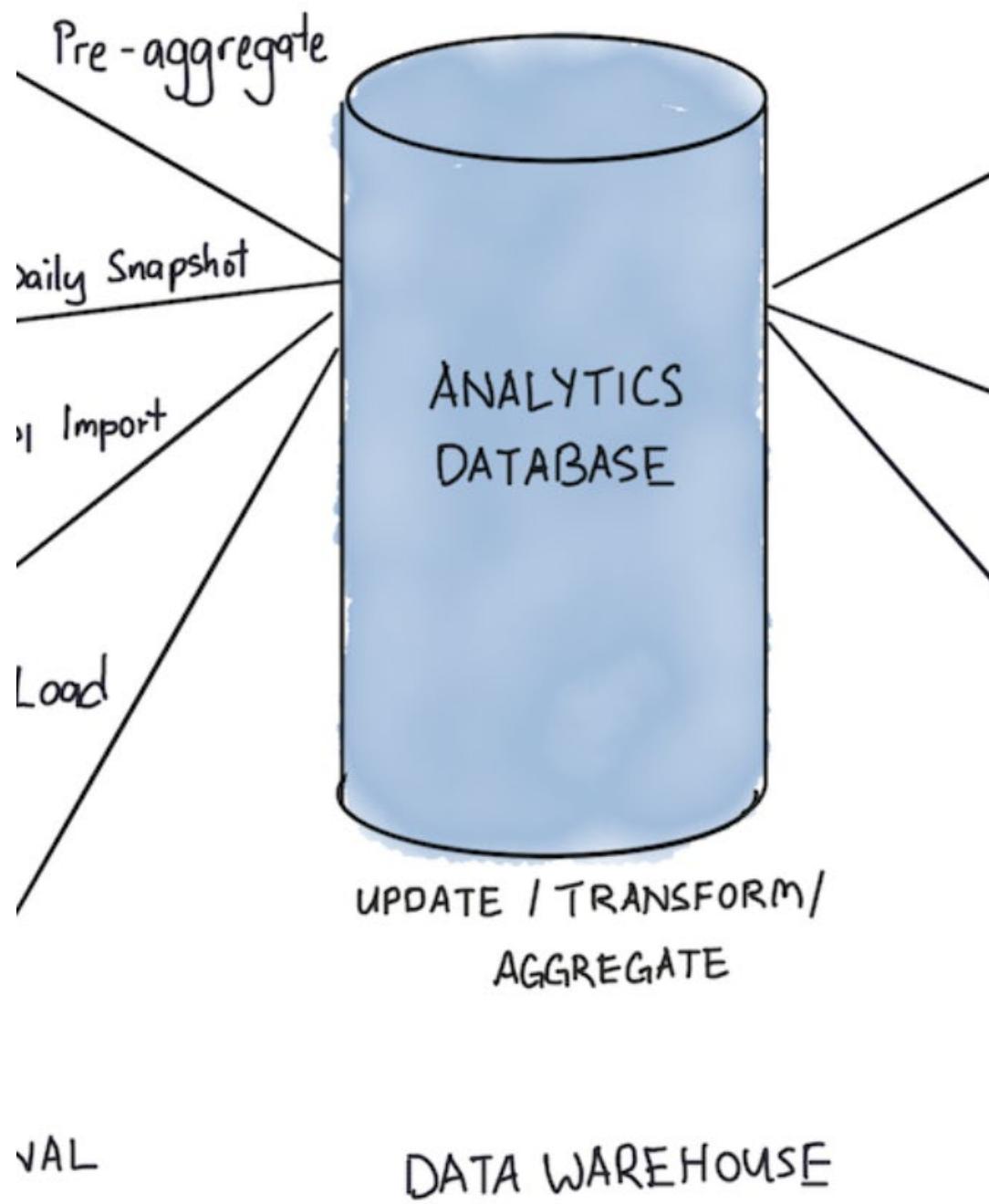


Chapter 5

Databases and Data Analytics

By the end of this lecture, you will be able to:

Discover the world of databases and data analytics. Learn about different types of databases, data analytics techniques, data warehousing, data visualization, and big data analytics.



Part 1: Introduction to Databases

Physical and logical views

Characters, fields, records, tables, and databases

Key fields

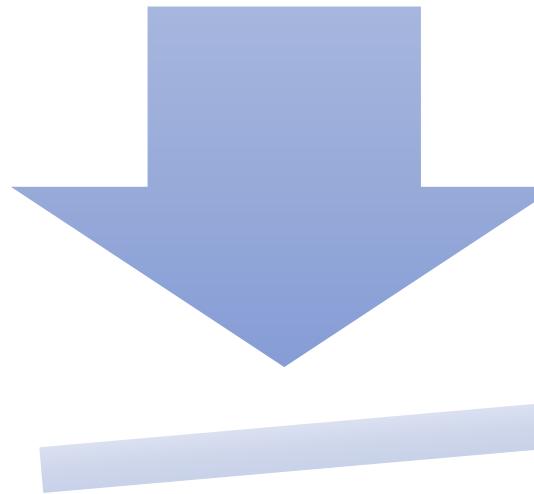
Batch processing and real-time processing

Database models

Individual, company, distributed and commercial databases

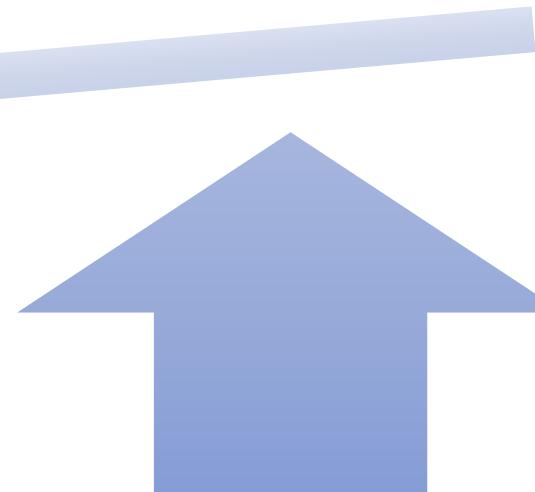
Database uses and security concerns

Introduction



A database is an electronic system that **allows data to be easily accessed, manipulated and updated**

Like a library, secondary storage is designed **to store information and an organized collection of data**



Data



Examples of data include:

- Facts or observations about people, places, things, and events
- Audio, music, photographs, and video

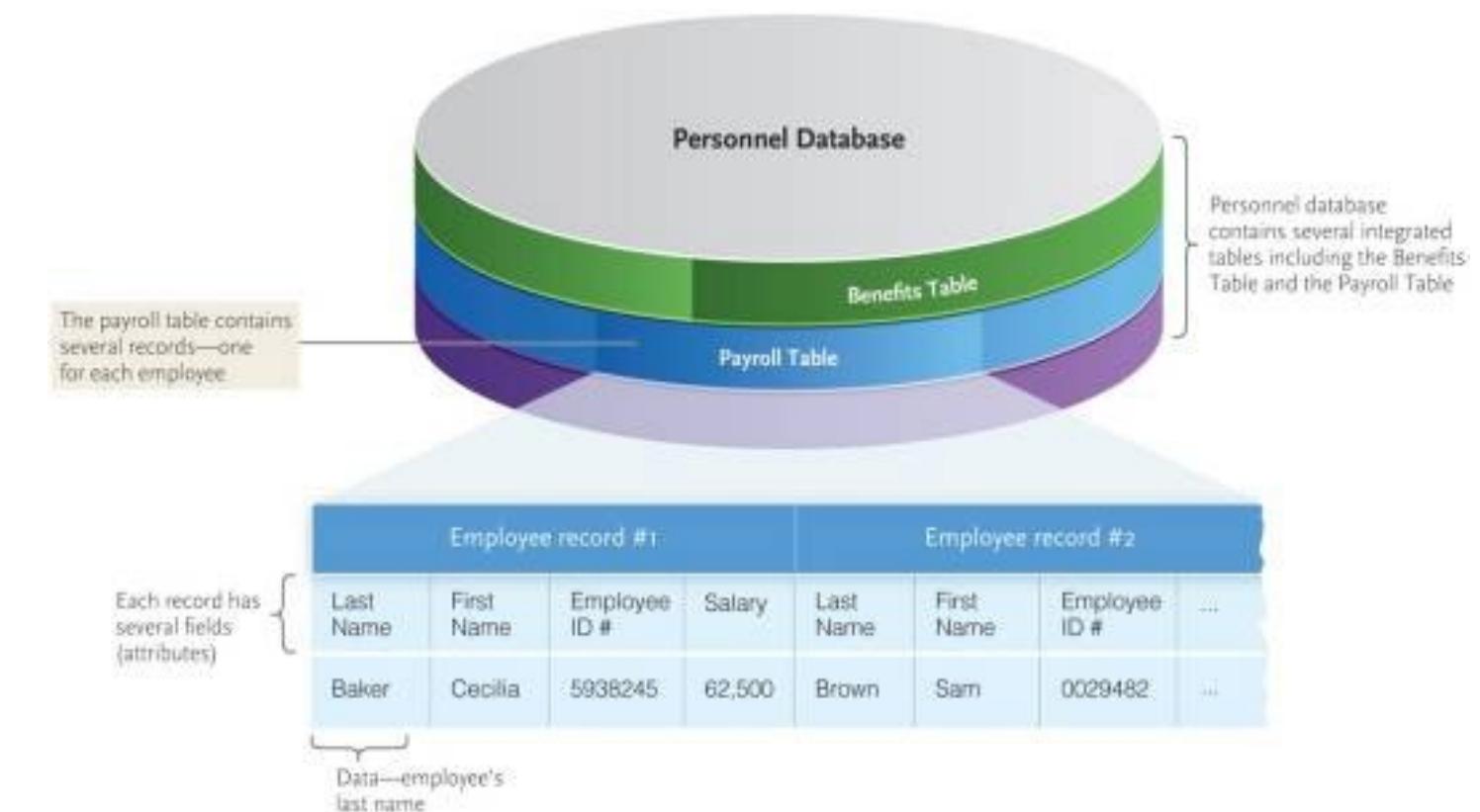


Type of data

- Structured data
- Semi structured data
- Unstructured data

Data Organization

- Character
- Field
- Record
- Table
- Database



Key Field



Unique identifier also known as **primary key**



Common examples:

- Social Security Numbers
- Student Identification Numbers
- Employee Identification Numbers
- Part Numbers
- Inventory Numbers

Batch Processing

- Batch processing:
 - Data is **collected over a period of time** and the processing happens **later all at one time**



