**Data Engineering & Azure Fundamentals**

**1. Introduction to Data Engineering**

**Definition**

Data Engineering is the discipline of designing, building, and managing systems that collect, store, process, and make data available for analysis. It ensures that organizations can derive insights and make data-driven decisions.

**Roles & Responsibilities of a Data Engineer**

* **Data Ingestion**: Building pipelines to bring data from multiple sources.
* **Data Transformation**: Cleaning, aggregating, and preparing raw data into usable formats.
* **Data Storage**: Designing and maintaining databases, data lakes, and warehouses.
* **Performance Optimization**: Ensuring efficient queries and reduced processing times.
* **Automation**: Using scripts and tools to automate workflows.
* **Collaboration**: Working with data scientists, analysts, and business teams.
* **Governance & Security**: Ensuring compliance, access control, and data protection.

**2. Types of Data**

1. **Structured Data**
   * Organized into rows and columns (relational databases).
   * Examples: Customer details in SQL database, financial transactions.
2. **Semi-Structured Data**
   * Contains some organizational structure but not strictly tabular.
   * Examples: JSON, XML, sensor logs.
3. **Unstructured Data**
   * No predefined model; difficult to store in relational databases.
   * Examples: Images, videos, audio files, social media posts.

**3. Batch vs Real-Time Processing**

**Batch Processing**

* Processes large volumes of data in groups (batches).
* Data is collected over a period, then processed.
* **Use Cases**: Payroll systems, monthly sales reports.
* **Pros**: Cost-effective, handles huge datasets.
* **Cons**: High latency; not suitable for immediate decisions.

**Real-Time Processing**

* Processes data continuously as it arrives.
* **Use Cases**: Fraud detection, stock trading, IoT sensors.
* **Pros**: Low latency, supports real-time insights.
* **Cons**: Expensive, complex architecture.

**4. Cloud Computing Basics**

Cloud computing delivers computing resources (servers, storage, networking, databases, etc.) over the internet.

**Models:**

* **IaaS (Infrastructure as a Service)**:  
  Raw infrastructure (VMs, storage). Users manage OS, apps.  
  *Example*: Azure Virtual Machines.
* **PaaS (Platform as a Service)**:  
  Pre-configured platform for app development.  
  *Example*: Azure App Service, Azure SQL Database.
* **SaaS (Software as a Service)**:  
  Fully managed applications delivered over the internet.  
  *Example*: Microsoft 365, Salesforce.

**5. Azure Global Infrastructure**

* **Regions**: Physical locations hosting data centers (e.g., East US, West Europe).
* **Availability Zones (AZs)**: Isolated data centers within a region to ensure high availability.
* **Global Network**: Provides redundancy, low latency, and disaster recovery.

**6. Azure Resource Hierarchy**

1. **Subscription**: The top-level container with billing and access controls.
2. **Resource Groups**: Logical containers for resources (VMs, storage, databases).
3. **Resources**: Individual services (VM, SQL DB, Storage Account).

**Best Practice**: Organize resources logically by application, department, or project.

**7. Navigating the Azure Portal**

* Use **search bar** for quick navigation.
* Pin frequently used services to the **dashboard**.
* Check **Notifications** for deployment status.
* Use **Resource Graph Explorer** for complex queries.
* Practice **tags** for resource organization.

**8. Azure CLI & PowerShell Basics**

**Azure CLI (Command Line Interface)**

* Cross-platform tool for managing Azure resources.
* Example commands:
* az login
* az group create --name myResourceGroup --location eastus
* az vm create --resource-group myResourceGroup --name myVM --image UbuntuLTS

**Azure PowerShell**

* PowerShell cmdlets designed for Azure management.
* Example commands:
* Connect-AzAccount
* New-AzResourceGroup -Name myResourceGroup -Location eastus
* New-AzVM -ResourceGroupName myResourceGroup -Name myVM -Image UbuntuLTS

**9. Cost Management Basics in Azure**

* **Azure Pricing Calculator**: Estimate cost of resources before deployment.
* **Azure Cost Management + Billing**: Monitor and optimize costs.
* **Budgets & Alerts**: Set thresholds to control spending.
* **Right-Sizing Resources**: Scale up/down based on demand.
* **Reserved Instances & Spot Pricing**: Reduce cost for predictable workloads.