

The Databricks Data Intelligence Platform

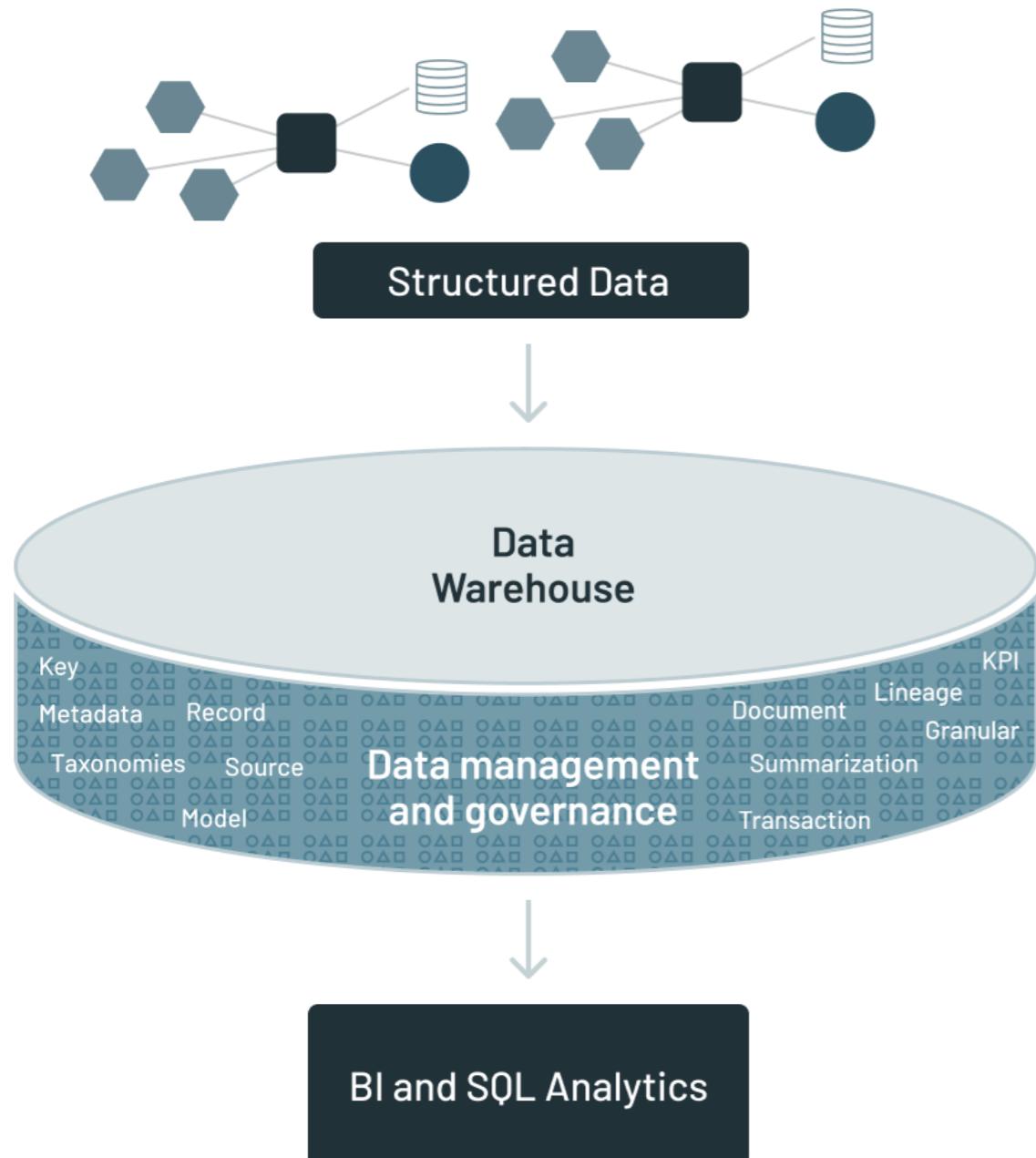
INTRODUCTION TO DATABRICKS



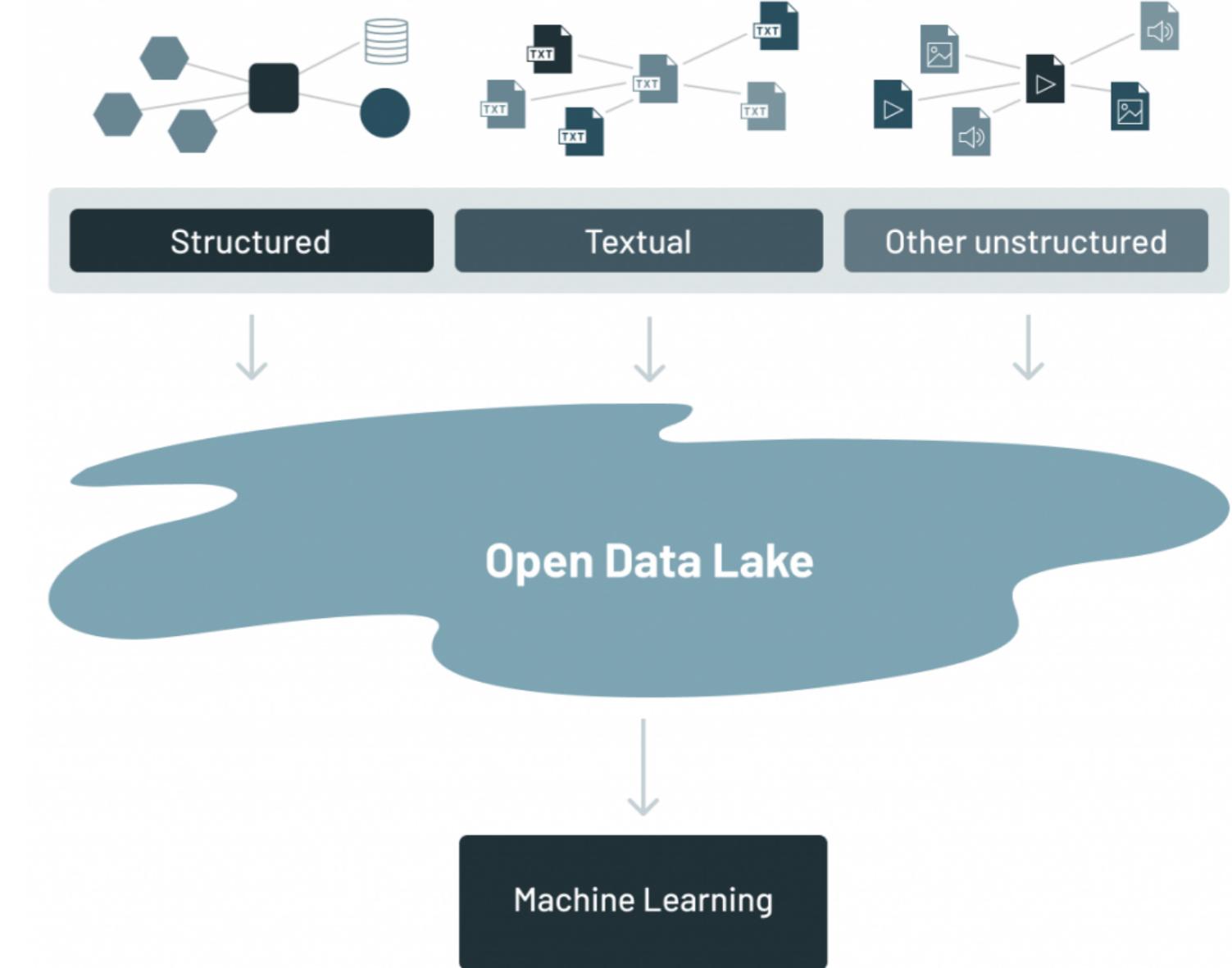
Kevin Barlow

Data Analytics Practitioner

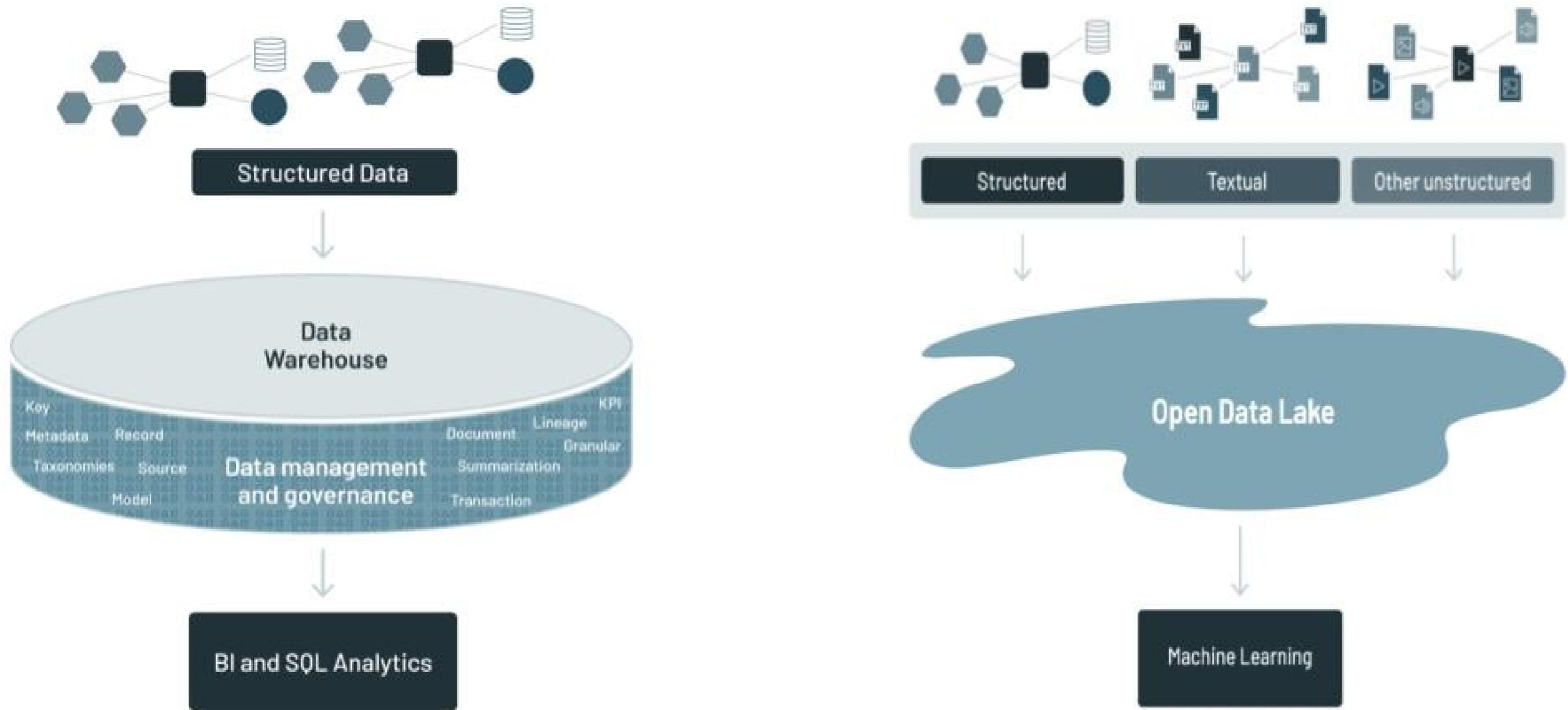
