Master Data Management: A Review

Saravjeet Singh[1*] and Jaiteg Singh[2]

[1,2]School of Computer Sciences, Chitkara University, India
www.ijcseonline.org

Abstract— Master data management helps companies to handle and maintain consistent and complete information of master data. Master data orchestration and knowledge extraction are the two biggest problem related to master data. Due to these two predicaments master data management is necessary for effective organization of master data. Master data is type of organizational data but not meta data and use for decision making and business intelligence. Master data in an enterprise is structured data and flows between different component of an enterprise under one's control. Common master data domains include person (customers, supplier, and controller), products (item), and location and other specific data. As an outcome of MDM, organizations have higher quality product, reliable data, consistent master data records and probable dedicated and interested consumer. This paper presents an overview of master data's content, scope, sample, advantages, master data management process, tool and techniques and challenges.

Keywords— Master Data, Master Data Management, People, Place, and Price

I. INTRODUCTION

For an enterprise all the operation like analysis and transactions are supported by Master Data. Master data is a type of operational data that have high value, defines core information that helps in critical decision making and handling business processes across the enterprise. Information about person (i.e. supplier, consumer, and controller, partners, employee), items(i.e. products and materials), price and more other specific information. Master data management is challenge due to advancement in technology and typicality of data generation. In the early days of computing, an organization depends on resources that housed all application and data in single module. Relational databases in the absence of data normalization led to significant data redundancy and data inconsistency.

Master Data Management can be of great help for customer relationship management, customer retention and to locate potential customer for a particular business domain. Every organization has their Master Data Management process and tool.

Master data management (MDM) is widely adopted data organization disciplines in modern eras. This is due to the consensus-driven definitions of commerce entities and the unswerving function of them across an enterprise. It act as significant triumph factors for key functional business activities. These activities includes business intelligence, operational quality, clients complete view, optimization of supply chain, mergers, authoritarian exposure, observance and acquisitions, and treating data as an enterprise asset. Master data management deals with core information of the organization.

Concept of master data comes in existence in early 20's and end of 90's. Desktop computing with relation data demands coherent and single copy data that can helps for customer data management and business intelligence. To provide such knowledge special type of data is require that act as master copy of all data's. Master data is a special data that is a part of transactional data but infrequent to changes. In 1995 Seibel puts his effort for operational and master data differentiation this leads to management of organization data that can be master data or operational data. Management of master data is not a different term but it is necessary for master data. As organization has system to handle master data but with increase in technology, organization size a separate module is require to handle and store master data. Complexity of master data and system growth gives the birth to master data management. In 2003 SAP start working on this concept. Many companies like oracle, Infosys, IBM, google, facebook, TCS, and so on start MDM in early 20's.

II. TYPES OF ORGANIZATION DATA

Anne Cleven and Felix Wortmann gives the brief introduction of master data and master data management. Master data Management (MDM) is a process and technology that describe, create, manage and maintain a reliable, sustainable, secure and consistent environment that provide single copy of consistent data. Organization data is categorized into metadata and domain data, and Domain data can be transactional data or master data, whereas informational metadata and operational metadata are type of metadata. The purpose of operational metadata is to enable design and technical operation of information systems [39]. Informational metadata on the other hand facilitates the understanding and access of domain data and is maintained for end users [39]. "Master data is core and critical entities of an enterprise that are repeatedly processed and use for future oriented and significant business decision [6]. Product, price, person, location, account detail are basic master data entities but these are vary from organization to organization. Daily invoice, transfer of funds, online purchasing, changes and alteration in account, sale/purchase information are type of transactional data sources and these type generated values represent transactional data. In comparison to master data transactional data is variant in nature and changes in its life

