Birla Institute of Technology and Science, Pilani

Work Integrated Learning Programme Division

BITS-Wipro Infotech Collaboration:: M. Tech. Systems Engineering

II SEMESTER   2015-2016

COMPREHENSIVE SEMESTER EXAMINATION (Make up)

Course Title	: Data Warehousing	**Date**	:
Course No.	: SEWP ZG514	**Nature**	: Open book
Weightage	: 60 marks	**Duration**	: 3 Hrs.

Note: Attempt all the questions

Q1. How can you apply the data to the warehouse? What are the modes?   [5]
Answer:
Data may be applied in the following four different modes: load, append, destructive merge, and constructive merge. Let us understanding of the effect of applying data in each of these four modes:

Load If the target table to be loaded already exists and data exists in the table, the load process wipes out the existing data and applies the data from the incoming file. If the table is already empty before loading, the load process simply applies the data from the incoming file.
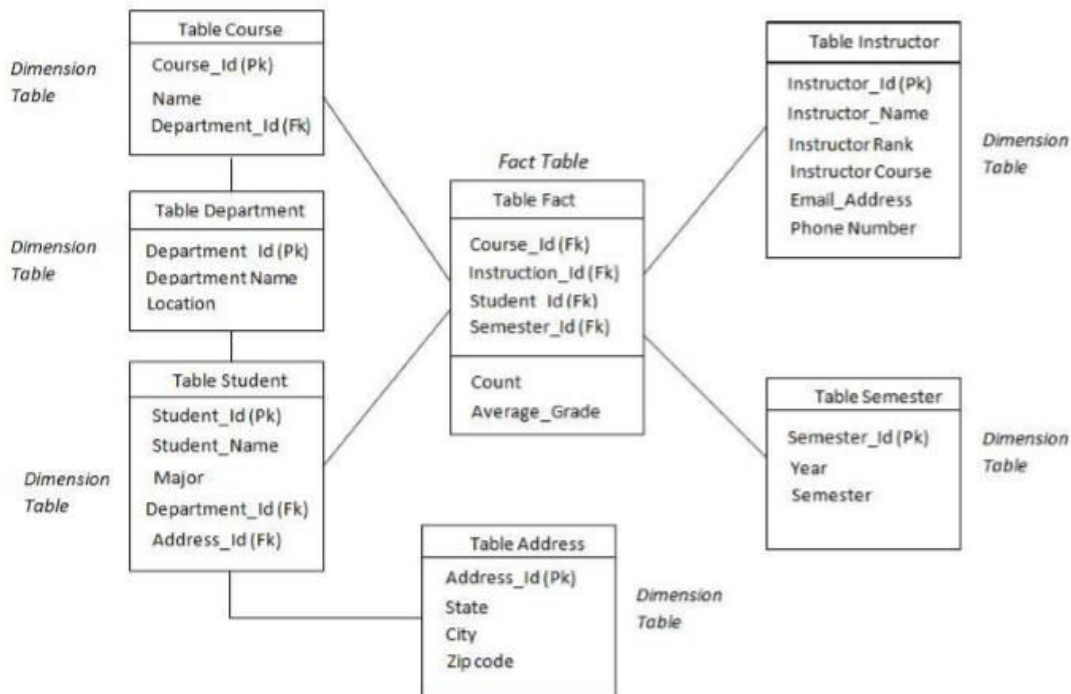
Append You may think of the append as an extension of the load. If data already exists in the table, the append process unconditionally adds the incoming data, preserving the existing data in the target table. When an incoming record is a duplicate of an already existing record, you may define how to handle an incoming duplicate. The incoming record may be allowed to be added as a duplicate. In the other option, the incoming duplicate record may be rejected during the append process.

Destructive Merge In this mode, you apply the incoming data to the target data. If the primary key of an incoming record matches with the key of an existing record, update the matching target record. If the incoming record is a new record without a match with any existing record, add the incoming record to the target table.

Constructive Merge This mode is slightly different from the destructive merge. If the primary key of an incoming record matches with the key of an existing record, leave the existing record, add the incoming record, and mark the added record as superseding the old record.

