FACTORS INFLUENCE DATA MANAGEMENT MODELS SELECTION

Dr. Gholam Ali Shaykhian, NASA

Ali 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

B.Sc of computer science from Khartoum university. Earned my masters from ottawa university in system science. My doctorate degree in information system from University of Phoenix(my dissertation was in MDM). I worked in IT industry over 20 years, 10 of them with Microsoft in different groups and for the last 4 years (while at Microsoft) was with business intelligence group. My focus was in Master Data Management. For the last 2 years I teach at universify of Najran in computer and information system college.

FACTORS INFLUENCE DATA MANAGEMENT MODELS SELECTION

Gholam Ali Shaykhian, Ph.D.

Mohd Abdelgadir Khairi, Ph.D.

Abstract

Data Management Models selection (Centralized Data Model or Federated Data Model) for managing organization data is influenced by many factors. This paper explains 21 factors that are useful to select an architectural model. Information technology leaders must understand the impact of factors such as cost, quality, and availability on the model selections. The study indicates that all factors except for Training, Reliability, Scalability, and Maintainability were found to be significantly contributed to the selection of the Data Management architectural model.

Introduction

Business Intelligence (BI) communities within Information Technology (IT) departments strive to select a suitable Data Management architectural model that help bridge the gap among their organizations, technologies, and their customers. Combined with tools for data quality management, this provides the trusted information foundation that companies base their analytics on" [14]. Wolter and Haselden [17] stated ". Data Management Models, centralized and federated architectural models play a major role in managing organization enterprise data. The centralized model allows an organization to organize and manage enterprise data into a central repository [8][17]. On the contrary, the federated model does not keep the enterprise data in one database instead allow users to query the data from multiple sources using appropriate data management tool [8][17]. The paper examines factors useful to consider when selecting centralized or federated architectural models. 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 [10], 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.

Mosley [10] encourages organizations to implement Data Management architecture to achieve better performance, agility, and cost-effective data warehouse architecture implementation. He added that almost all data warehouse systems tried to embed data management solution within its architecture of data warehouse [10]. Organizations are attempting to select the right architecture model (federated data management or centralize

data management) to implement data management solutions extended beyond the reliable version of the data to include business rules that ensure that the data validation, standardization, reconciliation, and alignment with organization's data governance [2].

Cost factor: According to Godinez, Hechler, Koenig, Lockwood, Oberhofer, and Schroeck [3], if the cost of data consolidation and replication is greater than the business benefits, the FDM model should be used. In addition, to reduce cost in the beginning FDM should be used as a short-term practice [5][6] to drive quick benefits to the organization before moving to longer-term strategy (CDM) [3].

Longman [4] considered FDM with respect to cost because operational data management (FDM) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success.

Delivery factor: Organization could use FDM model to submit their product data to global FDM system that allow quick data delivery to the connected manufactures and retailers [3]. The increase interest in grid and cloud computing made FDM suitable for managing these complex systems and provide scalable, dependable, and quick delivery solutions. FDM enable data access, dynamic migration of data for workload balancing, collaboration, and parallel processing in such complex systems [3]. However, FDM is not suited to handle retrieving and dealing with large amounts of data, hence CDM will be an alternative solution to deliver large data to applications [5][6].

CDM will benefit organizations for a longer-term strategy whereas FDM is useful for shorter-term strategy [3]. The disadvantage of CDM is that the MDM system is not up to date like FDM system and that may delay the prompt delivery to current master data to consumers [3]. Operational MDM (FDM) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success [4].

Performance Factor: FDM approach may not be a suitable to deal with large data because of the potential performance impact. However, FDM performance will improve if the MDM system queries small data from well-defined sources and was designed for short-term strategy. CDM is suitable MDM system for long-strategy and it performs better that FDM with respect to large data because CDM centralized the data in one repository and that make the master data retrieval more agile [5][6]. Sometimes the business requirements for performance may prevent from using CDM. One advantage of using FDM system over CDM is that FDM done quickly and that provide quick business result. Another disadvantage is the delay of master data propagation due to the CDM system is not up to date [3]. In addition, performance and data governance are the highest requirements for CDM implementation [3]. Loshin [5][6] believe CDM will properly

address the benefits of MDM expectation such as productivity improvement and cost reduction.

