



DATA ENGINEERING & ANALYTICS

Specialization

FOUNDATION PROGRAM



contact@regexsoftware.com



www.regexsoftware.com

ABOUT THE PROGRAM



At Regex Software, we offer a comprehensive 18–24 Month Data Science & AI Specialization Program designed especially for B.Tech, BCA, B.Sc, BBA, B.Com and other college students who want to build a strong career in the tech industry.

This program is ideal for 1st and 2nd year students who want to start learning future technologies early and gain a competitive edge in the industry. The course is designed to build strong fundamentals first and then gradually move toward advanced concepts in Artificial Intelligence, Machine Learning, Big Data, and Data Analytics.

At Regex Software, we believe in practical learning rather than only theoretical knowledge. Our training approach combines conceptual understanding with hands-on implementation through real-world projects, live datasets, industry case studies, and portfolio-building assignments. Throughout the 18–24 month journey, students continuously work on practical tasks that help them understand how technologies are used in real companies and business environments.

The curriculum covers Python programming along with advanced libraries such as NumPy, Pandas, and Matplotlib, helping students build a strong foundation in programming and data handling. Students also learn Statistics, Data Analysis, SQL, Database Management, Machine Learning, Deep Learning, Big Data technologies like Hadoop and Spark, Data Visualization, Cloud fundamentals, and MLOps basics. The entire learning path is designed according to current industry requirements so students become job-ready with practical industry skills.



MODE

Physical
(Jaipur & Ahmedabad)
or Online
(Google Meet)



DURATION

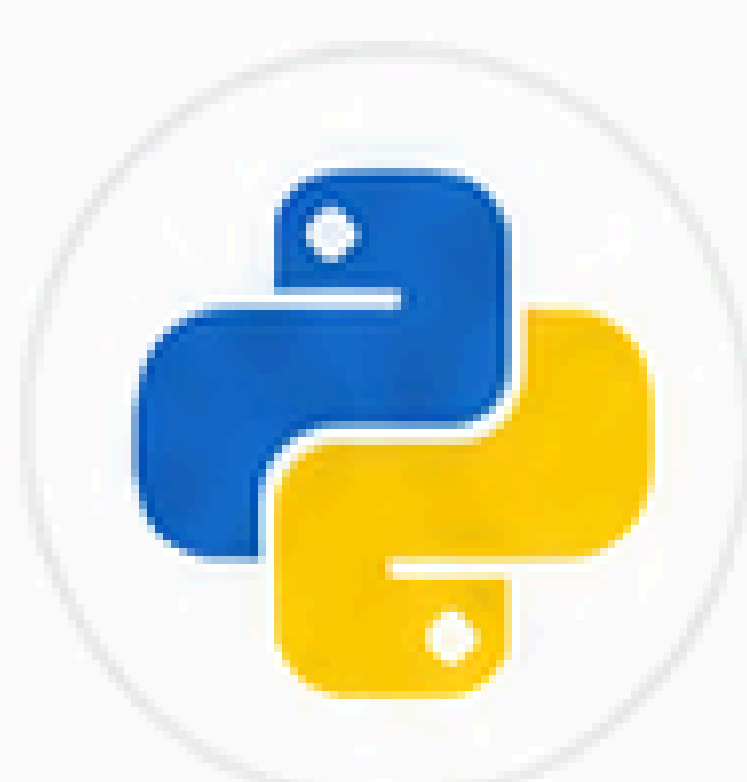
18–24 Months
+ 6 months support



PARTICIPANTS

18 - 20 Per
Batch

WHAT YOU WILL LEARN ?



Python



SQL



ETL & Data
Warehousing



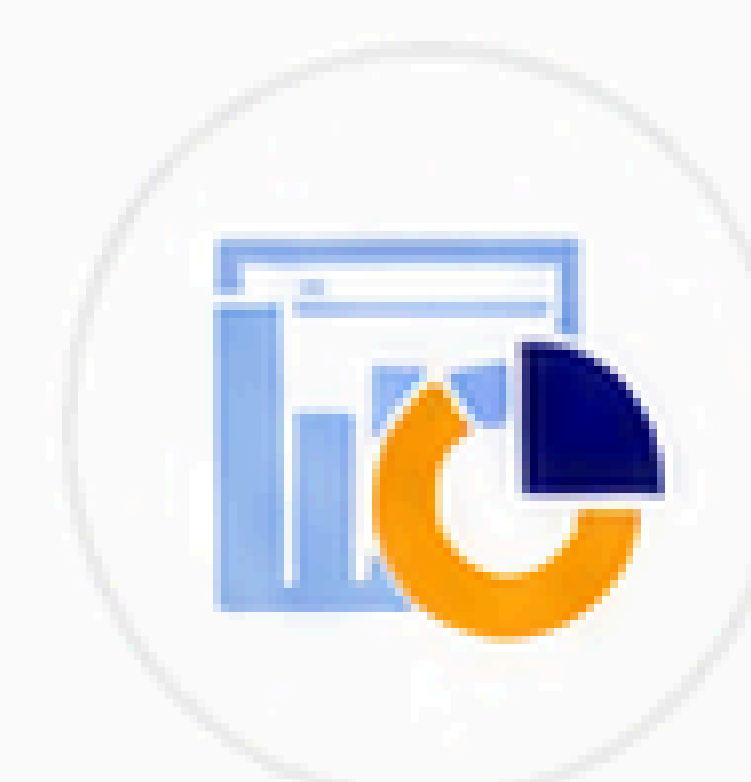
Hadoop



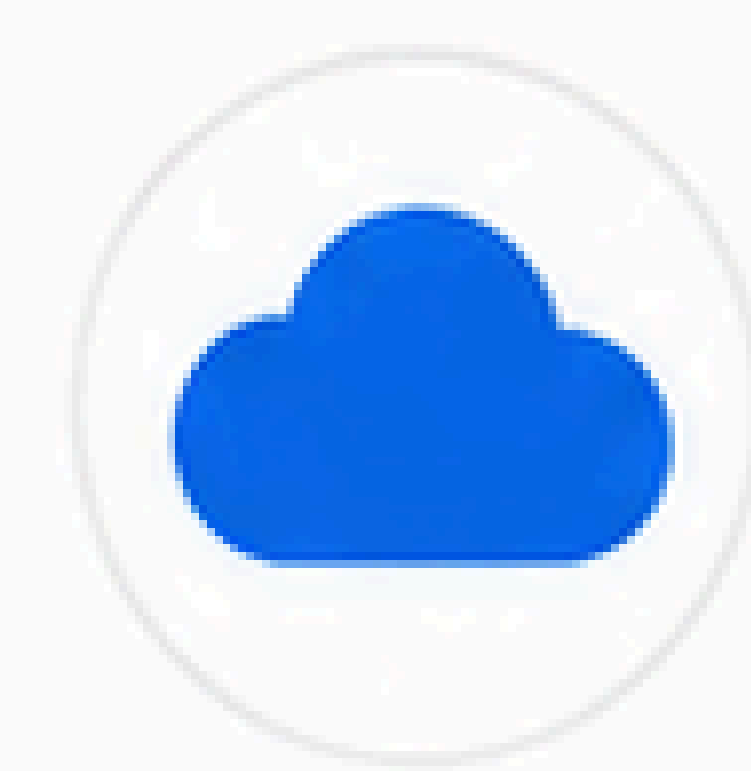
Apache Spark



Data Analytics



Data Visualization



Cloud
Fundamentals



Machine
Learning



Deep
Learning



MLOps
Basics



Real World
Projects



STUDY MATERIAL

- E-Notes
- Assignments & PDF test
- Live Video Lectures
- Access of Recordings & Study Material
- Mentorship Support
- Work on multiple Projects & Use Cases
- Work on Live Projects
- Interview Preparation Guidance



OUTPUT

- Build modern, scalable and end-to-end data pipelines
- Become job-ready Data Engineer / Data Analyst with real-world projects
- Understand Data Engineering & Analytics in depth
- Build projects on multiple domains
- Work on more than 25 Use CASES & Projects
- Learn to deploy data pipelines on cloud platforms
- Learn SQL, Python, ETL, Big Data Technologies, Data Warehousing, Visualization & more