cycle. Life cycle of transactional data include generation/creation, alteration, status changes, promotion, translation, transfer, deletion and uses. Volume of transactional data is high as compare to other data and its size remains increasing for ongoing business activity. Helps to define and explain other type o data. It act as clue for data understanding. Reference data includes predefined and agreed -upon set of values. Abbreviation of state, century, currency type of detail is included in reference data. [9]. Transaction data is very different from master data but master data matches with reference data. Reference data can exist without any other data, its volume is low and changes to reference data occur even more rarely. Nevertheless there is a fundamental difference to master data. While master data refers to core business entities, reference data is more fine granular and not limited to domain data. Figure 1 summarizes the argumentation and relates the constituent components of the introduced data taxonomy to each other.
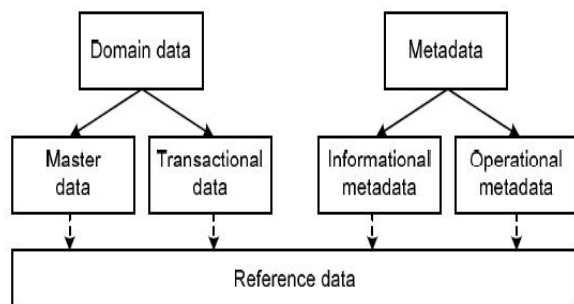

Fig. 1 Categorization of data

A. *Master data*

Master data is organization's core data that is essential BI operations in a specific organization. Information related to master data varies from one organization to another.

1) *Domains*

Each organization has their own set of master data requirements and characteristics. In its simplest form, master data is reference data about an organization's core business entities [13]. The common business master data entities are [9]:

➢ Party [5, 10, 11, 13]: This includes all business partner like persons, enterprise or organizations related to master data, for example customers, suppliers, producer, distributors, employees, citizens or other entities related to organization data. Typically party in master data includes name or other important detail of person or organization, e.g. having bank account is etc.

➢ Thing [9, 12, 13]: This domain includes data related to the products, services, items or assets a company owns and provide to globe. These entities can be product name, product id, price, product value etc.

➢ Location [9, 12, 13]: This domain contains all data which relates to area, places, sites, regions or address. A location can be a production house, factory, outlet address, sales territory, a city, an office, a production

facility or a shelf in a store. Location very really use individually, it often used with thing or party master data. Location master data combined with other data helps to answer questions like "Where is a sales territory?", "Where is an office?", "To which sales territory does a customer belong?", "address of particular office?" or "total sale particular outlet?"

➢ Other key entities [13]: Those sets of cohesive reference data that are of interest to the enterprise. Each enterprise will have its own unique set of these.

2) *MD characteristics*

The following characteristics are ascribed to master data:

a) *Behavior [4]*

Behavior describes how master data interact with other data. There exists noun/verb relationship between master data and transaction data. Transactional data includes the verbs, like daily sale and purchase, email etc whereas noun like person, place etc are master data element.

b) *Life Cycle [4]*

Life cycle of master data is called CRUD cycle that describes how master data is created, read, update and deleted. CRUD cycle is different for master-data element types and companies [16].

c) *Cardinality [4]*

Cardinality is the number of elements in a set. As cardinality decreases, the likelihood of an element being treated as a master-data element decreases.

d) *Lifetime [4]*

Life time of master data is high as compare to other data. Master data tends to be less volatile than transactional data. Master data does not frequently changes.

e) *Complexity [4]*

Simple entities rarely consider as master data because their management is not a challenge. Complexity of master data lies in its management.

f) *Value [4]*

Valuable the data element of the company is considered as master data element. Master data includes core business entities are have maximum value for an organization.

g) *Volatility [4, 6]*

Master data objects are least volatile in nature. During its life cycle master data values are un variant in nature and due to this quality it helps in critical decision making. In the same way completely stable data is of no use for decision making so volatility is must for critical decision but slould be in limited manner.

h) *Reuse [4]*

Master data is highly reusable due to its consistent and non volatile nature.

i)   Independent existence [5]

Master data does not require dependency on other entity for its existence. It can be independently created and destroy. For example an item can exist with its sale order but reverse is not true for transactional data.

j)   Constant data volume [6]

The number of master data objects keeps relatively constant compared to transactional data objects. Master data entity e.g. product remain constant whereas transactional data like sale data changes with time.

