KIMBALL GROUP KIMBALL UNIVERSITY

www.kimballgroup.com Number 111, April 1, 2009

Design Tip #111 Is Agile Enterprise Data Warehousing an Oxymoron?

By Ralph Kimball

In today's economy the data warehouse team is caught between two conflicting pressures. First, we need more immediate, impactful results about our customers and our products and services across the enterprise. In other words, integrate the enterprise's data NOW! But second, we need to allocate our scarce people and money resources more wisely than ever before. In other words, make sure that all our designs are extensible and universal, and avoid any future rework. We can't waste a dime!

If we yield to the first pressure to deliver results too quickly, then we deliver one-off spot solutions that pretty quickly get us in trouble. For example, suppose we show the marketing department a prototype application written by one our developers that provides new demographic indicators on some of our existing customers. Suppose that the application is based on a server side implementation driven through a web interface that calls PHP to pull data from a MySQL database on a small development server. At the time of the demo to marketing, anyone can access the application via a browser if they know the URL. Furthermore the testing was done through Internet Explorer. It doesn't seem to work quite correctly under Firefox, but "we'll fix that soon." In spite of the fact that you got marketing's attention, you are asking for trouble. What's wrong with this picture? If you slow down, and count to ten, you will realize that you are about to have a "success disaster." Your application has little or no data quality, governance, security, or extensibility. It also doesn't help that the person who wrote the PHP program has left the company…

On the other hand, if we revert to the safe IT tradition of using a "waterfall" design where we propose a comprehensive system architecture, write a functional specification, evaluate vendors, and design an enterprise data model, we won't deliver useful results in time to help the business. If IT management is competent, we won't be allowed to start such a project!

Is there a middle ground that we might call an "agile enterprise data warehouse?" Remember that agile development calls for small teams, a succession of closely spaced deliverables, acceptance of a cut-and-try mentality, delivery of code rather than documentation, and intensive interaction throughout the project from end users who effectively control the project. For more on the "agile development" movement, go to Wikipedia, and read Margy's 2005 Design Tip #73.

If organized effectively and in the right cultural environment, perhaps any IT initiative can be a candidate for agile development, but let's not get carried away. Here's my recommendation for a very useful project that will yield measurable results quickly, and is designed not to be politically controversial. Best of all, this is a "stealth project" that when successful, will lay the foundation for an architectural revolution in your environment. The acronym for this stealth project is LWDS-MDM: Light Weight Down Stream Master Data Management!

In Design Tip #106, I described the typical downstream MDM function that exists within most data warehouse environments, given that few of us have a full fledged centralized MDM driving our operational systems. The downstream MDM gathers incompatible descriptions of an entity such as Customer and publishes the cleaned, conformed, and deduplicated dimension to the rest of the data warehouse community. The subscribers to this dimension are almost always owners of fact tables who want to attach this high quality conformed dimension to their fact tables so that BI tools around the enterprise can do drill across BI reports and dashboards on the conformed contents of the dimension.

Your job is to assemble a small team of developers and end users to start building the conformed Customer dimension. This is an ideal target for incremental, adaptive development. Remember that the essential content of a conformed dimension is a small set of “enterprise attributes” that have a consistent interpretation across all the members of the dimension and can be used for constraining and grouping by all forms of BI. Furthermore, the agile development team can start by cleaning and conforming Customer records from a limited set of original sources around the enterprise. Don’t try to design the cosmic solution for Customer metadata! Instead, concentrate on making the first set of attributes and sources produce a usable result in two to four weeks! If you read about agile development, you will be shocked and amazed that you are expected to give a working demo of your system to the end users at the end of every such two to four week sprint!

Hopefully you see how this stealth project should be able to grow to become really powerful. Over time, each sprint adds more attributes and more sources. You are chipping away at the big issues of enterprise data integration, including an understanding of user requirements, data quality, and the need for data governance.

Write to me if you go down this path. Let’s take the agile approach away from the pure methodologists who really don’t understand data warehousing, and let’s show the world a real “use case.”