

Data Intelligence Platform - Analytics

INTRODUCTION TO DATABRICKS



Kevin Barlow
Data Practitioner

Why do organizations care about analytics?



Supported Languages

Scala

- Based on Java
- Generally used for data engineering



Python

- Used for all use cases



SQL

- Used for data engineering and BI

R

- Used for data science use cases

Databricks Notebooks

Based on Jupyter notebooks, Databricks Notebooks are an optimized and enhanced version.

The screenshot shows a Databricks Notebook interface with the following components:

- Left Sidebar:** A dark sidebar with various icons for navigation and management, including a file icon, a plus sign, a search icon, and a refresh icon.
- Top Bar:** Displays the title "Exploratory Analysis", the notebook name "test3", the language "Python", and status indicators like "Connected".
- Code Editor (Cmd 9):** Contains Python code for data processing and visualization. The code includes filtering rows for the United States, pivoting the DataFrame, replacing missing values, and changing the date column type to Datetime.
- Code Editor (Cmd 10):** Shows a line chart titled "Hospitalizations: 2020-2022" with the y-axis labeled "Patients" and the x-axis labeled "date". The chart displays six data series: Daily ICU occupancy (blue), Daily ICU occupancy per million (orange), Daily hospital occupancy (green), Daily hospital occupancy per million (purple), Weekly new hospital admissions (yellow), and Weekly new hospital admissions per million (cyan). The chart shows two major peaks in hospitalizations around January 2021 and January 2022.
- Right Panel:** A messaging sidebar with two messages:
 - isaac gritz (9/29/2022, 3:49:22 PM): This is really helpful, can we refactor this and share the code with the rest of the team?
 - rafi.kurlansik@datab... (9/29/2022, 3:52:44 PM): Sure thing, I'll create a Python module with these functions and work with Afsana on the unit tests.

SQL Editor

The screenshot shows the Databricks SQL Editor interface. On the left is a dark sidebar with various icons. The main area has a title bar with 'Catalog' and a search bar 'Type to filter'. Below this is a list of databases: 'For you' (hive_metastore, ml, samples, system) and 'All' (samples). The central part of the screen displays a query editor with the following code:

```
1 SELECT
2   o_orderdate AS Date,
3   o_orderpriority AS Priority,
4   sum(o_totalprice) AS `Total Price`
5 FROM
6   `samples`.`tpch`.`orders`
7 WHERE
8   o_orderdate > '1994-01-01'
9   AND o_orderdate < '1994-01-31'
10 GROUP BY
11   1,
12   2
13 ORDER BY
14   1,
15   2
```

Below the code is a results table titled 'Raw results' with columns: Date, Priority, Total Price. The data is as follows:

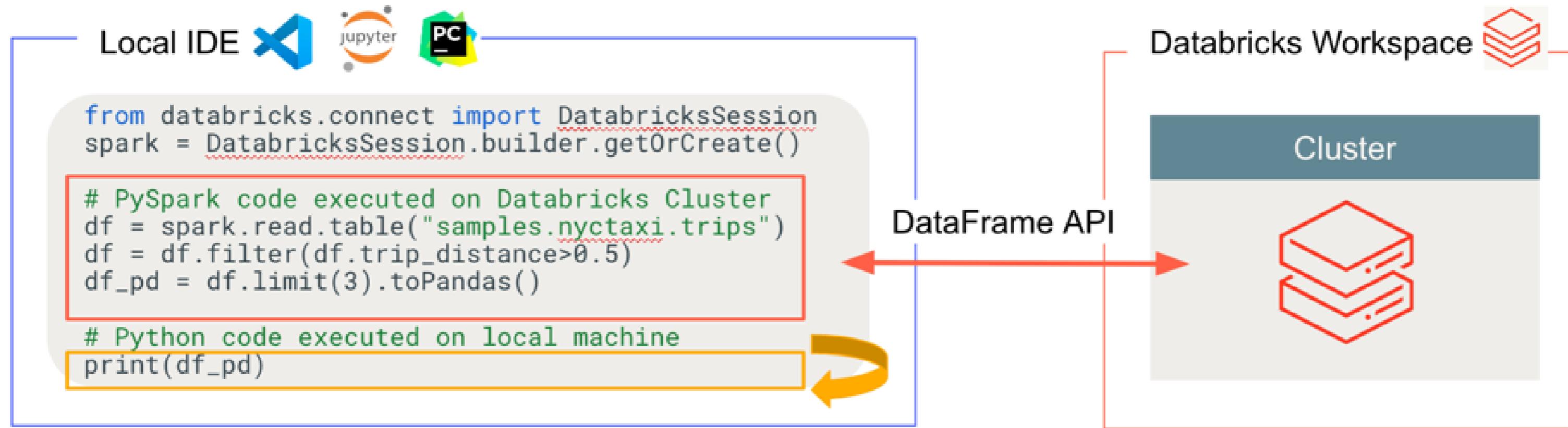
	Date	Priority	Total Price
1	1994-01-02	1-URGENT	96444609.82
2	1994-01-02	2-HIGH	93497904.94
3	1994-01-02	3-MEDIUM	88800085.02
4	1994-01-02	4-NOT SPECIFIED	97955477.98
5	1994-01-02	5-LOW	98015661.37
6	1994-01-03	1-URGENT	92534508.96
7	1994-01-03	2-HIGH	92286715.43
8	1994-01-03	3-MEDIUM	93521575.91
9	1994-01-03	4-NOT SPECIFIED	97569521.46