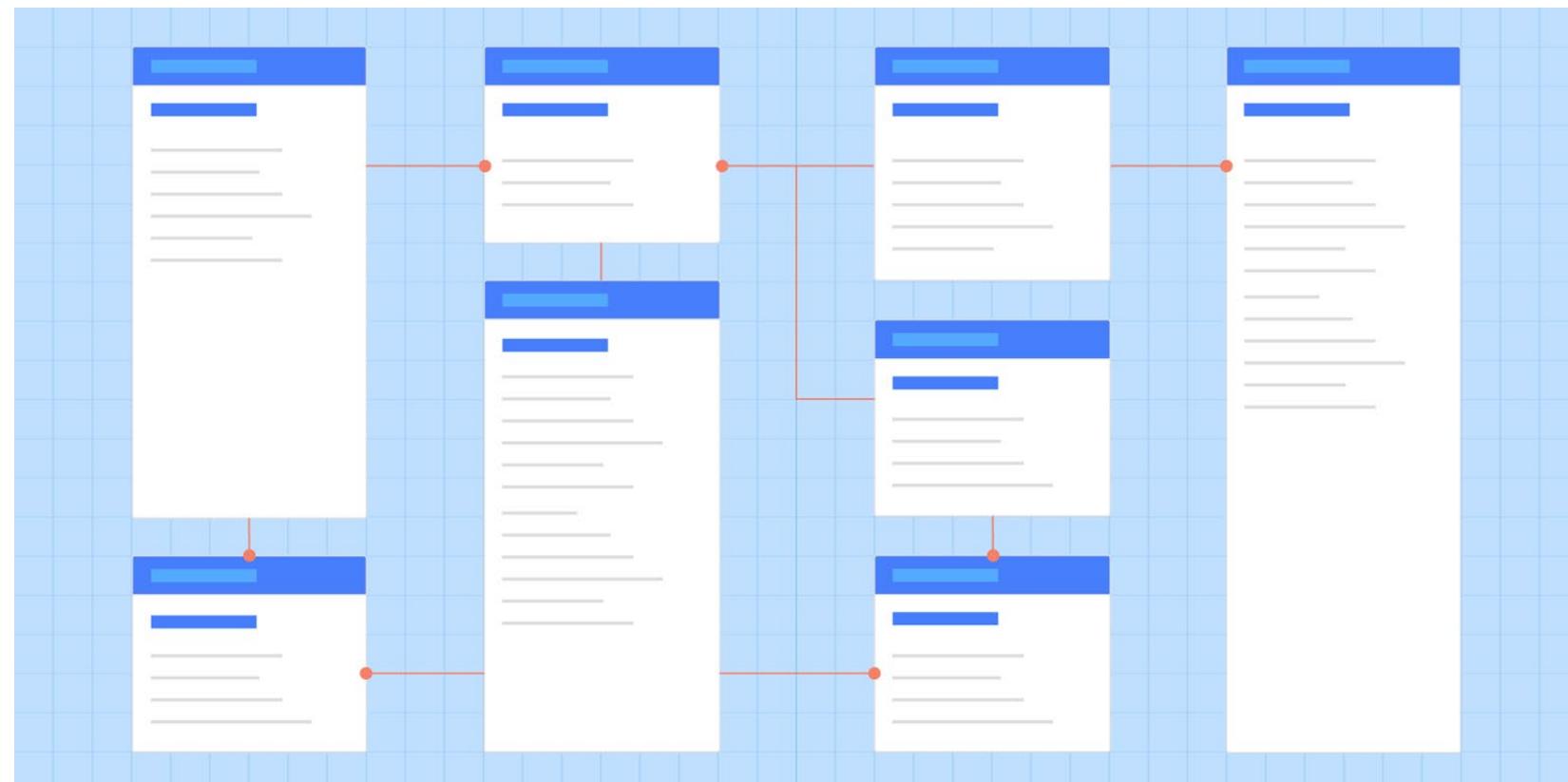
Real-time Processing

- Real-time processing:
 - Also known as **online processing** because it happens immediately during the transaction



Databases

- Collection of integrated data
 - Logically related files and records
- Databases address **data redundancy and data integrity**



Need for Databases

- Sharing
- Security
- Less data redundancy
- Data integrity

Employee Records2 - Employee Records

| Employee ID: | 02731 | Zip: | 92120-3741 | Contact Name: | Miranda Price | Home Phone: | (864) 938-2857 | Work Phone: | (864) 454-7734 |
|--------------|-----------------|-------------|---|--------------------|--|--|---|-------------|----------------|
| Hire Date: | 8/19/1999 | Phone: | (507) 555-6707 | Home address: | 4687 Stride Drive Greenville, SC 29602 | Respiration: | 9 | | |
| Last Name: | Marchant | Gender: | F | Pulse: | 86 | Temperature: | 97.9 F | | |
| First Name: | Roberta | Birth Date: | 5/13/1980 | Patient Allergies: | Pet dander, Pollen | Adverse Drug Reactions: | Sulfa drugs causes mild skin rash | | |
| Street: | 564 Palm Avenue | Photo: | Current Medications: Insulin, Diabinese | | | | Self Administrations of Medications: No | | |
| City: | Laird | | | | | Primary Medical Diagnosis: Admitted for: elevated blood sugar levels and dehydration. Diagnosed with: Type II Diabetes. Current status: blood sugar levels have decreased and he has been treated for his dehydration. | | | |
| State: | CA | | | | | | | | |

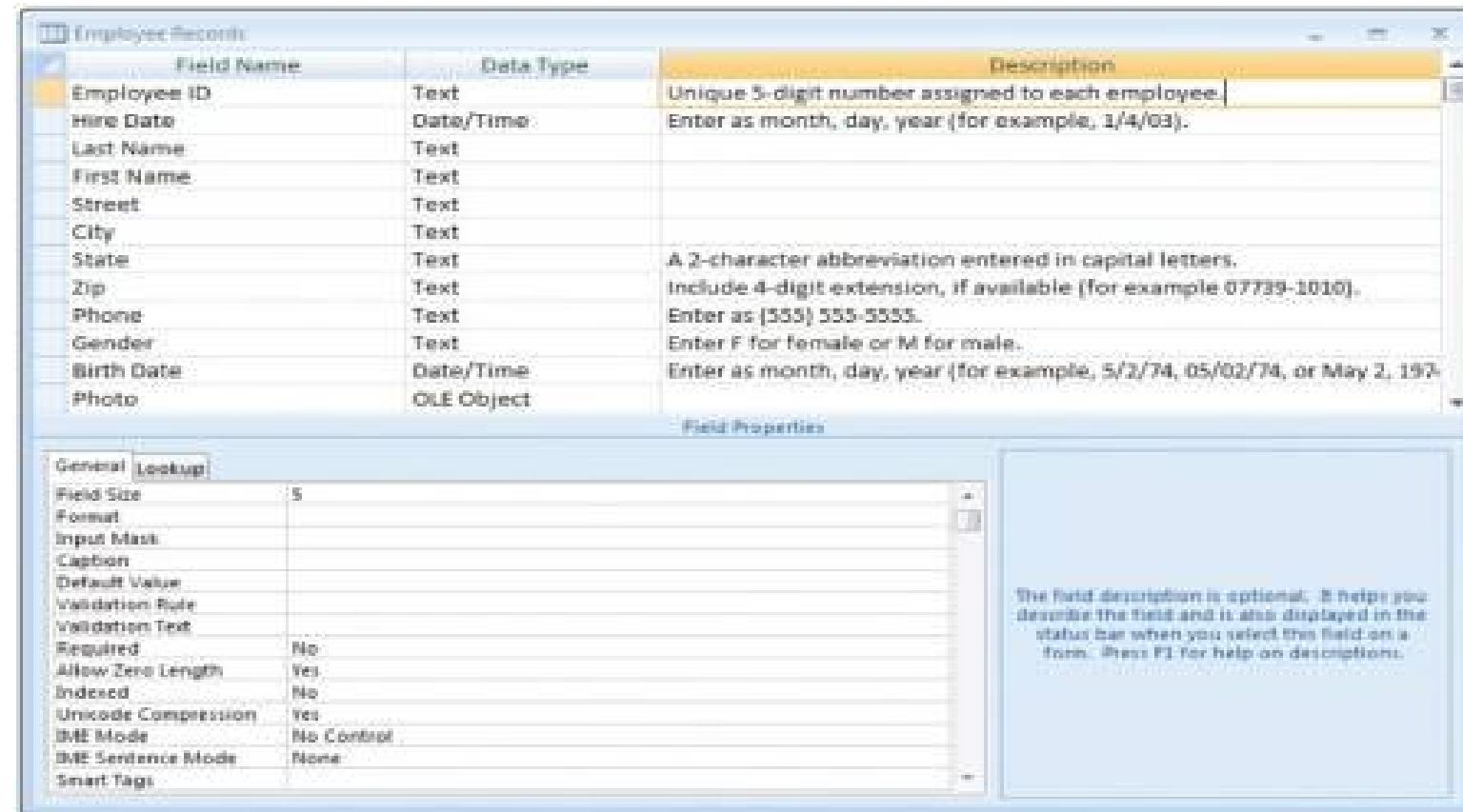
Record: 46 | 5 of 70 | < | > | << | >> | G | No Filter | Search |



| Hospital Patient Record | | |
|--|---|---|
| Patient Number: 1 | Date of Assessment: 2/15/2009 | |
| Completed By: Nancy Turner | Physician's Name: Dr. Andrea Wilson | |
| Basic Patient Information | | |
| Patient Name: Frank Davidson | | |
| Sex: Male | Date of Birth: 8/25/1946 | Social Security Number: 995065934 |
| Height: 6'2 | Weight: 203 | |
| Street Address: 1276 Antoninus Drive Greenville, SC 29601 | | |
| Home Phone: (864) 840-3225 | Work Phone: (864) 684-0095 | Religion: Christian |
| Patient Insurance Information | | |
| Primary Insurance: Blue Cross and Blue Shield | Name of Insuree: Frank Davidson | |
| Group Number: 2289765 | Insuree's Date of Birth: 8/25/1946 | |
| Emergency Contact Information | | |
| Contact Name: Miranda Price | Home Phone: (864) 938-2857 | Work Phone: (864) 454-7734 |
| Home address: 4687 Stride Drive Greenville, SC 29602 | | |
| Vital Signs | | |
| Blood Pressure: 130/83 | Respiration: 9 | |
| Pulse: 86 | Temperature: 97.9 F | |
| Patient Allergies: Pet dander, Pollen | Adverse Drug Reactions: Sulfa drugs causes mild skin rash | |
| Current Medications: Insulin, Diabinese | | Self Administrations of Medications: No |
| Primary Medical Diagnosis: Admitted for: elevated blood sugar levels and dehydration. Diagnosed with: Type II Diabetes. Current status: blood sugar levels have decreased and he has been treated for his dehydration. | | |

