

UNIT:5 Emerging Trends and Buzz Words

5.1 Emerging Trends and Buzz Words

5.1.1 Concepts of ERP (Fundamentals and importance)

Enterprise Resource Planning (ERP) is a system that helps organizations manage and integrate various business processes, including:

Financial management

Human resources management

Supply chain management

Manufacturing

Customer relationship management

Project management

ERP uses a centralized database to store data and information, making it accessible to all relevant departments. The goal of ERP is to improve efficiency and decision-making by providing real-time access to accurate and consistent data.

Modules of ERP system

1. Finance: In this Data is collected from various functional departments and generate financial reports ledger, Trail Balance, Balance Sheets etc.

2. Human Resource (HR) : HR Module routinely maintain a complete employee database including contact information, Salary details Attendance, Promotions of all employees.

- Produce pay check Reports
- Maintain personnel Record
- Training
- Time and Attendance Benefits

3. Purchasing: Purchasing module is tightly integrated with the inventory control and production planning modules

Importance of ERP System

The implementation of an ERP system can bring several benefits to an organization, including:

1. Improved data accuracy and consistency: ERP centralizes data storage, ensuring that all departments have access to the same information.
2. Increased efficiency: Automation of manual processes and integration of various business functions can lead to a more streamlined and efficient operation.

3. Better decision making: Real-time access to accurate and up-to-date information enables managers to make informed decisions.
4. Better visibility and control: ERP systems provide greater visibility into all aspects of the business, enabling managers to monitor performance and identify areas for improvement.
5. Improved collaboration: ERP systems allow departments to share information and work together more effectively, leading to better collaboration and increased productivity.
6. Cost savings: By streamlining processes and reducing manual labor, ERP systems can help organizations save time and money.
7. Increased customer satisfaction: Improved efficiency, accuracy, and communication can lead to better customer service and higher customer satisfaction.

Limitations of ERP

While ERP systems bring many benefits, they also have some limitations, including:

Implementation cost and complexity: Implementing an ERP system can be a complex and time-consuming process, requiring significant investment in terms of both time and money.

Resistance to change: Employees may resist the changes brought about by an ERP system, leading to lower adoption rates and decreased productivity.

Customization limitations: ERP systems may not always meet the specific needs of an organization, and customization can be limited.

Integration challenges: Integrating an ERP system with other systems and processes can be challenging, particularly for organizations with legacy systems in place.

Dependence on technology: An ERP system relies on technology and may be vulnerable to technical failures, which can disrupt business operations.

Data security concerns: Centralized data storage can increase the risk of data breaches, and organizations must ensure the security of sensitive information.

Inflexibility: ERP systems can be rigid, making it difficult to accommodate new business requirements or changes in business processes.

5.1.2 ETL Concepts (Extraction, transformation and loading)

Extract, Transform, Load (ETL) is a process used in data integration to extract data from various sources, transform it into a desired format, and load it into a target system for analysis and reporting.

Extract: The first step in the ETL process is to extract data from various sources, such as databases, spreadsheets, and text files. The data is usually extracted using specialized software tools or scripting languages, such as SQL.

Transform: Once the data has been extracted, it must be transformed into a format that can be loaded into the target system. This may involve cleaning the data, merging data from multiple sources, and converting data types.

Load: The final step is to load the transformed data into the target system. This is typically done using a database management system or data warehousing system, such as a relational database or a data mart.

Advantages of ETL Tool

The use of an ETL (Extract, Transform, Load) tool can bring several advantages to an organization, including:

Improved data accuracy and quality: ETL tools help to automate the process of data integration, reducing the risk of human error and improving data accuracy and quality.

Increased efficiency: ETL tools can automate many of the manual tasks involved in data integration, reducing the time and resources required to extract, transform, and load data.

Better data integration: ETL tools can seamlessly integrate data from multiple sources, regardless of format or structure, providing a unified view of the data.

Scalability: ETL tools can handle large amounts of data and can be scaled to meet the changing needs of an organization.

Flexibility: ETL tools can be customized to meet the specific needs of an organization, and can accommodate changes in data sources and data structures over time.

Data security: ETL tools can help to secure sensitive data during the integration process, helping to protect the confidentiality and integrity of the data.

Time savings: Automating the data integration process with an ETL tool can save time and reduce the workload for IT staff, allowing them to focus on other important tasks.

Better decision making: By providing a unified view of the data, ETL tools can help organizations make more informed decisions, based on accurate and up-to-date information.

