

Kimball Design Tip #60: Big Shifts Happening in BI

By Ralph Kimball

At last week's Business Intelligence Perspectives Conference hosted by *Computerworld* magazine in Palm Springs, two interesting themes were very evident that signaled some big shifts in the business intelligence (BI) world.

1. *Compliance is a Free Pass for BI*

A number of speakers marveled at how the new regulatory compliance requirements for financial disclosures, especially the Sarbanes Oxley Act, were opening the pocketbooks of companies to upgrade their BI environments. One speaker said, "All you have to do is mention compliance and the funding proposal is approved." But most of the speakers expressed simultaneous concern that no one knows just what the compliance requirements really mean. Not only are the requirements not spelled out in concrete database technical terms, but it appears that the practical impact of the compliance requirements may have to be played out in the courts, with IT departments defending their practices as "commercially diligent" and responsible.

Obviously, most IT departments interested in meeting the compliance requirements will try to err on the conservative side. And, of course, Sarbanes Oxley is not the only game in town. There are probably a dozen overlapping financial reporting statutes with similar requirements, depending on where you do business.

A conservative approach to meeting most compliance requirements would suggest the ability to:

- Prove (backward) lineage of each final measure and KPI appearing in any report
- Prove (forward) impact of any primary or intermediate data element on a final report
- Prove input data has not been changed
- Prove final measures and KPIs are derived from original data under documented transformations
- Document all transforms, present and past
- Maybe: re-run old ETL pipelines
- Maybe: show all end user and administrative accesses of selected data

This list is grist for ten more Design Tips!

2. *Sequential Behavior Analysis is BI's Mount Everest*

Some of the most interesting and scary case studies at the BI conference were descriptions of trolling huge databases to answer customer behavior questions. Andreas Weigend, Stanford Professor and former Chief Scientist at Amazon, described a study done at Amazon finding the delay (in days) between a customer first clicking on a product and then eventually buying that product. This is immensely difficult. Since most clicks do not result in purchases, you have to wait until a purchase is made and then look backward hours, days, or weeks in the blizzard of records in the clickstream to find the first click by that customer on that product.

The potential volume of data that entities like Amazon want to look through is staggering. Amazon stores every link exposure in their historical data. A link exposure is the presence of a link on a displayed page. It doesn't mean the user clicked the link. Amazon is capturing terabytes of link

exposure data each day!

Link exposures are just the beginning of a biblical flood of data, which will really get serious when RFIDs are deployed down to the individual stock item level. Not only is the volume of data horrifying, but the data is often captured in different servers, each representing “doorways” at different locations and times. These challenges raise the question whether the relational model and the SQL language are even appropriate. Yet the ad hoc questions people want to ask of this data demand the same kind of access that relational databases have been so successful in providing against much smaller data sources.

These two big shifts in BI have different personalities. The first (compliance) is like ballast, and the second (behavior) is like a balloon. But in my opinion, both are real and permanent. They will keep us busy.