Godinez considered CDM with respect to performance because improved business capabilities are an important benefit of having a one source of a company's data (CDM). Organizations are beginning to recognize that how they manage their key data is critical to their ability to improve agility and performance Lack of trusted, quality information is widely seen as a strategic barrier inhibiting such capabilities as agility and competitive differentiation [15].

Efficiency factor: The MDM professionals who participated in this study believe efficiency is an important factor influencing their selection of an MDM models (CDM or FDM). Loshin [5][6] stated that CDM system could formulate a unified view of the data that enables the organization to reduce operating costs and tasks. The retrieval of large amount of data from a different source reduces the efficiency and performance of FDM, hence, CDM is an alternative approach to reach such efficiency [3].

CDM host a single copy of master data that interact with MDM system whenever read or write to theses master data records needed. MDM system becomes a source of authoritative truth of master data for an organization (Godinez et al., 2010). Operational MDM (FDM) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success [4].

The focus of operational MDM (using FDM approach) is to make sure that data in multiple operational systems that should be the same is actually the same. The goal is to synchronize operational systems data so that you have consistency at the front end, such as in customer-facing systems, which is particularly important for organizations with a lot of customer contact [4].

Business problems are solved by MDM solution because it is managing master data that critical to the company (Shankar, 2008). Shankar suggested that to have an efficient MDM solution based on focusing on the technology approach, organization should start simple, FDM model, and small with one entity, this way analytical usage will be easy, project risk reduces, data governance relieved, and the scope and ROI will be limited. However, the technology focus will not solve all business problems which lead companies to pursue CDM option [12].

Limitation factor: FDM Compensate for SQL limitations at the data source by processing parts of a distributed request at the MDM federated server (Godinez et. al., 2010). Confounding factors such as legal constraints limit MDM across geopolitical

boundaries [3][5][6] do not allow CDM. Another limitation to use CDM is master data locking in Line of Business (LOB) packaged application.

A technology-focused approach begins with FDM (registry) model, one entity, and analytical usage. After the first step the MDM solution will become more mature to add more entities, different models such as CDM, and integrate the solution with operational systems. Shankar [13] found that starting with FDM approach allows organization to start quickly. However, this approach will be limited unless organization considers current business and possible future needs, this will allow organizations to get the right technology and have a better road map in place. Starting with CDM may be useful for the future growth and fast ROI [13].

Risk factor: The use of CDM provides complete master data view and enabling full risk assessment. CDM system allow assessing and understanding information assets risks such as operation risk (example denial of service), regularity and compliance risk, and reputational risk (for example customer data compromise lead to untrusted company) [3][5][6]. All downstream applications depend on CDM system availability. The risk is high in case of such CDM system failure unless avoided by using redundant component in order to make MDM system highly available.

A complete and up-to-date master data in CDM system at risk of being accessed by attackers, that requires centralized security and centralized access [3][5][6]. Loshin [6] added that CDM systems allow better risk management and improve productivity. Loshin points out the risk of implementing FDM instead of CDM "At a low degree of granularity, the more data points that applications must touch, the greater the potential for duplication, inconsistencies, or missing information to skew calculated risk assessment, either at the customer level or across lines of business. More trustworthy and consistent financial information improves the business's ability to manage enterprise risk" [5].

Operational MDM (FDM approach) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success [4]. Organization must consider the risk that the operational MDM (FDM approach) requires significant customization and custom coding of operational systems in order to enable the synchronization.

Training Factor: Professional participants do not agree that training plays a role in MDM model selection. Training may not have direct influence in CDM or FDM models. However, knowledge gained by training adds needed skills to understand technology and tools and hence contributes on model's selection decisions. "It is important to realize that becoming a skilled practitioner of data modeling requires a significant amount of training and experience, so it is critical to engage individuals with the proper skill sets to successfully complete this task" [6]. "Training for enterprise roll-out of tools and technology make capabilities available on a more widespread basis". Loshin [5][6]

stressed on the assessment of the training requirement to transfer from federated model to centralized model before starting transition in order to create smooth transition.

