# MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

MRCET CAMPUS

(AUTONOMOUSINSTITUTION – UGC, GOVT. OF INDIA)


## Department of CSE

## (Emerging Technologies)

## (Data Science)

**B.TECH( R-22 Regulation)**

**(III YEAR–I SEM)**

**(2024-25)**

# Data Warehousing &Business Intelligence

## (R22A6609)

# LECTURE NOTES

**MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY**
**(AutonomousInstitution–UGC,Govt. of India)**
Recognizedunder2(f)and12(B)ofUGCACT1956
(Affiliated to JNTUH, Hyderabad, Approved by AICTE-Accredited by NBA & NAAC – 'A' Grade - ISO 9001:2015 Certified)
Maisammaguda,Dhulapally(PostVia.Hakimpet),Secunderabad–500100,TelanganaState,India

**Department of Computer Science and Engineering**

# EMERGING TECHNOLOGIES

## Data Warehousing & Business Intelligence
## (R22A6609)

## LECTURE NOTES

**Prepared by**
**Dr. M.V.Kamal, HOD, Professor**
**&**
**P. Sreenivas, Associate Professor**

**On**
**10.07.2024**

## Department of Computer Science and Engineering

# EMERGING TECHNOLOGIES

## Vision

- "To be at the forefront of Emerging Technologies and to evolve as a Centre of Excellence in Research, Learning and Consultancy to foster the students into globally competent professionals useful to the Society."

## Mission

***The department of CSE (Emerging Technologies) is committed to:***

- To offer highest Professional and Academic Standards in terms of Personal growth and satisfaction.
- Make the society as the hub of emerging technologies and thereby capture opportunities in new age technologies.
- To create a benchmark in the areas of Research, Education and Public Outreach.
- To provide students a platform where independent learning and scientific study are encouraged with emphasis on latest engineering techniques.

## QUALITY POLICY

- To pursue continual improvement of teaching learning process of Undergraduate and PostGraduateprogramsin Engineering &Managementvigorously.
- Toprovidestateofartinfrastructureandexpertisetoimpartthequalityeducationand research environment to students for a complete learning experiences.
- Developing students with a disciplined and integrated personality.
- To offer quality relevant and cost effective programmes to produce engineers as per requirements of the industry need.

**For more information: www.mrcet.ac.in**

## *MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY*

## *DEPARTMENT OF COMPUTER SCIENCE - ET*

**DATA WAREHOUSING AND BUSINESS INTELLIGENCE**

B.Tech. III Year I Sem. (DS) L/T/ P/ C

3/ 0/ 0/ 3

**Course Objectives**:

1. This course is concerned with extracting data from the information systems that deal with the day-to-day operations and transforming it into data that can be used by businesses to drive high-level decision making

2. Students will learn how to design and create a data warehouse, and how to utilize the process of extracting, transforming, and loading (ETL) data into data warehouses.

UNIT - I : Data Warehouse, Data Warehouse Modelling, OLAP operations, Data Cube Computation methods ,ETL Operations in Data Warehouse.

UNIT - II : Business Intelligence Introduction – Definition, Leveraging Data and Knowledge for BI, BI Components, BI Dimensions, Information Hierarchy, Business Intelligence and Business Analytics. BI Life Cycle. Data for BI - Data Issues and Data Quality for BI.

UNIT – III: BI Implementation - Key Drivers, Key Performance Indicators and Performance Metrics, BI Architecture/Framework, Best Practices, Business Decision Making, Styles of BI-vent-Driven alerts-A cyclic process of Intelligence Creation. The value of Business Intelligence-Value driven and Information use.

UNIT - IV : Advanced BI – Big Data and BI, Social Networks, Mobile BI, emerging trends, Description of different BI-Tools (Pentaho, KNIME)

UNIT – V: Business Intelligence and integration implementation-connecting in BI systems- Issues of legality Privacy and ethics- Social networking and BI.

TEXT BOOKS:

1. Data Mining – Concepts and Techniques - JIAWEI HAN & MICHELINE KAMBER, Elsevier, 4th Edition.

2. Rajiv Sabherwal “Business Intelligence” Wiley Publications, 2012.

REFERENCE BOOKS:

1. Efraim Turban, Ramesh Sharda, Jay Aronson, David King, Decision Support and Business Intelligence Systems, 9th Edition, Pearson Education, 2009.

2. David Loshin, Business Intelligence - The Savy Manager's Guide Getting Onboard with Emerging IT, Morgan Kaufmann Publishers, 2009.

3. Philo Janus, Stacia Misner, Building Integrated Business Intelligence. Solutions with SQL Server, 2008 R2 & Office 2010, TMH, 2011. R22 B.Tech. CSE (AI and ML) Syllabus JNTU Hyderabad Page 97 of 147

4. Business Intelligence Data Mining and Optimization for decision making [Author: Carlo-Verellis] [Publication: (Wiley)]

5. Data Warehousing, Data Mining & OLAP- Alex Berson and Stephen J. Smith- Tata McGrawHill Edition, Tenth reprint 2007

6. Building the Data Warehouse- W. H. Inmon, Wiley Dreamtech India Pvt. Ltd.

7. Data Mining Introductory and Advanced topics – Margaret H Dunham, PEA.

## MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

### *DEPARTMENT OF COMPUTER SCIENCE - ET*

**INDEX**

| | | | |
|---|---|---|---|
| 13 | III | Business Decision Making | 41 |
| 14 | III | A cyclic process of Intelligence Creation | 41 |
| 15 | III | The value of Business Intelligence-Value driven and Information use. | 44 |
| 16 | IV | Advanced BI –Mobile BI, | 47 |
| 17 | IV | Social Networks | 49 |
| 20 | IV | Description of different BI-Tools (Pentaho, KNIME) | 50 |
| **21** | V | Business Intelligence and integration implementation | 52 |
| **22** | V | Issues of legality Privacy and ethics- | 54 |
| **23** | V | Social networking and BI. | 56 |

**MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY**

**Department of emerging technology**

## UNIT-I

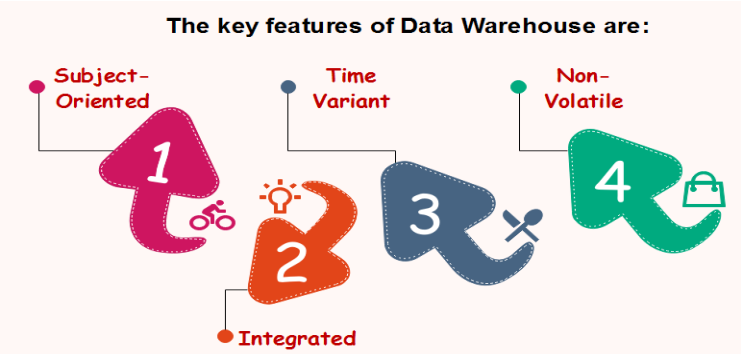
### Data Warehouse Definition :

A Data Warehouse (DW) is a relational database that is designed for query and analysis rather than transaction processing. It includes historical data derived from transaction data from single and multiple sources.

A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision-makers for data modeling and analysis.

A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.

It is not used for daily operations and transaction processing but used for making decisions



**Subject-Oriented :**

A data warehouse target on the modeling and analysis of data for decision-makers. Therefore, data warehouses typically provide a concise and straightforward view around a particular subject, such as customer, product, or sales, instead of the global organization's ongoing operations. This is done by excluding data that are not useful concerning the subject and including all data needed by the users to understand the subject.

