Data Management Model Selection: IT Professionals' preferences

Dr. Gholam Ali Shaykhian, -

Gholam Shaykhian has received a Master of Science (M.S.) degree in Computer Systems from University of Central Florida and a second M.S. degree in Operations Research from the same university and has earned a Ph.D. in Operations Research from Florida Institute of Technology. His research interests include knowledge management, data mining, object-oriented methodologies, design patterns, software safety, genetic and optimization algorithms and data mining. Dr. Shaykhian is a professional member of the American Society for Engineering Education (ASEE).

Dr. Mohd Abdelgadir Khairi, Najran University

I, Mohamed Khairi, my bachelor degree in computer science. I did my Masters in system science from University of Ottawa, Canada. My PH.D was in "Master Data Management" from University of Phoenix. I have over 20 years of experience in IT industry - ten of them with Microsoft in Redmond, WA. Currently I'm assistant professor at University of Najran. In addition of teaching and Research I'm coordinator of graduation projects and field training for computer and information system college.

Jinan Ziade, University of Phoenix

My name is Jinan Ziade. Currently a Doctoral student, ABD, at University of Phoenix working on my doctorate of management in Organizational Leadership, specializing in Information System Technology. I worked for AT&T and had 14 years of experience in the Advertising and Publishing industry. During that time, I held many positions including senior graphic design, team leader, and project coordinator. Drawing on the many years of experiences and the varied background in the design and layout techniques I helped the company made decisions about allocating resources by providing the best foundation to analyze the workflow of knowledge and find solutions to improve employees' productivities and performance. I have also lead the United Way/AT&T campaign for the five consecutive years (2008, 2009, 2010, 2011, and 2012). In this role, I was responsible for developing new innovative ideas to raise funds and gained excellent supervisory and leadership skills. It was a great experience seeing all the work being done to ensure the United Way was spending the dollars collected wisely. Contributing time and effort was a great investment in the community. I enjoy traveling around the world with my son and my husband to broaden our awareness of other cultures, pick up a language or two, and meet interesting people. We have had some wonderful travels and enjoyed learning new cultures and discovered new excitements.

Data Management Model Selection: IT Professionals' preferences

Gholam Ali Shaykhian, Ph.D.

Mohamed Abdelgadir Khairi, Ph.D.

Jinan Ziade, ABD

Abstract

Data Management Models selection (Centralized Data Model or Federated Data Model) for managing organization data is influenced by many factors. The goal is to determine the best architectural model for managing enterprise data; and help organizations to select an architectural model. The study compared and contrasted the federated and centralized data models within the context of business and technology requirements using a survey method. Each model is ranked using the following set of applicable factors: cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, deployment, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability.   The purpose of this paper is to compare/contract expert testimonies of data management with IT professionals' practical aspect of the data management technologies.

Introduction

Business Intelligence (BI) communities within Information Technology (IT) departments strive to select a suitable Master Data Management (MDM) architectural model that help bridge the gap among their organizations, technologies, and their customers. "Today, master data management (MDM) provides new tools, techniques and governance practices to enable businesses to capture, control, verify and disseminate data in a disciplined fashion. Combined with tools for data quality management, this provides the trusted information foundation that companies base their analytics on" [8]. The MDM centralized and federated architectural models play a major role in managing master data. The MDM centralized model allows an organization to organize and manage master data into a central repository [5], [9]. On the contrary, MDM federated model does not keep the master data in one database instead allow users to query the master data from multiple sources using MDM tool [3].

Determining which architectural model is suitable for an organization depends on several factors; including use of the master data, number of applications (domains) that will use

the master data [6], [7], development and availability costs, delivery schedule, performance, efficiency, limitations, risk, training, operations, compliances, deployment, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability.

The study compared and contrasted the federated and centralized data models within the context of business and technology requirements through conducting a survey. The survey ranked each model using the following set of applicable factors: cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, deployment, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability. The research approach was quantitative with qualitative embedded design research through the use of surveys and data analysis.

Many IT organizations such as Microsoft have many databases spread across several divisions with no apparent consensus about specific data model architecture or shared MDM principles. Some divisions use Federated Data Model (FDM) model and other divisions use Centralized Data Model (CDM) whereas the other divisions don't use MDM. This quantitative study report the IT Experts testimony and their preferences and experiences in selecting an architectural model.

