AKTU B.Tech 3rd Year

CS & CS Allied

Data Analytics & Visualization

Unit-1

in ONE SHOT

Part-I

AKTU PYQs Covered



By Shubhangi Ma'am

GATEWAY CLASSES

Gateway Classes

GATEWAY CLASSES

UNIT—1

Syllabus

GATEWAY CLASSES

Gateway Classes

Introduction to Data Analytics:

Sources and nature of data, classification of data (structured, semi-structured, unstructured), characteristics of data, introduction to Big Data platform, need of data analytics, evolution of analytic scalability, analytic process and tools, analysis vs reporting, modern data analytic tools, applications of data analytics.

Data Analytics Lifecycle:

Need, key roles for successful analytic projects, various phases of data analytics lifecycle – discovery, data preparation, model planning, model building, communicating results, operationalization.

Important Previous Year Questions

Q.1 Define Data Analytics and Explain its importance in decision-making. AKTU(2025-26)
Q.2 Differentiate between structured , semi-structured and unstructured data. AKTU(2025-26)
Q.3 Explain sources and characteristics of data with suitable examples. AKTU(2025-26)
Q.4 Analyze the role of Data Analytics in modern organization. AKTU(2025-26)
Q.5 What are the challenges and solutions associated with scaling analytics in modern organization. AKTU(2024-25)
Q.6 How do modern tools bridge the gap between analysis and reporting? AKTU(2024-25)
Q.7 Describe the key characteristics of Big Data . How do these characteristics necessitate the development of advanced analytics platforms? AKTU(2024-25)
Q.8 How has the need for data analytics evolved with the advent of IoT and real-time systems ? Provide relevant Examples . AKTU(2024-25)
Q.9 what are the features of Big Data ? AKTU(2024-25)
Q.10 Explain the need of Data Analytics. AKTU(2024-25)
Q.11 Discuss Classification of Data. Explain all categories in detail with examples.
Q.12 Write short note on the following : AKTU(2023-24)

Modern Data analytics tools
Application of Data analytics

Q.13 List the features of big data analytics platform. AKTU(2023-24)
Q.14 why there is need of data analytics ? AKTU(2023-24)
Q.16 discuss all the modern tools of Data analytics . AKTU(2022-23)
Q.17what is Data analytics ? AKTU(20225-26)
Q.18List any characteristics of big Data . AKTU(2025-26)
Q.19 Differentiate between Predictive and Prescriptive Data Analytics . AKTU(2024-25)
Q.20 Define the term data lake, data base , data warehouse . AKTU(2024-25)
Q.21 Explain the different categories of data analytics with examples. AKTU(2024-25)
Q.22 How have Advancements in technology contributed to the scalability of analytics ? AKTU(2024-25)
Q.23 What are the sources of data in the context of data analytics? AKTU(2024-25)
Q.24 Describe the characteristics of data that are relevant in the field of data analytics . How do these characteristics impact the analysis process ? AKTU(2022-23)
Q.25 Differentiate between analysis and reporting in the context of data analytics. How do these two aspects contribute to the overall understanding of data? AKTU(2022-23)
Q.26 Explore modern data analytics tools and their functionalities . How have these tools transformed the landscape of data analytics? AKTU(2022-23)

GATEWAY CLASSES
Gateway Classes

What is Data ?

- Data is a collection of raw facts and figures that do not have any meaning on their own until they are processed.
- **Data = unorganized information**

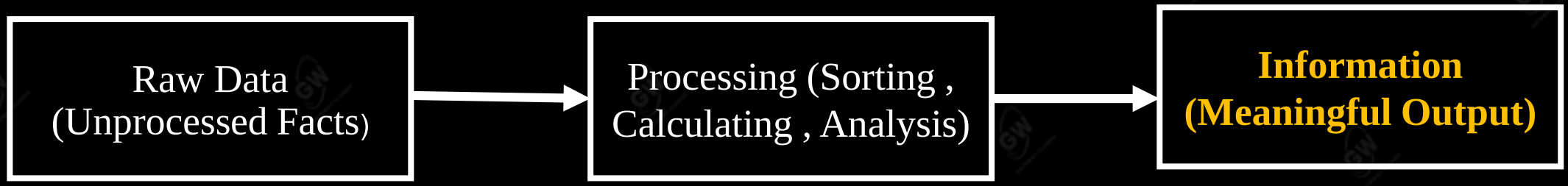
  It can be numbers, text, images, videos, etc.
- **Data:**

  Temperature readings of a city

  → 30°C, 32°C, 35°C, 33°C

Gateway Classes

“Individual temperature values are just data, but when we analyze them and understand the trend, it becomes information.”



Information

First we collect raw data, then we process it, and finally we get meaningful **information.”**

Digital Data

- Digital data refers to any data that is created, stored, and processed using digital devices and technologies such as smartphones, computers, and the internet.
- In today’s modern world, most of the data we generate through mobile phones, apps, and online platforms is digital data.

Examples :

a. Photos

b. Social media posts and messages

c. Online transactions

Gateway Classes

Sources of Data AKTU(2025-26)

- **Sources of Data** are the places from which data is collected for analysis in Data Analytics.
- Data Analytics depends on data collected from different sources. These sources provide the raw data that is processed and analyzed to generate useful insights and support decision-making.

Gateway Classes

Types of Data Sources :

1. **Internal Sources**
2. **External Sources**

Internal Sources	External Sources
Data collected from within an organization.	Data collected from outside an organization.
Example : • Sales records • Employee records • Customer database • Inventory records • Financial reports	**Example :** • Social media • Government reports • Market research • Websites • Weather data

Gateway Classes

Example :

An online shopping company collects data from:

- Customer orders
- Product reviews
- Website visits
- Payment records

The company analyzes this data to understand customer behavior and recommend products.

Nature of Data AKTU(2024-25)

- **Nature of Data** refers to the **type, form, and characteristics of data** used in **Data Analytics**.
- It explains how data is collected, represented, organized, and analyzed to generate useful insights.

Nature of Data

The nature of data in Data Analytics can be understood as follows :

1. **Qualitative Data**
2. **Quantitative Data**
3. **Structured Data**
4. **Semi-Structured Data**
5. **Unstructured Data**

Gateway Classes

Nature	Description
1. Qualitative Data	Qualitative data describes **qualities or characteristics**. It is non-numerical and cannot be measured using numbers. **Ex :** Gender ,Customer feedback
2. Quantitative Data	Quantitative data consists of numbers. It can be measured, counted, and analyzed mathematically. **Ex** : Age, Salary, Marks etc.
3. Structured Data	Structured data is organized in rows and columns and is easy to store and analyze. **Ex:** Student records , Employee Database etc.
4. Semi-Structured Data	Semi-structured data does not have a fixed table format but contains some organizational structure. **Ex:** XML files, Emails etc.
5. Unstructured Data	Unstructured data has no predefined format and is difficult to analyze using traditional databases. **Ex:** Images , videos etc.