At the bottom, it says '12 s 751 ms | 145 rows returned' and 'Refreshed a minute ago'.

¹ <https://docs.databricks.com/en/sql/user/sql-editor/index.html>

Databricks Connect

Bring your own IDE! Leverage the power of Databricks Clusters while working in your favorite coding environment.



¹ <https://docs.databricks.com/en/dev-tools/databricks-connect/index.html>

Let's practice!

INTRODUCTION TO DATABRICKS

SQL in the Data Intelligence Platform

INTRODUCTION TO DATABRICKS



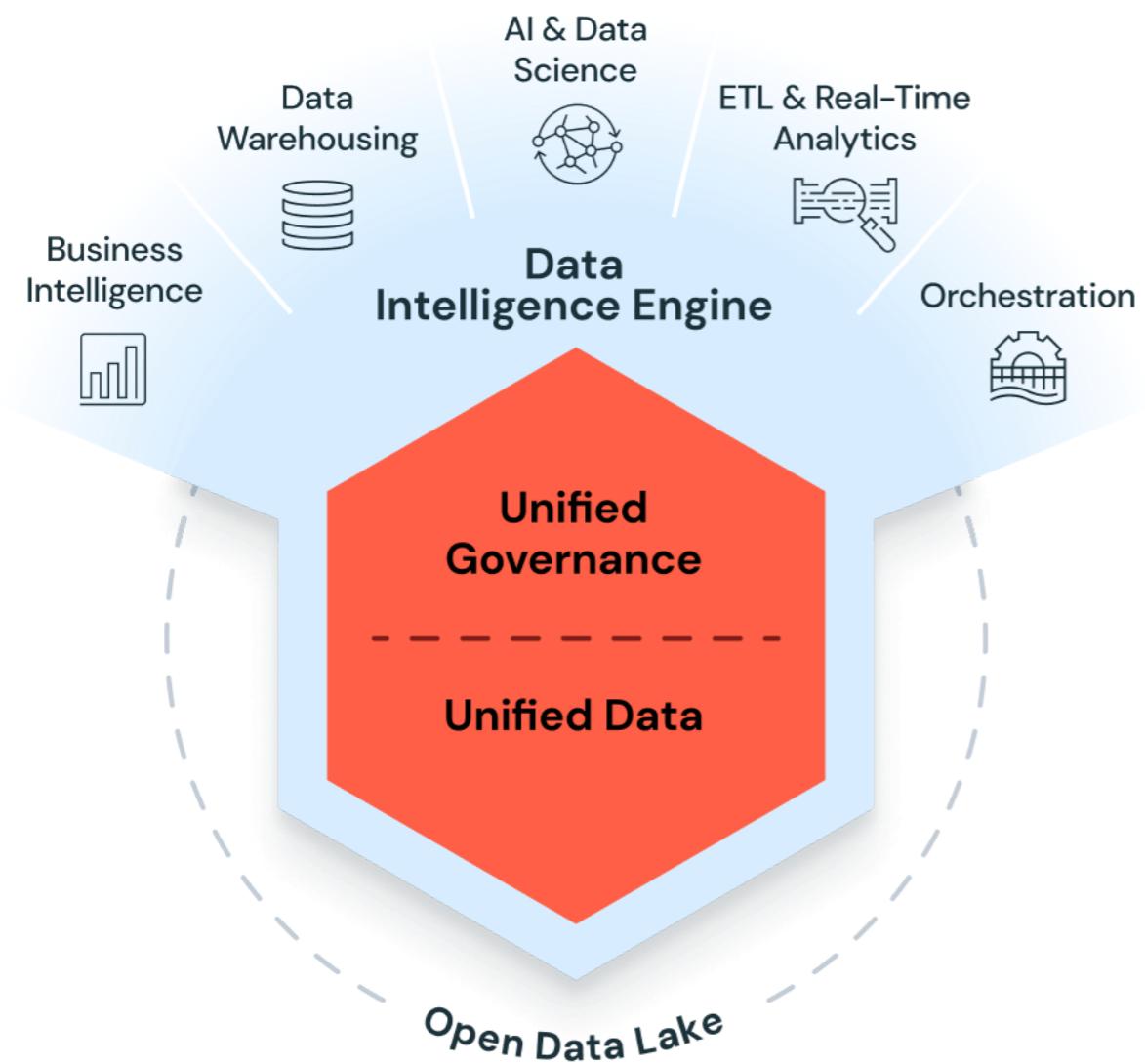
Kevin Barlow
Data Practitioner

Databricks SQL

The Data Intelligence Platform is designed to run your most challenging data warehousing workloads.



Databricks for SQL Users

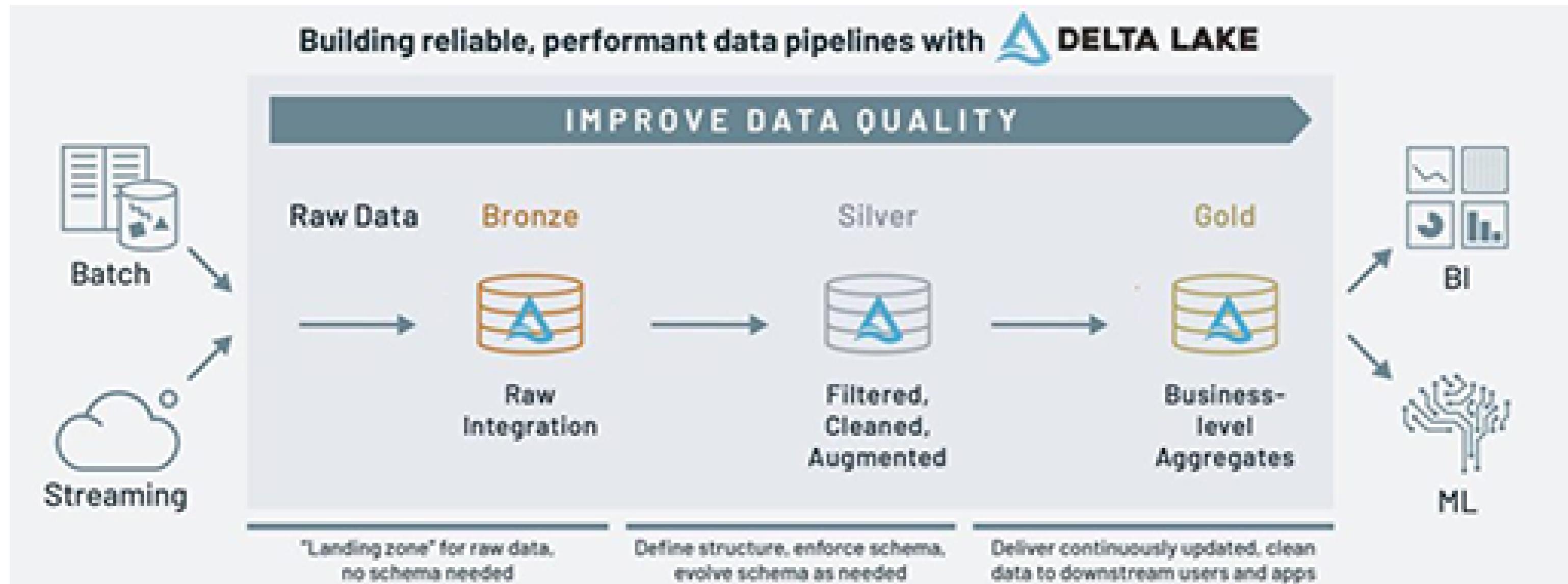


Databricks SQL

- Data Warehousing for the Lakehouse
- Familiar environment for SQL users
- SQL-optimized performance (Photon)
- Connect to your favorite BI tools

Comes built into the platform!