**Integrated :**

A data warehouse integrates various heterogeneous data sources like RDBMS, flat files, and online transaction records. It requires performing data cleaning and integration during data warehousing to ensure consistency in naming conventions, attributes types, etc., among different data sources

**Time-Variant :**

Historical information is kept in a data warehouse. For example, one can retrieve files from 3 months, 6 months, 12 months, or even previous data from a data warehouse. These variations with a transactions system, where often only the most current file is kept

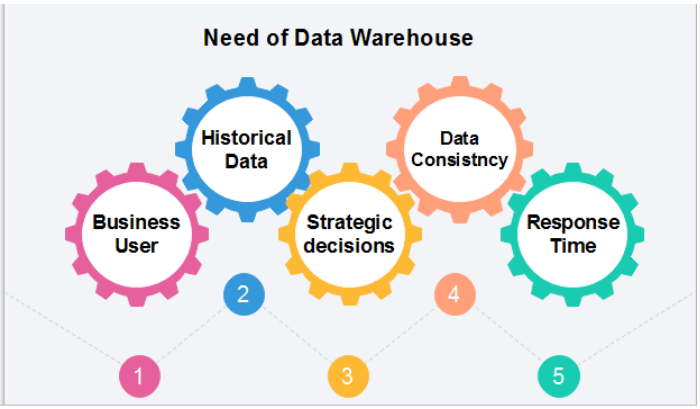
**Non-Volatile:**

The data warehouse is a physically separate data storage, which is transformed from the source operational RDBMS. The operational updates of data do not occur in the data warehouse, i.e., update, insert, and delete operations are not performed. It usually requires only two procedures in data accessing: Initial loading of data and access to data.

**Benefits of Data Warehouse :**

1. Understand business trends and make better forecasting decisions.
2. Data Warehouses are designed to perform well enormous amounts of data.
3. The structure of data warehouses is more accessible for end-users to navigate, understand, and query.
4. Queries that would be complex in many normalized databases could be easier to build and maintain in data warehouses.
5. Data warehousing is an efficient method to manage demand for lots of information from lots of users.
6. Data warehousing provide the capabilities to analyze a large amount of historical data.

## Need for Data Warehouse :




1. **Business User:** Business users require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.

2. **Store historical data:** Data Warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.

3. **Make strategic decisions:** Some strategies may be depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.

4. **For data consistency and quality:** Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.

5. **High response time:** Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time

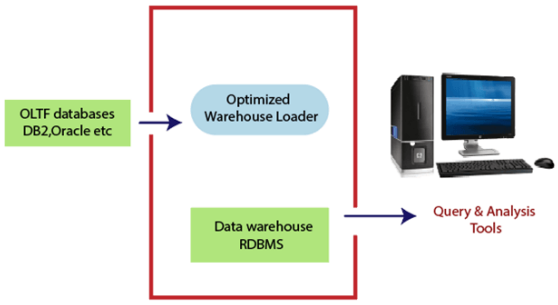
## Data Warehouse Modeling :

Data warehouse modeling is the process of designing the schemas of the detailed and summarized information of the data warehouse. The goal of data warehouse modeling is to develop a schema describing the reality, or at least a part of the fact, which the data warehouse is needed to support.

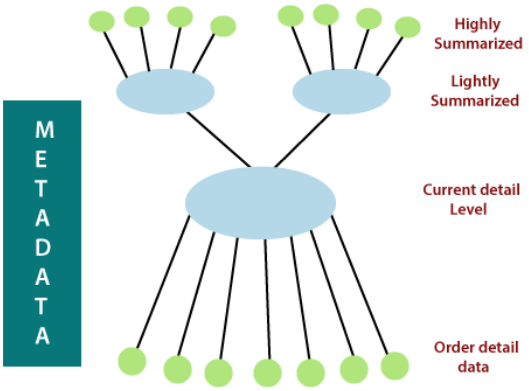
Data warehouse modeling is an essential stage of building a data warehouse for two main reasons. Firstly, through the schema, data warehouse clients can visualize the relationships among the warehouse data, to use them with greater ease. Secondly, a well-designed schema allows an effective data warehouse structure to emerge, to help decrease the cost of implementing the warehouse and improve the efficiency of using it.

Data modeling in data warehouses is different from data modeling in operational database systems. The primary function of data warehouses is to support DSS processes. Thus, the objective of data warehouse modeling is to make the data warehouse efficiently support complex queries on long term information.

In contrast, data modeling in operational database systems targets efficiently supporting simple transactions in the database such as retrieving, inserting, deleting, and changing data. Moreover, data warehouses are designed for the customer with general information knowledge about the enterprise, whereas operational database systems are more oriented toward use by software specialists for creating distinct applications.




**Data Warehouse Model**




**The Structure of data inside the data warehouse**

- Reflects the most current happenings, which are commonly the most stimulating.
- It is numerous as it is saved at the lowest method of the Granularity.
- It is always (almost) saved on disk storage, which is fast to access but expensive and difficult to manage.

**Older detail data** is stored in some form of mass storage, and it is infrequently accessed and kept at a level detail consistent with current detailed data.

**Lightly summarized data** is data extract from the low level of detail found at the current, detailed level and usually is stored on disk storage. When building the data warehouse have

to remember what unit of time is summarization done over and also the components or what attributes the summarized data will contain.

**Highly summarized data** is compact and directly available and can even be found outside the warehouse.

## OLAP operations :

**OLAP** stands for **Online Analytical Processing** (OLAP) could be a innovation that's utilized to organize expansive business databases and back business intelligence. OLAP databases are separated into one or more cubes, and each cube is organized and designed by a cube administrator to fit the way simply recover and analyze data so that it is less demanding to form and utilize the PivotTable reports and PivotChart reports that and just require.

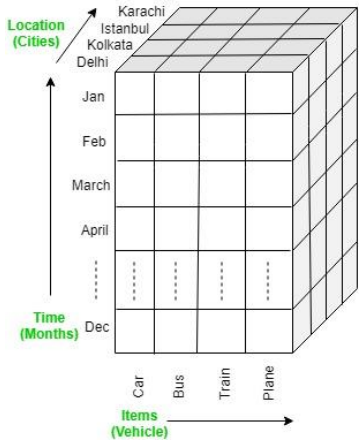
**OLAP operations:**

There are five basic analytical operations that can be performed on an OLAP cube:

**Drill down:** In drill-down operation, the less detailed data is converted into highly detailed data. It can be done by:

- Moving down in the concept hierarchy
- Adding a new dimension

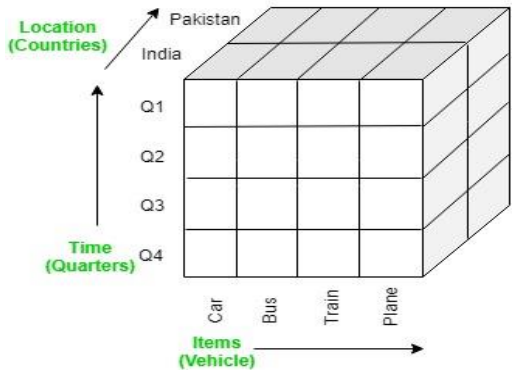
In the cube given in overview section, the drill down operation is performed by moving down in the concept hierarchy of *Time* dimension (Quarter -> Month).




**Roll up:** It is just opposite of the drill-down operation. It performs aggregation on the OLAP cube. It can be done by:

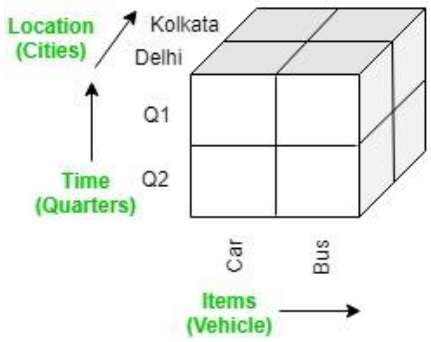
- Climbing up in the concept hierarchy
- Reducing the dimensions

In the cube given in the overview section, the roll-up operation is performed by climbing up in the concept hierarchy of *Location* dimension (City -> Country)

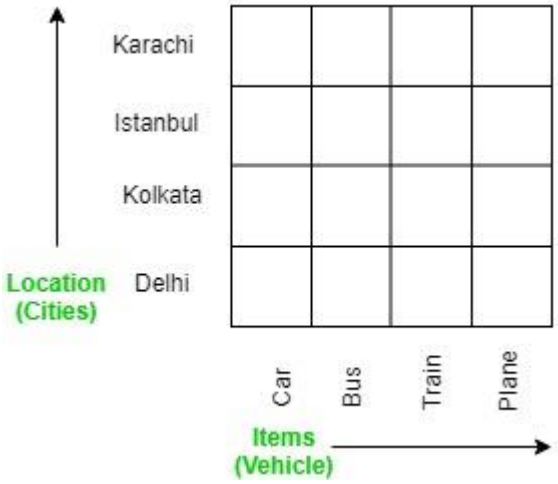



**Dice:** It selects a sub-cube from the OLAP cube by selecting two or more dimensions. In the cube given in the overview section, a sub-cube is selected by selecting following dimensions with criteria:

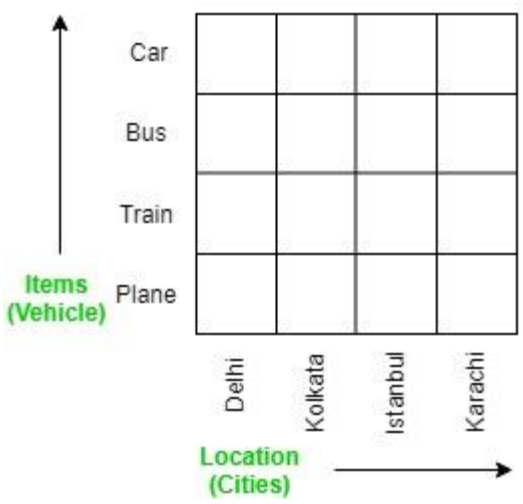
- Location = “Delhi” or “Kolkata”
- Time = “Q1” or “Q2”
- Item = “Car” or “Bus”




**Slice:** It selects a single dimension from the OLAP cube which results in a new sub-cube creation. In the cube given in the overview section, Slice is performed on the dimension Time = "Q1"




**Pivot:** It is also known as *rotation* operation as it rotates the current view to get a new view of the representation. In the sub-cube obtained after the slice operation, performing pivot operation gives a new view of it.

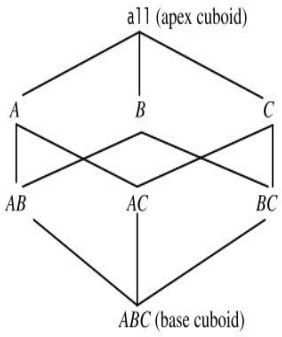



## Data Cube Computation methods :

### 1) Materialization of Cube: Full, Iceberg, Closed and Shell Cubes

A three-dimensional data cube with dimensions A, B, and C and an aggregate measure M. And can think of a data cube as a lattice of cuboids. Each cube is meant to symbolize a group. The fundamental cuboid, encompassing all three dimensions, is the ABC.

The aggregate measure (M) is calculated for each permutation of the three dimensions. There are six different cuboids that make up a data cube, with the base cuboid being the most specific. The apex cuboid is the most generalized cuboid. It stores a single number - the sum of all the tuples' measures in the base cuboid's measure M. From the topmost cuboid of the data cube example, we can descend into the lattice to access deeper levels of information




**Base Cell In Data Cube**

The term "*base cell*" refers to a cell in the base cuboid. Aggregate cells are cells that are not based on a cube. Each dimension that is aggregated in an aggregate cell is represented by a "" in the cell notation. Let's pretend we're working with an n-dimensional data cube.

Let each cell of the cuboids that make up the data cube be denoted by a = (a1, a2,..., an, measurements). If there are m (m n) values of a, b, c, d, e, f, g, h, I j, k, l, m, n, and o that are not "," then we say that an is an m-dimensional cell (that is, from an m- A is a base cell if and only if m = n; otherwise, it is an aggregate cell (where m n).

**Cuboids In Data Cube**

As an intriguing compromise between storage requirements and response times for OLAP, data cubes that are partially materialized are a viable option. We cannot compute the entire data cube, but rather only parts of it, called cuboids, each of which is composed of a subset of the cells in the full cube.

**Minimum Support And Iceberg In Data Cube**

Not only does this result in a more effective utilization of resources (namely, CPU time and disc space), but it also makes it possible to conduct more accurate analysis. There is a good chance that the non-passing cells are not important enough to warrant further investigation.

Cubes that only partially materialize are referred to as iceberg cubes, and this phrase is used to characterize such cubes. The term "minimum support," also abbreviated as "min sup" for short, describes the criteria that are the absolute minimum acceptable. It is common practise to refer to the effect of materializing only a fraction of the cells in a data cube as the "tip of the iceberg." In this context, "iceberg" refers to the entire cube including all cells.

**2) Roll-up/Drill-down-** This method involves aggregating data along one or more dimensions to create a summary of the dataset. It can be used to drill-down into specific areas of interest within the data. Roll-up/Drill-down is useful for quickly summarizing large datasets into manageable chunks while still maintaining important information about each dimension.

For example, if and have sales data for multiple products across several regions, and could use roll-up/drill-down to see total sales across all regions or drill-down into sales numbers for one particular product in one region.

**3) Slice-and-Dice -** This method involves selecting subsets of data based on certain criteria and then analyzing it using different dimensions. It is useful for identifying patterns that may not be immediately apparent when looking at the entire dataset.

Slice-and-Dice allows users to select subsets of data based on specific criteria such as time period or customer demographics which can then be analyzed using different dimensions like product categories or geographic locations. This helps identify patterns that may not be immediately apparent when looking at the entire dataset.

**4) Grouping Sets -** This method involves grouping data by multiple dimensions at once, allowing for more complex analysis of the dataset.Grouping Sets are useful when analyzing large datasets with multiple dimensions where users want to group by two or more dimensions at once. For example, grouping sets could show total revenue broken down by both product category and region simultaneously.

**5) Online Analytical Processing (OLAP) -** This method uses a multidimensional database to store and analyze large amounts of data. It allows for quick querying and analysis of the data in different ways.OLAP databases are specifically designed for analyzing large amounts of multi-dimensional data quickly through pre-aggregated values stored in memory making it ideal for real-time decision-making scenarios like stock market analysis.

**6) SQL Queries -** SQL queries can be used to compute data cubes by selecting specific columns and aggregating them based on certain criteria. This is a flexible method that can be customized based on the needs of the user.

## ETL Operations in Data Warehouse :

ETL stands for Extract, Transform, Load and it is a process used in data warehousing to extract data from various sources, transform it into a format suitable for loading into a data warehouse, and then load it into the warehouse. The process of ETL can be broken down into the following three stages:

1. **Extract**: The first stage in the ETL process is to extract data from various sources such as transactional systems, spreadsheets, and flat files. This step involves reading data from the source systems and storing it in a staging area.

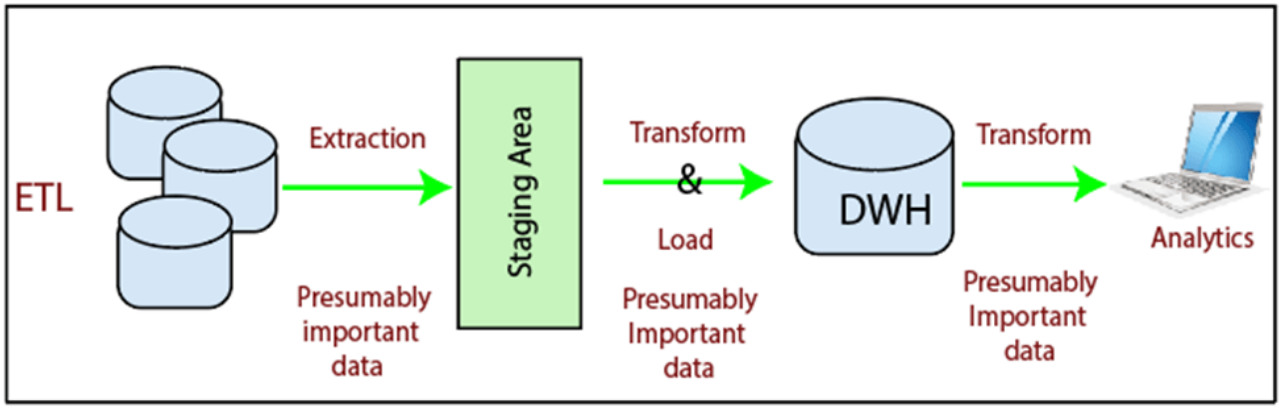
2. **Transform**: In this stage, the extracted data is transformed into a format that is suitable for loading into the data warehouse. This may involve cleaning and validating the data, converting data types, combining data from multiple sources, and creating new data fields.

