methodology adopted to study this subject. I argue that data-modeling should be considered a critical step in research involving secondary data, especially, when the data exceed sizes typically considered adequate for normal distribution assumptions. A data analytics project without a data model that accurately reflected the concepts and constructs of the domain of interest cannot be expected to produce outputs that are explainable or actionable by business analysts and executives.

Chapter 3: Research Method

Introduction

The purpose of this quantitative descriptive study was to examine the data model of a typical data analytics project in a big data environment for opportunities to improve the representation of information. I identified the data model as the primary focus of the study because the expression of information in data models is known to improve understanding and utilization of the data (Burch, 2018). I adopted nonexperimental DBR to study the available data and to map it to the cognitive models of the underlying management situation. This alignment of analytic outputs to cognitive models provided the basis for the acceptance and actionability of data analytics outputs (Okoli & Watt, 2018).

In this chapter, I discuss the details of the research method. This study emphasized design theories and concepts for extraction or extrapolation of knowledge from the available data. Therefore, discussion of the specific issues of the industry of the data source in this research was not relevant. Because the focus of the study was constructing the data model that captured underlying concepts for management decision problem-solving, requirements of population characteristics, sampling, and sampling procedures were not relevant. The discussion of threats to validity focused on construct validity since issues of external and internal validity or ethical considerations were not significant to the methodology.

Research Design and Rationale

I used a quantitative, nonexperimental, descriptive design format for this study to examine the data of the typical data analytics projects to find data model extensions that would improve the discovery, identification, specification, and resolution of management decision problems. The research questions guided the study:

Research Question 1: Can data model extensions improve the discovery of management scenarios from big data?

Research Question 2: Can data model extensions improve insights about the management scenarios?

Research Question 3: Can data model extensions express the complex constraints and rules needed to compose the acceptable and actionable solutions for analysts and executives?

The demonstration of the improvement in data analytics projects on the above questions would indicate an affirmative response to them. For this demonstration, I used secondary data from a typical enterprise data analytics project. In such a project scenario, the specific the needs of the users are vague or non-existent. Additionally, the current data analytics processes that occur in business intelligence, data mining, knowledge discovery from databases, deep learning, and artificial intelligence tended to create incomplete, nuisance and challenging insights and solutions. The secondary data used for the demonstration was made up of 140 datasets from five different sources with about 1000 data attributes, 1.75 billion rows totaling about 15 terabytes.

The selection of a design research methodology informed the focus of this study, which was building and evaluating data model designs to improve management analytics and decision making. It is typical for the type of problem, the research objective, and the expectations of the researcher to dictate research methodology and approach (Cooper, Hedges, & Valentine, 2009). The type of problem may be normative, descriptive, and prescriptive. The objective of research may be to test a proposition or hypothesis, explain an occurrence, or qualify the impact of structure or function. The expectation of the researcher may be to validate a theory, advance an acceptable explanation, and guide practice (Bakker & Van Eerde, 2012; Hussain, Elyas, & Nasseef, 2013; Leech & Dellinger, 2012; Reimann, 2011; Turner, 2010). The alignment of these factors was critical in the selection of research methodology and design of the study.

Cooper et al. (2009) argued that the impact of these factors were reflected in the classes of research methodology in the literature. There were three key classes of research methods as follows. The first class was inquiry driven by theoretical formulations (theory-based research), the contemporary scientific research approach. The second class was inquiry driven by epistemic needs (case-based research) which was made popular by social and behavioral sciences. The third class was inquiry driven by the need to improve design (DBR) which was made popular by design and engineering sciences. These classes of inquiry also determined the degree of interaction of the researcher with the subjects under investigation. In theory-based research, a high degree of direct interactivity and control is assumed to ensure that the measurements of the inputs capture the characteristics (or attributes) of the subject, unencumbered by nuances of the surrounding or the researcher. A high degree of indirect interactivity and control is necessary with case-based research because the inquiry focuses on understanding underlying epistemology (for example, phenomenology, ethnography, case study.). For inquiries driven by the need to improve design (so-called DBR), any interactivity or control biases the context (Kuechler & Vaishnavi, 2012). This study's research approach sought to capture the natural architecture and to determine changes in design to pursue to improve knowledge expressed by available data for management problem-solving. The researcher interacted with the scheme or configuration of elements (the design) in the domain of interest to understand the problem and propose solutions to them as needed. This research approach required the definition, construction, and test of the candidate designs of the subject of interest to achieve outcomes that did not result naturally (Kuechler & Vaishnavi, 2012).

The DBR approach was well established in Information Science and Engineering research (Bakker & Van Eerde, 2012; Kuechler & Vaishnavi, 2012), and was considered an important area of applied research for developing information, technology, and engineering solutions using existing knowledge and artifacts (Blessing & Chakrabarti, 2009). Chakrabarti (2011) argued that this form of research allow the researcher to develop new methods, constructs, and artifacts to simplify the application of knowledge and engineering rigor for consistent results. This approach differed from the contemporary research methodology of scientific evidence, which focused on the theoretical development and statistical hypothesis testing. Fortunately, work in the last 50 years has increased the acceptance of DBR methodology because of its success in driving advances in information and engineering disciplines (Kuechler & Vaishnavi, 2012).

The DBR to address several issues central to research studies as follows. The first issue was the nature of and approaches to a subject in the real world rather than the laboratory. It was also developed to address the issue of the use of a broader set of measures of the subject that emphasize competency rather than theoretical knowledge. The DBR was also positioned to address issues of the synthesis of recommendations for design or process improvement based on the formative evaluations, compared to summative evaluation of the classical research methodology. The DBR approach to research allowed proper integration of the “difficulty with the complexity of real work situations and their resistance to experimental control,” the availability of large amounts of data, and issues related to “comparing cross designs” (Herrington, 2012). The role of DBR in the researcher’s toolkit was to create practical knowledge to realize theoretical formulations. It also allowed for formative research to test and refine designs based on theoretical principles derived from practical measures, prior research, and progressive refinement through assessment (Hogue, 2013).

According to Blessing and Chakrabarti (2009), DBR methodology occurs in four parts. In the first part of the methodology, an existing design’s circumstances and constraints were presented and analyzed (the analysis phase of design research). In the second part, the researcher studied the interaction between the design and the results (the design phase of design research). In the third part, the researcher deliberately manipulated the design to change the interactions within the domain of interest by addressing design constraints that may be responsible for the results (the evaluation phase of design research). Finally, the researcher proposed and tested new designs and tools (the test phase of design research). Design-based studies required the analysis of designs within a robust framework of comparative inferences on structure and function. In this study, the designs were the data models for management analytics and decision-support problem-solving.

This research methodology required a quantitative format. The use of a quantitative format for a study created explicit links between theory and results, limiting the bias of the researcher (Creswell, 2011). In a quantitative study, the essential elements of the analysis are mathematical constructs. Quantitative research techniques differ from qualitative techniques, which use linguistic constructs (Creswell, 2011). Within the domain of quantitative research, there were five main research design types: randomized experiment, quasi-experiment, comparative, associational, and descriptive (Creswell, 2012). The first two types were experimental techniques, while the last three belong to the non-experimental class of techniques. According to Creswell (2012), there were four items to consider when selecting the approach to a quantitative design: random assignment of subjects, intervention or treatment by the researcher, structure of the criterion variables, and approach to the examination of relationships between variables. In comparative study design, there is no randomization of the assignment of subjects to the groups and no specific intervention against subjects in the study. However, there was the requirement to define numerical quantities for comparison of the items studied. In a descriptive study design, the expectation of comparison relaxed for the study to focus on the design.

Methodology

Based on the framework discussed above, the choice for this study was a descriptive approach. The secondary nature of data for the study dictated the selection of the descriptive approach. The availability of suitable secondary data allowed progress without the burden of collecting data. Secondary data also allowed a focus on the original attribution of the subject of interest as represented by the available data sources, but imposed constraints on the causal interpretation of the underlying effects and influences.

Population

In study design, the population refers to the group studied. The population in this study was not typical. The data used in this study was sales data from a medical product distribution company which captured purchasing habits of customers, selling characteristics of agents, market demand, pricing actions on products, and marketing actions to drive penetration of products within the marketplace. The scope of the data was enterprise-wide, which meant, it contained all the information captured by the organization from its business activities related to the products, pricing, sales, marketing, and customer servicing. Each of these areas had sets of concepts which were extracted to facilitate the analytic activities. I expanded on the Resources-Events-Agent (REA) ontology framework proposed for business. See Appendices F & G for details.

Sampling and Sampling Procedures

This study did not utilize sampling or sampling procedures. Analytics and decision-support problems in management required the participation of all data points in the analytic processing. The focus of the study was to construct the data model that leveraged every necessary data point in analysis and decision-support problem-solving. This approach was selected to overcome the challenges of existing data analytics processes where the use of sampling added complexity to the insights generated due to the concerns of representativeness of the sample compared to the entire population of items under consideration.

Archival Data

I used secondary data in this study. The data sources included SAP/R3 Enterprise Resource Planning (ERP) system order processing module, along with additional sources of product and market information gathered from second- and third-party sources. The dataset spanned three years, from 2007 to 2009. Appendix A shows the list of data sets included in the data use agreement approved by Institutional Review Board (IRB). The IRB approval number 10-28-15-0015433 was issued October 25, 2015. The use of historical data was deliberate and should not impact the outcome of the research. The data came from a data asset repository used for exploratory data analysis and analytic pilot projects. The data was completely de-identified. Randomly generated identifiers related the different segments of the data together. The next chapter contains the description of the data used to illustrate the data model extension approach. The chapter also covers the anomalies in the data set addressed to ensure accurate transformation into analytic and decision attributes.

Data Analysis Plan

A data analysis plan should present the description of the software, data cleansing and screening procedures, details of the statistical tests, procedures, variables, and how results will be interpreted. Because this study focused on data model designs for analytics and decision-support problem-solving in management, the emphasis was on discovering attributes of the data for managerial tasks. For this reason, I expanded this section to include the processes of ontology learning, data engineering, and analytic formulation which were critical to the data model extension methodology and the data analysis in large complex analytic and big data scenarios.

Data Model Extension Methodology

As mentioned above, the data available for data analytics projects in a big data environment came in different formats, data types, and data naming conventions. To conduct a proper analysis of the underlying data model, I constructed data asset diagrams at two levels: high-level and detail-level. The high-level data asset diagram provided a panoramic view of all the data asset available for the analytic exercise, and the links between the datasets. The detail-level data asset diagram showed the content of each data asset, including the logic for the links between the data assets when they existed. The dataset link logic was of three types: direct referential association (primary key – foreign key association), indirect reference association through matching or associative relation.

The typical dataset for big data analytics described above contained duplication, redundancy, inconsistencies, and other data issues. These data assets also embedded the critical data, process, and business rules that are helpful in the application of the data model to uncovering problem scenarios. To highlight these situations in the data, I refined that data asset diagram using a generalized entity relation recognition algorithm to restructure the data asset into a generalized entity relation model and generated an accompanying entity relation diagram for visualization of the data model. At this level, the data model applied all the normalization rules to ensure data quality and integrity in the data. This data model reflected the piece of the “real-world” expressed by the available data, as interconnected elements of a type system with one or more schema(s), devoid of artifacts of physical implementation as databases, data-files, and applications. (Puonti, Lehtonen, Luoto, Aaltonen & Aho, 2016).

For big data analytics, this real-world was complicated and contained tens of schemas. Each schema formed the collection of relations within a data model connected by association restrictions, including, domain, cardinality, and referential types. The data model at this point still embedded the functional and transitive associations. Additionally, the data model did not express the progression of the concepts with the data over time. This classical data model was also limited to relations to real entities within the schema as such the resulting design manifestation was known as the entity-relationship data model (Puonti et al., 2016).

In this classical data model, given properties, P_1, P_2,…, P_n, a relation, R, was defined by the n properties such that each instance or tuple had its first property from P_1, its second property from P_2, and so on. A relation, R, was the subset of Cartesian product P_1, P_2,…, P_n or $\prod_{j=1}^{n} P_j$. P_j is the j^{th} property of R with degree n, hence referred to as, n-ary relation, with the following characteristics (Kumari & Singh, 2017):

1. Each row was an instance of the relation or tuple of R,
2. Row ordering was not consequential,
3. All instances or tuples of R were distinct or unique,
4. Column ordering corresponds to the ordering of the set of attributes of R,
5. Term label corresponding to the set domain conveyed the significance of each attribute
6. The term labels applied to the attributes were unique and conferred some interpretative value to its content
7. The combination of attributes covered by R uniquely described an entity, subject, object, or class with the rows or tuples reflecting the membership in the collection
8. Property values assumed standard data types, including, integer, decimal, character, and currency, all of which were of scalar type.

9. Advanced data types like user data types (UDTs), algebraic data types (ADTs), statistical data types (SDTs) and functional data types (FDTs) were not allowed in data models.
10. A typical data model of the enterprise units contained hundreds of relations within tens of schemas.

The entity relations of the enterprise data model were not well differentiated. They could represent people, groups, events, actions, transactions, resources, place, and other ontological classes. To facilitate the translation of entity to ontology relations, I derived a list of 16 possible ontology commitments for management analytic and decision-support problem problem-solving shown in Appendix G. I used the list of candidate ontology commitments to refine the relations and properties within each data set into ontology classes and properties as data model extensions. These derived relations differentiated the entity relations into higher forms encompassing complex associations and constraints.

To derive data model extensions, I generalized the analytic continuity concept of functions to the relations through analytic elements. Analytic elements were projections of the properties of the relations beyond initial specifications, but which maintained logical continuation as follows. Considering the data model as a collection of properties, P, of relations, where each relation, R, was the generalized functional of a unique aspect of an ontology, O, of a domain, D, in the universe, U. Each property, P, was an analytic element, (a, l), where, a, was attribution represented by the property, and l was the function or logic on *a*. The original analytic element in a data model was (a_0, l_0) and subsequent derivation results in a matrix (a_1, l_1), …, (a_i, l_j) were extensions of each other through the connection component, σ, of the set $a_0 \cap a_1 \cap, \ldots, \cap a_x$ if $l_0| == l_1|$ σ, …, $== l_x|$ σ^*. The analytic element defined by the pair, (a, l), therefore, continued to the boundary point, ε, with $\partial a \subset D$. These elements continued the expression of the relation, R, beyond the defined scope. This universal cover was the original scope for extensions of the analytic elements of the relation, or the analytic space. The maximal analytic extension of the relations was, therefore, an unambiguous holomorphic functional of complex properties differentiable about every point in the domain. This new relation, σ^*, specified the region where the sum of terms of the sequences of the relation, or its infinite series, became divergent. It extended the point beyond which values existed, and when the expectation of the return of a single point was unrealistic (so-called mathematical singularity).

The use of complex numbers reinforced extensions, especially, when defined on more than one property of the relation. Complex numbers eliminated the need for mathematical singularity or isolated points and favored algebraic or geometric varieties with a mathematical plurality or cohomology. The dimension axioms defined expanded the analytic space of the relation as follows. Given that complex numbers were expressions of the form $y_0 + x_i$ where x and y were real numbers, and *i* was the imaginary unit, the solution to quadratic equation $y^2+1=0$, and satisfied the equation $y^2 = -1$. Complex numbers extended the dimensionality of real numbers which were, technically, points or zero-dimensional. For example, 1-dimensional number line was extendable to the 2-dimensional plane by using the horizontal axis for the real numbers and the vertical axis for the imaginary part of the complex number. The complex number $y_0 + x_i$ identified point coordinates (y, x) in the complex plane. A complex number whose real part was zero was said to be purely imaginary, whereas a complex number whose imaginary part was zero was a real number. Therefore, complex numbers were analytic extensions on ordinary real numbers, converging on an area defined by a range of real numbers within the domain, the germ, g, of the power series.

Assuming two germs, g_1 and $g_{2,}$ were the sets of vectors, when the absolute difference between the set of vectors was less than the radius of convergence of $g_{1,}$ and if the power series defined by g_1 and g_2 specify identity relations on the intersection of the domains, then g_2 was an extension of g_1, making up the point (or sheaf) of the extension. The union of the germ sheafs identified from the power series of the domain by sets, $Ur_{(g)}$, for all $r > 0$ and $g \in G$ defined the basis for an open set for the topology on G. Connected components of G, equivalence relations, formed the analytic extension map of space. A map defined by $\varphi g_{(h)} = h_0$ from $Ur_{(g)}$ to C where r was the radius of convergence of g, represented the chart of the extension. The set of such charts formed the atlas for G or the universal relation. Therefore, an analytic extension of the relation generalized the power series defined from the sequences of the underlying properties. They created objects within topological spaces of class equivalence with others of the same type with shared local properties. Therefore, the converging power series of the relation properties enhanced the underlying information and resulted in sets of vectors around the points or space of expression within the empirical domain.

Analytic extensions on relations provided the paradigm for extension of data models to include high-order logic. The extensions specified the analytic space through the derivation of algebraic varieties of the domain ontology and topology was possible. This algebraic variety ranged from simple ones like variables to intermediate types such as features, identity vectors, and eigenvectors to name a few. Complex algebraic varieties like tensors or co-dimensional entities resulted from the further projection of the intermediate algebraic variety. At the level of the complex algebraic varieties, the dimensions were degenerates of real dimensions with about half the number. That is, if the dimension of the complex algebraic variety were, d, its real dimension would be 2d. The real algebraic variety of equations with real coefficients became the dimension. The real dimension referred to the maximum number of manifolds contained in the set of its real points. Also, the real dimension was never greater than the complex dimension and equals it if the variety was irreducible and had real points that are singular. For example, the relation of a complex algebraic variety with dimension two would be a surface, but with a real dimension zero. It had only one real point, (0, 0, 0), which was singular. The relation representing a smooth complex hypersurface in complex projective space of dimensions, n, was a manifold of dimension $2(n - 1)$. The complex hyperplane did not separate a complex projective space into two components, instead, expressed them as having real co-dimension of 2.

Based on this methodology, analytic extensions on relations were, inherently, more robust than an analytic extension of functions, and enabled the differentiation of data using analytical geometry. It transformed scalar data into algebraic varieties. The additional properties improved the expressiveness of the relation for problem-solving based on the set points or boundaries of differentiation that would satisfy the analysis problem. Considering the properties of relations as analytic elements made them heuristic translators, mapping one property to another, within the domain defined.

Composing an enterprise from relations of algebraic varieties transformed the data model into empirical ontology with a well defined analytic topology. The progressive differentiation and integration of these varieties expanded the characterization of the domain but maintained a universal cover on the relation reducing analytic over-reach. Analytic continuation or extension reached its boundary when the data model captured very concepts of the domain in alignment with the intuitive cognitive model of the business analysts and executives.

Expanded Data Analytics Process

The data analysis process started with the arrangement of the available input datasets, followed by extraction of the underlying entities and relationships, and the reconciliation of the attribution of entities and relationships across the datasets to achieve a rationalized metadata model of the input data. Further abstractions of the metadata captured characteristics of the underlying data not explicitly expressed by its metadata.

Contemporary literature on data analysis assumes the creation of a data model should precede data collection. It also often assumes that the data is in a structured format, mostly from databases. In the light of big data, these assumptions were no longer valid. Relaxing these assumptions allowed data in any form and from any source to participate in analytic processing. In this scenario, the data-modeling became the dynamic process of discovering the characteristics and relationships between the available data sources. This study used secondary data gathered without an explicit data model of the analytic needs. I adopted a data analytics process made up of the five following steps.

The first step of this expanded data analysis plan was to create the catalog of the datasets that were available for analytic processing. This catalog specified the nature of the datasets along with the format, the number of attributes, and the nature of attributes available to establish the scope and boundaries of the analysis problem. The catalog mapped datasets to the analytic processing objective. This mapping exposed gaps, when they existed, within the available data sets. Connections between the datasets, included referential keys, common attributes, hierarchical, temporal, or spatial association types.

The second step was to expand the datasets into entities and attributes. This step involved a critical review of the structure of the datasets. It included the capture of user-friendly names, descriptions, database data types, analytic data types, measurement scales for each attribute in the dataset. The classical data models typically ignored measurement scales, creating confusion with the interpretation of numbers.

The third step was to expand the metadata catalog to include relationships within and between the datasets, such as extension (1:1), subsumption (1:M), and qualified (M:N) relationships. These were cardinality relationships within the data model, which were expanded to accommodate non-cardinality semantic, symbolic, and dimensional expressions between and within entities in the data sets.

In the fourth step, attributes were organized into unique subject areas or functional domains to understand the extent of representation of the subject area or functional domains, as well as its association and dependence on other subject areas or functional domains captured in the data sets. Analysis of the data within the entity structure determined whether there were dynamic components of the attributes indicating the data may be of a repeating nature, either longitudinal (single subject over time) or cross-sectional (multiple data points at the same time from different sources). This orientation of the data was critical to determining the appropriate types of data engineering and analytic formulation processes to apply to the data. This decomposition of the subject areas or functional domains into static and dynamic components furthered the data model.

The fifth step was to establish the measurement frames for the concepts and then generate new attributes or values needed to operationalize them with the available data. The step continued for all the concepts of the decision-support problem. In classical business analysis scenarios, the optimal data model should allow business users access the attributes of the subject of interest in a way that aligned to the cognitive or conceptual representation of the domain. This outputs from this data model should support business analysts and executive in the tasks of management - plan, organize, lead and control without any need to for technical knowledge of the analytic techniques or data engineering method requirements.

Threats to Validity

Because this study employed DBR, which does not require interaction between the researcher and the subject or the data collection process, the common threats to validity did not apply. In studies where the researcher has direct or indirect interactivity with the subject, it is critical to address threats to validity. Addressing these threats prevent issues with study design (external validity), subject selection (internal validity), and inference from sample to population (construct validity). External validity issues include reactivity, interaction effects, specificity of variables, and interference. Internal validity issues include self-selection, non-stationary effects, and subject retention. Construct validity issues are related to statistical conclusion requiring correction. In classical scientific studies, the validation techniques in these situations may be statistical. In this study, I took an analytic and decision-theoretic approach to validity.

Analytic and decision-theoretic techniques evaluated models in the context of specific analysis and decision needs of the users. The analytic and decision-theoretic approach to model validation, though relatively new, has shown promise in alleviating interpretation constraints imposed by pure statistical validation approaches (Welton & Thom, 2015). The analytic and decision-theoretic approaches focused on determining the strength of evidence for a model's empirical power in the context of the specific analytic and decision situation, without consideration of the generalizability (Jiang, Yuan, Mahadevan, & Liu, 2013).

External Validity

The nature of the business data used in this research does not present situations that would challenge external validity. The goal of the study was to support strategic decisions in a domain of responsibility by establishing a methodology for data model extension. The data-modeling exercise would explicitly define attributes to address the reactivity, the interaction between subjects, the specificity of attributes, and the interferences existing in the data to enhance analytics and decision-support requirement of managerial tasks.

Internal Validity

Issues of internal validity also did not apply to this study because of the nature of the data under consideration. The analysis provided insights for decision-support problem-solving in management. In business analysis, absolute precision was not necessary. However, management requires consideration of history, maturation, regression, churn, and interaction, which are part of the practical expression of the subject for decision-support.

Construct Validity

In this study, construct validation was limited to analytical and decision-theoretic conclusions based on the significance to management analysis and decision-support. An important criterion in construct validity was the degree of alignment of conceptual or cognitive expectation of the management analysts or decision makers of the enterprise functional unit or domain. For this reason, the measures used in this study focused on business alignment, for example, intelligence density, decision yield, cognitive gain, empirical lift, and Bayesian yield. When necessary, statistical reference attributes, including f-statistic, t-statistic, f-statistic, and others were computed to assist the interpretation of the strength of evidence.

Ethical Procedures

The source data used in the study came from the data warehouse I maintained for analytic model and decision algorithms development. Personally identifiable information was removed from the data. The study did not require human subjects or interviews. The data was more than five years old. It was large enough for the study of extensions of relational data-modeling needed to advance analytic processing in databases, beyond the current state allowed by online analytic processing (OLAP).

Summary

The goal of this research was to study data model extension for management analytics and decision-support using a DBR methodology. The choice of this methodology aligned with the purpose and nature of the study. The quantitative non-experimental descriptive research format provided the ideal approach to the study of the data model and the opportunities to improve them through the implementation of analytic extensions.

A critical component of this methodology was the modified data analysis plan, which addressed the complexity of the available data for analytic processing. The data analysis plan also addressed the links, both functional and non-functional, within and between the datasets. Using the extended data analysis plan led to the identification of conceptual data elements of the available data, which connected the available data to the intentions of the management analytics and decision-support. These abstract and conceptual data elements connected the measurement frames of subjects in the available data. These measurement frames were the quantitative expression of effects, influences, and other characteristics embedded in the data.

A point of the expanded data analysis plan was that large and complex domains of an enterprise required reconciled attributes to map to the ontology of the underlying subjects. The mapping from data to ontology helped align structural, formal, and resolution expectations at appropriate levels of analysis. The mapping also allowed measurement, estimation, inference, and forecast needed to resolve business questions and management problems. Implementing analytic extensions at the data level transformed data from raw input into attributes for cognitive processing. For this reason, the focus of evaluation of these analytic attributes was on the empirical measures of analytics and decision-support, such as intelligence density, cognitive gain, empirical power, and others. Statistical measures of evidence such as statistical power, confidence interval, p-value, parametric statistics, and others were secondary to the analytic and decision-theoretic measures.

Chapter 4 contains further discussion of the analytic extensions, the details of answers to the research questions, and an application in a big data analytics scenario of a medical product distribution company. In Chapter 5, I provide a discussion of the findings and the implications for research in applied management and decision science.

Chapter 4: Results

Introduction

The purpose of this quantitative descriptive study was to examine the data model of a typical data analytics project in a big data environment for design alternatives to address issues of misalignment of data analytics project outputs, available data, and the prevailing intuitive cognitive model of the problem and solution scenarios. The objective of the study was to improve the acceptance and actionability of data analytics outputs by business analysts and executives. I identified the data model as the primary focus of the study because the expression of information in data models was known to improve understanding and application of the data to management problem-solving (Burch, 2018).

The research questions of this study were as follows:

Research Question 1: Can data model extensions improve the discovery of management scenarios from big data?

Research Question 2: Can data model extensions improve the formulation of insights about the management scenarios?

Research Question 3: Can data model extensions express the complex constraints and rules needed to compose the acceptable and actionable solutions for analysts and executives?

In this chapter, I discuss the results of the data model extension methodology and the extended data analysis plan, as described in the previous chapter. I follow the discussion of the results with an application of this data model approach to a typical data analytics project in a big data environment.

Data Collection

Data collection in a classical research situation provides information on the data collected for the analysis of the research subject including recruitment rate, response rate, discrepancies from plan, baseline statistics, sample representativeness, and so forth. In this study, I used secondary data and discuss the data collection process for a big data project in a DBR context.

The data analytics projects in big data environment start with a list of available data assets and a vague description of the business objective of the analytic and decision-support exercise. The available data assets represented the universe of data for the formulation of the analytic problem under consideration. The vague description of the business objective stipulates the expectations of the analytic exercise. Detailed requirements were problematic due to the overwhelming availability of data and the complexity of the information about the business problem.

The key steps in big data analytics projects were the collection of all the datasets available for the analysis, preparing the data for analytic algorithms, running the analytic algorithms to generate analytic models, running the analytic models against new data to determine its performance, reporting the performance of the models, reviewing the results in the context of the business problem under consideration (Zicari et al., 2016). When the results do not provide satisfactory answers to the business questions, this process is iterated until an acceptable answer is produced. Using the data model extensions and extended data analysis plan to update the data model throughout all these steps ensures that progression of the data analytics process. The extensions of the data model created the cognitive breadcrumbs needed to adapt the analytic processing for complex business problem-solving. The data collection process went as follows.

Since the big data environment was the collection point for the available data in the organization, and it ingested and maintained the data as-is from the source systems, whether structured (data files, tables) or unstructured (documents, records, graphs, multimedia) formats. The first step in the data collection was capturing the names of the data assets, number of files in the set (if more than 1), partition logic (if multiple files), format, size, number of rows, and the number of columns in all the datasets and documents provided. For each of the files in data asset, I determined whether there were links between the files, and if so, which attribute(s) established the link. I generated a data set link inventory to sustain the connections between the data assets for further processing.

I processed each data set further to gather details of the content. This included data element labels, data type, number of unique values, and cardinality ratio. The data element label was either the first row in the data set, declarations at the beginning of the data set for filetypes like parquet or Avro, separate metadata files like flat files, or database catalog of the primary source system. Data types were primary scalar data types of numeric and non-numeric types. Variants of numeric data types like integer, big-integer, decimals, money, float, and non-numeric datatypes like character, variable character, text, binary, variable binary were derived when not explicitly provided with the initial data set. A row in the data set was either a record for structured data or a line for unstructured data. Unique values were derived by counting the unique occurrence of the values of the data element either in a row of structured data or line of unstructured data. The unique values were the tokens of the underlying subjects. I calculated the cardinality ratio as the number of unique values divided by the number of the rows in the data set. Using the cardinality ratio, I implemented the following relation identification algorithm

Table 5

Relation extraction algorithm

0	For each data set, set dataset name to dataset label
1	For all data elements in the data set
2	if there is a data element with cardinality ratio = 1 and the data element is not a timestamp; then assign this data element a relation property key status assign the label of the data element as the name of the relation;
3	if the sum of all cardinality ratios = 1 then assign all data elements relation property key status assign the concatenated label of the data elements as a name of relation;
4	If the sum of all cardinality ratios > 1 then find the combination of the sum of data element keys that add to 1 assign each combination a relation property key name assign each combination a relation name that combined the name of property names
5	if

	data element with cardinality ratio = 1 and the data element is a timestamp, then skip timestamp data element find the combination of the sum of data element keys that add to 1 assign each combination a relation property key name assign each combination a relation name that combined the name of property names
6	If the sum of all cardinality ratios < 1 then find a partition of the data with a combination of the sum of data element keys that add to 1 assign each combination a relation property key name assign each combination a relation name that combined the name of property names. If no partitions were found that meet this criterion, flag dataset for manual review
7	End;

The following relation types were collected through this process.