Database Management

- DBMS engine
- Data definition subsystem
- Data dictionary or schema

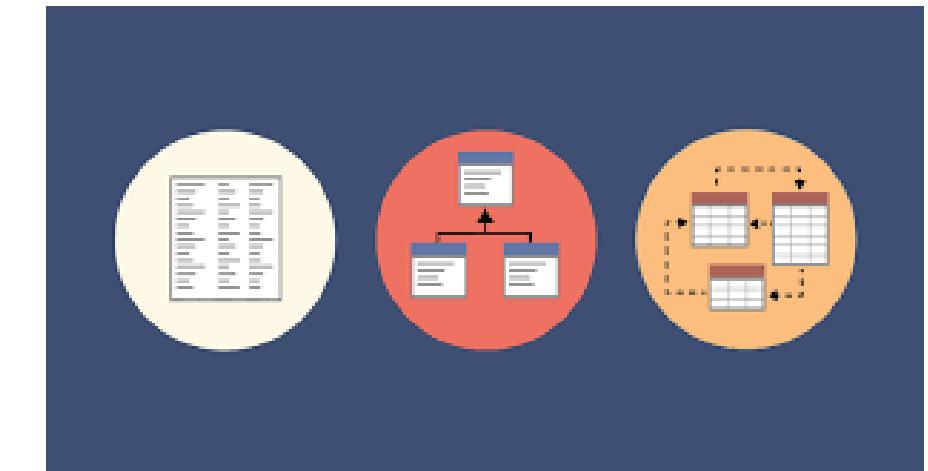


Database Management (Continue)

- Data manipulation subsystem
 - Query-by-example
 - Structured Query Language (SQL)
- Application generation subsystem
- Data administration subsystem
 - Database Administrators (DBAs)
 - Processing rights

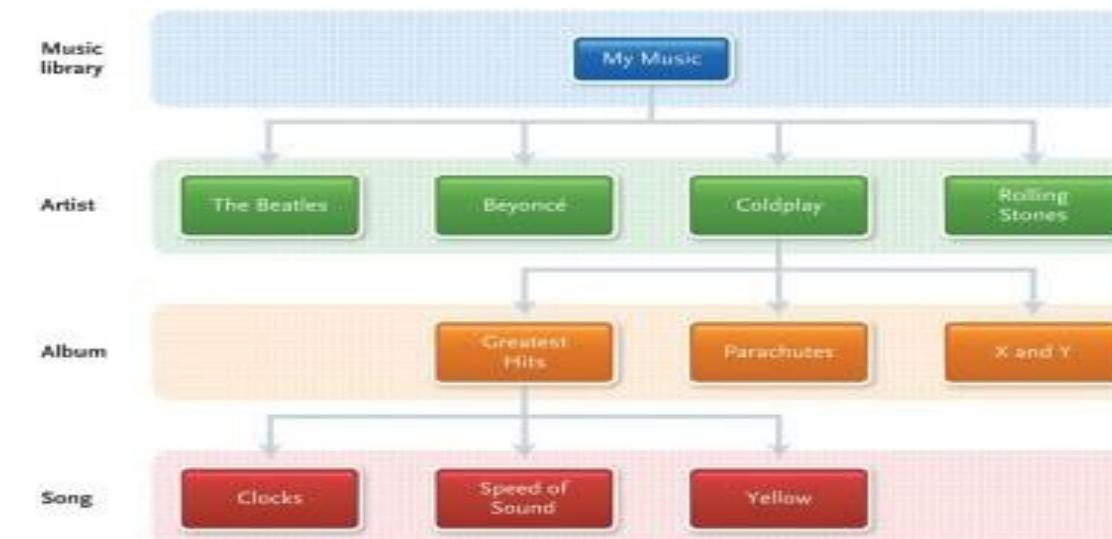
DBMS Structure

- Database model:
 - ❑ DBMS programs work with data that is logically structured or arranged
 - ❑ Model defined rules and standards for data in a database
- Five common data models:
 - ❑ Hierarchical database
 - ❑ Network database
 - ❑ Relational database
 - ❑ Multidimensional database
 - ❑ Object-oriented database



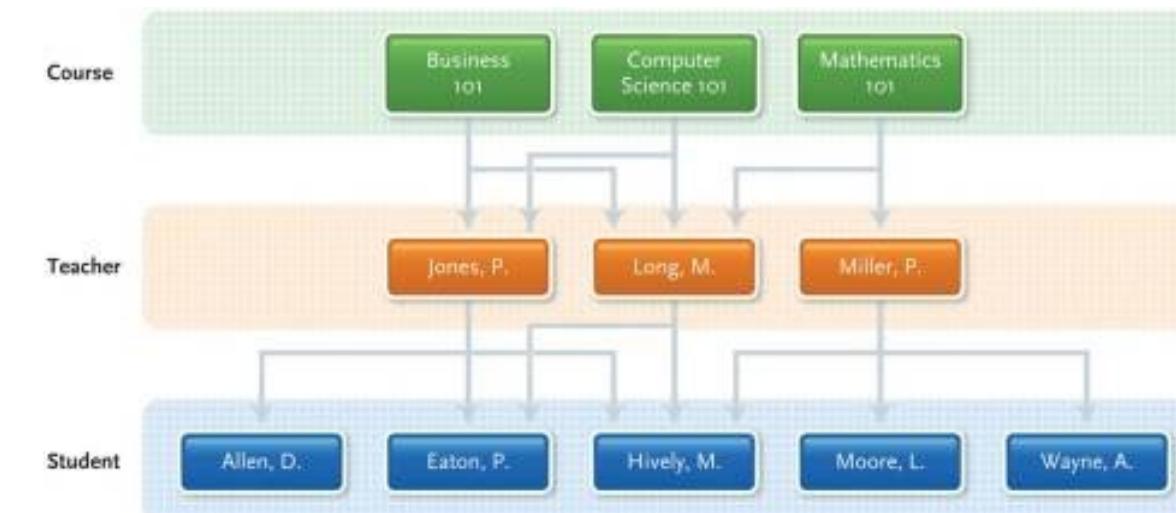
Hierarchical Database

- Fields or records structured in nodes
- Nodes
 - Points connected like branches of an upside-down tree
- One parent per node
- Parent can have several **child nodes**
 - One-to-many relationship



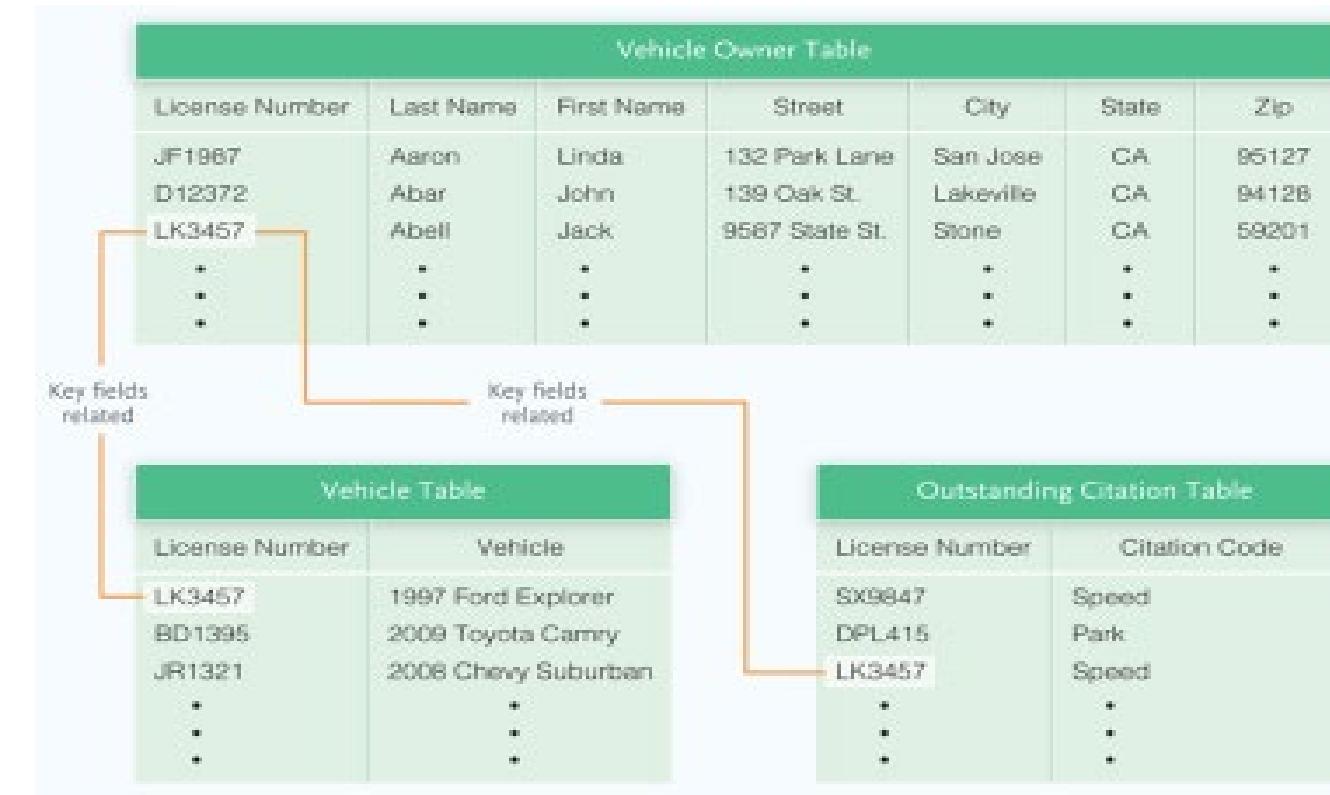
Network Database

- Hierarchical node arrangement
- Each child node may have more than one parent node (**many-to-many relationship**)
- **Pointers**
 - Additional connections between parent and child
 - Nodes can be reached through multiple paths



