databricks

Manage Data with Delta Lake

Module Agenda

Manage Data with Delta Lake

What is Delta Lake

DE 3.1 – Schemas and Tables

DE 3.2 – Version and Optimize Delta Tables

DE 3.3L – Manipulate Delta Tables Lab

DE 3.4 – Set Up Delta Tables

DE 3.5 – Load Data into Delta Lake

DE 3.6 – Load Data Lab

databricks

What is Delta Lake?

Delta Lake is an open-source project that enables building a data lakehouse on top of existing storage systems

Delta Lake Is Not...

- Proprietary technology

- Storage format

- Storage medium

- Database service or data warehouse

Delta Lake Is...

- [Open source](#)

- Builds upon standard data formats

- Optimized for cloud object storage

- Built for scalable metadata handling

Delta Lake brings ACID to object storage

- **A**tomicity

- **C**onsistency

- **I**solation

- **D**urability

Problems solved by ACID

1. Hard to append data

2. Modification of existing data difficult

3. Jobs failing mid way

4. Real–time operations hard

5. Costly to keep historical data versions

Delta Lake is the default for all tables created in Databricks