DATA ENGINEERING & ANALYTICS

with

AI Specialization

FOUNDATION PROGRAM



Apache Airflow



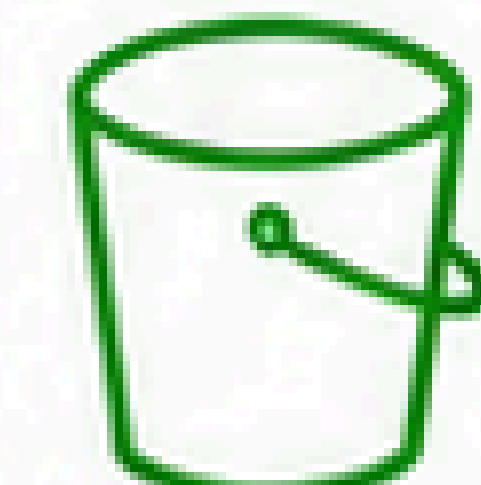
Apache Spark



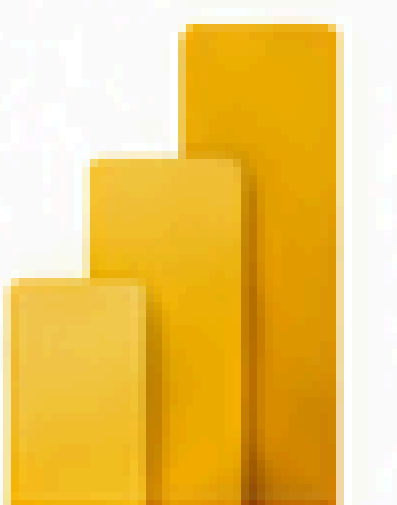
Databricks



Snowflake



Amazon S3



Power BI

PACKAGE OFFERED SO FAR

IT CANDIDATES



Minimum Package
4 LPA



Average Package
4 – 7 LPA



Overall Highest Package
36 LPA

NON-IT CANDIDATES



Minimum Package
3 LPA



Average Package
3.5 – 5.5 LPA



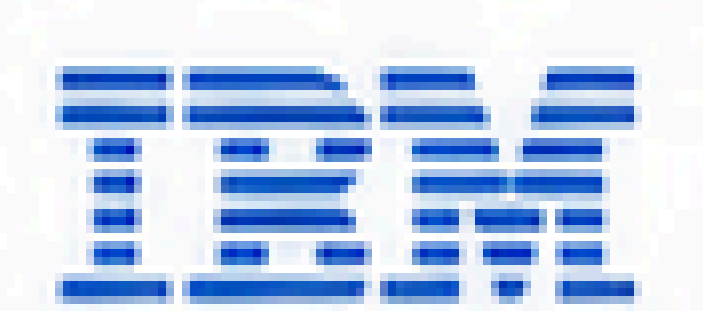
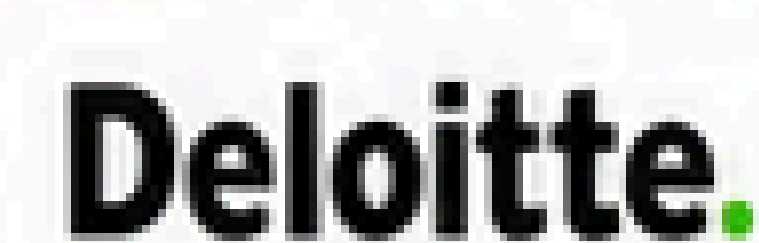
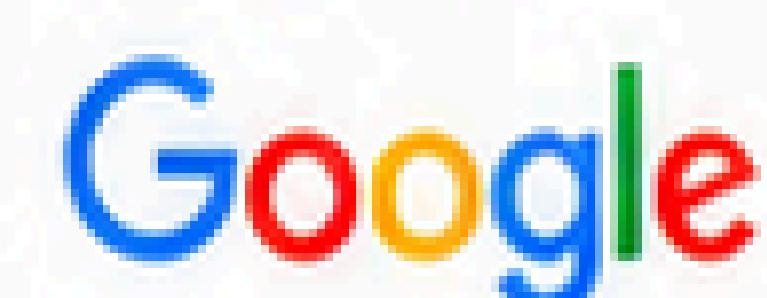
Overall Highest Package
14.5 LPA



EXTRA SESSIONS:

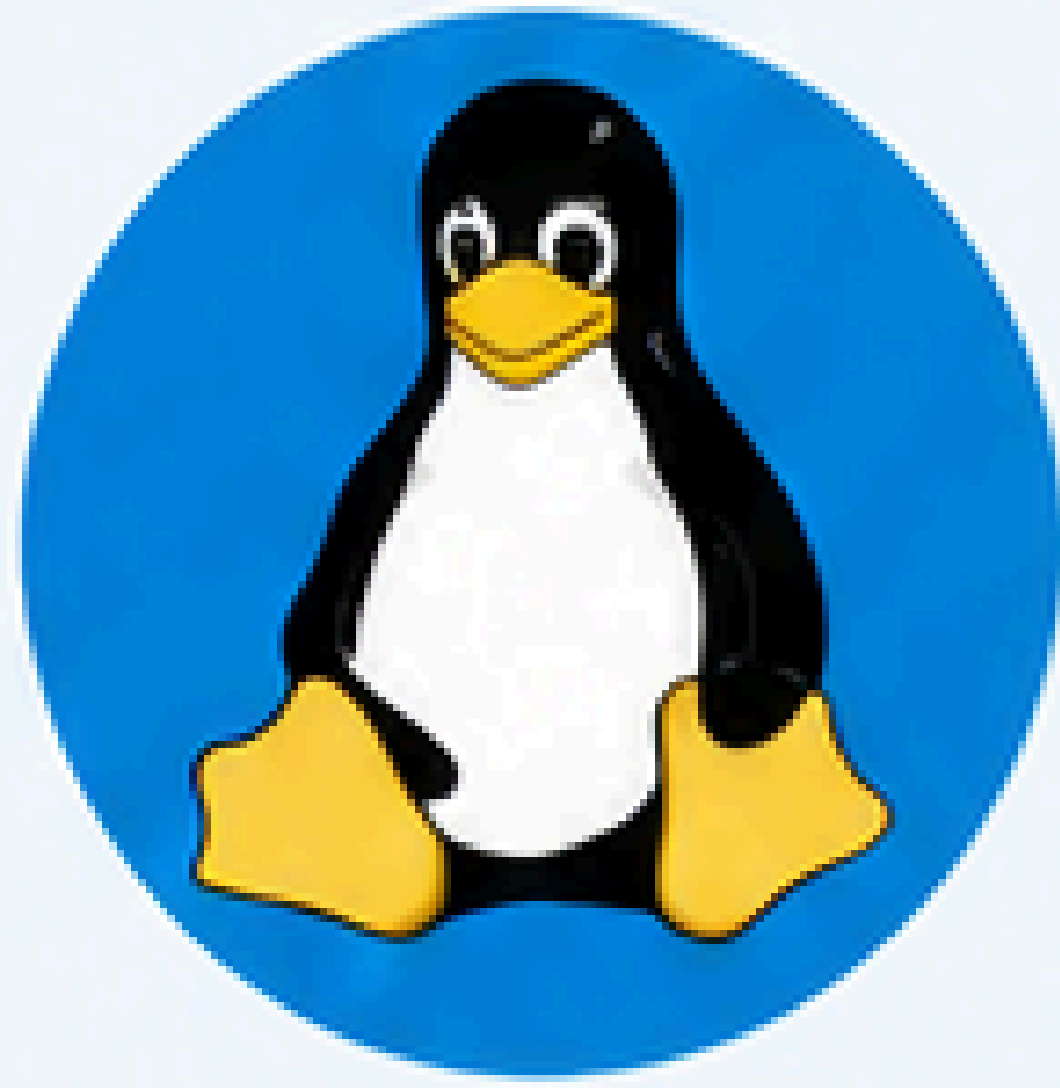
Additional Session on SQL, Python, Cloud (AWS), Databricks, Airflow, Spark, Power BI, Tableau, Git, Docker, and many more for all students.

OUR STUDENTS PLACED // PARTNERSHIP



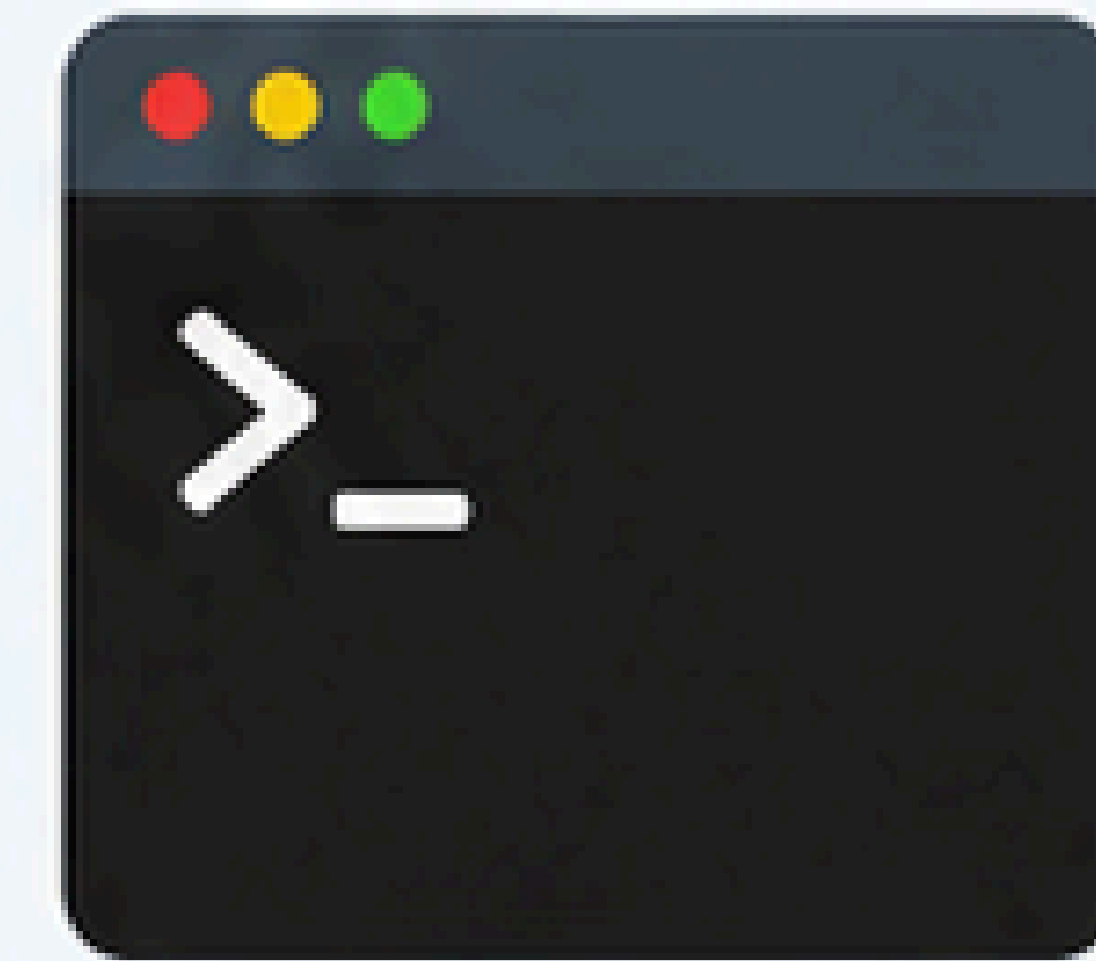
COURSE CONTENT:

PHASE 1



1. LINUX BASICS

- Introduction to Linux
- Basic Commands of Linux OS
- Vi Editor
- Tar Archive
- User Management and Permission



2. GITHUB

- Git vs GitHub
- Git Commands
- Git Branches
- Branching and Merging
- Git Push vs Pull Commands



3. C PROGRAMMING – CORE CONCEPTS

- Data Types
- Variables
- Variable Scope (Local, Global)
- Constants
- Operators
- Decision Making Statements (if Statement, if..else, Switch)
- Loops (While Loop, Do-While Loop, for-loop)



