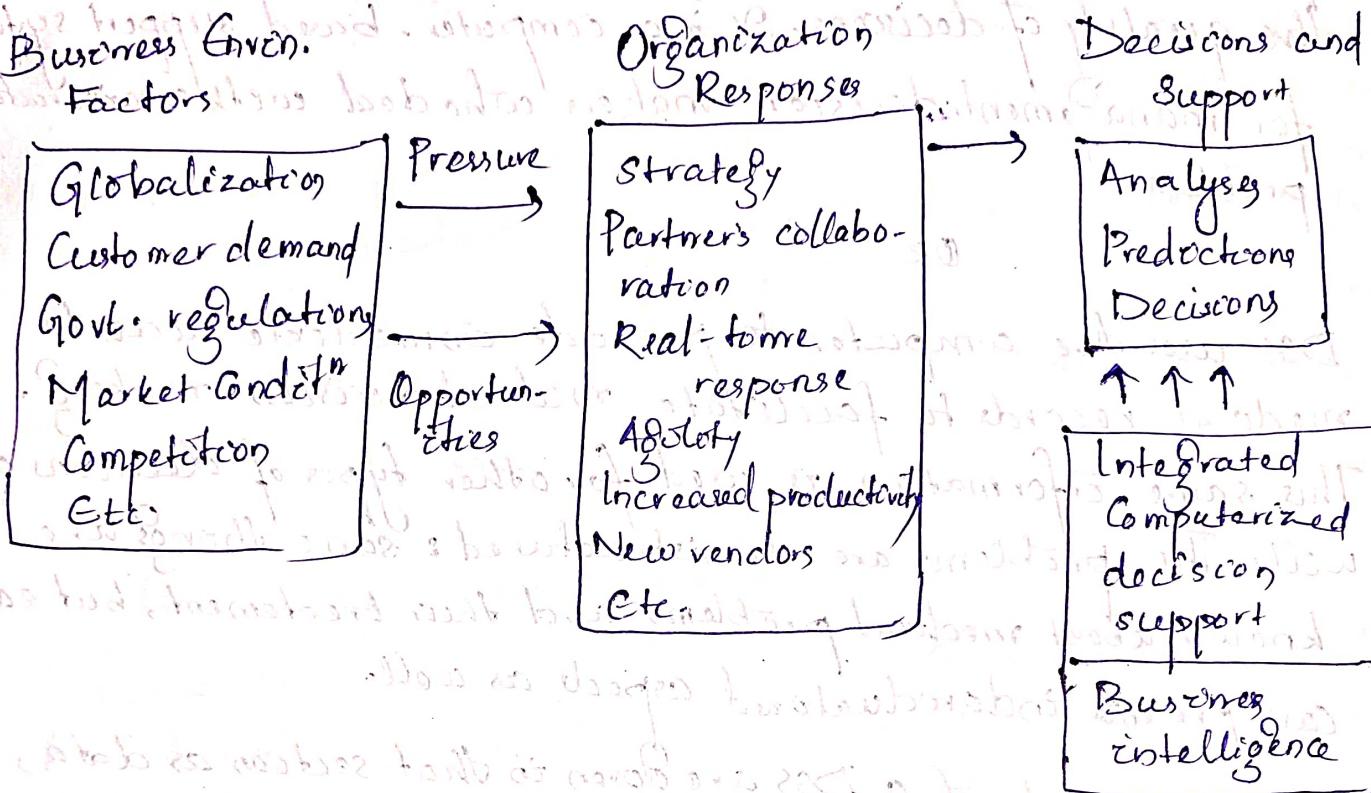


Q. List and Explain the components of Business Pressures -

Responses-Support model.

Ans. Business Given.



The components of the pressure-response-support model are business pressures, companies' responses to these pressures, and computerized support. The model suggests that responses are made to counter the pressures or to take advantage of opportunities, support facilitates monitoring the environment (e.g. for opportunities) and enhances the quality of the responses.

Q. Explain the DSS application and identify its components -
the data, models and user interface cited in the case.

Decision support system (DSS) couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems.

OR

DSS uses the computer to provide immediate access to medical records to facilitate medical decision making. This same information is used for other types of decisions as well. The problems are semistructured: some things are known about medical problems and their treatment, but each case presents individualized aspects as well.

