bit-byte-Data types-fields-record-files-database

DBMS - a software - organization, storage, management and retrieval of data

ddl,dml,dcl,tcl,

Key features:

* Data modeling
* Data storage and retrieval
* concurrent access
* Backup and recovery

DBMS minimizes

* Data redundancy
* Data inconsistency
* Data isolation
* Lack of flexibility
* Poor security
* Lack of data sharing and availability

DBMS maximizes

* Data security
* Data integrity
* Data independence

Database approach to Data management

(DBMS)

1. Relational DBMS:

* Data model
* Entity
* Record
* Attribute

Keys - primary, secondary, foreign

Operations - SELECT, JOIN, PROJECT

1. Object oriented DBMS
2. NoSQL DBMS
3. Cloud DBMS

Capabilities of DBMS

1. Data definition capability
2. Data dictionary
3. Data manipulation language
4. Report generation capability

How to design databases?

1. Designing databases - physical and logical
2. Design process identifies
3. Normalization

Cardinality symbols

* One-to-one
* One-to-many
* Many-to-many

1,2,3,bcnf

ER diagram

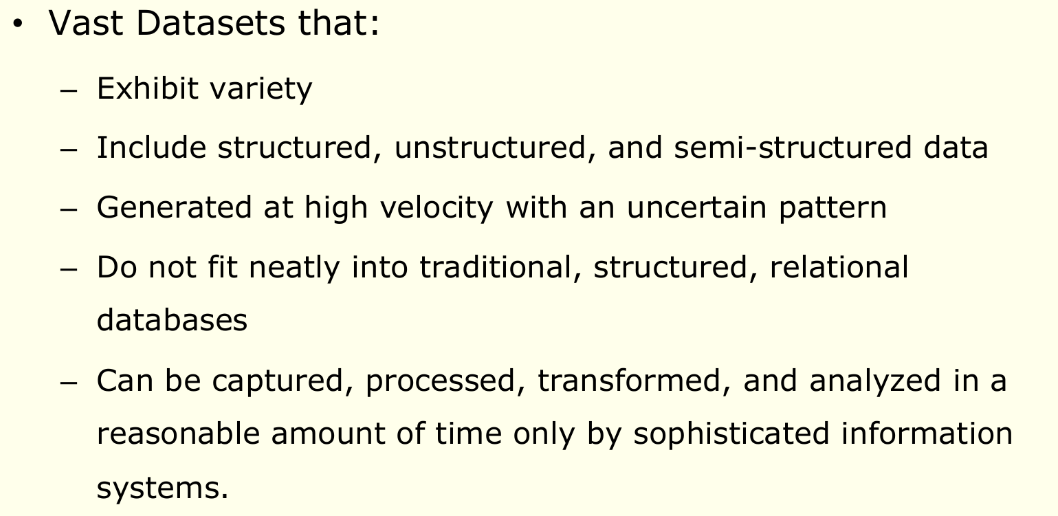
Difficulties of managing data:

1. Data Volume
2. Data variety
3. Data velocity
4. Data quality
5. Data security and privacy
6. Data integration
7. Data governance
8. Data backup and recovery
9. Data access control
10. Data interpretation