SQL in the Data Intelligence Platform



SQL in the Data Intelligence Platform

Benefits

- Fully integrated
- Scalable, performant compute
- Great UI for analysts

Key Features

- ANSI SQL
- Enhanced Photon engine
- Built-in visualizations

```
SELECT *
FROM json.`/Volumes/
catalog_name/schema_name/
volume_name/path/to/data`  
SELECT *
FROM catalog_name.schema_name.table_name  
CREATE TABLE
catalog_name.schema_name.table_name
AS
SELECT *
FROM ...
```

Let's review!

INTRODUCTION TO DATABRICKS

Congratulations!

INTRODUCTION TO DATABRICKS



Kevin Barlow
Data Practitioner

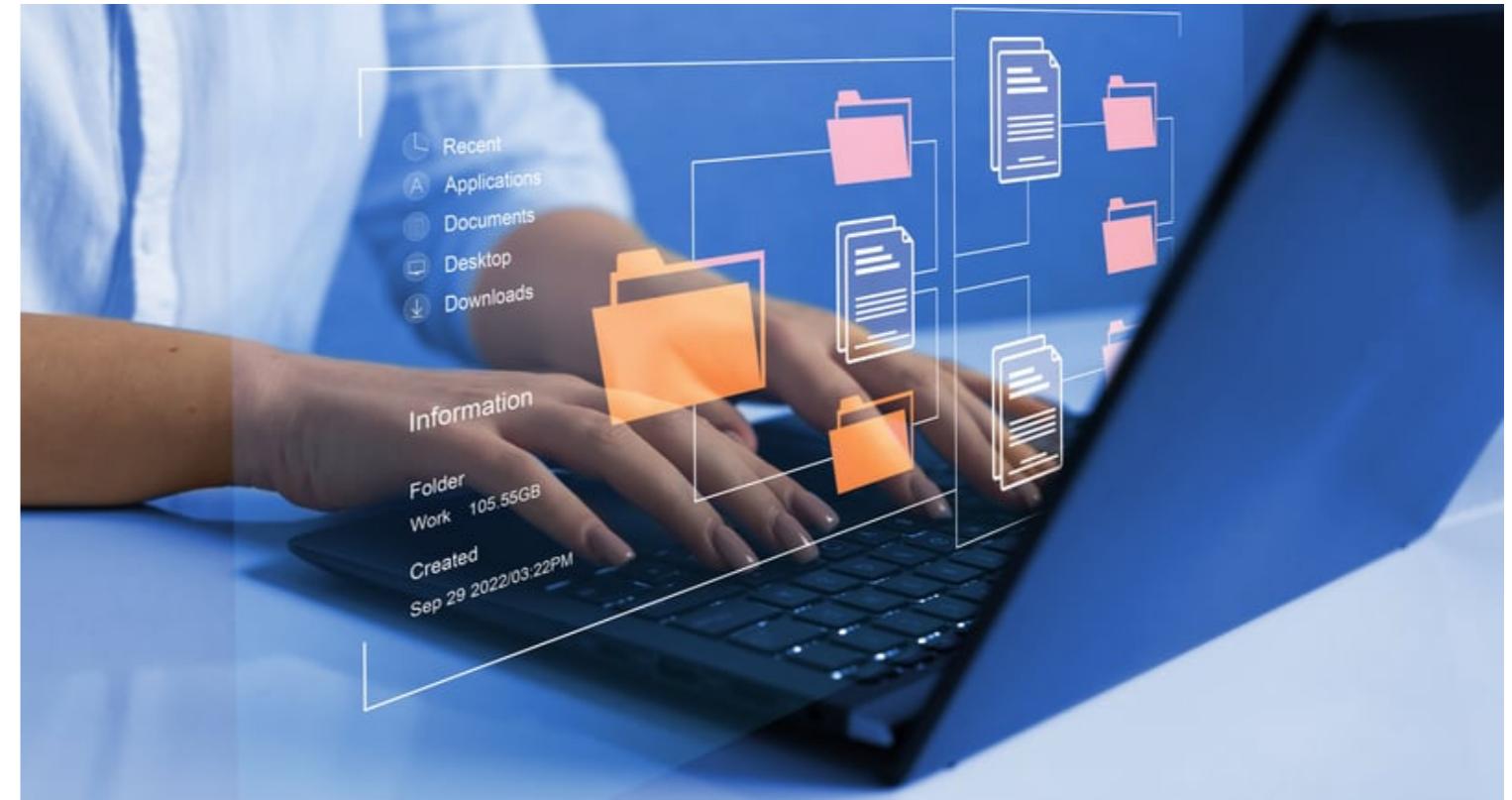
Chapter 1

- Introduction to the Data Intelligence Platform
- Key benefits of building on the platform
- Understanding the underlying architecture