An online web survey was available in multiple MDM communities Wikis and Blogs who work directly with Master Data; individuals were invited to provide responses anonymously. The online survey questions were related to MDM technologies, architectural attributes, and business requirements that influenced the selection of MDM model. The responses from approximately 80 MDM users were used for the analysis of this proposed research.

Data Collection

Federated and centralize data models are both process driven, tool centric architectural models that when implemented as prescribed can benefit business by reducing master data maintenance cost and improving efficiency. The selection of an architectural model is tightly coupled with the business and technology requirements. The number of concurrent applications (domains) of master data is another major factor that can affect several factors such as cost, performance, scalability, maintainability, reliability, and availability of master data management system.

This study focuses specifically on gathering information on fundamental key factors in implementation of an MDM; the key factors (cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, security, accessibility, dependability, data quality, deployment, stability, maintainability, reliability, availability, flexibility, scalability, and predictability metrics) were presented to experienced MDM practitioners (users and ETs) who have in-depth knowledge of master data management

and its importance to an organization. The survey involves a literature study of major IT organizations "how they manage their master data" augmented with quantitative design to present meaningful data set that pertains to IT organizations. Experts testimonies, individuals with deep knowledge about MDM and wrote well respected articles and books about MDM, serve as a key component to collect data and satisfy the qualitative analysis portion of an ongoing struggle.

Creswell stated that a quantitative study starts with a theory or an assumption and then looks for evidence to support the theory [1]. According to Creswell, quantitative research is appropriate for justifying the need for the research and to determine the direction of the study. The survey through quantitative study will not give the researcher an entire account of user preferences and experiences [1].

Data Analysis

Participants' responses to the survey questions on standards-based reform were recorded using yes and no responses, multiple choice responses, or a 5-point Likert-type response scale. In scoring the 5-point Likert-type response scale, each response was allocated 1, 2, 3, 4, or 5 points for each of the responses ranging from strongly disagree to strongly agree. For all other questions, participants responded to multiple choices answers resulting in a point scale from 1 to 5. The Likert –type rating scale was chosen to measure the strength of agreement with a clear statement (Delaney, 2004). In analyzing results of Likert – type responses, a median value can be calculated based on a 5 point scale wherein the closer the median is to 5, the more agreement there is to the response.

The questionnaire administered to the MDM professionals who are members of MDM wikis and blogs Participants accessed the questionnaire electronically through the Survey Monkey Web site. Data were tabulated on a different designated site for analysis using SurveyMonkey (SurveyMonkey, 2010). The use of SurveyMonkey for the return and initial collection of data provided a centralized collection point.

The potential responsiveness of members through use of the MDM Wikis and blogs supported an opportunity to obtain greater than 30 participants Creswell suggests findings with a number of participants of 30 or more is worthy of discussion [1]. Although the final count of 80 participants was more than the suggested number of 30, this study did present data points that related variables exceeding 30 and worthy of discussion for a correlational study.

The analysis phase included a data review to ensure no mistakes were made in the data extraction phase, the data entry phase, and the data connectivity phase. Descriptive statistics were generated using Microsoft Excel and SPSS. The descriptive statistics allowed for a quick review of the data for reliability, trends, anomalies, and possible outliers. In general, the analysis allowed the researcher to observe where trends existed.

Results

Participants in the survey were all IT professionals who directly work with MDM models. Table 1 shows the demographics of individuals who participated in the survey. The importance of the demographics in this study is to give the reader an idea about the background of professional participants in this study. The organization size they working on, the software they use, and their role with respect to MDM are the demographics data collected. The sampling frame for the study was selected using 80 IT professionals who directly use MDM models; these 80 IT professionals were from different sectors of the IT business. The count and the percentages of the respondents is shown in Table 1. The data reveals that 61.3 % of the respondents are users of MDM models. Users are considered as a customer of MDM application. They provide the requirements and understand what business needs. Their importance comes from their ability to drive requirements and business plans. Leaders take special care to their feedback for planning MDM implementation. Table 2 shows 62.5% of respondents are employed by large IT organizations (organizations with 500 or more employees). A large organization is capable of investing in training, software and hardware. This information is necessary to know how large organizations react to the factors in comparison to small one. The relationship between factors influencing model selection and organization size is not part of this study.