Data Governance

* Master data management
* Master data

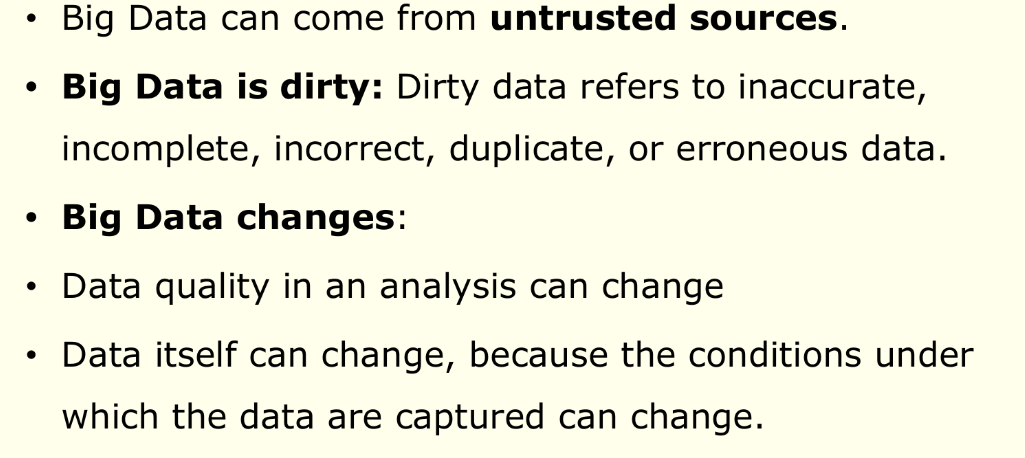
Big Data



Characteristics

* Volume
* Variety
* Velocity
* Veracity
* Value

Issues with big data



Managing big data

Big data can reveal valuable patterns and trends like tracking crime, spread of a disease or detecting fraud

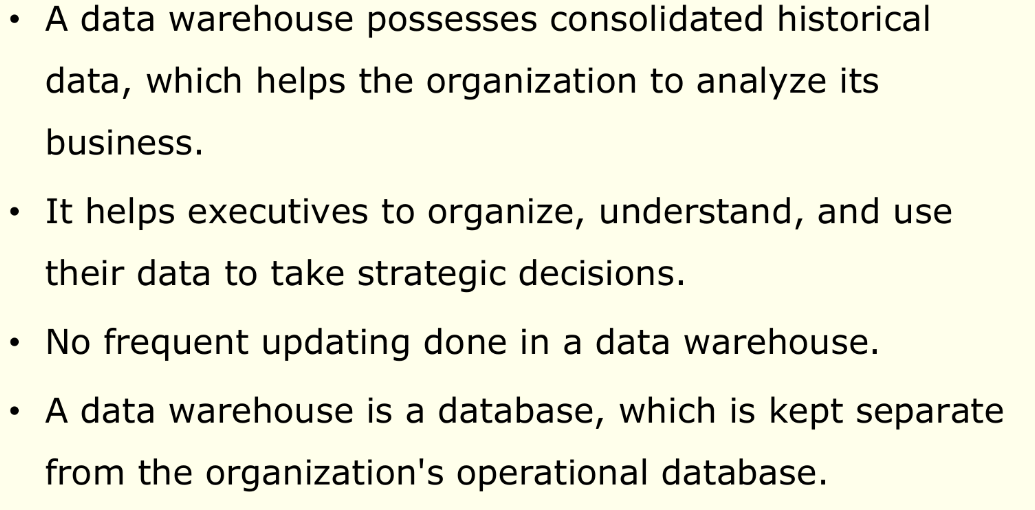
1. Define objectives and use cases
2. Data collection and ingestion
3. Data storage
4. Data cleansing
5. Data processing
6. Analytics
7. Visualization
8. Security
9. Inference