The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:

- Filtering – loading only certain attributes into the data warehouse.
- Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States, and America into USA, etc.
- Joining – joining multiple attributes into one.
- Splitting – splitting a single attribute into multiple attributes.
- Sorting – sorting tuples on the basis of some attribute (generally key-attribute)

3. **Load**: After the data is transformed, it is loaded into the data warehouse. This step involves creating the physical data structures and loading the data into the warehouse.

The ETL process is an iterative process that is repeated as new data is added to the warehouse. The process is important because it ensures that the data in the data warehouse is accurate, complete, and up-to-date. It also helps to ensure that the data is in the format required for data mining and reporting.



**Need of ETL :**

- It helps companies to analyze their business data for taking critical business decisions.
- Transactional databases cannot answer complex business questions that can be answered by ETL example.
- A Data Warehouse provides a common data repository
- ETL provides a method of moving the data from various sources into a data warehouse.
- As data sources change, the Data Warehouse will automatically update.
- Well-designed and documented ETL system is almost essential to the success of a Data Warehouse project.
- Allow verification of data transformation, aggregation and calculations rules.
- ETL process allows sample data comparison between the source and the target system.
- ETL process can perform complex transformations and requires the extra area to store the data.
- ETL helps to Migrate data into a Data Warehouse. Convert to the various formats and types to adhere to one consistent system.
- ETL is a predefined process for accessing and manipulating source data into the target database.
- ETL in data warehouse offers deep historical context for the business.
- It helps to improve productivity because it codifies and reuses without a need for technical skills.

**Selecting an ETL Tool :**

Selection of an appropriate ETL Tools is an important decision that has to be made in choosing the importance of an ODS or data warehousing application. The ETL tools are required to provide coordinated access to multiple data sources so that relevant data may be extracted from them. An ETL tool would generally contains tools for data cleansing, re-organization, transformations, aggregation, calculation and automatic loading of information into the object database.

An ETL tool should provide a simple user interface that allows data cleansing and data transformation rules to be specified using a point-and-click approach. When all mappings and transformations have been

defined, the ETL tool should automatically generate the data extract/transformation/load programs, which typically run in batch mode.

## UNIT-2

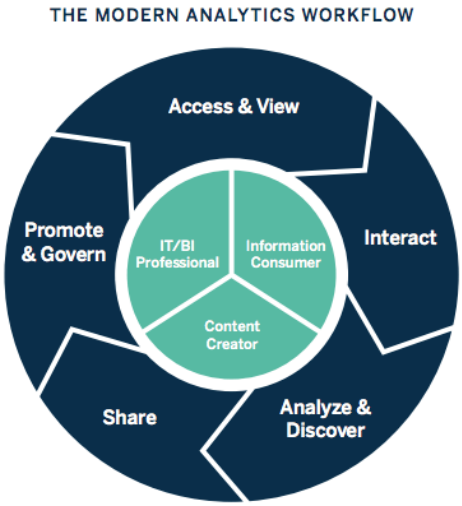
### Business Intelligence Introduction :

Business intelligence refers to a collection of mathematical models and analysis methods that utilize data to produce valuable information and insight for making important decisions.

Business intelligence (BI) is a technology-driven process for analyzing data and delivering actionable information that helps executives, managers and workers make informed business decisions. As part of the BI process, organizations collect data from internal IT systems and external sources, prepare it for

analysis, run queries against the data and create data visualizations, BI dashboards and reports to make the analytics results available to business users for operational decision-making and strategic planning.

Business intelligence (BI) is a set of technological processes for collecting, managing and analyzing organizational data to yield insights that inform business strategies and operations.




## Leveraging Data and Knowledge for BI :

The ability to leverage data requires that we develop an overall architecture and strategy for it. Ensuring that the data models which support the key functional domains are appropriate and that data quality is consistent is key. Likewise with respect to security, access control, roles and the stewarding of data in the organisation. Understanding who owns the data, which systems master it, which systems need access to that master data, how we enrich it and what events generate it, all help shape the overall strategy.

There may be an industry data model the organisation can leverage or should be working with. For some organisations there may be a competitive edge to be yielded from its data if it is leveraged the right way. Organisations need to understand the value of their data and see it as an asset.

When the data is understood, an organisation's ability to extract value from it is driven by their capability in integration terms. Organisations with a high level of integration capability possess a strategy for it, an overall integration framework, the technical skills with the tooling, the body of knowledge associated with the discipline of integration and the understanding of how/when to lean on the technology.

## BI Components :

**Main Components of Business Intelligence System:**

1. **Data Source**
2. **Data Mart / Data Warehouse**
3. **Data Exploration**
4. **Data Mining**
5. **Optimization**
6. **Decisions**

### 1. Data Source:

To begin, the first step is gathering and consolidating data from an array of primary and secondary sources. These sources vary in origin and format, consisting mainly of operational system data but also potentially containing unstructured documents like emails and data from external providers.

2. **Data Mart / Data Warehouse:**

Through the utilization of extraction and transformation tools, also known as extract, transform, load (ETL), data is acquired from various sources and saved in databases designed specifically for business intelligence analysis. These databases, commonly known as data warehouses and data marts, serve as a centralized location for the gathered data.

4. **Data Exploration:**

The third level of the pyramid offers essential resources for conducting a passive analysis in business intelligence. These resources include query and reporting systems, along with statistical methods. These techniques are referred to as passive because decision makers must first develop ideas or

establish criteria for data extraction before utilizing analysis tools to uncover answers and confirm their initial theories.

For example, a sales manager might observe a decrease in revenues in a particular geographic region for a specific demographic of customers. In response, she could utilize extraction and visualization tools to confirm her hypothesis and then use statistical testing to validate her findings based on the data.

5. **Data Mining:**

The fourth level, known as active business intelligence methodologies, focuses on extracting valuable information and knowledge from data. Part II of this book will delve into various techniques such as mathematical models, pattern recognition, machine learning, and data mining.

**5. Optimization:**

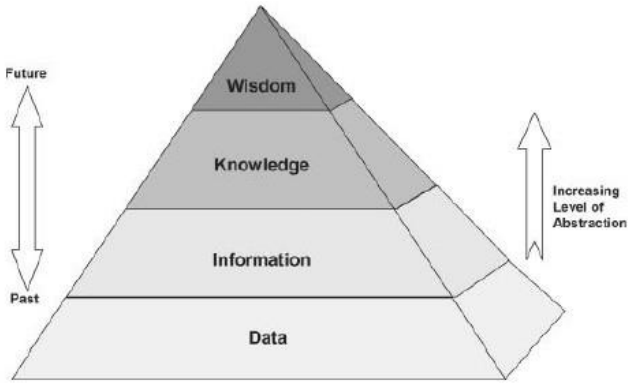
As and ascend the pyramid, and'll encounter optimization models that empower and to choose the most optimal course of action among various alternatives, which can often be quite extensive or even endless. These models have also been effectively incorporated in marketing and logistics.

**6. Decisions:**

At last, the pinnacle of the pyramid reflects the ultimate decision made and put into action, serving as the logical end to the decision-making process. Despite the availability and effective utilization of business intelligence methodologies, the decision still lies in the hands of the decision makers, who can incorporate informal and unstructured information to fine-tune and revise the suggestions and outcomes generated by mathematical models.

## Information Hierarchy :

The Information Hierarchy graph below organizes data, information, knowledge, and wisdom in layers, with an increasing level of abstraction and addition of knowledge, starting from the bottom-most data layer. Various analytical systems help to transform content from one layer to a higher one so as to be better comprehended by analysts.




Data Layer

Data are transactional, physical, and isolated records of activity (e.g., business transactions, customer interactions, facts or figures obtained from experiments or surveys). Data are, for example, numbers, texts, images, videos, and sounds, in a form that is suitable for storage or processing by a computer. Data are the most basic level and by themselves have little purpose and meaning.

Information Layer

Information is the semantic interpretation of data, and may represent relationships among data with meaning and purpose (e.g., a customer called for a specific type of support). Information therefore has been organized and presented in a systematic fashion to clarify the underlying meaning.