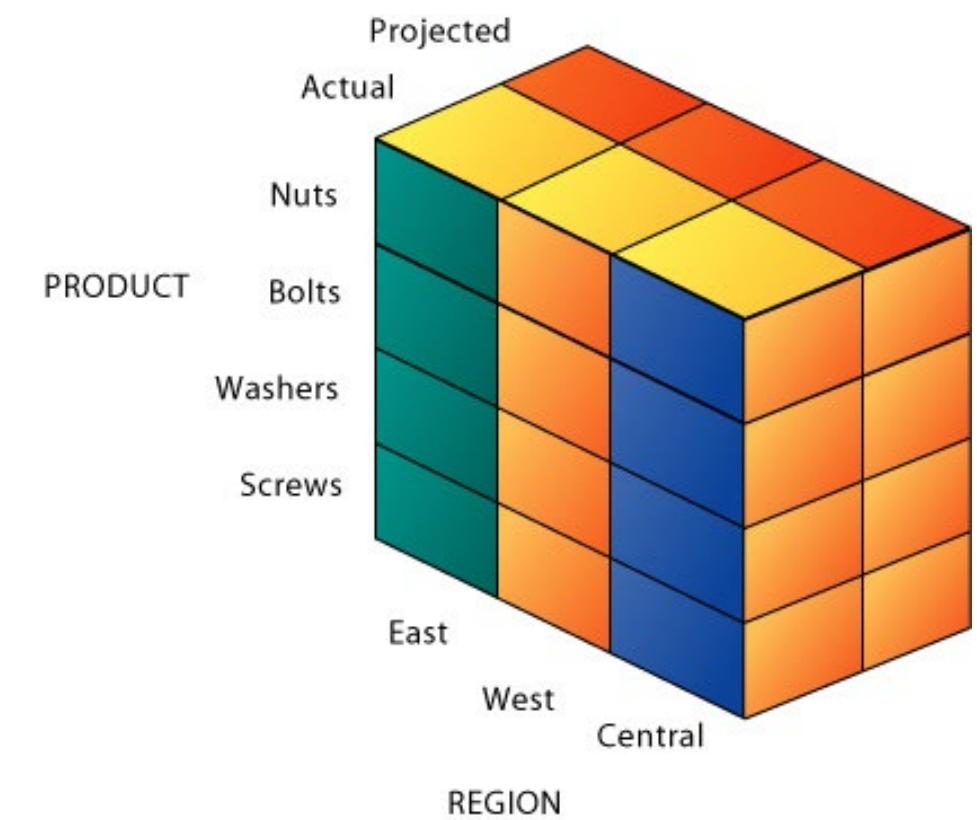
Relational Database

- More flexible
- Data stored in table called a **relation**
- Tables consist of **rows and columns**
- Tables related via a **common data item / key field**



Multidimensional Database

- A variation and an extension of the relational model to include additional dimensions, sometimes called a **data cube**
- Good for representing complex relationships
- Advantages over relational
 - Conceptualization
 - Processing speed



Object - oriented Database

- Works with unstructured data
 - Photographs
 - Audio
 - Video
- Objects contain both data and instructions
- Organize using **objects, classes, entities, attributes, and methods**



Types of Databases

- Individual
- Company or shared
- Distributed
- Commercial

| Type | Description |
|-------------|--|
| Individual | Integrated files used by just one person |
| Company | Common operational or commonly used files shared in an organization |
| Distributed | Database spread geographically and accessed using database server |
| Commercial | Information utilities or data banks available to users on a wide range of topics |

Types of Databases (Continue)

Relational Databases

The most popular type of database, used for storing structured data. Examples include MySQL, Oracle, and Microsoft SQL Server.

NoSQL Databases

Used for storing semi - structured and unstructured data. Examples include MongoDB, Cassandra, and Amazon DynamoDB.

Graph Databases

Used for storing interconnected data, such as social networks, recommendation engines, and fraud detection systems. Examples include Neo4j and Amazon Neptune.

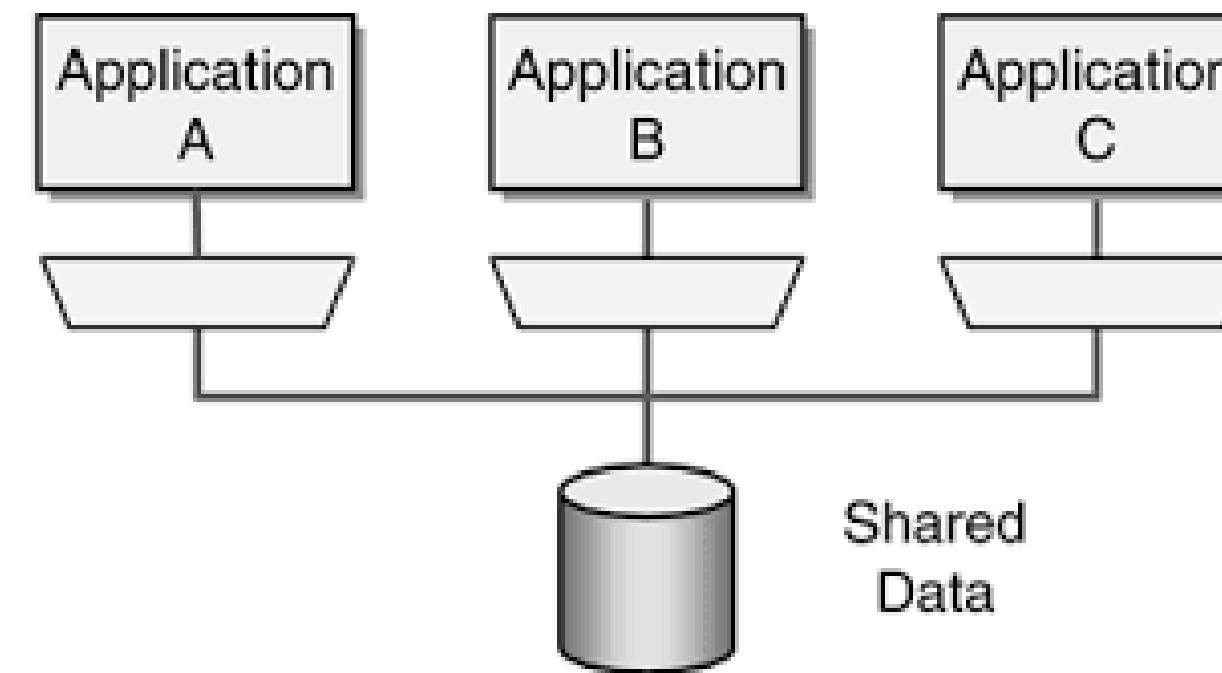
Individual Databases

- Also called a **microcomputer database**
- Integrated file collection for one person usually under the person's direct control
- Generally stored on the user's hard -disk drive or on a LAN file server



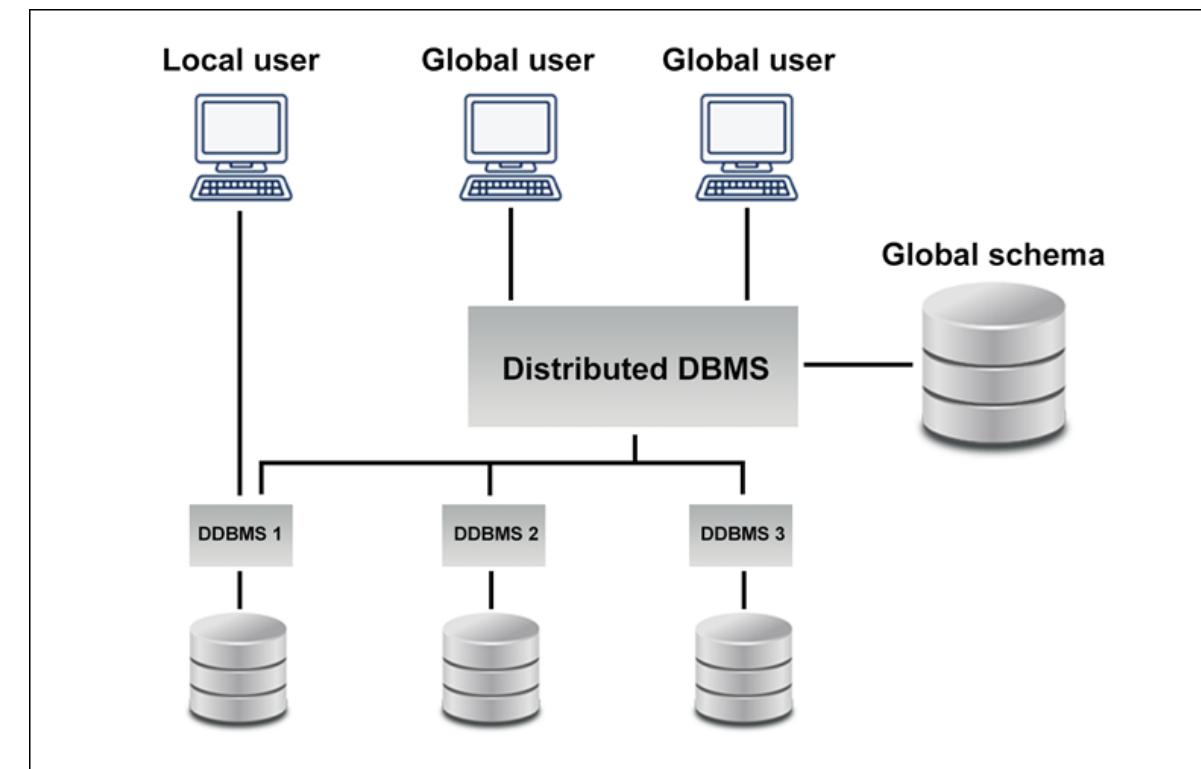
Company or Shared Databases

- Usually stored on a **central database server** and managed by a **database administrator**
- Users throughout a company can access the database through the **company's networks**



Distributed Databases

- Database is located in a place or places other than where users are located
- Typically, database servers on a client/server network provide the link between users and the distant data



Commercial Databases

- Enormous database developed by an organization to cover particular subjects
- Access is offered to the public or selected individuals for a fee
- Most designed for organizational and individual use
- Also referred to as **information utilities** or **data banks**



Database Uses and Issues

- Strategic uses
 - Special type of database called **data warehouse**
 - **Data mining** is used to search databases for information and patterns
- Security
 - Databases are valuable
 - Protection necessary



Security:
Electronic
fingerprint scanner

Careers in IT

- Database administrators
 - Determine the most efficient ways to organize and access a company's data
 - Responsible for database security and backing up the system
- Employers look for
 - Bachelors degree in Computer Science
 - Technical experience
- Database administrators can expect to earn \$48,500 to \$85,000 annually



Part 2: Introduction to Data Analytics

What is Data Analytics?

