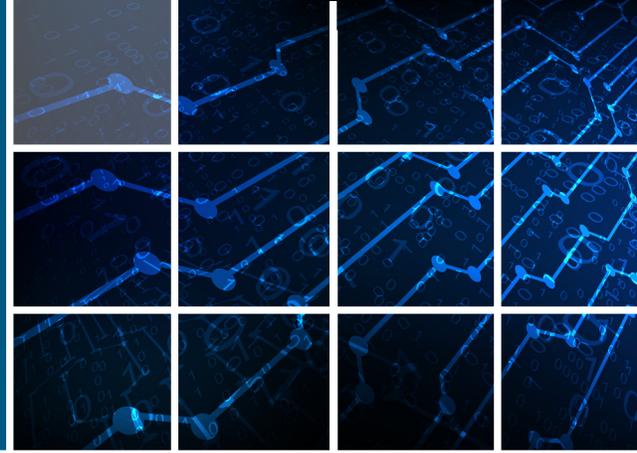


MRE Insights

Using Data Marts to Resolve Data Quality Issues



Data warehouses can be found just about everywhere. Most businesses have some sort of data repository which enables them to produce operational reporting. Most of these warehouses are **multi-generational** by nature, meaning they have been developed and populated by multiple generations of business owners, developers, business analysts, data analysts, architects, etc. Many of these repositories do not have or have lost their data governance. So, what happens when this data is incorrect or inaccurate? Do you throw away your multi-year data warehouse investment? Of course not.

Depending on how the data warehouse was structured and populated, it is possible to mitigate data inaccuracies and improve the overall data quality. This is not always an option. For example, if there is just too much bad data, there is little or no documentation on how the system was designed, or the folks who built it are no longer around. However, there are ways to attempt to solve this problem without having to redesign the entire data warehouse. The solution is to introduce a combination of data marts and business rules.

Data marts are operational reporting friendly databases (sometimes called cubes) and are generally used to store information which is then used to produce operational reporting. See Figure 1. However, they have other interesting applications and uses which ultimately help you to improve the data quality of your warehouse. A **business rule** layer is a logical layer which can be used to stage data prior to the data mart ingestion and perform functions which would lead to improved data quality.

High Level Data Warehouse Configuration

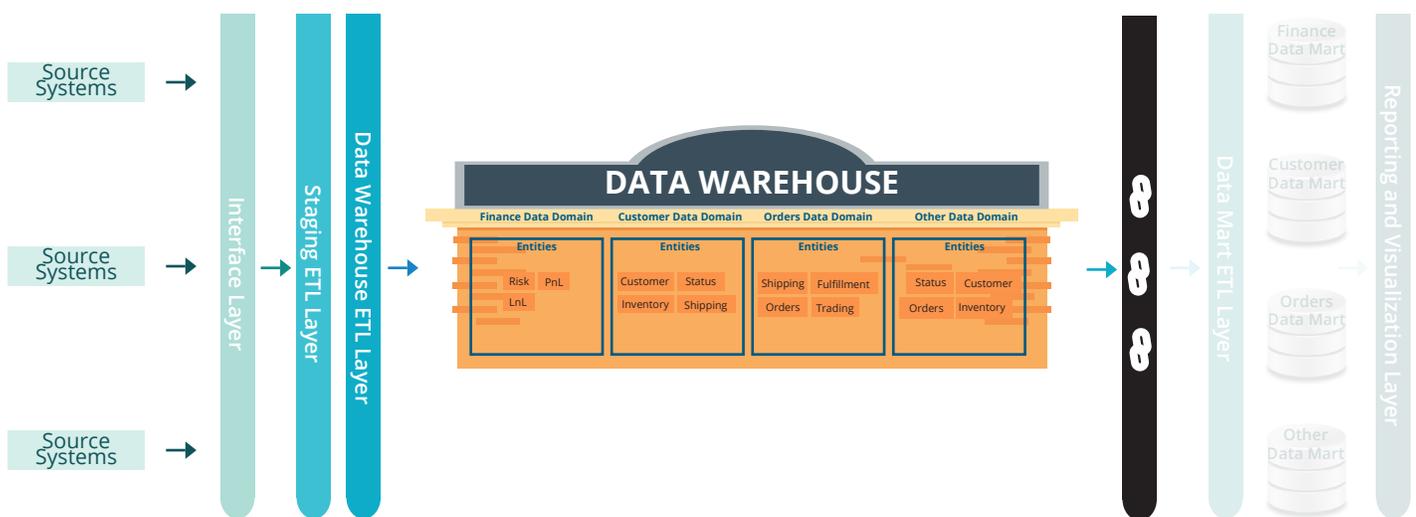


Figure 1. High level data warehouse configuration

The interesting use for data marts comes in when they are properly combined with a business rule layer and the results can potentially provide much-improved data quality and accuracy.

