**Lab Title: *Secure, Monitor, and Optimize Data Movement with Azure Data Factory***

**Objective**

To practice and understand key ADF capabilities:

* Integration Runtime setup
* Secure authentication (Managed Identity & Key Vault)
* Copy activity with performance and fault-tolerance
* Monitoring and cost optimization

**Pre-requisites**

* Azure subscription
* Contributor access to a Resource Group
* Storage Account (source and sink)
* Azure Key Vault (optional but recommended)

**Step 1: Create Azure Data Factory**

1. Navigate to Azure Portal → **Create a resource** → **Data Factory**
2. Fill details:
   * Subscription: your-subscription
   * RG: ADF-Lab-RG
   * Name: adf-lab-factory
   * Region: East US
3. Select **Enable Managed VNet** during creation.
4. Click **Create** and wait for deployment.

**Step 2: Configure Integration Runtime (IR)**

**2.1 Azure IR (default)**

1. Go to ADF Studio → Manage → **Integration Runtimes**
2. Observe the default Azure IR

**2.2 Self-hosted IR (optional if on-prem connectivity)**

1. Click **New** → Select **Self-hosted IR**
2. Download the IR client on a local machine and install
3. Register using the authentication key from ADF

**2.3 Managed VNet with Private Endpoints (optional advanced)**

1. Go to Manage → **Managed private endpoints**
2. Create a private endpoint to a **Blob Storage** resource

**Step 3: Create Linked Services**

**3.1 Using Azure Managed Identity**

1. Go to **Manage → Linked Services → New**
2. Select Azure Blob Storage
3. Use **Managed Identity**
4. Select your Storage Account
5. Click **Test Connection** and Save

**3.2 Using Azure Key Vault**

1. In Key Vault, store a secret:
   * Name: blob-conn-string
   * Value: Connection string of another Storage Account
2. In ADF → Manage → Linked Services → New
3. Choose Azure Key Vault → Provide vault name
4. In Linked Service for Blob, choose "Use Key Vault secret" and reference secret
5. Test and Save

**Step 4: Understand Dataset vs Linked Service**

**4.1 Create Linked Service**

* Already created in Step 3

**4.2 Create Dataset**

1. Go to Author → Datasets → New Dataset → Azure Blob → CSV
2. Choose the Linked Service created in Step 3
3. Set path and structure (delimiter, headers etc.)
4. Name it src\_csv\_dataset

**Step 5: Copy Activity with Options**

**5.1 Create Pipeline**

1. Author → Pipelines → New Pipeline
2. Name: copy-pipeline-basic
3. Add **Copy Data** activity

**5.2 Source:**

* Use src\_csv\_dataset
* Preview source data

**5.3 Sink:**

* Create a new dataset pointing to sink Blob container
* Choose file format (e.g., JSON, Parquet)

**5.4 Options:**

* In Copy Activity → Settings tab:
  + Enable **Fault tolerance**:
    - Skip incompatible rows
    - Log error rows
  + Enable **Staging** if large file (optional)
  + Enable **Parallelism** (if structured source supports)

**Step 6: Add Retry and Timeout Settings**

1. In the Copy Activity:
   * Retry: 3
   * Timeout: 00:10:00
2. Save and publish the pipeline

**Step 7: Trigger and Monitor Pipeline**

1. Trigger the pipeline manually
2. Go to **Monitor Hub**
3. Observe:
   * Pipeline run duration
   * Copy activity logs
   * Errors or warnings
4. Drill into activity to view data read/written

**Step 8: Setup Alerts via Azure Monitor**

1. Go to Azure Portal → Monitor → Alerts → New Alert Rule
2. Resource: your ADF instance
3. Condition: **Activity Failed Count > 0**
4. Action Group:
   * Email or Teams webhook (if already created)
5. Severity: Choose appropriate (e.g., Sev 2 or Sev 3)

**Step 9: Analyze Cost of ADF Usage**

1. Go to **Cost Management + Billing**
2. Filter by Resource: adf-lab-factory
3. View daily/monthly cost
   * Data movement
   * IR usage (vCore Hours)
4. Optional: Enable cost alerts