In addition to this, as shown in Table 3, a majority of respondents are currently using Siperian MDM Hub (21.3%) in their organizations. Siperian MDM solution provides support for FDM and CDM implementation. The result indicates the majority using Siperian is a multi-domain operational MDM vendor that Forrester named as a leader in last "Forrester Wave for Customer Hubs in Q3 of 2008" [2],[3].

Table 1
Count and Percentages for Individual's role

	Count	Percent
Developer	9	11.25
User	49	61.25
Architect	0	0
Quality Assurance	4	5
Data Consumer	2	2.5
Decision Maker	4	5
Program Manager	6	7.5
Solution Delivery	6	7.5
Others	0	0
Total	80	100.0

Majority of MDM professionals who participated in the survey prefer CDM model over FDM in respect to Cost (70%), Delivery (69%), Performance (75%), Efficiency (77%), Risk (72%), Training (69%), Operation (75%), Compliance (77%), Security (77%), Accessibility (73), Dependability *70%), Data Quality (77%), Stability (77%), Maintainability (56%), Reliability (73%), Scalability (69%), Predictability (68%), and Deployment (67%). Participants had no preferences as to Limitation (75%), Availability (82%), Flexibility (79%) factors. Figure 1 shows, the participants' responses for all 21 factors.

Table 2
Count and Percentages
Organization Size

	Count	Percent
Small (10-99)	20	25
Medium (100-499)	10	12.5
Large (More than 500)	50	62.5
Total	80	100.0

Table 3
Count and Percentages
Software Solution

	Count	Percent
D&B / Purisma Data Hub	5	6.25
Oracle Customer Data Hub	6	7.5
TIBCO CIM	3	3.75
Siperian MDM Hub	17	21.25
Orchestra Networks	3	3.75
Kalido MDM	6	7.5
Data Foundations One Data	5	6.25
Vision Ware MultiVue	4	5
IBM InfoSphere MDM Server	6	7.5
SAP NetWeave	9	11.25
DataFlux MDM	1	1.25
Others	1	1.25
Total	80	100.0

Conclusions and Recommendations for Further Study

Understanding the 21 factors explained in this paper can help to enhance the implementation of MDM models and could enable the IT Managers to recognize the importance of the MDM models early and factor it into their strategy.

One recommendation for further study is to explore the "why" question in depth to finding additional factors, conflicts in model selection, and the cloud computing. The recommendations for leaders were to use the guided chart for MDM model selection, to ensure that each factor is thoroughly examined when discussing the right model. The recommendation for future study included the factors' relationships and conflicts, and the cloud future effect of MDM models. Additional recommendations include possible extra factors (other than discussed in this research) that may influence MDM model choices.

