Custom Course

Day 2

Big Data Fundamentals

4 V's of Big Data:

1. Volume: Scale of data
2. Variety: Forms of data
3. Velocity: Analysis of data flow
4. Veracity: Uncertainty of data

2 types of data processing:

1. Batch Processing: Processes data in batches, Example: Transaction in bank
2. Stream Processing: Processes stream data/live data, Example: Cricket match

Stream analytics helps e-commerce business to get 360° data of the customers and improve their customer service.

Parallel and Distributive Processing:

In parallel computing multiple processors performs multiple tasks assigned to them simultaneously.

In distributed computing we have multiple autonomous computers which seems to the user as single system.

S.NO	Parallel Computing	Distributed Computing
1.	Many operations are performed simultaneously	System components are located at different locations
2.	Single computer is required	Uses multiple computers
3.	Multiple processors perform multiple operations	Multiple computers perform multiple operations

4.	It may have shared or distributed memory	It have only distributed memory
5.	Processors communicate with each other through bus	Computer communicate with each other through message passing.
6.	Improves the system performance	Improves system scalability, fault tolerance and resource sharing capabilities

Data WareHouse:

- same data as object (tables, views)
- structured data
- dimension modeling
- follows ACID properties
- follows fixed schema

Data Lake:

- supports all type of data like structured data, semi-structured data, unstructured data
- follows flexible schema
- ACID properties are not available here

Lake House:

it uses the properties of both data warehouse and data lake.

Data Security:

Authentication: We define who we are.

Authorization: We define what kind of data we have access to.

Cloud Fundamentals:

Private Cloud: Completely owned by an organization.

Public Cloud: Owned by cloud services or hosting provider.

Hybrid Cloud: Combines both private and public cloud.

Cloud Benefits:

1. High availability
2. Scalability
3. Global reach
4. Agility
5. Disaster recovery
6. Fault tolerance
7. Elasticity
8. Customer latency capabilities
9. Predictive cost considerations
10. Security

2 types of models are there:

1. Capital Expenditure (Capex): Up-front spending of money on physical infrastructure.
2. Operational Expenditure (Opex): Spend on products and services as needed, pay-as-you-go.

IaaS: Ex: Servers, Network firewalls, Data center plant

PaaS: Ex: OS, Dev Tools, DBMS, Business Analytics.

SaaS: Ex: Microsoft office, calendar

Serverless Computing: Cloud provider automatically provisions, scales, and manages the infrastructure required to run the code. Ex: Azure Functions, Azure Logic Apps

Azure Fundamentals:

Azure Subscription: Provides you with authenticated and authorized access to your Azure accounts.

Azure Resource Manager: Provides you the management layer to manage your Azure subscription.

Resource Group: Logical container to manage and aggregate resources in a single unit.