Chapter 2

- Learning about Unity Catalog and Delta
- Managing organizational data
- Utilizing clusters and SQL Warehouses for scalable compute power

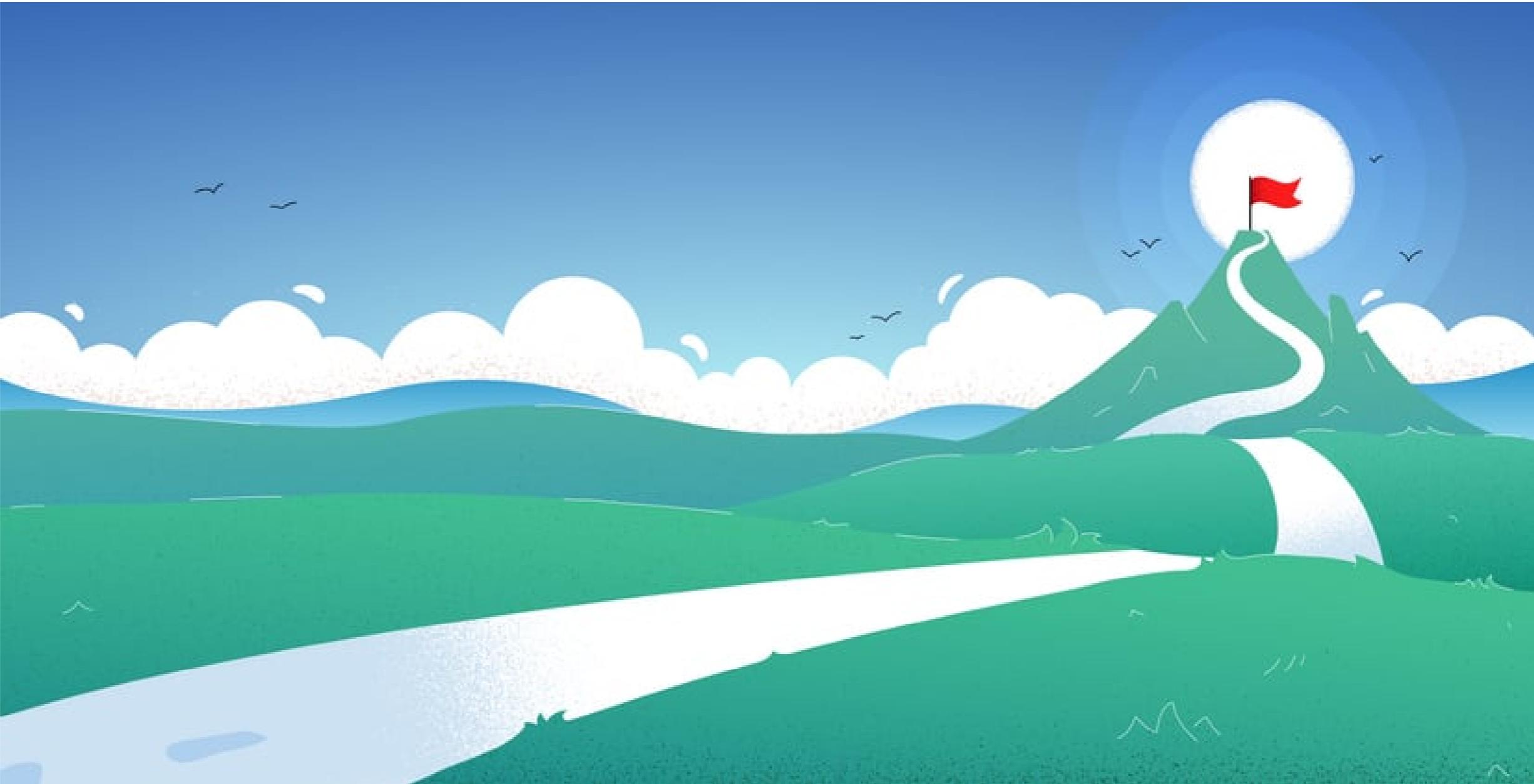


Chapter 3

- Conducting analytical processes in Databricks
- Using Databricks as an enterprise Data Warehouse
- Exploring Databricks SQL for analytics



Just the start



Congratulations!

INTRODUCTION TO DATABRICKS