Data Analysis vs. Data Analytics vs. Data Science

Use of Big Data in Data Analytics

Data Analytics Types

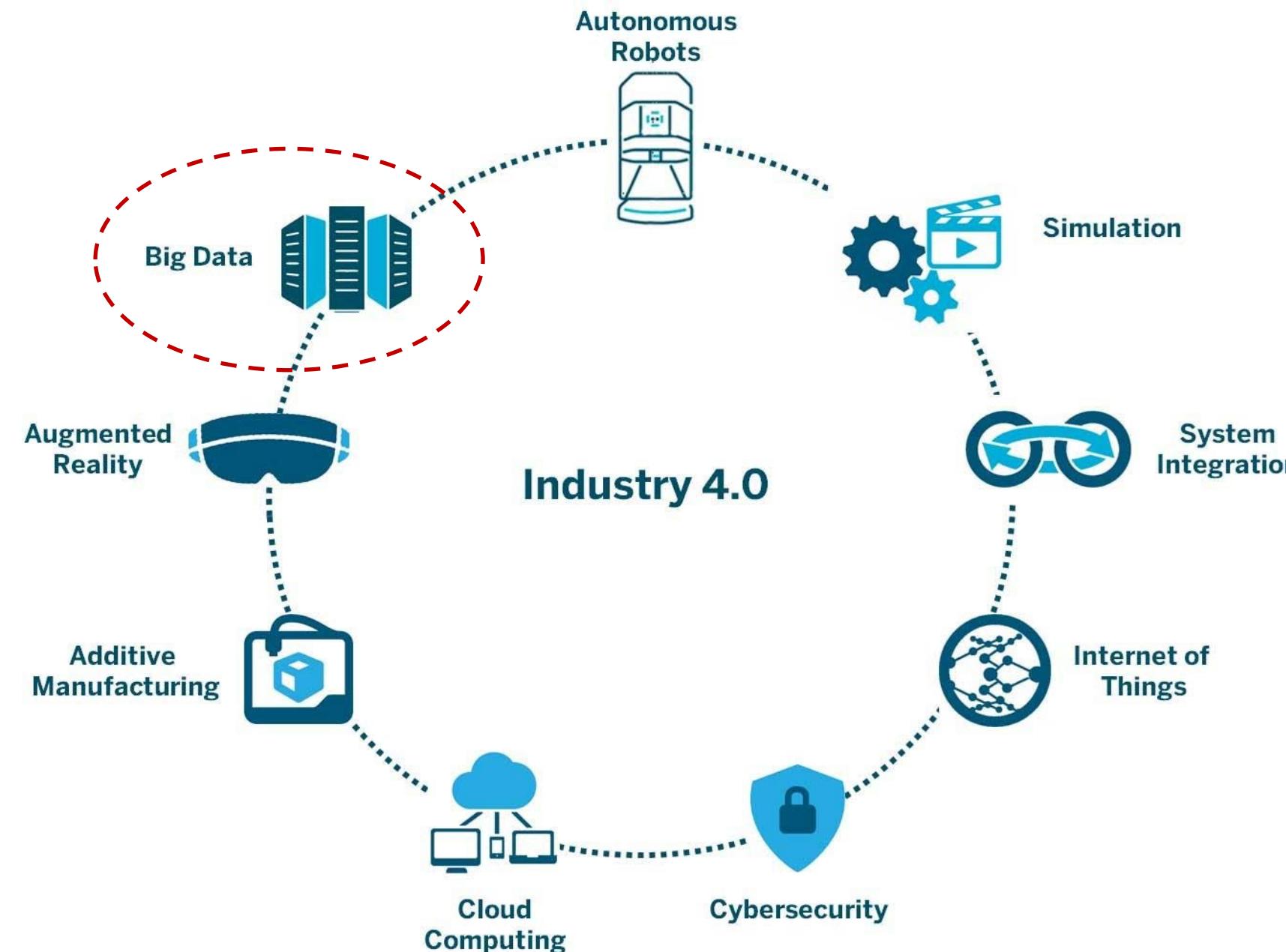
Data Analytics Techniques

Process of Data Analytics

Data Visualization & Data Warehousing

Role of Data Analyst in the Business

Where are Big Data Analytics in IR4.0 Technologies?



What is Data Analytics?

BIG DATA ANALYTICS

WHAT IS Data Analytics?

A series of techniques aimed at extracting relevant and valuable information from extensive and diverse sets of data gathered from different sources and varying in sizes

For examples:

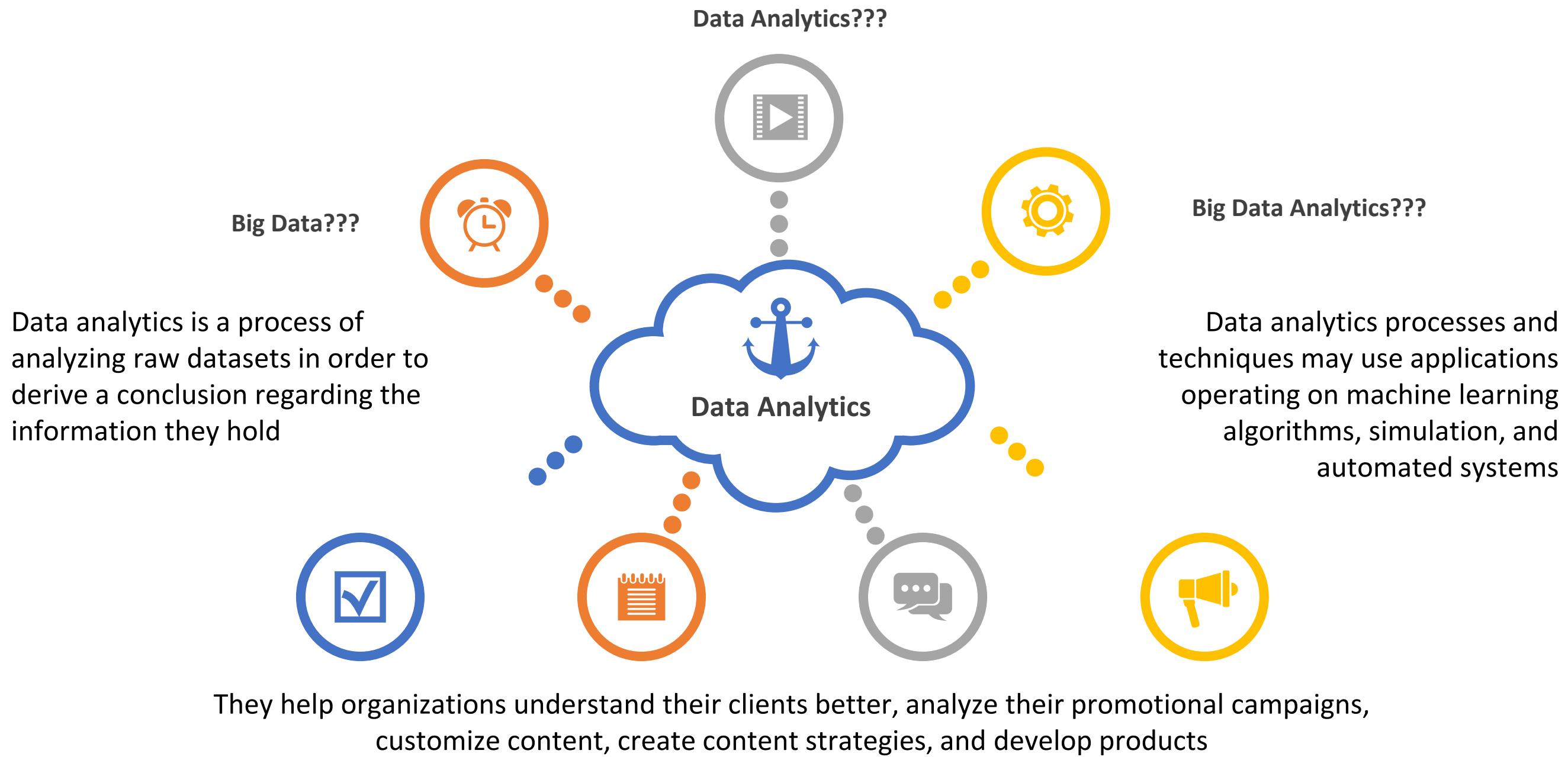
- content preferences
- different types of interactions with certain kinds of content or ads
- use of certain features in the applications
- search requests
- browsing activity
- online purchases

10100110100100001010100
111101110110110110101
00001110010101100101010
01110101000101010001011
01011011011010001010111
00010101000101000101110
10110001001101001101001
00001010100111101110110
11011010101000011100101
01100101010011101010001
01010001011010110110110
1001



Source:
<https://theappsolutions.com/blog/development/what-is-big-data-analytics/>

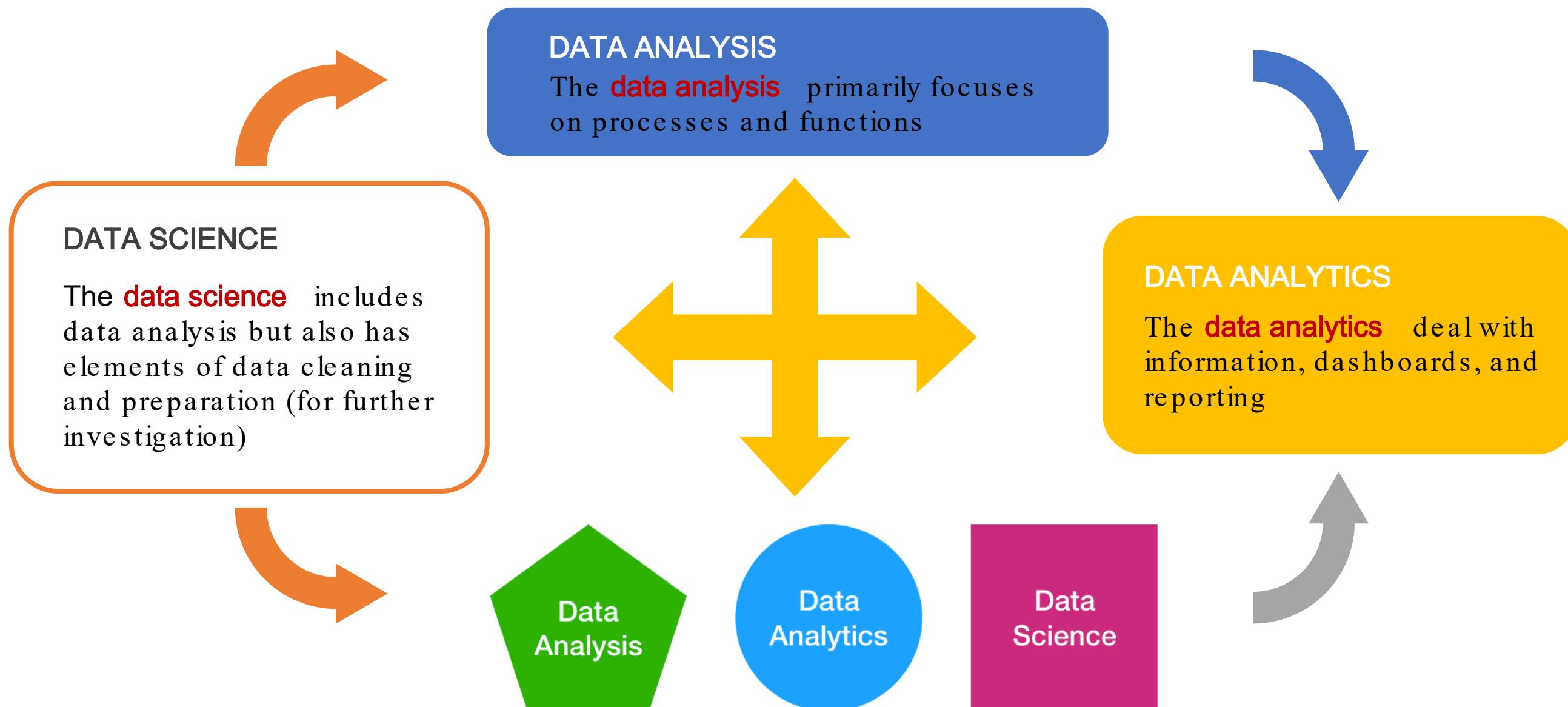
What is Data Analytics?