The components of a DSS are given in that section as data, models, knowledge, users and the user interface. In this system those are as follows:

- * **Data**: Primarily the contents of the medical records. Secondly, information about available resources such as drug inventories, room availability, and operating room schedules.
- * **Models**: Financial planning. There are presumably also patient care models in the system, such as drug interaction models, but those are not given in the case.
- * **Knowledge**: Knowledge management to be an optional DSS component.
- * **Users**: "Physicians, nurses, administrators, and other employees"- in other words, just about everyone in the hospital except the patients themselves.

* **User Interface:** There are several user interfaces here, each designed for a different task or user community. Those mentioned include one to bring a new patient's medical records "on a screen in 15 seconds," one to add tests to a patient's medical record, a wireless tablet interface for doctor data access, entry and communication; three bedside electronic chart with both manual & automated inputs; and barcode scanners to track medications. In addition to these patient care system interfaces, there are interfaces to the financial planning and analysis package and to the communications and collaboration.

Q. Explain briefly descriptive analytics?

Descriptive analytics is the process of using current & historical data to identify trends and relationships.

Descriptive analytics breaks down into 5 steps, including

Descriptive or reporting analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences.

This involves consolidation of data sources & availability of all relevant data in a form that enables appropriate reporting and analysis. From this data infrastructure we

can develop appropriate reports, queries, alerts and trends using various reporting tools & techniques. Visualization becomes a key player in descriptive analytics.

Business intelligence and data mining

Business intelligence and data mining

Q. When developing a successful data warehouse, what are the most important risks & issues to consider and potentially avoid?

Data warehouse is a pool of data produced to support decision making; it is also a repository of current and historical data of potential interest to managers throughout the organization.

DW projects have many risks. Most of them are also found in other IT projects, but data warehousing risks are more serious because data warehouses are expensive, time-and-resource demanding, large-scale projects.

When developing a successful data warehouse, it is important to carefully consider various risks and avoid the following issues:

- Starting with the wrong sponsorship chain.
We need an executive sponsor who has influence over the necessary to support and invest in the data warehouse, an executive project driver who has earned the respect of other executives & flexible and also need an IS/IT manager to head up the project.
- Setting expectations that you cannot meet.

Every data warehousing project has two phases: Phase 1 is the selling phase to access the needed resources and phase 2 is the struggle to meet the expectations as described in phase 1.

- Engaging in politically naive behavior.

Do not simply state that a data warehouse will help managers make better decisions.

- Loading the data warehouse with information just because it is available.
- Believing that data warehousing database design is the same as transactional database design.
- Choosing a data warehouse manager who is technology oriented rather than user oriented.
- Focusing on traditional internal record-oriented data and ignoring the value of external data and of text, images and perhaps, sound and video.
- Delivering data with overlapping and confusing definitions.
- Believing promises of performance, capacity, and scalability.
- Believing that your problems are over when the data warehouse is up and running.
- Focusing on ad hoc data mining and periodic reporting instead of alerts.

The natural progression of information in a data warehouse is as follows:

- (1) Extract the data from legacy systems, cleanse them, and feed them to the warehouse.
 - (2) Support ad hoc reporting until you learn what people want.
 - (3) Convert the ad hoc reports into regularly scheduled reports.
- This process of learning what people want in order to provide it seems natural, but it is not optimal or even practical.

Q. What are the major differences between a traditional data warehouses and an RDW?

Ans:-

TDW

RDW

- It is used for strategic decisions. → used for strategic and tactical ones.
- The results of using a TDW → are measured by operational data. can be hard to measure.
- refresh rates range from → data must be up-to-the daily to monthly. minute.
- TDW summaries are often → must supply detailed data. appropriate
- Supports few concurrent → supports many (over a thousand) users.
- typically use restrictive reporting to conform or check reporting needs flexible, ad hoc patterns, often predefined summary tables.
- consists of power users, → used by operational staff, knowledge workers, managers, call centers, perhaps external other internal users.

Q. What are the recent technologies that may shape the future of data warehousing? Why?

Ans:- Sourcing - Acquisition of data from diverse and dispersed sources.

- Open source Software - Use of open source software tools is increasing at an unprecedented level in warehousing, business intelligence and data integration.

