# LAB 01: Platform & Workspace Orientation

**Duration:** ~35 min | **Day:** 1 | **After module:** M01: Platform & Workspace | **Difficulty:** Beginner

---

## Scenario

> *"Your first day at RetailHub! Before diving into data engineering, familiarize yourself with the Databricks workspace, Unity Catalog structure, and the tools you'll use throughout the training."*

---

## Objectives

After completing this lab you will be able to: - Create and configure a Databricks cluster - Navigate the Unity Catalog hierarchy (Catalog → Schema → Table) - Explore external connections and volumes - Upload dataset files to a Databricks Volume - Use `dbutils` for file exploration - Read CSV files into DataFrames

---

## Prerequisites

- Access to the Databricks workspace
- Trainer has run `00_pre_config.ipynb` to provision your catalog

---

## Part 1: Create Your Cluster

1. Go to **Compute** → **Create Cluster**
2. Set the cluster name: `<your_name>_cluster`
3. Select **Single Node** mode
4. Choose **Runtime 15.4 LTS** or newer
5. Click **Create** and wait for the cluster to start

## Part 2: Explore Unity Catalog

1. Open **Catalog** in the left sidebar
2. Navigate: `retailhub_<your_name>` → `bronze` → Tables
3. Observe the three-level namespace: `catalog.schema.table`

## Part 3: External Connections

1. Explore the **External Data** section in Catalog
2. Understand how external connections link to cloud storage

## Part 4: Upload Files

1. Navigate to your catalog's **default** schema → **Volumes**
2. Upload the dataset files from `dataset/` folder
3. Verify files are accessible

## Part 5: Notebook Tasks

Open **`LAB_01_code.ipynb`** and complete the `# TODO` cells.

| Task | What to do | Key concept |
|---|---|---|
| **Task 1** | Verify Catalog Context | `SELECT current_catalog(), current_schema()` |
| **Task 2** | List Files in Volume | |

| Task | What to do | Key concept |
|---|---|---|
| | | `dbutils.fs.ls()` on dataset path |
| **Task 3** | Read CSV into DataFrame | `spark.read.csv(path, header=True)` |
| **Task 4** | Inspect Schema | `.printSchema()`, `.dtypes` |
| **Task 5** | Explore dbutils | `dbutils.fs.head()`, `dbutils.help()` |

---

## Summary

In this lab you: - Created your first Databricks cluster - Explored the Unity Catalog namespace - Uploaded datasets to a Volume - Read CSV files with Spark and inspected schemas - Used `dbutils` for file system operations

> **Exam Tip:** Unity Catalog uses a 3-level namespace: `catalog.schema.table`. `dbutils.fs.ls()` lists files in cloud storage. `spark.read.csv(path, header=True)` reads CSV with headers.

> **What's next:** In LAB 02 you will load data using explicit schemas, transform DataFrames, and save Delta tables.