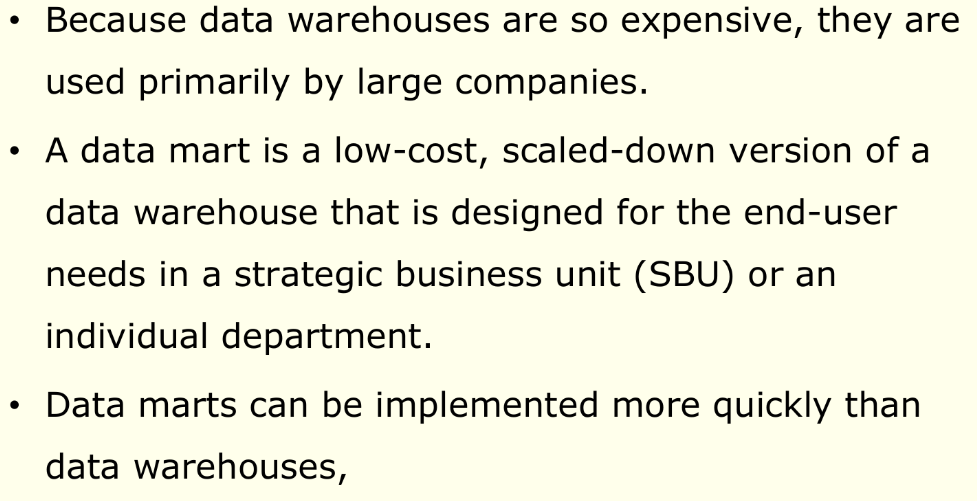
Putting Big Data to use

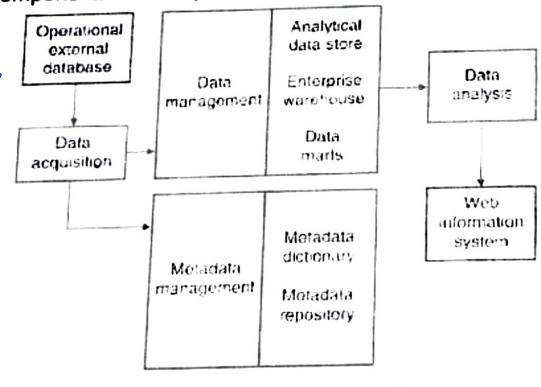
1. Making Big Data Available
2. Enabling Organizations to Conduct Experiments
3. Micro-Segmentation of Customers
4. Creating New Business Models
5. Organizations Can Analyze Far More Data

Data warehouse vs data marts

| **Characteristic** | **Data Warehouse** | **Data Mart** |
| --- | --- | --- |
| Scope | Large-scale, enterprise-wide data | Subset of data warehouse, department or specific business unit-focused. |
| Data Integration | Aggregates data from multiple sources and various departments. | Focuses on specific data for a particular department or business function. |
| Data Granularity | Typically contains detailed, historical, and summarized data. | Contains detailed or summarized data relevant to the specific business unit's needs. |
| Purpose | Provides a centralized repository for data analysis, reporting, and decision support for the entire organization. | Serves the needs of a specific group or department, enabling local analysis and reporting. |
| Schema Design | Utilizes a star or snowflake schema, which is often complex due to its comprehensive nature. | Employs a simplified star or snowflake schema, designed for specific requirements. |
| Data Volume | Contains a vast amount of data from across the organization, potentially petabytes of data. | Contains a subset of data, resulting in smaller data volumes compared to the data warehouse. |
| ETL Processing | Requires extensive ETL (Extract, Transform, Load) processes to consolidate, clean, and transform data from various sources. | May still require ETL processes but on a smaller scale due to a narrower focus. |
| Performance | Designed to handle complex queries and large datasets, optimized for analytical processing. | Typically provides faster query performance due to its smaller dataset and focused nature. |
| Cost and Complexity | Generally involves higher costs, both in terms of infrastructure and development effort, due to its enterprise-wide scope. | Tends to have lower implementation costs and development complexity compared to data warehouses. |
| Maintenance and Updates | Requires ongoing maintenance and updates to ensure data accuracy, consistency, and performance. | Relatively easier to maintain and update, with fewer data sources to manage. |
| Accessibility | Accessible to various departments and users, but data access might be subject to role-based access controls. | Primarily accessible to the specific business unit or department for which it was designed. |
| Data Governance | Data governance is crucial for maintaining data quality and consistency across the entire organization. | Data governance requirements are typically less complex, focusing on a smaller dataset. |
| Scalability | May require scaling to accommodate growing data needs and increasing concurrent users. | Typically easier to scale if additional data marts are created for different departments or functions. |