Future research needed for the factors' conflicts. For example cost and quality factors may conflict each other, increase quality may require increase of the cost.
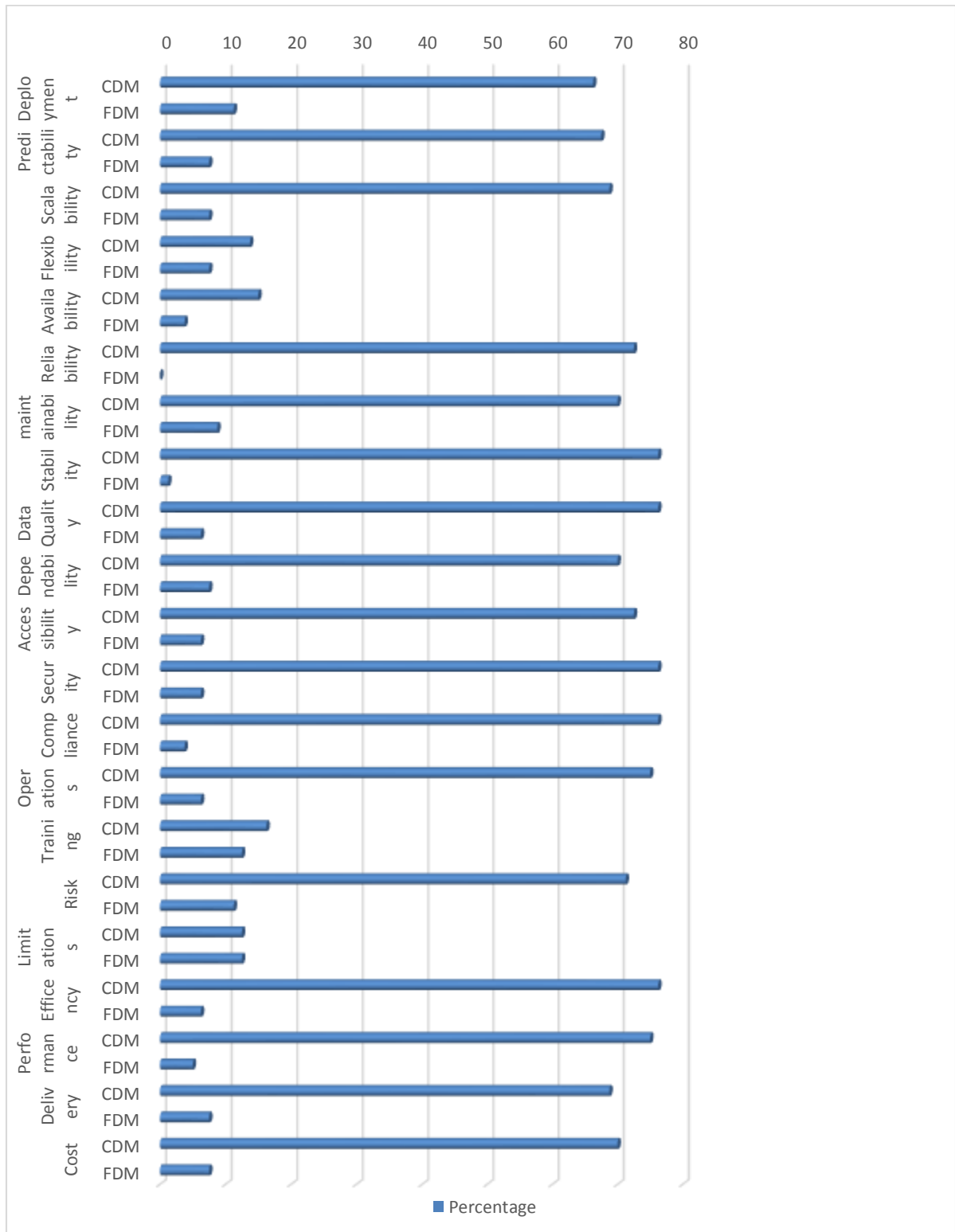

Figure 1. Survey participant's choices based on factors.

References

[1]   Creswell, J. W. (2004). Educational research planning, conducting, and evaluating quantitative and qualitative research (2nd ed.). Columbus, Ohio: Merrill Prentice Hall.

[2]   Karel, B. (2010). Introducing The MDM Market's Newest 800lb Gorilla: Informatica Acquires Siperian! . Retrieved from http://blogs.forrester.com/business_process/2010/01/introducing-the-mdm-markets-newest-800lb-gorilla-informatica-acquires-siperian.html.

[3]   Madhukar, N. (2009, June 24). Federated MDM data domains - A Perspective. Retrieved from http://www.infosysblogs.com/customer-relationship-management/2009/06/federated_mdm_data_domains_a_p_1.html.

[4]   Madhukar, N. (2009b). Implementing federated data domains. *Infosys.com*. Advance online publication. Retrieved from http://www.infosys.com/CRM/idea-center/Documents/federated-data-domains.pdf.

[5]   Moseley, M. (2009). Choosing the optimal multidomain MDM Architecture. *www.information-management.com*. Advance online publication. Retrieved from www.information-management.com/specialreports/2009_158/mulitdomian_mdm_master_data_management_centralized_-10015900-1.html.

[6]   Moseley, M. (2009). Eliminating Data Warehouse Pressures with Master Data Services and SOA. Business Intelligence Journal, 14(2), 33-43.

[7]   Moseley, M. (2009, October 15). Part 7: Centralized Models: Complete but Expensive. [Web log message]. Retrieved from http://blog.initiate.com/index.php/2009/10/15/part-7-centralized-models-complete-but-expensive/.

[8]   Tapscott, D. (2008). Business Intelligence: *Actionable Insights for Business Decision Makers*. Retrieved from http://www.businessobjects.com/campaigns/forms/q109/apj/everyone/tapscott/BI_for_Decision_Makers.pdf.

[9]   Wolter, R. (2007). Master Data Management (MDM) Hub Architecture. Retrieved from http://msdn.microsoft.com/en-us/library/bb410798.aspx.