- SaaS (Software as a service) - SaaS is a creative way of deploying information systems applications where the provider licenses the application system to customers for use as a service on demand.
- Cloud Computing - Cloud computing is a where numerous hardware and software resources are pooled & virtualized, so that they can be freely allocated to applications and software platforms as resources are needed.
- Data warehouse appliances - Definition referred to a holistic solution by simply providing a whole technology stack for data warehousing.
- Infrastructure - Architectural - hardware & software - enhancements
- Real-time data warehousing - Real-time datawarehouse (RTW) systems can achieve near-real-time update of data, where the data latency typically is in the range of minutes to hours.
- Data management technologies and practices - The future holds strong growth for master data management (MDM). This is relatively new but extremely important concept is gaining popularity for many reasons.
- In-memory processing (64-bit computing) a super computing. For data warehousing, the most compelling benefit of 64-bit systems is the large space of addressable memory allowing the deployment of an in-memory database for reporting or analytic applications that need very fast query response.

- New DBMS - DBMS is the component of the platform where most work must be done to implement a data model & optimize it for query performance.
- Advanced analytics - Some users choose advanced analytic methods based on data mining, predictive analytics, statistics, artificial intelligence and so on. Still, the majority users choose SQL-based methods.

Q. Why is Web accessibility important with a data warehouse?

The Web accessibility of a data warehouse is important because many analysis applications are web-based, because users often access data over the Web and because data from the Web may feed the DW.

Q. Describe the 3 steps of the ETL process. Why is the ETL process so important for data warehousing efforts?

Ans ETL is a process of data integration that encompasses 3 steps - extraction, transformation, and loading.

Extraction

In the first step, extracted data sets come from a source into a staging area. The staging area acts as a buffer between the datawarehouse & the data source. Since, the data may be coming from multiple different sources and directly transferring the data to the warehouse. The staging area is used for data cleaning & organization.

Transformation

The data cleaning & organization stage is the next & transformation stage. All of that data from multiple

- **New DBMS** - DBMS is the component of the platform where most work must be done to implement a data model & optimize it for query performance.
- **Advanced analytics** - Some users choose advanced analytic methods based on data mining, predictive analytics, statistics, artificial intelligence and so on. Still, the majority uses classic SQL-based methods.

Q. Why is Web accessibility important with a data warehouse?

Web accessibility of a data warehouse is important because many analysis applications are web-based, because users often access data over the Web and because data from the Web may feed the DW.

Q. Describe the 3 steps of the ETL process. Why is the ETL process so important for data warehousing efforts?

Ans ETL is a process of data integration that encompasses 3 steps - extraction, transformation, and loading.

Extraction

In the first step, extracted data sets come from a source into a staging area. The staging area acts as a buffer between the datawarehouse & the data source data. Since, the data may be coming from multiple different sources and directly transferring the data to the warehouse. The staging area is used for data cleaning & organization.

Transformation

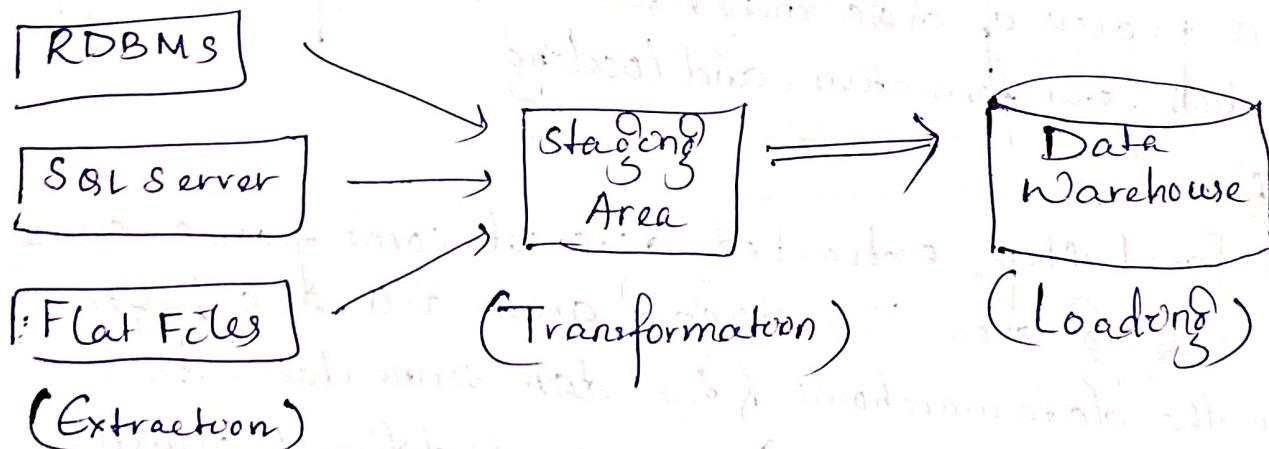
The data cleaning & organization stage is also called a transformation stage. All of that data from multiple