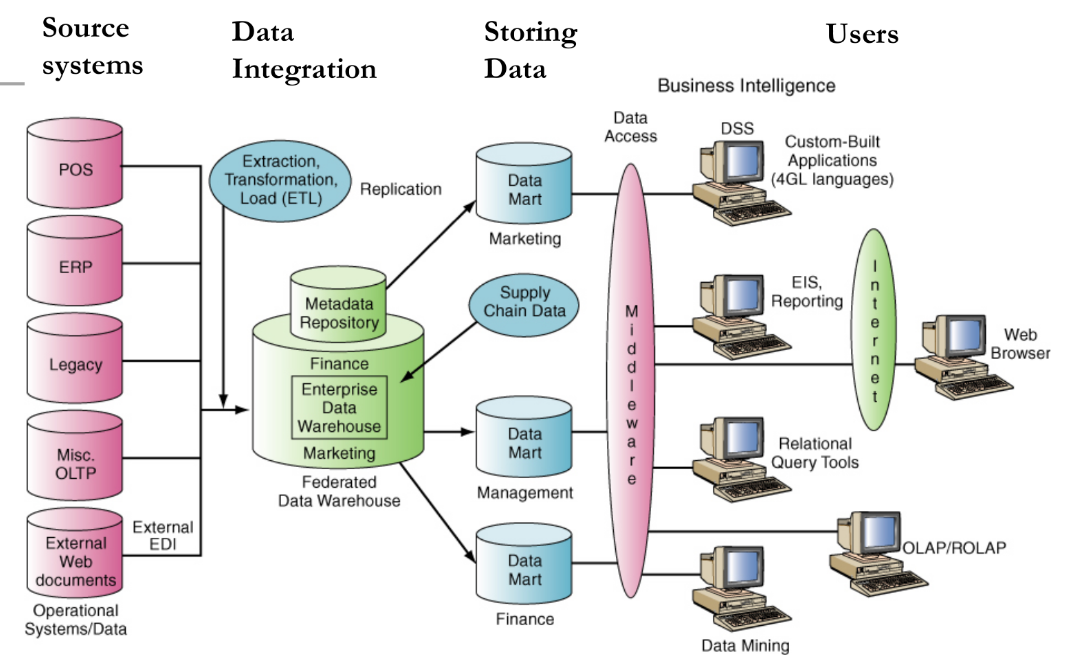


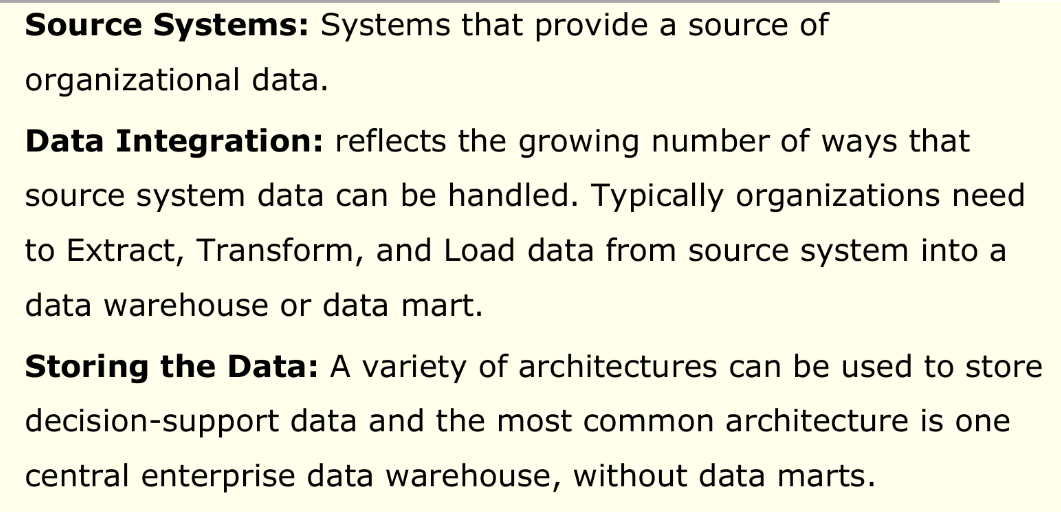


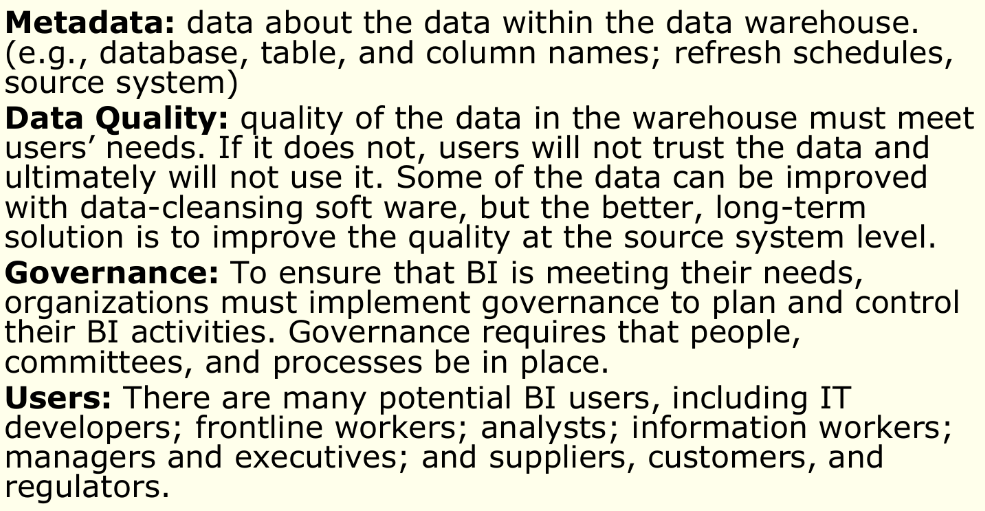


Characteristics of data warehouses/marts

1. Organized by business dimension
2. Use online analytical processing (OLAP)
3. Integrated
4. Time variant
5. Nonvolatile
6. Multidimensional



