THE DATA DOWNLOAD

As a data management and BI consulting company, we pride ourselves on being able to deliver on our projects as well as provide good quality content to our readers.

BUILDING THE FOUNDATION WITH A HIGH-LEVEL ARCHITECTURE

Develop a high level data flow architecture that is flexible and leverages repeatable design patterns for a best-in-class agile data warehouse environment.

PUBLISHED: MAY 30, 2016

WRITTEN BY: NANCY COUTURE

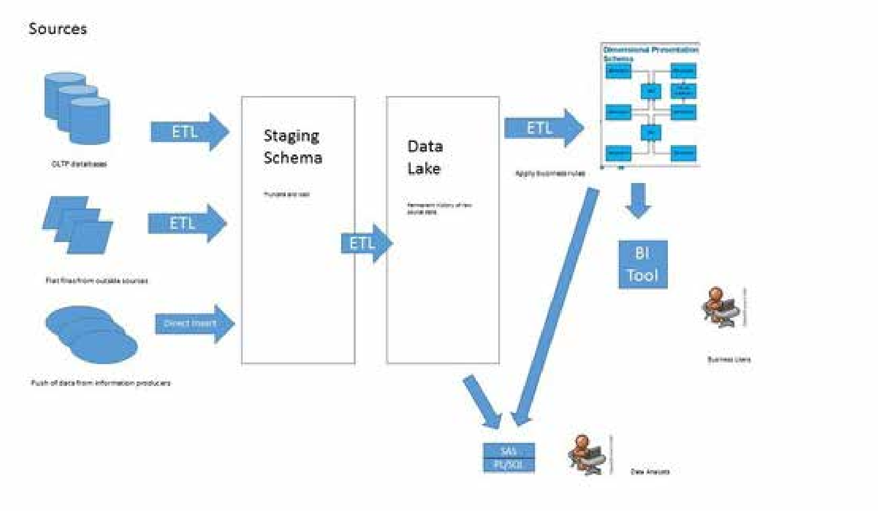
This is the third in a series of articles describing foundational steps to enable agile data warehouse development. This first series of articles describe foundational steps that enable agile data warehouse development – something that has been a challenge in enterprise data management for years. My prior articles published thus far describe how to develop a Business Conceptual Model as a starting point, then building a "grass roots" (at a minimum!) Data Governance capability.

The next focus for setting yourself up for a best in class agile data warehouse environment is to develop a high level data flow architecture that is inherently flexible and leverages repeatable design patterns.

In the end, every data warehouse has an architecture, composed of technical and data related components. The architecture is either planned, or it's developed without a plan. When data warehouses are developed without a predefined architecture, it can severely limit flexibility, and ultimately impact the amount of work it takes to enhance and maintain it. Without a planned architecture, subject areas don't fit together, connections lead to nowhere, and the whole warehouse is difficult to manage and even more difficult and time consuming to change. This can have an even larger negative impact when doing agile development.

The high level architecture should always be designed with an eye toward update and expansion. It should be based on the results of the initial interviews that led you to the business conceptual model, and reviewed by Data Governance, as described in my prior blogs. As a part of the interview process, you should have gotten a sense of the expected user base and usage.

For example, does your company have data scientists or data analysts who will use analytical tools against raw data? If so, your data architecture will need to take that into account. Will your data warehouse be updated with new records or with modifications to existing records? Will there be new data sources that need to be integrated into the data warehouse frequently? The answers to these questions will have an impact on your architectural design.



The architecture we designed in my last organization included our version of a Data Lake that allowed for a permanent history of raw data with very little modification. The Data Lake allowed us to retain a full version history of every source record to support "as is" and "as was" queries. Our data analysts were able to query against the Data Lake for exploration and predictive purposes. The Data Lake also has a number of technical advantages, such as supporting many load patterns and enabling very fast loads of new data so that our data analysts could obtain new source data quickly (agile in action)!

Our architecture included a number of pre-defined design patterns that allowed for faster development that supports agile more directly. These included design patterns for:

- Loading raw data (incremental, full, flat file, manual input, process push)
- Loading dimensions (type 1 and 2)
- Loading detailed fact tables
- Loading consolidated / summary tables

Reuse of design patterns supports agile development in many ways. It speeds development of similar features, minimizes reinvention, and enables new team members to be productive faster.

There are a lot of options when developing a data flow and data architecture for your data warehouse. **These are just a few examples of ways you can design the architecture to support incremental, agile development. As a result of our architecture design:**

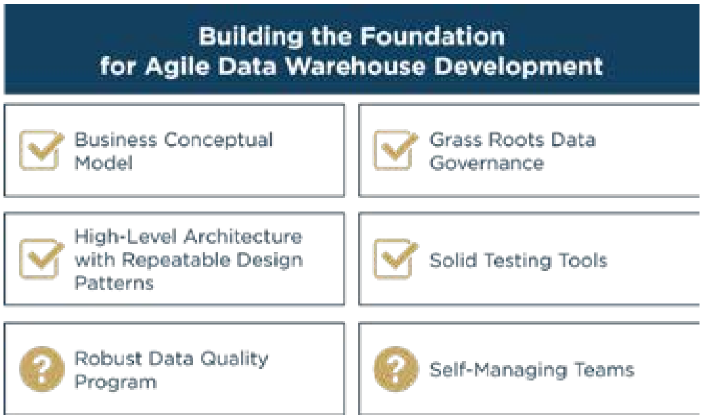
- Most refactoring / reloading has been of a scale that can be completed quickly
- Design patterns provide guidance for new development and speed the orientation of new team members.
- Full history in the Data Lake supports historical reloads, "as was" queries,

experimentation, research, prototyping, and also trouble shooting the source OLTP systems.
- The Data Lake has adapted to multiple loading patterns, including direct push from information producers. This is an architectural pattern we developed that enables some very innovative data management practices (more to come on this topic in future articles).

To sum up, there are many benefits to having a predefined data warehouse architecture. Some of these include:

- Provides an organizing framework - the architecture draws the lines on the map in terms of what the individual components are, how they fit together, who owns what parts, and priorities.
- Improved flexibility and maintenance - allows you to quickly add new data sources, and add / modify data from existing sources.
- Faster development and reuse - warehouse developers are better able to understand the data warehouse process, data base contents, and business rules more quickly.
- Coordinated parallel efforts - multiple, relatively independent efforts have a chance to converge successfully.

All of these benefits also allow you to leverage agile development more readily.



Building the Foundation
for Agile Data Warehouse Development

✓ Business Conceptual Model	✓ Grass Roots Data Governance
✓ High-Level Architecture with Repeatable Design Patterns	✓ Solid Testing Tools
? Robust Data Quality Program	? Self-Managing Teams

Additional steps in building this foundational approach to agile data warehouse development include:

- Ensuring solid testing and tools
- Implementing a robust data quality program
- Giving the development team the ability to self manage their agile development approach, incorporating continuous improvement
- I will cover these remaining steps in the next few upcoming articles.

DATASOURCE CONSULTING
an EXL company

CONTACT US
888.4LEANBI(453-2624) **P**
888.453.2624 **F**

DENVER
2399 Blake Street
Suite 170
Denver, CO 80205

SAN DIEGO
990 Highland Drive
Suite 110-M
Solana Beach, CA 92075