Data Warehouse



Data Lake

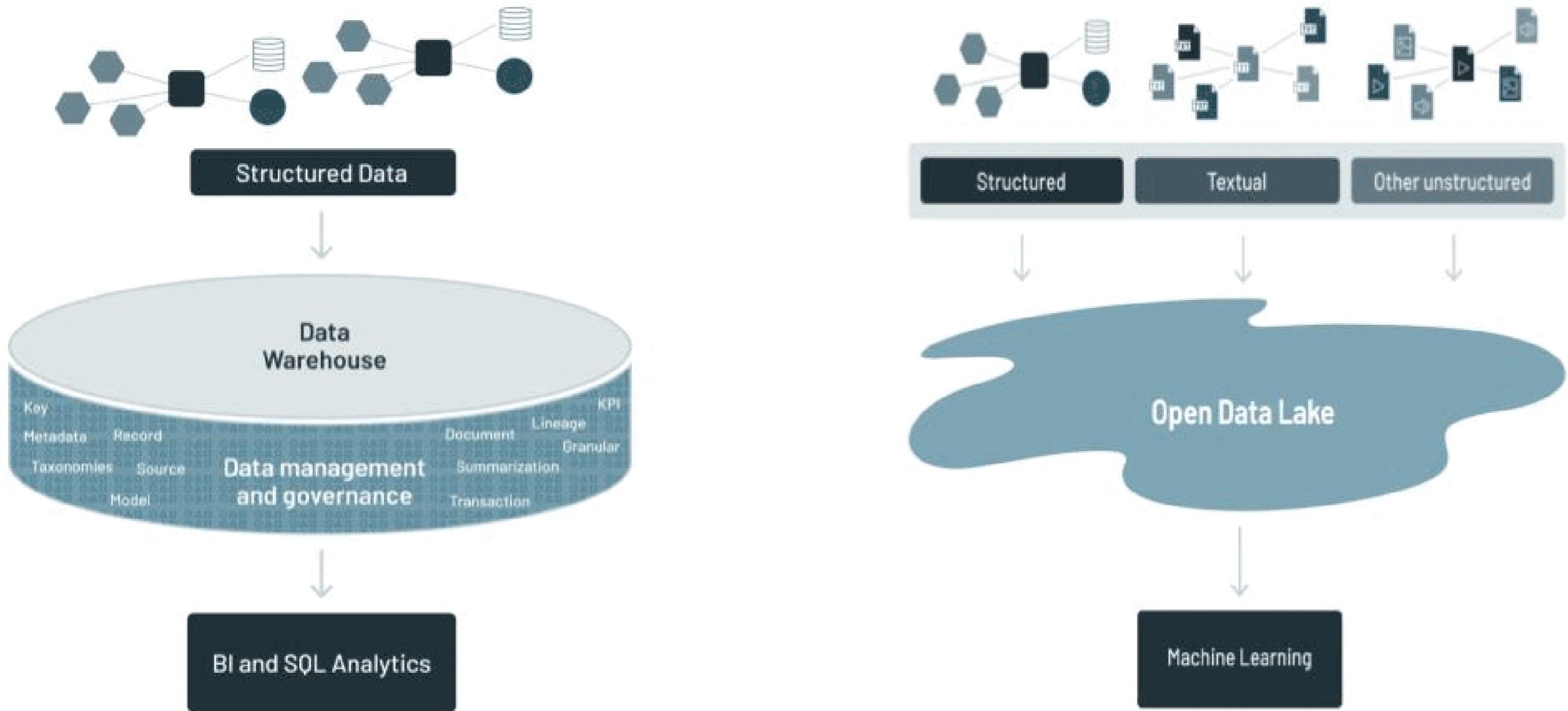


Birth of the Lakehouse



¹ <https://www.databricks.com/blog/2021/05/19/evolution-to-the-data-lakehouse.html>

Birth of the Lakehouse



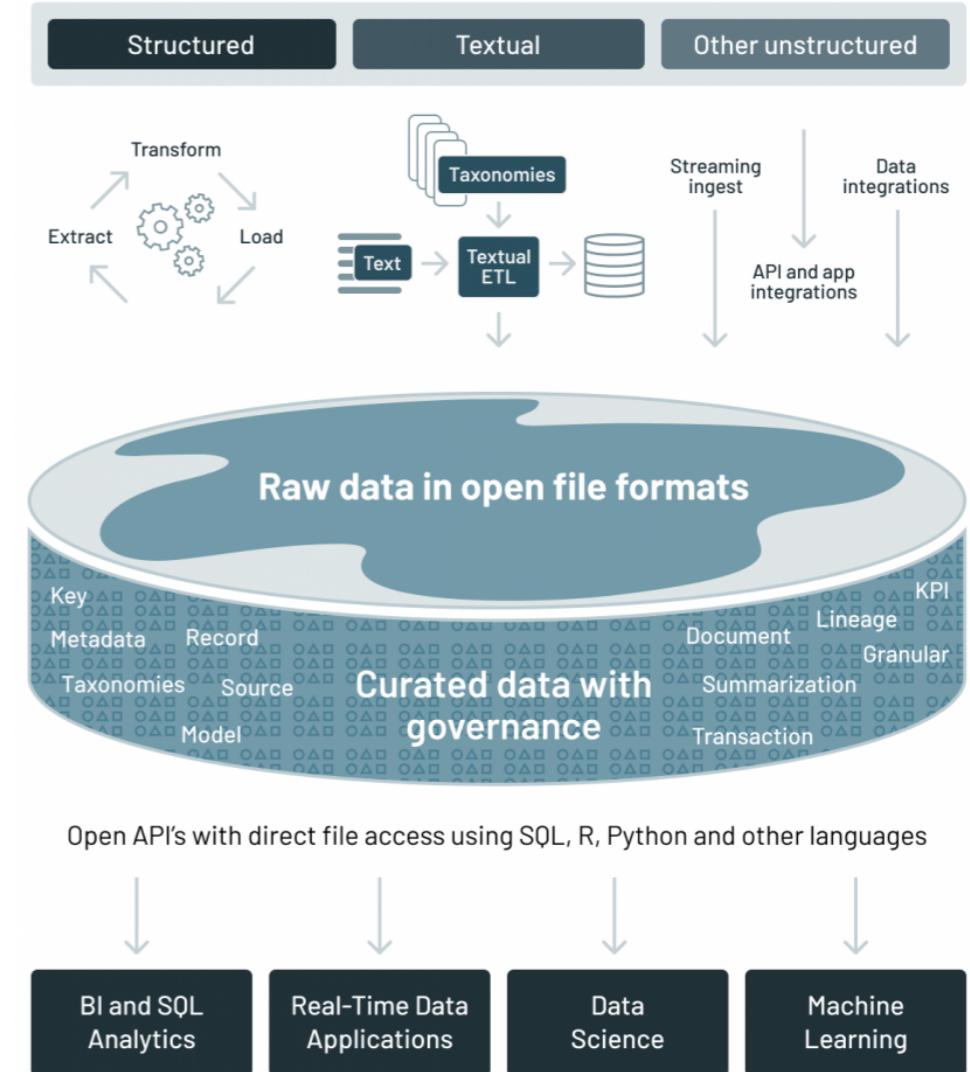
¹ <https://www.databricks.com/blog/2021/05/19/evolution-to-the-data-lakehouse.html>

The Databricks Lakehouse

The Databricks Lakehouse Platform

- Single platform for all data workloads
- Simplified architecture
- Collaborative environment

