**DMW(ASSIGNMENT-4)**

Q-1: Explain different Backup & Recovery Model in data warehouse.

A-1: Following are different backup and recovery models:-

1. Full recovery Model:- It provides the most flexibility for recovering database to an earlier point in time.
2. Bulk logged recovery model:- Bulk logged recovery model provides higher performance than lower log space consumption for certain large scale options.
3. Simple recovery model:- Simple recovery model provides the highest performance and lowest log space consumption but with the significant exposure to data loss in the event of system failure.

Q-2: Describe the basic similarities and difference among ROLAP,MOLAP,HOLAP.

A-2: Similarities between ROLAP,MOLAP,HOLAP:-

These three OLAP servers are used to implement data warehouse and they are related to the logical model used to represent data.

Difference between ROLAP,MOLAP,HOLAP:

|  |  |  |  |
| --- | --- | --- | --- |
| **Basics** | **ROLAP** | **MOLAP** | **HOLAP** |
| Storage location for detail data. | Relational database. | Multidimensional  database. | Relational database. |
| Storage location for summary aggregations. | Relational database. | Multidimensional  database. | Multidimensional  database. |
| Storage space requirement. | Large | Medium | Small |
| Latency | Low | High | Medium |

Q-3: Describe the following in details:

1. Data Mining Interface
2. Testing of Data warehouse
3. A-3: a) **Data Mining Interface** provides the medium that allows users to communicate with data mining process. It is difficult to use data mining query languages. It consists of:-
4. Data Collection and data mining query composition allows user to specify task relevant data sets and to compose data mining queries.
5. Presentation of discovered patterns allows the display of discovered patterns in various forms like tables, graphs, charts and other visualization techniques.
6. Hierarchy of data mining primitives allows to do the specification of concept hierarchy, either manually or automatically.
7. Interactive Multilevel Mining allows roll-up or drill down operations on discovered patterns.

b) **Testing of Data Warehouse** is very important for data warehouse system to make them work correctly and efficiently.

1)Unit Testing: In unit testing, each component is separately tested , i.e. procedure,program,SQL Script,Unix shell is tested.

2)Integration Testing: In integration testing, the various modules of the application are brought together and then tested again the number of inputs.

3)System Testing: In system testing, the whole data warehouse applications is tested together. System testing is performed by the testing team.

Q-4: Explain how query performance can be improved by cascading the operation.

A-4: To speed up the query processing are constructed with following procedures.

1. Determining which operations should be performed on the available cuboids:
2. This involves transformations of operations specified in the query into the corresponding SQL &/or OLAP operators.
3. These operations including roll-up, drill-down, projection selection, etc.
4. Determining on which materialized cuboids the relevant operations should be applied: In this, all of the materialized cuboids are identified which may be useful for answering the query , pruning the relationship among the cuboids and selecting the cuboids with least cost.

Q-5: What is Web Mining? Differentiate between Web content mining , web structure & web usage mining.

A-5: 1) Web mining is an application of data mining techniques to find information pattern from the web data.

2) Web mining helps to improve the power of web search engine by identifying the web pages and classifying the web document.

3) There are three types of web mining:

a) Web content mining

b) Web usage mining

c) Web structure mining

Difference:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criterion** | **Web content mining** | | **Web structure mining** | **Web usage**  **mining** |
|  | IR view | DB view |  |  |
| View data | a)Unstructured  b) Structured | a)Semi-structure  b)Website as DB | a)Link Structure | a)Interactivity |
| Main data | a)Text document  b)Hypertext document | a)Hypertext document | a) Link Structure | a)Server logs  b)Browser logs |
| Method | a)Machine learning  b)Statistical | a)Association rules  b)Proprietary algorithms | a)Proprietary algorithms | a)Machine learning  b)Statistical  c)Association rules |
| Applications categories | a)Categorization  b)Clustering | a) Web site schema discovery | a)Categorization  b)Clustering | a)Site construction  b)Adaption and management |

Q-6:Compare and contrast Spatial, Temporal mining with relevant examples.

A-6:

|  |  |
| --- | --- |
| **Spatial Mining** | **Temporal Mining** |
| Spatial mining is the extraction of knowledge spatial relationships and interesting measures that are not explicitly stored in spatial database. | Temporal mining is the extraction of knowledge about occurrence of an event or values weather they follow cyclic , random, seasonal variations etc. |
| It deals with spatial (location, geo-referenced) data. | It deals with implicit or explicit temporal content from large quantities of data. |
| It includes finding characteristics rules, discriminant rules , association rules and evaluation rules etc. | It amount mining new to unknown knowledge , which takes into ocurent the temporal aspects of the data. |
| For example:- Determination hotspots, unusual locations. | For example:- An association rule which looks like- “ Any person who buys car also buys steering lock”. |

Q-7: What are the applications of data ware housing?

A-7: Applications of DW are:-

1)Airline: In the Airline system, it is used for operation purpose like crew assignment, analysis of route , profitability, frequent flyer program promotion etc.

2)Banking: It is widely used in the banking sector to manage the resources available on desk effectively. Few bank also used for the market research, performance analysis of the product and operations.

3)Healthcare: Healthcare sector also used data warehouse to strategize and predict outcomes, generate patient’s treatment reports, share data with tie in insurance companies, medical aid services etc.

4)Public Sector: In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax record, health policy record , etc.

5)Investment and insurance sector: In this sector, the warehouse are primarily used to analyze data patterns, customer trends and to track market movements.

6)Telecommunications: A data warehouse is used in the sector for product promotions , sales decisions and to make distribution decisions.

Q-8: Define the term Aggregation, Historical Information and query facility.

A-8: Aggregation:

* Data aggregation is a process in which information is gathered and expressed is a summary form, for purpose such as statistical analysis.
* Data aggregation may be performed manually or though specialized software.

Historical Information: A Brief History of Data Warehouse are designed to support the decision making process through data collection, consolidation, analytics and research. They can be used in analyzing the specific subject area, such as “sales” and are an important part of modern Business Intelligence.

Query Facility: Data querying is a process of asking questions of data in search of a specific answer. Unlike many forms of search (i.e. Google) queries are normally structured and require specific parameters or code known as SQL(Structured Query Language).