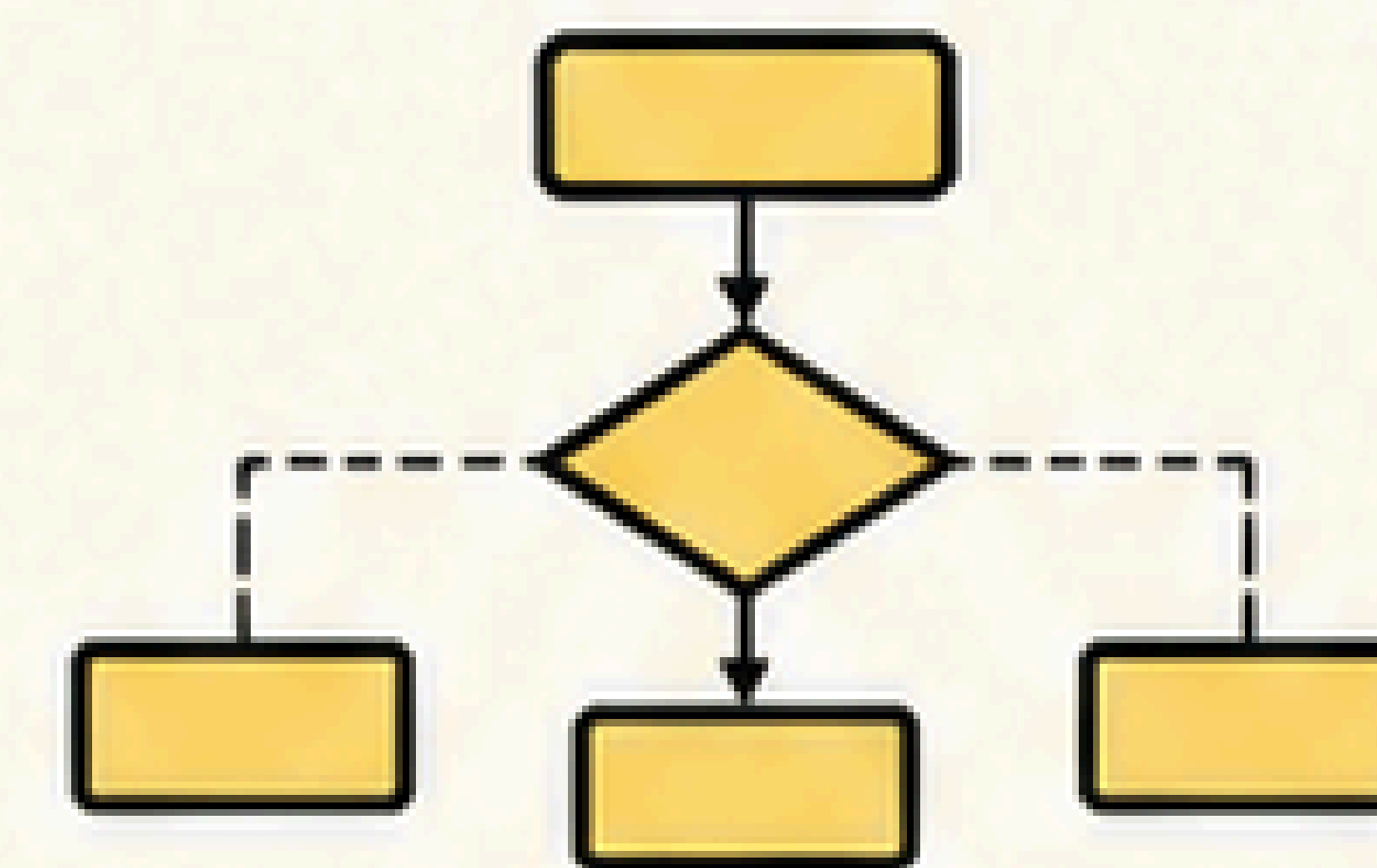
4. C PROGRAMMING – ADVANCED BASICS

- While Loop
- Do-While Loop
- for-loop
- Basic I/O Functions



5. LOGIC BUILDING

- Logical Thinking and Approach
- Flowcharts and Pseudocode
- Pattern Recognition
- Basic Problem-Solving



6. BASICS PROBLEM-SOLVING

- Understanding Problem Statements
- Breaking Problems into Steps
- Test Cases and Debugging
- Optimizing Solutions (Basics)



PHASE 2



1. C++ PROGRAMMING (OOP CONCEPTS)

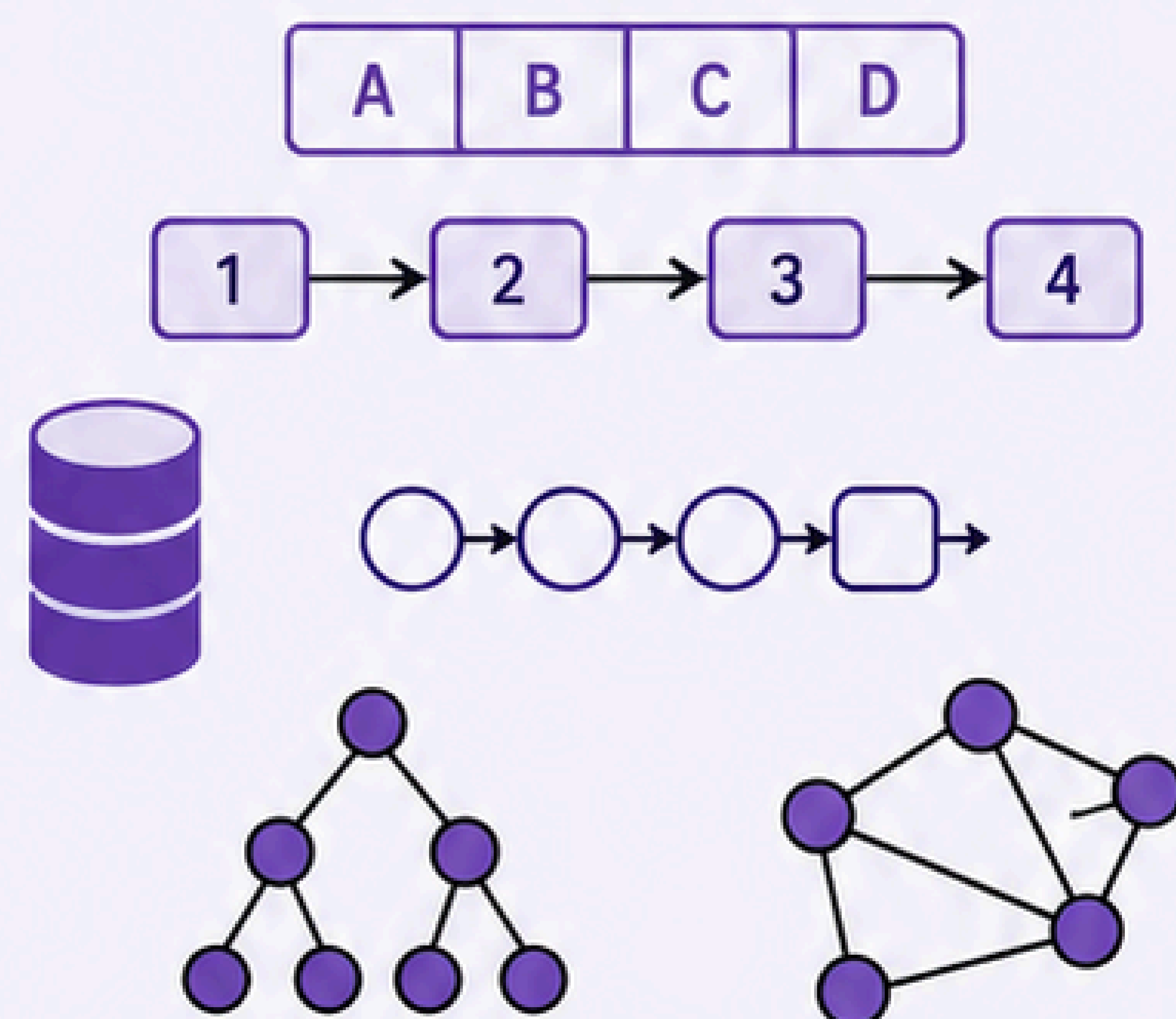
- Classes and Objects
- Constructors and Destructors
- this Pointer
- Access Specifiers: public, private, protected
- Getter and Setter Functions

```
class Student {  
public:  
    void set(int x);  
    int get();  
};
```



2. DATA STRUCTURES (ARRAY, LINKED LIST, STACK, QUEUE, TREES, GRAPHS)

- Arrays
- Linked List
- Stack
- Queue
- Trees
- Graphs



3. ALGORITHMS (SORTING, SEARCHING, RECURSION)

- Sorting Algorithms
- Searching Algorithms
- Recursion and Backtracking
- Time and Space Complexity



4. PROBLEM SOLVING (LEETCODE / CODEFORCES LEVEL)

- Problem Solving Techniques
- Pattern Recognition
- Practice on LeetCode
- Practice on Codeforces



PHASE 3

Python Programming



- Basics: Variables, Data Types, Operators, Input/Output
- Control Flow: If-Else, Loops
- Functions & Modules
- Data Structures: List, Tuple, Set, Dictionary
- OOPs: Classes, Objects, Inheritance, Polymorphism
- File Handling
- Exception Handling
- Popular Libraries: NumPy, Pandas, Matplotlib
- Mini Projects

```
def greet(name):  
    print(f"Hello, {name}!")  
  
name = input("Enter your name: ")  
greet(name)  
  
# Loop Example  
for i in range(1, 6):  
    print(i)
```

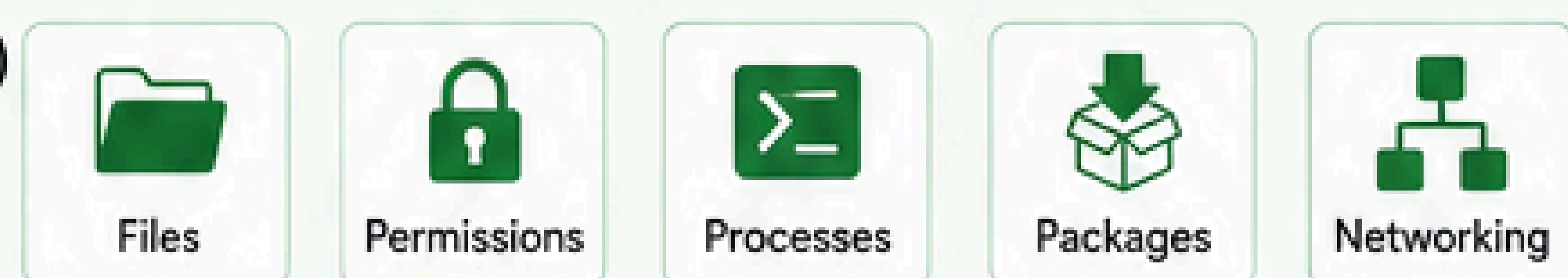


Linux Fundamentals



- Introduction to Linux
- Linux File System Hierarchy
- Basic Commands (ls, cd, pwd, cp, mv, rm)
- File and Directory Management
- Permissions and Ownership
- Process Management (ps, top, kill)
- Package Management (apt, yum)
- Text Editors (vi, nano)
- Shell Scripting Basics
- Networking Basics (ip, ping, netstat)
- User and Group Management