Q2. Suppose that a data warehouse for Big_University consists of the four dimensions student, course, semester, and instructor, and two measures count and avg_grade. At the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg_grade measure stores the actual course grade of the student. At higher conceptual levels, avg_grade stores the average grade for the given combination. Draw a snowflake schema diagram   [7]
Answer:

Table Course (Dimension Table)
- Course_Id (Pk)
- Name
- Department_Id (Fk)

Table Department (Dimension Table)
- Department_Id (Pk)
- Department Name
- Location

Table Student (Dimension Table)
- Student_Id (Pk)
- Student_Name
- Major
- Department_Id (Fk)
- Address_Id (Fk)

Table Fact (Fact Table)
- Course_Id (Fk)
- Instruction_Id (Fk)
- Student_Id (Fk)
- Semester_Id (Fk)
- Count
- Average_Grade

Table Instructor (Dimension Table)
- Instructor_Id (Pk)
- Instructor_Name
- Instructor Rank
- Instructor Course
- Email_Address
- Phone Number

Table Semester (Dimension Table)
- Semester_Id (Pk)
- Year
- Semester

Table Address (Dimension Table)
- Address_Id (Pk)
- State
- City
- Zip code

Q3. Based on current trends in technology need to design information systems . Explain the points to be taken care with respective traditional operational systems and the newer informational systems that need to be built? **[5]**

Answer: The fundamental reason for the inability to provide strategic information is that we have been trying all along to provide strategic information from the operational systems. These operational systems such as order processing, inventory control, claims processing, outpatient billing, and so on are not designed or intended to provide strategic information. If we need the ability to provide strategic information, we must get the information from altogether different types of systems. Only specially designed decision support systems or informational systems can provide strategic information.

We find that in order to provide strategic information we need to build informational systems that are different from the operational systems we have been building to run the basic business. It will be worthless to continue to dip into the operational systems for strategic information as we have been doing in the past. As companies face fiercer competition and businesses become more complex, continuing the past practices will only lead to disaster.

➢ *Watching the wheels of business turn*
➢ **Show me the top-selling products**
➢ **Show me the problem regions**
➢ **Tell me why (drill down)**
➢ **Let me see other data (drill across)**
➢ **Show the highest margins**
➢ **Alert me when a district sells below target**

	OPERATIONAL	INFORMATIONAL
Data Content	Current values	Archived, derived, summarized
Data Structure	Optimized for transactions	Optimized for complex queries
Access Frequency	High	Medium to low
Access Type	Read, update, delete	Read
Usage	Predictable, repetitive	Ad hoc, random, heuristic
Response Time	Sub-seconds	Several seconds to minutes
Users	Large number	Relatively small number

We need to design and build informational systems
† That serve different purposes
† Whose scopes are different
† Whose data content is different
† Where the data usage patterns are different
† Where the data access types are different

Q4. 2-D data pulled out from the data cube. 3M

Product ID	Location ID	Number Sold
1	1	10
1	3	6
2	1	5
2	2	22

Represent the above into 3-D format, focussing majorly on product-id and sales

Answer:

Product ID	Location ID			Total Sold
1	10	-	6	16
2	5	22	-	27
Total	15	22	6	43

Q5.(a)What is a OLAP cube? [5]
   Answer: An OLAP data cube is a representation of data in multiple dimensions, using facts and dimensions. It is characterized by the combination of information according to it's relationship. It can consist in a collection of 0 to many dimensions, representing specific data. There are five basic operation to perform on these kind of data cubes:
   ▪ Slicing
   ▪ Dicing
   ▪ Roll-Up
   ▪ Drill-Up and Drill-Down
   ▪ Pivoting
Q5 (b).Why is dimensional normalization not required? [5]

Answer: Dimensional normalization allows the database related problems to be solved. It is used to remove the redundant attributes that are used as de-normalized dimensions. Dimensions consist of sub-dimensions that are joined together in one. The dimensional normalization is not used due to the fact that it makes:

- Data structure more complex and due to which the performance can be degraded as it requires lots of joining of tables and keep the relations intact.
- The space is not utilized properly and use of more space is required.
- The query performance suffers when aggregating or retrieving many dimensional values. This requires proper analysis and making of operational reports are necessary

Q5 (c).Write two differences between ER Modeling and Dimensional Modeling?   [5]
     Answer: ER modeling is used for normalizing the OLTP database design. Dimensional modeling is used   for de-normalizing the ROLAP/MOLAP design. Eg for each

Q5 (d).Explain Data Warehouse paradigm of Bill Inmon and Ralph Kimball.   [10]
Answer:
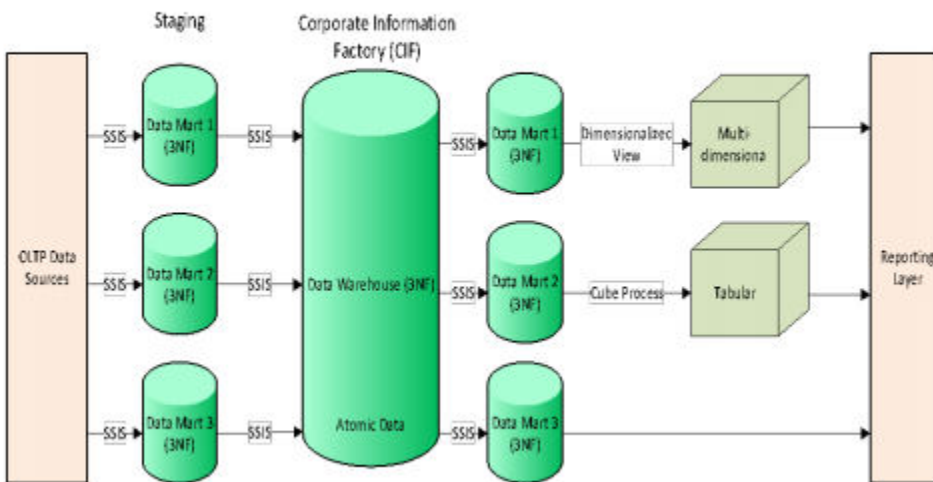
<u>Bill Inmon's paradigm:</u> Data warehouse is one part of the overall business intelligence system. An enterprise has one data warehouse, and data marts source their information from the data warehouse. In the data warehouse, information is stored in 3rd normal form.