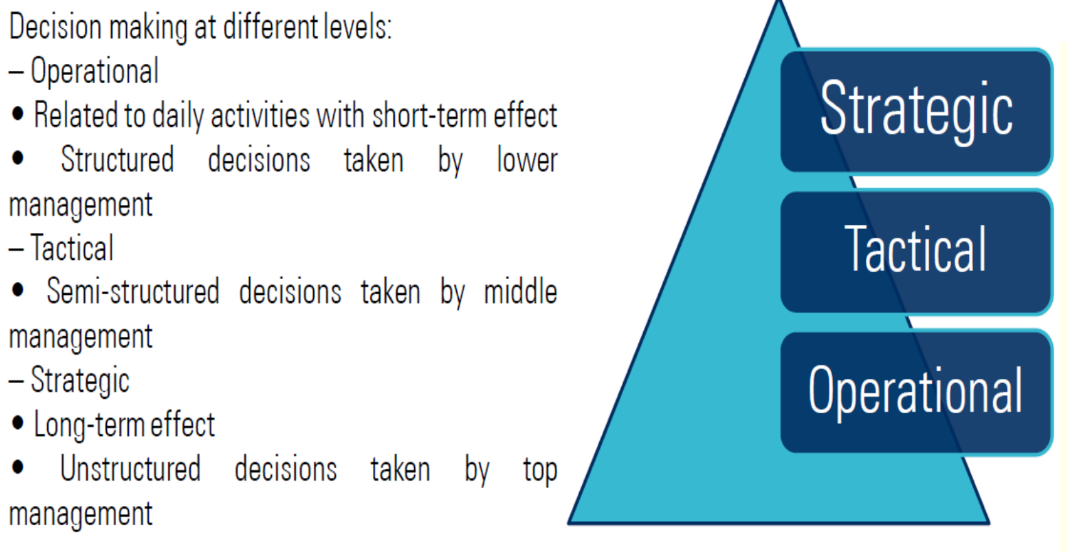


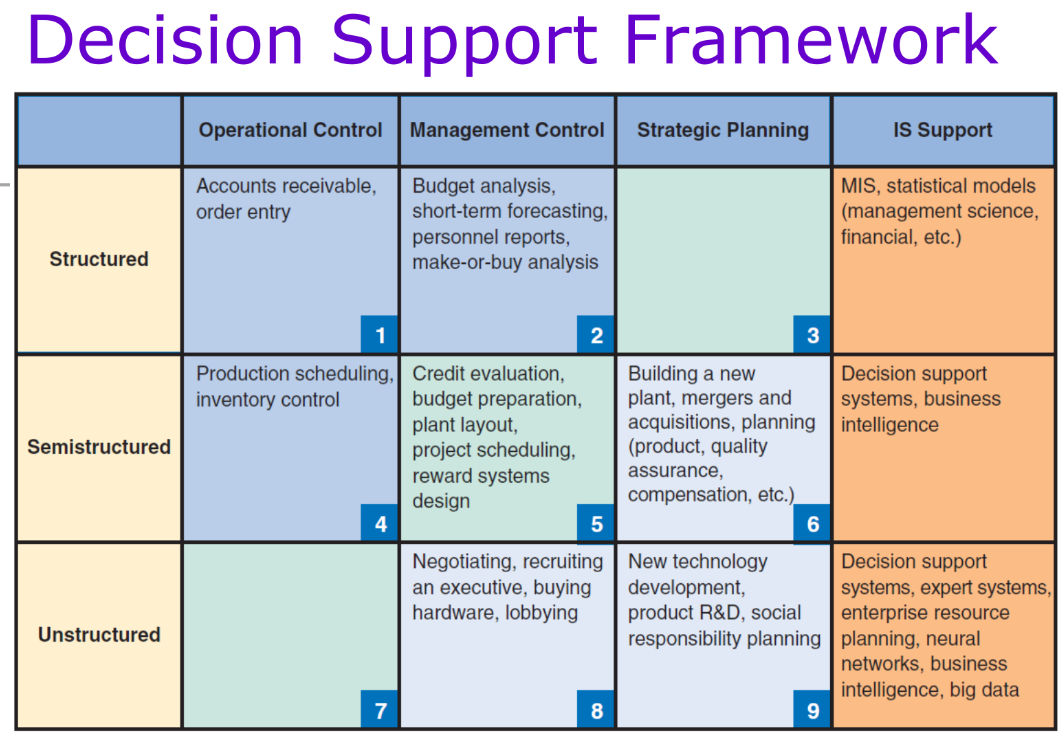


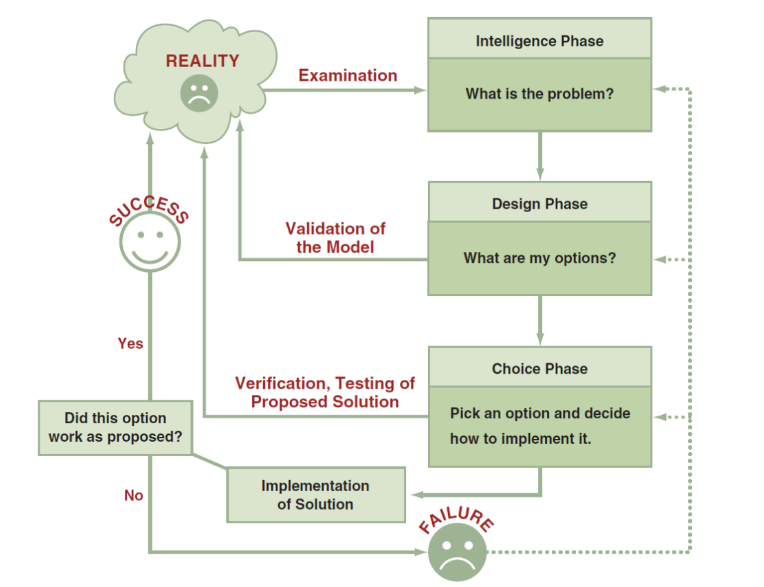
DECISION MAKING

Backbone - knowledge and information

* Problem identification
* Finding alternatives
* Making a choice



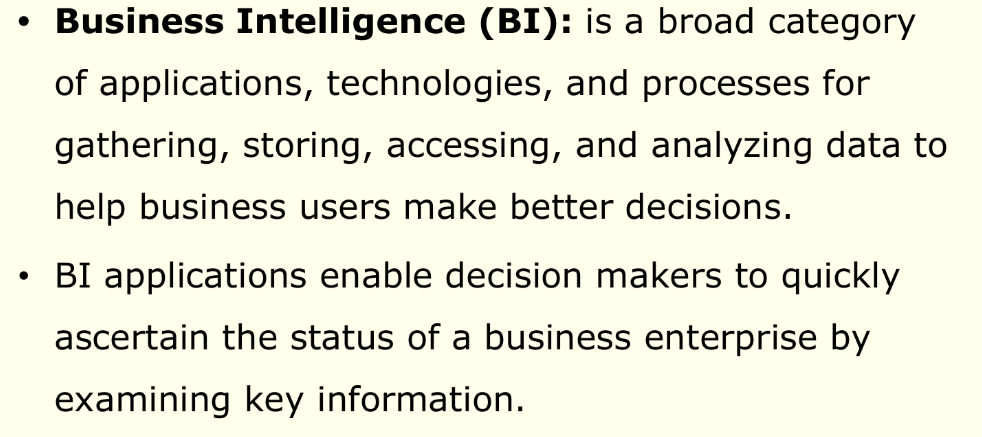


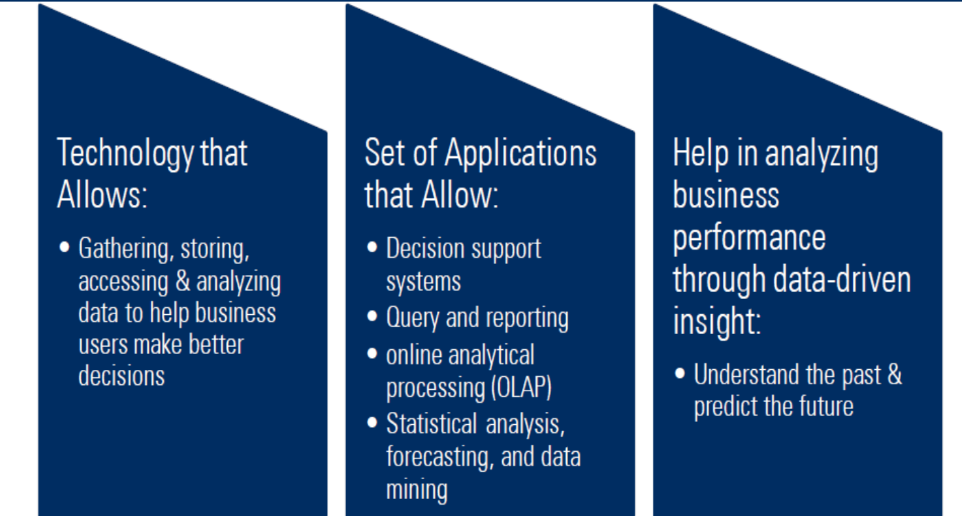


Phases

1. Intelligence phase
2. Design phase