GATEWAY CLASSES
Gateway Classes

GATEWAY CLASSES

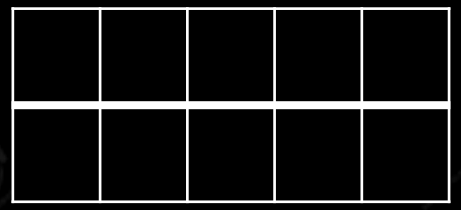
Classification of Data

AKTU (2024-25) / (2023-24)/ (2024-25)

GATEWAY CLASSES

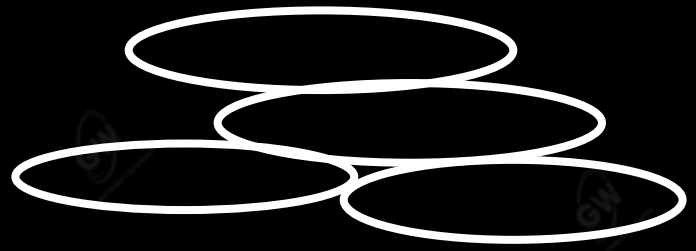
Structured

Easily displayed in Relational Databases , Tables , Rows , Columns.



Unstructured

The Data isn’t Structured therefore it Can’t be displayed in Rows & Columns.



Semi – Structured

Sometimes they look Structured & Sometimes they look Unstructured.



Gateway Classes

1. Structured

- Any data that can be **stored , accessed , and processed in the form of fixed format** is termed as a '**Structured**' data.
- It is well-organized and follows a clear structure, usually in the form of **rows and columns.**
- Each field in the data has a defined data type and meaning. Because of its organized nature, it is easy to search, filter, and analyze.
- This type of data also supports fast processing and efficient querying using standard tools like SQL.
- Even when the size of such data becomes very large, it is still considered structured if it maintains its organized format.

Gateway Classes

GATEWAY CLASSES

An **Employee table in a Database** is an example of Structure data –

Employee ID	Employee_Name	Gender	Department	Salary
2365	Rahul	Male	Finance	65000
53398	Neha	Female	HR	70000
4214	Shweta	Female	Marketing	80000
6547	Meena	Female	Finance	65000
9076	Parul	Female	HR	70000

Gateway Classes

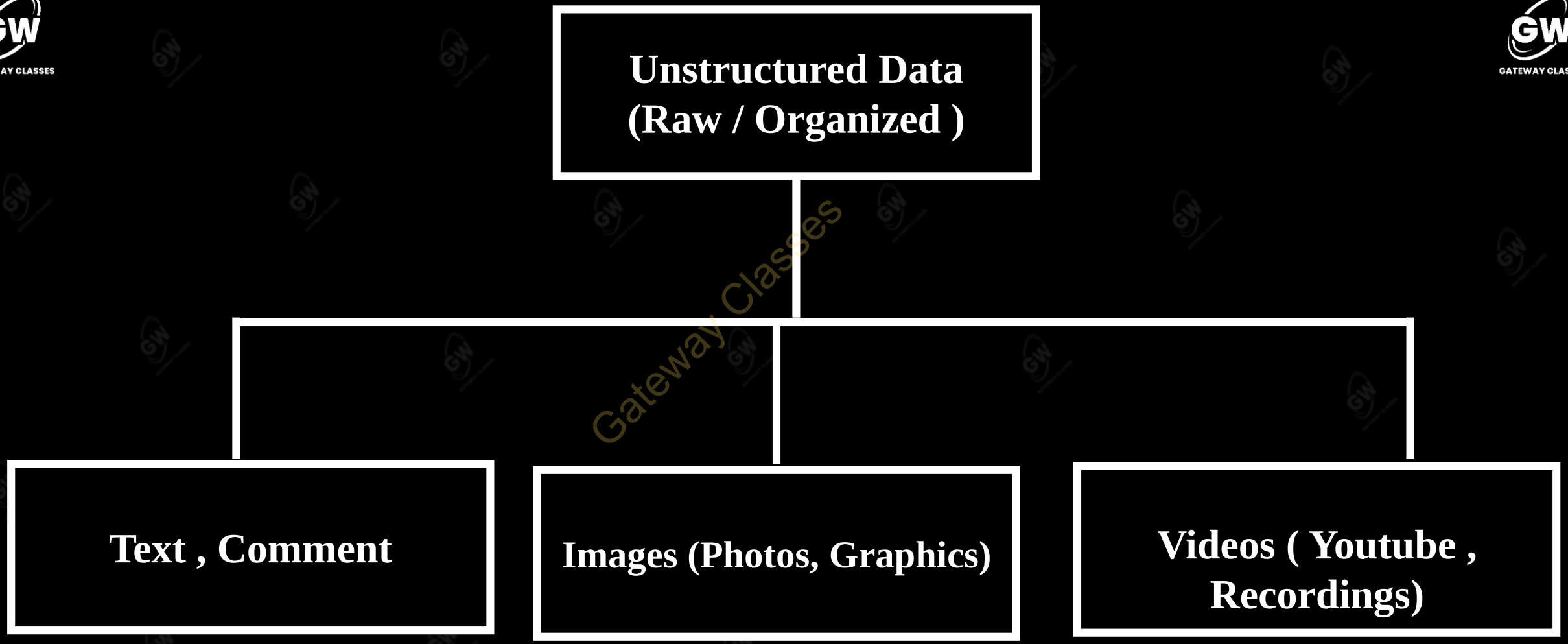
2. Unstructured

- Unstructured Big Data is data that does not have a fixed format or schema and cannot be easily stored or analyzed using traditional data processing tools.
- Unstructured Big Data does not have a specific structure or organization.
- It does not follow any fixed schema, which makes it flexible but difficult to manage.
- This type of data is usually large in volume and is generated at a high speed.
- It requires advanced tools and technologies for processing and analysis, such as Big Data analytics and machine learning techniques.

Common examples of Unstructured Big Data include **images, videos, audio files**, and social media content like posts, comments, and messages.

Gateway Classes

GATEWAY CLASSES



Gateway Classes

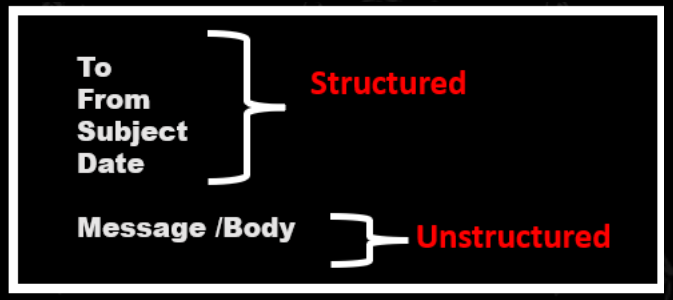
GATEWAY CLASSES

3. Semi – Structured

- Semi – Structured data can contains both the forms of data.
- Semi-structured data does not follow a rigid structure like rows and columns, but it is not completely unorganized either. It contains partial structure using tags or labels. This type of data is flexible and can handle changes easily.
- It is easier to process than unstructured data but more complex than structured data.

