



Day 1 | Databricks Basics Platform Setup & First Steps

14 DAYS
AI CHALLENGE

DAY 01

Topic:
Platform Setup & First Steps

Challenge:

1. Create Databricks Community Edition account
2. Navigate Workspace, Compute, Data Explorer
3. Create first notebook
4. Run basic PySpark commands

#DatabricksWithIDC



Why Databricks?

Before Databricks ↴

- ✗ Pandas → works on single machine
- ✗ Hadoop → powerful but complex

Databricks solves both

- ✓ Big data processing
- ✓ Simple SQL & Python
- ✓ Scalable & collaborative





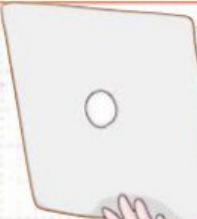
What is Databricks?

Databricks is a cloud-based data platform that helps teams:

- *Store data*
- *Process large data*
- *Analyze using SQL / Python*
- *Build ML models*



All in one unified platform





Lakehouse Architecture (Simple)

Traditional systems 

- *Data Lake → storage*
- *Data Warehouse → analytics*

 *Separate systems = complexity*

Lakehouse = Best of both

-  *Low-cost storage*
-  *Fast analytics*
-  *One source of truth*

**Data Lake + Data WareHouse
=LakeHouse**





Databricks on AWS (High Level)

Databricks runs on top of AWS

- Data stored in Amazon S3
- Compute powered by Apache Spark
- Managed by Databricks Workspace

👉 You focus on data, not infrastructure

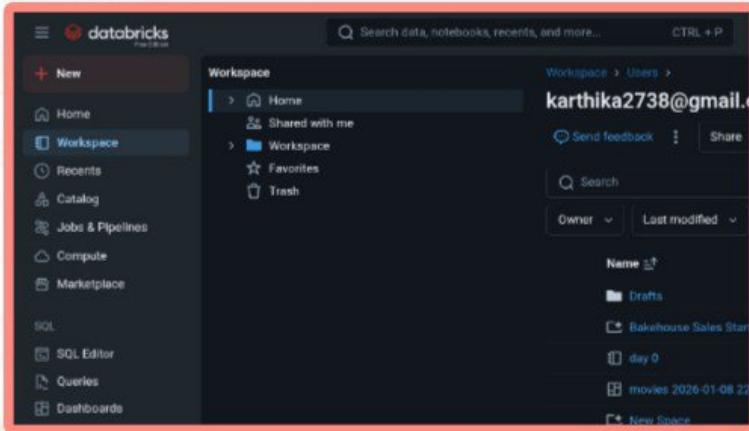




Databricks Workspace

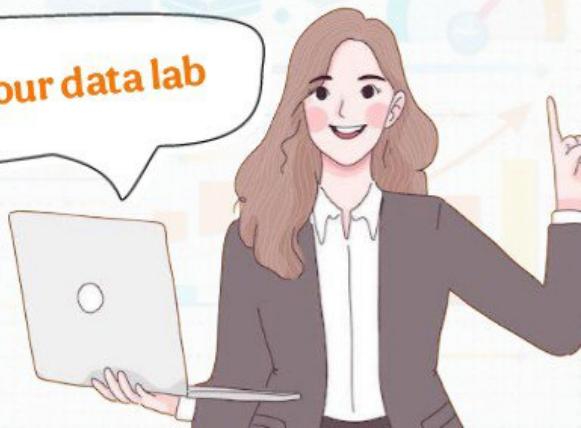
Workspace is where you work daily 

-  **Notebooks** – write SQL / PySpark
-  **Collaboration** – share notebooks
-  **Organized folders**



The screenshot shows the Databricks workspace interface. On the left, there's a sidebar with options like 'New', 'Home', 'Workspace' (which is selected), 'Recents', 'Catalog', 'Jobs & Pipelines', 'Compute', 'Marketplace', 'SQL', 'SQL Editor', 'Queries', and 'Dashboards'. The main area has a search bar at the top. Below it, there's a breadcrumb navigation 'Workspace > Users > karthika2738@gmail.com'. A 'Send feedback' button and a 'Share' button are also present. The main content area lists notebooks under 'Name' (sorted by name). The list includes 'Drafts', 'Bakehouse Sales Star', 'day 0', and 'movies 2026-01-08 22:00:00'. There's also a '+ New Space' button.

👉 Think of it as your data lab





Compute (Clusters)

Compute = engines that run your code

- *Spark clusters*
- *Can scale up/down*
- *Needed to run notebooks*

The screenshot shows the Databricks interface with the 'Compute' tab selected in the sidebar. The main view is titled 'Compute' and displays a list of 'SQL warehouses'. A single entry is visible: 'Serverless Starter Warehouse' created by 'karthika2738@gmail.com' (2X-Small, 0 / 1 Active / Max, Serverless). The sidebar also includes links for Home, Workspace, Recents, Catalog, Jobs & Pipelines, Marketplace, SQL, SQL Editor, Queries, Dashboards, Genie, and Alerts.

Status	Name	Created by	Size	Active / Max	Type
✓	Serverless Starter Warehouse	karthika2738@gmail.com	2X-Small	0 / 1	Serverless

👉 No cluster = no execution





Day 1 Tasks Completed ✓

- ✓ *Created Databricks Community Edition account*
- ✓ *Explored Workspace & Compute*
- ✓ *Created first notebook*
- ✓ *Ran basic PySpark commands*

Small steps → Strong foundation 🎉





Key Takeaway

- 📌 *Databricks is not just a tool*
- 📌 *It's a complete data platform*
- 📌 *Day 1 = Understanding the ecosystem*



Ready for Day 2!





- 📖 *Learning Databricks from official docs*
- 🔧 *Practicing hands-on daily*
- 📢 *Sharing my journey openly*

Follow along for
Day 2

