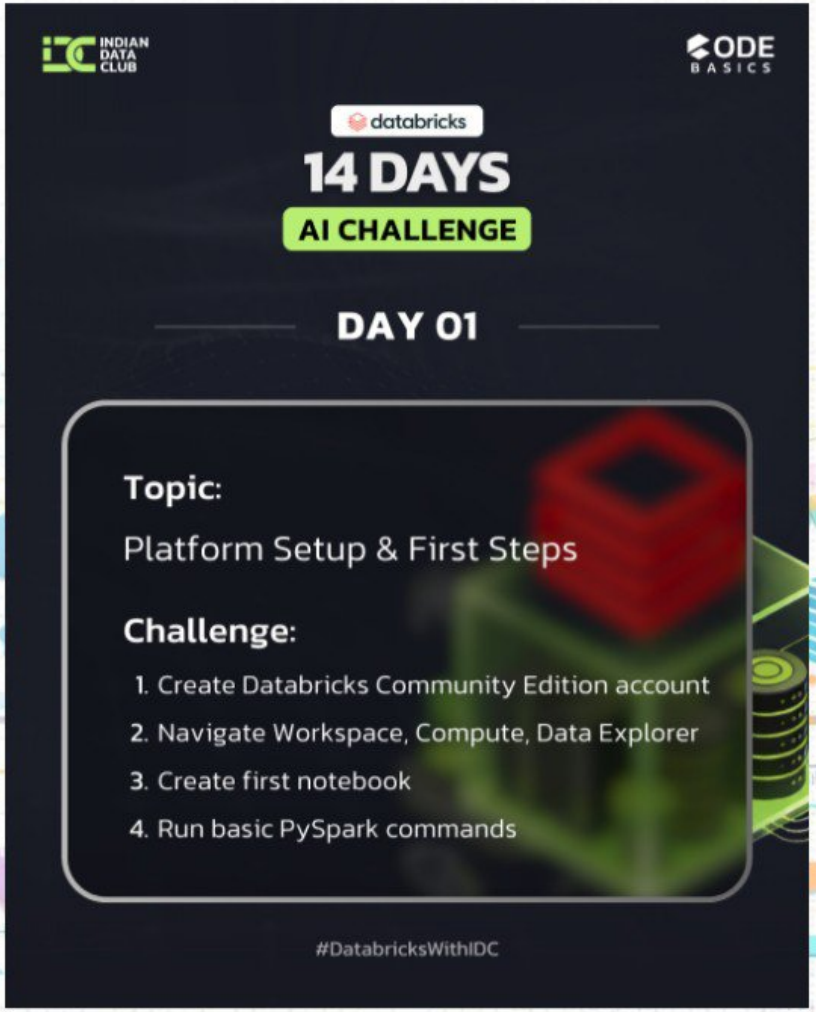
Day 1 | Databricks Basics
Platform Setup & First Steps



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

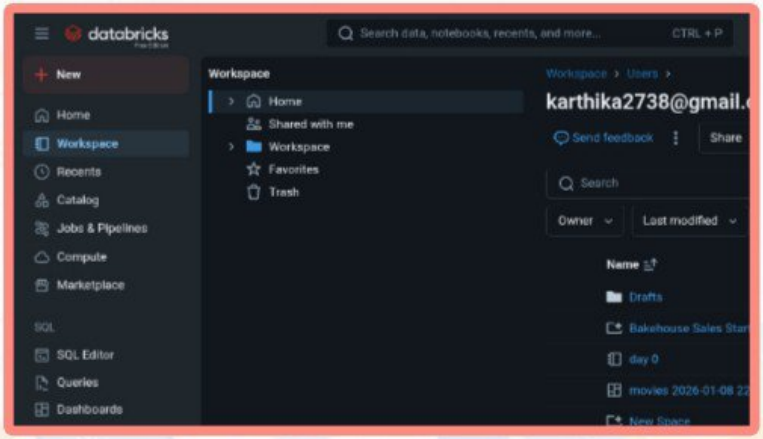
- *Data stored in Amazon S3*
- *Compute powered by Apache Spark*
- *Managed by Databricks Workspace*

👉 You focus on data, not infrastructure

Databricks Workspace

Workspace is where you work daily 👇

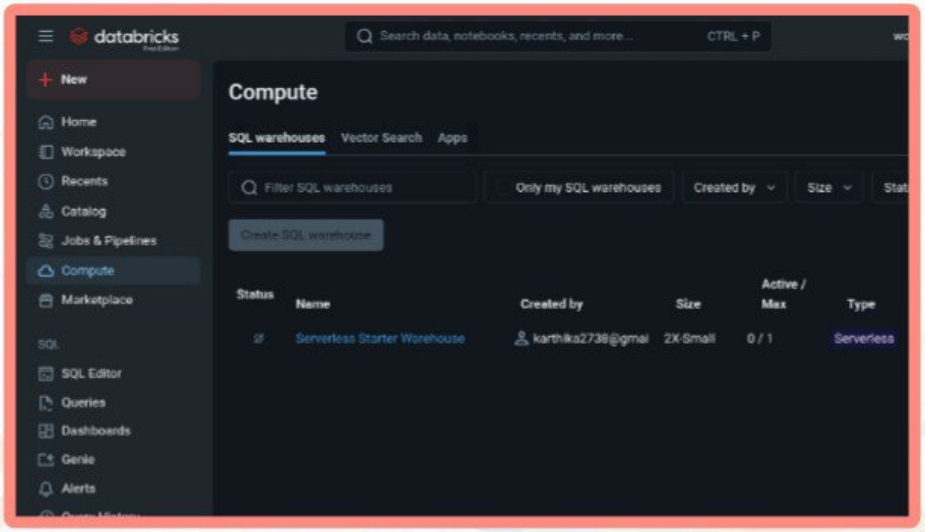
- 📓 *Notebooks – write SQL / PySpark*
- 👥 *Collaboration – share notebooks*
- 📂 *Organized folders*



Compute (Clusters)

Compute = engines that run your code

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



Day 1 Tasks Completed ✅

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

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*