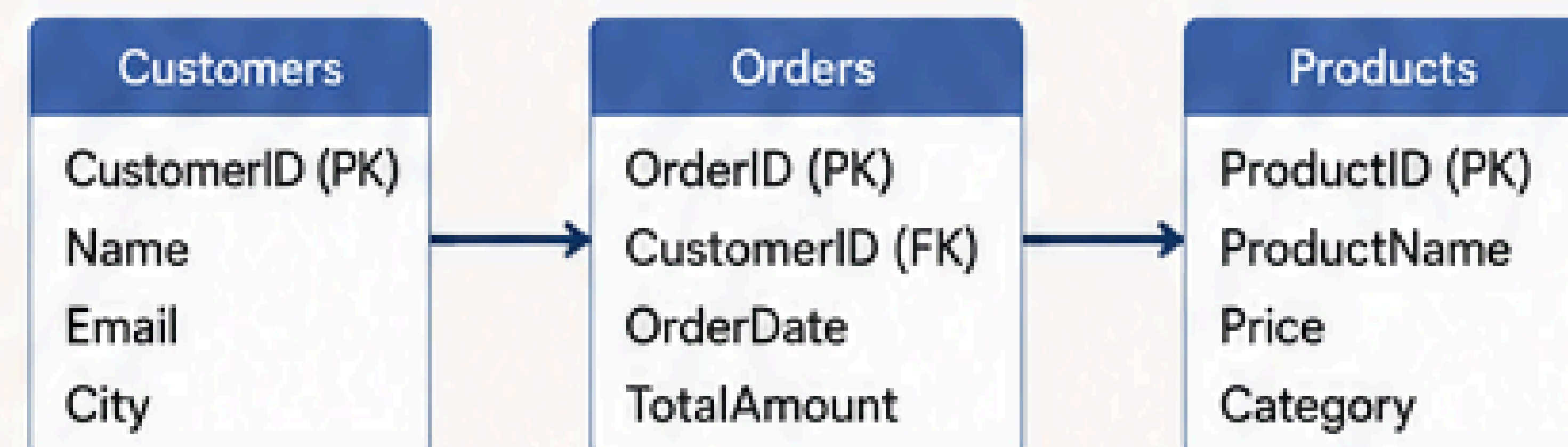
```
user@linux:~$ ls -l  
total 24  
drwxr-xr-x  2 user user 4096 May 20 docs  
-rw-r--r--  1 user user 1024 May 20 file.txt  
drwxr-xr-x  2 user user 4096 May 20 scripts  
  
user@linux:~$ pwd  
/home/user  
  
user@linux:~$ df -h  
Filesystem      Size  Used Avail Use% Mounted on  
/dev/sda1       50G   15G   33G   32% /  
/dev/sda2      100G   40G   60G   40% /home  
  
user@linux:~$ _
```



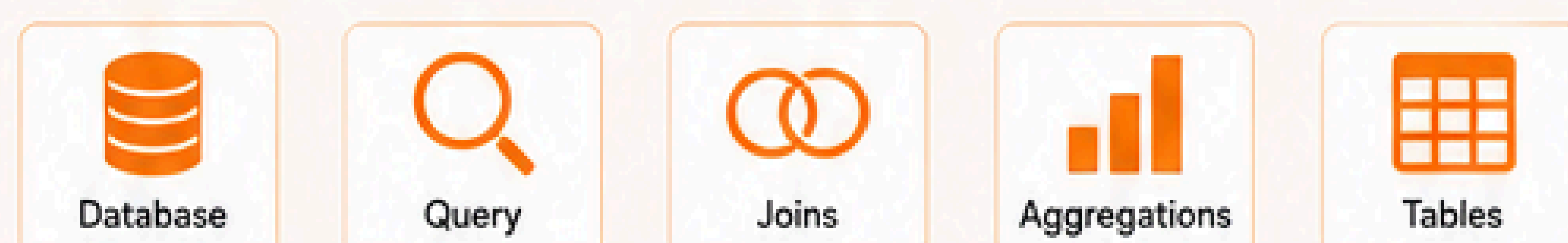
SQL & Database



- Introduction to Databases
- SQL Basics: SELECT, INSERT, UPDATE, DELETE
- Filtering Data: WHERE, ORDER BY, LIMIT
- Aggregations: COUNT, SUM, AVG, MIN, MAX
- Joins: INNER, LEFT, RIGHT, FULL JOIN
- Subqueries
- Group By and Having
- Database Normalization Basics
- Views, Indexes, Constraints
- Hands-on Projects



```
SELECT c.Name, o.OrderDate, p.ProductName, o.TotalAmount  
FROM Customers c  
JOIN Orders o ON c.CustomerID = o.CustomerID  
JOIN Products p ON o.ProductID = p.ProductID  
WHERE o.TotalAmount > 1000  
ORDER BY o.OrderDate DESC;
```



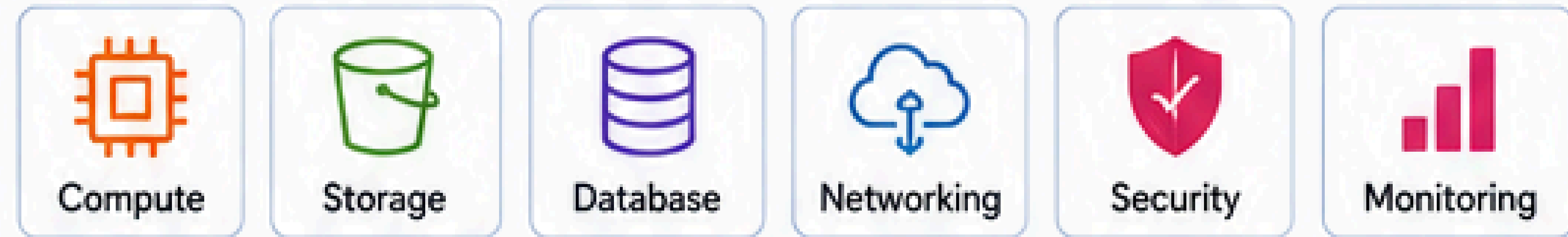
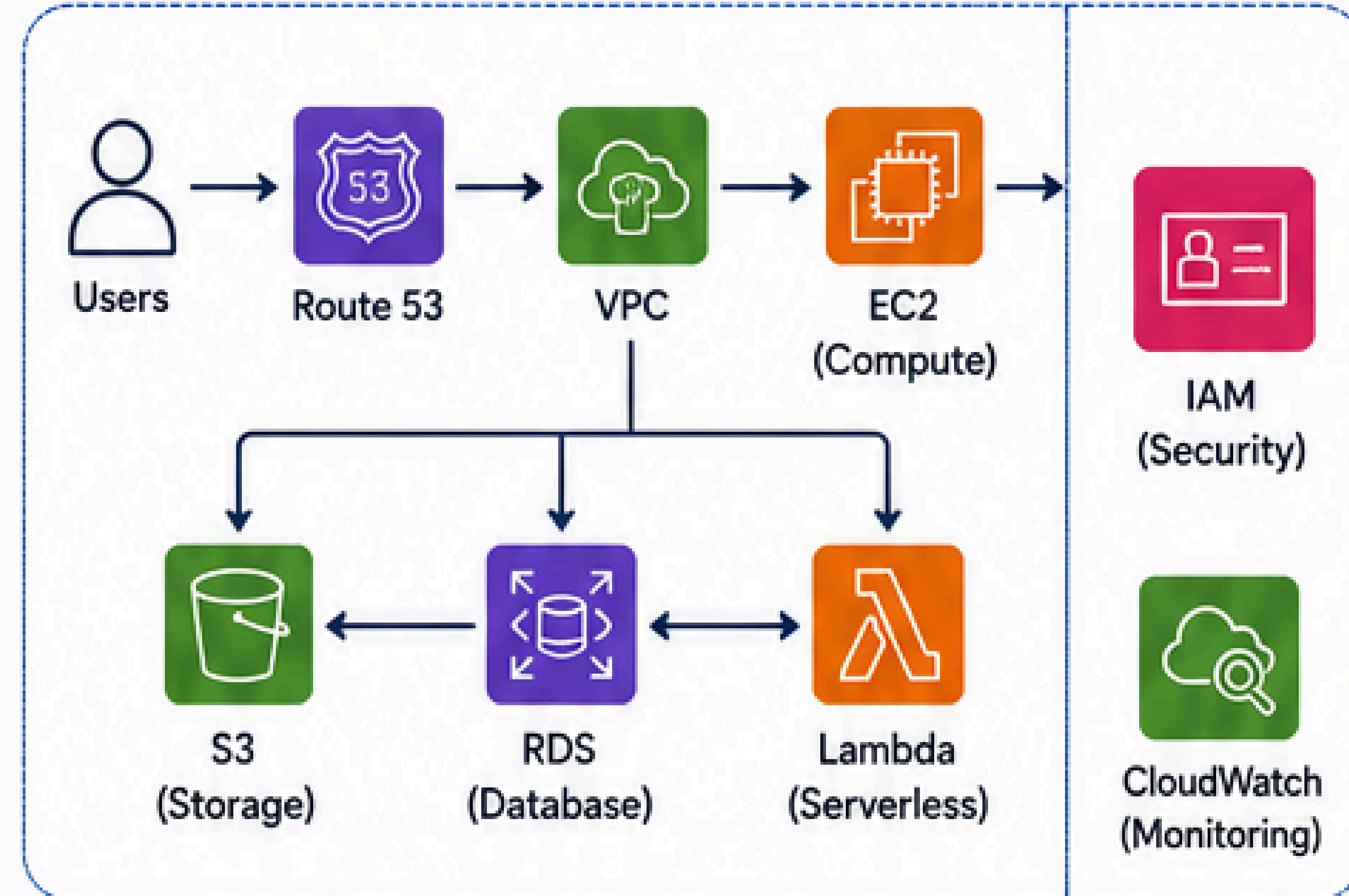
PHASE 4



• AWS Cloud

- Cloud Computing Concepts
- AWS Core Services (EC2, S3, IAM, VPC, RDS)
- Storage: S3, EBS, Glacier
- Compute: EC2, Lambda, ECS
- Networking: VPC, IAM, Route 53
- Databases: RDS, DynamoDB
- Data Integration: Glue, Athena
- Monitoring: CloudWatch
- Security Best Practices
- Cost Optimization
- Hands-on Projects