Operation factor: FDM implemented with two operational patterns: data integration and aggregation runtime; and availability and disaster recovery [3]. Godinez [3] added that in case of FDM the source systems operational is critical even if the MDM system is not. In the case of CDM high availability of variation of operational multi-tier for critical data pattern is needed. MDM services requests executed in MDM system using CDM in a transactional fashion and "thus part of the operational fabric of an IT environment" [3]. Operational risk such as Denial of Service (DoS) could be mitigated by using CDM which has centralized place to enforce security [3]. CDM is a dedicated system that stores a golden version of master data. In operational uses using the CDM approach, the master data published to downstream systems while source systems (upstream) maintain sovereignty to manage and own data.

CDM Approach – consolidation approach – viewed in analytical use cases, clean and consistent master data published to data warehouse and BI applications [12]. In addition, [1] states that the CDM transcends operational and analytical use and into global governance on data management throughout the enterprise. Shankar and Menon [13] noted that having a compatible system between applications such as CRM and MDM system will lead to rapid operation. They cited an example of a large IT company which bought CRM and MDS systems from the same vendor and that approach promoted the compatibility and smooth operation. Unlike operational MDM, analytical MDM does not try to synchronize the master data in the operational systems. Instead, analytical MDM seeks to create a harmonized golden copy of data in a repository (CDM), mapping the data and correcting inconsistencies in the master data repository (CDM) rather than in the operational systems (FDM approach) [4].

Compliance factor: Professional participants in this study believe compliance is an important factor affecting MDM model selection. Godinez [3] believe CDM is the "ideal world" for implementing MDM system but legal constraints for geopolitical boundaries are critical factors that do not allow CDM because of the risk of violating regularity and compliance for information assets. For the sake of managing enterprise compliance and mitigate the risk of compliance violation the FDM is the suitable solution [5][6]. Loshin [6] gave an example of CDM system in banks that was successful on complies with an anti-money laundering (AML) which is required by US PATRIOT Act.

The system was designed to ensure policies and procedures are in place, in compliance with Bank Secrecy Act, and allow outside and independent party to test the compliance. Since CDM systems own the master data, the accountability and responsibility of managing such data includes communicating the data rules that impact compliance, and determining the business goals, data quality assurance, and information policies [5]. Shankar [12] prefers to have centralized system (CDM) that could evolve over time to solve other business problems and still provide compliance monitoring and reporting.

Security factor: FDM system could be a solution to the issue of not allowing the data to be copied out from its source due to security and licensing requirements [3]. However, if source systems allow copying the master data to MDM system, CDM repository will provide secure environment for this data to be accessed by authorized users and applications. Enforcing security for MDM system is critical for business success and manages risks [5][6]. Security in addition to performance, data governance, and high availability are most important aspects in implementing CDM because MDM in this model represents a single point of failure for applications accessing the master data. Improper security to a complete and up-to-date master data could be compromised by any attack [3]. Centralized security of master data is one of the advantages choosing CDM system [3]. Loshin [5][6] suggested include access control as part of any governance structure in order to oversee the security management and identifying the access roles.

Accessibility Factor: FDM system allows accessing data anywhere in the organization, without worrying about the format, vendor, new database creation, and disruptive changes to existing database using standard SQL or any tools supports Open DataBase Connectivity (ODBC) or Java DataBase Connectivity (JDBC) [3]. CDM as system of record provide centralized access, security, and governance. CDM can act as service consumer, service provider, or both because it gets data from different sources that could be used by a requester.

Dependability Factor: Godinez [3] stated that MDM system deployed in FDM depends on availability of the source systems. The availability of the source systems decreases when it is numbers grows in FDM deployment. In contrary, all applications depend on availability of MDM system deployed in CDM. The advantage of CDM over FDM is that there is no CDM dependency on the source systems like FDM. FDM is much preferable in the grid and cloud computing because the complexity of data management arises from the scale, vitality, undependability, and delivery of data sources [5][6]. One advantage of CDM implementation is that it is independent or not intrusive to the existing applications [3]. However, because CDM is a single point of enterprise-wide failure, all applications depend on its availability as oppose to FDM which depends on sources availability.

Data Quality Factor: FDM may not be suitable approach when retrieving large data from the source with data quality problems [3]. In addition, source system's data that retrieved by FDM may not reach the quality bar that necessary for MDM system which lead to collaboration difficulties between the sources and MDM system [3][5][6]. The advantage of MDM system that implemented using CDM approach has ability to clean, valid, and matches the data to maintain high quality of master data Loshin [5][6] asserts that MDM deployed in CDM is responsible for the data quality because enterprise applications depends on to master data retrieved from MDM system and decisions made out of these data in contrary to CDM, FDM system is not responsible in data quality because the data quality is the sources' responsibility; this condition lower the cost of quality assurance.