Gateway Classes

Emails are also a good example, as they have some structured parts like sender, receiver, and subject, along with unstructured message content.



Example : E-Mail

Characteristics of Data AKTU(2024-25)

Characteristics of Data are the important features or qualities that make data useful, reliable, and suitable for analysis in **Data Analytics**.

Gateway Classes

Characteristics	Description
1. Accuracy	Accuracy means the data should be **correct and free from errors**.
2. Completeness	Completeness means **all required data should be available**. Missing data can affect the analysis.
3. Consistency	Consistency means the data should have the **same format and values** across all systems.
4. Reliability	Government census data is more reliable than unverified online data.
8. Uniqueness	Each data record should appear **only once**. Duplicate records should be avoided.

Introduction to Big Data

"Big Data is a collection of extremely large, complex, and rapidly growing data that cannot be stored, managed, or processed efficiently using traditional data processing tools.

"Big Data" is also a data but with a huge size.

- **'Big Data'** is a term used to describe collection of data that is huge in size and yet growing exponentially with time.
- Data which are very large in size is called Big Data.
- Big Data refers to very large amounts of data that cannot be easily processed, managed, or analyzed using traditional methods.
- Big Data is extremely large and complex data that traditional systems cannot handle.
- It requires advanced technologies and techniques to handle, manage, and extract useful information from it.

Gateway Classes

Big Data = Large + Complex + Fast Data (beyond traditional processing)

Some Examples of Big Data :

1. Social Media

- Platforms like Instagram, Facebook, Twitter
- Millions of users share photos, videos, and messages daily

2. Online Shopping

- Websites like Amazon, Flipkart
- Track user searches, purchases, and preferences

3. Video Streaming

- Platforms like YouTube, Netflix
- Store and analyze millions of videos and user views

Gateway Classes

Some Important Terms AKTU(2024-25)

Database	Datawarehouse	Data Lake
A **Database** is a collection of organized data stored in tables (rows and columns). It is used to store, manage, and retrieve data easily.	A **Data Warehouse** is a large storage system that collects data from different databases and stores historical data for analysis and reporting.	A **Data Lake** is a storage system that stores **raw data** in its original form. It can store structured, semi-structured, and unstructured data.
Stores **structured data**. Data is organized in rows and columns. Easy to search, update, and delete data.	Stores **historical data**. Collects data from many databases. Used for reporting and business analysis.	Stores raw data. Handles all types of data. Very large storage capacity.
A school stores student names, roll numbers, and marks in a database.	A company collects sales data from all its branches into one data warehouse to analyze yearly sales.	Netflix stores videos, user clicks, search history, and customer reviews in a data lake.

GATEWAY CLASSES

Introduction to Big Data Platforms AKTU(2023-24)

A Big Data Platform is a combination of hardware, software, and technologies used to store, process, manage, and analyze very large and complex datasets that cannot be handled efficiently by traditional data processing systems.

Need for Big Data Platforms

- Store huge amounts of data.
- Process data at high speed.
- Support structured, semi-structured, and unstructured data.
- Handle continuously growing data.
- Help organizations make better decisions.
- Support real-time data analysis.

Gateway Classes

Features of Big Data Platforms

Features	Description
Scalability	Easily handles increasing amounts of data.
High Storage Capacity	Stores large datasets efficiently.
Fast Processing	Processes data quickly.
Reliability	Provides secure and fault-tolerant data storage.
Distributed Processing	Uses multiple computers to process data simultaneously.
Cost-Effective	Reduces processing and storage costs.

GATEWAY CLASSES
Gateway Classes

Examples of Big Data Platforms :

Apache Kafka	Used for data ingestion (collecting real-time data)
Apache Hadoop HDFS	Used for storing large data across multiple machines
Apache Hadoop MapReduce	Used for processing batch processing of large datasets
Apache Spark	Used for fast and real-time data processing
Apache Hive	Used for querying & Analyzing data using for SQL

Gateway Classes

Applications of Big Data Platforms

- Healthcare
- Banking and Finance
- E-commerce
- Social Media
- Education
- Transportation
- Weather Forecasting

Characteristics of Big Data / 5 V's of Big Data

AKTU (2022 23)/(2024-25)

Big Data is defined by five important characteristics known as the **5 V's**. These help us understand the nature and challenges of handling big data.

The 5 V's are:

1. **Volume**
2. **Velocity**
3. **Variety**
4. **Veracity**
5. **Value**

Gateway Classes

1. Volume

Volume refers to the **large amount of data** that is generated every day. In today's digital world, organizations collect data from many sources like social media, websites, machines, and business systems. This data can be in huge size such as terabytes or petabytes. Because of this large volume, special storage and processing systems are required.

Example:

Social media platforms like Instagram or Facebook generate millions of posts, images, and videos daily.

Gateway Classes

2. Velocity

Velocity refers to **how fast data is generated and how quickly it moves and is processed**. Organizations need data to flow quickly so that they can take decisions on time. Data is continuously coming from sources like mobile apps, sensors, and social media, and it often needs to be processed in real-time or near real-time.

Example:

In online payments (UPI), transactions happen instantly and must be processed immediately.

Gateway Classes

3. Variety

Variety refers to the **different types and formats of data**. Data is not only in tables but also in many other forms like text, images, videos, emails, and JSON files. This makes data handling more complex.

Example:

A company collects customer data (structured), emails (semi-structured), and social media posts or videos (unstructured).

Gateway Classes

4. Veracity

Veracity refers to the **quality and accuracy of data**. Sometimes data can be incorrect, incomplete, or duplicate, which affects analysis.

Organizations need to clean and verify data before using it.

Example:

Fake reviews on e-commerce websites or duplicate customer records reduce data reliability.

Gateway Classes

5. Value

Value refers to the **usefulness of data in making decisions**. Collecting data is not enough; it should provide meaningful insights that help organizations improve their performance.

Example:

E-commerce websites analyze user data to recommend products and increase sales.

GatewayClasses

Importance of 5 V's of big data

1. Importance of Volume

- Volume helps organizations understand the **amount of data they need to handle**.
- It ensures proper **storage systems** are used.
- Helps in designing **scalable solutions** .
- Allows companies to store large historical data for analysis.
- Without handling volume properly, systems may crash or lose data.

2. Importance of Velocity

- Velocity is important for handling **fast-moving data**.
- Helps in **real-time decision making**
- Important for applications like **online transactions, fraud detection**
- Ensures quick response to changing situations
- Without velocity, data becomes outdated and less useful.

Gateway Classes

3. Importance of Variety

- Variety helps in managing **different types of data**.
- Allows integration of **multiple data formats**
- Helps organizations get a **complete view of information**
- Improves analysis by combining different data sources
- Without variety, important insights from non-structured data may be missed.