Knowledge Layer

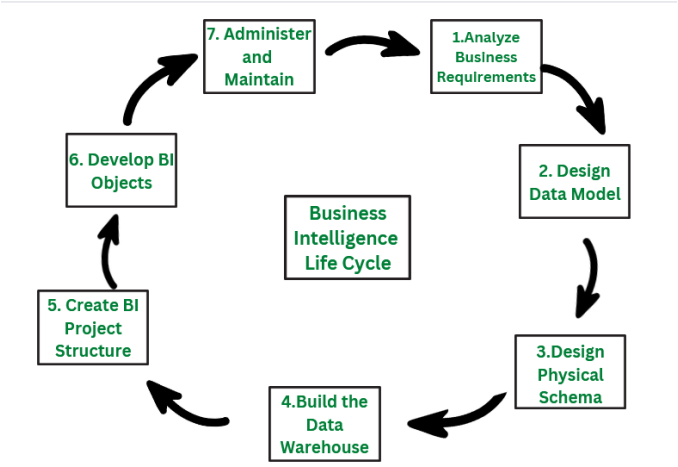
Knowledge is the general awareness or possession of information, facts, ideas, truths, or principles. Knowledge is generally personal and subjective. Knowledge can also be extracted from underlying information by applying some form of induction.

Wisdom Layer

Wisdom is the knowledge of what is true or right coupled with just judgment as to action. Wisdom requires a specific kind of knowledge and experience to make the right decisions and judgments in actions.

## BI Life Cycle :

Business Intelligence is the process of analyzing unprocessed data and turning it into knowledge that the company can use to make decisions. Organizations can better understand their marketing strategies, earnings and losses, effective development and management, and market trends and consumer behavior by using business intelligence.




**Business Intelligence Life Cycle**

***Phase 1: Analyze Business Requirements***

The first step in the Business Intelligence life cycle is to analyze the business requirements. The user identifies the business requirements in order to determine the type of analysis that the user then needs to perform. Identifying the requirements, let the user decides the further action to be performed.

For example, any retail company can analyze the sales data to figure out the products that are top-selling and the products that least sell.

***Phase 2: Design Data Model***

Once the requirements are identified the user needs to design the logical model according to the requirements. This logical model helps the user to analyze the relationships that exist within the data entities.

For example, For any retail company, the data model consists of products, their customers, and the sales data

***Phase 3: Design the Physical Schema***

Once the logical model is prepared the next step is to design the physical schema using the data model. The physical schema describes the structure and the content of the data warehouse.

For example, in any retail company, physical schema consists of sales-related facts, product-customer relationships, and the sales transactions

***Phase 4: Build the Data Warehouse***

Once the logical and physical schema is designed, the next step is to build the data warehouse. The design of a data warehouse depends on the physical and logical schema. After the design of the data warehouse, the data and the content from the source system are loaded into the data warehouse for further steps.

For example, for the retail system, designing the data warehouse consists of developing a database that would store the details of customers, products, and other requirements for the business.

***Phase 5: Create the Project Structure (Metadata)***

The next step after designing the data warehouse is to create a project structure also known as metadata. With the help of this created project structure, the mapping of the tables and data in the data warehouse is easier. Creating the project structure describes the further steps and types that need to be implemented.

For example, The project structure of the retail company consists of the attributes of the data, the design, and the working flow of the system. This project structure or metadata gives a brief idea about the working of the system.

***Phase 6: Develop The BI Objects***

The next step is to develop the BI objects such as metrics, attributes, dashboards, reports, and facts. This step consists of developing the reports and dashboards that can be used to analyze the data in the data warehouse.

For example, the retail company can develop reports and statistics charts that can describe the profit and loss margins.

***Phase 7: Administer and Maintain the Project***

The last step is to administer and maintain the project continuously as it undergoes changes. The project needs to be monitored to maintain the changes, security, and performance of the system.

For example, the retail company needs to monitor the reports and statistics accordingly to increase the profit of the sales.

## Data Issues and Data Quality for BI :

**Data Issues and Data Quality** is a crucial factor for the success of any business intelligence (BI) project. BI relies on accurate, consistent, and timely data to provide insights and support decision-making. However, data quality issues can arise from various sources and affect the BI outcomes in different ways. In this article, we will explore some of the common data quality issues that affect BI outcomes and how to prevent or resolve them.

**Incomplete data** refers to missing or insufficient information in the data sources. This can happen due to human errors, system failures, data integration issues, or lack of standards. Incomplete data can affect the BI outcomes by reducing the coverage, reliability, and validity of the analysis.

For example, if some customer records are missing or incomplete, the BI reports may not reflect the true customer behavior, preferences, or satisfaction. To avoid incomplete data, and should define clear data requirements, validate and monitor the data sources, and implement data quality rules and checks.

**Inconsistent data** refers to data that does not match or conform to the expected format, structure, or values. This can happen due to different data definitions, formats, or standards across the data sources, or due to data entry errors, duplication, or manipulation. Inconsistent data can affect the BI outcomes by creating confusion, ambiguity, and errors in the analysis.

For example, if some product names are spelled differently or have different codes in different data sources, the BI reports may not show the correct sales, inventory, or profitability. To avoid inconsistent data, should establish and enforce common data definitions, formats, and standards, and use data cleansing and deduplication tools.

**Invalid data** refers to data that does not meet the business rules, logic, or expectations. This can happen due to data entry errors, data corruption, data manipulation, or data quality issues in the external data sources. Invalid data can affect the BI outcomes by producing inaccurate, misleading, or irrelevant results.

For example, if some sales transactions have negative or zero values, or some customer records have invalid email addresses or phone numbers, the BI reports may not show the correct revenue, conversion, or retention.

## UNIT-3

## BI Implementation :

**Key Drivers :**

For executives and decision-makers, the need to digitize their processes becomes more pronounced as the competitive advantages that can be accrued from implementing solutions like business intelligence.

**Tech Drivers**

When it comes to technology needs driving the adoption of business intelligence, there are some clear commonalities between organizations as a whole that appear to be pushing the growth of the market.

Firstly, there is a frequent desire among businesses to adopt business intelligence for its technology capabilities and what it can provide in terms of analytics capacity.

significant is the fact that these trends have persisted over a long period of time as key drivers as reported by decision-makers and executives, suggesting their long-term significance with regard to driving BI adoption.

**Business Drivers**

Business drivers for business intelligence adoption among organizations have long been the biggest forces pushing the growth of the industry and show no sign of letting up.

In the simplest terms, rather than specific tech use cases being the primary drivers of adoption, the desire to modernize organizations in terms of how they are structured from

a data and analytics standpoint is more often than not the guiding influence in driving BI strategies.

**Financial Drivers**

Technology is often implemented in an organization for better efficiency and to improve a company's bottom line, and business intelligence is no different.

In this regard, the adoption of business intelligence and a primary driver of BI in companies is the return on investment they hope to see from the better utilization of data.

Business intelligence is a significant and disruptive part of the business tech landscape today.

By leveraging data and information effectively, companies can substantially improve their operational capabilities, and become more data-driven, and more competitive as a result.

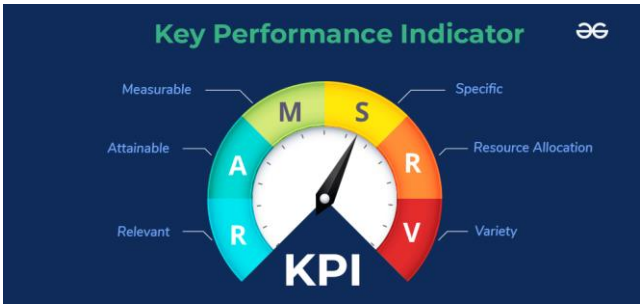
Adoption of business intelligence is typically driven by the three key factors we've looked at today—the desire to be more advanced from a tech standpoint, the need to make better use of data for operational purposes, and the improved productivity that can be reaped from BI implementation.

**Key Performance Indicators and Performance Metrics :**

Key performance indicators are data that show and just how good and are at attaining and business goals. Meanwhile, metrics track the status of and business processes. With KPIs, and will know if and're hitting and overall business targets, while metrics focus on the performance of specific business processes.