Shankar and Menon [12] noted that FDM approach maturity level of data quality is low in compare to CDM because FDM depends on source systems to provide quality data

whereas CDM processes the data quality validations such as cleaning and duplication reduction. Data quality in recent years has become important for integrating data from different sources within the organization and to create one version of its entities such as customer management, data warehouse, business intelligence, compliance initiatives.

The focus of operational MDM (using FDM approach) is to make sure that data in multiple operational systems that should be the same is actually the same. The goal is to synchronize operational systems data so that you have consistency at the front end, such as in customer-facing systems. This goal is particularly important for organizations with a lot of customer contact.

Master data are typically needed by multiple business units and maintained in heterogeneous systems in dispersed locations within a supply chain. Master data may not be maintained in a single central repository (CDM); therefore, the possibility exists for duplicated, inconsistent, and inaccurate master data.

Stability Factor: Loshin and Godinez [3][4][5] recommended on using FDM for short-term until the MDM is stable and gradually moves to CDM as a long-term strategy. FDM system is stable in a sense that data could be accessed from anywhere and anytime without creating new databases or without disruptive changes on existing ones, using standard SQL or any tools that support JDBC or ODBC.

Maintainability Factor: FDM approach does not require sources data consolidation, therefore the maintenance of master data in FDM is minimal and it is up to the data source to maintain it [3]. Moseley [9] argued that maintainability of master data in CDM is not possible because master data needs it in multiple business systems and it could be maintained better in these systems. However, CDM application provides features to the users to maintain master data without data expert support attendance.

Reliability Factor: Godinez [3] stated that companies have a fundamental problem in their ability to create reliable and accurate customer data for daily business processes. One major reason for data unreliability comes from incorrect and changeable customer reference data duplicated across the various IT systems within an enterprise. The unreliable sources impact the FDM approach negatively.

A reliable master data management solution is to obtain master data from different sources and consolidate data into one center (CDM approach): that accurate and consistent and could be used and consumed by all applications and users who needs it [12][13].

In order to reach a reliable 360º view of master data, the accuracy of system of record (CDM approach) for master data must be maintained to improve customer relationship and suitable customer [12][13]. The decisions are mainly done properly based on reliable master data. MDM solution either depend on reliable master data to create one view of data (FDM could be a solution) or consolidate the master data in centralized system

(CDM) for data quality processing before make the reliable one view of master data available for business intelligence uses [12][13].

Availability Factor: All systems enterprise-wide depends on availability of MDM (CDM Style) regarding master data because MDM in CDM style create a single point for master data to be consumed by all master data consumer in the enterprise applications (Godinez et al., 2010). Contrary to FDM, CDM does not depend on availability of sources. High availability, performance, security, and data governance are the highest requirements for MDM implemented in CDM style [3][5][6]. CDM may help data warehouse vendors who are working to overcome scalability, replication, and availability concerns as systems grow into petabytes of storage, but have not yet provided solutions with the flexibility and benefits of an MDM SOA approach [9].

Flexibility Factor: MDM (FDM style) is a flexible system because is allow flexible ways to ad hoc query master data from multiple sources. The correlation of the data from remote data sources and local tables in FDM is seamless as all data stored in FDM database. The flexibility of MDM (in FDM style) appears in ability of accessing data anywhere in the organization, regardless of data format or vendor, without new database creation, and affecting or disturbing existing databases, and using variety of access tools such as SQL, JDBC, and ODBC [3]. CDM may help data warehouse vendors who are working to overcome scalability, replication, and availability concerns as systems grow into petabytes of storage, but have not yet provided solutions with the flexibility and benefits of an MDM SOA approach [10][11].

Scalability Factor: With the interest on cloud computing growing, increases in grid applications in large scale of distribution and size, force organization to consider MDM in FDM style as a way to manage master data [3].

Dagon [1][16] sees the urgent needs for scalable MDM system in CDM Style to handle complex business process, keep track of continual changes, and increasingly M&A activities. Scalable MDM system is able to transform systems to SaaS(software as a Service) by cultivating it to deal with multitenant environment as well as making sure scalability and configuration capability.