3. Choice phase
4. Implementation phase

BUSINESS INTELLIGENCE





Scope for BI

* The Development of One or a Few Related BI Applications
* The Development of Infrastructure to Support Enterprisewide BI
* Support for the Organizational Transformation

Business Intelligence Applications for Data Analysis

1. Multidimensional Analysis or Online Analytical Processing

<https://www.geeksforgeeks.org/olap-operations-in-dbms/>

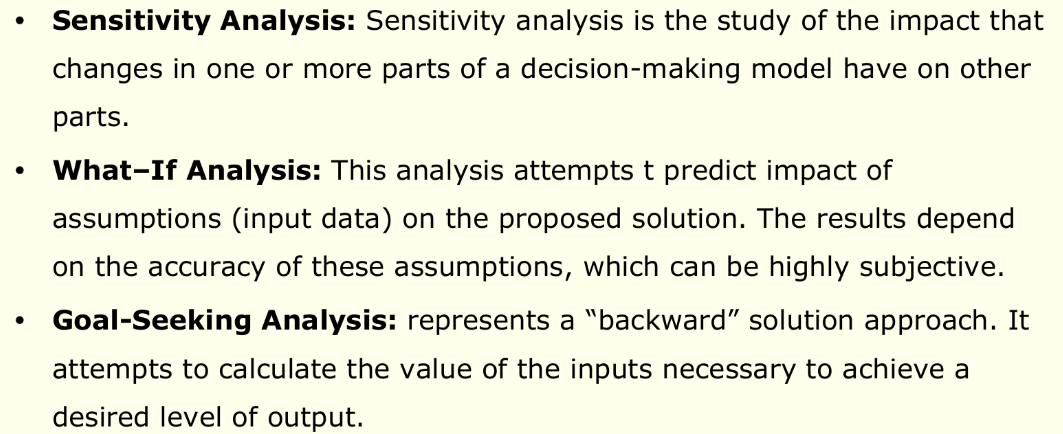
1. Data Mining

(1) predicting trends and behaviors

(2) identifying previously unknown patterns.

1. Decision Support Systems

Capabilities



Business Intelligence Applications for Presenting results

1. Dashboards
2. Data visualization

Technologies

* GIS
* Reality mining

1. Real time business intelligence