DWBI : Unit 5 Imp Questions

Q1. Explain different types of BI applications . (Explain 2 In Detail)

Ans.

The most important BI application types include the following:

1. Direct access queries: the classic ad hoc requests initiated by business users from desktop query tool applications.

2. Standard reports: regularly scheduled reports typically delivered via the BI portal or as spreadsheets or PDFs to an online library.

3. Analytic applications: applications containing powerful analysis algorithms in addition to normal database queries. Pre-built analytic applications packages include budgeting, forecasting, and business activity monitoring (BAM).

4. Dashboards and scorecards: multi-subject user interfaces showing key performance indicators (KPIs) textually and graphically.

5. Data mining and models: exploratory analysis of large "observation sets" usually downloaded from the data warehouse to data mining software. Data mining is also used to create the underlying models used by some analytic and operational BI applications.

6. Operational BI: real time or near real time queries of operational status, often accompanied by transaction write-back interfaces.

Q5. Explain Analytic Cycle for BI application

Ans.

Business analysis follows a common process from monitoring activity to identifying a problem or opportunity and determining an action to take, and finally back to monitoring the results of that action. The analytic cycle breaks the process into five distinct stages as below.

1)Stage 1: Monitor Activity

Standard reports are the starting point for the analytic cycle. Users work with these reports to examine current results versus previous periods or plan in order to provide a report card on the state of the business. BI application requirements in the monitor activity stage focus on the presentation layer and include technologies such as dashboards, portals, and scorecards.

2)Stage 2: Identify Exceptions

This stage focuses on the identification of "what's the matter?" or "where are the problems?" The emphasis in this stage is to identify the exceptions to normal performance, as well as the opportunities. The identify exceptions stage requires additional capabilities, such as distribution servers that distribute alerts to users' devices of choice based upon exception triggers and visualization tools to view the data in different more creative ways, including trend lines, spark lines, geographical maps, or clusters.

3)Stage 3: Determine Causal Factors

This stage tries to understand the "why" or root causes of the identified exceptions. Identifying reliable relationships and interactions between variables that drive exceptional performance is the key. Successfully supporting users' efforts in this stage will require your DW/BI system architecture to include additional software including statistical tools and/or data mining algorithms, such as association, sequencing, classification, and segmentation, to quantify cause-and-effect.

4)Stage 4: Model Alternatives

In this stage, you build on cause-and-effect relationships to develop models for evaluating decision alternatives. The data warehouse is a treasure trove of historical insight; many times, you can predict the effect of a potential decision by finding a similar situation in the warehouse's historical data. Your data warehouse architecture may also need to accommodate additional technologies in the model alternatives stage, including statistical tools and data mining algorithms for model evaluation, such as sensitivity analysis, Monte Carlo simulations, and goal seeking optimizations.

5)Stage 5: Take Action and Track Results

The results of the action should be captured and analysed in order to continuously fine tune the analysis process, business rules, and analytic models. This brings you right back to the monitor activity stage to start the cycle all over again. These stages occur in a circular process. The result of each analytic cycle becomes input to the next cycle. The results of one decision must be monitored, analysed for exceptions, and so on through the five stages

Q16. What is the need of following query formulation capabilities in Query Tools 1), 2), 3), 4)

Q6. What are query formulation capabilities required in query tool for Business Intelligent applications? (Explain 2 In Detail)

Ans:

1) Multipass or multiset queries. To calculate comparisons or correctly calculate non additive measures in report break rows, the query tool must break the report down into a number of simple queries that are processed separately by the RDBMS. Some tools formulate these queries as separate queries and combine the results based on a local join of the separate sets. Other tools create SELECT statements in the FROM clause for each separate query and combine the results as part of the query using an outer join.

Multipass queries also allow drilling across to different fact tables in several conformed business process dimensional models, potentially in different databases

2) Alerts As data volumes grow exponentially, the query tool needs to help the user identify records that stand out from the others, usually based on a comparative rule.They highlight exceptional entries that are too high or too low when compared to an expected value.

3) Successive constraints. The results of one query are used as a limit or filter on subsequent queries. This is a particularly important capability for creating behavioural study groups where you identify a cohort and examine its behaviour as a unit.

4) Semi-additive summations. There is an important class of numeric measures in common business fact tables that are not completely additive. Anything that is a measure of intensity is usually not additive across all dimensions, especially the date dimension. This feature makes all applications involving inventory levels, account balances, and other measures of intensity significantly simpler.