The MDM system (in CDM style) also allows greater scalability and memory management by allowing users to open many expired versions as finalized and they are balanced across multiple read only engines. CDM may help data warehouse vendors who are working to overcome scalability, replication, and availability concerns as systems grow into petabytes of storage, but have not yet provided solutions with the flexibility and benefits of an MDM SOA approach [9][10].

Predictability Factor: MDM implemented in CDM style provides predictable performance and high availability for master data consumption [3][5][6]. If the number of

sources feeds FDM with master data increases, the performance and availability becomes more unpredictable and that may impact the MDM solution.

Deployment Factor: The deployment of MDM in FDM style carry a risk as the number of sources grows because the availability can decrease quickly [3]. All applications in the enterprise depend on the availability of the MDM solution deployed in CDM style. MDM solution deployed by FDM should provide business a metadata that provide the key linkages between critical data elements in the source systems [3][4][5]. Deploying an MDM system in FDM style can be done quickly because source systems still manage most of the master data. The MDM deployed in CDM environment allows the synchronization between an MDM system and other systems (master data consumers) unidirectional or bidirectional [3]. MDM is designed around the concept of a (virtual or physical) central repository (CDM) to store and manage master data and can be implemented according to various architectural styles [7].

MDM is increasingly concerned with the notion of "multiples" - multiple data domains, the multiple relationships among them and multiple usage styles. Even a tactical MDM project will require facets of an enterprise MDM solution set (CDM); for example, a financial service provider will need to master more than "customer" as it looks to master "product" in post-phase one stage.

Summary

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 paper focused specifically on gathering information on fundamental key factors in implementation of an Data Management; 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) are discussed in this paper.

The overall conclusion is that CDM is the best option with respect to factors such as cost and availability. However, FDM may be considered as short term solution. To reduce the cost at the beginning, FDM should be used as a short-term practice to drive quick benefits to the organization before moving to longer-term strategy – CDM.

References

[1] Dagan, B. (2008). Master data management systems to be useful for ccs, other functions. *Natural Gas, 24*(10), 25-29. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=31872515&site=ehost-live

[2] Dyché, J., & Levy, E. (2006). Serving many masters: a closer look at MDM; as important to business strategy as the EDW was, the processing of the item master was ultimately much more critical to business operations. *swtuopproxy.museglobal.com*. Advance online publication. Retrieved from swtuopproxy.museglobal.com/MuseSessionID=fd925b9615e67115f7e6173a6599d7e2/MuseHost=proquest.umi.com/MusePath/pqdweb?index=0&did=1454942261&SrchMode=2&sid=1&Fmt=3&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1258783964&clientId=13118

[3] Godinez, M. ; Hechler, E., Koenig, K., Lockwood, S., Oberhofer, M., Schroeck, M. (2010). The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight. Indianapolis, Indiana: IBM Press. Retrieved From http://pqtechbus.safaribooksonline.com/book/databases/business-intelligence/9780137054947/enterprise-information-architecture-operational-model/ch06lev1sec1#searchlist

[4] Longman, C. (2008). Why Master Data Management is Such a Challenge. DM Review, 18(11), 18-20

[5] Loshin, D. (2008). Master Data Management. Morgan Kaufmann, CA: San Francisco

[6] Loshin, D. (2008). Master data management. San Francisco, CA: Morgan Kaufmann.

[7] Lucas, A. (2010). TOWARDS CORPORATE DATA QUALITY MANAGEMENT. Portuguese Journal of Management Studies, 15 (2), 173-195

[8] 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

[9] 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

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

[11] 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/.

[12] Shankar, R. (2008). Master Data Management Strategies to Start Small and Grow Big.Business Intelligence Journal, 13 (3), p37-47, 11p

[13] Shankar, R., & Menon, R. (2010). MDM Maturity Pragmatism Business Challenges and the Future of MDM. BUSINESS INTELLIGENCE Journal. 15(3) p19-25

[14] 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

[15] White, A. and Y. Genovese. (2006). "Your Business Process Platform Needs and Enterprise Information Strategy," GartnerGroup Research, April 26, ID# G00139332.

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

[17] Wolter, R., & Haselden, K. (2008). The What, Why, and How of Master Data Management. *msdn.microsoft.com*. Advance online publication. Retrieved from msdn.microsoft.com/en-us/library/bb190163.aspx