DWBI : Unit 2,3 Imp Questions

Q4. Explain Factless fact tables and consolidated fact tables .

Ans:

A] Factless Fact Tables : Sometimes organizations have business processes where an event occurs, but no quantifiable measurements are created or generated. These events may be very important to the business because they represent the existence of a relationship between several dimensions, despite the fact that no counts or amounts are generated when the many-to-many collision of dimensions occurs. These factless events result in factless fact tables. Factless fact tables are the preferred method for recording events in a data warehouse where there is no natural numeric measurement associated with the event.

Exmple : *You can imagine a factless fact table for capturing student attendance events. You load a row into the fact in Figure 6-24 whenever a student shows up for a class. The grain is one row per class attended by a student each day.We typically include a dummy counter fact whose value is always 1. You don't even need to physically store this fact; it merely facilitates counting to ask questions such as "how many students attended a class taught by this professor today?" If the grain of the fact table was one row for each student enrollment (not attendance) in a class each day, then the attendance counter could take on a 1 value if the student attended class or zero otherwise. In this scenario, the fact table is no longer factless because the 1 or 0 valuation is a measurement*

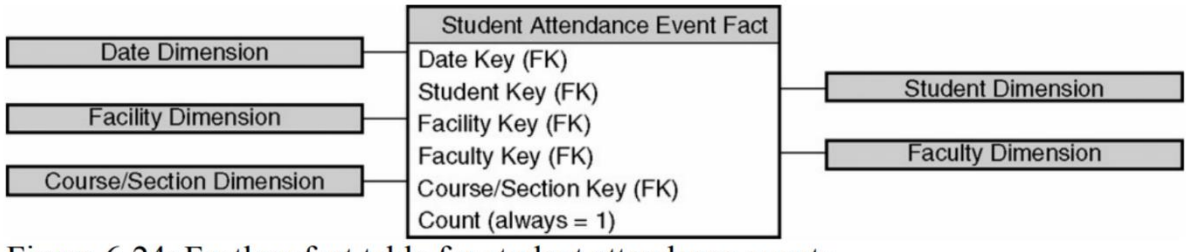

Figure 6-24: Factless fact table for student attendance events

B] Consolidated Fact Tables : Several single source dimensional models have been implemented, it is reasonable to potentially combine them into a consolidated dimensional model. Consolidated fact tables merge measurements from multiple processes into a single fact table. Because the measurements physically sit side-by-side in a fact row, they need to be rolled up to a single common level of granularity. As shown in Figure 6-25, the consolidated actual versus forecast schema brings the measurements together for simplified variance analysis.you could issue separate queries to a forecast fact table and a detailed fact table of actuals, and then combine the results on common dimension attributes; however doing this consolidation once in the ETL system rather than at query time will result in faster query response and a less complicated presentation.

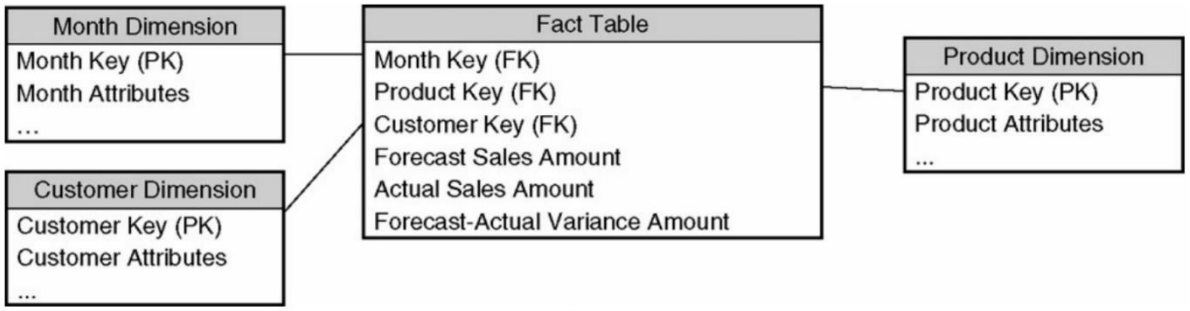

Figure 6-25: Consolidated fact table combining actual and forecast data

Q8. Explain Snowflaking and Outriggers.

