# Implementing an Azure Databricks Environment with Microsoft Azure

## IMPLEMENTING AN AZURE DATABRICKS ENVIRONMENT

**Michael Bender**
AUTHOR EVANGELIST, PLURALSIGHT

@MichaelBender

# Overview

Introduction to Azure Databricks

Fundamental Components of Azure Databricks

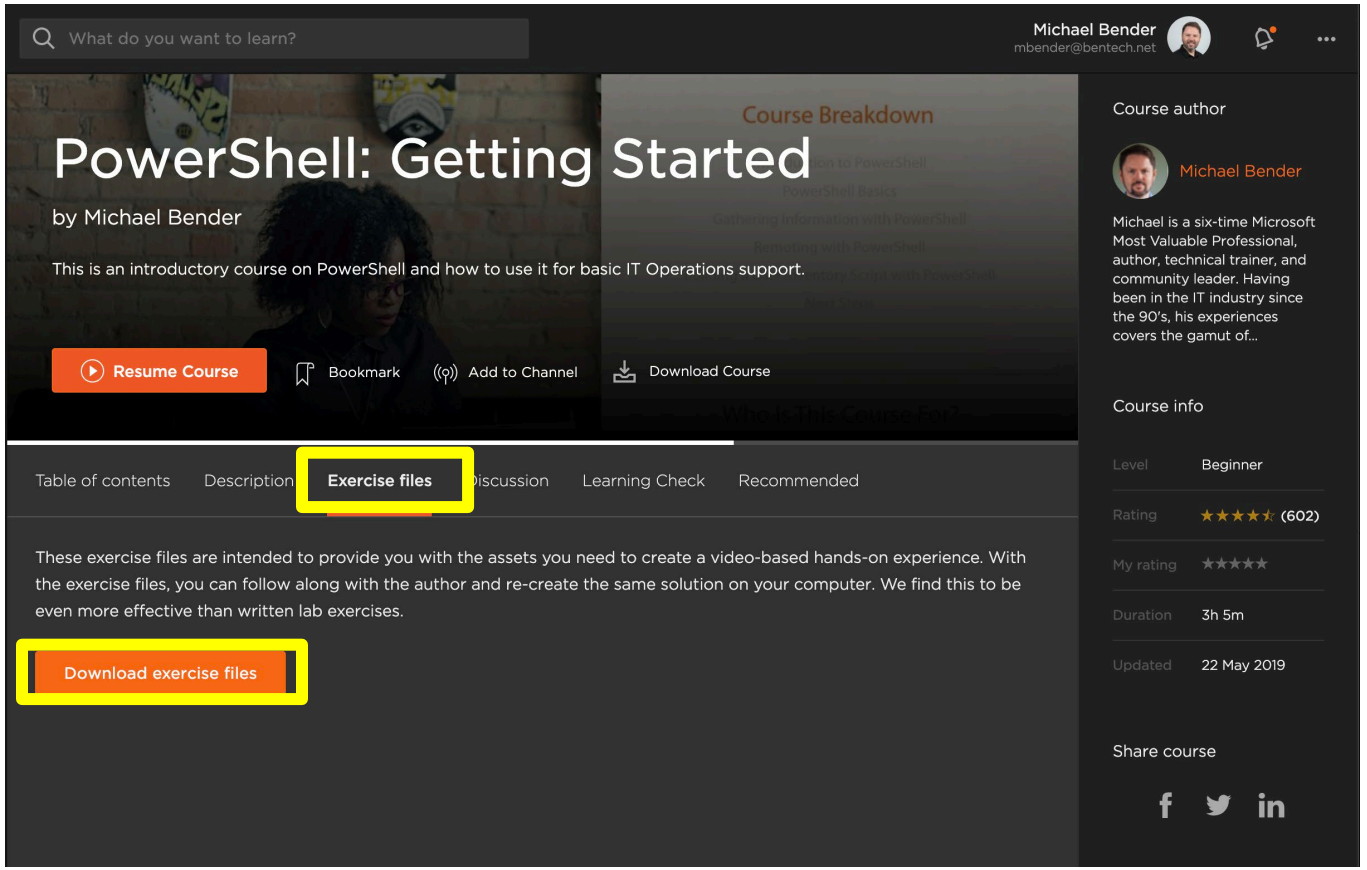
Working with workspaces, notebooks, tables and jobs