Kimball Model

DW Bus Architecture

Ralph Kimball's paradigm: Data warehouse is the conglomerate of all data marts within the enterprise. Information is always stored in the dimensional model.
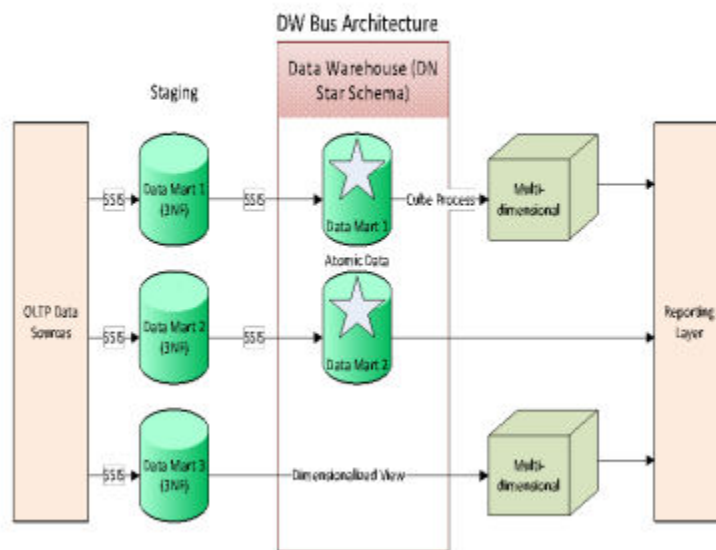
Q5 (e). What are the steps involved in creating dimensional modeling process?   **[5]**
Answer: The business process of the dimensional modeling includes:
(a) Choose The Business Process: In this, 4-step design method is followed that helps to provide the usability of the dimensional model. This allows the business process to be more systematic in representation and more helpful in explaining it as well. It includes the use of Business Process Modelling Notation (BPMN) or Unified Modelling Language (UML).

(b)Declaring The Grain: After choosing the business process, the declaration of the model comes that consists of grains. The grain of the model provides the accurate description of the dimensional model and allows the focus should be shifted there.

(c)Identify The Dimensions: In this phase, the identification of the dimension takes place in the dimensional model. The dimensions are defined in the grain that is defined in the declaration part above. Dimensions acts as a foundation of the fact table where the data gets collected that comes under the fact.

(d) Identify The Facts: Defining the dimensions provides a way to create a table in which the fact data can be stored. These facts are populated on the basis of the numerical figures and facts.

Q6. Describe Kimball Business lifecycle Model.
Answer

The Business Dimensional Lifecycle

```
                      ┌─────────────────────┐                    ┌──────────────────────┐
                      │     Technical       │───────────────────▶│  Product Selection & │──────────┐
                      │   Architecture      │                    │     Installation     │          │
                      │      Design         │                    └──────────────────────┘          │
                      └─────────────────────┘                                                       │
 ┌──────────┐  ┌──────────────┐                                                                     ▼
 │ Project  │◀▶│   Business   │  ┌─────────────┐   ┌───────────┐   ┌────────────┐   ┌────────────┐  ┌──────────┐
 │ Planning │  │ Requirements │─▶│ Dimensional │──▶│  Physical │──▶│    ETL     │──▶│ Deployment │  │  Growth  │
 └──────────┘  │              │  │  Modeling   │   │  Design   │   │  Design &  │   └────────────┘  └──────────┘
               │  Definition  │  └─────────────┘   └───────────┘   │ Development│                   ┌──────────────┐
               └──────────────┘                                    └────────────┘                   │ Maintenance  │
                                ┌─────────────┐   ┌─────────────┐                                   └──────────────┘
                                │     BI      │──▶│     BI      │
                                │ Application │   │ Application │
                                │Specification│   │ Development │
                                └─────────────┘   └─────────────┘
```

Project Management

Legend

| Plan | Build | Manage |