Data Engineering: Python in Jupyter vs Python in Databricks

Feature 1 - Execution

Local/Cloud Python

Cloud-based with Spark

Data Engineering: Python in Jupyter vs Python in Databricks

Feature 2 - Integration

Requires separate Spark setup

Built-in Spark & distributed processing

Data Engineering: Python in Jupyter vs Python in Databricks

Feature 3 - Scalability

Limited to local machine

Scales with clusters

Data Engineering: Python in Jupyter vs Python in Databricks

Feature 4 - Collaboration

Single user (or GitHub sharing)

Real-time multi-user editing

Data Engineering: Python in Jupyter vs Python in Databricks

Feature 5 - Magic Commands

`%timeit`, `%matplotlib inline`

`%sql`, `%scala`, `%md`, `%fs`

Data Engineering: Python in Jupyter vs Python in Databricks

Feature 6 - Storage

Local or database connection

DBFS, S3, Azure Blob

Data Engineering: Python in Jupyter vs Python in Databricks

Feature 7 - Job Scheduling

Needs external tools like Airflow

Built-in job scheduling

Feature 8 - Dependency Management

`pip` or `conda`

`%pip install`, Cluster Libraries