source systems will be normalized & converted to a single system format - improving data quality and compliance. ETL yields transformed data through these methods:

- cleaning
- filtering
- Joining
- Sorting
- Splicing
- Deduplication

>Loading

Finally, data that has been extracted to a staging area and transformed is loaded into data warehouse. The exact nature of the loading will depend upon the data source, ETL tools, and various other factors.



- * ETL is extremely important for data integration as well as for data warehousing.
- * The purpose of the ETL process is to load the warehouse with integrated and cleansed data.
- * The process of migrating data to a data warehouse involves the extraction of data from all relevant sources.

- * The process of loading data into a data warehouse can be performed either through data transformation tools that provide a GUI to aid in the development and maintenance of business rules or through more traditional methods, such as developing programs or utilities to load the data warehouse, using programming languages such as PL/SQL, C++, Java or .NET Framework languages.
- * Even though it is possible for programmers to develop software for ETL, it is common to use an existing ETL tool.
- * The following are some of the imp. criteria in selecting an ETL tool.
 - Ability to read from and write to an automated no. of data source architectures
 - Automatic capturing and delivery of metadata
 - A history of conforming to open standards
 - An easy-to-use interface for the developer & the functional user.

- Q. If a data mart can replace a data warehouse or complement it. Compare and discuss these options.?
- For a data mart to replace a data warehouse, it must make the DW unnecessary. This would mean that all the analyses for which DW would be used can instead be satisfied by a DM. If this so, it can be much less expensive, in terms of development and computer resources, to use multiple DMs instead of an overall DW.

⇒ In other situations, a data mart can be used for some analyses which would in its absence use the DWH, but not all of them. For those, the smaller DM is more efficient.

Q. Define Data mining. What factors have increased the popularity of data mining recently? [Ans. not asked]

Ans. Data mining is used to explore increasingly large databases and to improve market segmentation. By analysing the relationships between parameters such as customer age, gender, tastes, etc., it is possible to predict behaviour in order to direct personalized loyalty campaigns.