# Exercise Files

Slides

Code

Links to Resources



# What is Azure Databricks

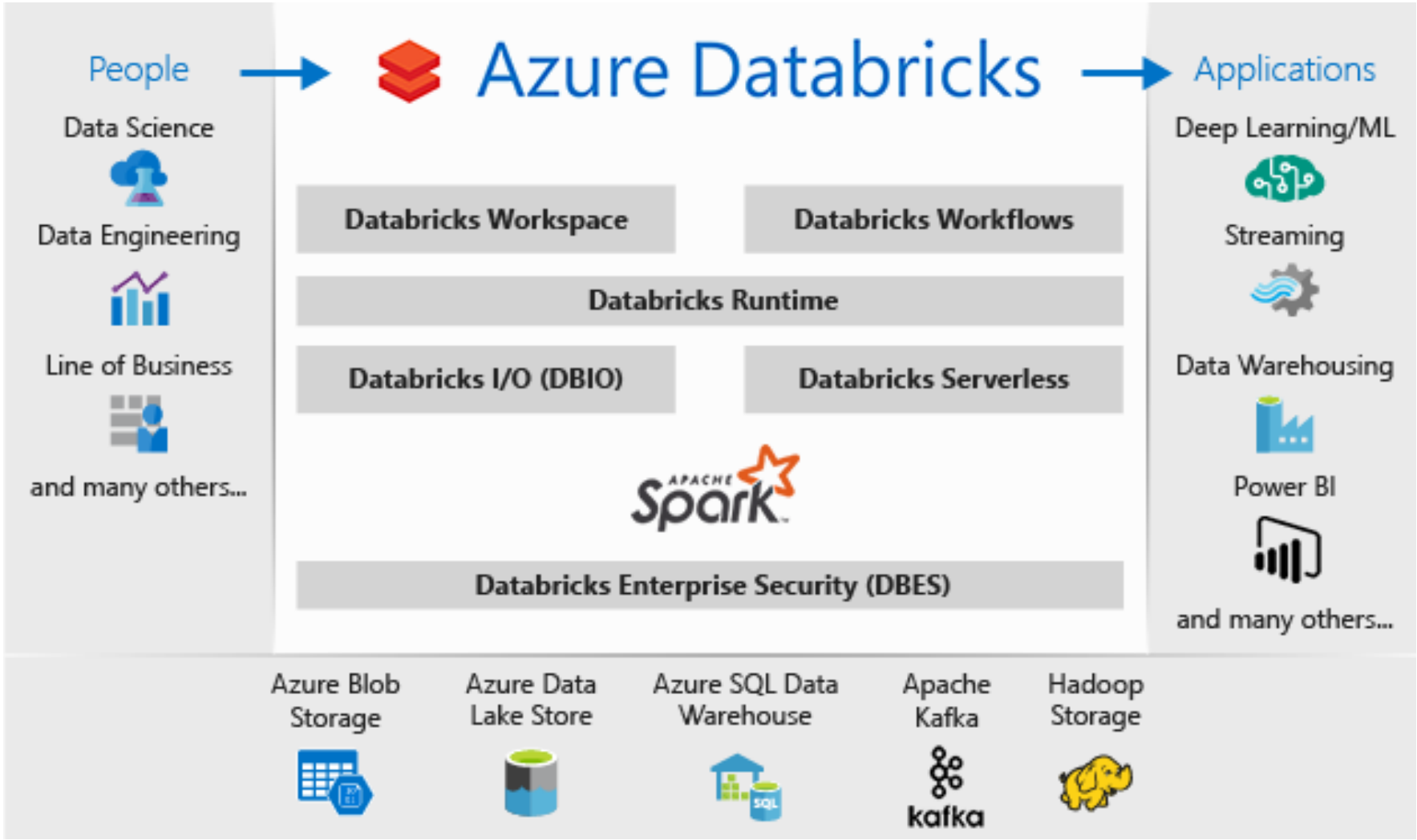
Scalable analytics platform in Azure

Based on Apache Spark

Workflows and workspaces for data users

Native integration with other Azure services

# What is Azure Databricks



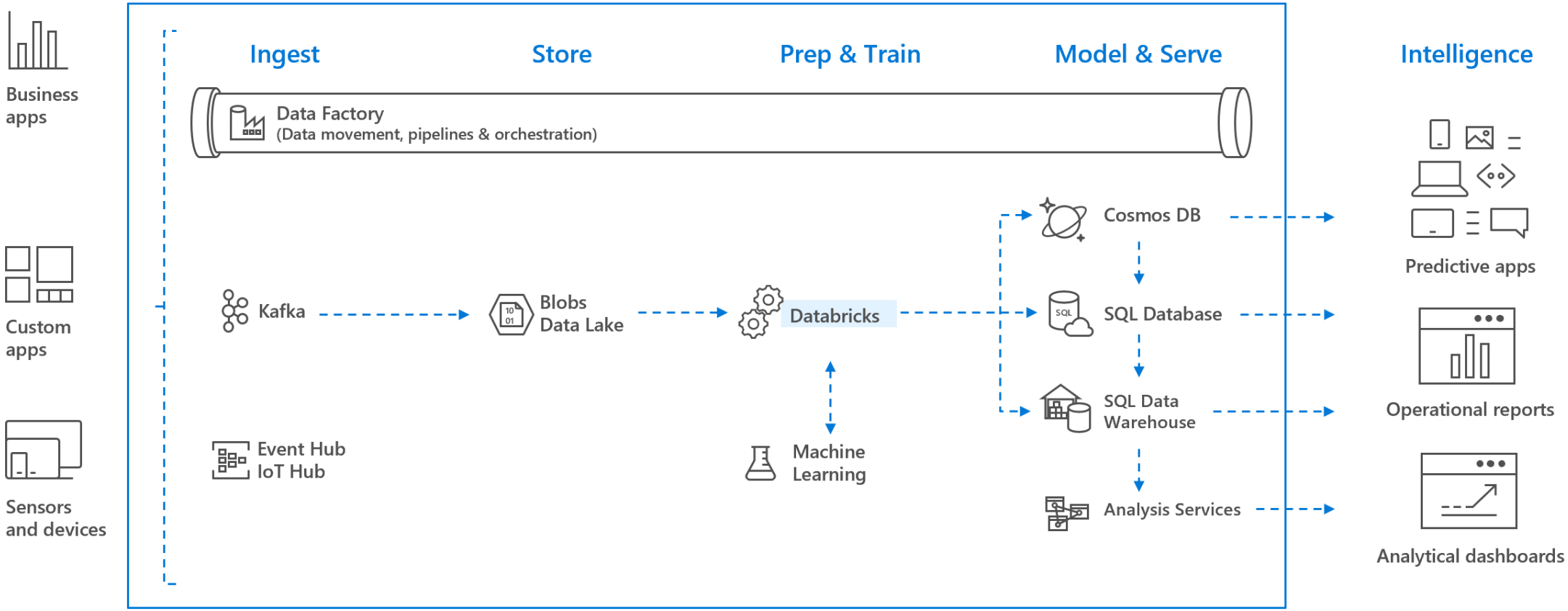