4. Importance of Veracity

- Veracity ensures **data quality and accuracy**.
- Helps in making **reliable decisions**
- Reduces errors and wrong analysis
- Improves trust in data systems
- Poor veracity leads to incorrect results and bad decisions.

GATEWAY CLASSES
Gateway Classes

5. Importance of Value

- Value focuses on **extracting useful insights from data**.
- Helps organizations **improve performance and profit**
- Supports **better decision-making**
- Converts raw data into meaningful information
- Without value, data is useless even if it is large and fast

Gateway Classes

IMPORTANCE OF BIG DATA

Better Decision Making	Big Data helps in better decision-making by providing **accurate and real-time insights**. It allows organizations to **analyze large amounts of data and identify patterns and trends**.
Improved Customer Experience	Organizations analyze customer data to understand preferences and needs. This helps in providing personalized services.
Real-Time Processing	Big Data allows quick analysis of data as it is generated. This is useful in areas like stock markets and online services.
Innovation and Growth	It helps companies discover new opportunities and improve products. This leads to business growth and innovation.

GATEWAY CLASSES
Gateway Classes

Big Data : Features AKTU(2024-25)

1. Security

Big Data security ensures that data is protected from unauthorized access. It uses techniques like encryption and authentication. Security is important because big data contains sensitive information. It helps prevent data breaches and cyber attacks. A secure system builds trust among users.

2. Compliance

Compliance means following legal rules and regulations related to data. Organizations must follow data protection laws. It ensures that data is used in a proper and ethical way. Non-compliance can lead to penalties. It is important for maintaining legal standards.

Gateway Classes

GATEWAY CLASSES
GATEWAY CLASSES

3. Auditing

Auditing refers to tracking and monitoring data usage. It records who accessed data and what changes were made. This helps in detecting errors and misuse. Auditing improves transparency in the system. It is useful for security and accountability.

Gateway Classes

4. Protection

Protection ensures data safety from loss, damage, or theft. It includes backup, recovery, and disaster management. It helps maintain data integrity. Proper protection ensures continuous system operation. It is essential for reliable data management.

Big Data : Applications AKTU(2024-25)

Different industries use Big Data in different ways:

APPLICATIONS OF BIG DATA	
Healthcare	Big Data is used to analyze patient records and improve treatments. It helps in disease prediction and better diagnosis.
Banking and Finance	Used for fraud detection and risk management. It helps in secure and fast financial transactions.
E-commerce	Companies analyze customer behavior to recommend products. It improves sales and customer satisfaction.
Social Media	Platforms analyze user activity to show relevant content and advertisements. It helps in user engagement.

Gateway Classes

What is Data analytics ?? AKTU(2025-26)

- **Data Analytics** is the process of collecting, cleaning, organizing, analyzing, and interpreting data to discover useful information, identify patterns, and support better decision-making.
- **Data Analytics** helps convert raw data into meaningful information. It identifies trends, patterns, and relationships that help organizations improve performance, solve problems, and make informed decisions.

Gateway Classes

Objectives of Data Analytics

- Find useful information from data.
- Identify patterns and trends.
- Improve decision-making.
- Increase business performance.
- Predict future outcomes.

Need of Data Analytics AKTU(2024-25)

Q: Importance of Data Analytics in modern organization.

Need/Importance	Description
Better Decision-Making	Data Analytics helps organizations make accurate and informed decisions based on facts instead of assumptions.
Understand Customer Behavior	It helps businesses understand customer needs, preferences, and buying habits, which improves customer satisfaction.
Identify Trends and Patterns	It discovers hidden patterns and market trends that help organizations improve products and services.
Predict Future Outcomes	By analyzing historical data, organizations can forecast future sales, customer demand, and business growth.
Reduce Cost and Save Time	Analytics identifies unnecessary expenses and improves resource utilization, saving both time and money.

GATEWAY CLASSES
Gateway Classes

Big Data Analytics AKTU (2023-24)

- Big Data Analytics is the process of analyzing large and complex data sets to find useful patterns and information. It helps in making better decisions.
- It deals with very large amounts of data generated from different sources.
- It helps organizations understand trends and user behavior.
- It is used to make fast and accurate decisions.
- It plays an important role in business and technology today



GATEWAY CLASSES
Gateway Classes

Types of Data Analytics AKTU (2024-25)

Types	Meaning
1. Descriptive Analytics It explains what has happened in the past. It summarizes data into simple reports and charts. It helps in understanding historical trends. **Example: Monthly sales report.**	**3.Predictive Analytics** It predicts what is likely to happen in the future. It uses past data and patterns. It helps in forecasting trends. **Example: Predicting future sales**
2.Diagnostic Analytics It explains why something happened. It finds the reasons behind a problem or event. It uses data comparison and analysis. **Example: Why sales decreased last month.**	**4.Prescriptive Analytics** It suggests what actions should be taken. It gives recommendations based on data analysis. It helps in decision-making. **Example: Suggesting best marketing strategy.**

GATEWAY CLASSES
Gateway Classes

Evolution of analytic Scalability AKTU(2024-25)

Q :How have advancements in technology contributed to the scalability of analytics?

Analytic Scalability is the ability of an analytics system to efficiently process, store, and analyze increasing amounts of data without reducing its performance.

Advancements in Technology Have Improved Analytic Scalability :

1. Big Data Technologies

Big Data technologies such as **Apache Hadoop** and **Apache Spark** allow organizations to process very large datasets quickly. These technologies divide data into smaller parts and process them on multiple computers simultaneously.

2. High-Speed Storage Systems

Modern storage technologies such as SSDs and distributed storage systems provide faster access to large amounts of data.

Gateway Classes

GATEWAY CLASSES

3. Artificial Intelligence (AI) and Machine Learning (ML)

AI and ML technologies automatically analyze large datasets, identify hidden patterns, and make accurate predictions without much human effort.

4. Cloud Computing

Cloud platforms such as **Amazon Web Services (AWS), Microsoft Azure, and Google Cloud** provide unlimited storage and computing resources. Organizations can increase or decrease resources according to their needs.