KPI stands for key performance indicator, a quantifiable measure of performance over time for a specific objective. KPIs provide targets for teams to shoot for, milestones to gauge progress, and insights that help people across the organization make better decisions.

The best performing businesses work their way through numbers, and there's no reason why and shouldn't give more flesh to and business decisions as they do. Business intelligence key performance indicators or KPIs allow and to glean the overall health of company, any one of departments or even how and customers perceive and company. And nowadays, don't have to tread the business intelligence game manually. And just have to invest in an excellent business intelligence tool to weave the magic numbers for and, a feat that is severely limited or altogether forbidding in the past. If and are resourceful enough, should know which business intelligence tools is best for and unique needs.



1. Financial Metrics :

From a tool like and accounting software, and should be looking at and cash flow, and balance sheet and income statement to derive and financial metrics. One look at any of these metrics should tell and if and business is healthy moneywise, which means and are generating revenue and astutely handling and finances. If and are planning to steer and company into new growth direction or to spark interest from potential investors, and will be presenting these financial KPIs to them for proof of investment value.

2. Marketing Metrics :

Business-wise, next only to financial metrics in the ladder of importance. Marketing metrics reveal the numbers to inform whether your latest marketing campaigns are delivering the rates have set to reach. Capable marketing software tools should give the values, for example, on new content approach riding latest marketing campaigns across multiple channels.

3. Project Management Metrics :

Productivity

Productivity is a straightforward computation of how much you are getting from all those that you put into a project. By dividing all the input units with the output units, you get a good picture of how efficient you are using the resources available to you for each project.

Return on Investment (ROI)

Return on investment tells you how much you get in return for each dollar you invested in a project. It's a simple equation that involves the project net benefits divided by the costs, then multiplying the result by 100.

**4. Customer Service Metrics**

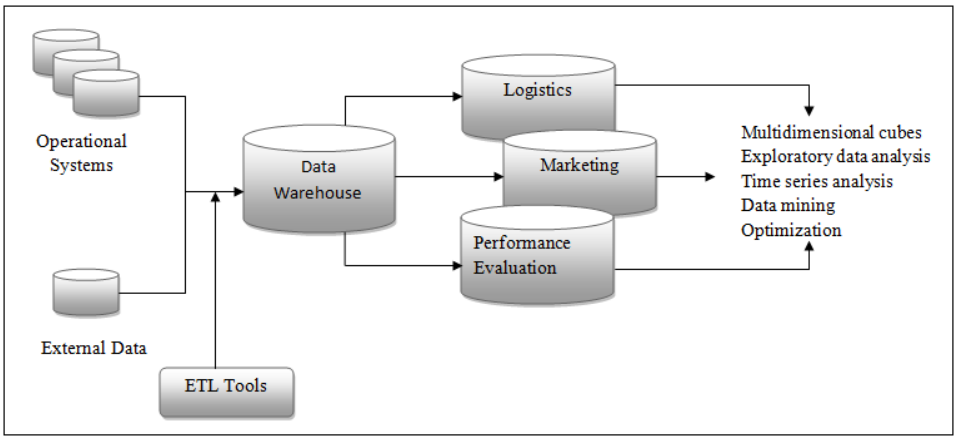
Net Promoter Score (NPS)

Sourced from a single-question survey, this gold standard customer experience metric gives you a singular picture of customer loyalty. Customers respond with values ranging from 0 to 100, where you will be aiming for the higher end of the range. get promoters from those who score 9 or 10, passive from those who score 8 or 9 and detractors from those who give a score of 0 to 6.

**BI Architecture :**

Business intelligence architecture consists of a collection of principles and guidelines that empower business organizations to leverage computer-based methods and technologies for generating meaningful insights. These methodologies are significantly for business intelligence design systems that streamline online data visualization, reporting, and analysis.

The fundamental purpose of business architecture is to close the gap that exists between an organization's strategic objectives and its operational realities. It entails comprehending the purpose, vision, and goals of the company and converting them into real-world systems and procedures that facilitate efficient execution. Business architecture strives to improve operations, boost efficiency, and spur innovation by coordinating the many parts of an organization.




**Components of BI Architecture :**

Data management

As the name implies, it is concerned with how you gather, keep, and access data. In order to support BI activities, a solid data management system must be in place. This system should be tailored to the organization's unique requirements.

Data management happens at all levels of the company, from sensor and device data to personnel data, everything is recorded into databases. External sources of data including social media, market research organizations, and government authorities are also stored.

Data Analytics

The process of transforming data into insights is known as data analytics. Businesses may acquire a better knowledge of their consumers, operations, and industry trends by utilizing data analytics.

There are several sorts of data analytics, but they all have the same goal: to assist organizations in making decision processes to get better results.

Data analytics may be classified into three types: descriptive analytics, predictive analytics, and prescriptive analytics.

- Descriptive analytics provides a solution to the inquiry, "What happened?"

Its objective is to comprehend historical events and patterns.

- Predictive analytics: Predictive analytics provides a solution to the query, “What will happen?”

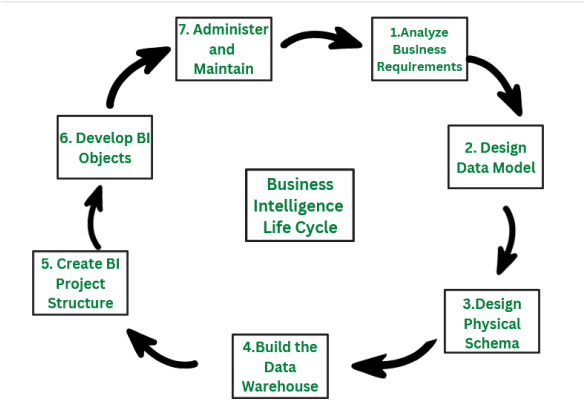
It’s useful to predict future events and trends.

- Prescriptive analytics: Prescriptive analytics provides a solution to the query, “What should we do?”

**Business Decision Making :**

A business decision, called an operational decision, is any choice made by a business professional that determines short-term or long-term company activities. Professionals make business decisions in response to a variety of different situations, including determining which job candidate to hire, how to distribute department budgets, when to expand into a new product market, if they should merge branches and other situations that require well-thought out actions.

**A cyclic process of Intelligence Creation** :




Business Intelligence Life Cycle

***Phase 1: Analyze Business Requirements***

The first step in the Business Intelligence life cycle is to analyze the business requirements. The user identifies the business requirements in order to determine the type of analysis that the user then needs to perform. Identifying the requirements, let the user decides the further action to be performed.

For example, any retail company can analyze the sales data to figure out the products that are top-selling and the products that least sell.

***Phase 2: Design Data Model***

Once the requirements are identified the user needs to design the logical model according to the requirements. This logical model helps the user to analyze the relationships that exist within the data entities.

For example, For any retail company, the data model consists of products, their customers, and the sales data

***Phase 3: Design the Physical Schema***

Once the logical model is prepared the next step is to design the physical schema using the data model. The physical schema describes the structure and the content of the data warehouse.

For example, in any retail company, physical schema consists of sales-related facts, product-customer relationships, and the sales transactions

***Phase 4: Build the Data Warehouse***

Once the logical and physical schema is designed, the next step is to build the data warehouse. The design of a data warehouse depends on the physical and logical schema. After the design of the data warehouse, the data and the content from the source system are loaded into the data warehouse for further steps.

For example, for the retail system, designing the data warehouse consists of developing a database that would store the details of customers, products, and other requirements for the business.

***Phase 5: Create the Project Structure (Metadata)***

The next step after designing the data warehouse is to create a project structure also known as metadata. With the help of this created project structure, the mapping of the tables and data in the data warehouse is easier. Creating the project structure describes the further steps and types that need to be implemented.

For example, The project structure of the retail company consists of the attributes of the data, the design, and the working flow of the system. This project structure or metadata gives a brief idea about the working of the system.

***Phase 6: Develop The BI Objects***

The next step is to develop the BI objects such as metrics, attributes, dashboards, reports, and facts. This step consists of developing the reports and dashboards that can be used to analyze the data in the data warehouse.