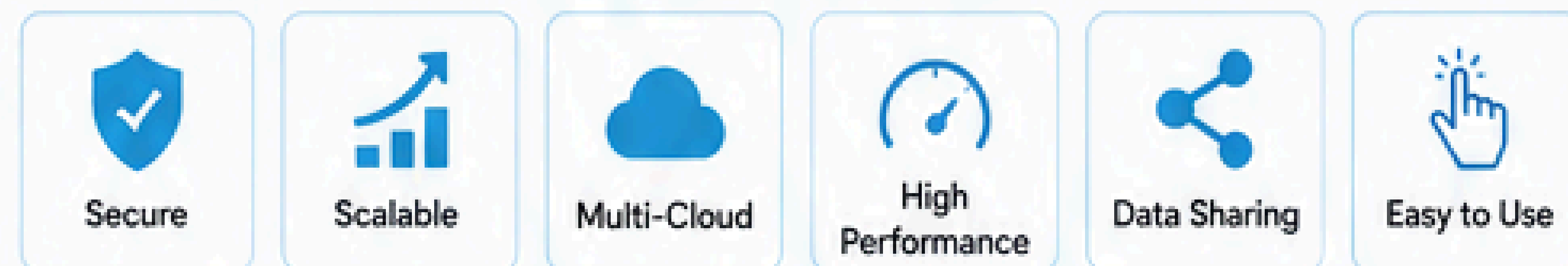
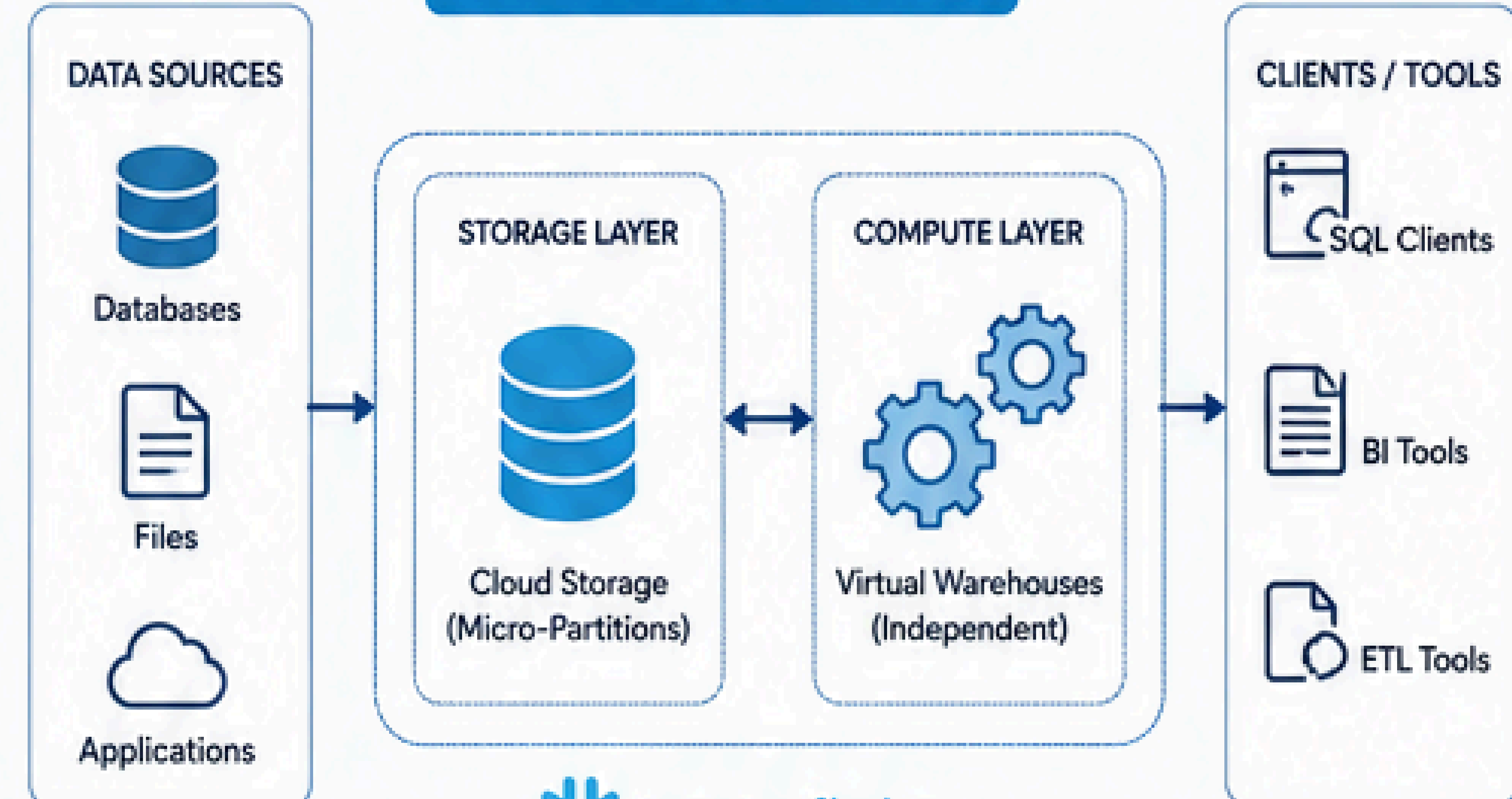
AWS ARCHITECTURE



• Snowflake

- Introduction to Snowflake
- Snowflake Architecture
- Data Loading: COPY INTO, Snowpipe
- Databases, Schemas, Tables, Stages
- Virtual Warehouses
- Time Travel & Fail-safe
- Data Sharing
- Clustering & Micro-Partitioning
- Performance Optimization
- Role-Based Access Control (RBAC)
- Hands-on Projects

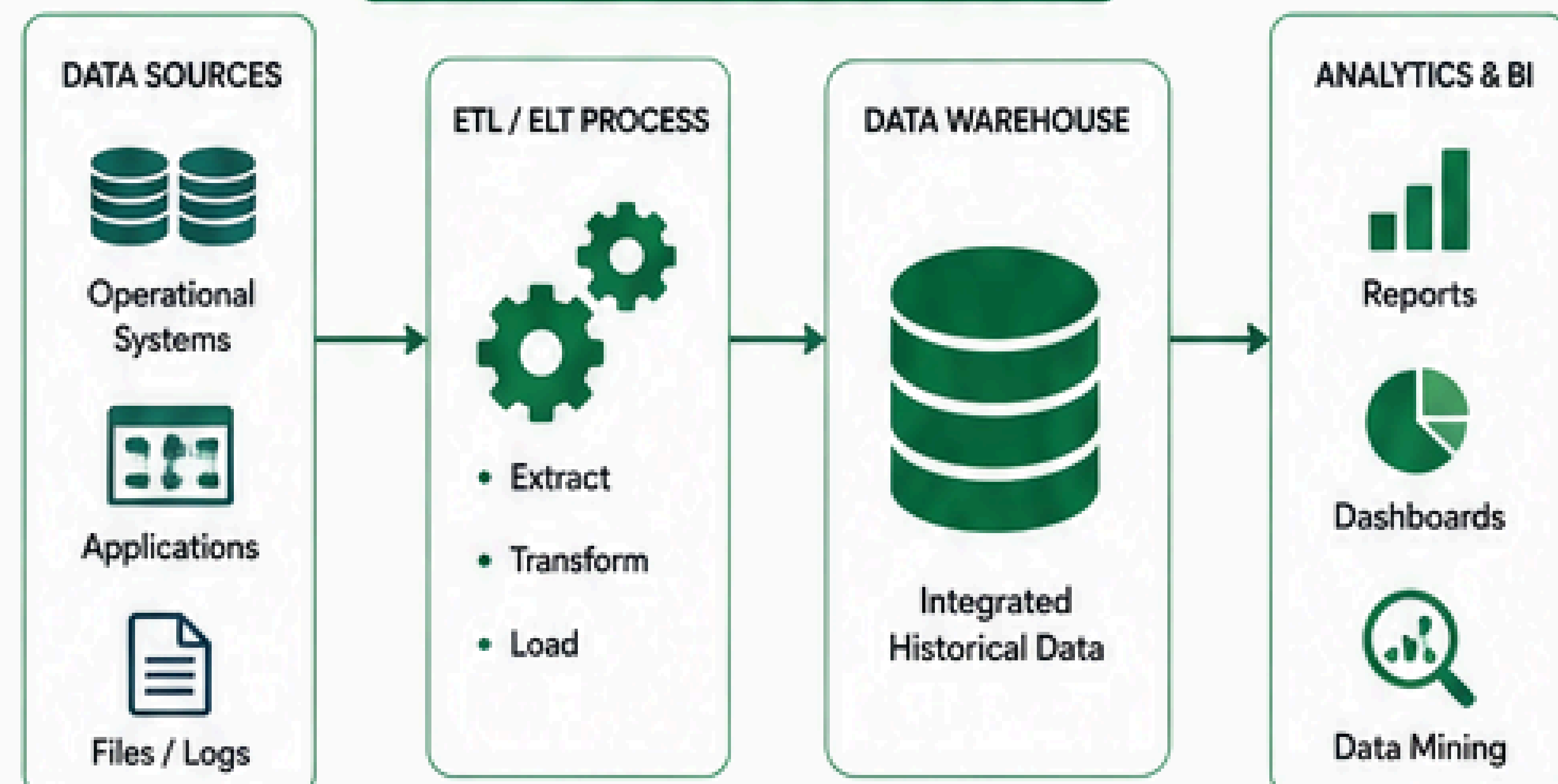
SNOWFLAKE ARCHITECTURE



• Data Warehouse

- Introduction to Data Warehousing
- OLAP vs OLTP
- Star Schema & Snowflake Schema
- Fact & Dimension Tables
- ETL/ELT Processes
- Data Modeling
- Data Quality & Governance
- Data Warehousing Best Practices
- Real-world Data Warehouse Projects