5. Distributed Computing

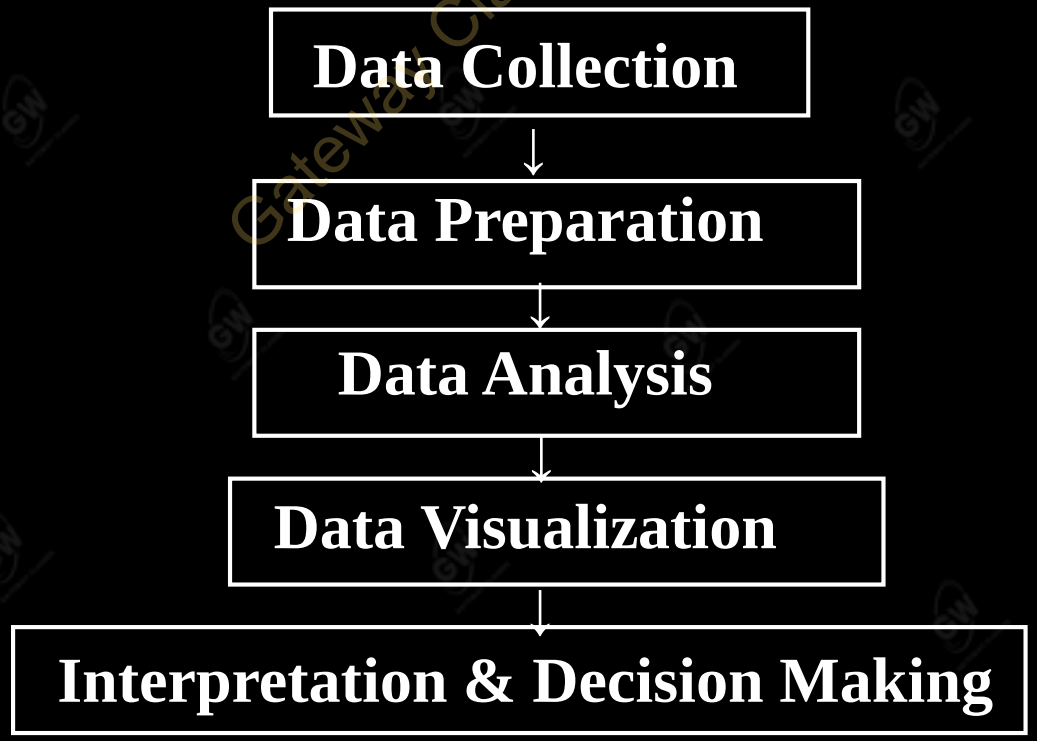
Instead of processing data on a single computer, distributed computing uses multiple computers working together. This reduces processing time and improves system performance.

Gateway Classes

Analytic Process

Analytics Process is a systematic sequence of steps used to collect, prepare, analyze, and interpret data to obtain meaningful insights and support better decision-making.

Steps of Analytic Process :



Gateway Classes

Step 1 : Data collection : Data is collected from different sources according to the business requirement. **Common Sources :** Databases ,Websites ,Social Media etc.

Step 2: Data Preparation : The collected data is cleaned and organized before analysis.

Step 3: Data Analysis : The prepared data is analyzed using statistical methods and analytical tools.

During analysis : Identify patterns, Find trends, Compare data ,Discover relationships, Make predictions.

Step 4: Data Visualization : The analyzed data is presented in a simple and visual form.

Common visualization methods: Charts, Graphs , Dashboards , Tables

Step 5: Interpretation and Decision Making : The final step is to interpret the results and use them to make business decisions.

Organizations use these insights to: Improve products and services ,Increase customer satisfaction .Reduce costs , Plan future strategies

Gateway Classes

Difference Between Analytics & Reporting AKTU (2023-24)

Basics	Analytics	Reporting
Meaning	Process of analyzing data to extract insights	Process of presenting data in organized form
Purpose	Supports decision-making and prediction	Supports monitoring and performance tracking
Focus	Future trends and patterns	Past data and results
Nature	Exploratory and detailed	Descriptive and summary-based
Complexity	Complex analysis using techniques	Simple charts, tables, reports
Output	Insights, patterns, predictions	Reports, dashboards, summaries
Questions /Answered	Why it happened & what will happen	| What happened

GATEWAY CLASSES
Gateway Classes

Modern Data Analytics Tools AKTU(2023-24)

Python

Python is used for data analysis and machine learning. It provides libraries like Pandas and NumPy for easy data handling

R

R is mainly used for statistical analysis and data visualization. It helps in handling complex data and creating graphs.

SQL

SQL is used to store, manage, and retrieve data from databases. It is important for working with structured data.

Tableau

Tableau is used to create interactive charts and dashboards. It helps in understanding data visually.

Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

Apache Spark

Spark is used for fast data processing and real-time analytics. It handles large datasets efficiently.

Apache Hadoop

Hadoop is used for distributed storage and processing of big data. It works across multiple machines to handle large data

Gateway Classes

Application of Data analytics AKTU(2023-24)

1. Healthcare	Data Analytics helps hospitals and healthcare organizations improve patient care and medical services.
2. Banking and Finance	Banks use Data Analytics to improve financial services and reduce risks.
3. E-Commerce	Online shopping companies use Data Analytics to understand customer behavior.
4. Marketing	Companies use Data Analytics to improve marketing strategies.
5. Education	Educational institutions use Data Analytics to improve teaching and learning.

Gateway Classes

Important Previous Year Questions

Q.1 Define Data Analytics and Explain its importance in decision-making. AKTU(2025-26)
Q.2 Differentiate between structured , semi-structured and unstructured data. AKTU(2025-26)
Q.3 Explain sources and characteristics of data with suitable examples. AKTU(2025-26)
Q.4 Analyze the role of Data Analytics in modern organization. AKTU(2025-26)
Q.5 What are the challenges and solutions associated with scaling analytics in modern organization. AKTU(2024-25)
Q.6 How do modern tools bridge the gap between analysis and reporting? AKTU(2024-25)
Q.7 Describe the key characteristics of Big Data . How do these characteristics necessitate the development of advanced analytics platforms? AKTU(2024-25)
Q.8 How has the need for data analytics evolved with the advent of IoT and real-time systems ? Provide relevant Examples . AKTU(2024-25)
Q.9 what are the features of Big Data ? AKTU(2024-25)
Q.10 Explain the need of Data Analytics. AKTU(2024-25)
Q.11 Discuss Classification of Data. Explain all categories in detail with examples.
Q.12 Write short note on the following : AKTU(2023-24)

Modern Data analytics tools
Application of Data analytics

Q.13 List the features of big data analytics platform. AKTU(2023-24)
Q.14 why there is need of data analytics ? AKTU(2023-24)
Q.16 discuss all the modern tools of Data analytics . AKTU(2022-23)
Q.17what is Data analytics ? AKTU(20225-26)
Q.18List any characteristics of big Data . AKTU(2025-26)
Q.19 Differentiate between Predictive and Prescriptive Data Analytics . AKTU(2024-25)
Q.20 Define the term data lake, data base , data warehouse . AKTU(2024-25)
Q.21 Explain the different categories of data analytics with examples. AKTU(2024-25)
Q.22 How have Advancements in technology contributed to the scalability of analytics ? AKTU(2024-25)
Q.23 What are the sources of data in the context of data analytics? AKTU(2024-25)
Q.24 Describe the characteristics of data that are relevant in the field of data analytics . How do these characteristics impact the analysis process ? AKTU(2022-23)
Q.25 Differentiate between analysis and reporting in the context of data analytics. How do these two aspects contribute to the overall understanding of data? AKTU(2022-23)
Q.26 Explore modern data analytics tools and their functionalities . How have these tools transformed the landscape of data analytics? AKTU(2022-23)

GATEWAY CLASSES
Gateway Classes

Gateway Classes

GATEWAY CLASSES

Empowering Learners, Transforming Futures

Gateway Classes

THANK YOU

GET IT ON Google Play

Download on the App Store

Helpline No- 7819 0058 53 , 9259833200

AKTU B.Tech 3rd Year

CS & CS Allied

Data Analytics & Visualization

Unit-1

in ONE SHOT

Part-II

AKTU PYQs Covered



By Shubhangi Ma'am

GATEWAY CLASSES

Gateway Classes

GATEWAY CLASSES

UNIT—1
Syllabus

GATEWAY CLASSES

Introduction to Data Analytics:

Sources and nature of data, classification of data (structured, semi-structured, unstructured), characteristics of data, introduction to Big Data platform, need of data analytics, evolution of analytic scalability, analytic process and tools, analysis vs reporting, modern data analytic tools, applications of data analytics.