1. The primary relation which enumerated instances of items with similar characteristics, for example, customer, product, sales representative, time, location, and others.
2. The composite relations which captured the interaction of the primary relations in the data model, for example, a sale became the relationship relation of the interaction of product, customer, sales representative, time, space, price, and so on.
3. Detail primary relation with high cardinality attributes of the primary relation properties.

4. Detail composite relation with high cardinality relation properties of composite relations.

The cardinality ratios separated the relations into two groups. The third normal form relational data model that resulted from this process captured the universe of attributes available for processing the goals and objectives of the data analytics project.

The collection of data about the data model extensions used the four types of relations – primary, primary detail, composite, and composite detail, captured with the relation recognition algorithm as input. Noting that the detail relation types extended the primary relations, the connections between these relations were extensions.

For primary relations, the connection could be one-to-one, union (full or partial), or none. If one-to-one, the relations were combined without any loss of data. If the full union, that meant the two primary relations could be concatenated together without a change of the number of properties in the resulting combined data set. If partial union, the relations had common properties which were concatenated such that each relation had properties that were unique to them and no values for the relation properties that were not common. When the properties of two primary relations did not map to a single set of properties, I combined their properties and allowed the attributes unique to each relation to remain as missing values. The missing value treatment such as elimination or imputation was applied at a later stage in the process as deemed appropriate. A relation connection type of “None” indicated complete independence of the two primary relations. Before declaring that the relations are completely independent, I searched for non-natural connections between the relations, such as, geographical location or timestamp, and so forth.

For each relation property within the relations of the data model, I created a new property to represent the association between the relation properties. The realization of these associations depended on the analytic formulation needed to quantify the association. The applicable analytic formulation process depended in the assumptions about the association which can be linear or non-linear, continuous or non-continuous (see Appendix F). As such, I adopted a process of appending the name of the analytic process to the name of the association. When multiple analytic formulations are applicable, I used as many of the techniques as needed to facilitate capture the data about the representativeness of the technique. I evaluated relation property values to discover transformation that would result in new relation properties. I also evaluated each pair of relation property values for form expressions of the association between them. The form expression between these attributes resulted in new relation properties. Recalling from Chapter 2 on the expression of large analytic domains that the formulation of form expressions for every combination of attribute resulted in very large matrix, I limited the discovery of form expressions to those that would advance the ontology learning of the objective of the analytic processing.

With new relation properties, I labeled and integrated them into the data model. This process continued until all the functional associations were discovered and integrated into the data model. The universe of attributes expressed by the relations in the data model at this point would be comprehensive for the discovery, identification, specification, and resolution tasks for the data analytics project. I used the conceptual model generated from the process to evaluate the appropriateness of the data for the analytic exercise.

To evaluate the information within the data model, I used information entropy calculation. For each relation and the relation property of the data model, the information entropy was calculated as the sum of the proportions of the values multiplied by the normal log of the proportions of the values. The relation or the relation property would be 1 when there was an equal proportion of all values and approach zero the more the variation of the proportions of values are in the relation. For the ontology learning process, relation properties with large information gain calculation had low ontology classification. Also, low information gain represented high ontology class. This situation of the ontology class was an indication of the generalizability of the concept across the domain. The information gain distribution of the relation properties was used in the determination of the ontological commitment for each property value in the data model and the derivation of data model extensions.

Study Results

Using the data collected from the process described above and the information entropy calculation discussed above, I grouped the relation property values into three classes. Property values with information entropy > 0.75 were specialized ontology concepts which were helpful in defining the dimensions of the subjects of the data model.

Those with analytic significance metric between 0.35 to 0.75 contained ontology concepts that had moderate generalizability and were helpful in extending the data model for specific analytic situations. Finally, the property values < 0.34 were helpful in extending the data model for general expression. This data model extension approach preserved the operations of relational algebra and calculus, including union, difference, product, selection, projection, logic, and arithmetic. It, also, sustained the connection of input data to the analytic expressions constructed for knowledge discovery, business intelligence, and decision-support in management.

Recall that the analytic elements of the properties of the relation had an attribute component (*a*) and the logic component (*l*), where the attribute element had value (*v*) and scale (*s*) sub-components. The manipulation of the *a, l* components of the analytic element resulted in the differentiation of the property or properties of the relation. These extensions were higher-order conceptual relations or properties derived from lower-order classical relation or properties using established data engineering and analytic formulation techniques (Foster & Stein, 2013). The catalog of ontological element types and analytic formulations are shown in Appendix F and G respectively. Further discussion of the extensions follows.

Semantic Extension

Semantic extension formed interpretation continuity on attributes of the relation. This extension involved manipulating the v-subcomponent of the *a*-compoentt of the analytic element pair, (*a*,*l*), discussed above. This manipulation expanded the expression of the relation property into a relation that captures all forms of representation based on the value expression (Krogstie, 2012; Feilmayr & Wöß, 2016). Semantic extensions of the data model transformed the scalar quantities of the relation properties into vector and matrix formulations of the relation. The extension impacts non-numeric attributes the most since their vectorization requires coding or transformation, such as dummy coding, threshold coding, proportionality coding, probability coding, the weight of evidence coding, Rasch coding, Dempster-Shafer coding, Likert scale coding, and so on and so forth. The implementation of semantic extensions identified the cases and the states of expression of characteristics of subjects within the data.

Semantic extension expanded the data model into its first-order logical expression. First-order logic system were sets of propositions on concepts conferring meaning to the underlying objects and subjects arising from the interaction of objects. In turn, propositions were sets of atomic predicates connected with logic connectives into compound predicates. The atomic predicates formed the units of expression of the object or subjects and offered interpretation context(s) reflected in the term or name reference of the predicate, the value assignment, data type, and scale of this value. The semantic relation became the collection of propositions for the expression of the instances of the relation with both conjunctive and disjunctive logic. Note that the typical entity relation is of conjunctive normal form only.

Semantic extension transformed the entity-attribute-value structure of the data model into the subject-predicate-variable structure of first-order logic. Semantic relations were made up of elements whose instances were variables of well-formed atomic formulas of first-order predicate logic with bounded values. For example, attribute, A, was the predicate logic for the instance or variable of A, p (?a), within attribute domain of A given entity, |a|, such that,

A: p(?a, |a|) = true & p(?a, not |a|)=[false | undefined]

Every attribute value was, therefore, the predicate logic assertion. For example, if attribute A was “year”, its variable, ?year, with a domain, |year|= |2010, 2011, 2012, 2013|, an assertion for predicate, p(?year, 2010) = true and p (? year, 2014) = false (closed world interpretation) or undefined (open world interpretation). The number of distinct predicates became fewer than the number of variables. Essentially, every instance or set of instances of attributes transformed into first-order predicate semantic relation.

The semantic expansion described above, based on first-order predicate logic, transformation resulted in many variables in a relation, especially when dealing with qualitative attributes of high cardinality or quantitative attributes whose magnitude had significance. For these attributes, semantic continuity was established with count, order, or ratio measures respectively. This improved expressivity of the semantic relations, allowing aggregation of lower-order predicates over well-defined sets, supersets, and higher levels of predicate cardinalities using following qualifiers:

- the many-sorted logic to partition instances into populations, groups, and types;
- intuitionistic type logic to link variables to proof and dependency types;

- Logic modal qualifiers, such as
    - alethetic for the states of possibility, impossibility, necessity
    - temporal for the timestamp, time span, time horizon
    - spatial for point, area, space
    - deontic for mandatory, obligatory, permissible,
    - epistemic for propositional, hypothetical, theoretic, proven
    - doxastic for temporal-spatial, situational, positional, and
    - fuzzy logic for heuristic categories

Essentially, semantic extensions of the data resulted in atomic concepts of the underlying data in a fully specified form. As noted above, this organization of data favored the derivation of scalar quantities representing the magnitude of the properties of the subjects as expressed in the data. The degree of connectivity of logic represented the complexity of expression. The associations or correlations between the variables provided the guide needed to answer complex questions about the subjects using the available data. The estimate of the size of the semantic database was the product of the sum of arities of all semantic relations and the number of relations.

The implementation of this extension of the data model required the following transformations:

1. For attributes with assignment devoid of order or interval, each assignment became a variable with binary values, commonly coded as one if present and zero if absent respectively.

2. For attributes with an assignment with order but devoid of the interval, the value is translated into a rank order further transformed, normalized, regularized as needed to capture the magnitude of expression.
3. For attributes with an assignment with order and interval, the value transformations eliminated covariation (standardization), scale issues (regularization and normalization) as needed
4. For attributes with characteristic assignment devoid of order with a large number of values, were grouped
5. For attributes with a characteristic assignment with order and a large number of values, were rank ordered
6. For attributes with numeric assignment devoid of order with a large number of values, were grouped
7. For attributes with a numeric assignment with order and many values, were rank ordered

Symbolic Extension

The symbolic extension formed subject-object relations which layered expression of entity types and groups into conceptual boundaries for similarity, discriminant, or other quantitative distance measures (Diday, 2012). They were extensions on the *l*-complement of the (a,l) pair of the analytic element. With these extensions, the concept of CUSTOMER became vector or spectrum of expression based on figurative characteristics like purchase frequency, price sensitivity, lifetime value, churn probability, to name a few. This vector could compare to the PROSPECT vector to show the flow between them in the PARTY vector of vectors. This result was the map of the journey of a PARTY through the enterprise from the cradle to the grave. Superimposing the actions of management responsible for changes in the characteristics of these vectors over the lifetime of the PARTY brought clarity to the consequence of management decisions and actions.

While the semantic extension evolved the data model into the relation of variables and captured the logical association behind the expression of the subjects within the domain of interest, the symbolic extension quantifies this expression for the comparison of the subjects (Jiao, Zhou, & Chu, 2016). Symbolic extension of the data model adds a transformation to records of the subject of interest to represent the distance from each other, or to a normative reference. This extension derived attributes of the subject of interest as a specific form of lossless encoding of the characteristics of the subject. The results were mappings, π, from a space of expression, y, to that which defined it, x, represented as follows:

$$\pi : y \rightarrow x$$

$y = \pi(x)$, that the subject of interest, y, encoded by the set of variables of the subject of interest, x. It was not necessary for this mapping to be injective, that is, for the same expression of the subject of interest transformed similarly by the same set of inputs (one-to-one mapping between expression and input(s)).

An emphasis on this extension was the capture of the space of expression that was identifying the subjects of interest to make similarity or discriminant assessment quantitative. Each element $y \in Y$ determined a unique set $x = \pi(y) \in X$, where every $x \in X$ was representative of one or more elements of $y \in Y$. Space, Y, was the Cartesian product of finite metric spaces defining the boundaries of the topology of the subject of interest. This space simplified the assessment of similarity compared to using the natural metrics within the space, X.

With this extension, the enterprise transformed into an abstract topological dynamical system, $\mathcal{Y} = (Y, T)$ where Y was a metric space and T is a continuous function within the metric space. That is, $\mathcal{Y}$, had n-dimensional axis reflecting states of as subset of the entire space, $\mathcal{Y} = (Y, \sigma)$. This space consisted of finite form expression $Y=(y_n)_n \in \mathcal{Y}$ over a finite relation, which was transformed using a shift map $\sigma(y)=(y_{n+1})_n \in \mathcal{Y}$. The relation between $\mathcal{Y}$ and (Y, T) was the factor map which controls the translation of Y into $\mathcal{Y}$ dynamical system, such that the lower layer was made up of fine grain attributes that encoded general characteristics of the subjects of the domain, while higher layers were responsible for specific characteristics of the subject.

The first *k* attributes jointly represent the union of a unit of order *k*, with every unit translating into a factor of the subject, independently. The units of the first order form into disjoint blocks or groups, each unit of order k included a piece of the k^{th} attribute and a finite collection of attributes of lower orders. The result was an association or concept hierarchy based on criterion induced by order of expression which was subtler than a strict hierarchical or graphical association based explicit criteria in the information.

To further explain the data model extension, let $y_1,...,y_p$ be the set of variables, D_j be the underlying domain of Y_j and $|Y|_j$ the range of Y_j for j=1,...,p., set of values for Y. Given a symbolic relation, S, with p-tuple $(y_1,...,y_p)$ with $y_j \in |Y|_j$ for j=1,...,p. S ={ $s_1,...,s_n$} were the relation instances, then $Y_j(s_i) \in |Y|_j$ for j =1,...,p, and i=1,...,n. Therefore, the data array consisted of n relations, one for each instance $s_i \in$ S, such that $(Y_1(s_i),...,Y_p(s_i))$ for i =1,...,n. The data types of the symbolic variables took on additional forms besides intervals and count integers of the semantic space, such as arrays of different dimensions, functional expression/mapping, nominal and ordinal values, modal values, standard numeric valued data types (Noirhomme-Fraiture & Brito, 2011).

An important relationship in symbolic extension was the concept of full or partial dependence. A symbolic relation, S_1 may be fully or partially dependent on another relation, S_2, if it could only be applied when S_2 takes expression within the all or given set for S_1. The relation, S_1, was dependent on the relation, S_2, if S_1 made no sense for some values of S_2, and hence became non-applicable.

Dimension Extension

Dimension extension established the accurate and orthogonal axes of expression of the abstract interrelated analytic spaces within the domain of interest given semantic and symbolic differentiation and integration. This extension “dimensionalizes” the domain into invariant quantities with the absolute significance of relative magnitude, per

Buckingham π theorem and Bridgman's principle (Shen, Davis, Lin, & Nachtsheim, 2014). Applied to the example of the PARTY, the complex dimension extension, defined the multiple paths for the PARTY vector of vectors which connects to the different outcomes within the domain, for example, customer tenure, type of order, purchase frequency, willingness to pay threshold, to name a few. Dimensional extensions allowed the definition of metrics like party conversion velocity and acceleration useful in determining the progression towards management targets.

Mathematically, a dimension is an axis or aspect of the expression of a subject or object defined in the geometric form. It is the derived extent on measures, metrics, moments, and coefficients of expression such as length, breadth, volume, height, to name a few of a subject or objects assuming its geometric realization. Dimensional extension represented these extents in the form of invariant quantities (or points) in an abstract space allowing accurate inference, projection, simulation, and optimization of the characteristics of the realized geometric form. In this sense, this extension applied to subjects and algebraic rules governing quantities, such that, calculations and derivations maintained correspondence with the properties of the subjects represented. It assumed an abstract space of expression defined by the dimensions inherent in the available data in which each record occupied a point in the space. The dimensional extensions provided the abstract coordinate system for analytic exploration of a subject of interest based on available data, allowing the manipulation of the numbers without concern of the units of measurement underlying the properties under consideration.

The construction of a dimensional extension to the data model required the abstract assignment of input and output features of the subject of interest, based on the provenance of the subject or the objective of the analytic exercise. Output features represent the result of the interaction of inputs. Consider, the input dimension elements denoted as $X_1,..., X_p$ and the resulting response element denoted as Y_0, the conventional dimensional model became:

$$Y_0 = f(X_1,...,X_p),$$

X_is were the symbolic attributes or the features of the subject of interest standardized to avoid mathematical issues with unit differences. *f* was the function expressing the association of the two sides of the relationship.

Additional assumptions of dimensional extension included the base quantities which constitute a subset of the inputs, denoted $X_1,...,X_t$, where $t \leq p$, to satisfy non-basis quantities, $[X_0],[X_{t+1}],...,[X_p]$ expressed by the combinations of the dimensions of the base quantities, $[X_1],...,[X_t]$, in the form of the power law. It is important to note that basis quantities cannot combine dimensions of other base quantities. Furthermore, assume that $[X_0]$ can be expressed by the combinations of $[X_i]$ for $i = 1,...,p$, otherwise it violates dimensional homogeneity. This assumption led to the existence of the basis quantities that may not be unique or independent. To address this, the transformation of the attributes uses basis quantities based on Buckingham’s Π-theorem. For example,

$$[X_i] = [X_1]d_{i1}...[X_t]d_{it} \text{ for } i = 0,t + 1,t + 2,...,p.$$

Consequently, the transformed quantities are

$\Pi_i = Xi((X^{-d_{i1}})_1 \cdots (X^{-d_{it}}))_t$ for i = 0, t+1, t+2, ... , p

where

$[\Pi_i] = [((X_i X^{-d_{i1}})_1 \cdots (X^{-d_{it}}))_t] = [X_1]^{d_{i1}} \ldots [Xt]^{d_{it}} [X_1]^{-d_{i1}} \ldots [X_t]^{-d_{it}} = 1$

The response function can be rewritten:

$Y_0 = f(X_1,\ldots,X_t,X_{t+1},\ldots,X_p)$.

Using Π_i instead of Y_i, the following expression resulted

$\Pi_0((Xd_0)_1 \cdots (Xd_{0t})_t) = f(X_1,\ldots,X_t,\Pi_{t+1}Xd_{t+1,1\ 1} \cdots Xd_{t+1,t\ t},\ldots,\Pi_p Xd_{p1\ 1} \cdots Xd_{pt\ t})$,

and

$\Pi_0 = (X^{-d_{01}})_1 \cdots (X^{-d_{0t}})_t \cdot f(X_1,\ldots,X_t,\Pi_{t+1}X_{dt+1})_1 \cdots (Xd_{t+1},\ldots,\Pi_p X_{dp1})_1 \cdots (Xd_{pt})_t)$

where f is the function to be estimated.

Or,

$\Pi_0 = g(X_1,\ldots,X_t,\Pi_{t+1},\ldots,\Pi_p)$

where

$\Pi_i, i = 0, t + 1,\ldots,p$ are quantities and

$X_1,\ldots, X_t$ are considered independent of one another

Based on Buckingham's theorem, Π_i represented the final expression of the output regarding the dimensions of the input.

Resolving the Research Questions

The research questions are restated here:

Research Question 1: Can data model extensions improve the discovery of management scenarios from big data?

Research Question 2: Can data model extensions improve the formulation of insights about the management scenarios?

Research Question 3: Can data model extensions express the complex constraints and rules needed to compose the acceptable and actionable solutions for analysts and executives?

Research question 1. To answer the research question 1, I examined the datasets that resulted from the use of the extended data model for the management scenarios. Within this data model, management scenarios were connections which expressed associations between sets of data elements about resources (for example, products), and agents (for example, sales representatives) in transactions (for example, sales transactions) leading to business outcomes (for example, profit margin). These were multi-dimensional association matrices of the semantic, symbolic, and dimension attributes of the available data. Within the multi-dimensional association matrices, the semantic attributes expressed degrees of similarity or dissimilarity to other values of the property, symbolic extensions established the congruence of properties to each other, and the dimensional extensions established the distance between relation instances.

Management scenarios, G*, such that each management scenario, G, was represented as

$$(\mathrm{V}, E = \{E_0, E_1, \ldots, E_m \subseteq (\mathrm{V}\ \mathrm{x}\mathrm{V})\}), \tag{11}$$

a multi-dimensional relational matrix where

$\mathrm{A} \in \{0 \ldots 1\}^{n.n.n}\{0 \ldots 1\}$ and

$$A_{i,j}^{k} > 0 \text{ if } (i, j) \in E_m: 1 \leq k \leq m \text{ or } 0 \text{ otherwise}$$

Thus, each extension data element E represented an adjacency matrix, and the combination of m adjacent matrices formed a complete expression of the management scenarios. Within this scenario space, the multidimensionality of the associations characterized the different conditions that apply to the scenario. Each condition became the unique path of connections with a unique set of functions and constraints which mapped to unique outcomes. Using the analytically extended data model representing all relevant multi-relational paths within the available data between all data elements, the paths with the same starting point and end-point formed the scenarios when there are discriminating combinations of initial value and end outputs. Based on this analysis, an analytically extended data model provided the relevant scenarios for management analysts and executives, through the extension attributes which connected inputs to outputs. Simply, a scenario was the path from a specified input point to a specified output point within the data model. The data model extension improved the discovery of management scenarios from big data.

Research Question 2. To answer research question 2, I studied the type of information any management analyst or executive would consider an actionable insight. Important requirements of insight were explainability, interestingness, and relevance. An insight was considered explainable if it was capable of being understood within the domain of interest. The capability to understand represented the alignment of the analytic results to the intuitive conceptualization of the subject of interest. That is, an insight was considered explainable if it connected well-established concepts within the domain of interest. When an insight revealed new connections and new concepts, the insight was found to be interesting. An insight was considered relevant by management analysts and executives when the connection to resources and agents had utility since these were the items the management analyst or executive could manipulate to solve the management case.

The extensions that connected the resource (for example, product, price, marketing, customers), agent (sales representatives, customer service representatives, product development, Pricing analyst) and transaction (sales) ontology elements addressed explainability, interestingness, and relevance. At the data level, the insights were the quantities that qualified the associations between the transactions and the interaction of resources and agents that provided the management analyst with clarity on where optimization opportunities existed in achieving the satisfactory outcome.

The extended data model captured the different levels of insights depending on the analytic objective of the exercise. The insight was descriptive if it provided perspective on current and historical occurrences. It was inferential when it provided information on one situation for the estimation of another. For example, inferential insight was diagnostic when it used existing information to provide a reason for an ongoing condition. It was predictive inferential insight when it used information of a current and past situation to make guesses about the future. The insight was considered a forecast when it took into account a point in the time in the future for which these assumptions could be realized all things being equal. The data model extensions improved the formulation of insights about management scenarios.

Research Question 3. As noted above, constraints and rules were expressions of different forms of logic. Constraints were logic of limits, and rules were logic of associations or projections. Complex constraints and rules are, therefore, n-order logic, in which lower order predicates or propositions were nested to create higher order predicates and propositions. As stated above, the classical relation could be considered the collection of first-order predicate logic at the attribute level, and first-order propositional logic at the tuple level. Data-model extensions grouped and nested the predicate and propositional logic in different combinations for the discovery, identification, specification, and resolution of management problem scenarios. In the extended data model, the logic of limits, associations, and projections were data points. The data model extensions captured complex constraints and rules needed to improve the acceptance of analytic outputs by management analysts and executives.

Application

This section contains the application of data model extension and extended data analysis methodologies to a big data analytics project scenario in a U.S. based global medical equipment manufacturer and distributor. A brief overview of the company and management needs set the context for the analytic exercise. The primary goal of the data analytics exercise was to identify the contributors to profit margin and overall growth of the company. The company consistently missed its annual revenue and profit goals, so management wanted the project to discover what was responsible for the situation and provide a recommended remediation plan.

Case Overview

The company was a huge manufacturer and distributor of medical supplies, uniquely positioned to provide products, education, and support services across the continuum of healthcare. It marketed as much as 100,000 products, including hospital furniture (bed, mattresses, seats, and tables of all types and specifications), durable medical equipment, housekeeping supplies, exam gloves and garments, and many others. Its customer base included hospitals, long-term care facilities, physician offices/practices, home health providers, and retail outlets. For large customers, they offered inventory, supply chain, logistics, technology and analytics and equipment customization solutions. Its more than 11,000 sales representatives marketed the products and services through some 200 distribution centers in 13 countries in North America, Europe, and Asia / Oceania. It operated a delivery fleet for high throughput routes and used delivery services to drop ship purchases as necessary. The company operated manufacturing plants in the China and Singapore. Its manufacturing plant in the United States closed about five years before this analysis.

Executive management was concerned about its stagnation of revenues and profits. The profit margins were much lower than peers in the same industry. The executive management considered its poor performance on key market valuation metrics, for example, price per share, earning per share, price to earnings ratio, price change, price change percentage, and market capitalization as indicative of the erosion of profit margins. They believed that their forward guidance of the market was responsible for the pessimistic view of the company by the market, the result being its low market valuation and stock pricing. Management believed that a better attribution of the profit margin would provide the tool to manipulate their operational activities to achieve higher levels of profitability which should command market valuation that was better than its peers.

The prevailing belief was that the company had a pricing issue, much more than a cost issue. An earlier profit strategy study proposed revenue estimates, targets and forecasts based on price increases and commission reductions only. Management wanted validation of these strategic proposals, quantitatively. They also wanted to design programs to achieve a consistent growth of the company and increase product footprint in existing and new customers over time which would translate into a higher market valuation of the company. They needed a comprehensive solution that can achieve profit margin expansion while minimizing the downside impact of price and commission changes on the customers and sales force.

Data

The enterprise transaction processing system (SAP/R3 Enterprise Resource Planning (ERP) system) of the company was the primary data source for the study. Additional data sources included data from GHX Market Intelligence, Distribution Feedback Reports, and Health Product Information System (HPIS) data which was the source for standardized health care product codes, along with competitive and complementary product information. Also, reports generated to support sales, customer, pricing, and product management contained information needed for analysis. Many of these reports were monthly and quarterly snapshots which had been saved off as documents for management use. Examples were active account reports provided as active_account.xls, Credit analysis by Reason Code report, provided as Credit analysis by Reason Code report.pdf, and so on. The data were representative of the complexity of enterprise data in modern organizations. Table 5 contains the summary of data assets in the study.

Table 5

Study Data Overview

Item	Value	Comment
The total size of data	15 terabytes (TB)	Considered Very Large Data Set (VLDS) for analytic modeling
Total number of fields in all data sets	700	A large number of fields means that the dimensionality of the dataset would be very high, which would make computability difficult
Number of datasets	137	All data sets would be linked to compose a complete universe of data asset for the analytic modeling exercise
Number of sales transaction records	1.7 billion	A large number of sales transactions means that the observation set for the analytic case of very robust, and it is likely to reflect the different mechanisms and subjects that underlie the data generation process completely
Timespan	3 years	More than one business cycle for analysis since systems aligned well with the calendar year

% of numeric attributes	30	As much as 210 numeric data elements are available as candidate variables
% of character attributes	70	As much as 490 characteristic variables are available candidate variables
The range of character attribute levels	1 – 5,000	A large number of levels of the natural classes, and potential explosion of dimensionality and contraction of degrees of freedom
Range of numerical attributes	-100,000 to +5,000,000	A large number of the cardinality of numerical attributes. Negative numbers meant quantities or money flowing in the opposite direction of what was expected. For example, negative order quantity was order quantity returned by customers, negative payment amount was amounts returned to the customers

The total data size was 15 terabytes, made of 1.7 billion sales transactions over three years. A total of 139 datasets and documents were available for the analysis, with 358 data elements. A ratio of 3:7 of numeric to characteristic attributes. The characteristic attribute levels ranged from 1 to 5,000, while numeric attributes ranged from -100,000 to + 5,000,000. Table 6 shows this diversity of data assets.

Table 6

Data Formats in Input Dataset and Documents

Data asset Type	Data Asset description	File source	# of data assets
DAT	Database output file	ERP data archive	1
DBF	Database file	ERP system	59
Mdb	Microsoft Access Database file	User Application	1
Xlsx	Microsoft Office 2000 Excel File format	Extracts from ERP and other systems	1
Xls	Microsoft Office Excel File format	Extracts from ERP and other systems	20
CSV	Character separated values file	Extracts from ERP and other systems	2
PDF	Adobe Acrobat Portable document format file	Internal and third party reports	16
RAR	Compressed file	Data file archive	14

SPOOL	Database or Application output file	ERP and other applications	1
DOC	Microsoft Office Word File format	Reports and analysis	1
DOCM	Microsoft Office Word Macro File format	Reports and analysis	1
TXT	Text file	Extracts from ERP and other systems	23
		Total number of data assets	140

As noted in the table above, available data also included several documents with unstructured or semi-structured information. The documents contained information needed for the analysis, so it became necessary to extract this information from the documents. I converted PDF, DOC, DOCM documents to unformatted texts documents before using the Open SourceText Mining algorithms within Pentaho Data Integration to process and ingest the data into the PostgreSQL database, which was capable of handling structured, unstructured, and hierarchical data representation. The conversion of the different data formats were vital activities of the data analysis process. The meticulous process of converting the non-structured data assets into structured data facilitated their integration with the structured data was an essential and significant undertaking. It is worth noting that the need for a document database like MongoDB or Graph database like Neo4j did not arise as the relational database selection had capabilities of handling these structures as relational constructs.

Appendix A displays the data use agreement obtained from the study showing the detail list of available data. Table 6 summarizes the data formats used in the study, and

Figure 3 shows the data diagram of the data-files, documents, and database extracts. The data diagram guided the arrangement for further data model development, by organizing 140 different data assets (data files, database tables, and documents) into a scheme with the linkages between the data assets. The diagram highlighted data assets from the SAP R/3 ERP system, which contributed most of the data. The non-SAP R/3 data assets included those from GHX Market Intelligence, for example, Distribution Feedback Reports and HPIS product data.

The data diagram represented groups of data assets with similar structure and source. The items in the box were the instances of data assets. For example, SALES ORDER data were in data sets of monthly data because of the size of the files, so were pulled into a single group. In this situation, it was necessary to break up the extraction process into monthly chunks for massively parallel extraction routines which minimized the window needed for the extraction process.