MRE Insights

Using Data Marts to Resolve Data Quality Issues

High Level Data Warehouse Configuration with Business Rules

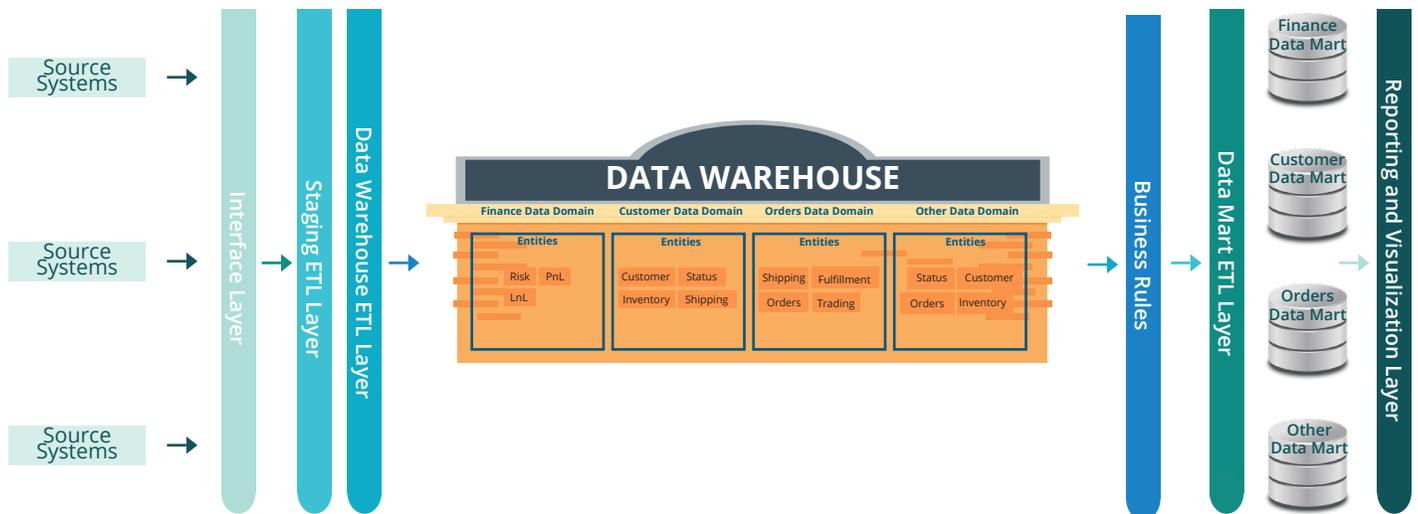


Figure 2. Proposed data warehouse and data mart configuration with business

By introducing a layer of business rules processing between the data warehouse and the data mart layer, we have found that it can help dramatically improve data quality. See Figure 2. A business rules layer has the capability to introduce the necessary business logic which can perform the task of discriminating between correct and incorrect data and filter such data accordingly. This strategy allows only valid and approved data to pass from the data warehouse to the data mart. Also, due to this strategy, the correct and accurate data is then made available by the reporting and visualization tools for operational reporting purposes.

The business rule layer has many capabilities to help improve the quality of data. Some of these capabilities include:

- ▶ Data filtering
- ▶ Date range validation
- ▶ Entity validation
- ▶ Currency validation
- ▶ Transaction validation
- ▶ Data reconciliation between data marts and data warehouse

To build this type of strategy, we recommend starting by gathering the operational report requirements first and working backwards towards the design of the data marts. Once you know what your end users want to report on you can then design the data models for your data marts. Sometimes this process may seem counterproductive as you may not have all the elements available in the warehouse, however, this is not necessarily a negative. This process will help you discover missing data elements which can be added to the warehouse to meet the operational reporting requirements. Even if you can provide between 70% to 80% of the business user requirements and keep working iteratively on the next 20% to 30% this will be a great start.

The following are some helpful reminders along your data quality journey. While you go through this process, ensure you have access to personnel whom are intimately familiar with the data. It will be critical to partner with internal business users to assist you along the way to ensure the data is accurate and trustworthy. Remember, it is very important to make a great first impression when it comes to presenting accurate data to the business community. It is also important to implement data governance strategies and practices to maintain the reputation, quality, accuracy, and overall integrity of your data.

MRE Insights

Using Data Marts to Resolve Data Quality Issues

Your business data is an incredibly valuable asset, but only if you can rely on it to make strategic and tactical decisions. It is worth investing time in performing exploratory data mining sessions to identify areas where your data may not be optimal or where you may have gaps. Once these areas are identified, you can then apply practices such as the business rule layer to assist in correcting any deficiencies. Other options may include applying a transformation rule layer prior to the warehouse ingestion to avoid ingesting incorrect data. This approach works as well, however, it requires careful design to ensure valuable data is not excluded from your warehouse. Additional strategies include data governance, data reconciliation to constantly validate the quality of your data, and the creation and maintenance of your data catalog. Also, implementing machine learning algorithms could help you identify invalid or incorrect data before you do.

Next time you are thinking about restarting your data warehouse project from scratch, give these ideas a try. It might just help you avoid the costly investment of redesigning your entire data warehouse.



Emilio Chemali, Director of Business Intelligence & Analytics, MRE Consulting

Emilio is a technology subject matter expert, respected thought leader and CIO100 Award Winner. With over 18 years of experience, Emilio has helped clients in multiple industries create business value through Business Intelligence, Data Analytics, DataOps, DevOps, IoT, Application Integration, Enterprise Mobility, Enterprise Architecture, Software Development, Infrastructure Management, Cloud Strategies, Server Virtualization, and Application Performance Tuning initiatives.

If you have any questions or concerns regarding your business data [contact us!](#)