III.   MDM

Master data management (MDM) is one of the most widely adopted data management disciplines of recent years. Lifecycle of master data is called CRUD cycle that describes how master data is created, read, update and deleted. CRUD cycle is different for master data element types and companies. MDM is define as "a process that include technique, tool and authority for matching and managing the system of record and system of entry for the purpose of master data. MDM include creation of master data, understanding of data, filtering and processing of data and accessing critical sets of data in an enterprise. [13].

A.   MDM process

MDM is refer to as a process that provide single source of trust for organization master data. MDM provides solution for long term to an organization that does not changes frequently. MDM process has 5 basic step, that includes Profiling, consolidation, cleansing and govern, share and leverage. [8] MDM process has 5 basic step, that includes Profiling, consolidation, cleansing and govern, share and leverage. [8]

1)   Profile [2]

Profile is the first step for MDM process. Profiling means gather and create data for common use. This provides processing of master data at the client premises but make ready for its storage at a central place. E.g. data completeness, value occurrence distribution, acceptable rang of values etc. After implementation MDM solution profiling will remain continue and provide assurance for data quality by providing effective method of data processing at client end. However complete quality is not only dependent on profile but other elements listed below.

2)   Consolidate [2, 16]

Consolidation is a process of collecting or gathering quality data that comes from different client after the process of profiling. This is a step next to the profile in MDM process. This is a basic key step of MDM. Collection of huge record and combination of historical data at central is making possible by only consolidation. This is the 1st fundamental prerequisite for true master data fortification.

3)   Cleanse and Govern [2]

After consolidating the master data we need to remove duplicate records, inconsistency of data and unwanted data. This can be done through cleansing. Cleansing provides error detection and correction, removal of duplication,

standardization, toning and expansion of the data. Governing is done after cleansing the data and it involves handling data quality issues, data regulation issues, data definition, privacy, rule and dictatorial policies, and admittance controls. These two sets the quality and overall standard of MDM.

4)   Share [2]

If data is ready at its maximum quality at central pint now next step is to propagate data to different clients for the effective MDM. Sharing of Master data is important for business extension and to make effective use of master data applications. After cleansing and governing master data needs to be shared among the silos so that they get consistent master data.

5)   Leverage [2]

MDM provides a single version of the truth and 360 degree view of user data. This 360 degree view needs to maintain for complete life cycle and this is possible only if we are maintaining MDM completely. Leverage provides the facility to handle MDM after creating it and up to the end that is complete life cycle.

B.   Steps to implement MDM

R V Heikkinen and S Pekkola gives an approach to establish master data management function for an organization. Steps for establishing MDM functions are following [7]

- Identifying the need and objective
- Identifying the organization's core data and processes that use it
- Defining and identify roles and responsibilities
- Data standard definition
- Define Master data management metrics
- Planning of master data management architecture and components
- Communication and training planning
- Formation of check point and creation of MDM development road map
- Identification of application that uses master data and defining operation data for those applications.

C.   Architecture of MDM system [14,19]

J Kokemüller, A Weisbecker gives Architecture of commercial MDM system that is divided into 4 common categories that are centralized architecture,

Leading system architecture; Directory based and peer to peer based architecture (as shown in Figure 2). In centralized architecture all master data of one or more classes is stored in one system. This system provides then processes for data alteration and monitors data changes. As the integrated system may require that the MDM system provides it with objects using the IDs of the integrated system, the MDM system usually implements a mapping table mapping MDM IDs to foreign IDs. Additionally, the MDM system may keep track of the object's version deployed to the integrated systems. Whether a system keeps track of several versions or just one is vendor dependent, yet always exactly one dataset

of an object is marked as the best version. This version is known under several names, the most common are:

30 "Single Version of Truth" and "Gold Copy". All new information is integrated into this version and it is ensured, that this version is always the best in terms of data quality. Starting from it, the views for the integrated systems on the information are calculated and deployed.
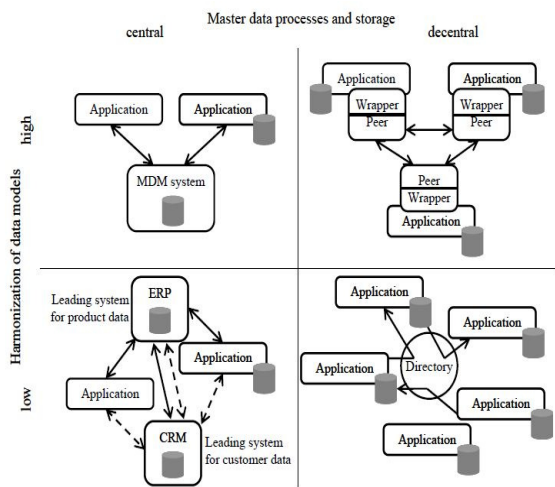

Fig. 2 MDM System Architecture [19]

The second variant is a leading system. Here, the data is not integrated in a separate database of an MDM system, but in one of the integrated systems. Usually, the system with the highest expressiveness for the particular data class is chosen. The transformations of the MDM Workflow are then carried out either by the leading system itself or a MDM system wrapping the leading system.

Directory based approach is a lightweight way to integrate master data is by using a directory. In the directory references to certain master data objects are stored. The objects themselves remain distributed over several systems. This approach is the only one that offers a purely virtual integration. It cannot enforce data quality algorithms and cannot generate a gold copy.

The 4th architecture for master data management is a Peer-to-Peer (P2P) based approach. In this master data system are wrapped by peers. They work in a networked structure where all participants are equal with respect to what they are able to do. This Peer-to-Peer collaboration reflects the organizational structure of autonomous enterprises that directly and equitable share information and are responsible for the integration to their neighbors. P2P integration provides functional components for semantic information integration and acts as a framework for efficient implementation of complex information integration scenarios. Peers provide flexibility towards the integrated systems and allow to this side heterogeneous behavior.

D. 3.4 Component of MDM system [15]

C White of SAP introduced and explains MDM system and its components. In an enterprise MDM system, all master data is maintained and processed by core business users and other IT systems and organizations uses MDM applications. This application keeps the information about master data, its meta data and its management activities according to business prospective. MDM application only related to master data these are separate from other application and not deal with transaction data. Master data is collected from historical and transaction data and after processing it is distributed over transactional applications. Component of typical MDM system includes MDM applications, MDS, MMS and MDI services as shown in figure 3.
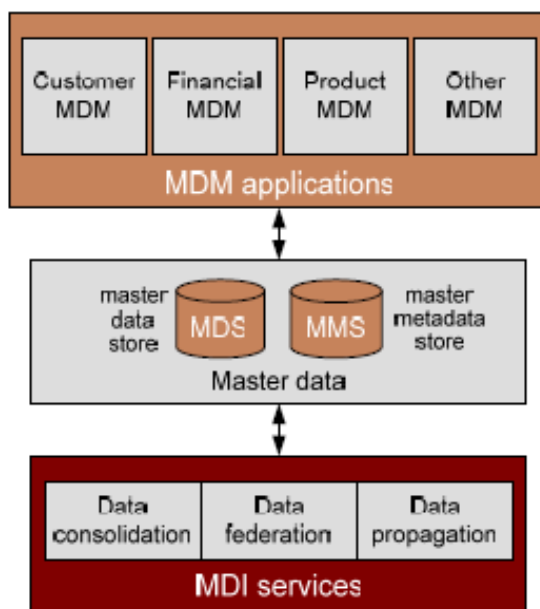

Fig. 3 Components of MDM

• MDM applications use for creating, processing and publishing master data and metadata.