Gateway Classes

Data Analytics Lifecycle:

Need, key roles for successful analytic projects, various phases of data analytics lifecycle – discovery, data preparation, model planning, model building, communicating results, operationalization.

Previous Year Important Questions

Q.1 List and briefly explain the phases of Data Analytics lifecycle. (2025-26)
Q.2 Explain Data Analytics lifecycle with a neat diagram and example. (2025-26)
Q.3 Define the term ‘ analytics lifecycle’. How do the roles of stakeholders vary across its different phases. (2024-25)
Q.4 Explain the phase of Data Analytics lifecycle in detail. (2023-24)
Q.5 Describe all the phases of Data Analytics life cycle. (2022-23)
Q.6 Explain the phases of the Data Analytics Lifecycle with objectives of each phase. (2025-26)
Q.7 What are the common open source tools for the model building phase ? (2022-23)

Gateway Classes

Data Analytics Lifecycle AKTU(2025-26)

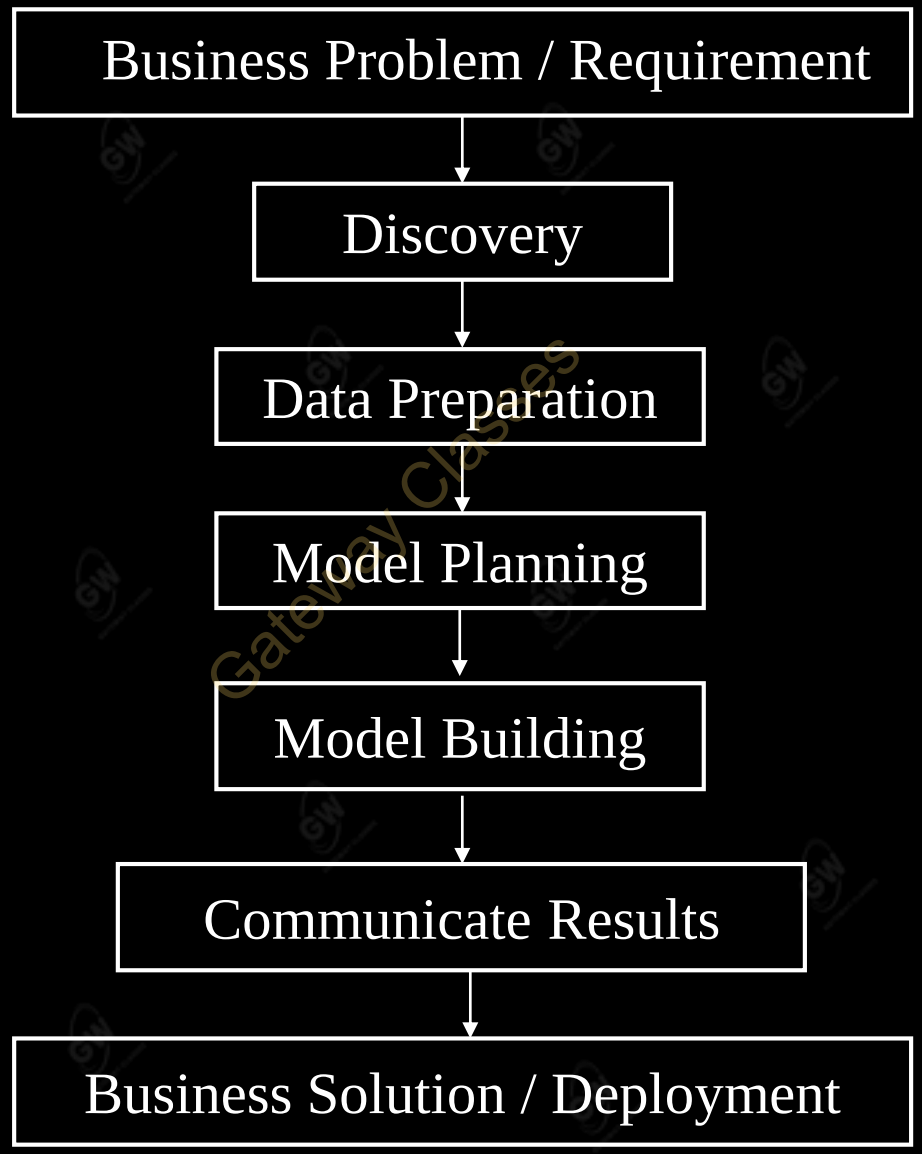
- Data Analytics Lifecycle is a step-by-step process used to solve business problems using data.
- It helps an organization collect data, prepare it, analyze it, and use the results to make better decisions.