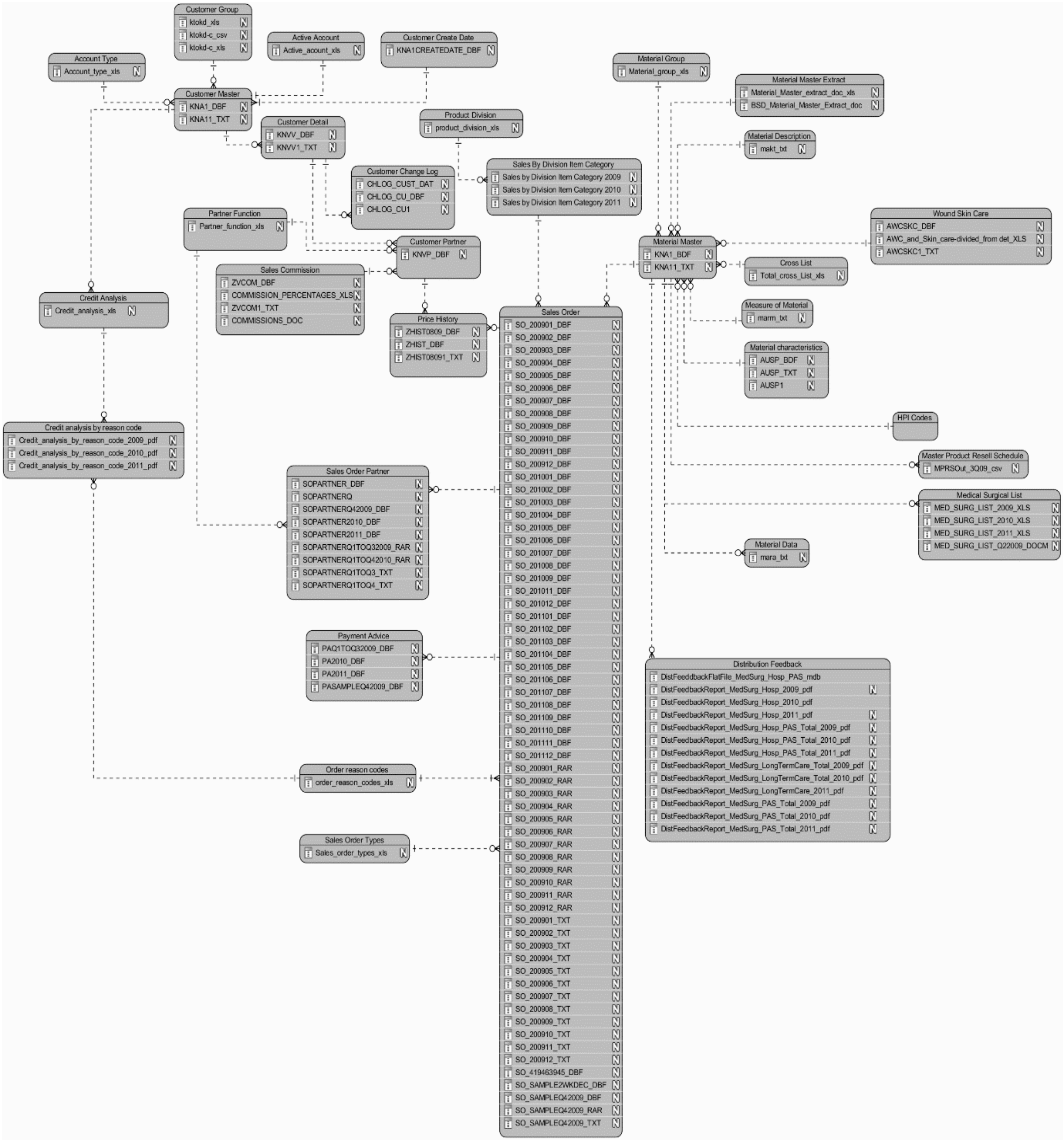

Figure 3. Study data asset diagram.

A significant number of the data files provided for the analysis contained sales order data (59 data files). Distribution feedback reports from an external source came in 14 files, and price history data came in 9 files. These three data file groups made up the top 3 datasets for the analysis.

Data Modeling

Base Relational Model. Each data file group was further processed to identify entity or entities it contained using the entity recognition (ER) algorithm described in the data collection section above. The ER algorithm produced classical and associational relations in a 3rd normal form with its primary key column and any foreign key columns identified. This output generated the data model made up of 30 entity relations, connected by 32 referential constraints. For the data dictionary of the data model, see Appendix B.

The generated data model addressed cardinality between the entities of the data sets within and between the datasets, as well as duplicates in the join keys that can inadvertently result in Cartesian unions. The data model brought together the available data from different sources and domains into a rationalized framework addressing data integration issues.

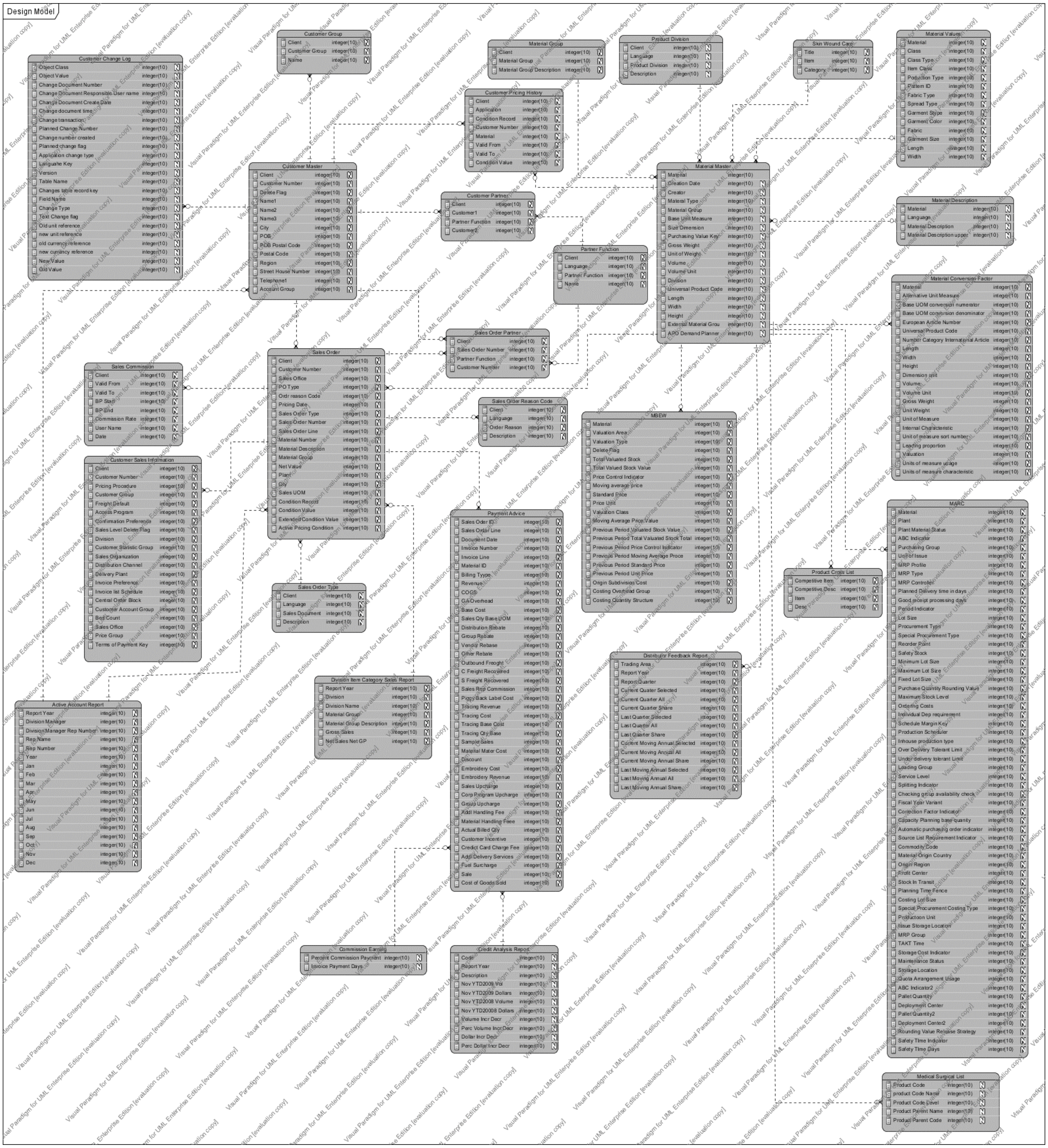

Figure 4. Study data model.

Data model extensions

Semantic extension of the data model expanded the attribution of relations for characterizing profit margin of sales from the universe, shown in Appendix E. I derived semantic extensions for each of the non-numeric attributes. I encoded the attribute relations to capture occurrence and non-occurrence of the attribute value. The figure below illustrates an example of semantic relations for PAYMENT ADVICE and the SALES ORDER PRICING relations including, BILL TYPE, BILL DATE, SALES ORDER TYPE, SALES OFFICE, SALES ORDER REASON, PLANT, PRICING CONDITION CODE, PRICING DATE, and SALES UNIT OF MEASUREMENT. The use of these semantic relations generated matrices of Boolean, nominal, ordinal, or ratio scale values of the underlying subjects within the sales domain.

I further extended the data model by adding symbolic elements. As noted in the section above, these symbolic elements are specific to the analysis problem under consideration. I determined items derived from domain of interest, listed in Appendix D. For example, the symbolic extension for the ORDER entity captured the additional attributes, for example, price blocks, returns (orders with negative sale amount), promotion sales (sale with special price type), samples (sales with zero amount and pricing type is sample), new sales reps (sales reps with tenure less than 1 month), specialized sales reps (sales reps with doctorate degrees for specialized equipment demonstrations and sales), and so on and so forth.

Finally, the dimension extensions captured the further expression of the enterprise as abstract units. For example, the frequency of the order, order to order size change, order to order price change, change in price impact on customers, price change impact on orders, and so on were dimensional extensions. It is typical for these expressions to be dimensionless (i.e., devoid of units) so that their applications are not constrained, and so that the quantities represent the absolute value of relative quantities. Ratios, percentages, coefficients, moments were used to formulate them in the data model. The final analytic data model resulting from the implementation of the extensions discussed above for margin expansion and growth was extensive, and too large for display here. Using the extended data model, I constructed datasets that made the profit margin the subject of interest or the outcome (target or dependent) variable of the data model. All other classical attributes and the analytical extensions derived from the available data were the indicator or input variables. Appendix I shows the catalog of analytic processes that are useful for continued formulation the management problems and solutions to arrive at the results in the management analysis and recommendation discussed below.

Management Analysis and Recommendations

Figure 5 below shows the conceptual determinants of profit margin, along with their interaction effects. I constructed the determinant from analytic extensions of the base data model. To determine the contribution of the different attributes to the profit margin, I utilized a random forest regression method. This analytic formulation method was selected since profit margin had a noncontinuous, nonlinear association with attributes within the management domains. As shown in the figure, there were no dominant contributors to the profit margin levels in the available data. The factors contributed between 1.98% and 3.50%, as such management intervention had to be broad to achieve any effect.

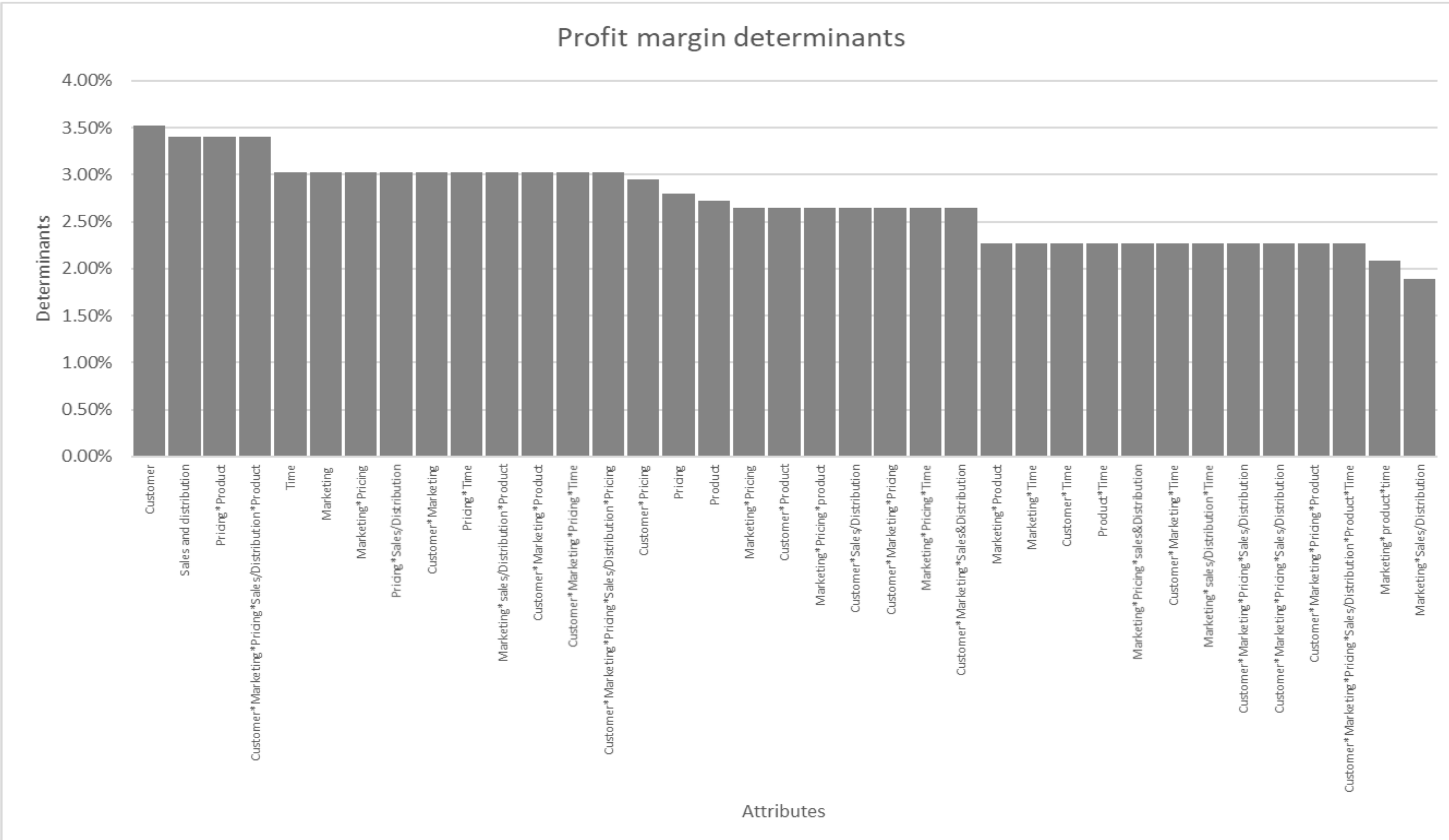

Figure 5. Profit margin determinants

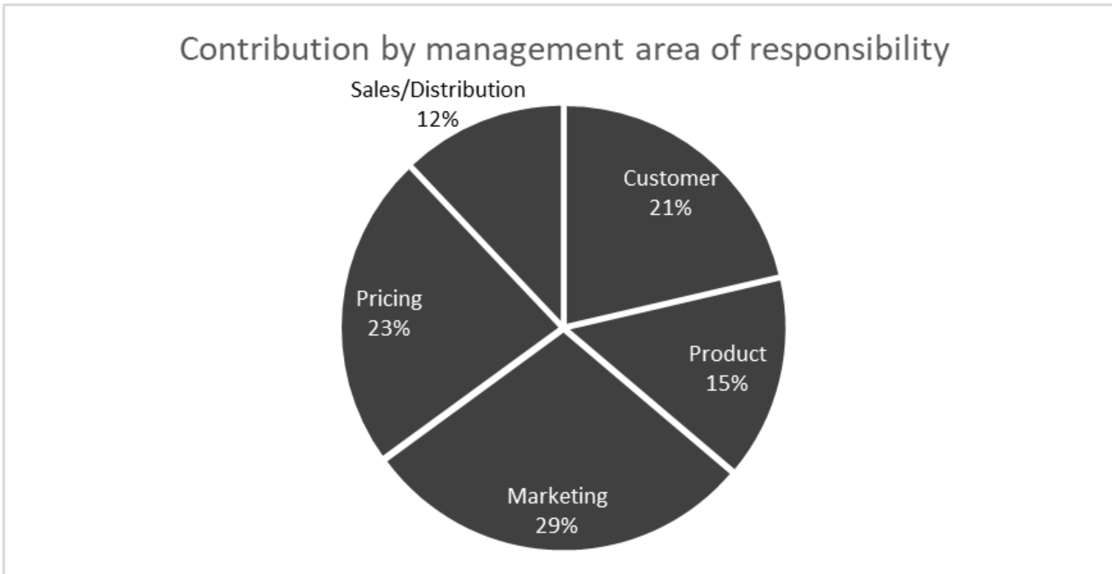

Figure 6. Summary of determinants by management area

The summary of the determinants of the profit margin by the different management areas showed marketing the lead contributor. The others area in order were pricing, customer and finally product management areas.

The profit margin coefficients estimated the impact of each of the concepts in the figure below. I, also, extracted these contributions using a random forest regression algorithm.

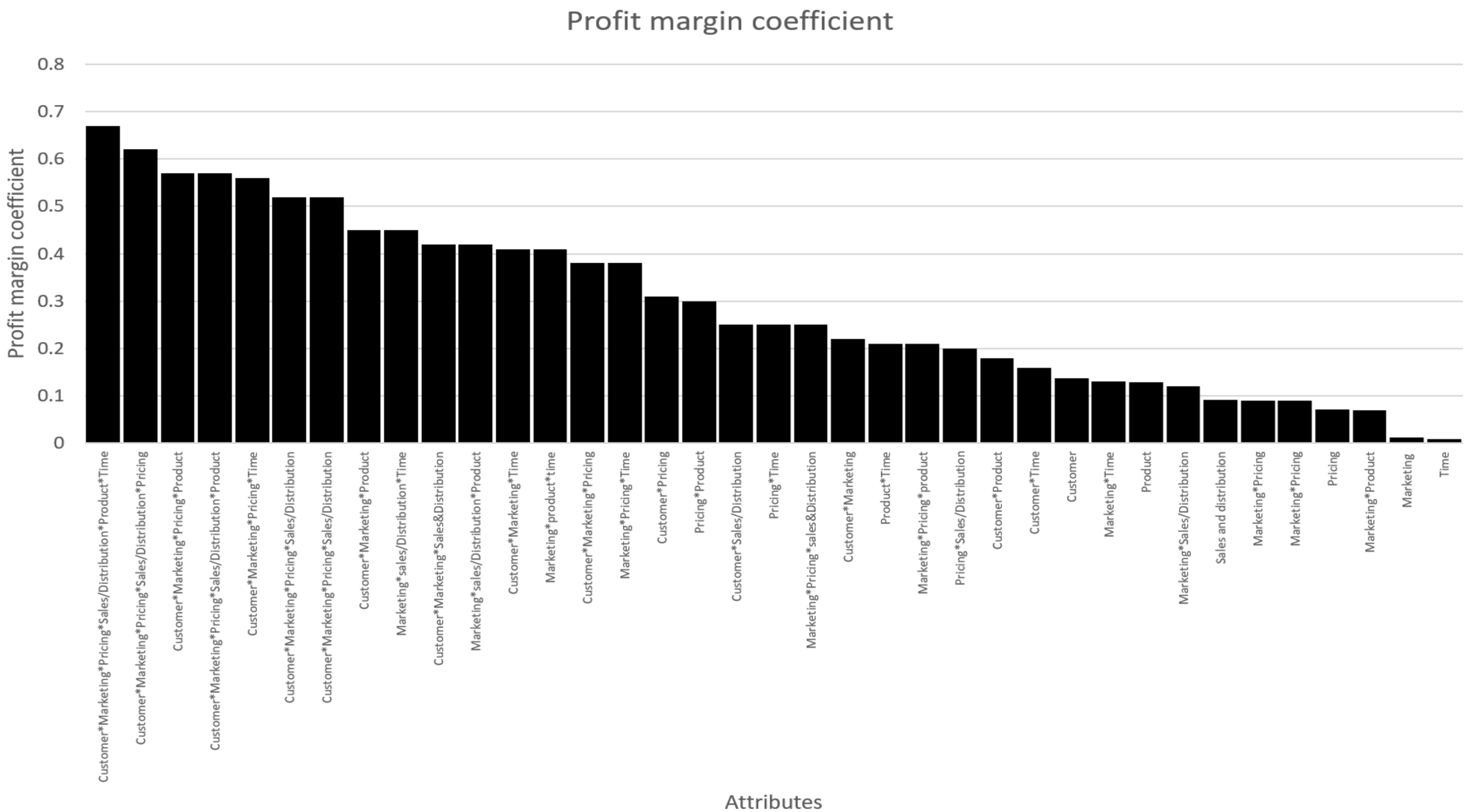

Figure 7. Profit margin coefficients

Figure 8. Profit margin coefficient by management area

Figures above show the impact of the concepts on profit margin level (determinants) and the profit margin change (coefficients). The determinants identified the critical aspects of the business driving profit margin levels. The profit management coefficients identified the contribution of each management area to the increase in the profit margin. Combining these two measures created the following order of influence on profit margin: marketing, customer, product, pricing, and sales/distribution management areas as shown in the pie chart below.

Figure 9. Management area influence of profit margin

The following charts show further decomposition beyond the management areas based on the key concepts derived from the ontology learning, data engineering and the analytic formulation methods against the available data.

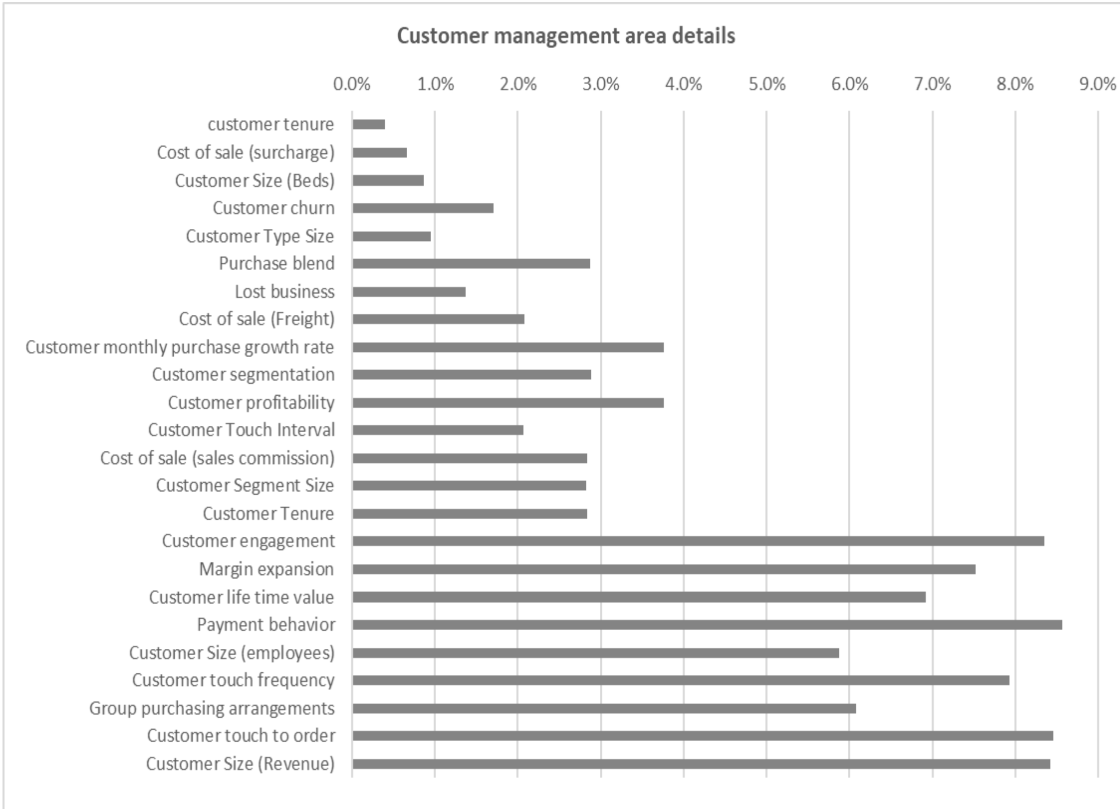

Figure 10. Bar chart of customer management area details

Figure 12 above represents the decomposition of the customer management area into the concepts discovered in the area. Within the customer management domain, the most important contributor to the profit margin was the customer size based on their revenue. There was also the tendency for sales agent engagement with the customers to produce orders, especially, when the company collaborated with the customer to develop specific products for the customer, for example, special hospital beds for geriatric patients and disabled patients and so on.

Figure 11. Bar chart of pricing management area detail

The pricing management area decomposition identified the impact of different pricing concepts. The price elasticity, relative price of a product to comparable products, the type size price differential as well and the revenue leakage based on pricing policies all impacted profit margins and profit growth.

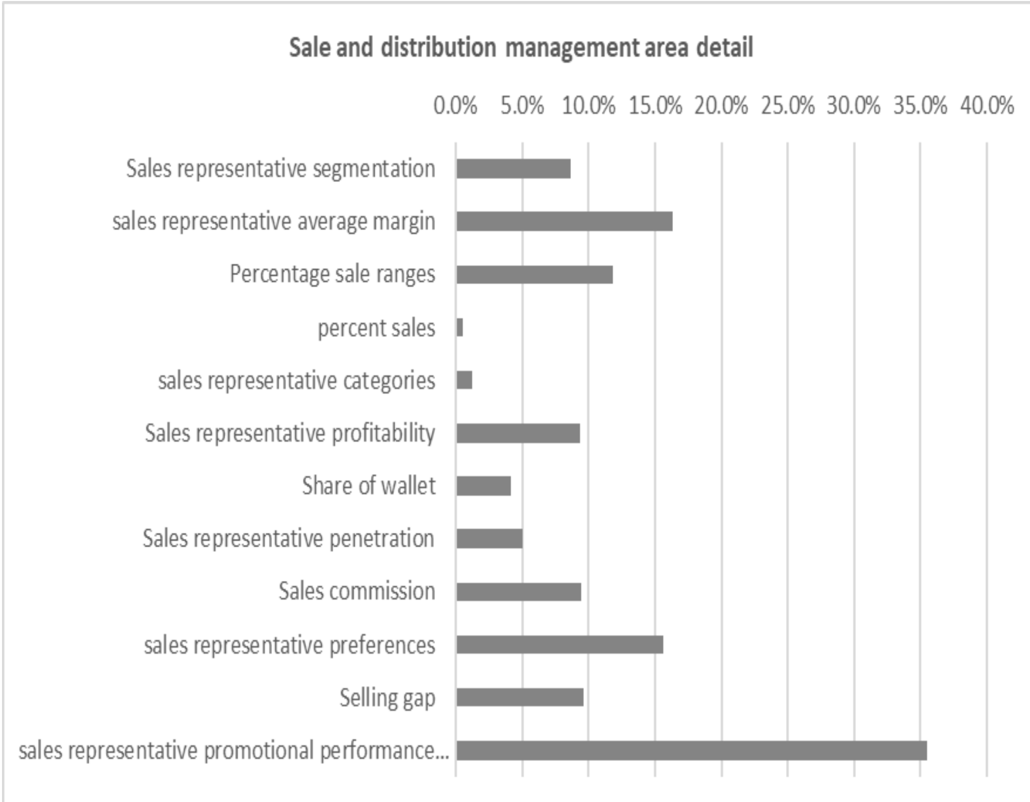

Figure 12. Bar chart of sales and distribution management area details

A look at the sales and distribution management area revealed that the primary concept impacting the profit margin was the sales representative promotional performance. These were promotions initiated directed by the sales agents in collaboration with the sales and marketing team. The average margin per sales representatives was significant the overall profit margin. Also important was the preferences expressed by the sales agents related to the products they were responsible for driving sales.

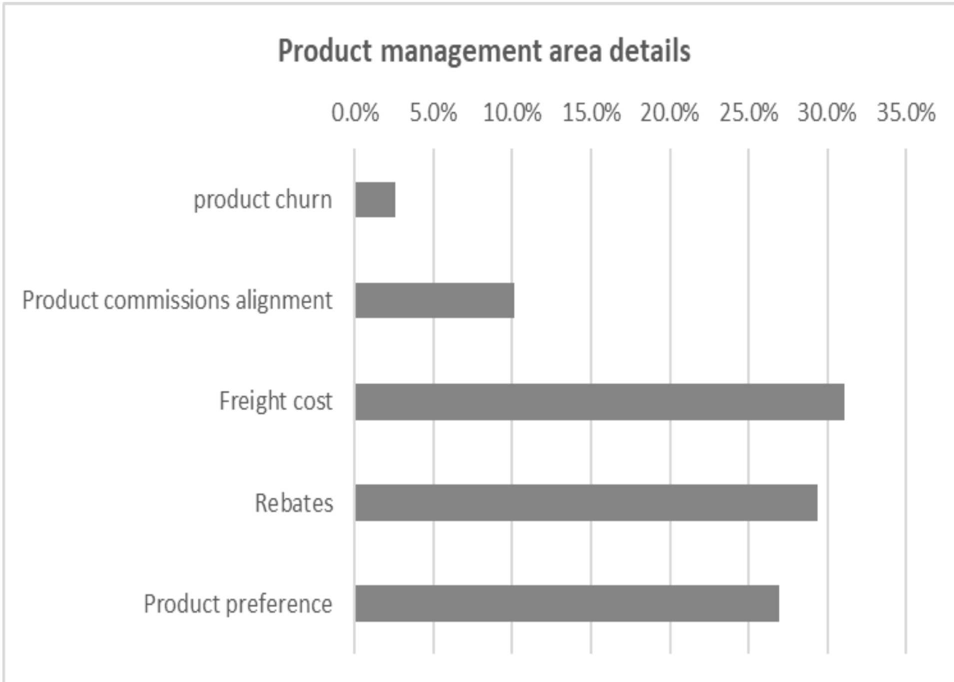

Figure 15. Bar chart of the product management area

The product management concepts of importance where the freight cost which seemed to add to the overall invoiced cost of sales, rebated products was also critical to the profit margin improvement of the company. Rebated products were products that were distributed by the company. The rebated products turned out to better priced than products the company manufactured.