DATA WAREHOUSE ARCHITECTURE



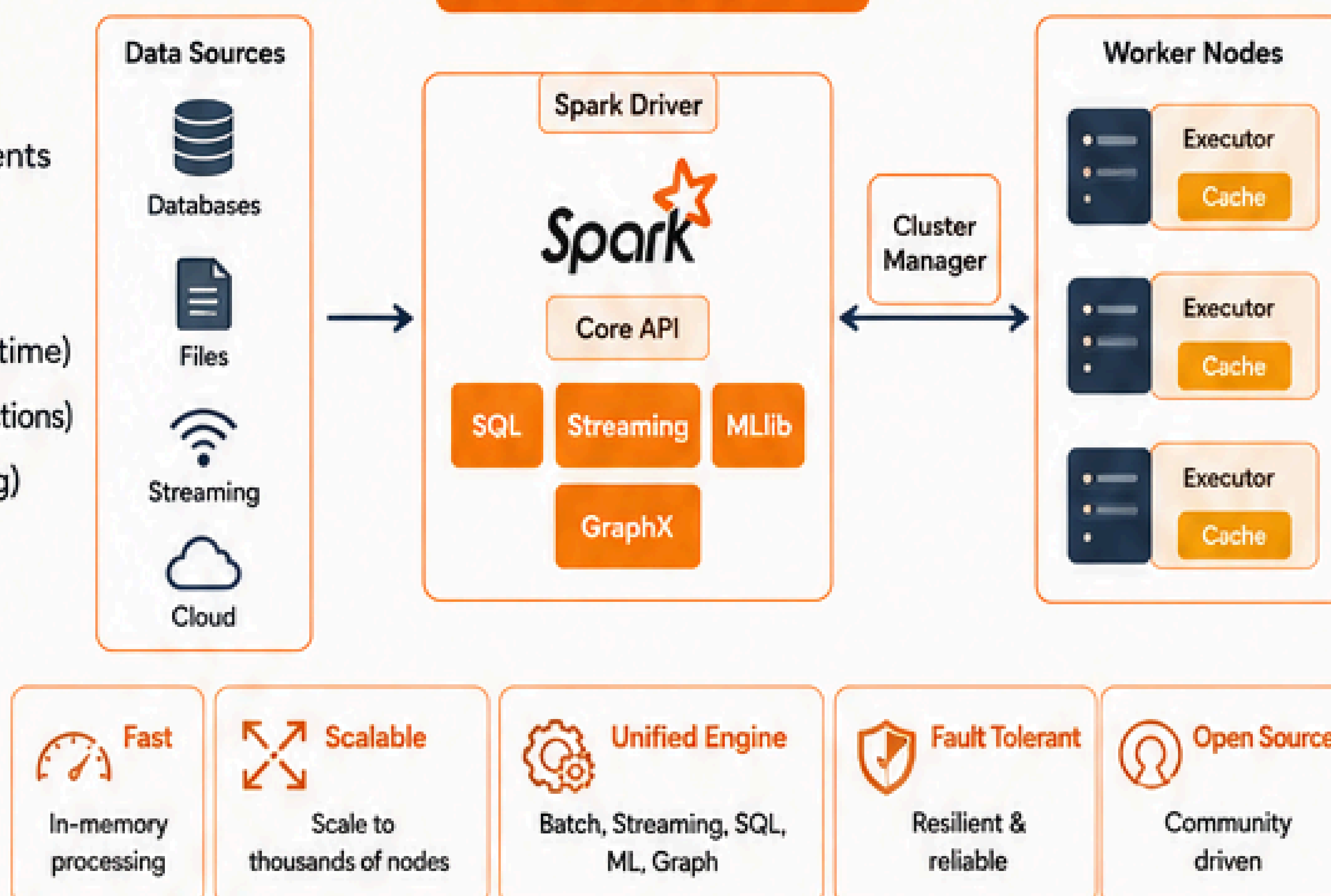
PHASE 5



Apache Spark

- Introduction to Apache Spark
- Spark Architecture & Components
- RDDs, DataFrames & Datasets
- Spark SQL
- Data Processing (Batch & Real-time)
- Spark Core (Transformations & Actions)
- Spark MLlib (Machine Learning)
- Spark Streaming
- Performance Tuning
- Use Cases & Best Practices
- Hands-on Projects

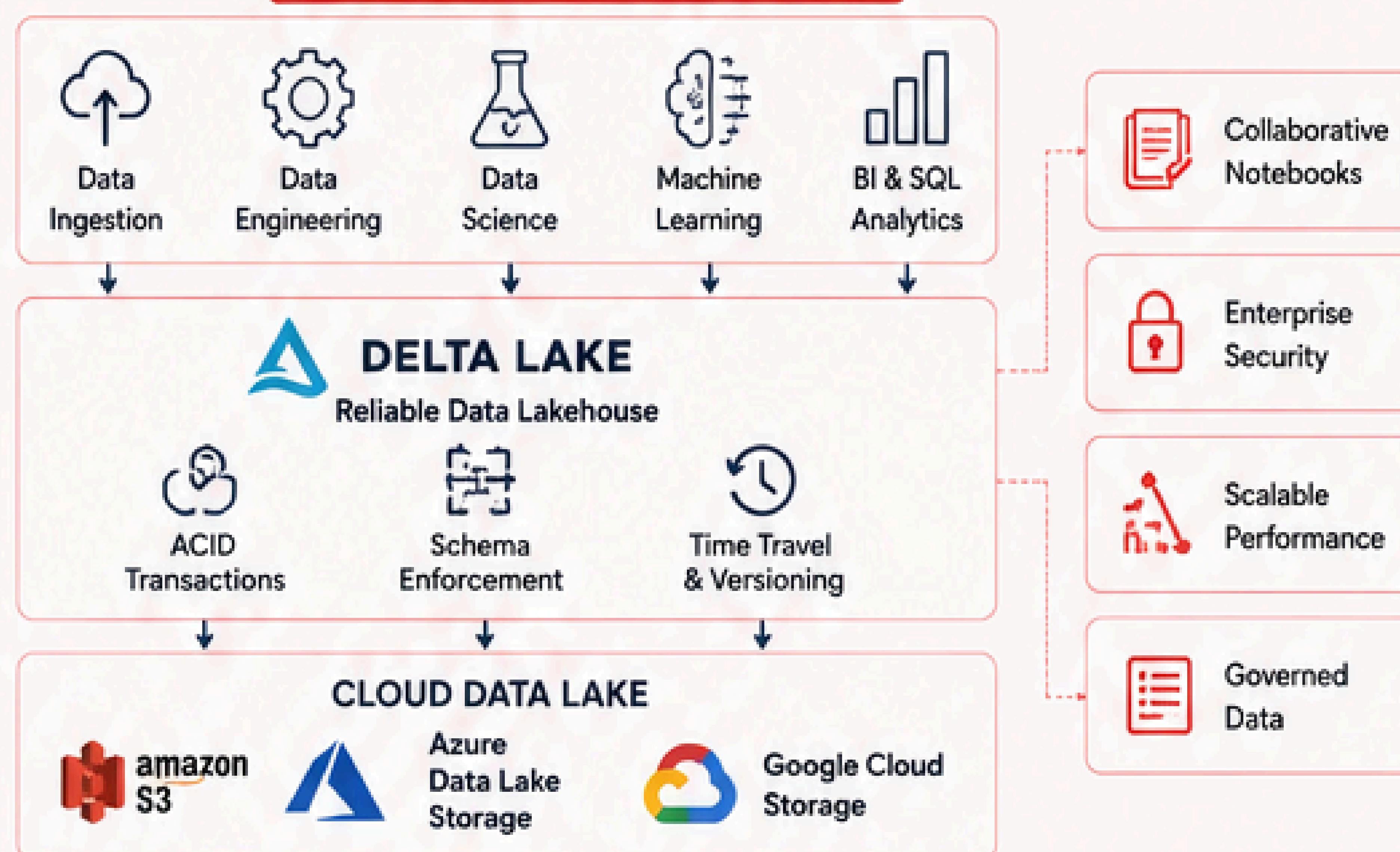
SPARK ARCHITECTURE



Databricks

- Introduction to Databricks
- Databricks Lakehouse Platform
- Clusters, Notebooks & Jobs
- Data Engineering Workflows
- Delta Lake Integration
- Unity Catalog (Governance)
- Data Sharing & Collaboration
- MLflow (ML Lifecycle)
- Auto Loader & Delta Live Tables
- Performance Optimization
- Hands-on Projects