Popular ETL tools are:

Some of the most popular ETL tools include:

Talend

Informatica PowerCenter

Microsoft SQL Server Integration Services (SSIS)

Oracle Data Integrator (ODI)

Apache Nifi

AWS Glue

Google Cloud Dataflow

Alteryx

Hevo

Pentaho Data Integration

Each of these tools has its own unique features and capabilities, and the best choice will depend on the specific needs and requirements of an organization. Factors such as cost, scalability, ease of use, and level of customization may influence an organization's decision when choosing an ETL tool.

5.1.3 Concepts of data warehousing

What is Data warehouse?

A data warehouse is a large, centralized repository of data that is used for storing, managing, and analyzing data from various sources. The purpose of a data warehouse is to provide a single, consolidated view of an organization's data that can be used for reporting and analysis.

Data warehouses are optimized for read-intensive operations and are designed to handle large amounts of data, making it possible to analyze data from multiple sources, over a long period of time. The data in a data warehouse is typically

organized in a way that makes it easy to perform complex queries and analysis, such as multidimensional analysis and data mining.

A data warehouse typically includes:

Data from multiple sources: A data warehouse can include data from a variety of sources, such as transactional systems, operational databases, and external data sources.

Centralized repository: Data is consolidated into a single location, making it easier to manage and analyze the data.

Historical data: A data warehouse is designed to store historical data, enabling organizations to analyze trends and patterns over time.

Subject-oriented data: Data is organized around specific subjects, such as customers, products, or sales, making it easier to perform specific types of analysis.

Non-volatile data: Data in a data warehouse is non-volatile, meaning that it is not updated in real-time. This allows organizations to perform analysis without affecting the performance of transactional systems.

Data warehouses play a critical role in supporting business intelligence and decision-making, enabling organizations to make informed decisions based on accurate, up-to-date information.

Database Vs Data warehouse:

A database and a data warehouse are both systems for storing data, but there are some key differences between them:

Purpose: A database is used for storing and managing transactional data, such as customer orders, product information, and financial transactions. A data warehouse, on the other hand, is designed for storing, managing, and analyzing large amounts of historical data for business intelligence purposes.

Data Volume: A database is typically optimized for handling a high volume of transactions, while a data warehouse is optimized for handling large amounts of data.

Data Structure: Databases use a transactional data structure, which is optimized for inserting, updating, and deleting data. Data warehouses use a dimensional data structure, which is optimized for complex queries and analysis.

Data Integration: A data warehouse often integrates data from multiple sources, including transactional systems, external data sources, and legacy systems. A database, on the other hand, typically contains data from a single source.

Data Consistency: A database is designed to ensure data consistency and integrity, with strict rules for updating data. A data warehouse, however, allows for more flexible data updates, as the focus is on providing a consolidated view of the data for analysis.

Performance: A database is optimized for fast, real-time transactions, while a data warehouse is optimized for fast, complex queries and analysis.

In summary, a database is typically used for managing transactional data, while a data warehouse is used for analyzing historical data for business intelligence purposes. The choice between a database and a data warehouse will depend on the specific needs and requirements of an organization.

Why Data warehouse is required?

Data warehouses are required for a number of reasons, including:

Consolidated view of data: A data warehouse provides a single, consolidated view of an organization's data, making it easier to analyze data from multiple sources and make informed decisions.

Improved data analysis: A data warehouse is optimized for complex queries and analysis, enabling organizations to perform multidimensional analysis and data mining, and uncover hidden patterns and trends in their data.

Historical data: A data warehouse stores historical data, allowing organizations to analyze trends and patterns over time and make informed decisions based on historical data.

Increased business intelligence: A data warehouse provides organizations with a centralized repository of data that can be used for business intelligence and decision-making, enabling them to make informed decisions based on accurate, up-to-date information.

Integration of disparate data sources: A data warehouse integrates data from multiple sources, including transactional systems, external data sources, and legacy systems, providing a single source of truth for an organization's data.

