Box 23: Data Integration Techniques

Box 23b: Data Integration Challenges

Precis of – Data Integration Techniques and Their Challenges Blog by DATA INTEGRATION NINJAS

Data Integration

The combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. […] There are several organizational levels on which the Data Integration can be performed:

1. Manual Data Integration. Technically speaking, this is really not a Data Integration. In this approach, a web-based user interface or an application is created for users of the system to show them all the relevant information by accessing all the source systems directly. There is no unification of data in reality.

2. Middleware Data Integration. A middleware data integration solution is essentially a layer between two disparate systems allowing them to communicate. Middleware integration can act like a glue that holds together multiple legacy applications, making seamless connectivity possible without requiring the two applications to communicate directly.

3. Data Virtualization Integration. Data Virtualization allows us to leave data in the source systems while allowing to create a new set of unified views. […] A lot of organizations today prefer this approach because of the benefits and technologies that exist today to support this approach. The main benefit of the virtual integration approach is near real time view of data from the source systems. It eliminates a need for separate data store for the consolidated unified data. […] [But] that doesn't mean it's the best way to do Data Integration although it certainly has a short term benefit. The drawbacks of this approach include limited […] data history availability or data version management and extra load on the source systems which may have an adverse effect on performance.

4. Data Warehouse Approach. This is the most commonly known approach[…] [especially] if you have read Ralf Kimball and/or Bill Inmon. This approach requires the creation of a new Data Warehouse (of Data Marts) that stores a unified version of data extracted from all the source systems involved and manages it independently of the original source systems. The benefits of this approach include the ability to easily manage history of data (or data versioning), ability to combine data from very disparate sources (mainframes, databases, flat files, etc.) and to store them in a central repository of data.

Design Challenges

1. Having a good understanding of data.

It is very important to have a person (or a team of people) who understand the data assets of the enterprise and also the source systems.[…] Call them Data Champions. The data champions should be able to lead the discussions about the long-term data integration goals in order to make them consistent and successful.

2. Understanding of objectives and deliverables. What is the business purpose behind the data integration initiative? Understanding of the objectives and deliverables for the project is critical to the next steps. What are the source systems? Do the source systems have data to support the business requirements? What are the gaps between data and the requirements? These questions should be answered adequately.

3. Analysis of source systems and extraction. Having a good understanding of the options of extracting data from the source systems is critical to the overall success. Things like frequency of extracts, extent of data extraction (full extract or incremental), and quality of the data in the source systems affect the timeline and overall direction of the project. In addition, it's important to know about the volume of data extract to be able to plan the integration approach along with a knowledge of source system's backup schedules and any specific maintenance windows etc. that may impact the data integration process.

4. Implementation Challenges. Tool selection is also very important to the overall success of any data integration project. A feasibility study should be performed to select the right toolset for the job. Small enterprises or companies who are just starting a data warehousing initiative sometimes find making that decision isn't easy considering the number of options available today. A feasibility study helps map out which tools are best suited for the overall data integration objective for the organization. Sometimes, for organizations that have built their data warehouse and a lot of ETL processes using a tool which is unable to scale or the tool is no longer relevant (for example OWB, which Oracle decided to do away with it to promote a better tool, ODI). In these cases, even matured organizations need to do a feasibility analysis to estimate what it takes to upgrade the existing Data Integration infrastructure to the new toolset.

Source: www.dataintegration.ninja/data-integration-techniques-and-its-challenges
Kumar, G, 'Data Integration Techniques and Their Challenges', Feb 2015