xD - Dimensional Modeling

Rules for fact table design:

1. Primary key of your fact table has minimum number of columns as possible with no surrogate keys. Should be made up of FK’s and degenerate dimensions. Fk’s should be generated ID’s.
2. Referential integrity is a must, foreign key in the fact table must have a value.
3. Avoid NULL values in foreign key by using FLAGS (e.g. no shopper card).
4. The granularity of fact table must be the lowest.????
5. Each fact should be additive across all dimensions. Each fact should be the same granularity. Identify what each row of the fact table would mean.

Rules of dimension table design:

1. Verbose attribute values: should be descriptive as possible.
2. Descriptive columns: Should be easy to tell what the column means.
3. Complete: No Null/Empty values in any of the attributes.
4. Discrete values: One business entity per row.
5. Quality Assured: Data is clean and consistent
6. Should always contain a business/natural key or legacy PK
7. Always have a surrogate primary key. Don’t have dependent PK from another system.

Date dimension

1. Leave -1 for unknown value in date dimension
2. PK’s can be YYYYMMDD
3. Explain
   1. conformed dimensions Video 4.6.1
      1. These are master or common reference dimensions and a key part of the enterprise bus technical architecture. Shared across business process aka the fact tables in the DW.
      2. Contain a superset of attributes required by all the fact tables.
      3. Types of conformed dimensions:
         1. Identical dimensions: exactly the same dimensions
         2. Perfect subset of an existing dimension: Convert one row per day into one row per month/year. For same business process but measured through different point in time.
      4. Retrieve values from multiple tables to answer one business process
   2. role-play dimensions. Video 4.6.2
      1. Same dimension is used more than once
      2. Common in date dimension
      3. Stored in the same table, just aliased as a view.
      4. Implemented as multiple FK’s in the fact table to the same dimension table
      5. E.g. Date: Order date, shipping date, delivery date -> same date
      6. E.g Address: Ship to, bill to -> same address dimension
4. Review (SCD) Slowly Changing Dimensions. Videos 4.4.1 – 4.4.4
   1. Dimensions that change infrequently, if they change they need to be addressed
   2. Popular strategies:
      1. Type 1: Overwrite the existing attribute.
         1. Correcting mistakes or errors in data
         2. Change historical associations that don’t matter
         3. Old value has no significance. A customer email address. Change of an employee name.
      2. Type 2: Add a new dimension row
         1. Natural/business keys are repeated
         2. Old and new values are stored along with effective dates. Starting date to expiration date.
         3. Simpler construct than Type 3
      3. Type 3: Add a new dimension attribute. Adding columns
         1. Infrequently used. Preserves history.
         2. Useful for soft changes, where users want to choose between the new and old attribute.
         3. New sales territory and old sales territory. Add new value to existing column and the old value is stored in the new column.
5. Explain (RCD) Rapidly Changing Dimensions. Video 4.5.1
   1. Not often and with no consistency.
   2. A change in customer’s street address.
   3. Change in a customer’s shirt color at the time of a product purchase.
6. Explain
   1. degenerate dimensions Video 4.5.2
      1. Another form of RCD. Values that change so frequently.
      2. There are too many of them for their own dimension. E.g. Order number, flight number, customer age
      3. Degenerate dimensions are business keys. They end up as primary key of the fact table.
      4. Any attribute in the fact table that is not a dimension key or fact is considered a degenerate dimension.
   2. mini-dimensions Video 4.5.3
      1. Attributes change frequently, consider placing them in their own, mini-dimensions.
      2. Separate all the SCD’s and RCD’s separately. Which combines to make the main dimension table for a customer.
7. Explain
   1. Junk dimensions Video 4.6.5
      1. Miscellaneous low-cardinality flags.
      2. Does not fit within any other dimension. Do not make mini dimension tables, instead combine all them to make Junk Dimensions.
   2. factless facts. Video 4.3.4
      1. Don’t include aggregate values in the fact table.
      2. Don’t add additive facts.
      3. Don’t have values that are derived from other columns as a separate column.
      4. Factless fact tables:
         1. Business process that do not generate quantifiable measurements
            1. E.g Student attendance, college admissions
         2. Can be easily converted into traditional fact tables by adding an attribute, Count which is always equal to 1.
         3. An event occurs and there is no quantifiable value, except for the fact that the event happened.