

## **Design Tip #77 Warning – Summary Data Alone Is Hazardous To Your Health**

By Margy Ross

An overarching false statement about dimensional models is that they're only appropriate for summarized information. Some people maintain that data marts with dimensional models are intended for managerial, strategic analysis and therefore should be populated with summarized data, not operational details.

We strongly disagree! Dimensional models should be populated with the most detailed, atomic data captured by the source systems so business users can ask the most detailed, precise questions possible. Even if users don't care about the particulars of a single transaction or sub-transaction, their "question of the moment" requires summarizing these details in unpredictable ways. Of course, database administrators may pre-summarize information, either physically or via materialized views, to avoid on-the-fly summarization in every case. However, these aggregate summaries are performance-tuning complements to the atomic level, not replacements.

If you restrict your dimensional models to summarized information, you will be vulnerable to the following shortcomings:

- Summary data naturally pre-supposes the typical business questions. When the business requirements change, as they inevitably will, then both the data model and ETL system must change to accommodate new data.
- Summary data limits query flexibility. Users run into dead ends when the pre-summarized data can't support an unanticipated inquiry. While you can roll up detailed data in unpredictable ways, the converse is not true – you can't magically explode summary data into its underlying components.

When critics authoritatively state that "dimensional models pre-suppose the business question, are only appropriate for predictable usage, and are inflexible," they're conveying the hazards of pre-summarization, not dimensional modeling. If dimensional models contain atomic data, as we advocate, then business users can roll-up or drill-down in the data ad infinitum. They can answer previously unexpected questions without any change to the database structures. When new attribute or measure details are collected by the source system, the atomic data model can be extended without disrupting any existing business intelligence (BI) applications.

Some people advocate an approach where the atomic data is stored in a normalized data model, while summary data is stored dimensionally. The atomic details are not completely ignored in this scenario; however accessing them for user consumption is inherently restricted. Normalized structures remove data redundancies to process transactional updates/inserts more quickly, but the resulting complexity causes navigational challenges and typically slower performance for BI reports and queries. While normalization may save a few bytes of storage space, it defeats the users' ability to seamlessly and arbitrarily traverse up, down and across the detailed and summary data in a single interface. In the world of DW/BI, query functionality/performance trumps disk space savings.

As the architect Mies van der Rohe is credited with saying, "God is in the details." Delivering dimensional models populated with the most detailed data possible ensures maximum flexibility and extensibility. Delivering anything less in your dimensional models undermines the foundation necessary for robust business intelligence (and is hazardous to the health and well-being of your overall DW/BI environment).