# Users of Azure Databricks

**Data Engineers**

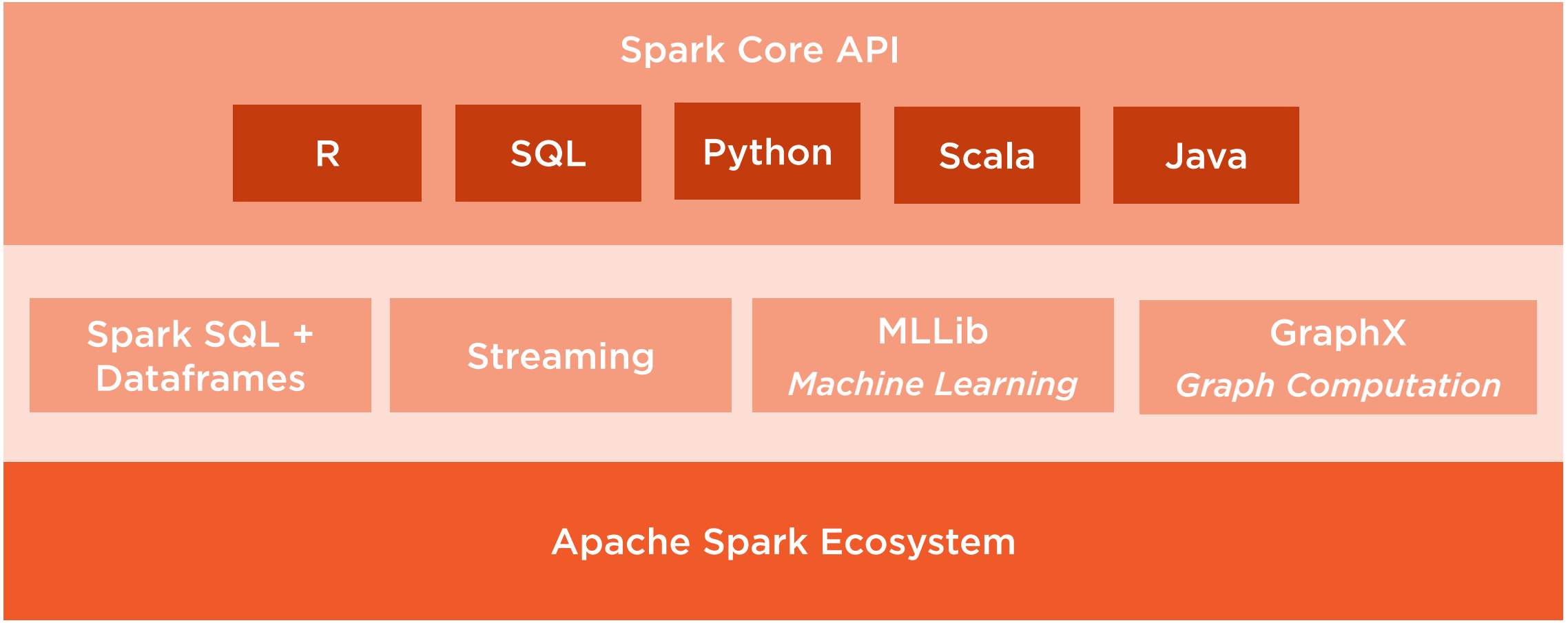
**Data Scientists**

**Business Users**

# Data Pipeline and Azure Databricks



# Apache Spark-based analytics platform



# Typical use cases for Azure Databricks

**Interactive analytics**

**Data integration**

**Machine learning**

**Stream processing**

# Fundamental Azure Databricks Components

Workspaces

Clusters

Notebooks