Data Lakehouse



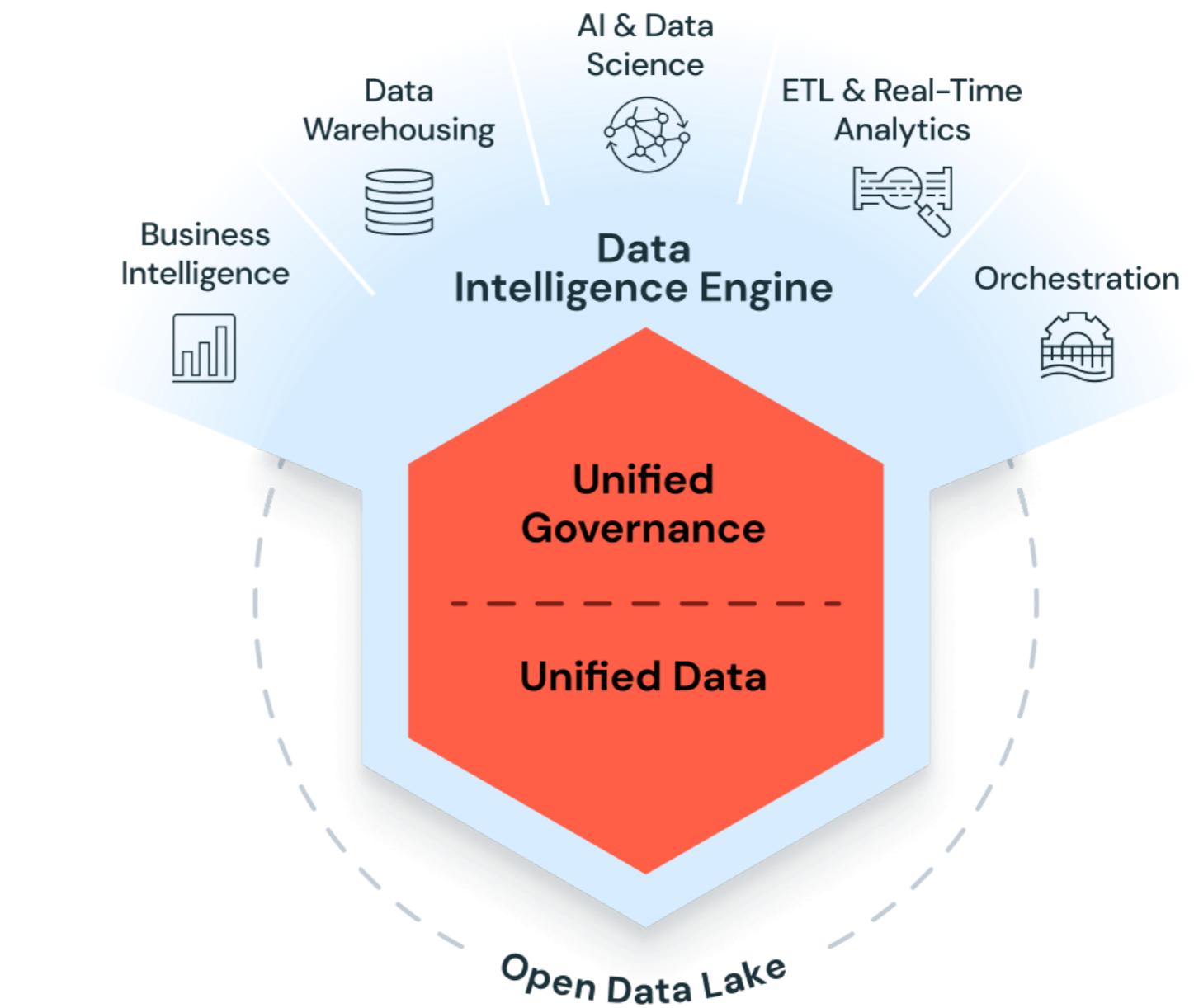
¹ <https://www.databricks.com/blog/2021/05/19/evolution-to-the-data-lakehouse.html>

The Databricks Data Intelligence Platform

Evolution from the Lakehouse

Databricks has expanded the Lakehouse vision to create the first *Data Intelligence Platform*

- Same core architecture of the Lakehouse
- Built-in AI
- First-class support for custom AI applications



¹ <https://www.databricks.com/product/data-intelligence-platform>

Databricks Architecture Benefits

Unification

- Every use case from AI to BI
- Benefits of data warehouse and data lake



Multi-Cloud

- Bring powerful platform to your data
- No lock-in to a specific cloud platform



Databricks Development Benefits

Collaborative

- Every data persona
- Ability to work in same platform in real-time



Open-Source

- Underpinned by *Apache Spark*
- Support for most popular languages (Python, R, Scala, SQL)

A blurred screenshot of a computer monitor displaying a code editor. The code is written in a programming language, likely Python or Scala, and includes functions like `WP_Query` and `WP_Query1`. The code is color-coded for syntax highlighting, with numbers on the left indicating line counts from 365 to 370.

Let's practice!

INTRODUCTION TO DATABRICKS

Setting up a Databricks workspace example

INTRODUCTION TO DATABRICKS



Kevin Barlow
Data Practitioner

Let's practice!