• A master data store (MDS) holds master data in a consolidated system.

• A master metadata store (MMS) containing the master data business model, and master data rules and definitions. This model documents master data entities, attributes, relationships and their business meaning.

• A set of master data integration (MDI) services to combine, amalgamate and circulate master data. Business users employ commissioned MDM applications to admittance and sustain master data in the MDS. The MDS represents the SOR for enterprise wide master data.

MDM Solutions			
Attribute	SAP Netweaver	IBM InfoSphere	Oracle MDM Suite
Latest Version	7.4 sp2	9.1.2	8.2 customer hub(11.2 of Integration

			architecture)
Introduced time	2004	2008	2004
Components	Console (including MDS, MDIS, and MDSS), Data Manager, Import Manager, syndicator Publisher + MDLS, Image Manager, CLIX Commands, Extractor for Reference Data	Data authoring UI, Import-export Workflow engine, Data Stewardship UI , Enterprise Viewer, MDM Inspector, MDM Product Maintenance UI, Reference Data management Hub, Unstructured Text Correlation	Fusion Middleware, Analytical MDM apps, Operational MDM Apps, Enterprise data quality, Customer Hub Site Hub, Product Hub
Domains	Domain based or Domain neutral : Product, Customer, Employee, Supplier, Material, Business	Domain based or Domain neutral: Party, Product, Account	Domain Based: Product, Customer, Account, Site
Method of use	Operational and analytical collaboration method is used when used with SAP BPM	It uses Operational and analytical collaboration	Operational and analytical collaboration method is used with other application like Business Process Management
Interpretability, Consistency, Ease Of Understanding	Dropdown Fields quantifier	Domain model mapping on import and export	Word replacement mapping on import and export

Table 1 MDM existing Solution Comparison

MDI services provide the way to integrate and consolidate through MDM applications.

Key requirements include:

• Data quality management

• Metadata integration and propagation

• Synchronous and asynchronous data propagation with guaranteed delivery

• Change data capture and data transformation

• Data federation

E. Existing MDM Solution

MDM solution is already provided by many organization like Microsoft, oracle, SAP, IBM etc, and they have their own prospective and architecture. In this section we compare 3 solutions that are SAP Net weaver, IBM infospere and Oracle MDM hub. There are several aspect of comparison that can be considered for analyzing these solution but we considered few of them that are covered in table 1.

F. Benefits of MDM

MDM provides the tools required to rationalize the data needed to improve business processes for industries [16].MDM is beneficial for an organization because it provides following benefits:

1) Data redundancy [18]
Without MDM, each system, application, and department within an organization has their own version of data. This leads to redundant master data and poor data quality [15].

2) Improved customer service
Master data objects provides improvements in meeting customer expectations like availability, accuracy, and responsiveness and helps to manage transparent customer data [17].

3) Data consistency [15, 18]
Reliance on the data generated by master data management process provide consistency to overall decision [17].

4) Improved operational efficiency and reduced costs [17]
Replication of the same data often is linked to replication of activities associated with managing those data sets. Formulating a unified view of the data enables the organization to reduce operating costs and tasks.

5) Supporting Business change [15]
Organizations are constantly changing as new products and services are introduced and withdrawn, companies are acquired and sold, andnew technologies appear and reach maturity. These disruptive events cause a constant stream of changes to master data, and without a way of managing

these changes, the issues of data redundancy, data inconsistency, and business inefficiency are exacerbated.

6) Single Customer View principle [18]

MDM provide single view of customer data by consolidating data to single place and providing consistent data to different applications.

IV.   CHALLENGES

Without effective tools and solutions for master data management, different challenges can arise [18]:

- Multiple heterogeneous data sources and different data structures
- Redundancy in master data and system
- Difference of quality level
- Incorrect matching
- Lack of 360 degree of customer view
- "all at once" approach to deployment.