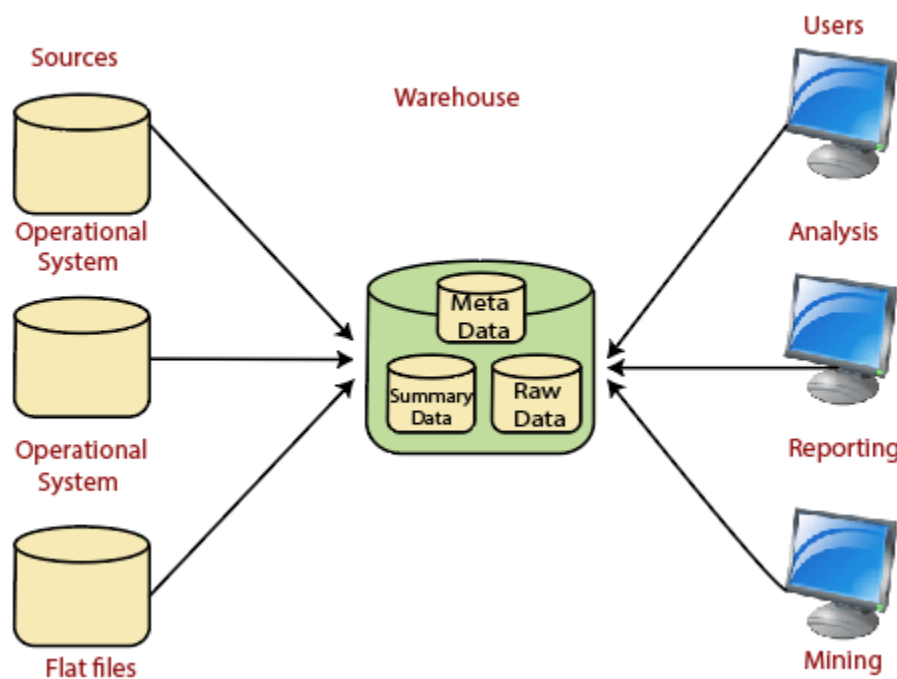
Improved data quality: Data in a data warehouse is typically cleaned, transformed, and enriched, improving the quality of the data and reducing the risk of incorrect analysis.

Performance: A data warehouse is optimized for fast, complex queries and analysis, making it possible to perform large-scale data analysis quickly and efficiently.

In summary, a data warehouse is a critical component for organizations that want to make informed decisions based on accurate, up-to-date information. By providing a consolidated view of data, improved data analysis, and increased business intelligence, a data warehouse can help organizations make better decisions, improve operational efficiency, and achieve a competitive advantage.

Data Warehouse Architecture:

Architecture of a Data Warehouse



Features of a data warehouse:

The main features of a data warehouse include:

Scalability: Data warehouses are designed to scale to accommodate large amounts of data and provide fast, complex queries and analysis.

Data Integration: Data warehouses integrate data from multiple sources, including transactional systems, external data sources, and legacy systems.

Data Cleansing: Data in a data warehouse is typically cleaned and transformed to improve data quality and reduce the risk of incorrect analysis.

Multidimensional Analysis: Data warehouses are optimized for multidimensional analysis, making it possible to analyze data from multiple perspectives and uncover hidden patterns and trends.

Historical Data: Data warehouses store historical data, allowing organizations to analyze trends and patterns over time.

Data Security: Data warehouses typically have robust security features to ensure that sensitive data is protected and only accessible by authorized users.

Performance: Data warehouses are optimized for fast, complex queries and analysis, enabling organizations to perform large-scale data analysis quickly and efficiently.

Reporting and Analytics: Data warehouses provide organizations with the ability to generate reports and perform advanced analytics, including predictive analytics and data mining.

OLAP (Online Analytical Processing): Data warehouses support OLAP, which provides a multidimensional view of data that makes it easier to perform complex queries and analysis.

Data Marts: Data warehouses often include data marts, which are smaller, specialized data warehouses that contain data for specific business functions or departments.

In summary, a data warehouse provides organizations with a centralized repository of data that can be used for business intelligence and decision-making. With its ability to scale, integrate data from multiple sources, perform multidimensional analysis, and provide fast, complex queries and analysis, a data warehouse is a critical component for organizations that want to make informed decisions based on accurate, up-to-date information.

5.1.7 Concept of Cloud:

Cloud computing is a model of delivering computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale.

In cloud computing, users can access technology resources on-demand, without having to manage infrastructure, install software, or worry about capacity planning. They pay only for what they use, typically on a utility (pay-per-use) basis.

There are three main types of cloud computing services:

Infrastructure as a Service (IaaS): IaaS provides virtualized computing resources over the internet, including servers, storage, and networking. Customers can provision and manage these resources as needed, without having to invest in physical hardware.

Platform as a Service (PaaS): PaaS provides a platform for building, deploying, and managing applications and services over the internet. PaaS includes development tools, databases, and middleware as well as application servers.

Software as a Service (SaaS): SaaS provides customers with access to applications over the internet, without having to install or run the applications on their own computers. SaaS includes productivity applications, such as email and customer relationship management, as well as industry-specific applications.