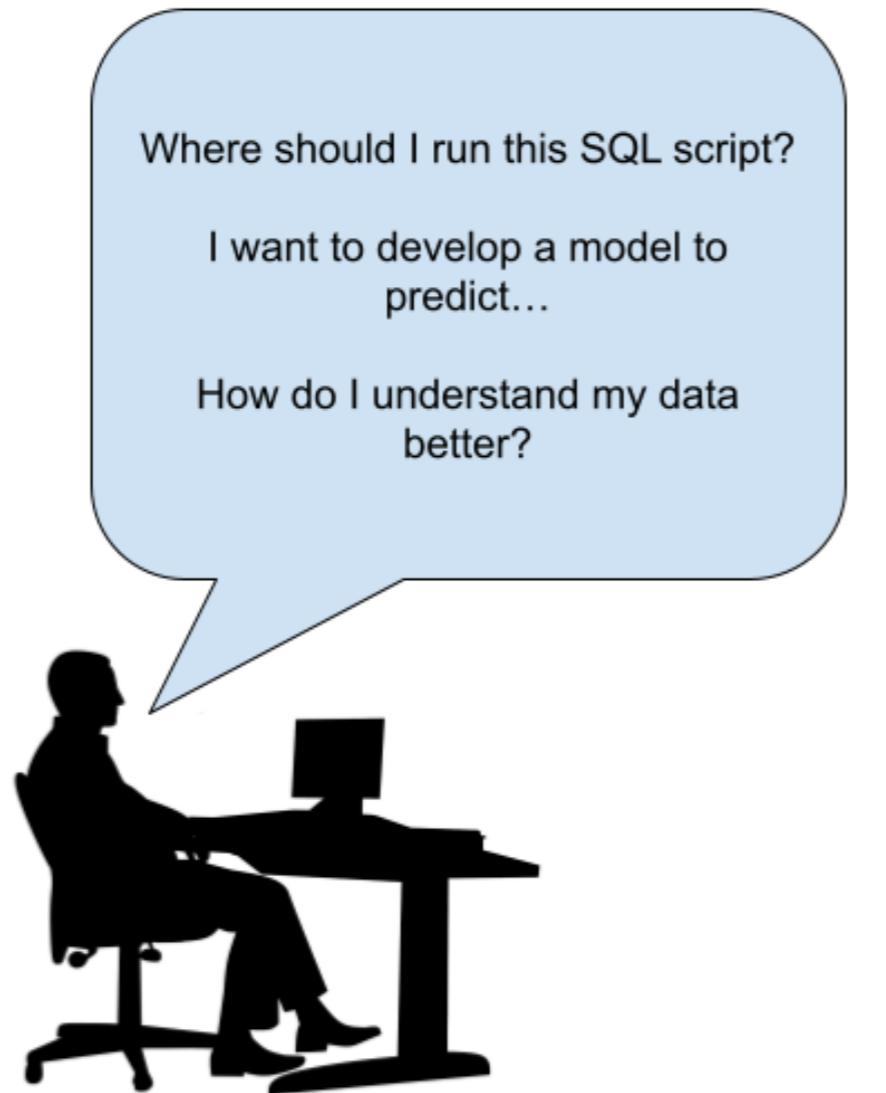
INTRODUCTION TO DATABRICKS

Databricks Architecture

INTRODUCTION TO DATABRICKS



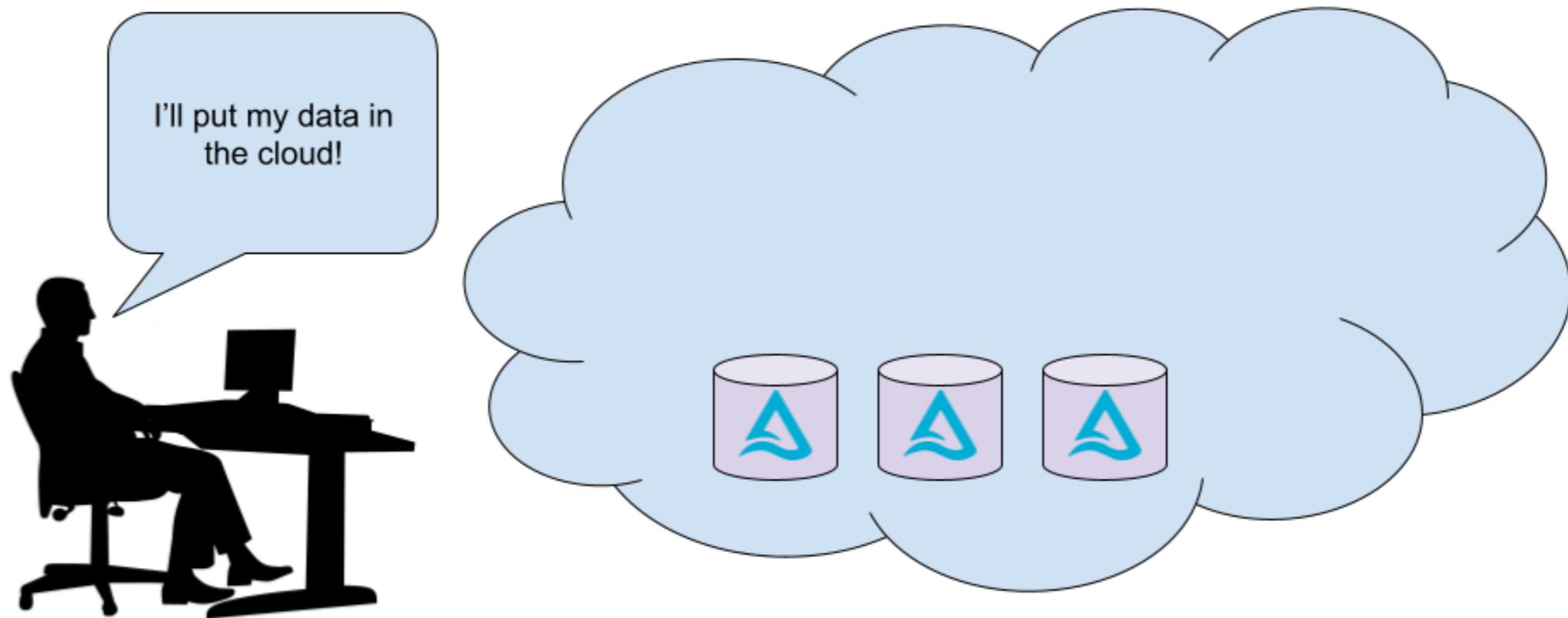
Kevin Barlow
Data Analytics Practitioner