Data Analysis vs. Data Analytics vs. Data Science

Data Analysis vs. Data Analytics vs. Data Science



Big Data and Data Analytics

1 Introduction

Big data refers to large volumes of structured and unstructured data that cannot be processed using traditional database and analytics tools.

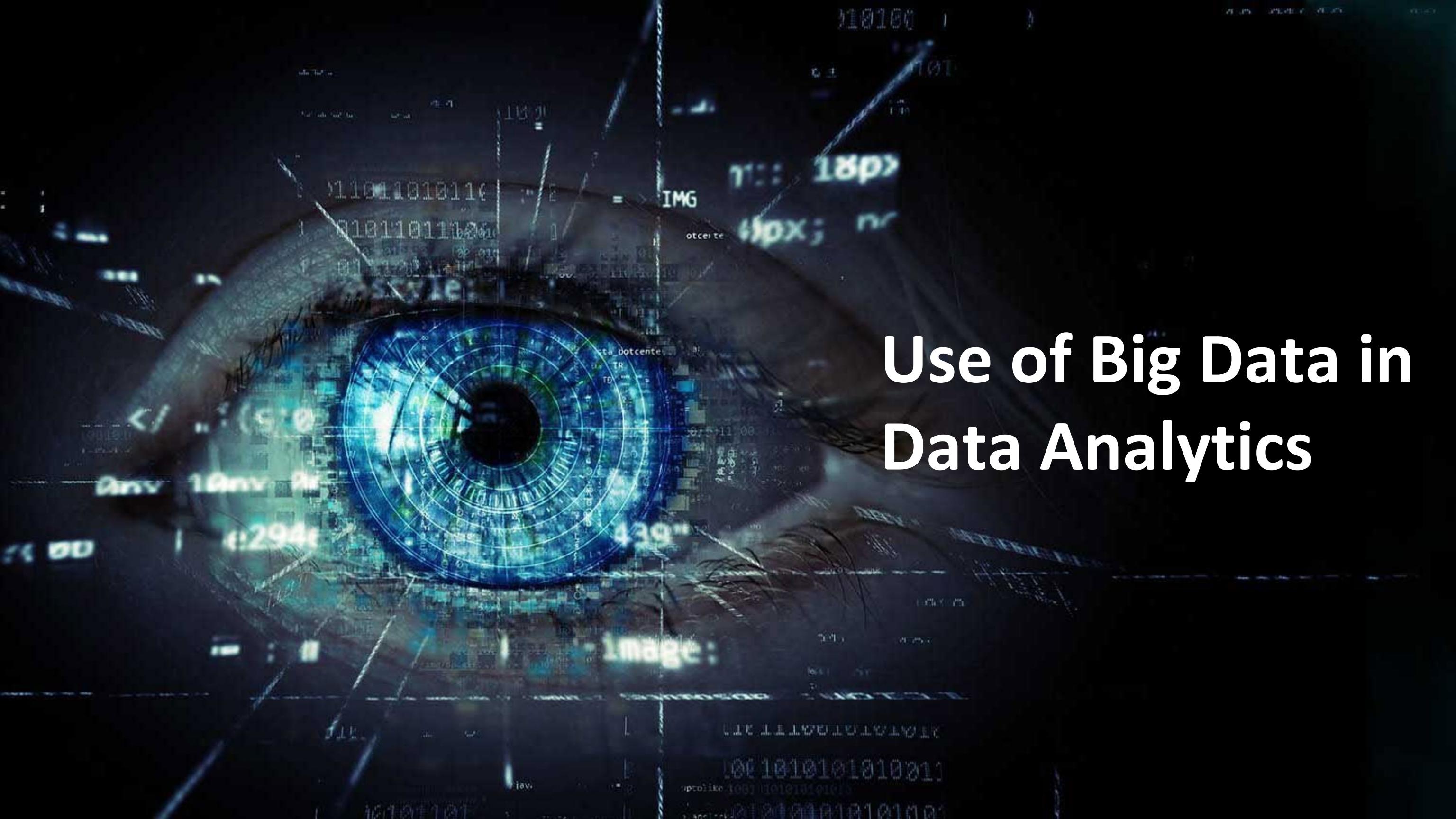
2 Challenges and Opportunities

Big data comes with challenges such as data quality, privacy, security, and scalability, but also provides opportunities for innovation and competitive advantage.

3 Technology and Tools

Big data technologies and tools include Hadoop, Spark, NoSQL databases, data lakes, and cloud services such as AWS and Azure.

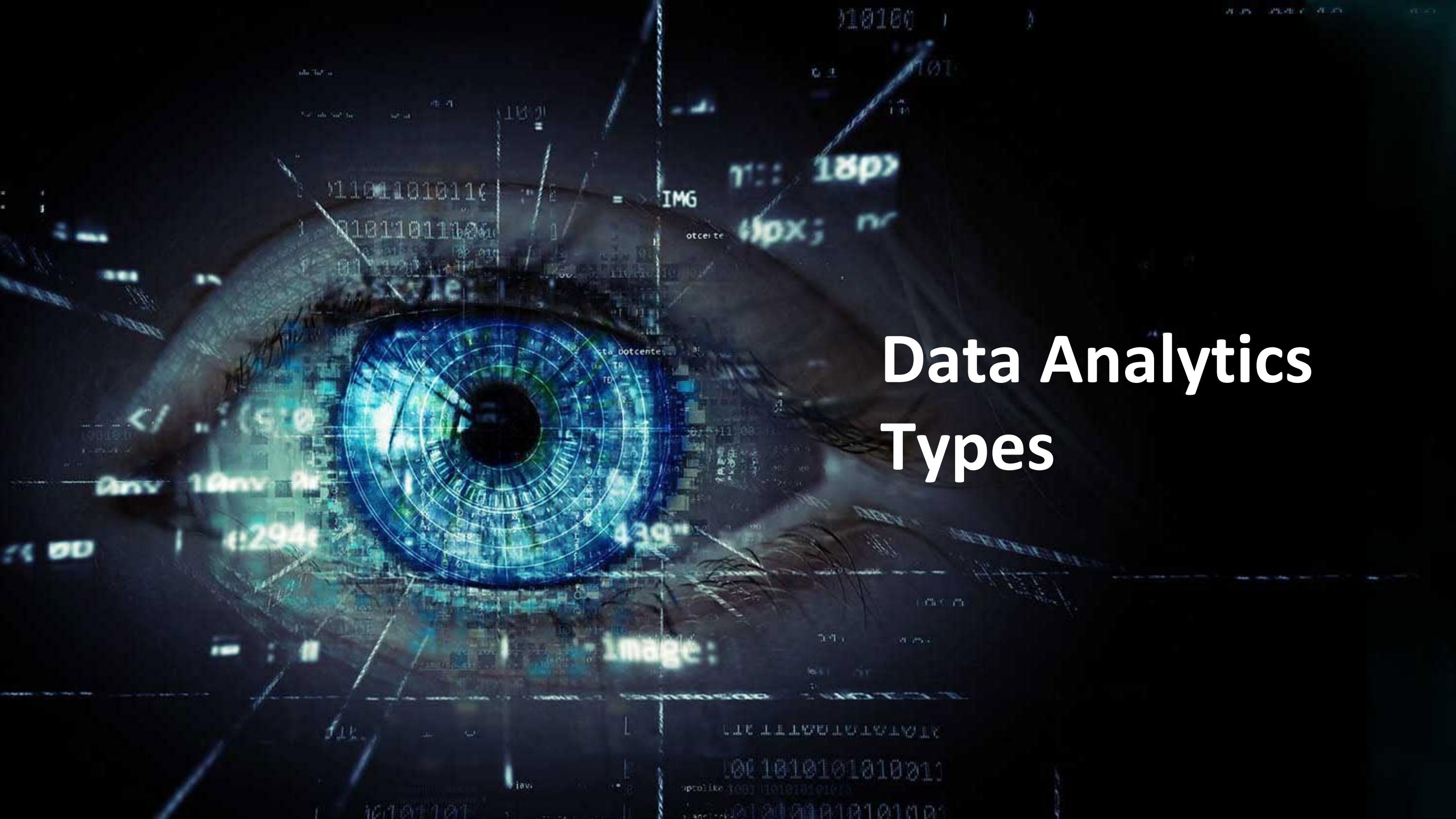
Use of Big Data in Data Analytics



Use of Big Data in Data Analytics



Data Analytics Types



Data Analytics Types



Descriptive Analytics

describes the happenings over time, such as whether the number of views increased or decreased and whether the current month's sales are better than the last one



Diagnostic Analytics

focuses on the reason for the occurrence of any event. It requires hypothesizing and involves a much diverse dataset. It examines data to answer questions, such as "Did the weather impact the selling of umbrella?" or "Did the new ad strategy affect sales?"