Simple Meaning

Raw Data → Process → Useful Information → Better Decisions

Gateway Classes

Phases of Data Analytics lifecycle



GATEWAY CLASSES
Gateway Classes

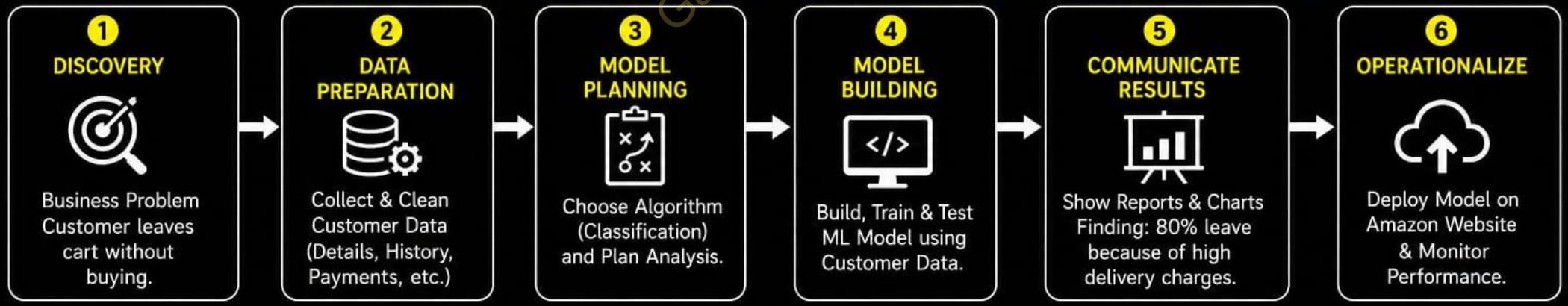
GATEWAY CLASSES

DATA ANALYTICS LIFECYCLE – EXAMPLE, FLOW & TOOLS

Phase	What happens?	Amazon Example	Tools Used
1. DISCOVERY	Understand the business problem and set goals.	Amazon notices that many customers add products to the cart but do not complete the purchase.	MS Excel, Google Sheets, JIRA, Trello, Microsoft Teams
2. DATA PREPARATION	Collect, clean and organize data.	Collect customer details, cart history, purchase history, payment method, etc. Remove duplicate and missing data.	SQL, Python (Pandas), R, Excel, Apache Spark
3. MODEL PLANNING	Decide how to analyze the data and choose the algorithm.	Decide to use a Classification model to predict whether a customer will purchase or not.	Python, R, Jupyter Notebook, Scikit-learn
4. MODEL BUILDING	Build, train and test the machine learning model.	Train the model using customer data and check its accuracy.	Python, R, Weka, KNIME, Orange, TensorFlow, PyTorch, Apache Spark MLlib
5. COMMUNICATE RESULTS	Present findings to stakeholders.	Show that 80% of customers leave the cart because of high delivery charges.	Power BI, Tableau, Excel, Google Data Studio
6. OPERATIONALIZE	Deploy the model and monitor performance.	Amazon starts showing discount coupons and free delivery offers to customers likely to abandon their carts.	AWS, Azure, Google Cloud, Docker, Flask, Kubernetes

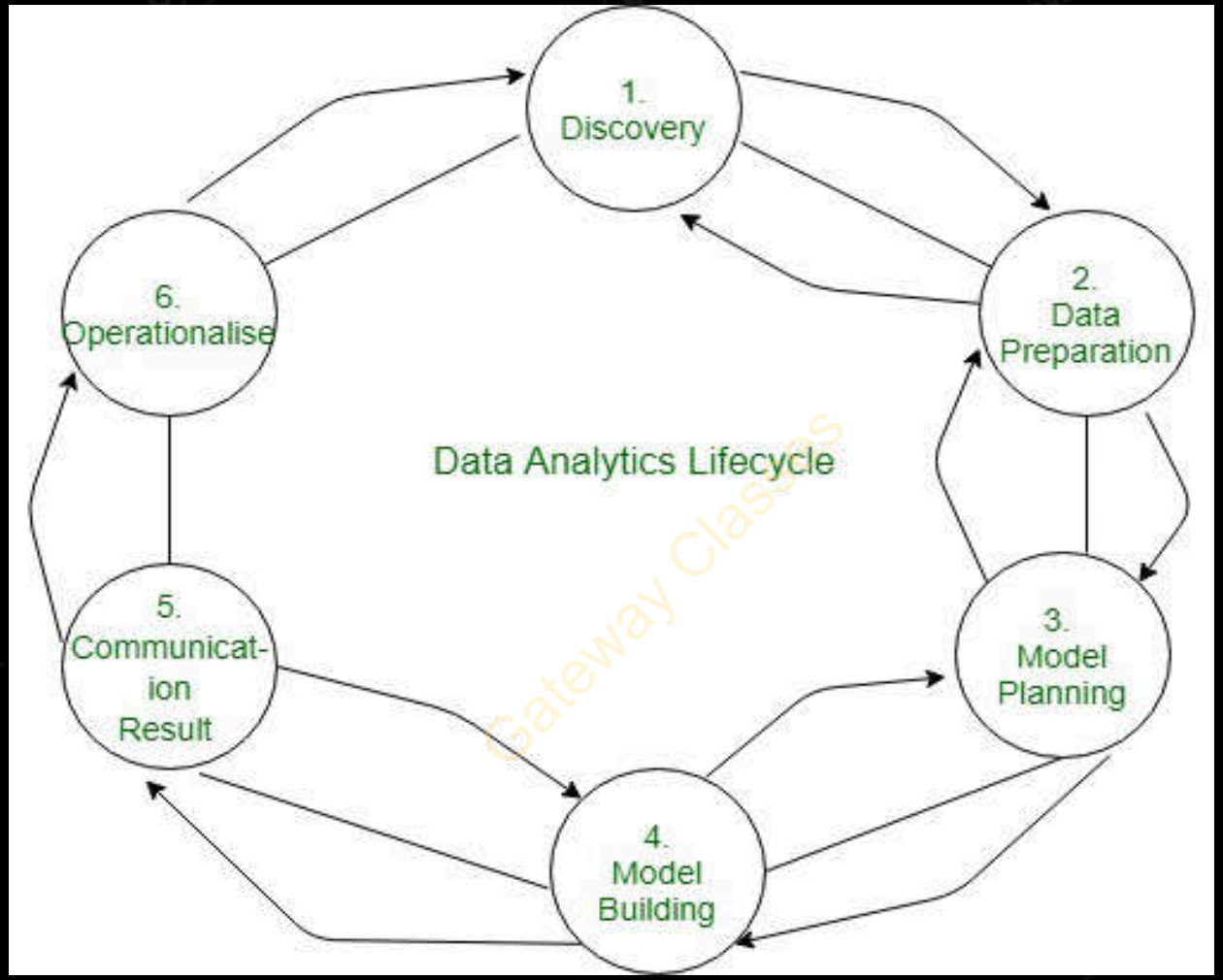
Gateway Classes

DATA ANALYTICS LIFECYCLE – FLOW (AMAZON EXAMPLE)



IN ONE LINE: Data Analytics Lifecycle is a step-by-step process that starts with understanding a business problem and ends with implementing a data-driven solution.

GATEWAY CLASSES



GATEWAY CLASSES

Discovery

The Discovery phase is the first phase of the Data Analytics Lifecycle. In this phase, the team understands the business problem, project goals, available data, and required resources.

Example :-

An online shopping company wants to increase its sales. The team first studies customer complaints, sales reports, and business goals before starting the project.

Data Preparation

The Data Preparation phase is used to collect, clean, organize, and prepare data before analysis.

Example :-

A bank removes duplicate customer records and fills missing information before analyzing customer data.

GATEWAY CLASSES
Gateway Classes

GATEWAY CLASSES

Model Planning

The Model Planning phase is the process of selecting the best method, tools, and techniques to analyze the prepared data.

Example :-

A company chooses a sales prediction model to forecast future sales.

Model Building

The Model Building phase is the process of creating, training, and testing a model using the prepared data.

Example :-

Netflix builds a recommendation model to suggest movies based on users' watching history.

GATEWAY CLASSES

GATEWAY CLASSES

Communicate Results

The Communicate Results phase is the process of presenting the analysis results to stakeholders through reports, charts, and dashboards, So they can understand the results better and make better decisions .

Example :-

A company presents a report showing that online sales increased after a new marketing campaign.