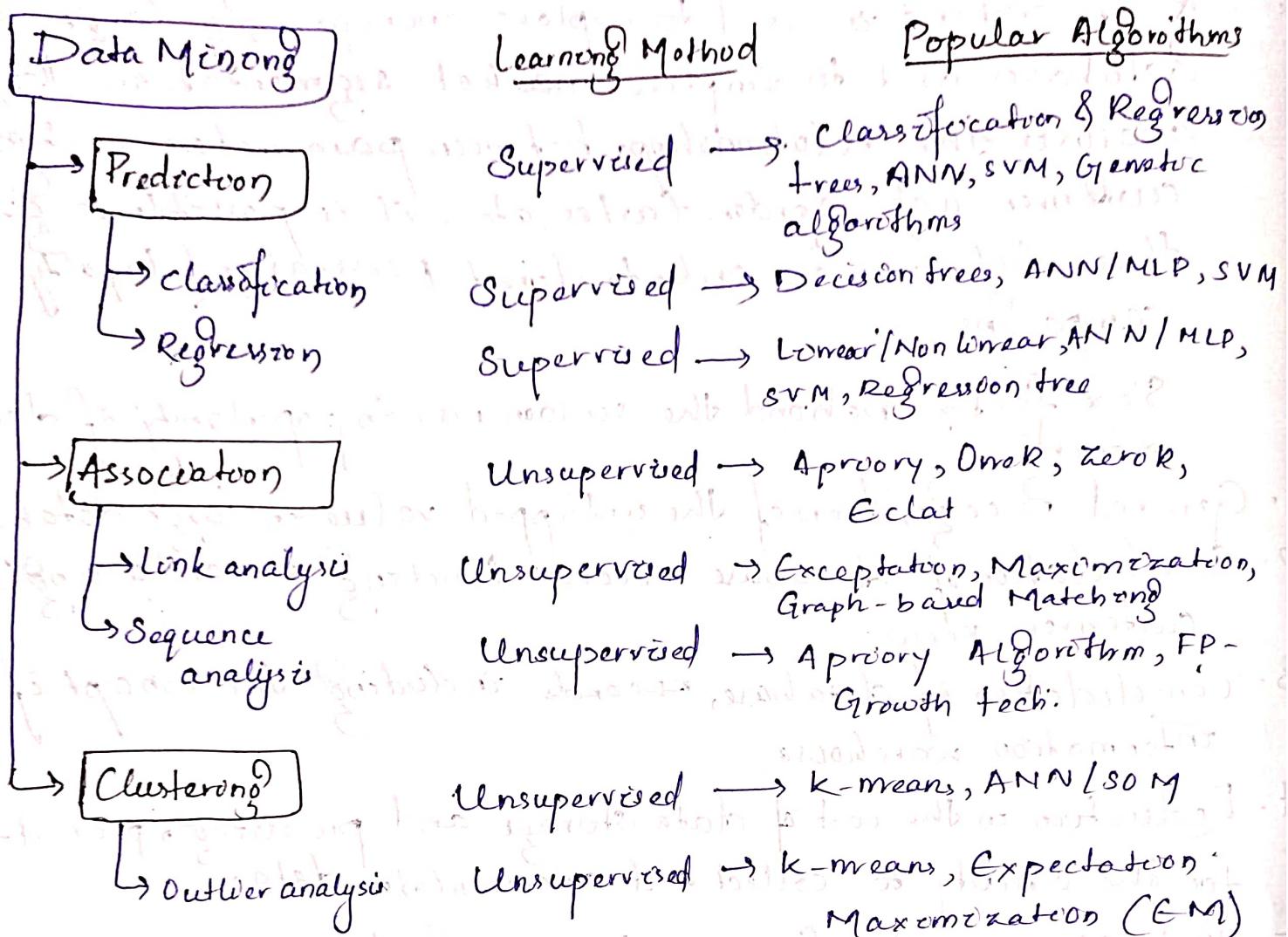
Six factors behind the sudden rise in popularity of data mining:

1. General recognition of the untapped value in large database;
2. Consolidation of database records tending toward a single customer view;
3. Consolidation of database records including the concept of an information warehouse;
4. Reduction in the cost of data storage and processing, providing for the ability to collect and accumulate data;
5. Intense competition for a customer's attention in an increasingly saturated marketplace;
6. The movement toward the ole-massification of business practices.

Some measure characteristics of data mining

- Automatic discovery of patterns
- Prediction of likely outcomes
- Creation of actionable information
- Focus on large data sets and databases

Q. What are some of the main categories of data mining technologies? Define at least 3 of them.



* Prediction

- This technique is generally used to predict the relationship that exists between both the independent & dependent variables as well as the independent variables alone.
- It can also use to predict profit that can be achieved in future depending on the sale.

* Clustering

- This technique is used to create meaningful object clusters that contain the same characteristics.
- Unlike classification that collect the obj. into predefined classes, clustering stores obj. in classes that are defined by it.

* Association

- It can be considered a method that can help us identify some interesting relations between different variables in large databases.
- It also helps to unpack some hidden patterns in the data, which can be used to identify the variables within the data.
- Helps in detecting the co-occurrence of different variables that appear very frequently in the dataset.

Q. Why is the popularity of text mining as a business intelligence tool increasing? Write down some popular application areas of text mining.

Ans. Text mining as a BI is increasing because of the rapid growth in text data & availability of sophisticated BI tools.

Some popular application areas of text mining

1. Information extraction: Finding key phrases and relationships in text by looking for predefined sequences via pattern

2. Topic tracking: Based on a user profile and documents that a user views, text mining can predict other documents of interest to the user

3. Summarisation: Summarising a document to save time for the reader

4. Categorization: Identifying main themes of a document, then putting the document into a predefined set of categories based on those themes.

5. clustering: Grouping documents without having predefined categories.