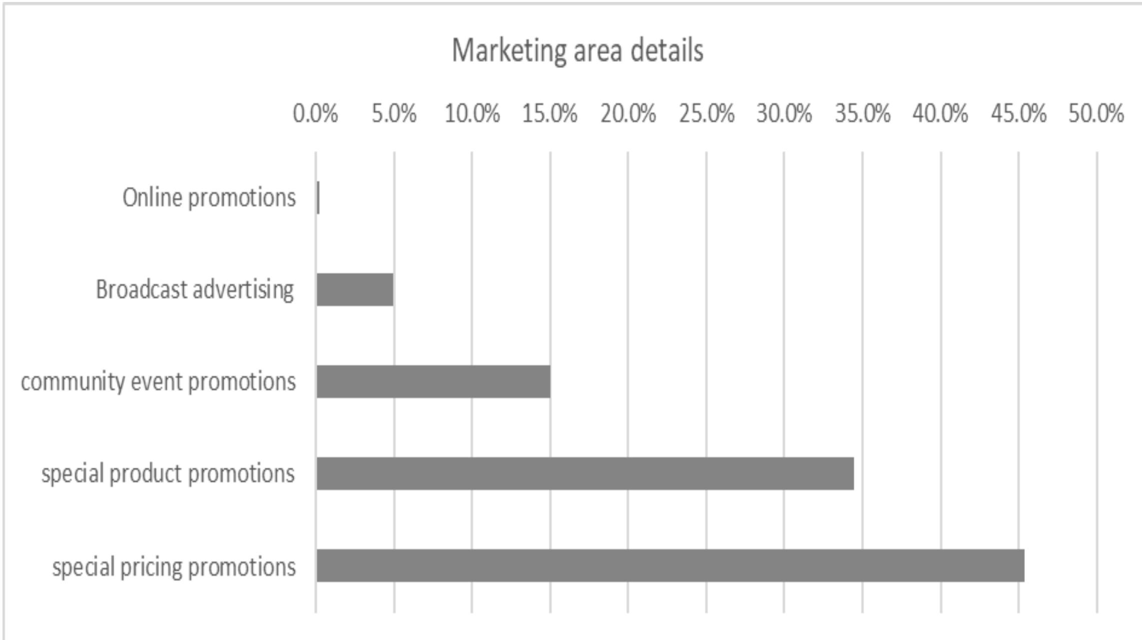

Figure 14. Bar chart of the marketing area details

Marketing area details showed most important concept in this area was the special pricing promotions conducted by the company, followed by the special products promotions. Also important was promotional activities at community events.

Based on the analysis conducted on these management area concepts the following programs were recommended for implementation. Each program had well-defined outcomes expectations:

Special pricing and product promotions with sales representatives – The management analysis above indicated that marketing contributed poorly to the profit margin growth. It also highlighted the most effective marketing promotions in achieving improvements in profit margin to be special pricing and special products promotions conducted with the sales agents within each of the sales territories. This program proposed a marketing process that gathered input from the sales representatives in each territory to determine the best approach, the product and the potential prospects to target. The manager of each of the sales territories established targets for the number of promotions to complete and the return on investment to target for continued investment on marketing within that sales territory. This program also identified companies to target for marketing, sales, and investment activities. It included actual investment in customers by extending products credits and allowing tiered payment cycle of 3, 6, 9, 12, 18, and 24 months to help improve the cash flow situation and growth of customers.

Group pricing arrangement and rebating program –This program administered group discount pricing arrangement to ensure compliance. The program monitored the volume, identifying customer groups that did not meet agreement for a rebalancing of the price to the actual volume. If the group exceeded the volume arrangement, rebate or credits were triggered. Not meeting the conditions of volume arrangement triggered reverse rebate or debits from the customer to the company. This program improved margins by 13% in the first year of implementation.

Onsite supply management program expansion: This program was implemented for the very large clients, to ensure retention. The company accepted responsibility for supply management, for exclusive multi-year supply arrangements. This program became so popular that it became a standard offering of the company. This program resulted in 20% increase in product penetration in existing clients, as well as 80% retention of rapidly growing clients

Pricing block improvement program: This program addressed the problem of the price blocks which occurred when the offer price is much below the standard price. The sales representatives used this to lower the price so they can get a volume that allowed for their commission to be competitive. Price blocks also delayed the delivery of items to the customer who added to challenges of customer satisfaction. This program established the policy that all price blocks should be cleared within 1 hour of the occurrence of the block by the sales representative or escalated to the Regional pricing manager. This lead to the recapture of an average of 7% of the profit margin which was eroded by price blocks.

Product manufacturing improvement program – this program targeted manufactured products to determine how to make the manufacturing more competitive. The target was for products manufactured by the company to be cheaper by about 20% so these products can compete effectively. This lead to many manufacturing strategic decisions, including outsourcing of manufacturing operations which allowed the achievement of the objective of getting manufactured products to 20% of the cost of comparable distributed products

The case overview showed that using analytic extensions to the data model to derive semantic, symbolic and dimensional attributes related to the profit margin problem, it was possible to apply advanced analytic processing techniques to discover the management scenarios underlying the problem. Though the business problem was vague, using the analytic extensions to enhance the data model, I was able to construct analytic attributes to represent the concepts described in Appendix D. These concepts like customer size, customer tenure, customer payment behavior, and many others were better at representing the management scenarios responsible for the problem. In relatiom to research question 1, these data model extensions improved the discovery of the management scenarios of underlying problems from big data.

Using the management scenarios discovered, it was not very difficult to connect the management scenarios to management insights needed to address these challenges imposed by the scenario. In the case above, the insights about marketing that led to the recommendation of special pricing and product promotion campaign with the sales representative at key clients was from the finding that this integrative approach to marketing contributed more to profit margin than other forms of promotion. About research question 2, the data model extensions and the additional analytic processing improved the insights about the management scenarios, and provided credible explanations and solutions for the problem under consideration.

The use of analytic extensions enabled the construction of attributes that captured complex rules and constraints needed to represent the domain knowledge. Analytic attributes like order to order interval, order to order quantity change, order to order price change, and many others allowed the capture of the complex business rules and constraints related to the behavior of the different participants in the transactions. Also special policies related to price block management were reflected in the data by identifying transactions in which these policies contributed negatively to the management situation under consideration. Therefore, the answer to the research question 3 was that it was possible to use data model extensions to represent complex constraints and business rules needed for the composition of acceptable and actionable solutions for analysts and executives.

Summary

In the study, I demonstrated that the use of the analytic extensions improved discovery of management scenarios, insights about these scenarios, and the representation of complex business rules and constraints needed compose acceptable and actionable solutions for business analysts and executives   The use of analytic extensions supported the realization of quantities for management analysis. The approach expanded the representation of information for management analysis and reduced the complexity of the model. Using different analytic formulations, I was able to define and operationalize critical concepts within the management domain needed to formulate solutions for management analysis and decision-support. The concepts I derived and quantified using analytic extensions to the data model captured difficult and complex conditions and constraints existing in the domain of interest for analytic problem-solving. Management problem-solving required the design and execution of business and technology programs to address the conditions and constraints within the enterprise preventing the achievement of desired outcomes. The need to improve the utilization of data in the design of management processes continued to increase with improvements in data gathering, storage, and retrieval techniques. Through significant work had been done in the construction of statistical databases for very large datasets as well as in knowledge discovery from databases, using data models to formalize the data architecture for these solutions remained a gap.

In this study, I worked on extending the classical relational data model with attributed with specific ontological commitments using semantic, symbolic and dimensional expression forms. While the classical data model saw the attributes as a primitive expression of the subjects within the enterprise domain, this approach of implementing extensions to the data fostered the capture of concepts which represented patterns, profiles, features, and facets directly within the data model.

This approach to the extension of the attribute space simplified the analysis of the contribution of the different elements to the behavior of the domain of interest. An illustration of this approach to management problem solving in the medical products distribution company led to recommendations that were well accepted by analysts and executives in the business. The programs included special pricing and product promotion campaigns with the sales representatives to expand market share, group pricing arrangement and rebate program monitoring to minimize profit leakage. Other recommendations included Onsite supply management program to increase customer loyalty, active price block administration to minimize inadvertent underpricing and overpricing scenarios, and product manufacturing process evaluation to target manufacturing cost for some of the products that were being cannibalized by rebated distributed products.

These recommendations aligned to the intuitions of the business analysts and executives. The approach avoided the issue of the use of esoteric technical and assessment methods with limited business and management value. In empirical management analysis, there was no value in comparing the results to chance or theoretical distributions to determine the significance of the problem or the outcome expectation. In classical research, the statistical power and significance of the variables are basic requirements. Using the data model, I was able focus analysis and recommendations on business impact of the attributes within the management domain. The validation of business effects of attributes was critical for the executive decision maker. These business effect estimates were important drivers of the design, execution, and administration of management programs that transformed the company to profitability.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The purpose of this quantitative nonexperimental descriptive DBR was to examine data model of a typical enterprise data analytics project to determine data model extensions that would improve the formulation of management problems for analytic processing. I focused on a typical data analytics project in a modern data-rich organization. These projects dealt with very large and complex analytic scenarios expressed with big data. The management analysis and decision-support requirements were ambiguous and sometimes unknown. This situation made the classical data analysis process and analytic processing techniques unsatisfactory. Hence, there were high levels of failure of these projects in the fulfillment of management needs to resolve business problems through well-informed recommendations that were acceptable and actionable by management analysts and executives.

In this chapter, I interpret the findings and the limitations of the study, followed by recommendations for further studies into the business knowledge discovery and modeling for management analytics and decision-support research. I conclude the chapter with a summary of further research opportunities for data analytics and decision-support in management.

Interpretation of Findings

Contribution to Knowledge and Research

In Chapter 2, I reviewed the literature on data-modeling for analytic processing. I also discussed the challenges of increasing complexity of the data and the size of analytic scenarios in data analytics projects.

The data analytics started with the static composition of data as reports and the use of reporting databases. The static outputs evolved to more functional expression in data warehouses, data marts, and business intelligence systems. In the last decade, more sophisticated analytical expression based on statistical and mathematical methods in software packages provided important advancements to the data analytics practice. The most recent progress has been in programmatic or computational expression using algorithms to evolve logic from associations within the data.

Despite this progress in analytic solution development, challenges remained in knowledge discovery, business intelligence, and decision-support for management problem-solving. Significant gaps existed between data, management problems and analytic solutions proposed. In this study, I demonstrated an approach to the problem with big data analytics with progressive transformation of the data and the creation of extensions to the data model for management problem formulation. This approach also allowed management analysts to apply the analytic insights in the composition of solutions for management problems. In this study, I emphasized the data model to establish the boundaries of analytic transformations and search for associations in the data. The absence of this analytic transformation boundary was the critical gap with existing algorithmic analytic processing approaches, which tended to create solutions that were difficult to translate to management programs which were needed to address business problems.

The confluence of big data, advances in analytic algorithms, and abundance of computational power provided the opportunity for transparent enterprise empirical modeling for intelligent management. This situation buffered issues, like (a) concerns of representativeness from using part of the data (sampling), (b) the need for theoretical distribution to estimate parameters or probabilities (curve-fitting), (c) the curse of dimensionality requiring variable selection, (d) the need for data fabrication or imputation of missing values to fill gaps in the data, and (e) the need for data reduction to match computational power availability that are important considerations on existing data-analytics projects. With these advances, the primary challenge of research in applied management and decision science becomes the design of data analytics processes that overcome legacy scenarios of limited data and computability. The approach to the study of this problem was to extend the data model to enhance the expressiveness of the underlying schema for the formulation of management problems and the design of solutions to these problems. The analysis indicated that this approach improved all types of analytic solutions developed to support management.

Data solutions supported the creation of the exact schema for the problem and solution scenarios. Analytic extensions to data solutions improved heuristic solutions by enabling the discovery of exact rules to replace approximate rules of heuristics. It also benefited analytical solutions which depended on exact theorems (or formulae, functions) by identifying the right combinations of propositions that make up the theorems. Numerical and computational solutions' dependence on exact procedures and algorithms required on proper representation of the information in the data model to support the different permutations of logic that make up the algorithm. In general, data models and their extensions enhanced the creation of relevant schemas with relevant rules and theorems and connections, which improve the algorithms and computational solution generation.

Enterprise data was fraught with complexity imposed by the data generation process including the lack of explicit connection between cause and effects, functional dependencies and associations. Data model extensions provided the tools to realize these embedded features for management problem-solving. Accurate insights on the performance of enterprise functions on value delivery to the marketplace were critical to sustaining viability. To this end, the perspective of enterprise outcomes should neither be completely random nor completely systematic. If the former were the case, management would be at the mercy of nature, locked in a game of chance, governed completely by the statistical and probabilistic processes. Conversely, if the latter were the case, management would be a pure game of quantitative choice governed by deterministic mathematical and numerical processes.

The findings of this study support the consensus in the literature that management is a game of strategy involving the creative design of enterprise programs to guide the interaction of resources and agents to create events and transactions for the fair exchange of goods and services (Colman, 2016; Weirich, 2017). The formulation and execution of this game were, therefore, the most critical activity of management, and defined the management actions in specific problem-solving situations.

Technically, management problems are constrained optimization problems of the form:

$$\int \begin{array}{l} \text{min|max } f(a,\ b,\ \ldots) \\ \text{subject to } g(a, b, \ldots) >= 0 \end{array} \qquad (10)$$

These were problems of integration of functions that minimize or minimize multiple objectives subject to constraints. The resulting complex Lagrangian functional represented the generalized coordinates with partial derivatives expressing changes in underlying variables and interactions over time. Figure 17 shows a conceptual diagram of its data model.

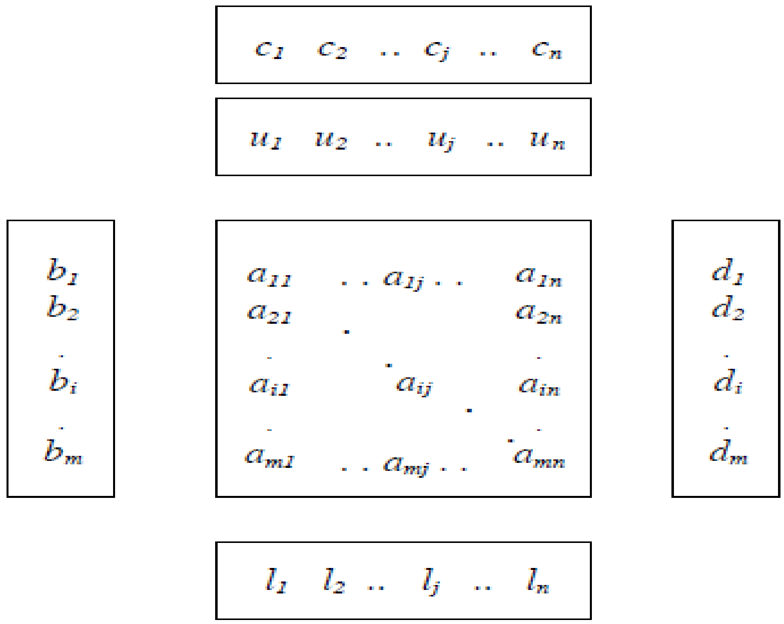

Figure 14. Generalized data model for the management problem formulation.
Note: a – attributes; c – coefficients, b, d – constraints; u, l – boundary data

Essentially, the data model was sets of attributes, coefficients, constraints, constants, and controls for each objective within the management domain. Because many of these were not the natural attributes of the domain of interest, their derivation depended on evolving them from data available, hence the need for the data model extensions.

Contribution to Data Analytics

As noted earlier, the use of schema-based analytic solutions was responsible for the rapid adoption of data warehousing and business intelligence systems in the last decade. The functional reorganization of data resulted in the adoption of analytics and decision-support technology in the enterprise. Differentiation of relations into fact and dimension relations provided rearrangement of the data for exploration. Unfortunately, their implementation in OLAP tools limited the application of advanced analytic programming.

The approach of analytic extensions to the relational data model discussed in this study overcame the constraint imposed by OLAP. With these extensions, the reorganization of the database schema was unnecessary. The additional translations of the data were layered onto the basic relational data schema, to enhance the representation of the underlying information. These layers of transformation contain the semantic, symbolic, and dimensional attributes needed to express similarity among values of the property of a relation, the congruence between two or more properties of a relation or the association between two or more relations in the data model.

The semantic extension focused on the logical continuation of values of the attributes and expressed the atomic concepts of the data. As discussed in the previous section, this involved implementing a data encoding process to derive variables which continue expression of concepts as arrays or vectors. This extension eliminated the fixed fact and dimension relations. Semantic extension considered any attribute as a fact or a dimension depending on the objective of the analysis. Analytic problem solving became much more flexible than currently possible with the OLAP multi-dimensional data model. These extensions allowed question-answering regarding values of the attributes represented by the data.

The focus of the symbolic extension was on connecting classes to alternative intentional logic to expand their expression. This extension was useful in imposing equivalence over property expression space to answer complex questions. That is, this extension organized data for the interpretation of association of the sets with breaks in semantic continuity but where there was congruence. The extension was akin to organizing characteristics of various levels or states of expression of a dynamical system, such that each level or state was a shift from another level. Symbolic extension fostered innovative aggregations of data allowing sophisticated description and redescription processing for profile classification, niche finding, analogical reasoning, story construction, schema matching to name a few.

With the dimension extension, the focus was on identifying the empirical dimensionality of the subject of interest based on the data. For example, the distribution of customers at every price point became the customer dimension of the price. This perspective of dimension was different from the classical definition of dimension in multidimensional modeling or dimensional analysis. In multidimensional modeling, the concept of a customer dimension for the price was not achievable at the attribute level.

These data model extensions made the answering of questions using the data directly possible at the all levels of knowledge and business intelligence: strategic, tactical, operational, and transactional. The schemes reflected the precise empirical ontology of the enterprise as proof systems (theorems) or automatic procedures (algorithms) for problem-solving. Using them, business analysts and executives could compose programs, test their feasibility, assess the expectations, and estimate the benefits. Data-driven and result-oriented management program development created predictability and efficiency in the practice of management.

Contribution to big data management research

Another contribution of this study was the application of the design-based methodology to data analytics in management research. The typical management research methodology advocated a process of identifying the problem, formulating the research questions, operationalizing the research questions as hypotheses, and identifying the variables for which data can be collected to test the hypothesis. Where necessary, the researcher designed the experiment and created the measurement instrument for the research. The researcher then gathered data, applied analytic techniques to fit the data to theoretical distributions, and determined whether the evidence in the data was significant.

This DBR started with the data and then learned from the data what was useful in solving the problems presented by the interaction of factors within the data generation process. With DBR, the problem did not need specification at the start of the research. The requirement was to learn the problems and the solutions from the data or direct manipulation of the data generation process or the learning environment. The learning requirement made the availability of big data, advanced analytic algorithms, and computing power critical to the advancement of this emerging research methodology. This methodology was robust to address the data analytic problems that increase in complexity with the nature of analytic problems. This research approach challenged variable selection, model selection processes, sampling, data reduction, data treatment applied to achieve better performance in model results, and many other research practices in the contemporary scientific inquiry. It also challenged the use of mathematical solutions and statistical routines. Mathematical solutions were needed when there was no data to express complex function. Statistical routines were useful when the data was not enough to support an assumption of accurate population representation. In the modern data-rich organization, none of these situations existed.

The use of DBR in this study illustrated the opportunity in using the data to discover and express issues existing in any domain of interest. It also demonstrated the use of the same data to seek solutions to the problem that would satisfy the end users of the analytics. Through the iterative transformation of the data, it was possible to quantify many of the concepts for cognitive processing of the domain of interest. The use of analytic extension eliminated the need to persist logic in the form of mathematical expressions. Rather these can be converted into attributes in a data model that can be analyzed and used in decision making. The use of the data form rather than the functional (or mathematical) form of the expression improved the interpretation of the results and the acceptance of the recommendations that were derived.

Limitations of the Study

The limitations that arose from the study regarding the generalizability, validity, and reliability of the research design, research methodology, and the study outcomes are discussed in this section. These were the issues discovered only after all the data had been analyzed.

The use of data model to broaden the characterization of the domain of interest for management problem-solving limited the solution scope to the available data. Influences that could not be extracted directly or indirectly from the available data were not considered. Anecdotal evidence that could not be substantiated with the business data could not be included in the analysis. For example, in the illustration, management analysts and executives believed that the nature of group purchasing contracts and arrangements contributed to revenue leakage. Since contract data was not available for analytic processing, their potential influences were not reflected in the recommendations and the resulting management programs.

The data model was the consequence of the data generation processes and the controls established within them to ensure the accuracy of the information captured. The nature of the data generation process, also, determined the representativeness of the underlying mechanisms and observations about the subjects within the captured data. Therefore, results of the analytic processes were limited by the context, content, and the relationships within the data generation process. These in turn limited the solution proposed for management problem-solving. This limitation was moderated by the use of big data which ensured the inclusion of all the data elements gathered about the subject of interest without consideration of methodological and computational needs.

The selection of the data from medical equipment manufacturing, supply, and distribution company as the source of the data for illustration of the data modeling approach was a consequence of the objective of the study which required the use of big data. The complexity manifested in the large number of products marketed by sales representative and the different classes of customers and markets in the United States. This selection of this industry was a natural and unavoidable limitation of the study. However, this selection limited the relevant business concepts to those of the industry. For example, there were different types of medical equipment and supplies for many different medical management scenarios. Some of them were used for therapeutic and others for diagnostic purposes. The equipment required different levels of skills from the sales representatives. These differential characteristics of the products had to be explicitly modeled for management analysis and decision-support. However, only those differential characteristics that were influential within the data set were reflected in the analytic results and management action recommendations. The specific extended data model constructed for the management problem-solving may not generalize to other management problem-solving scenarios within the industry or to other industries. However, the modeling approach which expanded the conceptualization of subjects and their alignment to the cognitive model of the domain of interest improved the management analytics and decision-support problem-solving in general.

Furthermore. the data model for management analytics and decision support for profit margin problem solving may only be limited to similar situations of profit margin optimization with complex interaction of products, pricing, customer, sales and marketing characteristics. However, the description of the methodologies adopted for ontology learning from available data, the application of data engineering to quantify abstract concepts, and the use of analytic formulation techniques to determine the functional association between sets of attributes have broader application.

The selection of the data was representative of a typical big-data environment with data size of more than 1 terabyte. The selection of large data meant that concepts would occur at a frequency that were statistically powerful, and therefore, relevant for management analytics and decision support. The approach of analyzing all the available data, instead of a subset of the data required the construction of a plethora of measures and metrics at different levels, such that one level can be linked to the next. This resulted in an architecture for the measures and metrics comparable to neural networks (Zelinka et al., 2011). The difference was that with this approach of extending the data model, the analysts would have control of the types of transformations within each layer. Although, this may prevent erroneous transformations, it also limited the transformations applied to those that are interpretable in business terms. As such, in situations where unconstrained transformations were allowed, different analytic outcomes and recommendations could result. Experience with unconstrained transformation was that they sometimes included transformation that do not have management analytics and decision support value (Zelinka et al., 2011).

Apart from the limitations of the study discussed above, there were no other limitation of the study that arose from the study. The acceptance of the management programs that resulted from the analysis process indicated that that the limitations discussed above did not materially impact the quality of the analytic recommendation derived from the data model constructed from the data available for analytic processing.

Recommendations

The use of analytic extensions addressed the different levels of information expression (measured, estimated, inferred, and forecasted) necessary for management analytic and decision-support problem-solving. This analytic extension of data models provided avenues to incorporate complex data elements of higher order logic into analytic processing and programming framework. The derived data was made available to the end user through the traditional analytic application user interfaces. The analytic extension of the data model led to an information representation scheme that aligned with the cognitive model of the domain of interest. It supported identification and classification of objects of interest within the domain. It also supported the abstraction of these information artifacts to the level needed for analytics and decision-support in management. The study identified three levels of data model transformation or analytic continuity concepts: semantic, symbolic, and dimensional extensions

The methods applied to the transformations at each level were also driven by specific theories. The theories of measurements (metrology) which advocated the formulation of measurement scales and instruments in all scientific disciplines drove semantic extensions. These theoretical formulations enabled scales developed for quantitative expression of non-physical quantities in nature, for example, key performance indicators (KPIs), balanced scorecard (BSC) metrics, customer lifetime value, customer churn, intelligent quotient, to name a few. The symbolic form addressed issues related to the optimal specification of the objects in a specific analytic context. At this level, accurate statistical and heuristic abstraction of data was necessary. The dimensional form addressed characteristics of subjects of interest mapped to abstract geometric forms.

The goal was to answer complex management decision questions directly from available data. For example, these transformations integrated the determinants and coefficients of expression within the domain of interest. Analytic extensions allowed reasoning about problems using the data model, rather than analytic algorithms. Therefore, in data analytics with big data where there would be many attributes, attribute levels, and analytic models, the use of data model extensions to persist these artifacts is recommended. The output of the analytic algorithm should also be captured in the data model and within the analytic application to enable real time comparison of analytic expectations to actual.

Implications

As mentioned in the previous section, as much as 50% of efforts to develop decision-support systems for management fail. The implication of such failures was the misalignment of fundamentals and market value of organizations. The situations of undervaluation or overvaluation had a long-term impact on organizations, as has been demonstrated by the technology industry burst of 2002 and financial industry failure and subsequent global market turmoil beginning in 2008.

Market economies depend on the accurate valuation of companies, which in turn depends on the predictability of the management activities of public and private companies. Since capitalism has become the dominant national economic philosophy in the world, the private sector plays an important role in national economic productivity, market efficiency, and overall societal prosperity. The productivity of organizations are important to the economic and social well-being of the society.

This study contains a schema-based approach to analytic problem-solving in management. That is, the solution to management analytics and decision-support problems lies in building a good data model to support analytic processing at all levels of the organization. In recent years, the focus on algorithms which are black-box solutions created the cognitive gap between empirical situation and analytic solutions. This schema-based solution provided a new layer of solutions that ensured proper application of analytical and numerical solution techniques.

About the contribution to social change, the company in the case illustration was a major distributor of medical equipment and a supplier to health care organizations. Discovering the causes of profitability issues, such as wasteful manufacturing processes, low value products, pricing discrepancies, product development partnership opportunities and so on brought the company closer to its customer base and the communities they served. The partnerships with Non-profit Community Hospitals and Health Centers in urban inner-city communities in Chicago, Detroit, Memphis, Atlanta, and many others opened up avenues for involvement in Community wellness and disease management programs. The company established incentives for the sale agents to participate in social programs within the communities they covered. The company established a foundation to support Health care facilities whose primary patients were Medicaid recipients to help cover losses from under-reimbursement for services from the U. S. government. The company also reached out to medical missions to South America, Africa, and Middle East to provide medical equipment and medical supply donations that these missions depended on for the free services they offered to very needy patients. The social change that could be realized through all these activities was improvement in health conditions of many communities, improvement in the daily activities of patients and residents of health care institutions served by the company, and support for non-profit organizations that were active in improving conditions of patients around the world.

Conclusions

The need for data-driven decision-making in organizations will only increase as data gathering, storage, and retrieval techniques improve. Significant work has been done in the construction of statistical databases for very large databases as well as in knowledge discovery from databases. While statistical databases lack the scalability of relational databases, relational databases based on classical data models were not able to provide the level of knowledge representation made possible by statistical databases. This study’s goal was to evolve data-modeling beyond an organizing framework for data. The goal was device tools and method to model data for higher levels of information representation necessary for business knowledge and intelligence discovery.

The discussion above showed that further extensions of the relational data model allowed a fundamental redefinition of the concept of the dimension from one popularized by OLAP community to one that was much more aligned to the mathematical interpretation (Hart, 2005). While the classical multidimensional model had attributes like dimensions, the dimensional extension approach provided a higher-order logic for constrained optimization and simulation problem-solving in management. The approach avoided the issue of the use of theoretical statistical distributions since comparison to chance was not valuable for management analytics and decision making.

In this study, I examined the use of analytic extensions to the data model to improve the discovery of management scenarios, insights, and complex business rules and constraints from big data. I established that the use of these analytic extensions was not just necessary but important to align available data to the intuitive concepts within the domain of interest. By using a combination of ontology learning, data engineering, and analytic formulation techniques in the derivation of these analytic extensions, the resulting data model was the concise and compact representation of the management scenarios, insights, and the rules that the management analyst and executive would optimize to achieve predictable outcome states. The use of analytic extensions closed the gap between analytic solutions and intuitive cognitive models of the business problems.

This study has the potential to increase (a) the acceptance of big data analytics outputs by business analysts and executives, (b) the return on investment for big data analytics projects, and (c) the overall efficiency of data-driven management analytics and decision-support. The social change implication is an increase in management engagement in social programs to sustain good corporate citizenship within stakeholder communities, including sponsorship of community events and social programs.

References

Adam, F. (2012). 20 years of decision making and decision support research published by the Journal of Decision Systems. *Journal of Decision Systems*, *21*(2), 93-99. doi:10.1080/12460125.2012.695890

Affenzeller, M., Winkler, S. M., Kronberger, G., Kommenda, M., Burlacu, B., & Wagner, S. (2014). *Genetic programming theory and practice XI*. New York, NY: Springer. doi:10.1007/978-1-4939-0375-7_10

Ahmad, M. F., Ariff, M. S. M., Zakuan, N., Takala, J., & Jusoh, A. (2013, April). *Business Engineering and Industrial Applications Colloquium (BEIAC), 2013 IEEE* (pp. 22-27). IEEE. doi:10.1109/BEIAC.2013.6560120

Ahmed, U., Tchounikine, A., & Miquel, M. (2014). Dynamic cubing for hierarchical multidimensional data space. *Journal of Decision Systems*, *23*(4), 415-436. doi:10.1080/12460125.2014.940241

Aiken, P. (2016). Experience: Succeeding at data management—BigCo attempts to leverage data. *Journal of Data and Information Quality*, *7*(1-2), 8. doi:10.1145/2900000/2893482

Al-Aqrabi, H., Liu, L., Hill, R., & Antonopoulos, N. (2015). Cloud BI: Future of business intelligence in the cloud. *Journal of Computer and System Sciences, 81*(1), 85-96. doi:10.1016/j.jcss.2014.06.013

Amblard, P. O., & Michel, O. (2014). Causal conditioning and instantaneous coupling in causality graphs. *Information Sciences*, *264*, 279-290.

doi:10.1016/j.ins.2013.12.037

Amelung, D., & Funke, J. (2013). Dealing with the uncertainties of climate engineering: Warnings from a psychological complex problem solving perspective. *Technology in Society*, *35*(1), 32-40. doi:10.1016/j.techsoc.2013.03.001

Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2015). *An introduction to management science: Quantitative approaches to decision making*. Boston, MA: Cengage Learning.

