After the Fact Data Integration Methods are now Obsolete

Currently, all data integration is attempted after the source databases are designed, instantiated and populated. These data integration methods are referred to as "after the fact" data integration. The distinction is important since the data integration by design methods are implemented at the time the database is designed. The fact that the Data Reintegration Methodology integrates data in the design phase and as the data is populated into the database eliminates the need for any "after the fact" data integration.

Currently, most data warehouse projects are used to integrate multiple data sets from existing transactional database systems. While the data warehouse was originally designed to be a reporting system, it has evolved today to become an "after the fact" data integration methodology as well.

The Master Data Management (MDM) approach is to resolve differences in the master data of existing data sets. Master data is extracted from the source data systems and transformed into a common data structure of the MDM database. The master data records are consolidated and duplicate master data records are removed. The consolidated enhanced master data records are then transformed back into the original source data system.

Data federation and data virtualization use sophisticated middleware to interact with existing databases allowing for report generation from these "isolated" data sets. Unintentional data commonality that exists in these databases is used to join the data return sets retrieved from these databases. Data federation vendors freely admit that their tools should be limited in their use as the return sets from these" isolated" data systems may not have the commonality needed to support more complex queries.

In light of the data integration by design paradigm, what is the impact upon these "after the fact" data integration methods? The data warehouse is no longer required to provide data integration functionality. Current data warehouse integrations are very inefficient when compared to data integration by design. Current data warehouses also become isolated information silos that are not integrated with other data warehouses. However, data warehouses may still be utilized to retain historically accurate data and to support sophisticated and complex reporting. With the Data Reintegration Methodology, any data warehouses will be integrated with all other data warehouses. The data warehouse will now be populated by simple data replication and data aggregation.

Within the Data Reintegration Methodology, Master Data Management will no longer be used as a data integration method. The Data Reintegration methodology supports multiple layers of data governance. Master Data Management will be used to manage master data that is only of significance to the organization. Other master data, used to support data integration across organizational borders, will be governed by the data providers. For each element of an organization's governed master data, the metadata will be the same in each database. Therefore, the master data management will not require transformations. The master data will be managed with simple master data replication.

Data federation will be greatly aided by the Data Reintegration Methodology. Each database enhanced by the Data Reintegration Methodology will support multiple data access paths with other so enhanced databases. The lack of metadata commonality and of data commonality that plagued data federation in prior art databases has been eliminated. So data federation will no longer be a data integration tool but will become a much improved reporting tool.

Get you free copy of the Data Reintegration Methodology whitepaper for more details.

RETURN to the Data Reintegration Methodology page!