Ans:

1) Snowflaking :

Dimensions are snowflaked when the redundant attributes and decodes are removed to separate tables and linked back into the original table with artificial keys. In other words, a snowflaked dimension is normalized. An example snowflaked product dimension table is shown in Figure 6-13.

Generally, snowflaking is not recommended for your dimensional models because it almost always makes the user presentation more complex and less legible.Snowflaking also makes most forms of browsing among the dimension attributes slower.

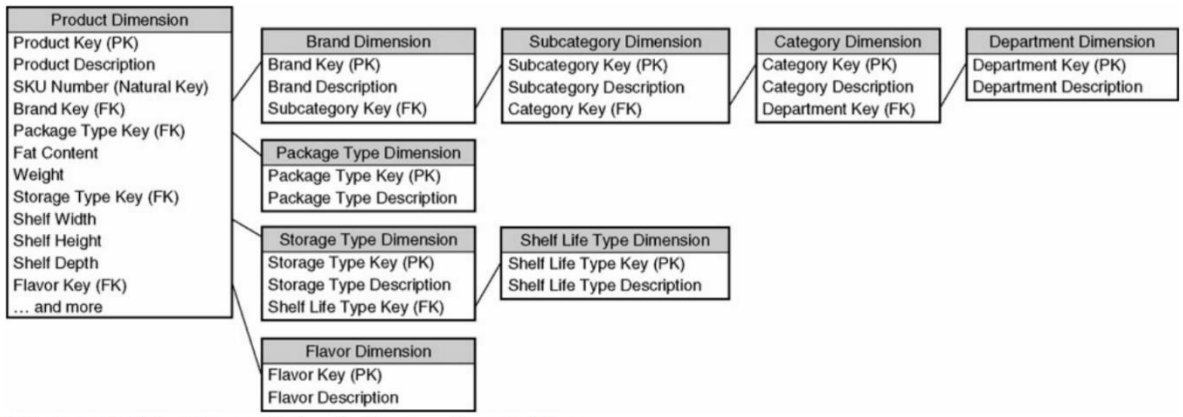

Figure 6-13: Snowflaked dimension tables

2) Outriggers :

Q12. What is a Fact Table granularity? Explain different types of Fact Tables?

Ans:

Fact Table Granularity :

1) The fact table's grain is the business definition of the measurement event that produces the fact row. While the grain determines the fact table's primary key, granularity itself is always expressed in business terms.

2) All the rows in a fact table should have a single uniform grain. The grain is determined by the physical realities of the data's source; its declaration becomes clear when you visualize the measurement process. The grain is your anchor: you must include enough dimensional context in your design to implement the grain, and you must exclude dimensional context that violates the grain.

3) Fact tables should contain the lowest, most detailed atomic grain captured by a business process. There is tremendous power, flexibility, and resiliency in granular atomic fact data. It allows queries to ask the most precise questions possible.

4) Granular atomic fact tables are also more impervious to changes; they can be gracefully extended by adding newly sourced facts, newly sourced dimension attributes, and by adding entirely new dimensions.

Read : 1. Transaction Fact Tables. 2. Periodic Snapshot Fact Tables

3] Accumulating Snapshot Fact Tables

1) The third type of fact table is the least frequently encountered, however it serves a very powerful function in certain applications. accumulating snapshots are used to represent activity over an indeterminate time span for processes that have a well defined beginning and end.

2) In addition to the typical quantities and amounts, accumulating snapshots often contain two additional types of facts. The first is the milestone counter; there is often a lengthy list of counters corresponding to the key milestones in the pipeline, which take on a 1 or 0 value. Accumulating snapshots often have lag facts representing the length of time between milestones, such as the elapsed time between the first and second milestone, or the first milestone and the final event in the pipeline.