Argyres, N., & Zenger, T. (2012). Capabilities, transaction costs, and firm boundaries. *Organizational Science, 23*(6), 1643 – 1657. doi:10.1287/orsc.1110.0736

Arnott, D., & Pervan, G. (2014). A critical analysis of decision support systems research revisited: the rise of design science. *Journal of Information Technology, 29*, 269-293. doi:10.1057/jit.2014.16

Aruldoss, M., Travis, M. L., & Venkatesan, V. P. (2014). A survey on recent research in business intelligence. *Journal of Enterprise Information Management, 27*(6), 831-866. doi:10.1108/JEIM-06-2013-0029

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2014). Big data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing, 75*(13), 156-175. doi:10.1016/j.jpdc.2014.08.003

Badinger, H., Mühlböck, M., Nindl, E., & Reuter, W. H. (2013). Theoretical vs. empirical power indices: Do preferences matter? *European Journal of Political Economy, 36*, 158-176. doi:10.1016/j.ejpoleco.2014.07.009

Bai, X., White, D., & Sundaram, D. (2011, January). *System Sciences (HICSS), 2011 44th Hawaii International Conference* (pp. 1-10). Institute of Electrical and Electronic Engineers (IEEE) Computer Society. doi:10.1109/HICSS.2011.353

Bakker, A., & Van Eerde, D. (2012). An introduction to design-based research for master and Ph.D. students. *Doing (qualitative) research: Methodology and methods in mathematics education. ZDM research handbook series: Advances in mathematics education*. Berlin, Heidelberg: Springer.

Beroggi, G. (2010). *Designing and evaluating e-management decision tools: The integration of decision and negotiation models into internet-multimedia technologies*. New York, NY: Springer Science + Business Media.

Bendre, M. R., & Thool, V. R. (2016). Analytics, challenges and applications in big data environment: a survey. *Journal of Management Analytics, 3(3)*, 206-239. doi:10.1080/23270012.2016.1186578

Beynon, M. (2011). Shafer-Dempster Theory. In L. Moutinho & G. D. Hutcheson (Eds.). *The SAGE dictionary of quantitative management research* (pp. 80–83). Thousand Oaks, CA: Sage.

Bhattacharyya, K., Datta, P., & Maitra, A. (2013). Resource dynamics on service effectiveness: Evidence from the small business service industry in the United States. *Journal of Service Science Research*, *5*(1), 1-33. doi:10.1007/s12927-013-0001-1

Blessing, L., & Charkrabarti, A. (2009). *DRM: A design research methodology*. London, UK: Springer-Verlag. doi:10.1007/978-1-84882-587-1

Block, B. (2012). Analysis paralysis. *The Journal of Wildlife Management*, *76*(5), 875-876. doi:10.1002/jwmg.408

Bohland, J. W., Myers, E. M., & Kim, E. (2014). An informatics approach to integrating genetic and neurological data in speech and language neuroscience. *Neuroinformatics*, *12*(1), 39-62. doi:10.1007/s12021-013-9201-6

Boland, L. A. (2014). *The methodology of economic model building: Methodology after Samuelson*. New York, NY: Routledge.

Bond T. G., & Fox, C. M. (2013). *Apply the Rasch model: Fundamental measurements in the human sciences*. New York, NY: Routledge.

Boone, W., Staver, J., & Yale, M. (2014). *Rasch analysis in the human sciences*. Amsterdam, Netherlands: Springer Netherlands. doi:10.1007/978-94-007-6857-4_1

Bosch, O. J., Nguyen, N. C., & Buckle-Henning, P. (2014, April). Where is this so-called "Fifth Discipline" if project failures, blown-out budgets, decision disasters and poor investments continue to plague our society? In *Proceedings of the 57th Annual Meeting of the ISSS-2013 Haiphong, Vietnam, 1*(1), 1-6. Retrieved from http://journals.isss.org/index.php/proceedings57

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world.* MIT Press. doi:10.1080/01639374.2016.1234531

Boulil, K, Le Ber, F., Bimonte, S., Grac, C., & Cernesson, F. (2014). Multidimensional modeling and analysis of large and complex watercourse data: An OLAP-based solution. *Ecological Informatics*, *24*, 90-106. doi:10.1016/j.ecoinf.2014.07.001

Burns, N., & Jindra, J. (2013). Political spending and shareholder wealth: The effect of the US Supreme Court ruling in Citizens United. *American Politics Research*, November 7. doi:10.1177/1532673X13508976.

Caron, E. A. M. (2013). *Explanation of exceptional values in multi-dimensional business databases.* (Doctoral dissertation, Erasmus Research Institute of Management (ERIM), 2013, Retrieved from URL: https://www.erim.eur.nl)

Caron, E. A. M., & Daniels, H. A. M. (2008). Explanation of exceptional values in multi-dimensional business databases. *European Journal of Operational Research*, 188(3), 884–897. doi:10.1016/j.ejor.2007.04.039

Calude, C. S., & Longo, G. (2016). The deluge of spurious correlations in big data. *Foundations of Science*, 1-18. doi:10.1007/s10699-016-9489-4

Cao, S., Dehmer, M., & Kang, Z. (2017). Network entropies based on independent sets and matchings. *Applied Mathematics and Computation, 307,* 265-270. doi:10.1016/j.amc.2017.02.021

Castle, J. L., Doornik, J. A., Hendry, D. F., & Nymoen, R. (2014). Misspecification testing: Non-invariance of expectations models of inflation. *Econometric Reviews, 33*(5-6), 553-574. doi:10.1080/07474938.2013.825137

Carroll, R. J., Primo, D. M., & Richter, B. K. (2016). Using item response theory to improve measurement in strategic management research: An application to corporate social responsibility. *Strategic Management Journal, 37(1)*, 66-85. doi:10.1002/smj.2463

Ceci, M., Cuzzocrea, A., & Malerba, D. (2013). Effectively and efficiently supporting roll-up and drill-down OLAP operations over continuous dimensions via hierarchical clustering. *Journal of Intelligent Information Systems*, 1-25. doi:10.1007/s10844-013-0268-1

Cegielski C., Allison Jones-Farmer, L., Wu, Y., & Hazen, B. (2012). Adoption of cloud computing technologies in supply chains: An organizational information processing theory approach. *The International Journal of Logistics Management, 23*(2), 184-211. doi:10.1108/09574091211265350.

Certo, S. T., Busenbark, J. R., Woo, H. S., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*. *37*(13), 2639-2657. doi:10.1002/smj.2475

Chakrabarti, A., & Blessing, L. T. (2014). *An anthology of theories and models of design*. London, UK: Springer-Verlag. doi:10.1007/978-1-4471-6338-1_1

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188. Retrieved from http://www.misq.org/

Chen, K., Zhang, X., Petersen, A., & Müller, H. G. (2017). Quantifying infinite-dimensional data: Functional data analysis in action. *Statistics in Biosciences*, *9*(2), 582-604. doi:10.1007/s12561-015-9137-5

Chen, M., Feixas, M., Viola, I., Bardera, A., Shen, H. W., & Sbert, M. 2016). *Information theory tools for visualization*. Boca Raton, FL: CRC Press.

Colman, A. M. (2016). *Game theory and experimental games: The study of strategic interaction*. Elmsford, NY: Elsevier.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

Creswell, J. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. New York, NY: Sage.

Creswell, J. W. (2011). Controversies in mixed methods research. *The Sage Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.

Crapo, A. W., & Gustafson, S. (2016). Semantics: Revolutionary breakthrough or just another way of doing things? *Semantic Web*, pp. 85-118. doi:10.1007/978-3-319-16658-2_5

Cronholm, S., & Göbel, H. (2015). *New horizons in design science: Broadening the research agenda*. NY, New York: Springer International Publishing. doi:10.1007/978-3-319-18714-3_40

Curşeu, P. L., Jansen, R. J., & Chappin, M. M. (2013). Decision rules and group rationality: Cognitive gain or standstill? *PloS one*, *8*(2), e56454. doi:10.1371/journal.pone.0056454

Cury, A., Crémona, C., & Diday, E. (2010). Application of symbolic data analysis for structural modification assessment. *Engineering Structures*, *32*(3), 762-775.

doi:10.1016/j.engstruct.2009.12.004

Cushing, J. B. (2013). Beyond big data? *Computing in Science and Engineering*, *15*(5), 4-5. doi:10.1109/MCSE.2013.102

Cuzzocrea, A. (2011). *Scientific and statistical database management*. Berlin, Germany: Springer. doi:10.1007/978-3-642-22351-8_43

Cuzzocrea, A., Bellatreche, L., & Song, I. Y. (2013, October). *Proceedings of the sixteenth international workshop on Data warehousing and OLAP* (pp. 67-70). Association for Computing Machinery, New York, NY. doi:10.1145/2513190.2517828.

Cuzzolin, F. (2012). On the relative belief transform. *International Journal of Approximate Reasoning*, *53*(5), 786-804. doi:10.1016/j.ijar.2011.12.009

Dabhi, V. K., & Chaudhary, S. (2014). Empirical modeling using genetic programming: A survey of issues and approaches. *Natural Computing*, 1-28. doi:10.1007/s11047-014-9416-y

Daraio, C., Lenzerini, M., Leporelli, C., Naggar, P., Bonaccorsi, A., & Bartolucci, A. (2016). The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality. *Scientometrics*, 1-15. doi:10.1007/s11192-016-1913-6

Davis, D. F., Golicic, S. L., Boerstler, C. N., Choi, S., & Oh, H. (2013). Does marketing research suffer from methods myopia? *Journal of Business Research*, *66*(9), 1245-1250. doi:10.1016/j.jbusres.2012.02.020

De Smedt, P. (2013). Interactions between foresight and decision-making. *Participation and interaction in foresight: Dialogue, Dissemination and Visions, 17*. doi:10.4337/9781781956144.00008.

Deng, Y. (2017). Fuzzy analytical hierarchy process based on canonical representation on fuzzy numbers. *J Comput Anal Appl*, *22*(2), 201-228. Retrieved from http://www.eudoxuspress.com/

Deng, Y., Lu, Q., Chen, J., Chen, S., Wu, L., & Tang, L. (2014). Study on the extraction method of deformation influence factors of flexible material processing based on information entropy. *Advances in Mechanical Engineering*, *6*, 547-947. doi:10.1155/2014/547947

Denis, B., Ghrab, A., & Skhiri, S. (2013, October). A distributed approach for graph-oriented multidimensional analysis. In *Big Data, 2013 IEEE International Conference* (pp. 9-16). IEEE. doi:10.1109/BigData.2013.6691777

Dezert, J., & Tchamova, A. (2014). On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule. *International Journal of Intelligent Systems, 29*(3), 223-252. doi:10.1002/int.21638

Diamantini, C., Potena, D., & Storti, E. (2013). *Advanced information systems engineering workshops.* Berlin, Germany: Springer. doi:10.1007/978-3-642-38490-5_26

Diday, E. (2012). An introduction to symbolic data analysis and its application to the SODAS project. *Revista de Matemática: Teoría y Aplicaciones*, *7*(1-2), 1-22.

doi:10.15517/rmta.v7i1-2.177

Diday, E. (2016). Thinking by classes in data science: The symbolic data analysis paradigm. *Wiley Interdisciplinary Reviews: Computational Statistics, 8*(5), 172-205. doi:10.1002/wics.1384

Downarowicz, T., Travisany, D., Montecino, M., & Maass, A. (2014). Symbolic extensions applied to multiscale structure of genomes. *Acta biotheoretica*, *62*(2), 145-169. doi:10.1007/s10441-014-9215-y

Fasel, D. (2014). *Fuzzy data warehousing for performance measurement*. Marly, Switzerland: Springer International Publishing. doi:10.1007/978-3-04226-8_4

Federer, H. (2014). *Geometric measure theory*. Springer. Hamburg, Germany.

Feilmayr, C., & Wöß, W. (2016). An analysis of ontologies and their success factors for application to business. *Data & Knowledge Engineering, 101*, 1-23. doi:10.1016/j.datak.2015.11.003

Fish, A. N. (2012). *Knowledge automation: How to implement decision management in business processes*. Hoboken, NJ: John Wiley & Sons.

Foster, D., & Stine, R. (2013). *Statistics for business: Decision making and analysis*. Boston, MA: Addison Wesley.

Fried, D., Jansen, P., Hahn-Powell, G., Surdeanu, M., & Clark, P. (2015). Higher-order lexical semantic models for non-factoid answer reranking. *Transactions of the Association for Computational Linguistics, 3*, 197-210. Retrieved from https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/index

Gartner, Inc (2013). *Predicts 2014: Business intelligence and analytics will remain CIO's top technology priority*. Retrieved from http://www.gartner.com/

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. *Academy of Management Journal, 59*(6), 1880-1895. doi:10.5465/amj.2016.4007

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1493-1507. doi:10.5465/amj.2016.4005

Gijo, E. V., & Scaria, J. (2013). Application of statistical techniques for improving yield of a manufacturing process. *International Journal of Business Excellence*, *6*(3), 361-375. doi:10.1504/IJBEX.2013.053616

Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal, 37*(1), 167-176. doi:10.1002/smj.2459

Gomes L. (2014). Machine learning masetro Michael Jordan on the delusions of big data and other huge engineering efforts. IEEE Spectrum. Retrieved from http://spectrum.ieee.org.

Gosain, A., & Singh, J. (2015, January). Conceptual multidimensional modeling for data warehouses: A survey. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)* 2014 (pp. 305-316). Springer International Publishing. doi:10.1007/978-3-319-11933-5_33

Graf, T., Erskine, A., & Steele Jr, G. D. (2014). Leveraging data to systematically improve care: coronary artery disease management at Geisinger. *The Journal of Ambulatory Care Management*, *37*(3), 199-205. doi:10.1097/JAC.0000000000000038

Graves, N., Wloch, C., Wilson, J., Barnett, A., Sutton, A., Cooper, N., ... & Lamagni, T. (2016). A cost-effectiveness modeling study of strategies to reduce risk of infection following primary hip replacement based on a systematic review. *Health Technology Assessment*, *20*(54). doi:10.3310/hta20540

Gregory, R., Arvai, J., & Gerber, L. R. (2013). Structuring decisions for managing threatened and endangered species in a changing climate. *Conservation Biology*, *27*(6), 1212-1221. doi:10.1111/cobi.12165

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*. doi: 10.1016/j.jsis.2017.07.003

Hand, D. J. (2012). Empirical model building: Data, models, and reality, by James R. Thompson. *International Statistical Review*, *80*(1), 192-192. doi:10.1111/j.1751-5823.2012.00179_16.x

Hamarat, C., Kwakkel, J. H., & Pruyt, E. (2013). Adaptive robust design under deep uncertainty. *Technological Forecasting and Social Change*, *80*(3), 408-418. doi:10.1016/j.techfore.2012.10.004

Heeney, C. (2016). An ethical moment" in data sharing. *Science, Technology & Human*

Values, 1-26. doi:10.1177/0162243916648220.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems, 47*, 98-115. doi:10.1016/j.is.2014.07.006

Hendry, D. (2009). *Dynamic econometrics*. Oxford, Great Britain: Oxford University Press. doi:10.1093/0198283164.001.0001

Herrington, J. (2012, June). Design-based research: Implementation issues in emerging scholar research. *World Conference on Educational Multimedia, Hypermedia and Telecommunications* (Vol. 2012, No. 1, pp. 1-6). Retrieved from http://www.editlib.org/j/EDMEDIA/v/2012/n/1/

Hoang, H. H., Jung, J. J., & Tran, C. P. (2014). Ontology-based approaches for cross-enterprise collaboration: a literature review on semantic business process management. *Enterprise Information Systems, 8*(6), 648-664. doi:10.1080/17517575.2013.767382

Hoerl, R. W., Snee, R. D., & De Veaux, R. D. (2014). Applying statistical thinking to ‘Big Data’ problems. *Wiley Interdisciplinary Reviews: Computational Statistics*, *6*(4), 222-232. Retrieved from http://onlinelibrary.wiley.com

Hsu, D. F., Ito, T., Schweikert, C., Matsuda, T., & Shimojo, S. (2011). *Active media technology*. Berlin, Germany: Springer. doi:10.1007/978-3-642-23620-4_4

Hubbard, R., & Lindsay, R. M. (2013). From significant difference to significant sameness: Proposing a paradigm shift in business research. *Journal of Business*

Research, *66*(9), 1377-1388. doi:10.1016/j.jbusres.2012.05.002

Hussain, M. A., Elyas, T., & Nasseef, O. A. (2013). Research paradigms: A slippery slope for fresh researchers. *Life Science Journal*, *10*(4), 2374-2381. Retrieved from http://www.lifesciencesite.com

Iivari, J. (2014). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems 24*(1). doi:10.1057/ejis.2013.35

Jasper, J. D., Bhattacharya, C., Levin, I. P., Jones, L., & Bossard, E. (2013). Numeracy as a predictor of adaptive risky decision making. *Journal of Behavioral Decision Making*, *26*(2), 164-173. doi:10.1002/bdm.1748

Jareevongpiboon, W., & Janecek, P. (2013). Ontological approach to enhance results of business process mining and analysis. *Business Process Management Journal*, *19*(3), 459-476. doi:10.1108/14637151311319905

Jiang, X., Yuan, Y., Mahadevan, S., & Liu, X. (2013). An investigation of Bayesian inference approach to model validation with non-normal data. *Journal of Statistical Computation and Simulation*, 83(10), 1829-1851. doi:10.1080/00949655.2012.672572

Jiao, R. J., Zhou, F., & Chu, C. H. (2016). Decision theoretic modeling of affective and cognitive needs for product experience engineering: key issues and a conceptual framework. *Journal of Intelligent Manufacturing*, 1-13. doi:10.1007/s10845-016-1240-z

Johnson, B. (2014). Big data: Architecture and its enablement. In K. Marconi & H. Lehmann (Eds.), *Big data and health analytics* (pp. 156-175). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Kaisler, S. H., Espinosa, J. A., Armour, F., & Money, W. H. (2014). Advanced Analytics--Issues and Challenges in a Global Environment. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 729-738). IEEE. doi:10.1109/HICSS.2014.98.

Kalou, K., & Koutsomitropoulos, D. (2014). *Artificial intelligence applications and innovations*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-662-44722-2_34

Kaytoue, M., Kuznetsov, S. O., Napoli, A., & Polaillon, G. (2011). Symbolic data analysis and formal concept analysis. In *XVIIIeme rencontres de la Société Francophone de Classification-SFC 2011*, Orleans, France. Retrieved from https://hal.inria.fr/hal-00646457/document

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., ... & Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal, 2014*. doi: 10.1155/2014/712826

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85. doi:10.1145/2500873

Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: The complete guide to dimensional modeling*. New York, NY: John Wiley & Sons.

Kimpel J. (2013). Critical success factors for data warehousing: A classical answer to modern question. *Issues in Information Systems, 14*(1), 376-384. Retrieved from iacis.org/iis

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 2053951714528481. doi:10.1177/2053951714528481

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society, 3*(1), 1-10. doi:10.1177/2053951716631130

Koit, M., & Oim, H. (2016). An approach to the modeling of natural reasoning. In *Artificial intelligence IV: Methodology, systems and applications,* (pp. 93-102). New York, NY: Elsevier.

Klahr, D., & Kotovsky, K. (Eds.) (2013). *Complex information processing: The impact of Herbert A. Simon*. New York, NY: Psychology Press.

Krogstie, J. (2012). Modelling languages: Perspectives and abstraction mechanisms. In *Model-based development and evolution of information systems,* (pp. 89-204). London: Springer.

Kuechler, W., & Vaishnavi, V. (2012). A framework for theory development in design science research: Multiple perspectives. *Journal of the Association for Information Systems, 13*(6), 395-423. Retrieved from http://aisel.aisnet.org/jais/

Kuntz, K. M., Russell, L. B., Owens, D. K., Sanders, G. D., Trikalinos, T. A., & Salomon, J. A. (2016). Decision models in cost-effectiveness analysis. *Cost-effectiveness in health and medicine* (pp. 105). New York: NY, Oxford University

Press.

Kumari, A., & Singh, V. (2017). Challenges of modern query processing. *In Proceedings of the First International Conference on Computational Intelligence and Informatics* (pp. 423-432). Singapore: Springer.

Kuznetsov, S. D., & Kudryavtsev, Y. A. (2009). A mathematical model of the OLAP cubes. *Programming and Computer Software, 35*(5), 257-265. doi:10.1134/S0361768809050028

Kwakkel, J. H., & Pruyt, E. (2013). Using system dynamics for grand challenges: The ESDMA approach. *Systems Research and Behavioral Science, 32*(3), 358-375. doi:10.1002/sres.2225

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. W. J. (2010, April). From predictive modeling to exploratory modeling: how to use non-predictive models for decision making under deep uncertainty. In *Proceedings of the 25th Mini-EURO Conference on Uncertainty and Robustness in Planning and Decision Making (URPDM2010)*, 15-17 April. Retrieved from http://www.inescc.pt/urpdm2010/

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387-394. doi:10.1016/j.ijinfomgt.2014.02.002

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2013). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, Winter, 21. Retrieved from http://sloanreview.mit.edu/

Law, J., & Lien, M. (2012). Slippery: Field notes on empirical ontology. *Social Studies of Science, 43*(3) 363–378. doi:10.1177/0306312712456947

Leech, N. L., & Dellinger, A. (2012). *Validity: mixed methods. The encyclopedia of applied linguistics*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781405198431.wbeal1244

Leonard, A., Mitchell, T., Masson, M., Moss, J., & Ufford, M. (2014). *SQL Server 2012 integration services design patterns*. New York, NY: Apress. doi:10.1007/978-1-4842-0082-7_13

Levchuk, G., Ortiz, A., & Yan, X. (2014, May). Analysis of large-scale distributed knowledge sources via autonomous cooperative graph mining. *In SPIE sensing technology+ applications* (pp. 91190K-91190K). International Society for Optics and Photonics. doi:10.1117/12.2050836.

Liu, H., Guo, J., Yu, W., Zhu, L., Liu, Y., Xia, T., ... & Gardner, R. M. (2016). The design and implementation of the enterprise level data platform and big data driven applications and analytics. In *Transmission and Distribution Conference and Exposition (T&D), 2016 IEEE/PES* (pp. 1-5). IEEE. doi:10.1109/TDC.2016.7520032

Liu, G., Liu, Q., & Li, P. (2017). Blessing of dimensionality: Recovering mixture data via dictionary pursuit. *IEEE transactions on pattern analysis and machine intelligence, 39*(1), 47-60. doi: 10.1109/TPAMI.2016.2539946

Lu, J., Niu, L., & Zhang, G. (2013). A situation retrieval model for cognitive decision support in digital business ecosystems. *IEEE Transactions on Industrial Electronics*, *60*(3), 1059-1069. doi:10.1109/TIE.2012.2188253

Ma, X., Zheng, J. G., Goldstein, J. C., Zednik, S., Fu, L., Duggan, B., ... & Fox, P. (2014). Ontology engineering in provenance enablement for the National Climate Assessment. *Environmental Modelling & Software, 61*, 191-205. doi:10.1016/j.envsoft.2014.08.002

MacLeod, M., & Nersessian, N. J. (2018). Modeling complexity: cognitive constraints and computational model-building in integrative systems biology. *History and philosophy of the life sciences*, *40*(1), 17. doi: 10.1007/s40656-017-0183-9

MacDonald, C. M. (2013). Learning and teaching information architecture: The current state of IA education. *Bulletin of the American Society for Information Science and Technology*, *40*(1), 28-35. doi:10.1002/bult.2013.1720400109

Mahadevan, S. (2013, January). *Proceedings of the international symposium on engineering under uncertainty: Safety assessment and management (ISEUSAM-2012)* (pp. 97-117). New Delhi, India: Springer. doi:10.1007/978-81-322-0757-3_5

Mancas, C. (2016). On database relationships versus mathematical relations. *Global Journal of Computer Science and Technology*, *16*(1). doi:10.17406/gjcst

Mancas, C. (2016). Algorithms for database keys discovery assistance. In *International Conference on Business Informatics Research* (pp. 322-338). Springer

International Publishing. doi:10.1007/978-3-319-45321-7_23

Martinez-Cruz, C., Blanco, I. J., & Vila, M. A. (2012). Ontologies versus relational databases: Are they so different? A comparison. *Artificial Intelligence Review, 38*(4), 271-290. doi:10.1007/s10462-011-9251-9

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*, 60-66. Retrieved from http://hbr.org/

Metz, D. (2014). *The concept of a real-time enterprise in manufacturing*. Wiesbaden, Germany: Springer Fachmedian. doi:10.1007/978-3-658-03750-5_6.

Michalewicz, Z., Schmidt, M., Michalewicz, M., & Chiriac, C. (2011). *Adaptive business intelligence*. Berlin, Germany: Springer-Verlag. doi:10.4018/978-1-59904-849-9.ch003

Mintz, O., & Currim, I. S. (2013). What drives managerial use of marketing and financial metrics and does metric use affect performance of marketing-mix activities? *Journal of Marketing*, *77*(2), 17-40. doi:10.1509/jm.11.0463

Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, *88*, 300-305. doi:10.1016/j.procs.2016.07.439.

Montgomery, D. C. (2013). *Applied statistics and probability for engineers* (6th ed.). New York, NY: Wiley.

Moole, B. R. (2005). *A probabilistic multidimensional data model and its applications in business management.* (Doctoral dissertation). Retrieved from ProQuest

Dissertations and Theses. (UMI No. 3180107)

Morard, B., Stancu, A., & Jeannette, C. (2012). The relationship between structural equation modeling and balanced scorecard: Evidence from a Swiss non-profit organization. *Review of Business & Finance Studies*, *3*(2), 21-37. Retrieved from http://www.ssrn.com/en/

Mortenson, M. J., Doherty, N. F., & Robinson, S. (2014). Operational research from Taylorism to terabytes: A research agenda for the analytics Age. *European Journal of Operational Research*, *24*(3), 583-601. doi:10.1016/j.ejor.2014.08.029

Mousavi, S., & Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, *67(*8), 1671-1678. doi:10.1016/j.jbusres.2014.02.013

Nadin, M. (2018). Meaning in the Age of Big Data. In Empirical Research on Semiotics and Visual Rhetoric (pp. 86-127). Hersey, PA: IGI Global

Nalepa, G. J. (2017). *Modeling with Rules Using Semantic Knowledge Engineering*. Springer, Cham. doi: 10.1007/978-3-319-66655-6

Niemi, T., Niinimäki, M., Thanisch, P., & Nummenmaa, J. (2014). Detecting summarizability in OLAP. *Data & Knowledge Engineering*, *89*, 1-20. doi:10.1016/j.datak.2013.11.001

Noirhomme-Fraiture, M., & Brito, P. (2011). Far beyond the classical data models: Symbolic data analysis. *Statistical Analysis and Data Mining*, *4*(2), 157-170. doi:10.1002/sam.10112

Ortega, D. A., & Braun, P. A. (2011). *Artificial general intelligence*. Berlin, Germany:

Springer. doi:10.1007/978-3-642-22887-2_28

Osuszek, L., Stanek, S., & Twardowski, Z. (2016). Leverage big data analytics for dynamic informed decisions with advanced case management. *Journal of Decision Systems*, *25*(sup1), 436-449. doi:10.1080/12460125.2016.1187401

Paganoni, A. M., & Secchi, P. (2014). *Advances in complex data modeling and computational methods in statistics*. Berlin, Germany: Springer. doi:10.1007/978-3-319-11149-0

Pardillo, J., & Mazon, J. (2011). Using ontologies for the design for data warehouses. *International Journal of Database Management System*, *3*(2), 7387. doi:10.5121/ijdms.2011.3205

Pedersen, T. B. (2013). *Business intelligence.* Berlin, Germany: Springer. doi:10.1007/978-3-642-36318-4_1

Peters, E. (2012). Beyond comprehension the role of numeracy in judgments and decisions. *Current Directions in Psychological Science, 21*(1), 31-35. doi:10.1177/0963721411429960

Pfaff, M. S., Klein, G. L., Drury, J. L., Moon, S. P., Liu, Y., & Entezari, S. O. (2013). Supporting complex decision making through option awareness. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 155-178. doi:10.1177/1555343412455799

Pourshahid, A., Richards, G., & Amyot, D. (2011). *E-Technologies: Transformation in a connected world*. Berlin, Germany: Springer. doi:10.1007/978-3-642-20862-1_7

Powell, W. W., & DiMaggio, P. J. (Eds.). (2012). *The new institutionalism in organizational analysis*. Chicago, IL: University of Chicago Press.

Power, D. J., Burstein, F., & Sharda, R. (2011). *Decision support*. New York, NY: Springer. doi:10.1007/978-1-4419-6181-5_2

Puonti, M., Lehtonen, T., Luoto, A., Aaltonen, T., & Aho, T. (2016). Towards Agile Enterprise Data Warehousing. ICSEA 2016, 241. Retrieved from https://www.researchgate.net/profile/Luigi_Lavazza/publication/307576316_ICSEA_2016_The_Eleventh_International_Conference_on_Software_Engineering_Advances/links/57c9a36a08ae3ac722af8728/ICSEA-2016-The-Eleventh-International-Conference-on-Software-Engineering-Advances.pdf#page=242

Qin, H., Guan, H., & Wu, Y. J. (2013). Analysis of park-and-ride decision behavior based on Decision Field Theory. *Transportation research part F: Traffic Psychology and Behaviour, 18*, 199-212. doi:10.1016/j.trf.2013.02.001

Ransbotham, S., Kiron, D., & Prentice, P. (2016). Beyond the hype: the hard work behind analytics success. *MIT Sloan Management Review, 57*(3). Retrieved from: https://sloanreview.mit.edu/

Reimann, P. (2011). Design-based research. In L. Markauskaite, P. Freebody & J. Irwin, (Eds.). *Methodological choice and design* (pp. 37-50). Rotterdam, Netherlands: Springer.

Resmini, A. (2012). Information architecture in the age of complexity. *Bulletin of the American Society for Information Science and Technology, 39*(1), 9-13.

doi:10.1002/bult.2012.1720390104

Reyna, V. F., & Brust-Renck, P. G. (2014). A review of theories of numeracy: Psychological mechanisms and implications for medical decision making. In B. Anderson & J. Schulkin (Eds.). *Numerical reasoning in judgments and decision making about health* (pp. 215-251). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139644358.011

Righi, A. W., & Saurin, T. A. (2015). Complex socio-technical systems: Characterization and management guidelines. Applied ergonomics, 50, 19-30.