Predictive Analytics

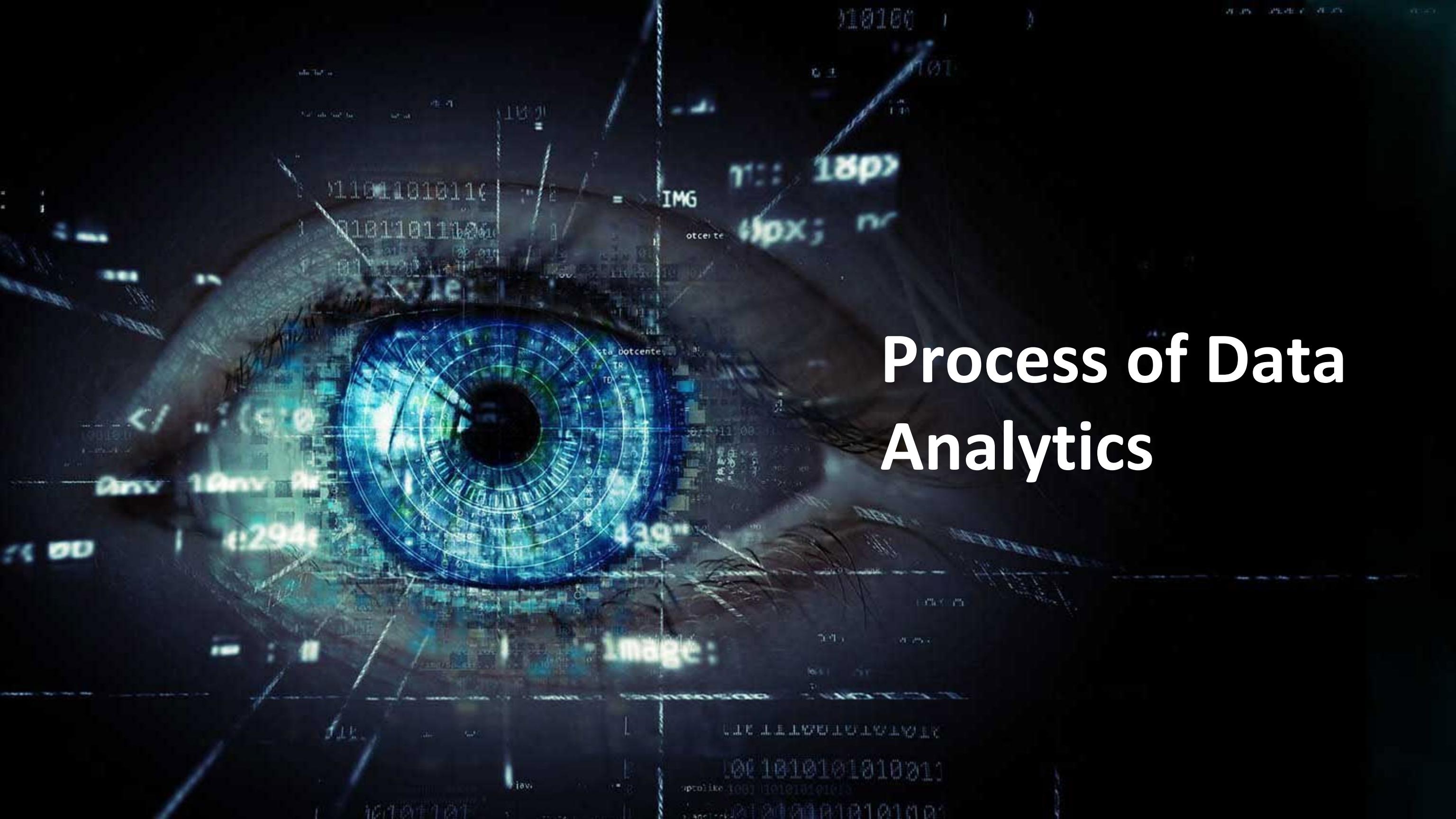
focuses on the events that are expected to occur in the immediate future. Predictive analytics tries to find answers to questions like, what happened to the sales in the last hot summer season? How many weather forecasts expect this year's hot summer?



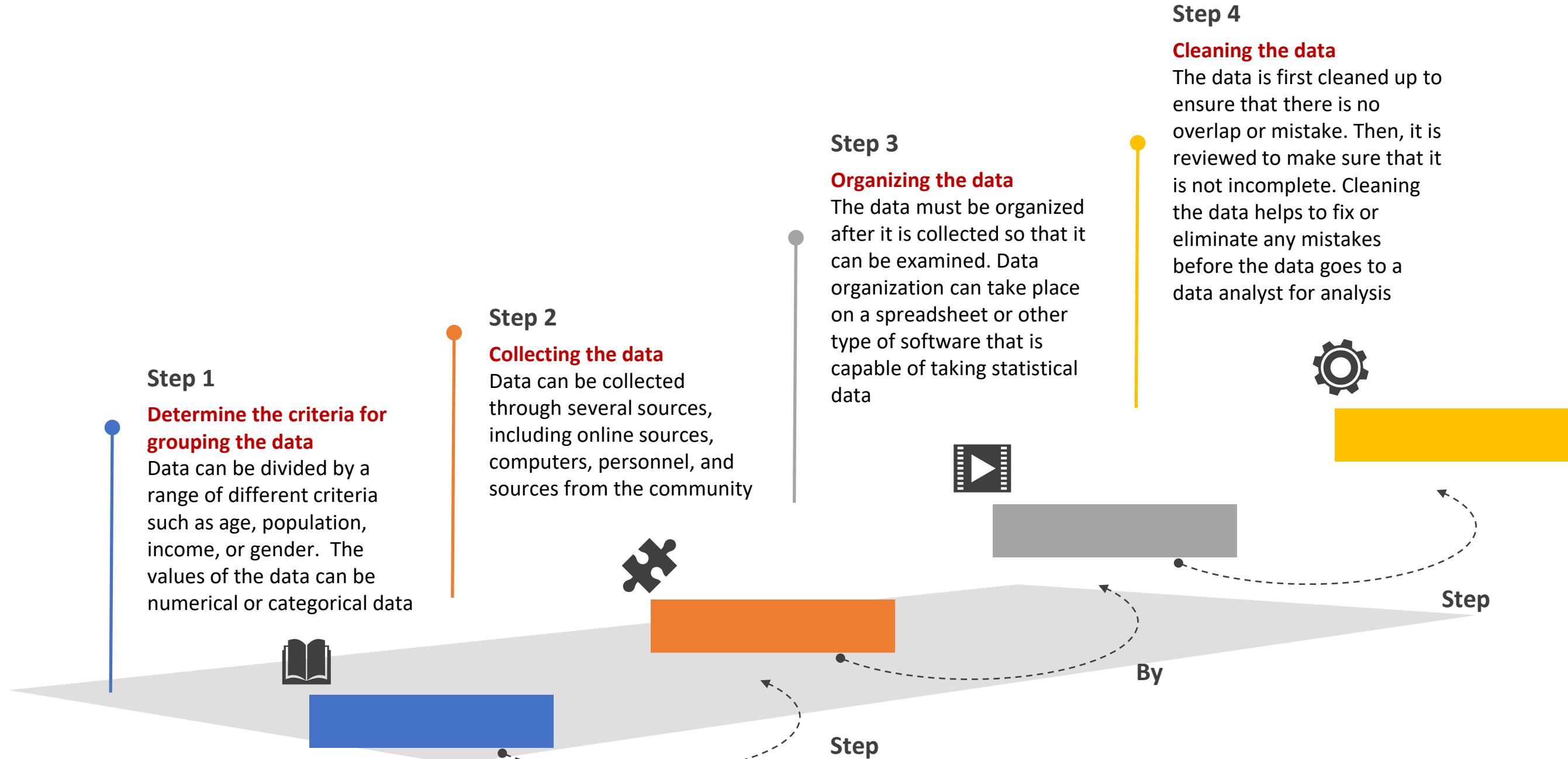
Prescriptive Analytics

indicates a plan of action. If the chance of a hot summer calculated as the average of the five weather models is above 58%, other than an umbrella, a rain coat should be considered to maximize the production

Process of Data Analytics



Process of Data Analytics



Data Visualization & Data Warehousing

Data Visualization

Importance

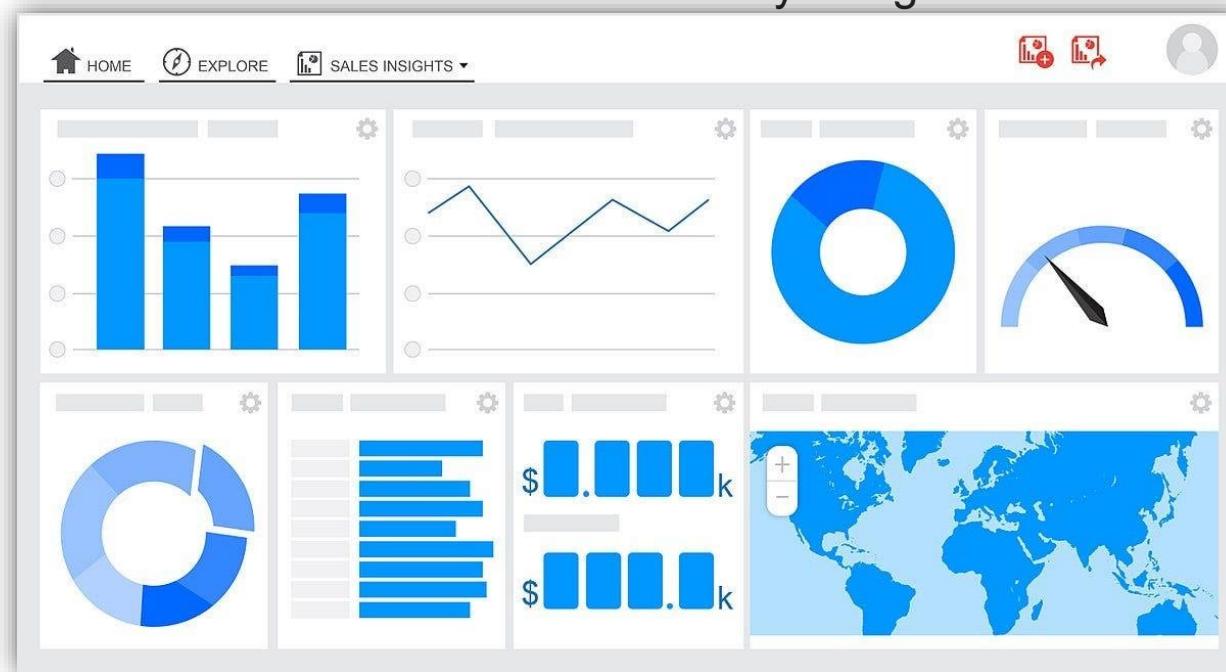
Data visualization helps turn complex data into insights and communicate them effectively to stakeholders.

Types of Tools

Data visualization tools can range from simple charting libraries to more advanced tools that allow for interactive dashboards and storytelling.

Best Practices

Effective data visualization requires understanding your audience, choosing the right type of visualization, using appropriate colors and labels, and avoiding clutter and complexity.



Data Warehousing



Definition and Purpose

Data warehousing is the process of storing and managing large volumes of data from different sources to support business decision-making.



