

Whitepaper

Data Mart Consolidation Process - What, Why, When and How?

Author : Pandian C Muneeswara

Data Mart Consolidation Process - What, Why, When and How?

Table of Contents

1. Introduction	03
2. What is Data Mart Consolidation?	03
3. Why Data Mart Consolidation?	04
4. When should you do Data Mart Consolidation?	04
5. How should you do consolidation?	05
6. Data Mart Consolidation – Data Model	05
7. Data Mart Consolidation – Data Integration	05
8. Data Mart Consolidation – Information Delivery	05

Author's Bio

Pandian C Muneeswara

Pandian C Muneeswara is the Principal Consultant from Hexaware Technologies, with more than 13 years of experience in leveraging metadata of BI systems. You can reach him at muneeswarap@hexaware.com

Data Mart Consolidation Process - What, Why, When and How?

1. Introduction

Data Mart consolidation at many instances is an immediate tool for saving cost in software and hardware infrastructure, but when executed effectively consolidation projects can also help in eliminating redundant and dormant process ensuring a new optimized system. There are still many questions in terms of What is consolidation? Why should I do it? When should I think of consolidation? and How do I go about doing consolidation? this article answers these questions. Also details in terms what needs to be looked at the metadata level for the consolidation process.

2. What is Data Mart Consolidation?

The process of merging the different components involved in a data mart into another data mart in order to rationalize the functionality, the hardware infrastructure and the software.

3. Why Data Mart Consolidation?

Consolidation in many organizations can be an ongoing process, some of the reasons for consolidation projects to be initiated are

- Organizations have independent departmental marts and having many common processes and data getting redundantly replicated among them
- Organization has recently acquired a company which has functionally overlapping analytical systems
- The need to build an Enterprise Data Warehouse to ensure single version of data throughout the organization
- As part of BI Platform standardization initiative
- The need to define a shared service model for ease of administration of the Data Warehouse components, this again needs consolidation activity to ensure common standards are followed across the environment for optimal servicing

The key benefits of Data mart Consolidation are

- Optimal infrastructure usage by consolidating servers to maximize system performance
- Save disk space by eliminating unwanted data copies
- Easier system maintenance with integrated impact analysis
- Integrated system security and user profiles
- Common metadata definitions across systems
- Common process frameworks related to data handling like error data management, process control and auditing etc
- Availability of reusable components
- Effective software license usage by eliminating unwanted data base, data integration and reporting product instances
- Standardization of technology platform and eliminating multiple skill needs
- Reducing the Data Integration window time and making data available at the earliest to the users
- Enabling integrated cross-functional data analysis

4. When should you do Data Mart Consolidation?

Following are the indicators which characterizes the need for data mart consolidation

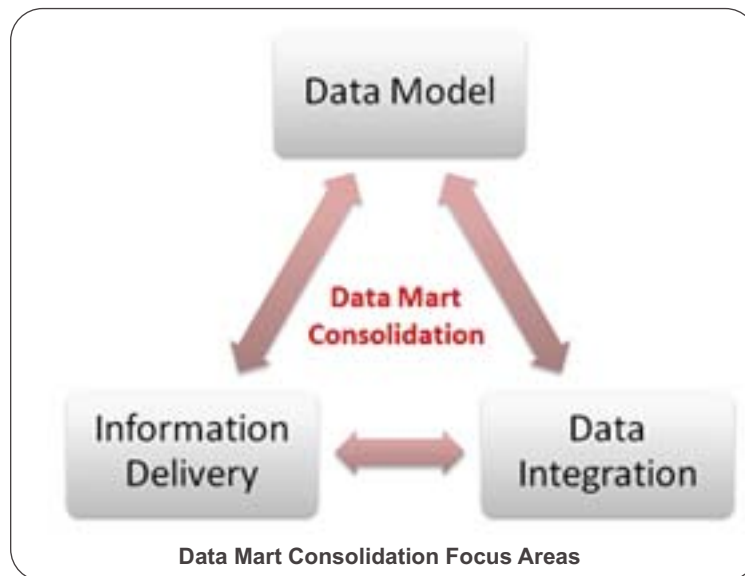
- Users extract data from two data marts and integrate them externally in tools like Excel
- Users have to go through different security mechanism to access the data
- Users are able to generate almost similar report from two different data marts
- Users find discrepancy in data that is shared between two data marts
- Same data is not being made available to all users at the same time
- Existence of redundant data copies and data integration process between data marts
- Existence of difference in data element definitions in terms of physical and metadata across data marts
- Existence of multiple independent servers for data integration and reporting
- Existence of multiple data integration and reporting products

Data Mart Consolidation Process - What, Why, When and How?