Rojas, L. A., & Macías, J. A. (2013). Bridging the gap between information architecture analysis and software engineering in interactive web application development. *Science of Computer Programming*, *78*(11), 2282-2291. doi:10.1016/j.scico.2012.07.020

Romero, O., & Abelló, A. (2011). *Scientific and statistical database management*. Berlin, Germany: Springer. doi:10.1007/978-3-642-22351-8_51

Romero, O., Marcel, P., Abelló, A., Peralta, V., & Bellatreche, L. (2011). *Data warehousing and knowledge discovery*. Berlin, Germany: Springer. doi:10.1007/978-3-642-23544-3_17

Saaty, T. L., & Peniwati, K. (2013). *Group decision making: Drawing out and reconciling differences*. New York, NY: RWS Publications.

Sankararaman, S., & Mahadevan, S. (2013). Bayesian methodology for diagnosis uncertainty quantification and health monitoring. *Structural Control and Health*

Monitoring, *20*(1), 88-106. doi:10.1002/stc.476

Savinov, A. A. (2012). *Data* (pp. 70-76). Retrieved from http://conceptoriented.org

Schuitema, G., Anable, J., Skippon, S., & Kinnear, N. (2013). The role of instrumental, hedonic and symbolic attributes in the intention to adopt electric vehicles. *Transportation Research Part A: Policy and Practice*, *48,* 39-49. doi:10.1016/j.tra.2012.10.004

Schutz, C., Neumayr, B., & Schrefl, M. (2013). *Advanced information systems engineering*. Berlin, Germany: Springer. doi:10.1007/978-3-642-38709-8_33

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., & McCulloch, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management, 11*(2), 78-88. doi: 10.1080/17509653.2016.1142191

Segura, L., & Sepulcre, J. M. (2015). A rational belief: The method of discovery in the complex variable. *Foundations of Science Online.* doi:10.1007/s10699-015-9412-4

Sharma, A., & Sood, M. (2013). Exploring model driven architecture approach to design star schema for a data warehouse. *Proc. of Advances in Engineering and Technology Series, 7*, 466-471. Retrieved from http://searchdl.org

Sharma, S., & Mittal, H. (2016). Data mining: Unblocking the intelligence in data. *Journal of Network Communications and Emerging Technologies*

(JNCET), 6(5). Retrieved from http://www.jncet.org

Shen, W., Davis, T., Lin, D. K., & Nachtsheim, C. J. (2013). Dimensional analysis and its applications in statistics. *Journal of Quality Technology*, *46*(3) 185-206. Retrieved from http://asq.org/pub/jqt/

Shi, H. X. (2014). *Entrepreneurship in family business.* New York, NY: Springer International Publishing. doi:10.1007/978-3-319-04304-3_3

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *Signal Processing Magazine, IEEE, 30*(3), 83-98. doi:10.1109/MSP.2012.2235192

Simiński, R. (2016). Multivariate approach to modularization of the rule knowledge bases. *Man–Machine Interactions 4* (pp. 473-483). Springer International Publishing.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70,* 263-286. doi:10.1016/j.jbusres.2016.08.001

Smirnov, P., & Kovalchuk, S. (2014). *Fuzzy Systems and Knowledge Discovery*. 2014 11th International Conference on (pp. 930-934). IEEE. doi:10.1109/FSKD.2014.6980964.

Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological* *review*, *120*(3), 589. doi:10.1037/a0033140.

Sofroniou, N. (2011). Rasch analysis. In L. Moutinho & G. D. Hutcheson (Eds.). *The SAGE dictionary of quantitative management research*. Thousand Oaks, CA: Sage.

Song, Q., Ni, J., & Wang, G. (2013). A fast clustering-based feature subset selection algorithm for high-dimensional data. *Knowledge and Data Engineering, IEEE Transactions on*, *25*(1), 1-14. doi:10.1109/TKDE.2011.181

Stoll, C. (2016). Data, Information and Knowledge in the Computer era. *Conceptual Data Modeling and Database Design: A Fully Algorithmic Approach, Volume 1: The Shortest Advisable Path*, 1. Boca Raton: FL. CRC Press.

Storey, V. C., & Song, I. Y. (2017). Big data technologies and Management: What conceptual modeling can do. *Data & Knowledge Engineering, 108,* 50-67. doi: 10.1016/j.datak.2017.01.001

Su, K. W., Yang, C. H., & Huang, P. H. (2016). The construction of an ontology-based knowledge management model for departure procedures. *Advances in Aerospace Science and Technology, 1*(01), 32. doi:10.4236/aast.2016.11004.

Sun, C. S., Cantor, S. B., Reece, G. P., Fingeret, M. C., Crosby, M. A., & Markey, M. K. (2014). Helping patients make choices about breast: A decision analysis approach. *Plastic and Reconstructive Surgery, 134*(4), 597-608. doi:10.1097/PRS.0000000000000514

Sun, Y., Lu, C., Bie, R., & Zhang, J. (2016). Semantic relation computing theory and its application. *Journal of Network and Computer Applications*, *59*, 219-229. doi:10.1016/j.jnca.2014.09.017

Syme, D., Granicz, A., & Cisternino, A. (2012). *Expert F# 3.0.* Apress. doi:10.1007/978-1-4302-4651-0_12

Thompson, J. R. (2011). *Empirical model building: Data, models, and reality*. New York, NY: John Wiley & Sons. doi:10.1002/9781118109656

Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering*, *22*(2), 127-151. doi:10.1007/s11518-013-5219-4

Tremblay, S., Gagnon, J. F., Lafond, D., Hodgetts, H. M., Doiron, M., & Jeuniaux, P. P. (2017). A cognitive prosthesis for complex decision-making. *Applied Ergonomics*, *58,* 349-360. doi:10.1016/j.apergo.2016.07.009

Truong, T. M., Amblard, F., Gaudou, B., Sibertin-Blanc, C., Truong, V. X., Drogoul, A., ... & Le, M. N. (2013, December). An implementation of framework of business intelligence for agent-based simulation. In *Proceedings of the Fourth Symposium on Information and Communication Technology* (pp. 35-44). ACM. doi:10.1145/2542050.2542069

Tsang, E. W. (2016). *Management research*. Thousand Oaks: CA. Sage.

Tufféry, S. (2011). *Data mining and statistics for decision making* (Vol 2). Hoboken: NJ: John Wiley & Sons.

Turner, K. L., & Makhija, M. V. (2012). The role of individuals in the information processing perspective. *Strategic Management Journal*, *33*(6), 661-680.

doi:10.1002/smj.1970.

Ullah, S., & Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology, 13*(1), 43. doi:10.1186/1471-2288-13-43.

Umanath, N., & Scamell, R. (2014). *Data modeling and database design*. Boston, MA: Cengage Learning.

Vallbé, J. J. (2015). *Frameworks for modeling cognition and decisions in institutional environments*. Amsterdam, Netherlands:Springer Netherlands. doi:10.1007/978-94-017-9427-5_3.

Verde, R., & Diday, E. (2014). Symbolic data analysis: A factorial approach based on fuzzy coded data. In J. Blasuis & M. Greenacre (Eds.), *Visualization and verbalization of data* (pp. 255-270). Boca Raton, FL: CRC Press.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77-84. doi:10.1111/jbl.12010

Welton, N. J., & Thom, H. H. (2015). Value of information: We’ve got speed, what more do we need? *Medical Decision Making*, *35*(5), 564-566. doi:10.1177/0272989X15579164

Weirich, P. (2017). Epistemic game theory and logic: Introduction. *Games 2017, 8*(2), 19. doi:10.3390/g8020019

Weinstein, M. C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C.,

& Luce, B. R. (2003). Principles of good practice for decision analytic modeling in health-care evaluation. Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health, 6*(1), 9-17. doi:10.1046/j.1524-4733.2003.00234.x

Weiner, M., Sherr, M., & Cohen, A. (2002). Metadata tables to enable dynamic data modeling and web interface design: the SEER example. *International Journal of Medical Informatics, 65*(1), 51-58. doi:10.1016/S1386-5056(02)00002-3

Werro, N. (2015). *Fuzzy classification of online customers.* Chan, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-15970-6_3

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020. doi:10.1037/xge0000014

Wickens, T. D. (2014). *The geometry of multivariate statistics*. New York, NY: Psychology Press.

Windschitl, M., Braaten, M., & Thomson, J. (2007). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, *92*(5), 941-967. doi:10.1002/sce.20259.

Wu, F., Priscilla, L., Gao, M., Caron, F., De Roover, W., & Vanthienen, J. (2012, January). *On the move to meaningful internet systems.* In *OTM 2012 Workshops* (pp. 525-533). Berlin, Germany: Springer. doi:10.1007/978-3-642-33618-8_69.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., . . . & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems, 14*(1), 1-37. doi:10.1007/s10115-007-0114-2

Xiong, L., & Liu, Y. (2016). Strategy representation and reasoning for incomplete information concurrent games in the situation calculus. *In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)* (pp. 1322-1329). Retrieved from http://www.ijcai.org

Xu, H., Luo, H., & He, J. (2013, December). What-if query processing policy for big data in OLAP System. *In Advanced Cloud and Big Data (CBD), 2013 International Conference* on (pp. 110-116). IEEE. doi:10.1109/CBD.2013.40

Yao, J., Bai, Z., & Zheng, S. (2015). *Large sample covariance matrices and high-dimensional data analysis*. Cambridge, UK: Cambridge University Press.

Yeoh, W., & Popovič, A. (2015). Extending the understanding of critical success factors for implementing business intelligence systems. *Journal of the Association for Information Science and Technology, 67*(1), 134-147. doi:10.1002/asi.23366

Zelinka, I., Davendra, D., Senkerik, R., Jasek, R., & Oplatkova, Z. (2011). Analytical programming: A novel approach for evolutionary synthesis of symbolic structures. In E. Kita (Ed.). *Evolutionary algorithms* (pp. 149-176). Rijeka: Croatia: InTech. doi:10.5772/16166.

Zicari R.V. et al. (2016) Setting Up a Big Data Project: Challenges, Opportunities, Technologies and Optimization. In: Emrouznejad A. (eds) Big Data Optimization:

Recent Developments and Challenges. Studies in Big Data, vol 18. Springer, Cham. doi:10.1007/978-3-319-30265-2_2.

Zojaji, Z., & Ebadzadeh, M. M. (2016). Semantic schema theory for genetic programming. *Applied Intelligence, 44*(1), 67-87. doi:10.1007/s10489-015-0696-4

Appendix A: Schedule A - Data Use Agreement

List of datasets and documents

S.No	Data file or document	Format	Data group	Type
1	DistFeedbackFlatFile_MedSurg_Hosp_PAS.mdb	mdb	Distribution Feedback	Data file
2	CHLOG_CUST.DAT	DAT	Customer Change	Data file
3	200901.DBF	DBF	Sales order	Data file
4	200902.DBF	DBF	Sales order	Data file
5	200903.DBF	DBF	Sales order	Data file
6	200904.DBF	DBF	Sales order	Data file
7	200905.DBF	DBF	Sales order	Data file
8	200906.DBF	DBF	Sales order	Data file
9	200907.DBF	DBF	Sales order	Data file
10	200908.DBF	DBF	Sales order	Data file
11	200909.DBF	DBF	Sales order	Data file
12	200910.DBF	DBF	Sales order	Data file
13	200911.DBF	DBF	Sales order	Data file
14	200912.DBF	DBF	Sales order	Data file
15	201001.DBF	DBF	Sales order	Data file
16	201002.DBF	DBF	Sales order	Data file
17	201003.DBF	DBF	Sales order	Data file
18	201004.DBF	DBF	Sales order	Data file
19	201005.DBF	DBF	Sales order	Data file
20	201006.DBF	DBF	Sales order	Data file
21	201007.DBF	DBF	Sales order	Data file
22	201008.DBF	DBF	Sales order	Data file
23	201009.DBF	DBF	Sales order	Data file
24	201010.DBF	DBF	Sales order	Data file
25	201011.DBF	DBF	Sales order	Data file
26	201012.DBF	DBF	Sales order	Data file
27	201101.DBF	DBF	Sales order	Data file
28	201102.DBF	DBF	Sales order	Data file
29	201103.DBF	DBF	Sales order	Data file
30	201104.DBF	DBF	Sales order	Data file
31	201105.DBF	DBF	Sales order	Data file
32	201106.DBF	DBF	Sales order	Data file
33	201107.DBF	DBF	Sales order	Data file
34	201108.DBF	DBF	Sales order	Data file

35	201109.DBF	DBF	Sales order	Data file
36	201110.DBF	DBF	Sales order	Data file
37	201111.DBF	DBF	Sales order	Data file
38	201112.DBF	DBF	Sales order	Data file
39	AUSP.DBF	DBF	Characteristic	Data file
40	AWCSKC.DBF	DBF	Advanced Wound	Data file
41	CHLOG_CU.DBF	DBF	Customer Change	Data file
42	KNA1.DBF	DBF	Customer Master	Data file
43	KNA1CREATEDATE.DBF	DBF	Customer Create	Data file
44	KNVP.DBF	DBF	Customer Partner	Data file
45	KNVV.DBF	DBF	Customer Master	Data file
46	PAQ1TOQ32009.DBF	DBF	Payment Advice	Data file
47	PA2010.DBF	DBF	Payment Advice	Data file
48	PA2011.DBF	DBF	Payment Advice	Data file
49	PASAMPLEQ42009.DBF	DBF	Payment Advice	Data file
50	SO419463945.DBF	DBF	Sales order	Data file
51	SOPARTNER.DBF	DBF	Sales order	Data file
52	SOPARTNERQ1TOQ32009.DBF	DBF	Sales order	Data file
53	SOPARTNERQ42009.DBF	DBF	Sales order	Data file
54	SOPARTNER2010.DBF	DBF	Sales order	Data file
55	SOPARTNER2011.DBF	DBF	Sales order	Data file
56	SOSAMPLE2WKDEC.DBF	DBF	Sales order	Data file
57	SOSAMPLEQ42009.DBF	DBF	Sales order	Data file
58	ZHST0809.DBF	DBF	Price History	Data file
59	ZHST.DBF	DBF	Price History	Data file
60	ZVCOM.DBF	DBF	Commission	Data file
61	2009 Sales by Div Item Category.xls	Xls	Sales by Division	Data file
62	2010 Sales by Div Item Category.xls	Xls	Sales by Division	Data file
63	2011 Sales by Div Item Category.xls	Xls	Sales by Division	Data file
64	2009 MED_SURG LIST.xls	Xls	Medical Surgical	Data file
65	2010 MED_SURG LIST.xls	Xls	Medical Surgical	Data file
66	2011 MED_SURG LIST.xls	Xls	Medical Surgical	Data file
67	Account Types.xls	Xls	Account types	Data file
68	active_acct.xls	Xls	Active accounts	Data file
69	AWC and Skin Care divided from det.xls	Xls	Advanced Wound	Data file
70	Commission percentages.xls	Xls	Commission	Data file
71	Credit-Analysis.xls	Xls	Credit analysis	Data file
72	ktokd.xls	Xls	Customer Group	Data file
73	ktokd-c.csv	Csv	Customer Group	Data file
74	ktokd-c.xls	Xls	Customer Group	Data file
76	Material Group.xls	Xls	Material Group	Data file

77	Material Master Extract_DOC.xls	Xls	Material Master	Data file
78	MPRSOut_3Q09.csv	csv	Master Production	Data file
79	Order Reason Codes.xls	xls	Order Reason	Data file
80	Partner Functions.xls	xls	Partner Functions	Data file
81	Product Division.xls	xls	Product Division	Data file
82	Sales Order Types.xls	xls	Sales Order Types	Data file
83	Total Cross List.xls	xls	Total Cross List	Data file
84	active_acct.pdf	pdf	Active account	Document
85	Credit Analysis by Reason Code Report -	pdf	Credit analysis by	Document
86	Credit Analysis by Reason Code Report -	pdf	Credit analysis by	Document
87	Credit Analysis by Reason Code Report -	pdf	Credit analysis by	Document
88	DistFeedbackReport_MedSurg_Hosp_20	pdf	Distribution	Document
89	DistFeedbackReport_MedSurg_Hosp_20	pdf	Distribution	Document
90	DistFeedbackReport_MedSurg_Hosp_20	pdf	Distribution	Document
91	DistFeedbackReport_MedSurg_Hosp_PAS_Total_2009.pdf	pdf	Distribution Feedback	Document
92	DistFeedbackReport_MedSurg_Hosp_PAS_Total_2010.pdf	pdf	Distribution Feedback	Document
93	DistFeedbackReport_MedSurg_Hosp_PAS_Total_2011.pdf	pdf	Distribution Feedback	Document
94	DistFeedbackReport_MedSurg_LongTer	pdf	Distribution	Document
95	DistFeedbackReport_MedSurg_LongTer	pdf	Distribution	Document
96	DistFeedbackReport_MedSurg_LongTer	pdf	Distribution	Document
97	DistFeedbackReport_MedSurg_PAS_Tot	pdf	Distribution	Document
98	DistFeedbackReport_MedSurg_PAS_Tot	pdf	Distribution	Document
99	DistFeedbackReport_MedSurg_PAS_Tot	pdf	Distribution	Document
100	200901.rar	rar	Sales order	Data file
101	200902.rar	rar	Sales order	Data file
102	200903.rar	rar	Sales order	Data file
103	200904.rar	rar	Sales order	Data file
104	200905.rar	rar	Sales order	Data file
105	200906.rar	rar	Sales order	Data file
106	200907.rar	rar	Sales order	Data file
107	200908.rar	rar	Sales order	Data file
108	200909.rar	rar	Sales order	Data file
109	PAQ1TOQ32009.rar	rar	Payment advise	Data file
110	PASAMPLEQ42009.rar	rar	Payment advise	Data file
111	SOPARTNERQ1TOQ3.rar	rar	Sales order	Data file
112	SOPARTNERQ1TOQ4.rar	rar	Sales order	Data file
113	SOSAMPLEQ42009.rar	rar	Sales order	Data file
114	KNVP.spool	spool	Cusromet partner	Data file

115	200901.txt	txt	Sales order	Data file
116	200902.txt	txt	Sales order	Data file
117	200903.txt	txt	Sales order	Data file
118	200904.txt	txt	Sales order	Data file
119	200905.txt	txt	Sales order	Data file
120	200906.txt	txt	Sales order	Data file
121	200907.txt	txt	Sales order	Data file
122	200908.txt	txt	Sales order	Data file
123	200909.txt	txt	Sales order	Data file
124	ausp.txt	txt	Characteristic	Data file
125	AUSP1.txt	Txt	Characteristic	Data file
126	AWCSKC1.txt	Txt	Advanced Wound	Data file
127	CHLOG_CU1.txt	Txt	Customer Change	Data file
128	KNA1_KNVV.TXT	TXT	Customer master	Data file
129	KNA11.txt	Txt	Customer Master	Data file
130	KNVP1.txt	Txt	Customer master	Data file
131	KNVV1.txt	Txt	Customer Master	Data file
132	makt.txt	Txt	Material	Data file
133	mara.txt	Txt	General Material	Data file
134	marm.txt	Txt	Measure of	Data file
135	mvke.txt	Txt	Material Sales	Data file
136	PAQ1TOQ32009.txt	Txt	Payment advise	Data file
137	PAQ1TOQ32009_1.txt	Txt	Payment advise	Data file
138	PASAMPLEQ42009.txt	Txt	Payment advise	Data file
139	SOPARTNERQ1TOQ3.txt	Txt	Sales order	Data file
140	SOPARTNERQ1TOQ4.txt	Txt	Sales order	Data file
141	SOSAMPLEQ420091.txt	Txt	Sales order	Data file
142	ZHST08091.txt	Txt	Price History	Data file
143	ZVCOM1.txt	Txt	Commissions	Data file
144	2Q09_MED SURG LIST.docm	Docm	Medical Surgical	Document
145	BSD_Material Master Extract.DOC	DOC	Material Master	Document
146	Commissions.doc	Doc	Commissions	Document

Appendix B: Study Data Dictionary

Relation name	Property Label	Property Full Name	Order	Length	Start	End	PrimaryKey Column
Material master	Material	Material	1	18	1	19	Material Number
Material master	Creation_date	Creation date	2	8	19	27	
Material master	Name_of_person_who_created_object	Name of person who created object	3	12	27	39	
Material master	Material_type	Material type	4	4	39	43	
Material master	Material_group	Material group	5	9	43	52	
Material master	Base_unit_of_measure	Base unit of measure	6	3	52	55	
Material master	Size/dimensions	Size/dimensions	7	32	55	87	
Material master	Purchasing_Value_key	Purchasing Value key	8	4	87	91	
Material master	Gross_weight	Gross weight	9	14	91	105	
Material master	Unit_of_weight	Unit of weight	10	3	105	108	
Material master	Volume	Volume	11	14	108	122	
Material master	Volume_unit	Volume unit	12	3	122	125	
Material master	Division	Division	13	2	125	127	
Material master	International_Article_Number/Universal_Product_Code	International Article Number/Universal Product Code	14	18	127	145	
Material master	Length	Length	15	14	145	159	
Material master	Width	Width	16	14	159	173	
Material master	Height	Height	17	14	173	187	
Material master	External_Material_Group	External Material Group	18	18	187	205	
Material master	Order_Unit_of_Measure	Order Unit of Measure	19	3	205	208	
Material master	Transportation_Group	Transportation Group	20	4	208	212	
Material master	Material_Group_–_Ship_Materials	Material Group – Ship Materials	21	4	212	216	
Material master	APO_Demand_Planner	APO Demand Planner	22	3	216	219	
Material master	Attribute_1	**Attribute_1**	23	5	219	224	
Material master	Attribute_2	**Attribute_2**	24	6	224	230	
Material master	Attribute_3	**Attribute_3**	25	1	230	231	
Material measure	Material	Material	1	18	1	19	Material Number

Material measure	Alternative_unit_of_measure	Alternative unit of measure	2	3	19	22	
Material measure	Numerator_for_conversion_to_base_UoM	Numerator for conversion to base UoM	3	5	22	27	
Material measure	Denominator_for_conversion_to_base_UoM	Denominator for conversion to base UoM	4	5	27	32	
Material measure	European_Article_Number_(EAN)_-_obsolete!!!!!	European Article Number (EAN) - obsolete!!!!!	5	13	32	45	
Material measure	International_Article_Number/Universal	International Article Number/Universal	6	18	45	63	
Material measure	Number_category_of_International_Article	Number category of International Article	7	2	63	65	
Material measure	Length	Length	8	14	65	79	
Material measure	Width	Width	9	14	79	93	
Material measure	Height	Height	10	14	93	107	
Material measure	Unit_of_dimension_for_length/width/height	Unit of dimension for length/width/height	11	3	107	110	
Material measure	Volume	Volume	12	14	110	124	
Material measure	Volume_unit	Volume unit	13	3	124	127	
Material measure	Gross_weight	Gross weight	14	14	127	141	
Material measure	Unit_weight	Unit weight	15	3	141	144	
Material measure	Unit_of_measure_contained_in_a_unit_of_measure	Unit of measure contained in a unit of measure	16	3	144	147	
Material measure	Internal_characteristic	Internal characteristic	17	10	147	157	
Material measure	Unit_of_measure_sort_number	Unit of measure sort number	18	2	157	159	
Material measure	Leading_proportion	Leading proportion	19	1	159	160	
Material measure	Valuation_based_on_the_proportion_quantity	Valuation based on the proportion quantity	20	1	160	161	
Material measure	Units_of_measurement_usage	Units of measurement usage	21	1	161	162	
Material measure	Unit_of_measurement_of_characteristic	Unit of measurement of characteristic	22	3	162	165	
Material sales	Material	Material	1	18	1	19	Material Number
Material sales	Sales_organization	Sales organization	2	4	19	23	
Material sales	Distribution_channel	Distribution channel	3	2	23	25	
Material sales	Material_Statistics_group	Material Statistics group	4	1	25	26	
Material sales	Volume_Rebate_Group	Volume Rebate Group	5	2	26	28	
Material sales	Commission_Group	Commission Group	6	2	28	30	

Material sales	Distribution-chain-specific_material_status	Distribution-chain-specific material status	7	2	30	32
Material sales	Date_from_which_distr.-chain-spec._material_status_is_valid	Date from which distr.-chain-spec. material status is valid	8	8	32	40
Material sales	Minimum_Order_quantity_in_base_UOM	Minimum Order quantity in base UOM	9	14	40	54
Material sales	Minimum_Delivery_quantity_in_delivery_no	Minimum Delivery quantity in delivery no	10	14	54	68
Material sales	Minimum_make-to-order_quantity	Minimum make-to-order quantity	11	14	68	82
Material sales	Delivery_unit	Delivery unit	12	14	82	96
Material sales	Unit_of_measure_of_delivery_unit	Unit of measure of delivery unit	13	3	96	99
Material sales	Sales_Unit	Sales Unit	14	3	99	102
Material sales	Item_category_group_from_material_master	Item category group from material master	15	4	102	106
Material sales	Delivery_plant	Delivery plant	16	4	106	110
Material sales	Material_Pricing_group	Material Pricing group	17	2	110	112
Material sales	Product_Division	Product Division	18	3	112	115
Material sales	Top_1001	Top 1001	19	3	115	118
Material sales	Product_Rep_type	Product Rep type	20	3	118	121
Material sales	Freight_Override	Freight Override	21	3	121	124
Material sales	Vendor_Code	Vendor Code	22	3	124	127
Material sales	Latex_Free	Latex Free	23	1	127	128
Material sales	Color_Required	Color Required	24	1	128	129
Material sales	Formulary_item_for_Home_Health_Orders	Formulary item for Home Health Orders	25	1	129	130
Material sales	Catalog_Database_4_Internet	Catalog Database 4 Internet	26	1	130	131
Material sales	ID_for_product_attribute_5	ID for product attribute 5	27	1	131	132
Material sales	ID_for_product_attribute_6	ID for product attribute 6	28	1	132	133
Material sales	ID_for_product_attribute_7	ID for product attribute 7	29	1	133	134
Material sales	ID_for_product_attribute_8	ID for product attribute 8	30	1	134	135
Material sales	ID_for_product_attribute_9	ID for product attribute 9	31	1	135	136
Material sales	ID_for_product_attribute_10	ID for product attribute 10	32	1	136	137
Material sales	Custom_item_category	Custom item category	33	4	137	141
Material sales	HCPCS_Code	HCPCS Code	34	30	141	171

Material sales	Material_Block_Group_1	Material Block Group 1	35	10	171	181	
Material sales	Material_Block_Group_2	Material Block Group 2	36	10	181	191	
Material sales	Material_Block_Group_3	Material Block Group 3	37	10	191	201	
Material sales	Material_Block_Group_4	Material Block Group 4	38	10	201	211	
Material sales	Material_Block_Group_5	Material Block Group 5	39	10	211	221	
Material sales	Canada_Maple_Leaf	Canada Maple Leaf	40	1	221	222	
Material sales	Do_Not_Reactivate	Do Not Reactivate	41	1	222	223	
Material sales	Direct_Only	Direct Only	42	1	223	224	
Material sales	To_Be_Discontinued	To Be Discontinued	43	1	224	225	
Material sales	Surplus_Flag	Surplus Flag	44	1	225	226	
Material sales	No_Re-route_Flag	No Re-route Flag	45	1	226	227	
Material sales	Preferred_Components	Preferred Components	46	1	227	228	
Material sales	Ship_300_Exclude	Ship 300 Exclude	47	1	228	229	
Material sales	Corporate_Controlled_Pallet_(CCP)	Corporate Controlled Pallet (CCP)	48	1	229	230	
Material sales	Custom_Product_Attribute_P	Custom Product Attribute P	49	1	230	231	
Material sales	Custom_Product_Attribute_Q	Custom Product Attribute Q	50	1	231	232	
Material sales	Custom_Product_Attribute_R	Custom Product Attribute R	51	1	232	233	
Material sales	Custom_Product_Attribute_S	Custom Product Attribute S	52	1	233	234	
Material sales	Custom_Product_Attribute_T	Custom Product Attribute T	53	1	234	235	
Material sales	Custom_Product_Attribute_U	Custom Product Attribute U	54	1	235	236	
Material sales	Custom_Product_Attribute_V	Custom Product Attribute V	55	1	236	237	
Material sales	Custom_Product_Attribute_W	Custom Product Attribute W	56	1	237	238	
Material sales	Custom_Product_Attribute_X	Custom Product Attribute X	57	1	238	239	
Material sales	Custom_Product_Attribute_Y	Custom Product Attribute Y	58	1	239	240	
Material sales	Custom_Product_Attribute_Z	Custom Product Attribute Z	59	1	240	241	
Material sales	Manufacturer_Code	Manufacturer Code	60	10	241	251	
Material sales	Manufacturer_Name_(from_table_ZMFR)	Manufacturer Name (from table ZMFR)	61	35	251	286	
Material sales	Manufacturer_Item_Number	Manufacturer Item Number	62	35	286	321	
Material description	Material	Material	1	18	1	19	Material Number
Material description	Language	Language	2	1	19	20	

Material description	Material_description	Material description	3	40	20	60	
Material description	Material_description_in_upper	Material description in upper	4	40	60	100	
Material plant	Material	Material	1	18	1	19	Material Number
Material plant	Plant	Plant	2	4	19	23	
Material plant	Plant_specific_material_status_from_MM	Plant specific material status from MM	3	2	23	25	
Material plant	ABC_indicator	**ABC indicator**	4	1	25	26	
Material plant	Purchasing_group	Purchasing group	5	3	26	29	
Material plant	Unit_of_Issue	**Unit of Issue**	6	**3**	29	32	
Material plant	Material_-_MRP_profile	Material - MRP profile	7	4	32	36	
Material plant	MRP_type	MRP type	8	2	36	38	
Material plant	MRP_controller	MRP controller	9	3	38	41	
Material plant	Planned_delivery_time_in_days	Planned delivery time in days	10	3	41	44	
Material plant	Good_Receipt_Processing_Days	Good Receipt Processing Days	11	3	44	47	
Material plant	Period_Indicator	Period Indicator	12	1	47	48	
Material plant	Lot_size_(materials_planning)	Lot size (materials planning)	13	2	48	50	
Material plant	Procurement_type	Procurement type	14	1	50	51	
Material plant	Special_procurement_type	Special procurement type	15	2	51	53	
Material plant	Reorder_Point	Reorder Point	16	14	53	67	
Material plant	Safety_stock	Safety stock	17	14	67	81	
Material plant	Minimum_lot_size	Minimum lot size	18	14	81	95	
Material plant	Maximum_lot_size	Maximum lot size	19	14	95	109	
Material plant	Fixed_lot_size	Fixed lot size	20	14	109	123	
Material plant	Rounding_value_for_purchase_order_qty	Rounding value for purchase order qty	21	14	123	137	
Material plant	Maximum_stock_level	Maximum stock level	22	14	137	151	
Material plant	Ordering_Costs	Ordering Costs	23	12	151	163	
Material plant	Dep._Requirement__Ind._For_Individual	Dep. Requirement Ind. For Individual	24	1	163	164	
Material plant	Schedule_Margin_Key	Schedule Margin Key	25	3	164	167	
Material plant	Production_Scheduler	Production Scheduler	26	3	167	170	
Material plant	In-house_production_type	In-house production type	27	3	170	173	
Material plant	Over_delivery_Tolerance_Limit	Over delivery Tolerance Limit	28	4	173	177	