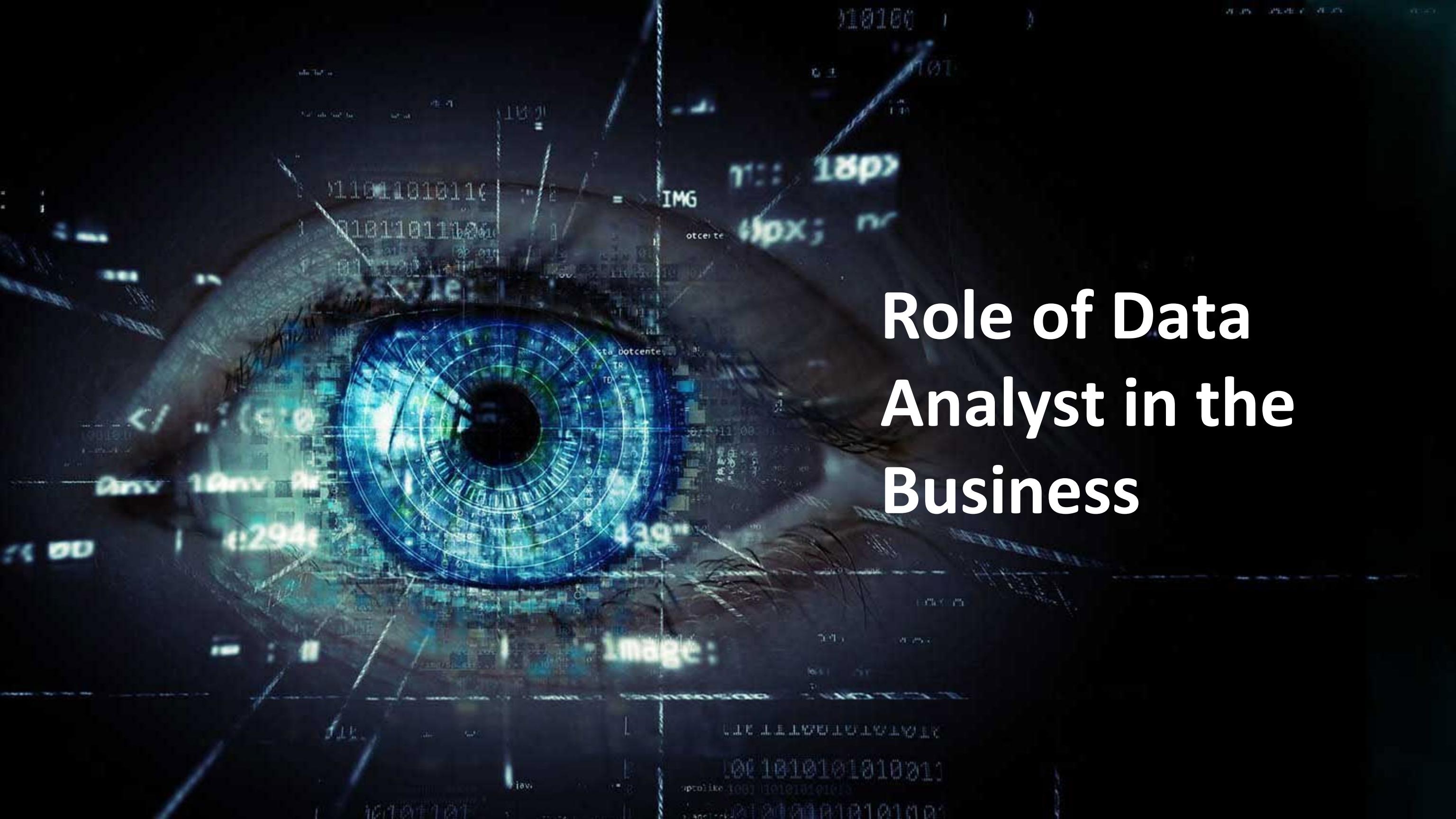
Extract, Transform, and Load (ETL) Process

The ETL process involves extracting data from various sources, transforming it into a consistent format, and loading it into a data warehouse.

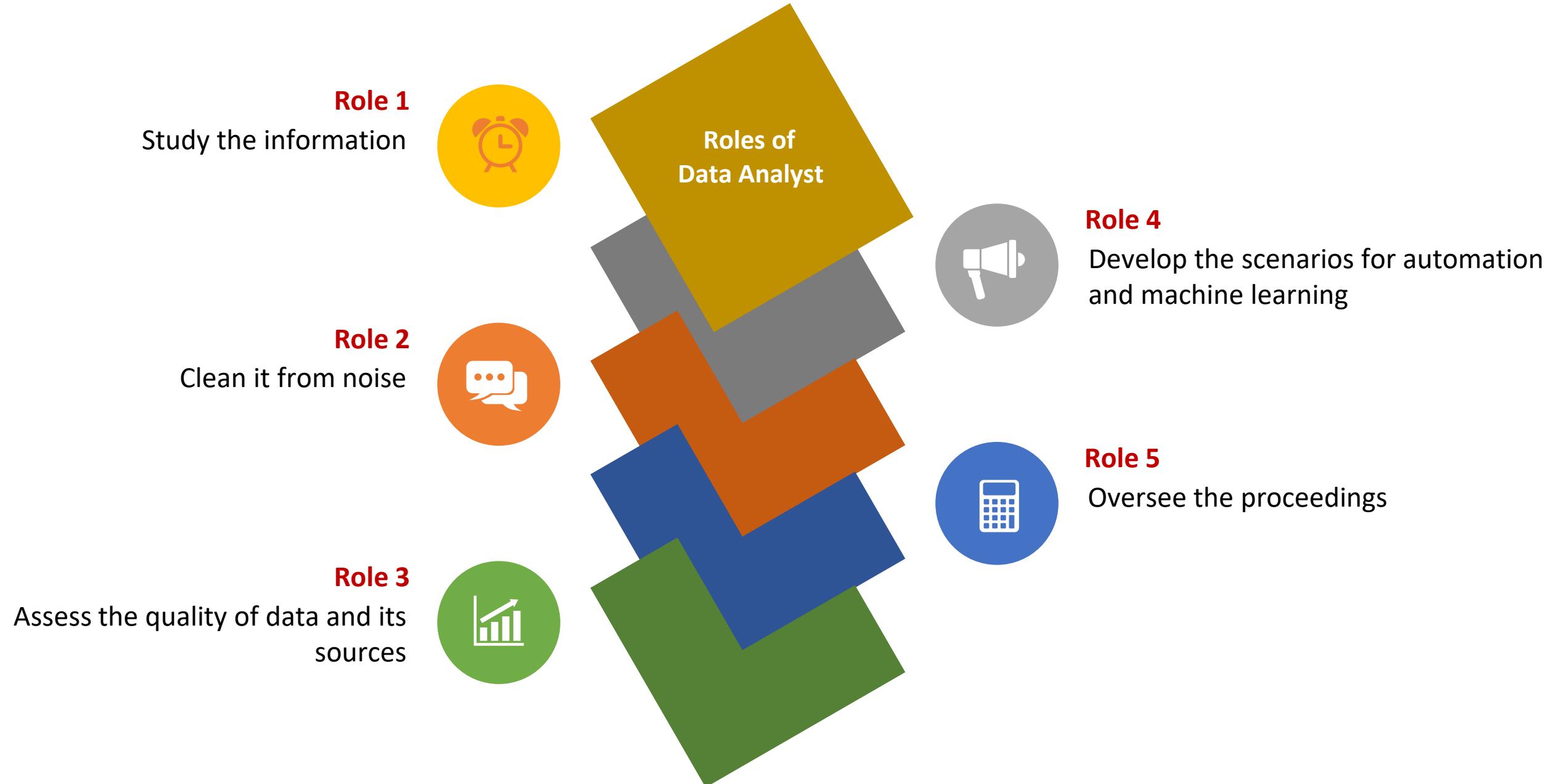
Benefits and Challenges

Data warehousing provides a centralized repository of information that can be used for analytics and reporting, but it also comes with challenges such as cost, complexity, and data integration.

Role of Data Analyst in the Business



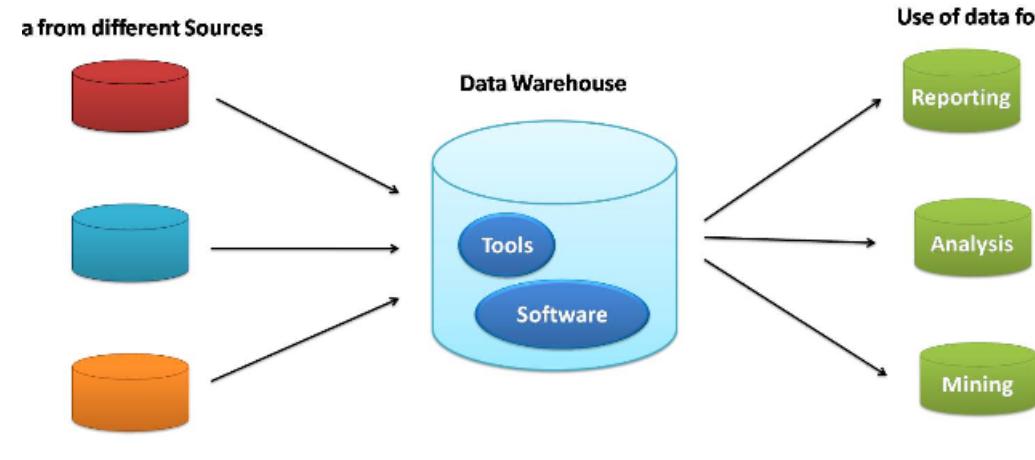
Role of Data Analyst in the Business



Data Scientist vs Data Analyst

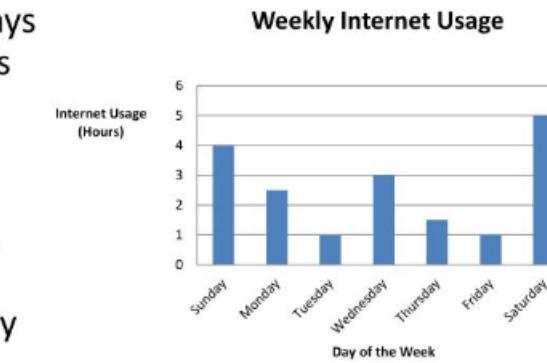
| Data Scientist vs Data Analyst | | |
|--|--|--|
| Features | Data Scientist | Data Analyst |
|  Background | A Data Scientist deals with various data operations. | A Data Analyst's role is related to data cleaning, transforming and generating inferences from data. |
|  Scope | Involved with several underlying data procedures | Involvement is limited to small data and static inferences. |
|  Type of Data | Handles structured & unstructured data | Deals with structured data only |
|  Skills | Possesses knowledge of mathematics, statistics & machine learning algorithms | Has problem solving skills, knowledge of basic statistics |
|  Tools | Proficient in SAS, Python, R, TensorFlow, Hadoop, Spark | Knows Excel, SQL, R (in some cases), Tableau |

Conclusion



Drawing Conclusions from Data

- *Go beyond analysis to find out what is behind the data (Games on Tuesdays & Fridays-so less time is spent on internet)
- *Use data to support (1 hour was spent on the internet on Tuesdays & Friday compared to at least 1.5 hours for every other day)



Key Points

Databases and data analytics are essential for modern businesses to make informed decisions. Data warehousing, visualization, and big data analytics are important components of data analytics.

Importance

The ability to effectively manage and analyze data is critical for success in today's world of information overload.