For example, the retail company can develop reports and statistics charts that can describe the profit and loss margins.

***Phase 7: Administer and Maintain the Project***

The last step is to administer and maintain the project continuously as it undergoes changes. The project needs to be monitored to maintain the changes, security, and performance of the system.

For example, the retail company needs to monitor the reports and statistics accordingly to increase the profit of the sales.

**The value of Business Intelligence :**

Business Intelligence (BI) brings together a variety of tools, technologies, applications, and practices to analyse your business data. The end result will be decipherable data that can clearly guide future actions for the business. Decision-makers of the business, including executives, managers and owners can use BI to their advantage.

Decision-making

The key purpose of any BI implementation is to convert company data into well-structured, analysable insights or real business intelligence that informs strategic decision-making.

At the heart of intelligent decision-making is having a single, centralised depository that pulls together data from all your business activities and customer interactions.

Great BI is having access to all your business data in a single unified place with a dashboard that includes data from different areas such as sales, finance and inventory control to provide a holistic view of the business, its customers and their interactions with the business.

This means that business decisions are based on facts rather than assumptions or conjecture.

Sales and marketing

BI technology delivers in-depth data to boost sales and support the performance of marketing activities. It provides sales staff with tools to measure sales activity and identify consumer trends and behaviour through improved visibility of sales information.

Specialised features also track, and measure marketing campaigns and promotions and this information can then be used to ensure future marketing initiatives are effective and profitable.

**Value driven and Information use :**

**1. Data sourcing**

The first step in the BI process is accessing the data. This stage deals with the storage, management, and accessing of raw data. Thus, data warehousing becomes an essential aspect of business intelligence. Sourcing data is of utmost importance as the quality and relevancy of data directly impact the quality of insights gained and the consequent decision taken by the leadership. Another aspect of data sourcing deals with identifying and exploring various data resources. It's crucial for the following reasons-

- Data can be related to each other
- It can be unstructured (in the form of text, images, or other forms of unstructured data)
- Data can have peculiar features (such as different data types etc.)

It is vital for a business intelligence analyst to know what kind of data is available and where and how they can access it.

**2. Data engineering and analysis**

After data sourcing, the next logical step is information engineering and analysis. To perform analysis on the data and to gain even the fundamental insights from the data, a BI analyst needs to have the data in a structure that is conducive for analysis. This requires data engineering that includes-

- Converting data into a structured (tabular) format
- Removing or imputing missing values
- Capping outliers
- Removing multicollinearity

Once the data is structured and ready, **Exploratory Data Analysis (EDA)** is performed through which useful information can be synthesized, such as-

- Summarizing data using descriptive statistics

- Finding associations in data

After EDA, Structured Data Analysis (SDA) is performed, which includes three types of analysis-

1. ***Trend Analysis***: It's used for identifying patterns in the data. This includes, for example, assessing sales based on different geographical regions, the volume of product sales over a stipulated period, etc.
2. ***Mathematical Analysis:*** It's used for calculating performance and growth using mathematics. This includes, for example, assessing the margin of sales and growth in absolute and in percentage.
3. ***Statistical Analysis***: Inferential statistics is used to assess the statistical significance of the patterns and peculiarities being identified in the data. Statistics is also used to perform predictions, analytics, and forecasting using regression and other methods. The statistical coefficients can also be used to perform prescriptive analytics, identifying reasons for certain business phenomena. Therefore, the model building can take place at this stage of Business Intelligence.

**3. Situation awareness**

The core of business intelligence is to provide concerned individuals in an organization with knowledge helping in situational awareness. This stage in Business Intelligence deals with report creation and presentations that provide the decision-makers with essential and relevant information to help them be aware of the events in and around the organization. This information can make them aware of, for example, government policies, upcoming market trends, market forces, etc.

**4. Decision-making**

Once the decision-makers know the 'what and why' of the events in and around the business, the next stage is proactive decision-making and its evaluation. The insights, knowledge, and intelligence gained from analytics enable decision-makers to make data-driven decisions. This can be, for example, coming up with strategic decisions such as management change, management of products and categories, opening or shutting of offices, launching of new products, or can be operational, such as promotional campaigns, upsell and cross-sell campaigns, etc.

**5. Decision support**

This stage deals with supporting the proposed decision by evaluating it. Evaluation includes identifying the risks, opportunities, benefits, profit, pitfalls, cost-to-benefit ratio, return on investment, and

estimating the business value of the proposed decisions. All of this helps in making efficient and effective decision-making.

## UNIT - IV

## Advanced BI

Business Intelligence (BI) is a set of technology-driven processes and technologies that convert raw data into useful information to drive profitable business actions.

Big data involves storing, processing, and visualizing a combination of structured, semi-structured, and unstructured data collected by companies to extract meaningful information and insights.

Big data analytics makes use of various advanced analytic techniques, such as predictive models, statistical algorithms, etc., to analyze and process large and diverse datasets from different sources and sizes.

The main goal of Big Data analytics and Business Intelligence is to summarize the data results so that businesses can uncover real insights and trends, thereby helping them make informed decisions.

**Big Data Applications in Business :**

- **Improving customer service**: Organizations can use big data analytics to better understand customer needs and preferences. This can help them improve customer service and provide a more personalized experience.
- **Enhancing product development**: Big data can be used to help businesses improve product development processes. By analyzing data from customer interactions and feedback, businesses can identify trends and understand what products and features customers want.
- **Enhancing marketing efforts**: Big data can help businesses improve their marketing efforts by providing insights into customer behavior. By analyzing data on customer interactions and interests, businesses can develop targeted marketing campaigns that are more likely to be successful.
- **Improving operations**: Big data analytics can be used to improve business operations by identifying inefficiencies and waste. By understanding how customers interact with the organization's products and services, businesses can make changes that improve efficiency and reduce costs.

Big data analytics helps companies collect, process, clean, and analyze large datasets so that they can uncover trends, patterns, and correlations from a large pool of raw data. This helps the companies make data-informed decisions, thereby promoting business growth.

Business Intelligence helps companies and businesses gather the necessary data, analyze it and determine which actions need to be taken to reach their goals. This process also helps them get answers to their queries and track their performance against these goals.

Business intelligence includes data analytics and business analytics, which help users conclude from data analysis. The data scientists use the data, along with advanced statistics and predictive analytics, to uncover patterns and predict future patterns. Business intelligence then uses these models and algorithms to break down the results into actionable language, thereby helping companies make the right business-related decisions that are based on the collected data.

**Social Networks :**

Social intelligence (not to be confused with social BI, which involves the sharing of reports and other visualizations generated by business intelligence platforms) involves collecting data from people who view or interact with your business on social media. This data can consist of demographics, location, number of times visitors view the page before making a purchase, etc.

By collecting this data, users can glean insights, interpret trends and make predictions based on said data in order to make data-driven decisions about their business practices. This can be anything from marketing changes in order to cater to key demographics, brand alignment with certain causes, price changes, etc.

Mobile BI :

BI delivers relevant and trustworthy information to the right person at the right time. Mobile business intelligence is the transfer of business intelligence from the desktop to mobile devices such as the BlackBerry, iPad, and iPhone.

The ability to access analytics and data on mobile devices or tablets rather than desktop computers is referred to as mobile business intelligence. The business metric dashboard and key performance indicators (KPIs) are more clearly displayed.

With the rising use of mobile devices, so have the technology that we all utilise in our daily lives to make our lives easier, including business. Many businesses have benefited from mobile business intelligence. Essentially, this post is a guide for business owners and others to educate them on the benefits and pitfalls of Mobile BI.

**Description of different BI-Tools (Pentaho, KNIME) :**

- **Spreadsheets:** Spreadsheets like Microsoft Excel and Google Docs are some of the most widely used BI tools.
- **Reporting software:** Reporting software is used to report, organize, filter, and display data.
- **Data visualization software:** Data visualization software translates datasets into easy-to-read, visually appealing graphical representations to quickly gain insights.
- **Data mining tools:** Data mining tools "mine" large amounts of data for patterns using things like artificial intelligence, machine learning, and statistics.
- **Online analytical processing (OLAP):** OLAP tools allow users to analyze datasets from a wide variety of angles based on different business perspectives.