5. How should you do consolidation?

In the process of Data Mart Consolidation, the layers that need to be assessed and consolidated are the

1. Data Model
2. Data Integration Process and
3. Reporting Layer



Let us define the following

- Parent Data Mart (PDM) – The Data Mart which is to get extended with the functionality from another (Child) Data Mart
- Child Data Mart (CDM) – The Data Mart which is to be merged with the Parent Data Mart

All the key layers can be assessed for system understanding and as well to determine the overlap and uniqueness between PDM and CDM systems through the metadata present in the systems. The following sections detail what metadata elements need to be analyzed in the three layers.

6. Data Mart Consolidation – Data Model

The key analysis that needs to be done in the process of Data Model consolidation are

- Collect the data elements from the Child Data Mart (CDM) data model
- Determine the source elements for each of the elements in the CDM. The Data Integration metadata has to be analyzed to determine the source elements
- Collect the tables and data elements from the PDM based on following points
 - Source systems that feed the CDM also feeds the Parent Data Mart (PDM)
 - PDM receives data elements from other source system that performs the same function as the identified source system of the CDM
- Determine the data elements from the PDM that has a overlap with the elements in CDM
- Determine any differences in the data type, format and data length for the similar elements in both the marts
- Determine the tables and elements in the CDM that are not used in the reporting layer. The identified tables could be intermediate staging or work tables and as well tables that can be possibly retired
- Determine the tables and elements in the CDM that do not have a data integration process defined. The identified tables could be candidates that are being currently maintained through a manual process and this will require an automation process to be designed

Once the outputs of the above steps are available then the process of defining and integrating data model elements between the two systems can be started.

Data Mart Consolidation Process - What, Why, When and How?

7. Data Mart Consolidation – Data Integration

The second area to focus is Data Integration (DI) Process. Much of the information gathered from the consolidation of Data Model exercise will be leveraged for the consolidation of DI process. The key aspects that needs to be gathered in the process of consolidating DI processes are

- Collect the source, target and their relationship details from the Data Integration processes from CDM and PDM
- Determine the overlapping DI processes between PDM and CDM. These would be scenarios of
 - Tables that are sources for both the Data Marts
 - DI processes that read data from PDM and loads into CDM
 - Same data transformation logic being applied on the incoming data or over the data already present in the Data Mart

This step helps determine the DI processes that can be eliminated or merged

- Determine how many of the DI process in the CDM have source tables that are not as sources for the PDM.
 - This step helps determine how many of the existing DI process from CDM needs to be added to the PDM DI process
- Determine the tables and elements in the CDM that are not used in the reporting layer. These tables can be intermediate staging or work tables as well dormant tables.
 - This step helps determine DI processes that can be possibly eliminated
- Determine how many of the DI processes in CDM are in a different technology like hand coded DI processes in SQL procedures, scripts or any other tools than that of PDM DI process
 - This step helps determine how many of the existing DI process in CDM needs to be recoded or converted to the platform of the PDM DI process
- Determine any common DI logic being used between the CDM and PDM processes or within the CDM DI.
 - This step helps in leveraging existing reusable component in PDM or help build new reusable component that can be used as part of the consolidation
- Determine how many of the tables in CDM are currently manually maintained, like entry of a file arrival status or date of arrival to trigger other DI processes
 - This step helps determine scenarios of automation and eliminate any manual interaction for DI.
- Determine the key performance intensive long running DI processes in CDM
 - This step helps determine DI processes that needs to tuned as part of the consolidation

Determine the schedule dependency of CDM DI process and PDM DI process. Collect the current PDM DI server utilization details.

- This step helps in preparing an integrated schedule-dependency timeline of the PDM and CDM DI process within the available DI Server window. This is a very critical task to ensure that as part of the consolidation the data being made available in PDM maintains the same state of availability as in CDM

8. Data Mart Consolidation – Information Delivery

The third area to focus is the Information Delivery process, there are three parts to it they are

1. Semantic Layer Integration
2. Security and User profile integration
3. Report consolidation

The key aspects that needs to be performed in the consolidation exercise across these areas are

Semantic Layer Integration

- Collect the metadata of the objects present in the semantic layer of both PDM and CDM
- Determine the common objects in the semantic layers
- Determine the unused objects in the semantic layers, also confirm with users on the objects that are not required in CDM and eliminate them
- For the required objects in CDM determine the placeholders in the PDM Semantic layer and confirm with users on the location of the new object in the PDM semantic hierarchy
- Merge the CDM semantic with the PDM Semantic

Data Mart Consolidation Process - What, Why, When and How?