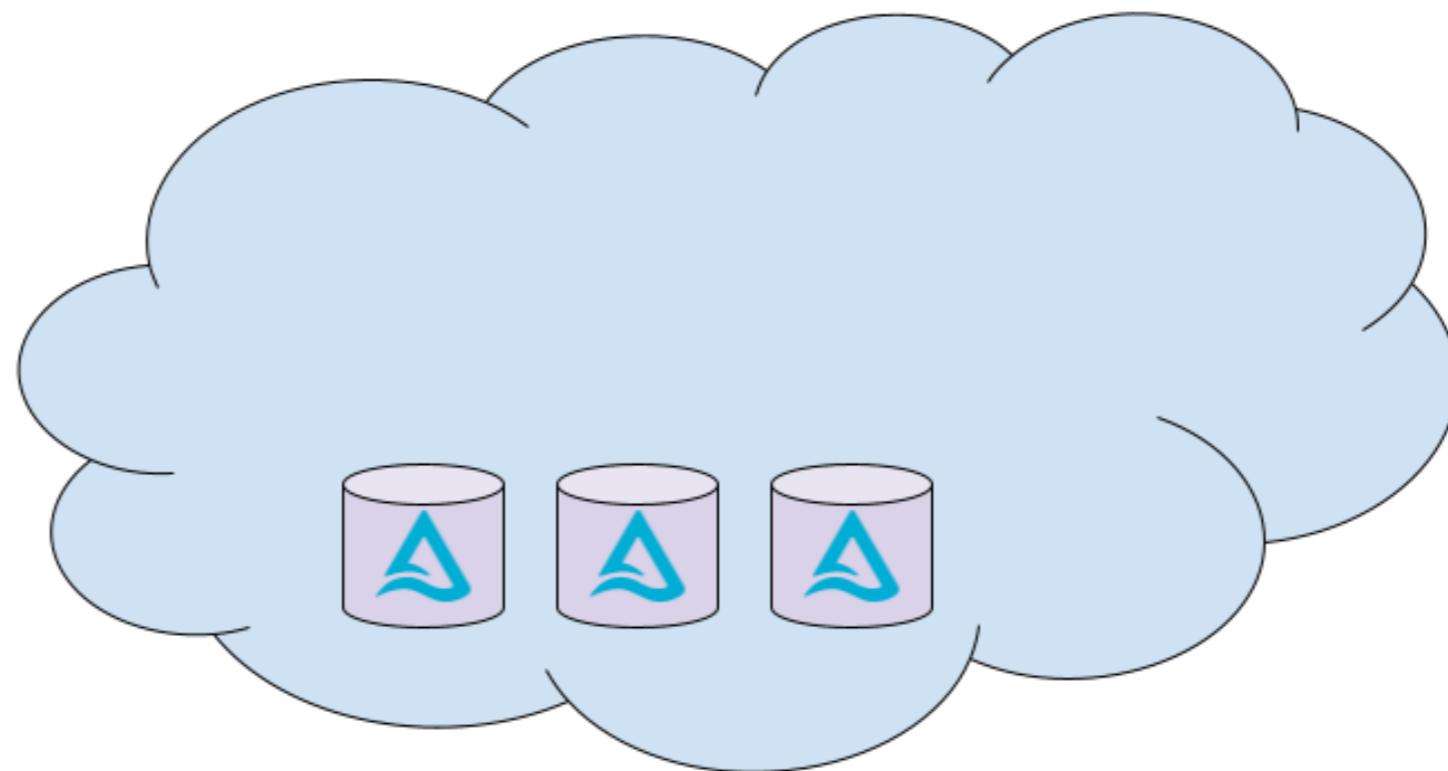
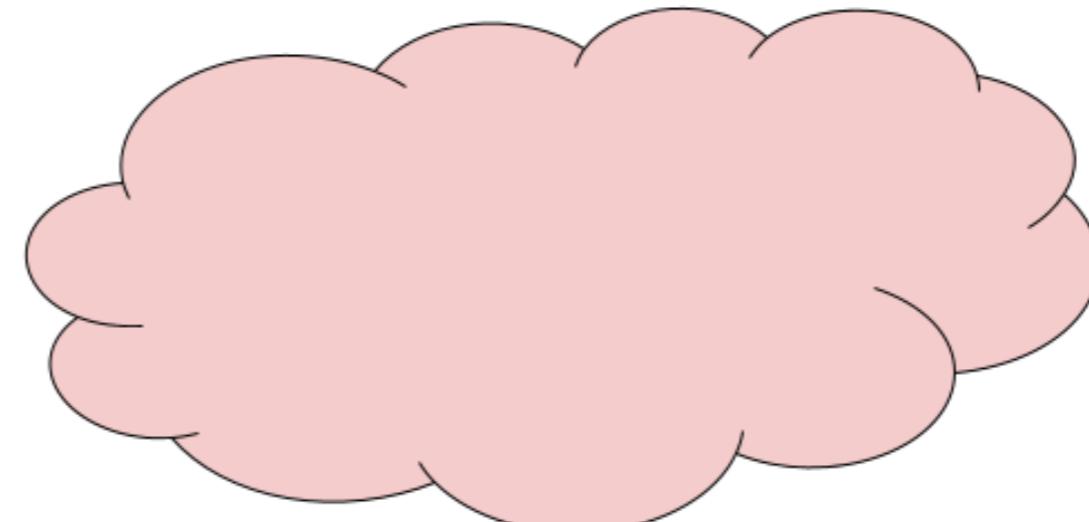


Where should I run this SQL script?

I want to develop a model to
predict...

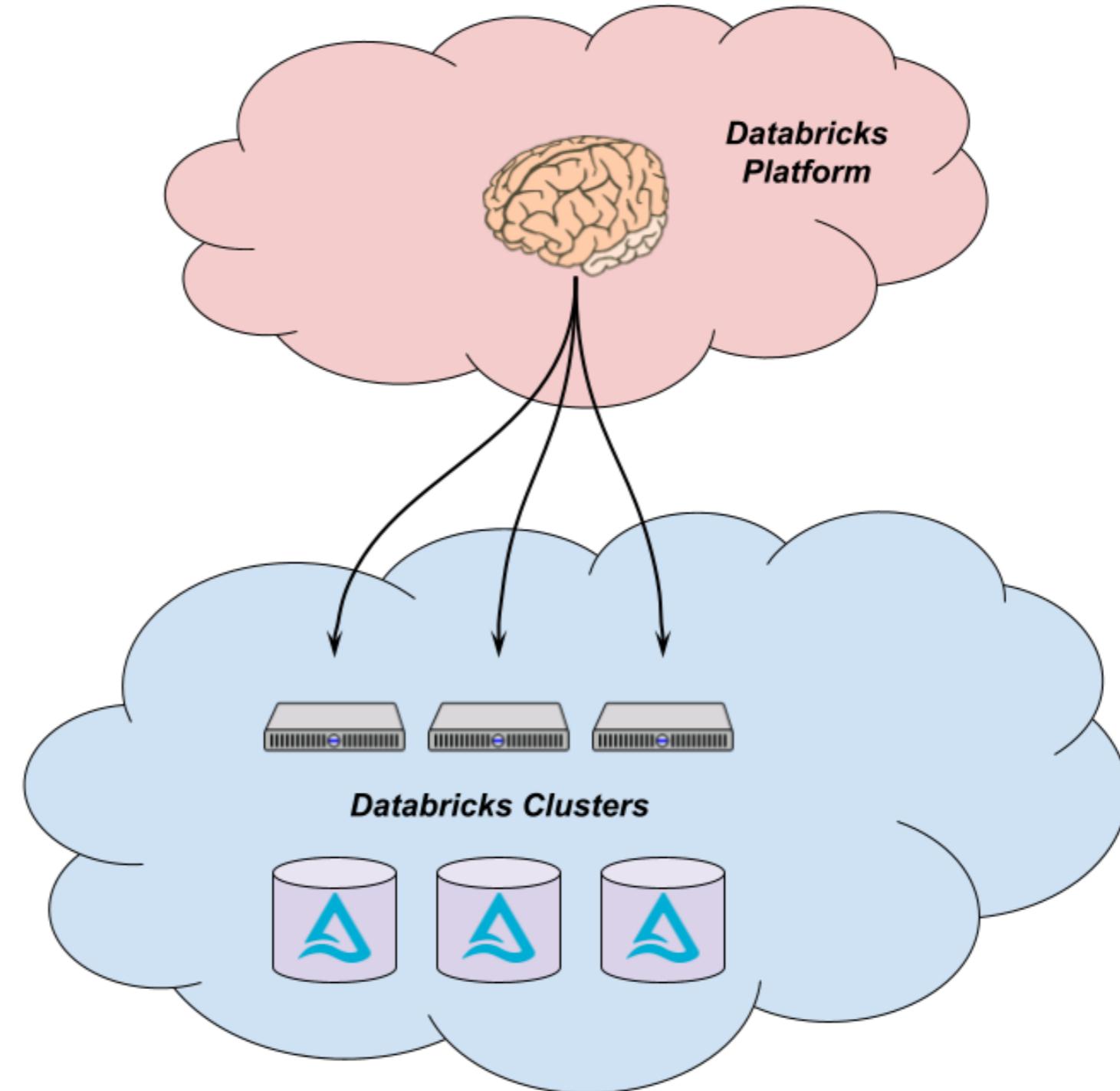
How do I understand my data
better?