KNIME Analytics Platform is an open-source data analysis platform that allows data scientists to create and manage visual data analytics workflows.

**Features/advantages:**

- Scalability via clever data management (intelligent automatic caching of data in the background while maximizing throughput performance)
- High extensibility using a well-defined API for plugin extensions
- Simple user interface
- Workflow import/export (for exchanging with other KNIME users)

Hitachi Data Systems, Pentaho and Hitachi Insight Group have merged into one company: Hitachi Vantara. The tool aims to assist organizations across numerous industries in extracting value from all of their data, including big data and IoT.

**Features/advantages:**

- Intuitive web-based data access wizard and interactive reporting.
- Robust data integration and federation for IT and developers, allowing them to access and integrate data from Excel to Hadoop.
- Interactive visual analysis.

**Price:** The software comes in free community and subscription-based enterprise editions.

**Application example:** flydubai, a United Arab Emirates transportation company, uses Pentaho for Analytics an

# UNIT – V

## Business Intelligence and integration implementation :

Data integration is an important part of any data technology solution. Data integration enables and to easily access all of the data and need to make informed business decisions and allows and to transform and combine data sources for accurate analysis.

The insights that result from having and data contextualized are invaluable because they allow business users to quickly determine what actions should be taken. Furthermore, these insights can give rise to innovative business ideas, more efficient operations, opportunities for expansion, and more.

It's important to have a solid foundation in place that allows and to easily access all of and data while simultaneously keeping costs under control. It's also best to look at BI data integration as an investment rather than an expenditure because it provides returns across the entire organization—from individual departments to executive leadership, it informs the decision-making process. When properly implemented, organizations see increased productivity thanks to streamlined processes and the availability of high-quality information whenever necessary.

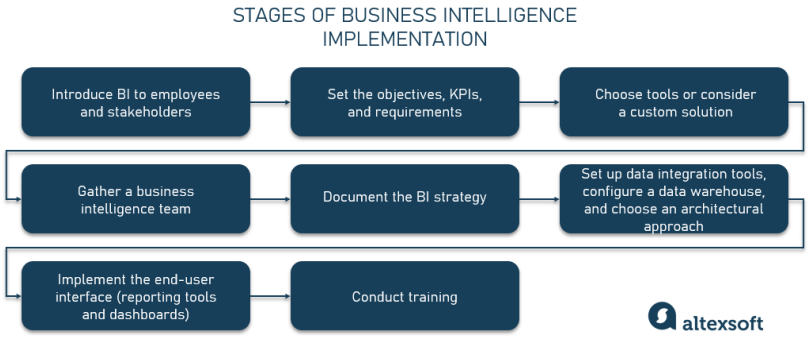
Cloud-based architecture is a key component of an effective data integration strategy. The reason for this is quite simple: cloud-based solutions allow companies to normalize their data across multiple systems and platforms and align their business on a single source of truth. A cloud-based tool also allows for real-time decision-making across multiple data sources.

Cloud-based architecture is a key component of an effective data integration strategy. The reason for this is quite simple: cloud-based solutions allow companies to normalize their data across multiple systems and platforms and align their business on a single source of truth. A cloud-based tool also allows for real-time decision-making across multiple data source.

**Visualize and analyze.** Once and have and data connected and transformed into something that is consumable, it's time to visualize the information in meaningful ways so decision-makers can understand what they are seeing and take intelligent action on accurate information.

A single visualization on its own is usually not enough to answer a complex business question. Interactive dashboards are useful in determining how a change in one area of and business impacts other areas. This type of visualization strategy allows for proactive management.

For business analysts and other stakeholders, the process of getting relevant data is expensive and time-consuming. Data integration powered by modern BI enables them to bring together their different datasets on a single platform, perform the necessary ETL, and build out actionable dashboard content without having to leave their BI tool.



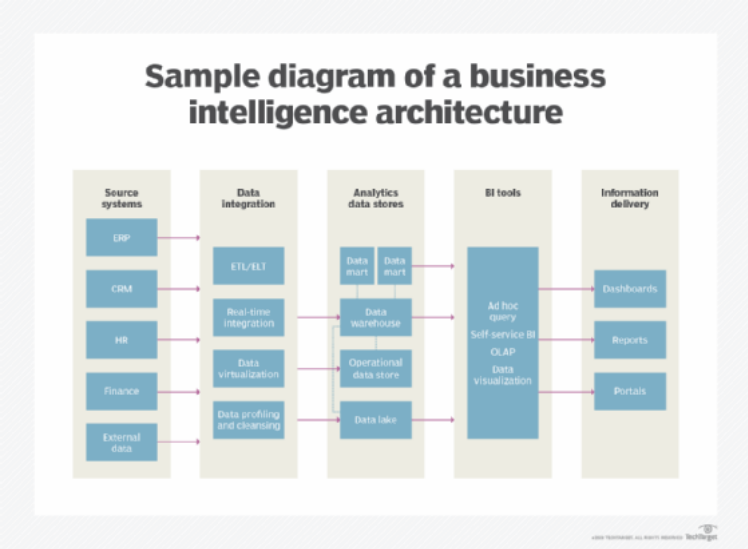
## Connecting in BI systems :

Connecting in BI System Forget switching back and forth between multiple applications or shuffling through myriad spreadsheets. With BI, can get a high-level view of and ERP data via dynamic dashboards that make it easy to understand the data working with. From there, and can get as granular as and want, even tracing issues to their root causes.

Modern BI tools can do more than process incoming data. They can also analyze historical data from and legacy systems, looking for insights that can help and business.

For instance, may want to view and sales performance over the past five years to see if and numbers fluctuate during certain months. As long as this data has been migrated to and new ERP system, it will be easy for BI tools to access and analyze.

predictive analytics, BI can also provide prescriptive analytics. This is where and system takes the predictions it has made and suggests actions for and business to take.



**Issues of legality Privacy and ethics :**

The type of ethics in business intelligence (BI) is the ethical principles of conduct that govern an individual in the workplace or a company in general.

It is also known as professional ethics and not to be confused with other forms of philosophical ethics including religious conviction, or popular conviction. Professional ethics according to is that profit is not the only important strategy of a business anymore. There is also more of a concern and motivator of companies to do what is right.

Companies must acknowledge that they have a common good to protect there local community, improve employee relations and promote informational press to the public. Government regulations are not changing fast enough to cover all the changes in technology that bombards users on day to day bases.
It is up to corporations to create a code of ethics, and to persistently be receptive to the needs of the public being served.
Everyday in BI management professionals may be at risk of making unethical practices in there decisions that regards the consumer, business and/or other employees data. Ethics is a touchy subject, there is always going to be controversy on how companies choose to handle business decisions.

There is no definite decision to make when it comes to ethical decisions. While sometimes it may involve illegal practices, other times it is just a decision that needs to be made in a company to promote a better way of life for all. Always be honest about who and are and where work if engaging in conversation with someone about project.

**Social networking and BI :**

A BI tool that supports mobile, self-serve data preparation, plug n' play predictive analysis and smart data visualization will provide business users with sophisticated tools and algorithms that are easy-to-use and provide access to data that is easy to share and personalize. Business users can use these tools to become Citizen Data Scientists and in so doing, an organization will begin to see the emergence of power users who take a creative, insightful approach to data analysis. When users can share this data in reports and dashboards, the next logical step is to allow those users to comment on, share, and recommend data, reports and analysis.

Just as social media users can 'Like' or 'Share' a provocative post from another user, a social business intelligence tool can support the same kind of sharing and learning within and organization. As power users, and data analysis, become 'popular', the enterprise will see improved user adoption of BI tools and, more importantly, will begin to understand what types of data users value, and how to provide and present data to improve data leverage, and increase confidence in decisions.

By balancing high-quality data with 'popular' self-serve data preparation, in a social/sharing environment, and organization can balance resources and measure and manage data quality vs. data popularity so that the social aspect of data analysis can work hand-in-hand with the quality data approach. Sharing, rating, ranking, commenting and popularizing data will usher in the new generation of business intelligence analysis and make every market more competitive and successful.

***