Topics for dwh test

- Bi tools landscape - self service,corporate, data mining
- dwh diff b/t oltp and olap
- characterization of funcs in the dwh
- benefits of dwh
- dwh proj (life cycle)
- dwh executing env (staging area, etl processes)
- fact table, measures, types of measures
- rule sof fact tables
- dimensions
    - 4 types of slow changing discussed in class
    - 0,1,2,3
- diff types of dim
    - role, conformed, degenerate
    - what are they how u use them
- star schema vs snowflake
- shape of dim table
    - rows, attr
- rules of dim table design
- data in the dimen table

- cube, query language, rollup, drilldown, slice ,dice

- tabular

- tabular vs dim model
- bi tools learning curve
- characteristics of in-memory db
- reasons to adopt tabular model
- smp vs numa architecture
- xvelocity engine diff bt traditional db
- imp of rsrc bt a multidimension and tabular server
- kpi
    - design
    - identify
    - charactersitcs of good kpi
    - not as much emphasize on kpi and etl

- etl

    - executing env
    - phases of etl process

○ data cleansing, types of data cleansing
        ○ loading data in the dwh

  • reporting life cycle

  • reporting srevices architecture

Test Format

  • 10mcq
  • 10tf
  • 5sa

Sample Test Questions

5 Short Answers (4 marks each)

1. Give four characteristics of an in memory database [4 marks]
2. Identify 4 problems with the following dimension table (see picture) [4 marks]
3. List at least 4 rules of Dimension table design
4. List at least 4 characteristics of a data warehouse
5. List the relative importance of the following rsrcs for a multidimensional server as opposed to a tabular server:
        ○ State whether it is critical, important, not important or not applicable:

Multidimensional Tabular

RAM
RAM Speed
Number of Cores
Core Speed
Disk Speed
SSD Disk Drive

10 MCQ (2 marks each)

1. What cube operation produces the cube on the right from the cube on the left a. slice b. rollup c. drill down d. cut across

2. Which statement is false? a. The granularity of your fact tables should be at the lowest, most detailed atomic grain captured by the business process. b. each fact should be Addtiive, or re-designed to be as additive as possible c. each fact must be of the same granularity d. multiple processses are combined in a single fact table

3. which of the following best describes a key performance indicator? a. a business metric used to evaluate performance factors b. a law enforced for employees in a business c. a way of calcualting how many employees were terminated in a year d. An approach to measuring a company's morale

4. Which is not a part of the *Data Cleansing* a. compute aggregates b. conversion and normalization c. special-purpose cleansing d. domain-independent cleansing

5. data warehouse is a. the actual discovery phase of a knowledge discovery process b. the stage of selecting the right data for a KDD process c. A subject-oriented integrated time variant non-volatile collection of data in support of management d. None of these

6. In which type of SCD(slowly changing dimensions) do we preserve history of data: a. Type Zero b. Type One c. Type Two d. None of these

7. Sequence of jobs to load data into warehouse: a. First load data into fact tables then dimension tables, then Aggregates if any b. First load data into dimension tables then fact tables, then Aggregates if any c. First aggregates then load data into dim tbls, then fact tbls d. does not matter if we load either of fact, dimensions or aggregates

8. A role dimension is: a. described in the same way across multiple facts and/or data marts ensuring consistent reporting across the enterprise b. combination of a number of misc, low-cardinality tags and indicators c. used on more than one axis e.g. Supply Date, Due Date, Order Date d. dimensions without a hierarchy

9. From the user point of view, the data warehouse is ____ a. read only a. write only a. read write only a. none

10. Which statement is false? The problems with current relational databases are: a. can no longer offer timely resp bc of high access latency b. cannot be used for apps like gaming, RT bidding, reacting to rapid changes in stock market c. cannot maintain concurrency control d. un-acceptable performacne for RT apps

10 T/F (1 mark each)

1. KPIs are directly linked to industry standards (T/F)
2. Self Service BI are based on formatted reports that are typically based upon approved corporate data
3. The language for the tabular model, DAX, is easier to learn than MDX
4. Data in the operational systems is loaded directly in the data warehouse
5. One can easily write data warehouse queries in SQL

6. Full pre-aggregation gives fast response time
7. OLTP is used to run the business
8. OLTP is subject oriented
9. KPIs are measurements that are the by-products of financial reports
10. In tabular computing, RLE is only applied when the size of the compressed data is smaller than the original data.