Material plant	Under_Delivery_Tolerance_Limit	Under Delivery Tolerance Limit	29	4	177	181
Material plant	Loading_group	Loading group	30	4	181	185
Material plant	Service_level	Service level	31	4	185	189
Material plant	Splitting_Indicator	Splitting Indicator	32	1	189	190
Material plant	Checking_group_for_availability_check	Checking group for availability check	33	2	190	192
Material plant	Fiscal_Year_Variant	Fiscal Year Variant	34	2	192	194
Material plant	Indicator:_Take_Correction_Factor_	Indicator: Take Correction Factor	35	1	194	195
Material plant	Base_quantity_for_capacity_planning	Base quantity for capacity planning	36	14	195	209
Material plant	Indicator:_Automatic_Purchasing_Order_	Indicator: Automatic Purchasing Order	37	1	209	210
Material plant	Indicator:_source_list_requirement	Indicator: source list requirement	38	1	210	211
Material plant	Commodity_Code/Import_......	Commodity Code/Import	39	17	211	228
Material plant	Material_Country_of_Origin	Material Country of Origin	40	3	228	231
Material plant	Region_of_Origin	Region of Origin	41	3	231	234
Material plant	Profit_Center	Profit Center	42	10	234	244
Material plant	Stock_in_transit	Stock in transit	43	15	244	259
Material plant	Planning_Time_Fence	Planning Time Fence	44	3	259	262
Material plant	Costing_Lot_Size	Costing Lot Size	45	14	262	276
Material plant	Special_Procurement_Type_of_Costing	Special Procurement Type of Costing	46	2	276	278
Material plant	Production_Unit	Production Unit	47	3	278	281
Material plant	Issue_Storage_Location	Issue Storage Location	48	4	281	285
Material plant	MRP_Group	MRP Group	49	4	285	289
Material plant	Takt_Time	Takt Time	50	3	289	292
Material plant	Storage_costs_indicator	Storage costs indicator	51	1	292	293
Material plant	Maintenance_Status_(Views_Created)	Maintenance Status (Views Created)	52	15	293	308
Material plant	Storage_Location_for_EP	Storage Location for EP	53	4	308	312
Material plant	Quota_Arrangement_Usage	Quota Arrangement Usage	54	1	312	313
Material plant	ABC_Indicator	ABC Indicator	55	1	313	314
Material plant	Pallet_Quantity	Pallet Quantity	56	4	314	318
Material plant	Deployment_Center	**Deployment Center**	57	**4**	318	322
Material plant	Pallet_Quantity	**Pallet Quantity**	58	**4**	322	326

Material plant	Deployment_Center	**Deployment Center**	59	**4**	326	330	
Material plant	Rounding_value_release_strategy	**Rounding value release strategy**	60	**13**	330	343	
Material plant	Safety_Time_Indicator	**Safety Time Indicator**	61	**1**	343	344	
Material plant	Safety_Time_Days	**Safety Time Days**	62	**2**	344	346	
Material characteristics value	Material	Material	1	18	1	19	Material Number
Material characteristics value	Class__(Class_)	Class (Class)	2	18	19	37	
Material characteristics value	Class_Type__(Klart)	Class Type (Klart)	3	3	37	40	
Material characteristics value	Item_Class	Item Class	4	10	40	50	
Material characteristics value	Production_Group	Production Group	5	6	50	56	
Material characteristics value	PATTERN_ID	PATTERN ID	6	10	56	66	
Material characteristics value	Fabric_Type	Fabric Type	7	20	66	86	
Material characteristics value	Spread_Type	Spread Type	8	20	86	106	
Material characteristics value	Style_of_Garment	Style of Garment	9	20	106	126	
Material characteristics value	Color_of_Garment	Color of Garment	10	20	126	146	
Material characteristics value	Fabric	Fabric	11	20	146	166	
Material characteristics value	Size_of_Garment	Size of Garment	12	20	166	186	
Material characteristics value	Dimension_1_–_Length	Dimension 1 – Length	13	20	186	206	
Material characteristics value	Dimension_2_–_Width	Dimension 2 – Width	14	20	206	226	
Material valuation	Material	Material	1	18	1	19	Material Number
Material valuation	Valuation_area	Valuation area	2	4	19	23	
Material valuation	Valuation_type	Valuation type	3	10	23	33	
Material valuation	Deletion_flag_for_all_material_data_of_a_valuation_type	Deletion flag for all material data of a valuation type	4	1	33	34	
Material valuation	Total_valuated_stock	Total valuated stock	5	15	34	49	
Material valuation	Value_of_total_valuated_stock	Value of total valuated stock	6	15	49	64	

Material valuation	Price_control_indicator	Price control indicator	7	1	64	65	
Material valuation	Moving_average_price/periodic_unit_price	Moving average price/periodic unit price	8	13	65	78	
Material valuation	Standard_price	Standard price	9	13	78	91	
Material valuation	Price_unit	Price unit	10	5	91	96	
Material valuation	Valuation_class	Valuation class	11	4	96	100	
Material valuation	Value_based_on_moving_average_price_(only_with_price_ctrl_S)	Value based on moving average price (only with price ctrl S)	12	15	100	115	
Material valuation	Total_valuated_stock_in_previous_period	Total valuated stock in previous period	13	15	115	130	
Material valuation	Value_of_total_valuated_stock_in_previous_period	Value of total valuated stock in previous period	14	15	130	145	
Material valuation	Price_control_indicator_for_previous_period	Price control indicator for previous period	15	1	145	146	
Material valuation	Moving_average_price/periodic_unit_price_in_previous_period	Moving average price/periodic unit price in previous period	16	13	146	159	
Material valuation	Standard_price_in_the_previous_period	Standard price in the previous period	17	13	159	172	
Material valuation	Price_unit_of_previous_period	Price unit of previous period	18	5	172	177	
Material valuation	Origin_as_subdivision_of_cost	Origin as subdivision of cost	19	4	177	181	
Material valuation	Costing_overhead_group	Costing overhead group	20	10	181	191	
Material valuation	Costing_W/_Quantity_Structure	Costing W/ Quantity Structure	21	1	191	192	
Customer master sale	Client	Client					
Customer master sale	Customer_number_	Customer number	3	26			Customer Number
Customer master sale	Pricing_procedure_assigned_to_this_customer__	Pricing procedure assigned to this customer	1	9			
Customer master sale	Customer_group_	Customer group	2	2			Customer Group
Customer master sale	Freight_Default_	Freight Default	4	12			
Customer master sale	Access_Program_	Access Program	5	2			
Customer master sale	Confirmation_Preference_	Confirmation Preference	6	3			
Customer master sale	Deletion_indicator_for_customer_(at_sales_level)_	Deletion indicator for customer (at sales level)	7	8			
Customer master sale	Division_	Division	8	4			
Customer master sale	Customer_statistics_group_	Customer statistics group	9	1			

Customer master sale	Sales_organization_	Sales organization	10	10	
Customer master sale	Distribution_channel_	Distribution channel	11	2	
Customer master sale	Delivering_plant_	Delivering plant	12	14	
Customer master sale	Invoice_Preference_	Invoice Preference	13	12	
Customer master sale	Invoice_list_schedule_(calendar_identification)_	Invoice list schedule (calendar identification)	14	2	
Customer master sale	Central_order_block_for_customer_	Central order block for customer	16	8	
Customer master sale	Customer_account_group_	Customer account group	18	4	
Customer master sale	Bed_Count_	Bed Count	32	8	
Customer master sale	Sales_Office_	Sales Office	33	4	Sales Office
Customer master sale	Price_group_(customer)_	Price group (customer)	34	16	
Customer master sale	Terms_of_payment_key_	Terms of payment key	35	4	
Customer master general data	Client	Client			
Customer master general data	Customer_number_	Customer number	19	10	Customer Number
Customer master general data	Central_deletion_flag_for_master_record_	Central deletion flag for master record	20	1	
Customer master general data	Name_1_	Name 1	21	35	
Customer master general data	Name_2_	Name 2	22	35	
Customer master general data	Name_3_	Name 3	23	35	
Customer master general data	Name_4_	Name 4	24	35	
Customer master general data	City_	City	25	35	
Customer master general data	Post_office_box_	Post office box	26	10	
Customer master general data	P.O._Box_postal_code_	P.O. Box postal code	27	10	
Customer master general data	Postal_code_	Postal code	28	17	
Customer master general data	Region_(State,_Province,_County)_	Region (State, Province, County)	29	12	
Customer master general data	Street_and_house_number_	Street and house number	30	35	

Customer master general data	First_telephone_number_	First telephone number	31	25
Customer master general data	Account_Group	Account Group		
Competitive Item mapping	Competitive_Item	Competitive_Item	1	11
Competitive Item mapping	Competitive_Desc	Competitive_Desc	2	44
Competitive Item mapping	Medline_Item	Medline_Item	3	10
Competitive Item mapping	Medline_dec	Medline_dec	4	46
Competitive Item mapping	Var5		5	1
Material production record	Dist_I_Num	Dist_I_Num	1	13
Material production record	d_mfg_prod	d_mfg_prod	2	13
Material production record	D_MFG_ID	D_MFG_ID	3	12
Material production record	UM	UM	4	4
Material production record	Dist_num	Dist_num	5	8
Material production record	MFG_ID	MFG_ID	6	8
Material production record	HPIS_Cat	HPIS_Cat	7	9
Material production record	Brand	Brand	8	22
Material production record	Cat_Desc	Cat_Desc	9	32
Material production record	UM_CONV	UM_CONV	10	8
Material production record	Mfg_name	Mfg_name	11	22
Material production record	Class	Class	12	8
Material production record	class_desc	class_desc	13	61
Material production record	Major	Major	14	8

Material production record	Maj_Desc	Maj_Desc	1 5	31	
Material production record	Interim	Interim	1 6	8	
Material production record	Int_desc	Int_desc	1 7	51	
Material production record	Sub	Sub	1 8	8	
Material production record	Sub_Desc	Sub_Desc	1 9	40	
Material production record	Minor	Minor	2 0	8	
Material production record	Min_desc	Min_desc	2 1	61	
Advanced wound skin care product	Title	Title	1	12	
Advanced wound skin care product	Item	Item	2	9	
Advanced wound skin care product	Category	Category	3	3	
Customer Group	Client	Client	1	8	Client
Customer Group	Customer_group	Customer group	2	3	Customer Group
Customer Group	Name	Name	3	24	
Commission rate	Client	Client	1	8	client
Commission rate	Valid_From	Valid_From	2	8	
Commission rate	Valid_To	Valid_To	3	8	
Commission rate	BP__Start	BP__Start	4	8	
Commission rate	BP__End	BP__End	5	8	
Commission rate	Commission_Rate	Commission_Rate	6	8	
Commission rate	User_Name	User_Name	7	8	
Commission rate	Date	Date	8	8	
Material group	Client	Client	1	8	client
Material group	Material Group	Material Group	2	8	
Material group	Matl_grp_descr_	Matl_grp_descr_	3	26	
Order reason	Client	Client	1	8	client
Order reason	Language	Language	2	1	
Order reason	Order Reason	Order Reason	3	12	

Order reason	Description	Description	4	52	
Partner function	Client	Client	1	8	client
Partner function	Language	Language	2	1	
Partner function	Part_Funct_	Partner Function	3	2	Partner Function
Partner function	Name	Name	4	25	
Product division	Client	Client	1	8	client
Product division	Language	Language	2	1	
Product division	Product_Division	Product_Division	3	8	
Product division	Description	Description	4	37	
Sales document	Client	Client	1	8	client
Sales document	Language	Language	2	1	
Sales document	Sales_Doc__T	Sales_Doc__T	3	5	
Sales document	Description	Description	4	20	
Sales order	Client	Client	1	3	Client
Sales order	Customer number	Customer number	2	10	Customer Number
Sales order	Sales Office	Sales Office	3	4	Sales Office
Sales order	PO Type (Order Method)	PO Type (Order Method)	4	4	
Sales order	Order Reason Code	Order Reason Code	5	3	
Sales order	Pricing Date	Pricing Date	6	8	
Sales order	Sales Order Type	Sales Order Type	7	4	
Sales order	Sales order Number	Sales order Number	8	10	
Sales order	Sales Order Line	Sales Order Line	9	6	
Sales order	Material Number	Material Number	10	18	Material Number
Sales order	Material Description	Material Description	11	40	
Sales order	Material Group	Material Group	12	9	
Sales order	Net Value	Net Value	13	8	
Sales order	Plant	Plant	14	4	
Sales order	QTY	QTY	15	8	
Sales order	Sales UOM	Sales UOM	16	3	
Sales order	Condition Record	Condition Record	17	4	
Sales order	Condition Value	Condition Value	18	8	
Sales order	Extended Condition Value	Extended Condition Value	19	8	

Sales order	Active Pricing Condition - will need a formula here	Active Pricing Condition - will need a formula here	20	1			
Sales order partner	Client	Client	1	3			Client
Sales order partner	Sales Order Number	Sales Order Number	2	10			
Sales order partner	Partner Function	Partner Function	3	2			
Sales order partner	Customer number	Customer number	4	10			Customer Number
Customer history	CLient	CLient	1	3			Client
Customer history	Application	Application	2	2			
Customer history	Condition Record	Condition Record	3	4			
Customer history	Customer number	Customer number	4	10			Customer Number
Customer history	Material	Material	5	18			Material
Customer history	Valid From	Valid From	6	8			
Customer history	Valid To	Valid To	7	8			
Customer history	Condition Value	Condition Value	8	8			
Customer change log	cdhdr-objectclas	Object Class	1	15	1	16	
Customer change log	cdhdr-objectid	Object value	2	90	16	106	
Customer change log	cdhdr-changenr	Document change number	3	10	106	116	
Customer change log	cdhdr-username	User name of the person responsible in change document	4	12	116	128	
Customer change log	cdhdr-udate	Creation date of the change document	5	8	128	136	
Customer change log	cdhdr-utime	Time changed	6	6	136	142	
Customer change log	cdhdr-tcode	Transaction in which a change was made	7	20	142	162	
Customer change log	cdhdr-planchngnr	Planned change number	8	12	162	174	
Customer change log	cdhdr-act_chngno	Change number of the document created by this change	9	10	174	184	
Customer change log	cdhdr-was_plannd	Flag that changes were generated from planned changes	10	1	184	185	
Customer change log	cdhdr-change_ind	Application object change type (U, I, E, D)	11	1	185	186	
Customer change log	cdhdr-langu	Language Key	12	1	186	187	
Customer change log	cdhdr-version	3-Byte field	13	3	187	190	
Customer change log	cdpos-tabname	Table Name	14	30	190	220	
Customer change log	cdpos-tabkey	Changed table record key	15	70	220	290	

Customer change log	cdpos-fname	Field Name	16	30	290	320
Customer change log	cdpos-chngind	Change type (U, I, E, D)	17	1	320	321
Customer change log	cdpos-text_case	Flag: X=Text change	18	1	321	322
Customer change log	cdpos-unit_old	Change documents, unit referenced	19	3	322	325
Customer change log	cdpos-unit_new	Change documents, unit referenced	20	3	325	328
Customer change log	cdpos-cuky_old	Change documents, referenced currency	21	5	328	333
Customer change log	cdpos-cuky_new	Change documents, referenced currency	22	5	333	338
Customer change log	cdpos-value_new	New contents of changed field	23	254	338	592
Customer change log	cdpos-value_old	Old contents of changed field	24	254	592	846
Customer partner	Client	Client				
Customer partner	Customer	Customer (typically sold to)				
Customer partner	Partner Function	Partner Function				
Customer partner	Customer	Customer (the byproduct of the sold to)				
Payment advice	Sales Order	Sales Order				
Payment advice	Sales Order Line	Sales Order Line				
Payment advice	Document Date	Document Date				
Payment advice	Invoice Number	Invoice Number				
Payment advice	Invoice Line	Invoice Line				
Payment advice	Material	Material				
Payment advice	Billing Type	Billing Type				
Payment advice	Revenue	Revenue				
Payment advice	COGS (VPRS Cost)	COGS (VPRS Cost)				
Payment advice	G&A Overhead	G&A Overhead				
Payment advice	Base Cost	Base Cost				
Payment advice	Sales Qty - Base UOM	Sales Qty - Base UOM				
Payment advice	Distributor Rebate	Distributor Rebate				
Payment advice	Group Rebate	Group Rebate				
Payment advice	Vendor Rebate	Vendor Rebate				
Payment advice	Corporate Rebate	Corporate Rebate				
Payment advice	Oth Rebate Receivabl	Oth Rebate Receivabl				

Payment advice	Outbound Freight	Outbound Freight
Payment advice	C Freight Recovered	C Freight Recovered
Payment advice	S Freight Recovered	S Freight Recovered
Payment advice	Sales Rep Commission	Sales Rep Commission
Payment advice	Piggyback Label Cost	Piggyback Label Cost
Payment advice	Tracing Revenue	Tracing Revenue
Payment advice	Tracing Cost	Tracing Cost
Payment advice	Tracing Base Cost	Tracing Base Cost
Payment advice	Tracing Qty (Base)	Tracing Qty (Base)
Payment advice	Sample Sales	Sample Sales
Payment advice	Matl Master Cost	Matl Master Cost
Payment advice	Discount	Discount
Payment advice	Embroidery Cost	Embroidery Cost
Payment advice	Embroidery Revenue	Embroidery Revenue
Payment advice	Sales Upcharge	Sales Upcharge
Payment advice	Corp. Prog. Upcharge	Corp. Prog. Upcharge
Payment advice	Group Upcharge	Group Upcharge
Payment advice	Adtl.Handling/DS Fee	Adtl.Handling/DS Fee
Payment advice	Material handling fe	Material handling fe
Payment advice	Actual billed qty	Actual billed qty
Payment advice	Customer Incentive	Customer Incentive
Payment advice	CREDIT CARD CRG FEE	CREDIT CARD CRG FEE
Payment advice	Addl Delv Services	Addl Delv Services
Payment advice	Fuel Surcharge	Fuel Surcharge
Payment advice	Sales	Sales=VVR00 + VVR50 + VVR51 + VVR02 + VVR03 + VVR52 + VVR54
Payment advice	COGS	Cost of Goods Sold=VVC01 + VVC02 - VVC50 + VVC04 + VVC13
Medical surgical product list	Text	Text
Medical surgical product list	product Code	product Code

Medical surgical product list	Product Code name	Product Code name
Medical surgical product list	Product Code Level	Product Code Level
Medical surgical product list	Parent Product Code	Parent Product Code
Medical surgical product list	Parent Product Name	Parent Product Name
Medical surgical product list	Parent Product Level	Parent Product Level
Distribution Feedback	Major	Major
Distribution Feedback	MajorDesc	MajorDesc
Distribution Feedback	Interim	Interim
Distribution Feedback	InterimDesc	InterimDesc
Distribution Feedback	Sub	Sub
Distribution Feedback	SubDesc	SubDesc
Distribution Feedback	Class	Class
Distribution Feedback	ClassDescription	ClassDescription
Distribution Feedback	MfgCode	MfgCode
Distribution Feedback	MfgName	MfgName
Distribution Feedback	Report_Group	Report_Group
Distribution Feedback	Market	Market
Distribution Feedback	Territory	Territory
Distribution Feedback	Dist_TQ_TY	Distribution total quarter to year
Distribution Feedback	Dist_LQ_TY	Distribution last quarter to yesr
Distribution Feedback	Dist_MAT_TY	Distribution material total quarter to year
Distribution Feedback	Dist_Mat_LY	Distribution material las quarter to year
Distribution Feedback	All_TQ_TY	All quarter to year
Distribution Feedback	All_LQ_TY	All last quarter to year
Distribution Feedback	All_MAT_TY	All material total quantity total year
Distribution Feedback	All_Mat_LY	All material last year
Distribution Feedback	Major	Major

Appendix C: Cognitive Conceptualization of Analytic Problem

Management domain	Concepts	Attributes	Propositions	Requirements	Key questions
Sales					
	Sales Coverage	Customer coverage, Sales Rep coverage, Customer product affinity, Sales Rep product preference, Number of Sales Reps Type of Sales Reps Number of Sales Channels Type of Sales Channels	1) There are specific sales reps with "identifiably" low sales of specific product categories for similar customer types. 2)We need to know why. Reps sell what they know 3) Reps need additional support or training to increase their share of the wallet 4)Reps sell what is profitable to them 5) We have too many products in each client handled by different reps, which means we too many reps on a single account. 6) What is the proper balance of product baskets of existing customers clients? Is customer type granular enough segmentation scheme?	- Define product categories with similar sales coverage - Define customer types - Calculate average margin for each product category/customer type combination - Calculate product category percentage of sales by rep by customer type - Define percentage ranges - Chart number of reps within each percentage range for each customer type/product category combination - Calculate opportunity based on raising low sales areas - Next phase: attempt to determine hypotheses/correlations between these unsold basket elements and rep characteristics (for example, training sessions attended, tenure)	Is customer type granular enough segmentation scheme?
Pricing					
	Price trend	General price erosion trends	Specific accounts, GPOs, pricing methods and reps trigger general price erosion	Identify price reduction (i.e., erosion) "events" and identify correlations to specific reps, accounts, and GPOs. Chart the distribution of price trends by product category to investigate erosion and inflation misconceptions? For just the top x% of revenue?	Should we chart the distribution of price trends by product category to investigate erosion and inflation misconceptions? For just the top x% of revenue?

Pricing	Repricing	Inadvertent GPO repricing	ZCEP and ZREP one-time prices are picked up by GPOs and then shared with other customers, and thus inadvertently eroding prices more broadly	Correlate ZCEP and ZREP applications with subsequent price erosion within same-GPO orders, or perhaps overall erosion	Can the effects be seen in the data? We assume it cannot be found using pricing procedures. Can we identify/isolate subsequent price reductions for customers under the same GPO? Can we even identify customers under the same GPO with current data?
Pricing	Type/Size pricing	Type/size pricing	Inconsistent pricing of same product for similar customers is leaving money on the table: (1) activity to minimize price erosion can lower leakage; (2) activity to minimize high-priced outliers can reduce churn	Measure price variability of same products for similar customer segments (type/size); overall margins for similar customers with similar baskets can also be compared (to nullify arguments regarding taking minimal profits for one product to win business in other areas) Variability by rep and GPO can also be measured Margin variability may also be measured	Should we run the top x products just to sum up an opportunity?
Customer	Customer loss	Advanced Wound Care Lost Business	The lack of customer touches is leading to lost business in advanced wound care; buying behavior may be used as a predictor	- Identify loss events/baskets of loss events - Define rules to flag loss events - Create data set w/ loss events (break into two sets) - Difference between customers' w/ losses and w/o losses - Determine impact on order, revenue, profit - Identify drivers of loss events (for example, regression, neural network, genetic algorithm, principal component analysis) - Develop alert conditions ** Compare to proposed analysis/hypothesis	How can we estimate customer touches (with dates) and types? Do the sales reps keep sales engagement logs or are sales calls captured in the CRM system?
Product	Business loss	Product churn	Same as advanced wound care lost business, but for other major products and without the specific "touch" hypotheses. Purely for identification	See advanced wound care	

			and measuring purposes at this point	
Pricing	Pro	Price Hikes	The lack of published across the board price hikes doesn't give the sales force the cover to raise prices to match the Charlie process. I.e., Charlie price hikes don't effectively make it to the customer price	Measure before and after average prices and compare percentages increases to cost increase percentages. Identify any correlations to product, customer, rep etc.
Pricing	Price Optimizat ion	Freight is a soft spot	Price controls are more extensive and visible for products than for freight, so freight is being used as a lever for winning business and masking product price erosion	Measure scale and variability in freight collections. Identify correlations to product, customer, rep, higher-priced products (to see if a trade-off is being made), etc.
Sales	Back-end Revenue Leakage	Cash application / short Pay	Cash rec'd isn't matched to orders (which may be OK), and cash rec'd doesn't foot to orders	Compare payments rec'd to orders placed, and develop hypotheses from there (i.e., correlate differences with other factors such as certain projects, distribution centers, late shipments, etc.)
Price	Price Optimizat ion	Align compensation	Based on their compensation (customer price - GM/GP), sales management has room to give, and is therefore loose with approving sub-optimal price requests. We're lowering the price (i.e., leaking revenue) unnecessarily. Capping sales mgmt discretion or aligning sales mgmt comp with sales rep comp would minimize this type of price erosion.	Measure frequency, scale and variability by manager of "low price" approvals (for example, how many are "batch" approved?)

Price	Price Optimization	Blocked prices	Blocks are being released "easily." Prices could remain higher to avoid revenue leakage	Measure frequency, scale and variability by approver of "low price" approvals (for example, how many are "batch" approved?)	
	Price Optimization	Tiers and Commitments	GPO tiered pricing is awarded, but not monitored. Commitments are not monitored	TBD	Are tiers or commitments captured in the systems?
Sales	Back-end Revenue Leakage	Rebates	Not always collected	TBD	How are rebates managed/administered? Who is getting paid to do what? Is the company getting or are customers getting paid? Both? At what level? Account level? Order level? Product level?
Marketing	Effectiveness of promotions and marketing campaigns	Sample sales Promotional sales	Customers whose first purchase is a sample sale or promotional sale are usually given special price. How many of those customers continue to make purchases after the initial investment?		Are there ways of determining whether sales reps follow up after promotional or sample sale?

