

Design Tip #92 Dimension Manager and Fact Provider

By Bob Becker

Conformed dimensions are the glue that ties together your enterprise data warehouse. To facilitate and manage the conforming process, we have identified two additional fundamental responsibilities for the DW/BI team: the dimension manager and fact provider. Typically these functions are performed by the ETL team working closely with the data stewardship organization.

The dimension manager is a centralized authority who prepares and publishes conformed dimensions to the data warehouse community. We first introduced this new data warehouse role and its partner role, the fact provider, in *The Data Warehouse ETL Toolkit*. A conformed dimension is by necessity a centrally managed resource: each conformed dimension must have a single, consistent source. It is the dimension manager's responsibility to administer and publish the conformed dimension(s) for which he has responsibility. There may be multiple dimension managers in an organization. The dimension manager's responsibilities include the following ETL processing:

- Implement the common descriptive labels agreed to by the data stewards and stakeholders during the dimension design.
- Add new rows to the conformed dimension for new source data, generating new surrogate keys.
- Add new rows for Type 2 changes to existing dimension entries (true physical changes at a point in time), generating new surrogate keys.
- Modify rows in place for Type 1 changes (overwrites) and Type 3 changes (alternate realities), without changing the surrogate keys.
- Update the version number of the dimension if any Type 1 or Type 3 changes are made.
- Replicate the revised dimension simultaneously to all fact table providers.

It is easier to manage conformed dimensions in a single tablespace in a single DBMS instance on a single machine because there is only one copy of the dimension table. However, managing conformed dimensions becomes more difficult in multiple tablespace, multiple DBMS, or multi-machine distributed environments. In these situations, the dimension manager must carefully manage the simultaneous release of new versions of the dimension to every fact provider. Each conformed dimension should have a version number column in each row that is overwritten in every row whenever the dimension manager releases the dimension. This version number should be utilized to support any drill across queries to assure that the same release of the dimension is being utilized. *The Data Warehouse ETL Toolkit* has more details on how to instrument this drill-across capability incorporating the dimension release version number.

The fact table provider is responsible for receiving conformed dimensions from the dimension managers. The fact provider owns the administration of one or more fact tables and is responsible for their creation, maintenance and use. If fact table are used in any drill across applications then by definition the fact provider must be using conformed dimensions provided by the dimension manager. The fact provider's responsibilities are more complex. They include:

- Receive or download replicated dimension from the dimension manager. In an environment where the dimension can not simply be replicated but must be updated locally, the fact provider must process dimension records marked as new and current to update current key maps in the surrogate key pipeline and also process any dimension records marked as new but postdated.
- Add all new records to fact tables after replacing their natural keys with correct surrogate

- keys.
- Modify records in all fact tables for error correction, accumulating snapshots, and late arriving dimension changes.
 - Remove aggregates that have become invalidated and/or recalculate affected aggregates. If the new release of a dimension does not change the version number, aggregates have to be extended to handle only newly loaded fact data. If the version number of the dimension has changed, the entire historical aggregate may have to be recalculated.
 - Quality assure all base and aggregate fact tables. Be satisfied that the aggregate tables are correctly calculated.
 - Bring updated fact and dimension tables on line.
 - Inform end users that the database has been updated. Tell users if major changes have been made, including dimension version changes, postdated records being added, and changes to historical aggregates.

In order for your conformed dimension strategy to be effective over the long haul, your conformed dimensions must remain in sync across all fact tables. Formally establishing dimension managers and fact providers proactively assigns responsibilities to specific individuals to ensure a sustainable conformed environment over time. We recommend that you assign a dimension manager to each major conformed dimensions and a fact provider for each fact table. Certainly, in many organizations, an individual may be responsible for multiple dimension or fact tables.