**Microsoft Purview Governance & Automation – Hands-on Lab Guide**

**Lab 1: Setting up Collection Hierarchy & RBAC**

**Objective**: Create collections and assign roles to users/groups.

**Steps**:

1. Open [Microsoft Purview Studio](https://web.purview.azure.com/).
2. Go to **Collections** > **New Collection**.
3. Name the collection (e.g., Finance-Data) and assign a parent if needed.
4. After creation, click on the collection and go to **Access Control (IAM)**.
5. Assign users/groups to roles: Collection Admin, Data Curator, or Data Reader.

**Expected Outcome**: A structured collection with assigned access roles.

**Lab 2: Configure and Run a Scan**

**Objective**: Scan metadata from an Azure Data Lake or SQL source.

**Steps**:

1. Go to **Sources** > **Register** a new source (e.g., Azure Blob Storage).
2. Set authentication method (Managed Identity or Key Vault).
3. After registration, click **New Scan**.
4. Set up rule set, scan scope, and schedule.
5. Run the scan and monitor progress in the **Monitor > Scan** section.

**Expected Outcome**: Metadata is extracted and available in the catalog.

**Lab 3: Create Custom Classification Rule**

**Objective**: Create a custom classifier using regex for PAN or Employee IDs.

**Steps**:

1. Go to **Management Center > Classification Rules**.
2. Click **+ New classification rule**.
3. Name it PAN-Classifier and add a regex like [A-Z]{5}[0-9]{4}[A-Z].
4. Set confidence and description.
5. Apply the rule to relevant scan rule sets.

**Expected Outcome**: Custom patterns are detected in future scans.

**Lab 4: Define and Assign Glossary Terms**

**Objective**: Create business glossary terms and link them to data assets.

**Steps**:

1. Go to **Glossary > New Term**.
2. Add name: Customer, description, and synonyms.
3. Save the term.
4. Go to **Browse assets**, select an asset, click **Manage terms**.
5. Link the Customer term to columns like CustomerID, CustomerName.

**Expected Outcome**: Assets are enriched with business definitions.

**Lab 5: Push Lineage from Databricks to Purview**

**Objective**: Capture transformation lineage from Databricks notebook.

**Steps**:

1. Configure Purview account access in Databricks (via SPN or MI).
2. Use REST API to send lineage data:
   * Source: raw file/table
   * Process: Databricks notebook/job
   * Target: cleansed table
3. Validate lineage in Purview under **Lineage** view.

**Expected Outcome**: End-to-end lineage graph from source to target.

**Lab 6: Monitor Governance Scorecard**

**Objective**: Review metadata and governance adoption metrics.

**Steps**:

1. Go to **Insights > Governance Scorecard**.
2. Explore:
   * Scanned asset coverage
   * Classification coverage
   * Glossary usage
3. Review trends and identify under-governed areas.

**Expected Outcome**: Governance health is visible with improvement areas.

**Lab 7: Automate Scan via DevOps Pipeline**

**Objective**: Trigger Purview scan from Azure DevOps pipeline.

**Steps**:

1. Store scan config in a parameterized ARM template.
2. In Azure DevOps, create a pipeline with PowerShell or REST API task.
3. Use a Service Principal with Purview permissions.
4. Run pipeline to deploy scan configuration.
5. Validate in Purview > Monitor > Scan history.

**Expected Outcome**: Scans are automated through CI/CD pipelines.

**Lab Tips**:

* Use dummy data files with known patterns for testing classification.
* Monitor activity using Azure Monitor or Log Analytics workspace.
* Reuse glossary terms across departments for consistency.