Cloud computing provides a number of benefits, including:

Scalability: Cloud computing resources can be scaled up or down as needed, without having to make upfront investments in hardware.

Cost Savings: Cloud computing eliminates the need to invest in expensive hardware and software, and provides a pay-per-use model that can reduce costs compared to traditional computing models.

Flexibility: Cloud computing offers users the ability to access technology resources from anywhere, at any time, and from any device.

Innovation: Cloud computing enables faster innovation, as users can access the latest technology resources and services as soon as they become available.

Disaster Recovery: Cloud computing provides built-in disaster recovery and business continuity capabilities, reducing the risk of data loss and downtime.

In summary, cloud computing is a model for delivering computing services over the internet, offering users the ability to access technology resources on-demand, without having to manage infrastructure or invest in physical hardware. With its scalability, cost savings, flexibility, and innovation, cloud computing is becoming an increasingly popular computing model for organizations of all sizes.

Advantages:

- Lower computer costs
- Improved performance
- Reduced software costs
- Instant software updates
- Improved document format compatibility
- Unlimited storage capacity
- Increased data reliability
- Universal document access
- Latest version availability
- Easier group collaboration
- Device independence

Disadvantages:

- Requires a constant Internet connection
- Does not work well with low-speed connections
- Features might be limited
- Can be slow
- Stored data can be lost
- Stored data might not be secure

Deployment models for cloud computing:

There are three main deployment models for cloud computing:

Public Cloud: Public clouds are owned and operated by third-party service providers and are open for use by the general public. Public clouds offer on-demand access to shared computing resources, such as servers, storage, and applications, over the internet. Examples of public cloud providers include Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

Private Cloud: Private clouds are owned and operated by a single organization for exclusive use by that organization. Private clouds offer the benefits of cloud

computing, such as scalability and cost savings, but with more control over the underlying infrastructure. Private clouds can be hosted on-premises or by a third-party service provider.

Hybrid Cloud: Hybrid clouds combine the benefits of public and private clouds by allowing organizations to run some applications and services in a public cloud, while others are run in a private cloud. Hybrid clouds offer organizations the flexibility to choose the best deployment model for each application or service, based on factors such as security, performance, and cost.

Each deployment model has its own set of benefits and limitations, and organizations can choose the model that best fits their needs based on their specific requirements and constraints.

In summary, public clouds offer the benefits of on-demand access and scalability, private clouds offer greater control and security, and hybrid clouds offer the best of both worlds by combining the benefits of public and private clouds.

5.1.4 Concepts of Data Science and its application area:

Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data. So, Data Science is primarily used to make decisions and predictions, It uses predictive causal analytics, prescriptive analytics (predictive plus decision science) and machine learning.

Predictive causal analytics: Predictive casual analytics apply to predict the possibilities of a particular event in the future.

For example: if you are providing money on credit, then the probability of customers making future credit payments on time is a matter of concern for you. Here, you can build a model which can perform predictive analytics on the payment history of the customer to predict if the future payments will be on time or not.

Prescriptive analytics: Prescriptive analytics is applied when we want system has the intelligence of taking its own decisions and the ability to modify it with dynamic parameters. In other terms, it not only predicts but suggests a range of prescribed actions and associated outcomes.

For example: Google's self-driving car, the data gathered by vehicles can be used to train self-driving cars. You can run algorithms on this data to bring intelligence to it. This will enable your car to take decisions like when

to turn, which path to take, when to slow down or speed up.

Machine learning for making predictions: It is applied when transactional data of a finance company and need to build a model to determine the future trend, then machine learning algorithms are the best bet. This falls under the paradigm of supervised learning. It is called supervised because you already have the data based on which you can train your machines.

For example - a fraud detection model can be trained using a historical record of fraudulent purchases.

Machine learning for pattern discovery: It is applied when don't have the parameters based on which you can make predictions, and then you need to find out the hidden patterns within the dataset to be able to make meaningful predictions. This is nothing but the unsupervised model as you don't have any predefined labels for grouping. The most common algorithm used for pattern discovery is Clustering.

For example : In a telephone company and you need to establish a network by putting towers in a region. Then, you can use the clustering technique to find those tower locations which will ensure that all the users receive optimum signal strength.

Application areas of Data Science

Internet Search

There are many search engines like Yahoo, Bing, Ask, Google etc. All these search engines make use of data science algorithms to deliver the best result for our searched query in fraction of seconds. Considering the fact that, Google processes more than 20 petabytes of data every day. Had there been no data science, Google wouldn't have been the 'Google' we know today.