V.   CONCLUSION

In this article, an overview of Master Data's content, domain, management process, characteristics, architecture of MDM, solutions of MDM, advantages and challenges have been reviewed. As MDM solution and architecture reviewed in this paper is related to relational data. Besides the improvement of these relational MDM solutions, object oriented or semantic solution will be discussed in future.

Although this paper clearly has not explained all era of Master Data Management and issues of Master Data Management, hopefully it has provided some useful information for researcher.

REFERENCE

[1] A Cleven and F Wortmann, "Uncovering four strategies to Approach Master Data Management", IEEE 43th HICSS, Jan 2010, pp 1-10.
[2] Master Data Management An Oracle White Paper September 2011
[3] J Kopcke, Master Data Management: Old Problem with a New Urgency
[4] Master Data Management outside of SAP –Standards based MDM for PLMT orbjörn Holm, Eurostep.
[5] A. Dreibelbis, E. Hechler, I. Milman, M. Oberhofer, P.van Run, and D. Wolfson, Enterprise Master Data Management: An SOA Approach to Managing Core Information, IBM Press, 2008.
[6] D. Loshin, Master Data Management, Morgan
    [1]   Kaufmann, 2008.
[7] R Heikkinen and S Pekkola, "Establishing an Organization's Master Data Management Function: A Stepwise Approach", IEEE 46th HICSS, Jan 2013, pp 4719-4728
[8] Rittman, "Introduction to Master Data Management" Mark
[9] Master Data Management (MDM Mastering the Information Ocean) Capgemini white paper.
[10] C. Beasty: "The Master Piece", CRM Magazine, Vol. 12,2008, pp. 39-42.
[11] A. Berson, and L. Dubov, Master Data Management and Customer Data Integration for a Global Enterprise, Mcgraw-Hill, 2007.
[12] A. Zornes: "The Fourth Generation of MDM", DM Review, Vol. 17, 2007, pp. 26-37.
[13] C Imhoff and C White, "Master Data Management: Creating a Single View of the Business", Copyright 2006, Powell Media, LLC, Intelligent Solutions and BI Research All rights reserved.
[14] J kokemüller, A weisbecker, "master data management: products and research"
[15] C White, "Using Master Data in Business Intelligence", BI Research March 2007 Sponsored by SAP
[16] "Building the Business Case for Master Data Management", An Oracle Thought Leadership White Paper April 2009.
[17] "Master Data Management" Book by David Loshin
[18] http://www.uniserv.com/en/data   management/master-data-management/
[19] J Kokemüller, A Weisbecker, " Master Data Management: Products and Research ".
[20] Walter, P., Werth, D. Loos, P. "Peer-to-Peer-Based Model-Management for Cross-Organizational Business Processes", Los Alamitos, CA, USA, June, 2006, pp. 255-260
[21] http://pic.dhe.ibm.com/infocenter/mdm/v11r0/index.jsp?topic=%2Fcom.ibm.mdmhs.overview.doc%2Farchitecture.html
[22] http://help.sap.com/saphelp_nwmdm71/helpdata/en/43/D7AED5058201B4E10000000A11466F/frameset.htm

AUTHORS PROFILE

Mr.Saravjeet Singh currently working in Chitkara University, Rajpure, Punjab.
He belongs to Ambala (Haryana). He done his B-tech from Kurukshetra University Kurukshetra With 78%. His masters in computer science from Chitkara University. His research areas are database, big data and semantic data.He works in Chitkara University as Assistant Professor and has been publish reseach paper, attended various seminar and workshops.

Dr Jaiteg singh currently working in chitkara university. He done tis Phd fron punjabi university and hs been published many research paper and books. His research areas are database, big data, data warehouse, data mining and testing. He works in Chitkara University as Associate Professor.