5) ANSI SQL 99 support. SQL 99 added basic OLAP capabilities to SQL, including a WINDOW construct that allows you to aggregate across a defined subset of a query to generate rolling averages, for example. Lots of interesting SQL92 capabilities, such as UNION, MINUS, and nested SELECT statements in various locations of a SELECT statement are supported by the database vendors, but are not fully supported by many BI tool vendors.

6) Direct query string entry. As a last resort, you will probably need to be able to view and alter the SQL or OLAP language generated by the tool. This includes creating complex queries and adding optimizer hints.

Q7. How Data Mining is used in DWBI Systems?

Q16. How DataMining is used for Estimation and Prediction, Anomaly Detection?

Q10. How Clustering and Affinity grouping is used for Datamining in DWBI Systems?

Ans:

1) Clustering

1. Clustering is a pure example of undirected data mining, where the user has no specific agenda and hopes that the data mining tool will reveal some meaningful structure.

2. An example of clustering is looking through a large number of initially undifferentiated customers and trying to see if they fall into natural groupings.

3. Clustering algorithms work well with all kinds of data, including categorical, numerical, and textual data.

4. It is not even necessary to identify inputs and outputs at the start of the job run. Usually, the only decision the user must make is to ask for a specific number of candidate clusters.

5. The clustering algorithm will find the best partitioning of all the customer records and will provide descriptions of the centroid of each cluster in terms of the user's original data.

6. Once the clustering model has been trained, you can use it to classify new cases by matching the new case to the "nearest" centroid. It often helps to first cluster customers based on their buying patterns and demographics.

2) Classifying

The classifier determines to which cluster centroid is nearest or most similar. Viewed in this way, clustering may well be a natural first step that is followed by classifying. Classifying in the most general sense is immensely useful in many data warehouse environments. A classification is a decision

Techniques that can be used for classifying include:

standard statistics,	memory-based reasoning,	genetic algorithms,
link analysis,	decision trees,	neural networks.

3) Affinity grouping

1. Affinity grouping is a special kind of clustering that identifies events or transactions that occur simultaneously.

2. A well-known example of affinity grouping is market basket analysis. Market basket analysis attempts to understand what items are sold together at the same time.

3. This is a hard problem from a data processing point of view because there are thousands of different products in a typical retail environment. It is pointless to enumerate all the combinations of items sold together because the list quickly reaches astronomical proportions.

4. The art of market basket analysis is to find the meaningful combinations of different levels in the item hierarchy that are sold together.

5. Specific techniques that can be used for affinity grouping include standard statistics, memory-based reasoning, link analysis, and special-purpose market basket analysis tools.

4) Estimating and Predicting

1. Estimating and predicting are two similar activities that normally yield a numerical measure as the result.

2. For example, you may find a set of existing customers who have the same profile as a candidate customer. From the set of existing customers, you may estimate the overall indebtedness of the candidate customer.

3. Prediction is the same as estimation, except that you are trying to determine a result that will occur in the future.

4. Specific techniques that can be used for estimating and predicting include standard statistics and neural networks for numerical variables, as well as the techniques described for classifying when predicting only a discrete outcome.

5) Anomaly detection

Anomaly detection can take advantage of any of the data mining algorithms. Anomaly detection involves a few extra twists in the data mining process. Often, it's necessary to bias the training set in favour of the exceptional events. Otherwise, there may be too few of them in the historical data for the algorithm to detect.

Q14 Explain need of following technical features for BI application and Query Tools 2), 3), 4), 5), 7), 8)

Q8. What are the technical features expected in Query Tool used in BI Systems? (Explain 2 In Detail)

Ans:

1) Multitasking - Users must be able to run other programs and create and run other queries while a query is running.

2) Cancel query - Users should be able to kill a single query they initiated while it is in process without killing all of them. This cancel should manage a clean break from the database server, and should not require rebooting the desktop machine.

3) Scripting - A scripting language and command line interface is critical for automating report execution.

4) Connectivity - Make sure you can get to all the database platforms desired. Connectivity includes connecting to other data sources — text, spreadsheets, XML files, other relational databases, and OLAP engines.

5) Scheduling - The tool needs to provide or take advantage of some kind of scheduling system. Users will want to defer queries for overnight processing or set them up for processing on a regular basis.

6) Metadata driven - The administrator should be able to define simple subsets of the warehouse, such as only those tables involved in a single business process dimensional model. This may include predefined join paths, column groupings, business descriptions, calculated columns, and pick list sources. This setup process should be simple and fast.

7) Software administration - This is a disappearing problem with the adoption of the web as an application platform. Until the transition is complete, make sure the vendor includes administration utilities that allow you to update any software, data models, local pick lists, and connectivity software from a central location.