Appendix D: Analytic Attribution of Concepts

Management domain	*Atttributes*	*Description*	*Analytic attribution*
Sales and Distribution	sales representative preferences	Expressed as a preference score for each sales representative for each product, derived from the rank order of the volume of the products sold across customers, at the sales representative type level, sales rep tenure as well as customer type and product type	Sale Sales rep Preference score Sales rep Preference likelihood/expectation Preference margin distance Preference trend
	Sales commission	The contractual amount paid to the sales representative as compensation for the sale, this varies with the type of product.	Commission % of commission over margin levels, sales rep type Commission likelihood / expectations Commission margin distance Commission trend
	Sales representative penetration	Percentage of sale by a sales rep compared to all the sales by all the reps, normalized by company size	Sales % sales for sales rep compared to all sales reps Sales rep penetration likelihood/expectation Sales rep penetration margin distance Sales rep penetration trend
	sales representative categories	Grouping of sales representatives by their selling patterns and characteristics	Sale Sale cycle—interval
	sales representative average margin	Profit margin generated by each representative	Margin Profit margin contribution
	percent sales	Sales attributes to a sales representative	Sale Percent sales qty
	Percentage sale ranges	Percent sale ranges by sales representatives	Sale Sale range for sales rep
	Sales representative segmentation	Sales representative regrouping based on selling performance and margin contribution	Sale Sales rep segment
	Share of wallet	Proportion of product class in a particular customer, where there are multiple sales rep on the account determines the breakdown by sales representatives	Sale Proportion of sale by rep for customer, product and product + customer compared to other reps
	Sales representative profitability	Overall profitability of the sales representative compared to peers	Margin Rep sales margin compared to total margin
Pricing	general price erosion trends	Price erosion are situations in which a product price stays below the recommended price because of a price reduction event	price Price and percentage of price for product
	Inadvertent group repricing events	Inadvertent group repricing event is a situation of price erosion due to group pricing activity preceding the purchase of an item	Price type Repricing indicator
	Type/size price index (product level)	This is the ratio of price paid by a customer for a product divided by the average price paid by the customer group for the product	Price, Type size
	Price change impact	Changes in volume or frequency accompanying price changes	Price
	Blocked prices events	This is a type of price erosion event that occurs when a price that is blocked for any reason is manually released.	Blocked price status

	Price elasticity	Price elasticity is the measure of the change in volume with price	Price, Qty
	Revenue leakage	This is the difference in quantity or volume arising from a low or high price	Price
	Relative price	Ratio of the quoted price compared to the actual price for the product	Price
marketing promotions and effectiveness	promotions on purchasing habits	The number of purchases made with pricing designated as promotional price	Promotional sale indicator
	special pricing	This a pricing designation for specific purposes or specific situations	Special pricing indicator
	customer tenure	The length of time a customer has been purchasing from the company	Customer create date
	sales representative promotional performance and commissions	This is the performance of the sales representative during promotional period	Commission
product design impact	product churn	Event in which there is a swift from one product to another when it can be detected	Product order Product churn indicator
	Freight cost	Cost of transporting the material to the customer	Freight cost
	Product commissions alignment	Commissions allocation for a product and type of sales representative when applicable	Commissions Sale rep type
	Rebates	Payments from manufacturers for products sold. Apply these rebates to determine the true revenue attributable to products.	Rebate
	Product profitability	Margin associated with particular products	Margin
Customer trend and behavior	Group purchasing arrangements	A type of pricing arrangement based on grouping consumers together to form a purchasing group	GPO status
	Payment behavior	Patterns of payment adopted by consumers, for example, full payment for shipment, partial payments for shipments, scheduled payments, etc	Payment status
	Customer engagement	This is the degree to which the customer is engaged with the company, determined by the number of purchases and sales contact	Number of purchases
	Lost business	This is sales that were not made either as a result of the loss of the customer or reduction in the quantity of purchase as a result of changes in prices	Number of sales not made
	Customer profitability	The margin contribution of each customer to the bottom line	Margin
	Customer segmentation	Classification of customers into groups based on their life time valuations	Customer group status
	Customer churn		Customer churn from purchase expectation
	Customer life time value		Customer tenure Projections of live time value
	Margin expansion	Degree to which the profit margin can be increased as a percentage of current margins	Margin expansion projections
	Purchase blend	Combination of products that occur consistently together	Sale Order basket
	Selling gap	The gap between expected and actual selling	Sale Expected sale Selling gap

Customer Tenure	Number of years customer has been with company	# years
Customer Size (employees)	Size of the customer to be inferred from the number of employee	# employees
Customer Size (Beds)	Size of the customer based on the number of beds	# bed
Customer Size (Revenue)	Size of the customer based on their annual revenue	Purchase
Customer Type Size	Size of the customer based on the type	Type size
Customer Segment Size	Size of a segment of the customer considered to have a differentiating characteristic	Segment status Segment size
Customer monthly purchase growth rate	Monthly growth of purchases by the customer	Sale Monthly purchase rate
Price change	Change in price for the same time for the same customer from purchase to purchase	Price Price change index
Customer touch frequency	Number of times the customer has interacted with the company with the sales rep or other persons in the company	# orders # interactions
Customer Touch Interval	The interval between touches	Interval between orders
Customer touch to order	The interval from the touch to other	Interval variability
Cost of sale (Freight)	Cost of freight	Freight
Cost of sale (sales commission)	Cost of sales	Commission
Cost of sale (surcharge)	Cost of sales	Surcharge

Appendix E: Relation Property Matrix

Property name	Data type	Occurrence	Material master	Material measure	Material sales	Material description	Material plant	Material characteristics value	Material valuation	Customer master sale	Customer master general data	Competitive Item mapping	Material production record	Advanced wound skin care product	Customer Group	Commission rate	Material group	Order reason	Partner function	Product division	Sales document	Sales order	Sales order partner	Customer history	Customer change log	Customer partner	Payment advice	Medical surgical product list
Material	Character	9	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Creation date	date	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Name of person who created object	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material type	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material group	Character	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
Base unit of measure	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Size/dimensions	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Purchasing Value key	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gross weight	Numeric	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of weight	Numeric	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Volume	Numeric	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Volume unit	Numeric	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Division	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Universal Product Code	Character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Length	Numeric	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Width	Numeric	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Height	Numeric	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
External Material Group	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Order Unit of Measure	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Transportation Group	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Group – Ship Materials	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APO Demand Planner	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Attribute_1	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Attribute_2	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Attribute_3	Character	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alternative unit of measure	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Numerator for conversion to base UoM	Numeric	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Denominator for conversion to base UoM	Numeric	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
European Article Number (EAN) - obsolete!!!!!	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
International Article Number/Universal	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number category of International Article	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of dimension for length/width/height	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit weight	Numeric	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of measure contained in a unit of measure	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Internal characteristic	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of measure sort number	Numeric	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Leading proportion	Numeric	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Valuation based on the proportion quantity	Numeric	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Units of measurement usage	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of measurement of characteristic	Character	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sales organization	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Distribution channel	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Statistics group	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Volume Rebate Group	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Commission Group	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Distribution-chain-specific material status	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Date from which distr.-chain-spec. material status is valid	date	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum Order quantity in base UOM	Numeric	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum Delivery quantity in delivery no	Numeric	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum make-to-order quantity	Numeric	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Delivery unit	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of measure of delivery unit	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sales Unit	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Item category group from material master	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Delivery plant	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Pricing group	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Product Division	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Top 1001	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Product Rep type	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Freight Override	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vendor Code	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Latex Free	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Color Required	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Formulary item for Home Health Orders	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Catalog Database 4 Internet	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID for product attribute 5	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID for product attribute 6	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID for product attribute 7	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID for product attribute 8	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID for product attribute 9	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID for product attribute 10	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom item category	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HCPCS Code	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Block Group 1	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Block Group 2	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Block Group 3	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Block Group 4	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Block Group 5	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Canada Maple Leaf	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Do Not Reactivate	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Direct Only	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
To Be Discontinued	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Surplus Flag	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
No Re-route Flag	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Preferred Components	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ship 300 Exclude	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corporate Controlled Pallet (CCP)	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Custom Product Attribute P	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute Q	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute R	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute S	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute T	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute U	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute V	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute W	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute X	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute Y	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Custom Product Attribute Z	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Manufacturer Code	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Manufacturer Name (from table ZMFR)	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Manufacturer Item Number	Character	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Language	Character	5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
Material description	Character	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Material description in upper	Character	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Plant	Character	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Plant specific material status from MM	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ABC indicator	Character	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Purchasing group	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit of Issue	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material - MRP profile	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MRP type	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MRP controller	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Planned delivery time in days	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Good Receipt Processing Days	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Period Indicator	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lot size (materials planning)	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Procurement type	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Special procurement type	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Reorder Point	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Safety stock	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum lot size	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum lot size	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fixed lot size	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rounding value for purchase order qty	Integer	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum stock level	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ordering Costs	Currency	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dep. Requirement Ind. For Individual	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Schedule Margin Key	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Production Scheduler	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
In-house production type	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Over delivery Tolerance Limit	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Under Delivery Tolerance Limit	numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Loading group	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Service level	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Splitting Indicator	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Checking group for availability check	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fiscal Year Variant	Date	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Indicator: Take Correction Factor	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Base quantity for capacity planning	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Indicator: Automatic Purchasing Order	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Indicator: source list requirement	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Commodity Code/Import	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Material Country of Origin	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Region of Origin	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Profit Center	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stock in transit	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Planning Time Fence	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Costing Lot Size	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Special Procurement Type of Costing	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Production Unit	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Issue Storage Location	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MRP Group	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Takt Time	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Storage costs indicator	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maintenance Status (Views Created)	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Storage Location for EP	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quota Arrangement Usage	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pallet Quantity	Numeric	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Deployment Center	Character	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rounding value release strategy	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Safety Time Indicator	Character	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Safety Time Days	Numeric	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Class (Class)	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Class Type (Klart)	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Item Class	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Production Group	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PATTERN ID	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fabric Type	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Spread Type	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Style of Garment	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Color of Garment	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fabric	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Size of Garment	Character	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dimension 1 – Length	Numeric	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dimension 2 – Width	Numeric	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Valuation area	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Valuation type	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Deletion flag for all material data of a valuation type	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total valuated stock	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Value of total valuated stock	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Price control indicator	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Moving average price/periodic unit price	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standard price	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Price unit	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Valuation class	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Value based on moving average price (only with price ctrl S)	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total valuated stock in previous period	Numeric	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Value of total valuated stock in previous period	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Price control indicator for previous period	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Moving average price/periodic unit price in previous period	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standard price in the previous period	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Price unit of previous period	currency	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Origin as subdivision of cost	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Costing overhead group	Character	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Costing W/ Quantity Structure	Numeric	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Client	Character	13	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0
Customer number	Character	5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
Pricing procedure assigned to this customer	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer group	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Freight Default	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Access Program	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Confirmation Preference	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Deletion indicator for customer (at sales level)	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Division	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer statistics group	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sales organization	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Distribution channel	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Delivering plant	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Invoice Preference	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Invoice list schedule (calendar identification)	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Central order block for customer	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer account group	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bed Count	Numeric	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sales Office	Character	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Price group (customer)	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Terms of payment key	Character	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Central deletion flag for master record	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Name 1	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Name 2	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Name 3	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Name 4	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
City	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Post office box	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.O. Box postal code	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Postal code	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Region (State, Province, County)	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Street and house number	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
First telephone number	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Account Group	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Competitive_Item	Character	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Competitive_Desc	Character	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Medline_Item	Character	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Medline_dec	Character	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dist_I_Num	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d_mfg_prod	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D_MFG_ID	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UM	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dist_num	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MFG_ID	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HPIS_Cat	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brand	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cat_Desc	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UM_CONV	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mfg_name	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Class	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
class_desc	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Major	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maj_Desc	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interim	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Int_desc	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sub	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sub_Desc	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minor	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Min_desc	Character	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Title	Character	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Item	Character	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Category	Character	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer group	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Name	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
Valid_From	Date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Valid_To	Date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
BP__Start	Date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
BP__End	Date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Commission_Rate	Numeric	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
User_Name	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Date	date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Matl_grp_descr_	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Order Reason	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Description	Character	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
Partner Function	Character	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0
Product_Division	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Sales_Doc__T	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
PO Type (Order Method)	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Order Reason Code	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Pricing Date	date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sales Order Type	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sales order Number	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
Sales Order Line	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Material Number	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Net Value	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
QTY	Integer	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sales UOM	Character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Condition Record	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Condition Value	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Extended Condition Value	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Active Pricing Condition - will need a formula here	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Application	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Valid From	Date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Valid To	Date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Object Class	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Object value	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Document change number	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
User name of the person responsible in change document	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Creation date of the change document	date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Time changed	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Transaction in which a change was made	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Planned change number	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Change number of the document created by this change	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Flag that changes were generated from planned changes	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Application object change type (U, I, E, D)	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Language Key	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
3-Byte field	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Table Name	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Changed table record key	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Field Name	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Change type (U, I, E, D)	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Flag: X=Text change	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Change documents, unit referenced	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Change documents, referenced currency	Character	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
New contents of changed field	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Old contents of changed field	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Cond.type	Character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amount	Currency	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cond.value	Currency	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer (typically sold to)	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Customer (the byproduct of the sold to)	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Sales Order	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Document Date	date	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Invoice Number	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Invoice Line	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Billing Type	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Revenue	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
COGS (VPRS Cost)	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
G&A Overhead	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Base Cost	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Sales Qty - Base UOM	Integer	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Distributor Rebate	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Group Rebate	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Vendor Rebate	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Corporate Rebate	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Oth Rebate Receivabl	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Outbound Freight	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
C Freight Recovered	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
S Freight Recovered	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Sales Rep Commission	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Piggyback Label Cost	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Tracing Revenue	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Tracing Cost	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Tracing Base Cost	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Tracing Qty (Base)	Integer	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Sample Sales	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Matl Master Cost	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Discount	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Embroidery Cost	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Embroidery Revenue	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Sales Upcharge	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Corp. Prog. Upcharge	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Group Upcharge	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Adtl.Handling/DS Fee	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Material handling fe	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Actual billed qty	Integer	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Customer Incentive	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
CREDIT CARD CRG FEE	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Addl Delv Services	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Fuel Surcharge	Currency	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Sales	Currency	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cost of Goods Sold	Currency	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Text	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
product Code	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Product Code name	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Product Code Level	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Parent Product Code	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Parent Product Name	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Parent Product Level	Character	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Totals		410	24	22	62	4	62	14	21	21	0	4	21	3	3	8	3	4	4	4	4	19	4	0	24	4	41	7

Appendix F: Ontology Learning

No	Ontology class	Context	Usage	Model Class	Structure identification	Formal specification	Resolution options
1	Concept, Variable, Attribute	Characteristics of a subject	Context for expression of properties	Concept or Variable Model	Cognitive Maps, Concept Maps	Heuristics, Policies, Rule of Thumb, Expert Rules	Explanatory, Consequence, Potency
2	Entities	"A thing" defined by related set of data attributes, concepts or variables	Expression of related concepts and variables	Entity Model	Entity structures (tuples)	Entity Normal Form	Entity resolution for example satisfiability, subsumption, etc.
3	Evidence / Facts	Indicative concepts, variables or attributes	Vectors and matrices related to concepts or variables	Multidimensional Model	Multidimensional structures (array of tuples)	Measures, cubes, D-structures, F-structures	Collinearity/orthogonality, dimension reduction, granularity determination
4	Effects	Impact of evidence / facts on characteristics of interest	Attribution of evidence on a characteristic of interest	Effect Model	effect model specification	Regression Equation, Factor (covariate) structure rules & logic, Ordinary Differential Equations (ODE)	Ordinary Least Squares, Regression Coefficients, ODE solution
5	Events	Occurrences of interest	Represents outcomes of interaction of concepts	Event Model	Event/fault trees, discrete/continuous event model	Multivariate statistics, Partial Differential Equations (PDE)	Generalized Least Squares, Generalized Regression Coefficients, PDE solutions
6	Influence	The impact of occurrences	Characterization of the size of impact of occurrences	Influence Model	Bayesian Network, Influence diagram	Classical and Bayesian Probability	Distribution Parameters, Maximum Likelihood Estimates, Odds Estimates,
7	Preference	Resolution of influences	Desired influence	Preference Model	Weighted / Modified Preference Diagrams	Probability, Weights, Scores	Agreement/disagreement assessment
8	Case	Logical organization of related influences	Homogeneous sets with similar experiences and characteristic	Case Specific Model	Constrained Bayesian Network, Constrained Influence Diagram	Constrained probability	Constrained mathematical programming, Constrained Evaluation / Evolutionary Algorithms,

							Case-based reasoning
9	Decision / Choices	Integration of preference, goal and activity resolutions	Determination of an approach to a situation of interest	Decision Models	Decision Trees, Networks & Forests, Influence Diagram	Utility, fuzzy logic, reasoning, learning	Simulation, Optimization, experimentation
10	Action / Activity	Work products and the outcomes	Performance profile of activities and actions	Action / Planning Model	Activity Tree, Network or Forests	Schedule, Sequence / Order	Program Evaluation and Review Technique, Critical Path Method, Marginal Cost
11	Resource / Entity	Concrete or abstract non-information objects within the space of interest	Players in the space of interest	Resource Model / Capacity Model	Resource / Capacity Charts	Relative workload estimates	Point of Failure Method, Resource Capacity
12	Function/ Task / Process	Unit of work or activity	Characterization of the Work efforts within the space of interest	Task / Process Model	Task / Throughput Charts	Relative throughput estimates	Point of resistance, etc
13	Goal	Defined expectations of behavior and outcomes	Characterization of expected behavior or outcomes	Goal Model	Benchmarks and Thresholds	Relative benchmarks and Threshold estimates	Actual to Goal variance
14	Cycle	Defined or expressed regularity in occurrence	Characterizes the reoccurrence of interest	Time series	Cycle time	Cycle effect	Cycle variances
15	Horizon	Defined or expressed period of regularity in cycle to cycle changes	Characterizes regularity in cycles	Time series	Horizon time	Horizon effect	Horizon variances
16	Emergence	Discovered irregularity in occurrence	Irregularity in occurrence	Multidimensional panel	Structural break	Formal break	(re)solution break

Note: Some terms, for example, unexplainably, subsumption, and others used in this appendix are technical so are not found in the English dictionary

Appendix G: Analytic Formulation

	Analytic formulation phase →	Structural			Formal			Resolution		
No	Model Level → Analytic formulation	Data	Domain	Decision	Data	Domain	Decision	Data	Domain	Decision
1	Value List	X								
2	Objective Hierarchy	X								
	Means-Ends Diagrams									
3	Relational structures	X								
4	Knowledge Chain		X							
5	Value Tree / Network		X							
6	Influence Diagram		X							
7	Decision Tree / Network			X						
8	Event Trees / Network			X						
9	Failure Tree / Network			X						
10	Fault Tree / Network			X						
11	Belief / Bayesian Networks			X						
12	Causal Loops Diagrams		X							
13	Causal Models		X							
14	Relevance Diagrams		X							
15	System Flow Diagrams		X							
16	Knowledge Maps		X							
17	Semantic Networks		X							
18	Discrete Event Model		X							
19	Systems Dynamics Model		X							
20	Statistical Moments				X					
21	Factor model				X					
22	Rule based derivation				X					
23	Weights				X					
24	Scores				X					
25	Arithmetic Functions					X				

24	Statistical Equations	X				
25	Mathematical Algorithms	X				
26	Utility Models		X			
27	Probability Models		X			
28	Fuzzy Logic Models		X			
29	Ordinary Least Square parameters			X		
30	Generalized Least Square parameters			X		
31	Maximum Likelihood parameters			X		
32	Backward Reasoning parameters			X		
33	Recursion Integration parameters				X	
34	Numerical Integration parameters				X	
35	Simulation parameters				X	
36	Mathematical Programming parameters					X
37	Evolutionary Algorithms parameters					X

Appendix H: Data Engineering Transformation Functions

No	Data	Transformation function
1	Numerical	as is
		range
		n-order moments
		Autogressive correlation
		Binning
		Ordering
2	Categorical	Count
		Distinct count
		Frequency
		Probability
		Conditional probability
		Unnormalized distribution
		Normalized distribution
3	Text	Word, topic count
		Distinct word, topic count
		Word vectors
		Sequence correlation
		Sub-sequence correlation
4	Timestamp	Millisecond
		Second
		Minute
		Hour
		Time of day
		Day
		Day of week
		Week
		Week of month
		Week of year
		Month
		Month of year
		Year
5	Series	Average
		Max
		Min
		Count
		Variance
		Recent(k)
		Fast fourier transformation
		Discrete Wavelent transformation
		Autocorrelation coefficients
6	Sequence	Count
		Distinct count
		Vector transformation
		Correlated subsequences
7	Matrix	Correlation
		Eigenvector, Eigen values
		Principal component
		Factor
		Perceptron
		Support vector functions
		Tensor decomposition

Appendix I: Analytic Formulation Catalog

No	Model Format	Model Name	Restriction
1	Model a	Univariate	One variable
2	Model y =x	Bivariate Correlation	Max of 2 variates at a time. Approach depends on the data type of the criterion and response variates, includes Spearman, Pearson, Krukal Wallis, Chi-Squared, ANOVA
3	Model y=x;	simple regression	Numeric dependent variable and numeric independent variate. Categorical variates have to be dummy coded
4	model y=x z;	multiple regression	Numeric dependent variable and numeric independent variate. Categorical variates have to be dummy coded
5	model y=x x*x;	polynomial regression	Numeric dependent variable and numeric independent variate. Categorical variates have to be dummy coded
6	model y=x z;	Multiple discriminant	Categorical dependent variable and numeric independent variable
7	model y1 y2=x z;	multivariate regression	Numeric dependent and independent variables
8	model y=a;	One-way ANOVA	Numerical dependent and categorical independent
9	model y=a b c;	main effects model	Numerical dependent and categorical independent

10	model y=a b a*b;	factorial model (with interaction)	Numerical dependent and categorical independent
11	model y=a b(a) c(b a);	nested model	Numerical dependent and categorical independent
12	model y1 y2=a b;	multivariate analysis of variance (MANOVA)	Numerical dependent and categorical independent
13	model y=a x;	analysis-of-covariance model	Numerical dependent and categorical or numeric independent
14	model y=a x(a);	separate-slopes model	Numerical dependent and categorical or numeric independent
15	model y=a x x*a;	homogeneity-of-slopes model	Numerical dependent and categorical or numeric independent
16	Model y1=a x11 x12; y2= a x21 x22; y3=a x31 x32	Structural Equation	dependent variates are numeric, while independent variates can be numeric or categorical
17	Model y1 y2 y3=a x1 x2 x3	Canonical Correlation	Most generalized form of all models. Dependent variables numeric or categorical and independent variables numeric or categorical
18	Model y=a b c	Conjoint model	Numeric dependent and categorical independent
19	Model y=x1 x2	Linear Probability model	Categorical dependent and numeric independent
20	Model (x)(a)	Factor Model	Categorical and numeric variates
21	Model (x)(a)	Principal Component	Categorical and numeric variates
22	Model (x)(a)	Cluster	Categorical and numeric
23	Model (x)(a)	Correspondence	Categorical variates
24	Model (x)(a)	Multidimensional Scaling	Categorical and numeric variates

25	Model y=y1 y2; y1=x11 x12; y2=x	Decision Tree	Categorical
25	Model y=y1 y2; y1=x11 x12; y2=x	Neural Network, Deep learning, Boltzman's machines, Support vector machine	Categorical
26	Model y=y1 y2; y1=x11 x12; y2=x	Genetic Algorithms, Evolutionary algorithms	Categorical
27	Model y=y1 y2; y1=x11 x12; y2=x	Markov Chain / System Dynamics Autoregressive models	Categorical or numeric
28	Model y=y1 y2; y1=x11 x12; y2=x	Simulation	Numeric
29	Model y=y1 y2; y1=x11 x12; y2=x	Optimization	Numeric
30	Model y=y1 y2; y1=x11 x12; y2=x	Mathematical / Numeric	Numeric

Appendix J: Analytic Results: Profit Margin

Transactions / actions	Margin growth coefficient	Determinant	Adjusted influence	Influence Proportion
Customer	0.138	0.93	0.12834	1.55%
Marketing	0.013	0.8	0.0104	0.13%
Pricing	0.0714	0.74	0.052836	0.64%
Sales and distribution	0.0913	0.9	0.08217	1.00%
Product	0.128	0.72	0.09216	1.12%
Time	0.009	0.8	0.0072	0.09%
Customer*Marketing	0.22	0.8	0.176	2.13%
Customer*Pricing	0.31	0.78	0.2418	2.93%
Customer*Sales/Distribution	0.25	0.7	0.175	2.12%
Customer*Product	0.18	0.7	0.126	1.53%
Customer*Time	0.16	0.6	0.096	1.16%
Marketing*Pricing	0.09	0.8	0.072	0.87%
Marketing*Sales/Distribution	0.12	0.5	0.06	0.73%
Marketing*Pricing	0.09	0.7	0.063	0.76%
Marketing*Product	0.07	0.6	0.042	0.51%
Marketing*Time	0.13	0.6	0.078	0.94%
Pricing*Sales/Distribution	0.2	0.8	0.16	1.94%
Pricing*Product	0.3	0.9	0.27	3.27%
Pricing*Time	0.25	0.8	0.2	2.42%
Product*Time	0.21	0.6	0.126	1.53%
Customer*Marketing*Pricing	0.38	0.7	0.266	3.22%
Customer*Marketing*Sales&Distribution	0.42	0.7	0.294	3.56%
Customer*Marketing*Product	0.45	0.8	0.36	4.36%
Customer*Marketing*Time	0.41	0.6	0.246	2.98%
Marketing*Pricing*sales&Distribution	0.25	0.6	0.15	1.82%
Marketing*Pricing*product	0.21	0.7	0.147	1.78%
Marketing*Pricing*Time	0.38	0.7	0.266	3.22%
Marketing*sales/Distribution*Product	0.42	0.8	0.336	4.07%
Marketing*sales/Distribution*Time	0.45	0.6	0.27	3.27%
Marketing*product*time	0.41	0.55	0.2255	2.73%
Customer*Marketing*Pricing*Sales/Distribution	0.52	0.6	0.312	3.78%
Marketing*product*time	0.41	0.55	0.2255	2.73%
Customer*Marketing*Pricing*Sales/Distribution	0.52	0.6	0.312	3.78%
Customer*Marketing*Pricing*Product	0.57	0.6	0.342	4.14%
Customer*Marketing*Pricing*Time	0.56	0.8	0.448	5.43%
Customer*Marketing*Pricing*Sales/Distribution*Product	0.57	0.9	0.513	6.21%
Customer*Marketing*Pricing*Sales/Distribution*Pricing	0.62	0.8	0.496	6.01%
Customer*Marketing*Pricing*Sales/Distribution*Product*Time	0.67	0.6	0.402	4.87%
Customer*Marketing*Pricing*Sales/Distribution*Product	0.63	0.7	0.441	5.34%
Customer*Marketing*Pricing*Sales/Distribution*Product	0.69	0.7	0.483	5.85%

Appendix K: Analytic Results: Profit Margin

Management domain	Features	margin growth	Determination coefficient	Adjusted margin growth influence	Contribution to growth within dimension	Adjusted contribution	Cumulative Contribution
product design impact	Product profitability	0.720176732	0.804778633	0.579582846	0.3253226	7.23%	7.2%
product design impact	Rebates	0.784295201	0.732871385	0.57478751	0.322630955	7.17%	14.4%
Sales and Distribution	sales representative preferences	0.766971255	0.456471808	0.350100755	0.136760911	6.44%	20.8%
product design impact	Freight cost	0.829797329	0.587367781	0.487396215	0.273577807	6.08%	26.9%
Customer trend and behavior	Customer Size (Revenue)	0.97134557	0.78899108	0.766382991	0.109471964	4.49%	31.4%
Customer trend and behavior	Customer touch to order	0.976191548	0.623446858	0.608603553	0.086934375	3.57%	35.0%
Customer trend and behavior	Group purchasing arrangements	0.701797356	0.800166203	0.561554526	0.08021378	3.29%	38.3%
Pricing	Relative price	0.880831772	0.751596344	0.66202994	0.214690634	2.73%	41.0%
Customer trend and behavior	Customer touch frequency	0.915325222	0.493966585	0.452140074	0.064584761	2.65%	43.6%
Customer trend and behavior	Customer Size (employees)	0.678012968	0.642257002	0.435458576	0.062201936	2.55%	46.2%
Pricing	Inadvertent group repricing events	0.717887563	0.806523872	0.578993457	0.187762615	2.39%	48.6%
Customer trend and behavior	Payment behavior	0.989040783	0.407705929	0.403237791	0.057599443	2.36%	50.9%
Customer trend and behavior	Customer life time value	0.79922891	0.495243224	0.395812703	0.056538825	2.32%	53.3%
Customer trend and behavior	Margin expansion	0.868642974	0.44524817	0.386761694	0.055245958	2.27%	55.5%
Customer trend and behavior	Customer engagement	0.963126931	0.398446482	0.383754538	0.054816408	2.25%	57.8%
Customer trend and behavior	Selling gap	0.470480162	0.779480092	0.36672992	0.052384572	2.15%	59.9%
Pricing	Revenue leakage	0.779845271	0.651157749	0.507802291	0.16467593	2.10%	62.0%
Customer trend and behavior	Price change	0.647851603	0.48741613	0.315773322	0.04510581	1.85%	63.9%
Pricing	Type/size price index (product level)	0.761868077	0.576749437	0.439406984	0.142495918	1.81%	65.7%
Sales and Distribution	Sales commission	0.463545188	0.535363865	0.248165343	0.096941575	1.59%	67.3%
Customer trend and behavior	Customer Tenure	0.326643911	0.798068163	0.260684107	0.037236736	1.53%	68.8%
product design impact	Product commissions alignment	0.26974362	0.436574053	0.117763066	0.066100967	1.47%	70.3%
Pricing	Price elasticity	0.959649552	0.346852661	0.332857001	0.107942672	1.37%	71.6%
Pricing	Price change impact	0.479536962	0.660346408	0.316660511	0.102690289	1.31%	72.9%
Customer trend and behavior	Customer Segment Size	0.326457979	0.633378032	0.206771312	0.029535705	1.21%	74.2%
Customer trend and behavior	Cost of sale (sales commission)	0.3265924	0.607595655	0.198436123	0.028345086	1.16%	75.3%
Sales and Distribution	Sales representative penetration	0.247365861	0.692465663	0.171292365	0.066912452	1.14%	76.5%
Sales and Distribution	Share of wallet	0.202296792	0.657784861	0.133067767	0.051980662	1.12%	77.6%

Sales and Distribution	Sales representative profitability	0.459534018	0.432210481	0.198615419	0.077585739	1.09%	78.7%
Sales and Distribution	sales representative categories	0.058897687	0.4677201	0.027547632	0.010761014	1.09%	79.8%
Customer trend and behavior	Customer Touch Interval	0.238844951	0.731310847	0.174669903	0.024950263	1.02%	80.8%
Customer trend and behavior	Customer profitability	0.434365807	0.394188528	0.171222018	0.024457759	1.00%	81.8%
Customer trend and behavior	Customer segmentation	0.332425217	0.486319918	0.161665004	0.023092613	0.95%	82.7%
Customer trend and behavior	Customer monthly purchase growth rate	0.434192384	0.364546904	0.15828349	0.022609589	0.93%	83.7%
marketing promotions and effectiveness	sales representative promotional performance and commissions	0.408793048	0.688797137	0.281575481	0.365784621	0.92%	84.6%
marketing promotions and effectiveness	special pricing	0.522116151	0.468043864	0.244373261	0.317456548	0.80%	85.4%
Customer trend and behavior	Cost of sale (Freight)	0.239856267	0.555664447	0.1332796	0.019037974	0.78%	86.2%
marketing promotions and effectiveness	promotions on purchasing habits	0.397142411	0.57631145	0.228877719	0.297326845	0.75%	86.9%
Customer trend and behavior	Lost business	0.157391907	0.769660468	0.121138329	0.017303686	0.71%	87.6%
Customer trend and behavior	Purchase blend	0.330830996	0.347797374	0.115062151	0.016435751	0.67%	88.3%
Pricing	general price erosion trends	0.239552237	0.567180792	0.135869427	0.044061291	0.56%	88.8%
Sales and Distribution	percent sales	0.026024291	0.770856954	0.020061006	0.007836491	0.48%	89.3%
Customer trend and behavior	Customer Type Size	0.109684201	0.723572093	0.079364427	0.011336603	0.46%	89.8%
Sales and Distribution	Percentage sale ranges	0.581785292	0.888589257	0.51696816	0.201944828	0.46%	90.3%
Pricing	Blocked prices events	0.142885531	0.770032482	0.1100265	0.035680651	0.45%	90.7%
Customer trend and behavior	Customer churn	0.197153717	0.365224963	0.072005459	0.010285431	0.42%	91.1%
Sales and Distribution	sales representative average margin	0.801100103	0.801185393	0.641829701	0.250719867	0.34%	91.5%
product design impact	product churn	0.068445838	0.321915726	0.022033792	0.012367672	0.27%	91.7%
Customer trend and behavior	Customer Size (Beds)	0.099365277	0.37929598	0.03768885	0.005383565	0.22%	92.0%
Sales and Distribution	Sales representative segmentation	0.424184744	0.59478652	0.252299368	0.098556461	0.20%	92.2%
Customer trend and behavior	Cost of sale (surcharge)	0.076889656	0.445357531	0.034243387	0.004891407	0.20%	92.4%
marketing promotions and effectiveness	customer tenure	0.045881863	0.326020975	0.01495845	0.019431986	0.05%	92.4%