databricks



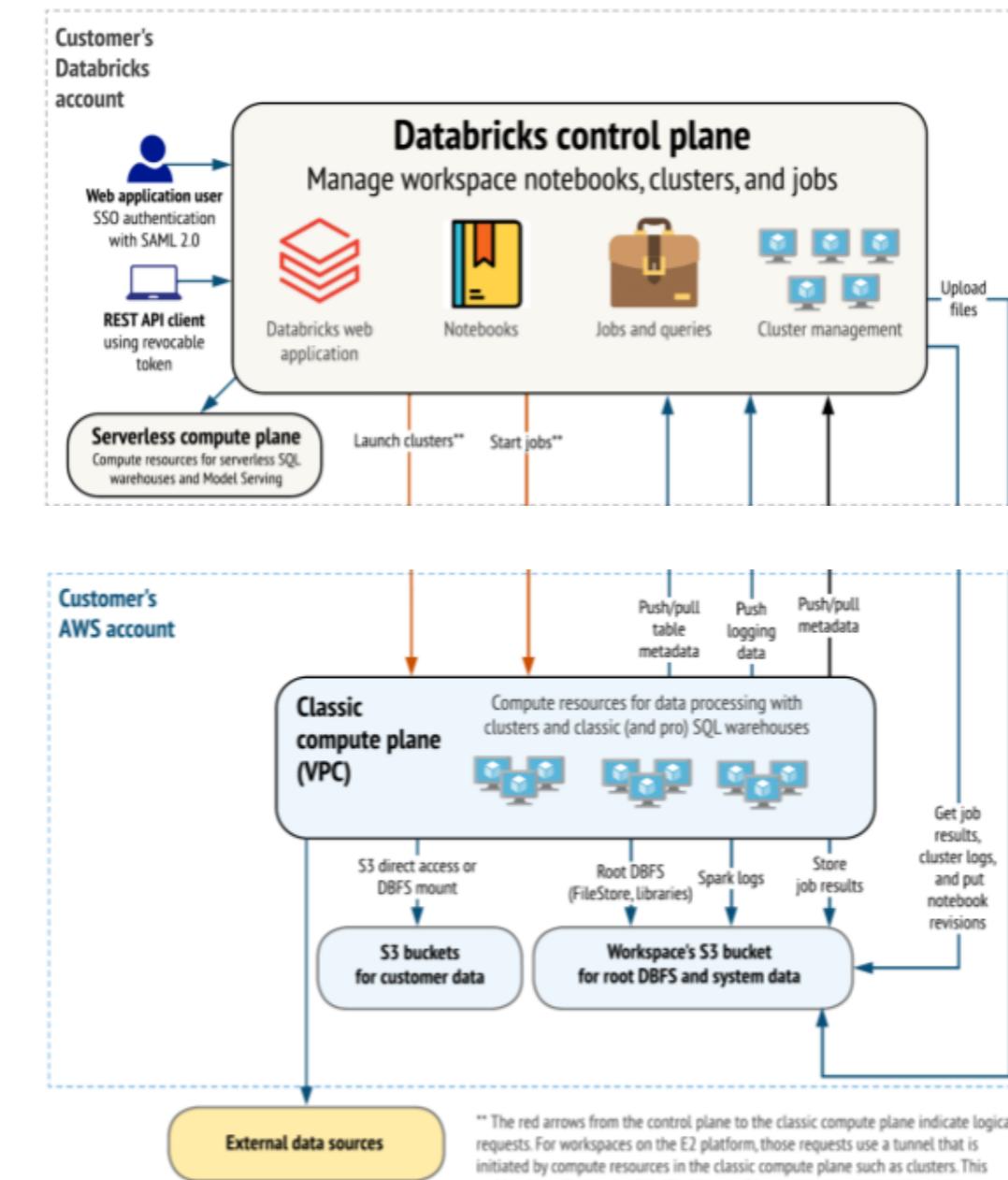
The Databricks Architecture

Control Plane

- Databricks owned environment in your cloud / region
- Hosts the UI, notebooks, and general code
- Orchestrates compute nodes for processing

Compute Plane

- Customer owned environment
- Location for data storage
- Customer networking, applications, etc.



¹ <https://docs.databricks.com/en/getting-started/overview.html#high-level-architecture>

Let's review!

INTRODUCTION TO DATABRICKS

Administering a Databricks workspace

INTRODUCTION TO DATABRICKS



Kevin Barlow
Data Practitioner

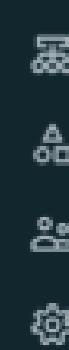
Account Administrators

Key Responsibilities:

- Creating and managing workspaces
- Governing access to workspaces
- Managing the account subscription



Account Console



Account console

Manage your Databricks account at scale



Workspaces

Configure workspace settings. Workspaces contain notebooks, libraries, queries, and workflows



Data

Manage metastores as your top-level container for data, catalogs, schemas (also called databases), views and tables



Users & groups

Manage identities for use with jobs, automated tools and systems



Settings

Configure your Databricks account user provisioning and other settings

Workspaces

The screenshot shows the Databricks Account console interface. On the left is a dark sidebar with four icons: a cluster, a user, a gear, and a gear with a plus sign. The main area has a header "Account console" and a subtitle "Manage your Databricks account at scale". Below are four cards:

- Workspaces** (highlighted with a red border): Configure workspace settings. Workspaces contain notebooks, libraries, queries, and workflows.
- Data**: Manage metastores as your top-level container for data, catalogs, schemas (also called databases), views and tables.
- Users & groups**: Manage identities for use with jobs, automated tools and systems.
- Settings**: Configure your Databricks account user provisioning and other settings.

