Kimball Design Tip #45: Techniques For Modeling Intellectual Capital

By Bill Schmarzo

One of the biggest challenges to deploying analytics is that the resulting intellectual capital gets buried inside the tools that are used to build the analytics. Once the intellectual capital gets locked inside one of these tools, it becomes difficult to share that decision making logic with users who might be using different business intelligence, query and reporting tools.

I define intellectual capital as the organizational best practices for a specific well-defined business activity. For example, what is the best way to analyze the introduction of a new product line extension such as the new flavor of tooth paste, in the consumer packaged goods market? This would include not only the data and metrics, but also what constitutes normal performance for each of the supporting metrics including the "baseline". Any performance outside of 2 to 3 standard deviations from the baseline would be considered exceptional. I would then try to understand which of the driving metrics such as the price gaps versus competitive set, sell-through, quantity and quality of retail support, any out-of-stocks or distribution metrics, inventory-on-hand, displays, coupons) were outside of acceptable guidelines.

Successful organizations learn to embed their intellectual capital into their data warehouse design using common dimensional modeling techniques instead of locking it into the tools. And these organizations leverage the 5-stage analytic lifecycle (publishing reports, identifying exceptions, determining causal factors, modeling alternatives and tracking actions) in the business requirements process to tease out their intellectual capital requirements.

In the first stage, publishing reports, the intellectual capital foundation starts with ensuring that you have the "right" data with the "right" grain. This stage is critical for setting the data warehouse foundation with standard metrics, conformed dimensions, common attributes, and the most atomic grain for supporting a common vocabulary across the organization. Dimensional modeling techniques that are useful in the publish reports stage include named hierarchies that support standard "state of the business" reporting as well as real time partitions to support real time reporting needs.

In the exceptions stage, the intellectual capital shows up as the richness and robustness of the dimension tables. For example, let's say that you are trying to identify the specific business entities that are causing performance problems. Dimensional modeling techniques like drill down with atomic grain, and drill across supported by the bus architecture, enable the users to identify those areas of the business that are the sources of exceptional performance.

In the determining causal factors stage, dimensional modeling techniques such as consolidated marts and accumulating snapshots help to understand the causes of exception performance. For example, let's say that we're trying to understand processing bottlenecks for business activities such as claims processing or order tracking. An application could be built using a data analysis tool to retrieve the milestone dates and calculate the time gaps. But instead we'd recommend designing that logic into the data warehouse using the accumulating snapshot technique so that the users can see the processing pipeline in one simple database format, regardless of which tool they are using.

In the evaluating alternatives and tracking actions stages, dimensional modeling techniques such as

mini-dimensions play a role in capturing and employing the analytic results. For example, let's say that your analytic process yields customer scores for up-selling, cross-selling, private label purchases, credit card fraud, web site visits, email responses, and customer attrition. We might want to track these scores over time in order to analyze the impact of a marketing program on these customer scores. We can make it easier to track and share this intellectual capital by designing these scores into the data warehouse design using demographic mini-dimensions instead of burying them in unpredictable places in the schema. We can also use architectural techniques such as a hot response cache to rapidly deliver the results in the operational environment, thereby "closing the analytic loop."

There are many practical dimensional modeling techniques that can be used to ensure that the intellectual capital that results from the analytic lifecycle gets designed back into the data warehouse instead of being locked into the tools used to build the analytics. This dramatically improves the organization's ability to capture, share and re-use their intellectual capital.