DATABRICKS LAKEHOUSE PLATFORM



Unified Platform
One platform for all data workloads

Lakehouse Architecture
Open data formats with ACID reliability

Open & Interoperable
Open source technologies

Built for Scale
Petabyte scale and beyond

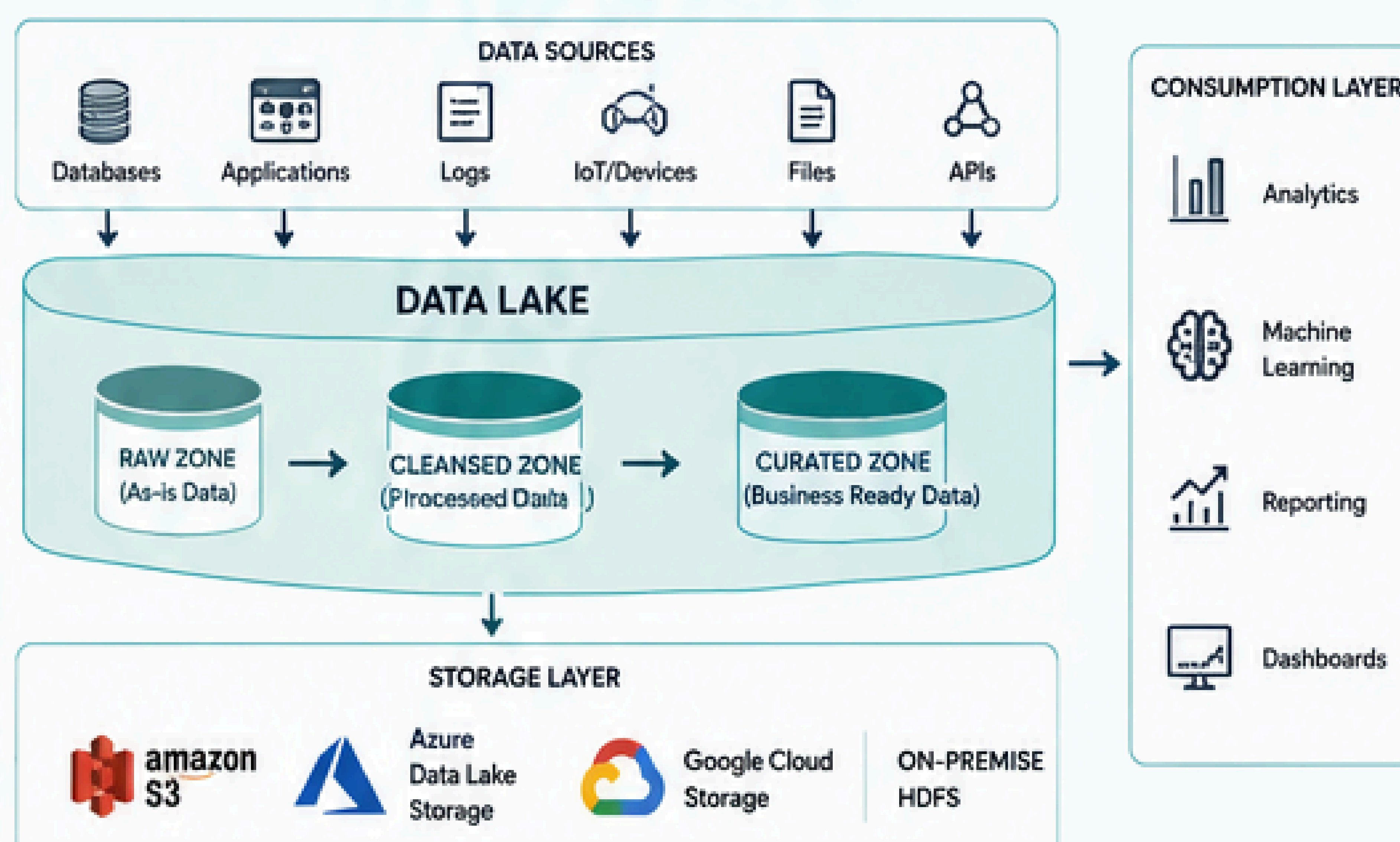
Cost Efficient
Optimized for cloud



Data Lake

- Introduction to Data Lake
- Data Lake vs Data Warehouse
- Data Lake Architecture
- Storage Layers (Raw, Cleansed, Curated)
- File Formats (Parquet, ORC, JSON, CSV)
- Data Ingestion Strategies
- Data Governance & Security
- Metadata Management
- Data Catalogs
- Use Cases
- Hands-on Projects

DATA LAKE ARCHITECTURE



Store Any Data
Structured, Semi-structured & Unstructured

Cost Effective
Leverage cheap cloud storage

Scalable
Scale to petabytes of data

Flexible
No schema on write, schema on read

Secure
Access control, encryption & governance

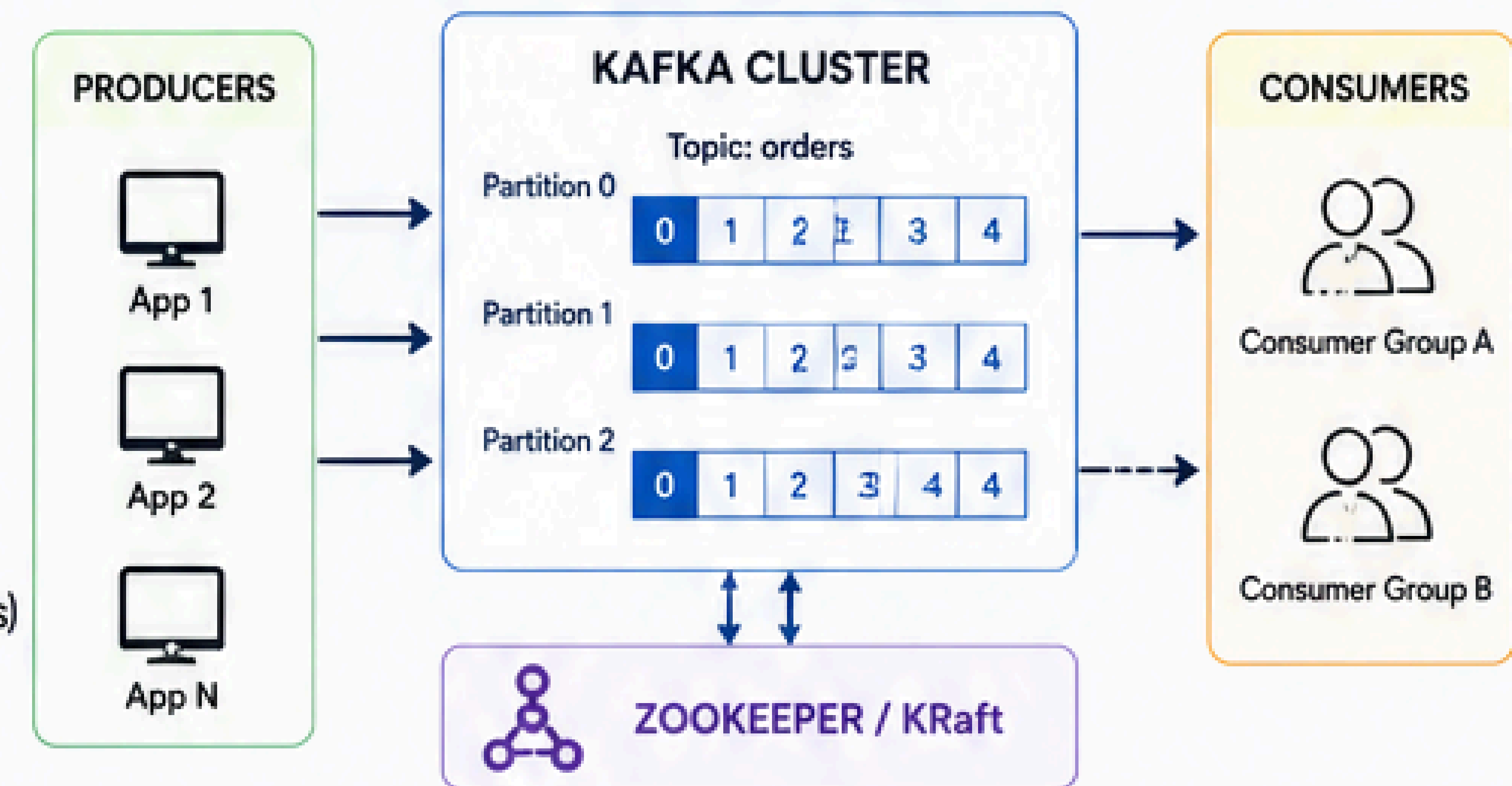
PHASE 6



• Kafka (ETC)

- Introduction to Kafka
- Kafka Architecture & Components (Broker, Topic, Partition, Offset)
- Producers, Consumers & Topics
- Kafka Brokers & Zookeeper / KRaft
- Partitions, Replication & ISR
- Consumer Groups
- Kafka Connect (Source & Sink Connectors)
- Kafka Streams & ksqiDB
- Schema Registry & Avro
- Exactly Once Semantics
- Monitoring with Kafka Tools
- Real-world Use Cases
- Hands-on Projects

KAFKA ARCHITECTURE



ECOSYSTEM



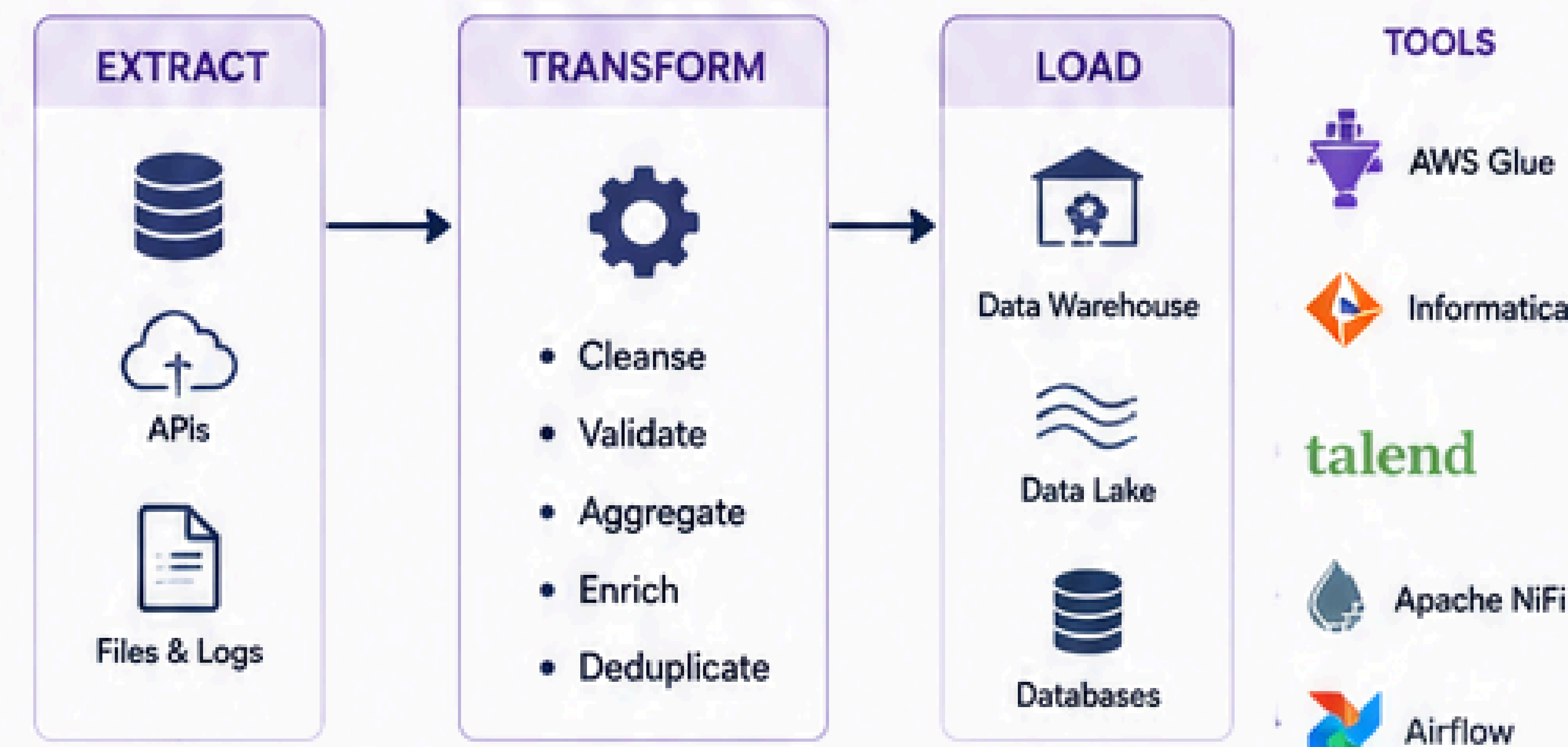
KEY BENEFITS



• ETL Processes

- Introduction to ETL
- Extract: Sources & Extraction Techniques
- Transform: Data Cleaning, Validation, Aggregation, Enrichment
- Load: Full Load, Incremental Load
- Types: Batch ETL vs Real-time ETL
- Handling Slowly Changing Dimensions (SCD Type 1, 2, 3)
- Error Handling & Logging
- Data Quality Checks
- ETL Tools Overview (Informatica, Talend, AWS Glue, etc.)
- Scheduling & Orchestration
- Hands-on ETL Project