<https://accounts.cloud.databricks.com/>

Data

The screenshot shows the Databricks Account console interface. On the left is a vertical sidebar with four icons: a network icon (top), a gear icon (second), a user icon (third), and a gear icon (bottom). The main area has a header "Account console" and a subtitle "Manage your Databricks account at scale". Below this are four cards:

- Workspaces**: Manage workspace settings. Workspaces contain notebooks, libraries, queries, and workflows.
- Data**: Manage metastores as your top-level container for data, catalogs, schemas (also called databases), views and tables. This card is highlighted with a red border.
- Users & groups**: Manage identities for use with jobs, automated tools and systems.
- Settings**: Configure your Databricks account user provisioning and other settings.

<https://accounts.cloud.databricks.com/>

Users & Groups

The screenshot shows the Databricks Account console interface. On the left is a vertical sidebar with icons for Workspaces, Data, Settings, and Help. The main area has four cards:

- Workspaces**: Configure workspace settings. Workspaces contain notebooks, libraries, queries, and workflows.
- Data**: Manage metastores as your top-level container for data, catalogs, schemas (also called databases), views and tables.
- Users & groups**: Manage identities for use with jobs, automated tools and systems. This card is highlighted with a red border.
- Settings**: Configure your Databricks account user provisioning and other settings.

<https://accounts.cloud.databricks.com/>

Workspace Administrators

Key Responsibilities:

- Managing identities in your workspace
- Creating and managing compute resources

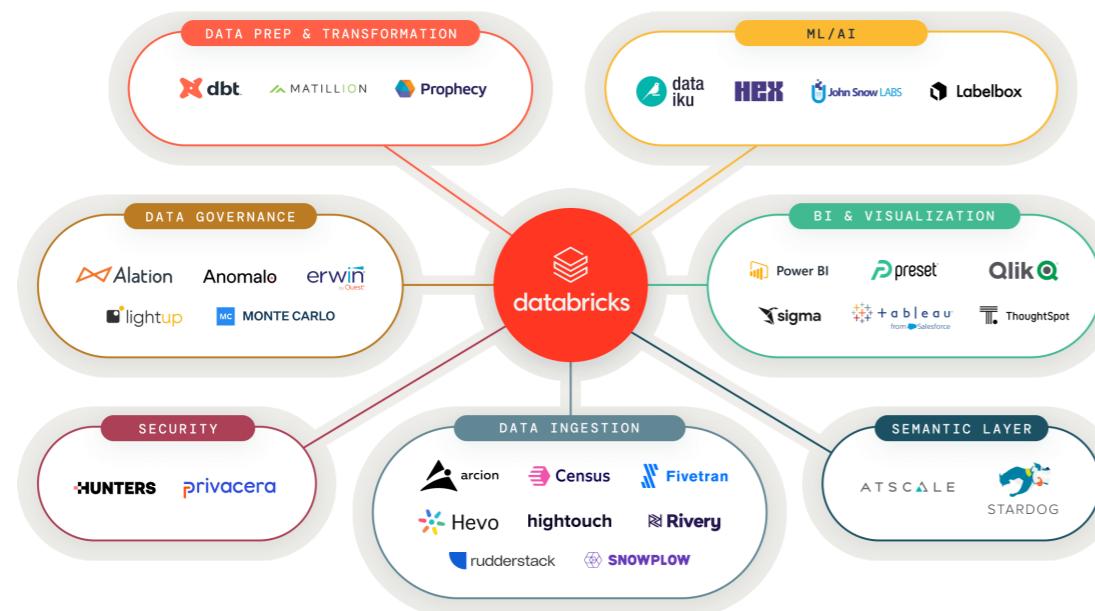
Admin Settings

[Users](#) [Service principals](#) [Groups](#) [Global init scripts](#) [Workspace settings](#) [SQL settings](#) [Notification destinations](#) [SQL warehouse settings](#)

Other Administrative Activities

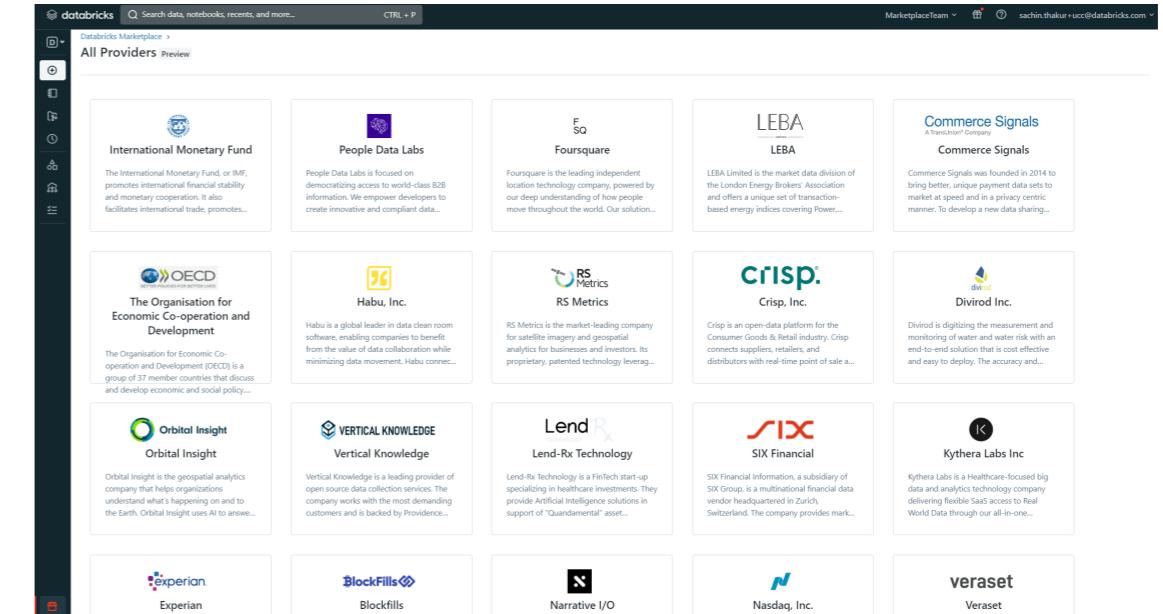
Partner Connect

- UI-based connection to partner technologies
- BI connections, ingestion tools, etc.



Databricks Marketplace

- Discover and access third-party datasets
- Integrate directly into your data catalogs



¹ <https://www.databricks.com/partnerconnect>

Let's review!

INTRODUCTION TO DATABRICKS