Gateway Classes

Operationalization

The Operationalization phase is the final step where the analytical model is deployed and used in the real business environment.

Example :-

Amazon deploys its recommendation system on its website to suggest products to customers.

Importance of Data Analytics Lifecycle

The Data Analytics Lifecycle is important because it provides a structured approach to complete an analytics project successfully. It helps organizations manage data efficiently and make better business decisions.

- Helps in understanding the business problem clearly.
- Ensures high-quality and accurate data before analysis.
- Supports better and faster decision-making.
- Improves the accuracy of analytical models.
- Saves time, cost, and resources.
- Reduces errors and business risks.
- Helps in deploying the final solution successfully.

Gateway Classes

Need of Data Analytics Lifecycle

- Helps identify and understand the business problem clearly.
- Provides a systematic and well-organized approach for analytics projects.
- Ensures the collection of relevant and high-quality data.
- Improves data quality through proper cleaning and preparation.
- Helps select the most suitable analytical methods and tools.
- Reduces errors and increases the accuracy of analysis.
- Supports faster and more effective decision-making.
- Saves project time, cost, and resources.
- Improves communication and collaboration among team members.
- Helps develop reliable and efficient analytical models.

GATEWAY CLASSES
Gateway Classes

Objectives of Data Analytics Lifecycle

- Convert raw data into useful information.
- Solve business problems effectively.
- Support better and faster decision-making.
- Improve the quality and accuracy of data.
- Identify useful patterns and trends from data.
- Build reliable analytical models.
- Reduce project time, cost, and risks.
- Improve business performance and productivity.
- Deliver meaningful insights to stakeholders.
- Ensure the successful implementation of analytics solutions.

GATEWAY CLASSES
Gateway Classes

Open Source tools for Model Building Phase AKTU(2022-23)

- Open Source Tools are free software tools whose source code is publicly available. Anyone can use , modify, and distribute these tools.
- In the Model Building Phase , these tools are used to create , train , test , and evaluate machine learning or Statistical models.

Common Open Source tools :

- R
- Python
- Apache Spark(Mlib)
- Weka
- KNIME
- Orange
- TensorFlow
- Pytorch

Short trick :

Ram Par Se Wahan Kahda, Orange Todkar Pahucha

Gateway Classes

Tools	Description
R	R is an Open –Source Programming language used for statistical analysis, Data visualization, & Machine learning **Example :** A Company uses R to predict next months sales.
Python	Python is an Open –Source Programming Language used for data analysis, Machine Learning, Artificial Intelligence. Popular Libraries are NumPy , Pandas etc. **Example :** Python is used to predict whether an email is spam or not.
Apache Spark(MLib)	Apache Spark is an Open – source big data framework . MLib is its machine learning library used to build models on large datasets. **Example :** An e –commerce company uses spark to analyze millions of customer records

GATEWAY CLASSES
Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

Tools	Description
Weka	Weka is an Open Source machine Learning tool with a graphical user interface (GUI). It allows users to build models without much coding. It is used in Classification, Clustering, Regression, Data Preprocessing etc. Examples : A student uses Weka to classify customer data.
KNIME	KNIME is an open – Source Data analytics platform. It provides a drag and drop interface for building machine learning workflows. It is used in Data Cleaning, Data Analysis , Machine Learning, Workflow Automation. Example : A business analyst builds a sales prediction model using KNIME without writing code.

Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

Tools	Description
Orange	Orange is an Open Source visual Data mining and machine Learning tool . It provides a simple drag-and drop interface. **Example** : A beginner uses Orange to create a prediction model Visually.
Tensor Flow	Tensor Flow is an open Source deep leaning framework developed by Google. It is used to build neural network models. **Example :** Tensor Flow is used in Face recognition systems.
Pytorch	Pytorch is an open source deep learning framework developed by Meta(Facebook) . It is popular for AI research and application development. **Example :** It is used to build chatbots and image classification models .

GatewayClasses

Key Roles for Successful Analytics Projects

1. Business User

The Business User is the person who identifies the business problem and defines the project requirements. They explain the business objectives, expected outcomes, and provide domain knowledge to the analytics team. Throughout the project, they review the results and ensure that the final solution meets the organization's business goals.

Key Responsibilities

- Identifies business problems.
- Defines project goals and requirements.
- Provides business knowledge.
- Reviews the final solution.

Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

2. Project Manager

The Project Manager is responsible for planning, organizing, and managing the entire analytics project. They coordinate with all team members, allocate resources, monitor project progress, and ensure that the project is completed on time and within budget.

Key Responsibilities

- Plans and manages the project.
- Assigns tasks to team members.
- Monitors project progress.
- Ensures timely project completion.

Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

3. Data Engineer

The Data Engineer collects, stores, and prepares data for analysis. They build data pipelines, integrate data from different sources, and ensure that the data is clean, accurate, and easily accessible for the analytics team.

Key Responsibilities

- Collects data from multiple sources.
- Cleans and prepares data.
- Maintains data pipelines.
- Ensures data quality and availability.

Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

4. Data Analyst

The Data Analyst examines the prepared data to discover useful patterns, trends, and insights. They create reports, charts, and dashboards that help managers understand the data and make informed business decisions.

Key Responsibilities

- Analyzes data.
- Identifies patterns and trends.
- Creates reports and dashboards.
- Supports business decision-making.

Gateway Classes

GATEWAY CLASSES

5. Data Scientist

The Data Scientist develops advanced analytical and machine learning models to solve complex business problems.

They analyze large datasets, build predictive models, and improve model performance to generate accurate business insights.

Key Responsibilities

- Builds predictive models.
- Uses machine learning techniques.
- Solves complex business problems.
- Improves model accuracy.

Gateway Classes

GATEWAY CLASSES

GATEWAY CLASSES

6. Database Administrator (DBA)

The Database Administrator (DBA) manages and maintains the organization's databases. They ensure data security, perform regular backups, control user access, and maintain database performance to keep the data safe and available.

Key Responsibilities

- Manages databases.
- Ensures data security.
- Performs backup and recovery.
- Maintains database performance.

Gateway Classes

Previous Year Important Questions

Q.1 List and briefly explain the phases of Data Analytics lifecycle. (2025-26)
Q.2 Explain Data Analytics lifecycle with a neat diagram and example. (2025-26)
Q.3 Define the term ‘ analytics lifecycle’. How do the roles of stakeholders vary across its different phases. (2024-25)
Q.4 Explain the phase of Data Analytics lifecycle in detail. (2023-24)
Q.5 Describe all the phases of Data Analytics life cycle. (2022-23)
Q.6 Explain the phases of the Data Analytics Lifecycle with objectives of each phase. (2025-26)
Q.7 What are the common open source tools for the model building phase ? (2022-23)

Gateway Classes

Gateway Classes

Empowering Learners, Transforming Futures

GATEWAY CLASSES

Thank You

Gateway Classes

GET IT ON Google Play

Download on the App Store

Helpline No- 9259833200