**Lab Guide: Data Engineering & Azure Fundamentals**

**Lab 1: Introduction to Data Engineering**

**Objective:** Understand what data engineering is, the role of a data engineer, and responsibilities.

**Steps:**

1. Open a new document/notebook (Word/Notepad/Jupyter).
2. Define **Data Engineering**: "Designing, building, and managing scalable data pipelines to enable analytics & machine learning."
3. List **Roles & Responsibilities**:
   * Collect & ingest data from multiple sources.
   * Transform & clean data for analytics.
   * Manage storage (Data Lake, Warehouses).
   * Ensure data governance & security.
4. Write down real-world examples:
   * Streaming click data from e-commerce.
   * Batch ETL for finance reporting.

**Lab 2: Types of Data**

**Objective:** Identify and classify structured, semi-structured, and unstructured data.

**Steps:**

1. Create a folder data\_types\_lab.
2. Inside it, create:
   * structured.csv → Add tabular sales data.
   * semi\_structured.json → Add nested JSON of an order with customer details.
   * unstructured.txt → Paste some free text reviews or images.
3. Discuss differences:
   * **Structured** → Fits in rows/columns (SQL).
   * **Semi-structured** → Flexible schema (JSON, XML).
   * **Unstructured** → Free form (text, images, audio).

**Lab 3: Batch vs Real-Time Processing**

**Objective:** Compare batch and streaming data flows.

**Steps:**

1. Create two files:
   * batch\_orders.csv → List of 100 daily orders.
   * stream\_orders.txt → Append one new order every 5 seconds (simulate real-time).
2. Open Python/Jupyter:
   * Read batch\_orders.csv with pandas.
   * Tail stream\_orders.txt using watch or tail -f in terminal.
3. Write down:
   * Batch Pros: Good for large historical data.
   * Batch Cons: Latency.
   * Real-time Pros: Low latency.
   * Real-time Cons: Higher infra cost.

**Lab 4: Cloud Computing Basics (IaaS, PaaS, SaaS)**

**Objective:** Understand different service models.

**Steps:**

1. In Azure Portal, create:
   * **VM (IaaS)** → Choose Ubuntu Server.
   * **Azure SQL Database (PaaS)** → Serverless option.
   * **Power BI (SaaS)** → Use trial/free version.
2. Note differences:
   * IaaS = Full control, infra responsibility.
   * PaaS = Managed infra, you focus on apps/data.
   * SaaS = Fully managed software.

**Lab 5: Azure Global Infrastructure**

**Objective:** Explore Azure regions and availability zones.

**Steps:**

1. Go to Azure Region Map.
2. Open Azure Portal → Check **Subscriptions → Properties → Region**.
3. Create a resource (VM/Storage) in East US and another in West Europe.
4. Document:
   * Region = Geographic location.
   * AZ = Redundancy inside a region.

**Lab 6: Azure Resource Hierarchy**

**Objective:** Understand how Azure resources are organized.

**Steps:**

1. In Azure Portal:
   * Create a new **Subscription** (if available, otherwise use existing).
   * Inside, create a **Resource Group** named RG-DataLab.
   * Inside, add a **Storage Account**.
2. Note hierarchy:
   * Subscription → Resource Group → Resource.

**Lab 7: Navigating Azure Portal**

**Objective:** Learn how to use Azure Portal efficiently.

**Steps:**

1. Log in at portal.azure.com.
2. Pin your frequently used services (VM, Storage, SQL).
3. Create custom dashboard with tiles.
4. Use **Search Bar** to quickly access services.

**Lab 8: Azure CLI Basics**

**Objective:** Run commands in Azure CLI.

**Steps:**

1. Install Azure CLI:
   * Windows: winget install Microsoft.AzureCLI
   * Linux/Mac: curl -sL https://aka.ms/InstallAzureCLIDeb | sudo bash
2. Run commands:
3. az login
4. az account list
5. az group create --name RG-CLI-Lab --location eastus
6. az storage account create --name datalabstorage123 --resource-group RG-CLI-Lab --location eastus --sku Standard\_LRS
7. Verify in Portal.

**Lab 9: PowerShell Basics**

**Objective:** Use PowerShell for scripting.

**Steps:**

1. Install Az PowerShell:
2. Install-Module -Name Az -AllowClobber -Scope CurrentUser
3. Connect-AzAccount
4. Run:
5. Get-AzSubscription
6. New-AzResourceGroup -Name RG-PS-Lab -Location "WestEurope"
7. New-AzStorageAccount -ResourceGroupName RG-PS-Lab -Name psdatalabstore -Location "WestEurope" -SkuName Standard\_LRS

**Lab 10: Cost Management**

**Objective:** Monitor and manage Azure costs.

**Steps:**

1. Go to **Azure Portal → Cost Management + Billing**.
2. Check **Cost Analysis** for your subscription.
3. Set up **Budget Alert**:
   * Example: Budget = $10.
   * Notification at 80% usage.
4. Document differences between **Pay-as-you-go** vs **Reserved Instances**.