

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
SECOND SEMESTER, 2017-2018
COMPREHENSIVE EXAMINATION
PART A (CLOSED BOOK)
SS G515 Data Warehousing

Date : 11-05-2018 (AN)

Max.Duration : 90 minutes

Max.Marks : 40(20%)

I. 1-20 are multiple choice type questions. [50% Negative Marking] [0.5*20=10 marks]

1. Fact tables are which of the following?
 - A. Completely denormalized
 - B. Partially denormalized
 - C. Completely normalized
 - D. Partially normalized
2. A star schema has what type of relationship between a dimension and fact table?
 - A. Many-to-many
 - B. One-to-one
 - C. One-to-many
 - D. All of the above.
3. Transient data is which of the following?
 - A. Data in which changes to existing records cause the previous version of the records to be eliminated
 - B. Data in which changes to existing records do not cause the previous version of the records to be eliminated
 - C. Data that are never altered or deleted once they have been added
 - D. Data that are never deleted once they have been added
4. A multifield transformation does which of the following?
 - A. Converts data from one field into multiple fields
 - B. Converts data from multiple fields into one field
 - C. Converts data from multiple fields into multiple fields
 - D. All of the above
5. In Star Schema Dimension tables are:
 - A. Short and Fat
 - B. Long and Thin
 - C. Long and Fat
 - D. Short and thin
6. Snowflaking means
 - A. Normalizing the data
 - B. Denormalizing the data
 - C. Customizing the data
 - D. None of these
7. Sequence of jobs to load data in to 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 dimension tables, then fact tables
 - D. Does not matter if we load either of fact, dimensions, or aggregates
8. In which type of SCD(Slowly changing dimensions) do we preserve history of data:
 - A. Type One
 - B. Type Two
 - C. Type Three
 - D. None of these
9. In datamarts stovepipe means:
 - A. Similar Data
 - B. Isolated data
 - C. Normalized Data

- D. None of above
10. During ETL load we generally have
- A. Unsorted data for Aggregator
 - B. Sorted data for Aggregator
 - C. Does not matter if we use Sorted or Unsorted data for Aggregation
 - D. None of Above
11. In 4 step dimensional process, declaring grain of business process is:
- A. First Step
 - B. Second Step
 - C. Third Step
 - D. Fourth Step
12. Drill Across generally use the following join to generate report:
- A. Self Join
 - B. Inner Join
 - C. Outer Join
 - D. None of these
13. _____ maps the core warehouse metadata to business concepts, familiar and useful to end users.
- A. Application level metadata.
 - B. User level metadata.
 - C. End user level metadata.
 - D. Core level metadata.
14. The star schema is composed of _____ fact table.
- A. one.
 - B. two.
 - C. three.
 - D. four.
15. The key used in operational environment may not have an element of _____.
- A. time.
 - B. cost.
 - C. frequency.
 - D. quality.
16. Data warehouse contains _____ data that is never found in the operational environment.
- A. normalized.
 - B. informational.
 - C. summary.
 - D. denormalized.
17. Detail data in single fact table is otherwise known as _____.
- A. monoatomic data.
 - B. diatomic data.
 - C. atomic data.
 - D. multiatomic data.
18. _____ is a good alternative to the star schema.
- A. Star schema.
 - B. Snowflake schema.
 - C. Fact constellation.
 - D. Star-snowflake schema.
19. The biggest drawback of the level indicator in the classic star-schema is that it limits _____.
- A. quantify.
 - B. qualify.
 - C. flexibility.
 - D. ability.

20. In multi-dimensional analysis, slicing can be achieved by:
- moving up a dimension hierarchy like city -> state, etc.
 - moving down a dimension hierarchy like state -> city, etc.
 - adding a new dimension
 - removing one or more dimensions
-

II. Match the columns:

[5]

- | | |
|---------------------------------|-------------------------------|
| 1. information package diagrams | A. determine data extraction |
| 2. need for drill-down | B. provide OLAP |
| 3. data transformations | C. provide data feed |
| 4. data sources | D. influences load management |
| 5. data aging | E. query management in DBMS |
| 6. sophisticated analysis | F. low levels of data |
| 7. simple and complex queries | G. larger staging area |
| 8. data volume | H. influence data design |
| 9. specialized DSS | I. possible pollution source |
| 10. corporate data warehouse | J. data staging design |
-

Section – III (True/False) [50% Negative Marking]

[0.5 * 10 = 5]

- The number of cells in a multidimensional data cube should equal the number of records in the corresponding relational table.
 - The results of both the CUBE and ROLLUP operators can be achieved without these operators, by using a number of SELECT statements connected by the UNION operator.
 - The CUBE, ROLLUP and GROUPING SETS operators can be combined in a SQL statement to provide exactly the summary totals needed.
 - The query rewriting process substitutes materialized views for fact and dimension tables in a query.
 - MOLAP has smaller storage requirements, but it also has a slower response time than ROLAP.
 - Independent data marts are often created because an organization focuses on a series of short-term business objectives.
 - Periodic data are data that are physically altered once added to the store.
 - Data scrubbing can help upgrade data quality; it is not a long-term solution to the data quality problem.
 - Technical metadata is usually less structured than business metadata.
 - Data warehouse architecture is just an overall guideline. It is not a blue print for the data warehouse.
-

IV. Differentiate the following :

[2*4=8]

- Two-Way Aggregates and Three –Way Aggregates
 - Slice and Dice Operation
 - Drill-through and Drill across
 - Constructive merge and Destructive merge
-

V. Short Question/Answer

[2*6=12]

- State three data delivery paradigms that require an end-to-end perspective in Real Time Data Warehousing.
- What are junk dimensions? How do you model them in dimensional modeling?
- State two main limitations of SQL to handle multi-dimensional analysis.
- Discuss the effect of Sparsity on Aggregation.
- State the role of Audit dimension in dimensional modeling. Give any two attributes of Audit dimension.
- How are the queries in Data Warehouse different from queries in Operational Systems?
