2. Explain steps involved in building Detailed Dimensional Model.

Ans:

Identify the Data Sources

An important component of the modeling process is determining the most appropriate data sources to populate the target design. Detailed data analysis is a significant part of the design effort. Several levels of analysis must occur. First, you must understand the data required to support the business requirements to help select the candidate data sources. Second, you must gain an in-depth understanding of each candidate data source, typically through a data profiling process. Part of the challenge of developing the dimensional model is aligning the business requirements with the unfortunate realities of the data available in the enterprise.

Understand Candidate Data Sources

After you have created the high level design, but before you can progress very far in creating detailed table designs, you need to make decisions about the data source for each table. The first step in this process is to understand all of the sources that are candidates for populating your models. Some candidate data sources may be business processes and dimensions identified in the bus matrix. Additional candidate sources can be identified by the IT veterans who are familiar with the data lurking in the various nooks and crannies of the organization. They also know about the history and challenges of these sources.

Profile and Select the Data Sources

From this list of candidate data sources, the design team needs to evaluate each and determine the best source. Gather and carefully review whatever documentation is available for the source systems, such as data models, file definitions, record layouts, written documentation, and source system programs. More advanced source systems may have their own metadata repositories with all this information already integrated for your convenience.

Establish Conformed Dimensions

During these detailed design sessions, the key conformed dimensions are being defined. Because the DW/BI system is an enterprise resource, these definitions

must be acceptable across the enterprise. The data stewards and business analysts are key resources that should reach out to achieve organizational consensus on table and attribute naming, descriptions, and definitions. The design team can take the lead in driving the process, but it is ultimately a business task to agree on standard business definitions and names for the design. This will take some time, but it is an investment that will provide huge returns in terms of users' understanding and willingness to accept the dimensional model. Don't be surprised if folks in the front office, like the business sponsor/driver or governance steering committee, must get involved to resolve conformed dimension definition and naming issues.

Identify Base Facts and Derived Facts

The data profiling effort will identify the counts and amounts generated by the measurement event's source system. However, fact tables are not limited to just these base facts. There may be many more metrics that the business wants to analyze that are derived from the base facts, such as year-to-date sales, percentage difference in claims paid versus prior year, and gross profit. You need to document the derived facts, as well as the base facts.

Document the Detailed Table Designs

The attribute and metrics list discussed earlier and shown in Figure 7-2 is used as a working deliverable in the early phases of the detail design process. This tool provides an easy place to capture the basic information about each table and column. We prefer this spreadsheet tool because it allows the data modeler to quickly move or copyattributes around the model as needed, so the documentation is efficiently updated between modeling sessions. Providing freshly updated documentation for each design session helps move the process along more quickly because there is less time spent rehashing changes already agreed upon. We usually use the simplified attributes list during the early design discussions.

Update the Bus Matrix

During the detailed modeling process, there is often new learning about the business process being modeled. Frequently, these findings result in the introduction of new fact tables to support the business process, new dimensions, or the splitting or combining of dimensions. It is important to keep the initial

bus matrix updated throughout the design process because the bus matrix is a key communication and planning device that will be relied upon by project sponsors and other team members. It also serves as a communication tool for the design team in discussions with other designers, administrators, and business users. The matrix is very useful as a high level introduction to the overall DW/BI system design and implementation roadmap. The matrix shown in Figure 7-7 is an illustration of a more detailed bus matrix that has been updated and extended by the design team to communicate additional information about the associated fact table's granularity and metrics.

Review and Validate the Model

Once you're confident in the model, the process moves into the review and validation phase identified in Figure 7-1. This phase involves reviewing the model with successive audiences, each with different levels of technical expertise and business understanding to get feedback from interested people across the organization. At a minimum, the design team should plan on talking to three groups:

- Source system developers and DBAs who can often spot errors in the model very quickly
- Core business or power users who were not directly involved in the model development process
- Finally, the broader user community.