8) Security - Ideally, the tool will participate in whatever user authentication system is available. Tool-based security is not that valuable in the DW/BI environment unless it participates with the network security service and the database. The BI tool may need to provide authorization functions, limiting users to only those reports or data subsets they are allowed to see.

9) Querying - Direct querying of the database should be supported without an administrative layer or with minimal work (i.e., initial setup of less than 10 minutes). This is especially valuable for the warehouse team because they are constantly examining new data sources, often on different platforms.

Q9. How Analytic applications are used in DWBI Systems?

Ans: Common analytic applications include:

• Promotion effectiveness - A promotion effectiveness analytic application might allow the user to investigate the impact of several variables on the response rate to a credit card promotion. The promotion analysis screen would include several panes, each one containing a graph or report.

• Web path analysis

• Affinity program analysis

• Shelf space planning

• Fraud detection

• Sales force management

• Category management

• Pre-Built Analytic Applications - Prebuilt analytic applications that focus on specific opportunity areas can incorporate domain expertise that would take years to develop in-house. Where your requirements line up with these offerings, it makes great sense to purchase and implement these pre-built apps. Other pre-built applications that focus more on basic reporting can be a good starting point, but in many cases are more work than you might expect

• Read / Write Analytic Applications - There is a separate class of analytic application that requires both read and write access. These applications include planning, budgeting, forecasting, and what-if modelling. These applications are essentially transaction systems, but often end up connected to the DW/BI system because the base level input data for these applications is fed from the data warehouse.

Q4. Explain Analytical and Presentation capabilities of Query Tools used in BI Systems

Q15. Explain following presentation and analytical Capabilities of query tool in BI applications (Explain 2 In Detail)

Ans:

2) Pivot the results Pivoting is the basis of multidimensional or cross tabulation analysis. The row-based results set that SQL generates almost always ends up being presented in a format with one or more dimensions displayed across the top of the report and one or more down the side. The pivot feature should be independent of the data source.

3) Drill down. This function allows users to add more detail to results set by including additional columns. These columns may be limited to a specific hierarchy or they may simply be attributes that support the analytic hypothesis under investigation.

4) Column calculations on pivot results. Some calculations, like showing one row value as a percentage of another row value, are useful. Share calculations and ratios rely on this capability.

5) Sorting. Sorting, especially by a non-displaying element, is important. For example, a financial report might show line items in a particular order that has nothing to do with the information displayed. Some tools allow you to physically specify the display order by dragging and dropping rows or columns.

6) Complex formatting. Formatting is often more important than it probably should be. You need to have a full range of graphic design tools, including lines, boxes, shading, images, fonts, and sizes, and you need to have pixel level control of the placement of report elements.

7) Compound documents. The dashboard interface is essentially a compound document made up of several individual reports, graphs, and highlighters. These all may be connected together by a shared parameter, like date, so changing the parameter will change the contents of all the sub-reports connected to it. In some cases, this linked set of sub-reports can be created in the portal tool.

8) User-changeable variables. User-changeable variables or parameters can be included anywhere in the query document, from the query filter to the report headings.

1) Basic calculations on the results set. This should include a range of math, statistical, string, sequential processing, conditional, and reporting functions. These calculations are often used to overcome other deficiencies in the tool. For example, it is possible to create a computed column using an IF or CASE statement that copies the description column if the rank ≤ 25 .

Q19. List and explain additional functionalities required for BI portal.

Ans:

1) Search- The search tool serves as an alternative report locator if the business process categories aren't adequate. A good search tool that indexes every document and page on the BI web site and all the report metadata, including report name and report description, can dramatically shorten the amount of time it takes a user to find what she wants.

2) Metadata browser- A metadata browser can be as simple as a few active web pages or reports that allow the user to browse through the metadata's descriptions of the databases, schemas, tables, columns, business rules, load statistics, report usage, and report content. In short, interested users can learn a lot about the DW/BI system through the metadata browser.

3) User forum- It may make sense to host a support-oriented discussion forum on the BI portal. This can be a good way for users to find help when they need it. It can also create a record of problems and their solutions for future reference. It takes a fairly large user community to generate the critical mass of activity needed to make a newsgroup successful. It also takes time and energy to moderate a user forum.

4) Personalization- Users should be able to save reports or report links to their personal pages, along with any parameter settings they've chosen. This personalization can be a powerful incentive for people to return to the portal every day.

5) Information centre- It helps keep things interesting to have new items appear on a regular basis. Offer a survey, have people register for tool training, or post a notice of an upcoming user forum meeting.