ETL WORKFLOW



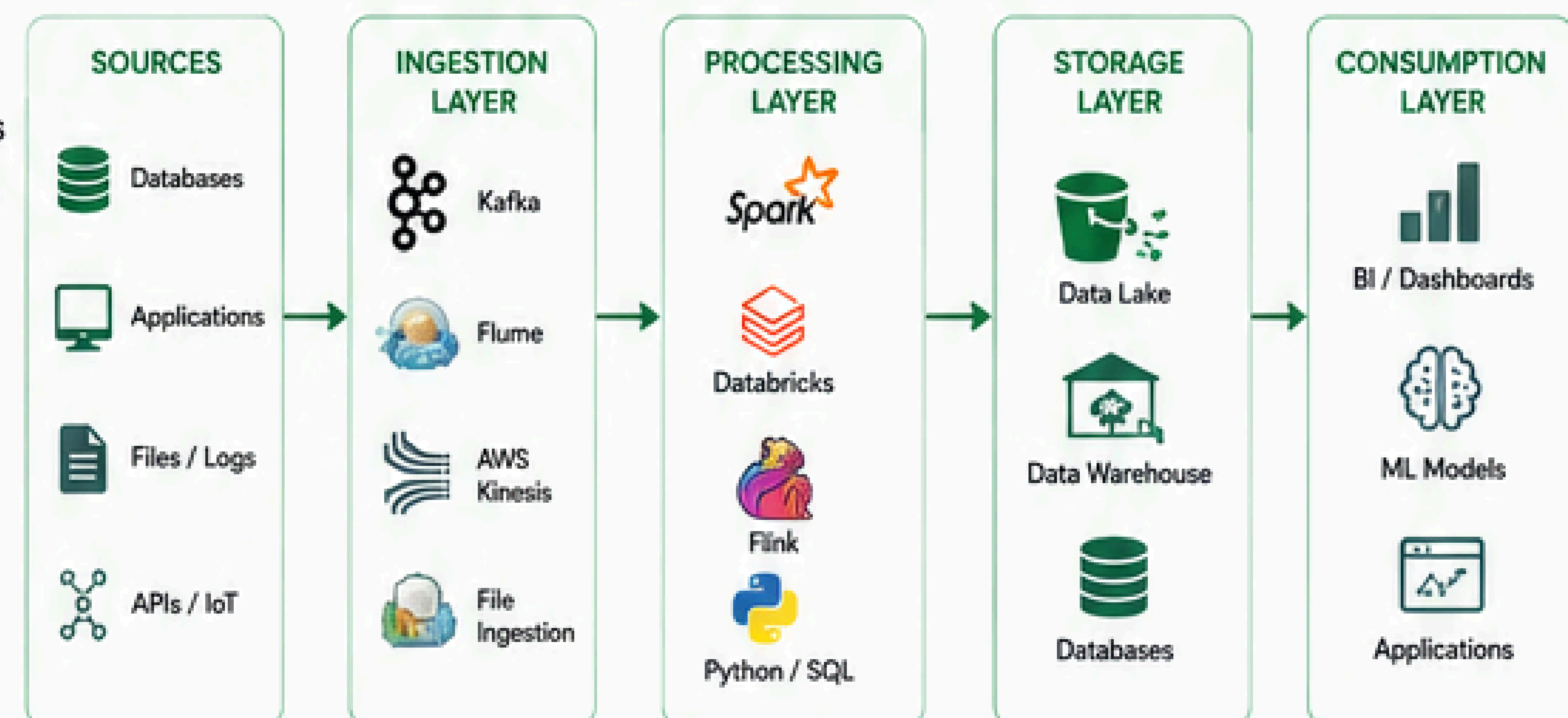
COMMON ETL PATTERNS



• Data Pipeline

- Introduction to Data Pipelines
- Batch vs Stream Processing Pipelines
- Pipeline Architecture & Components
- Ingestion Layer
- Processing Layer
- Storage Layer
- Orchestration & Scheduling
- Monitoring & Alerting
- Data Quality & Lineage
- Tools: Airflow, Kafka, Spark, Databricks, AWS, GCP, Azure
- End-to-End Pipeline Project

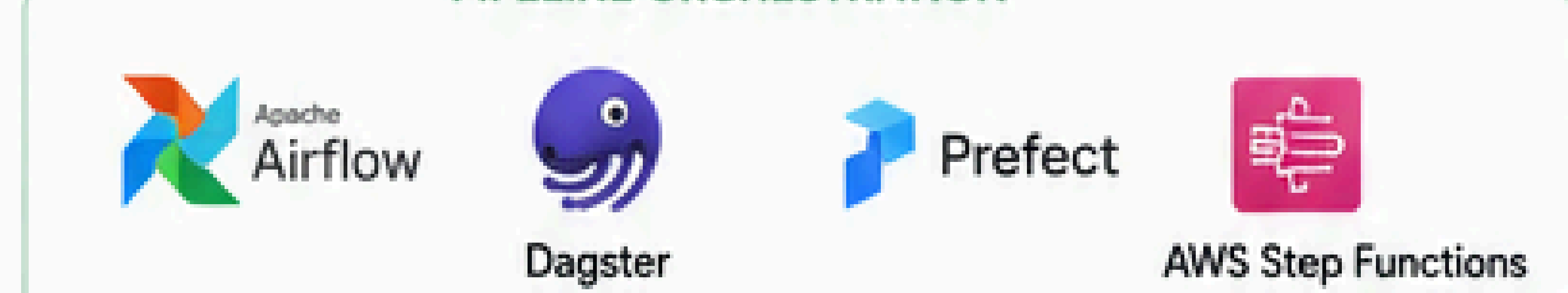
TYPICAL DATA PIPELINE ARCHITECTURE



PIPELINE BEST PRACTICES



PIPELINE ORCHESTRATION



PHASE 7



DATA ANALYTICS

• Data Analytics

- Introduction to Data Analytics
- Business Problem Solving
- Data Cleaning & Preparation
- Exploratory Data Analysis (EDA)
- Descriptive Statistics
- Inferential Statistics
- Data Analysis with Python (Pandas, NumPy, Matplotlib, Seaborn)
- KPIs & Metrics
- Trend Analysis & Forecasting
- Insight Generation & Business Recommendations
- Data Storytelling
- Tools: Python, Excel, SQL, Power BI

ANALYTICS WORKFLOW



SAMPLE ANALYTICS DASHBOARD



REAL WORLD INDUSTRY PROJECTS

• Real World Industry Projects

- End-to-End Project Development
- Work on Real Datasets
- Data Collection & Ingestion
- Data Cleaning & Processing
- Exploratory Data Analysis
- Data Modeling & Analysis
- Dashboard & Report Creation
- Business Insights & Recommendations
- Deployment (where applicable)
- Domain Exposure
- Project Documentation & Presentation

SAMPLE PROJECTS

Retail Sales Analysis Analyze sales trends, products, and customer behavior.	Customer Churn Prediction Build ML model to predict customer churn and reduce attrition.	E-commerce Analytics Analyze user behavior, conversion rate & revenue drivers.
Finance Analysis Analyze financial performance and risk metrics.	Marketing Campaign Analysis Evaluate campaign performance and ROI.	Supply Chain Analytics Optimize inventory, demand & delivery performance.



INTERVIEW QUESTIONS

• Interview Questions

- SQL Interview Questions
- Python Interview Questions
- Data Analytics & Statistics Questions
- Machine Learning Basics Questions
- Data Engineering Questions
- Cloud & Big Data Questions
- Scenario-Based Questions
- Behavioral Questions

QUESTION CATEGORIES

SQL	Joins, Subqueries, Window Functions, CTE, Aggregations
Python	Data Structures, Functions, OOPs, Pandas, NumPy
Statistics & Analytics	Mean, Median, Mode, Probability, Hypothesis Testing, Correlation
Machine Learning	Regression, Classification, Overfitting, Model Evaluation
Data Engineering	ETL, Data Pipelines, Kafka, Spark, Airflow
Cloud & Big Data	AWS, Snowflake, Databricks, Hadoop, Spark
Behavioral	Tell me about yourself, Strengths & Weaknesses, Teamwork, Problem Solving



MOCK INTERVIEW

• Mock Interview

- 1-on-1 Live Mock Interviews
- Role-based Interview Practice
- Technical + Behavioral Rounds
- Resume Review & Feedback
- Communication & Presentation Skills
- Performance Evaluation
- Improvement Feedback
- Build Confidence & Readiness

WHAT YOU GET

Real Interview Experience	Detailed Feedback	Performance Improvement	Expert Interviewers	Interview Readiness
---------------------------	-------------------	-------------------------	---------------------	---------------------

INTERVIEW ROLES COVERED

- Data Analyst
- Data Engineer
- Business Analyst
- Data Scientist
- ML Engineer

JOIN OUR COMMUNITY:



**Telegram
Channel**

<https://t.me/REGexSoftware>



**YouTube
Channel**

[https://www.youtube.com/
@REGexSoftware](https://www.youtube.com/@REGexSoftware)



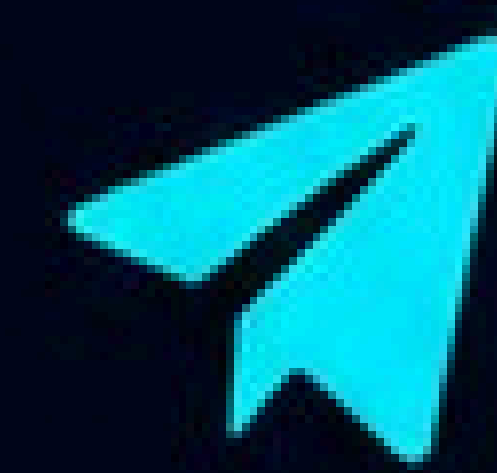
**Instagram
Profile**

[https://www.instagram.com/
regexsoftware/?hl=en](https://www.instagram.com/regexsoftware/?hl=en)



**LinkedIn
Company Page**

[https://www.linkedin.com/
company/regexsoftware/
posts/?feedView=all](https://www.linkedin.com/company/regexsoftware/posts/?feedView=all)



JOIN TELEGRAM



JOIN YOUTUBE



JOIN INSTAGRAM



JOIN LINKEDIN