Security and User Profile Integration

- Determine the security definition for the required objects in CDM
- Map the security setup in CDM and PDM, determine the gaps and work around
- Determine the common users and unique users between the system
- Create users who are not present in the PDM system
- Apply the required security definition over the new objects in PDM

Report Consolidation

- Collect the metadata of the objects present in the reports of CDM
- Determine the reports in PDM that carry similar objects to that present in CDM, these are reports in CDM that may not be required to be migrated from CDM to PDM
- Determine the set of unique reports that needs to be moved into the PDM system
- Re-point or redevelop and change the existing reports in CDM to the integrated semantic definition in PDM

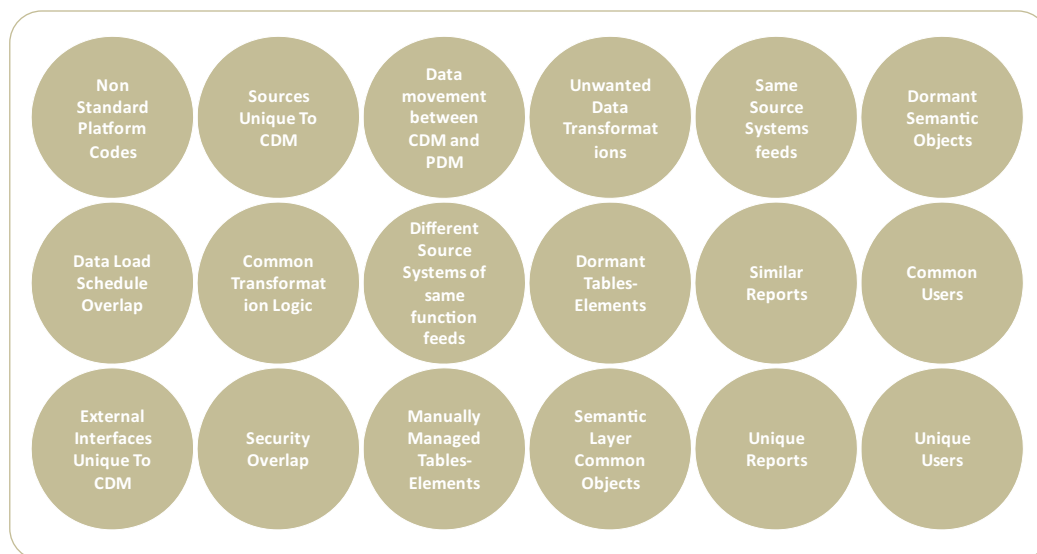


Figure 2: Data Mart Consolidation Key Metadata Analysis Areas

Focusing on the three different layers as part of the consolidation project should ensure that the critical challenges are addressed. Leveraging the metadata of the application to understand the different layers in terms of what objects are present, what overlaps, what is unique etc will be a key to success for a quicker and as well a accurate consolidation process.

Data Mart Consolidation Process - What, Why, When and How?

Address

1095 Cranbury South River Road, Suite 10, Jamesburg, NJ 08831. Main: 609-409-6950 | Fax: 609-409-6910

Safe Harbor

Certain statements on this whitepaper concerning our future growth prospects are forward-looking statements, which involve a number of risks, and uncertainties that could cause actual results to differ materially from those in such forward-looking statements. The risks and uncertainties relating to these statements include, but are not limited to, risks and uncertainties regarding fluctuations in earnings, our ability to manage growth, intense competition in IT services including those factors which may affect our cost advantage, wage increases in India, our ability to attract and retain highly skilled professionals, time and cost overruns on fixed-price, fixed-time frame contracts, client concentration, restrictions on immigration, our ability to manage our international operations, reduced demand for technology in our key focus areas, disruptions in telecommunication networks, our ability to successfully complete and integrate potential acquisitions, liability for damages on our service contracts, the success of the companies in which Hexaware has made strategic investments, withdrawal of governmental fiscal incentives, political instability, legal restrictions on raising capital or acquiring companies outside India, and unauthorized use of our intellectual property and general economic conditions affecting our industry.