Digital Advertisements (Targeted Advertising and re-targeting)

The entire digital marketing spectrum. Starting from the display banners on various websites to the digital bill boards at the airports – almost all of them are decided by using data science algorithms.

This is the reason why digital ads have been able to get a lot higher CTR (Click –Through Rate) than traditional advertisements. They can be targeted based on user's past behavior time.

Image Recognition

You upload your image with friends on Facebook and you start getting suggestions to tag your friends. This automatic tag suggestion feature uses face recognition algorithm. Similarly, while using whatsapp web, you scan a barcode in your web browser using your mobile phone. In addition, Google provides you the option to search for images by uploading them. It uses image recognition and provides related search results.

Speech Recognition

Some of the best example of speech recognition products is Google Voice, Siri, etc. Using speech recognition feature, even if you aren't in a position to type a message, your life wouldn't stop. Simply speak out the message and it will be converted to text.

Price Comparison Websites

At a basic level, these websites are being driven by lots and lots of data which is fetched using APIs and RSS Feeds. Price Grabber, Price Runner, Jungle, Shopzilla, Deal Time are some examples of price comparison websites. Now days, price comparison website can be found in almost every domain such as technology, hospitality, automobiles, durables, apparels etc.

Fraud and Risk Detection

One of the first applications of data science originated from Finance discipline. Companies were fed up of bad debts and losses every year. However, they had a lot of data which use to get collected during the initial paper work while sanctioning loans. They decided to bring in data science practices in order to rescue them out of losses. Over the years, banking companies learned to divide and conquer data via customer profiling, past expenditures and other essential variables to analyze the probabilities of risk and default. Moreover, it also helped them to push their banking products based on customer's purchasing power.

5.1.5 Data Analytics and its related Tools:

Data Analytics involves the use of techniques and tools to extract insights and knowledge from data. Key concepts in data analytics include:

1. **Data Collection & Preparation:** Gathering and cleaning data from various sources.
2. **Exploratory Data Analysis (EDA):** Analyzing data to discover patterns and relationships.
3. **Data Visualization:** Using visual aids like graphs, charts, and maps to represent data.

4. **Predictive Analytics:** Using statistical and machine learning methods to make predictions about future events.
5. **Data Mining:** The process of discovering patterns and knowledge from large data sets.
6. **Data Warehousing:** Storing and organizing large amounts of data for efficient retrieval and analysis.
7. **Business Intelligence:** The use of data analytics to inform business decisions and measure performance.

Related Tools:

1. **SQL:** A programming language used for managing and querying data stored in relational databases.
2. **R:** An open-source programming language and software environment for statistical computing and graphics.
3. **Python:** A general-purpose programming language used in data science, machine learning, and scientific computing.
4. **Tableau:** A data visualization and business intelligence tool used to create interactive dashboards and reports.
5. **Power BI:** A business intelligence tool used for data visualization and reporting.
6. **SAS:** A statistical software suite used for data analysis, predictive modeling, and business intelligence.
7. **KNIME:** An open-source data analytics platform used for data mining and machine learning.
8. **Hadoop:** An open-source software framework used for storing and processing big data.

5.1.6 Concepts of Business analytics

Business Analytics involves the use of data, statistical and quantitative analysis, and other fact-based methods to drive informed decisions and solve complex business problems. Key concepts in Business Analytics include:

Data Collection & Preparation: Gathering and cleaning data from various sources.

Data Warehousing: Storing and organizing large amounts of data for efficient retrieval and analysis.

Data Visualization: Using visual aids like graphs, charts, and maps to represent data.

Descriptive Analytics: Summarizing and describing data to understand the current state of a business.

Predictive Analytics: Using statistical and machine learning methods to make predictions about future events.

Prescriptive Analytics: Using optimization and simulation techniques to generate and evaluate alternative solutions to business problems.

Business Intelligence: The use of data analytics to inform business decisions and measure performance.

Dashboards and Reporting: A visual representation of business performance data and KPIs.

Customer Analytics: The use of data and analytics to understand and improve customer behavior and experience.

Marketing Analytics: The use of data and analytics to inform and optimize marketing efforts.

Related Tools:

SQL: A programming language used for managing and querying data stored in relational databases.

Tableau: A data visualization and business intelligence tool used to create interactive dashboards and reports.

Power BI: A business intelligence tool used for data visualization and reporting.

SAS: A statistical software suite used for data analysis, predictive modeling, and business intelligence.

IBM Cognos: A business intelligence and performance management tool used for data analysis and reporting.

Oracle Business Intelligence: An enterprise-level business intelligence platform used for data analysis, reporting, and dashboards.