Tables

Jobs

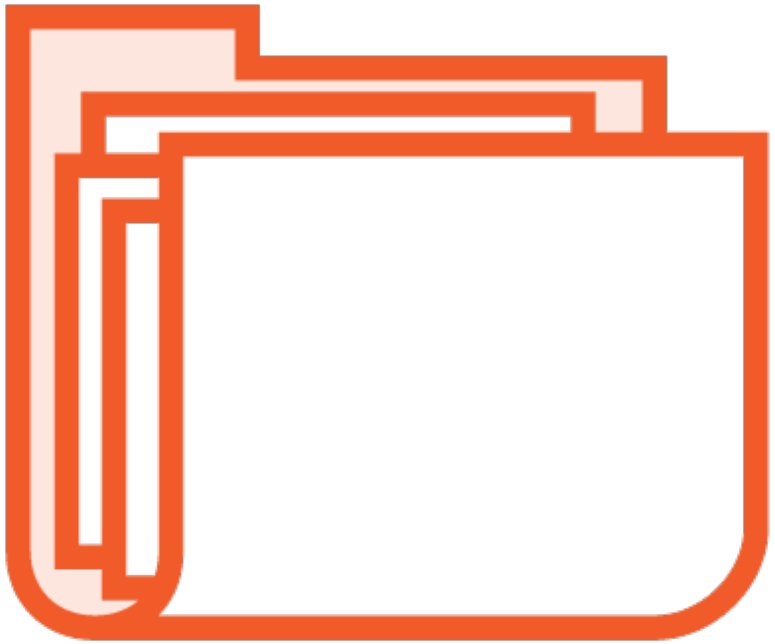
# Workspaces

Organizes all of your Azure Databricks assets

Control Access via Workspace Access Control

Create and manage using UI, CLI or Workspace API

# Folders



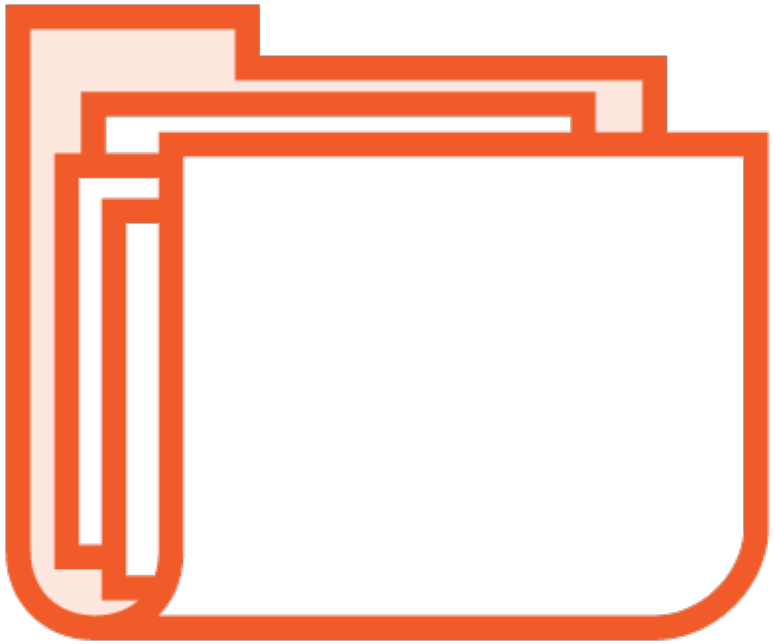
Storage for all workspaces assets

Check the icon for object type contained

Use access control to manage collaboration

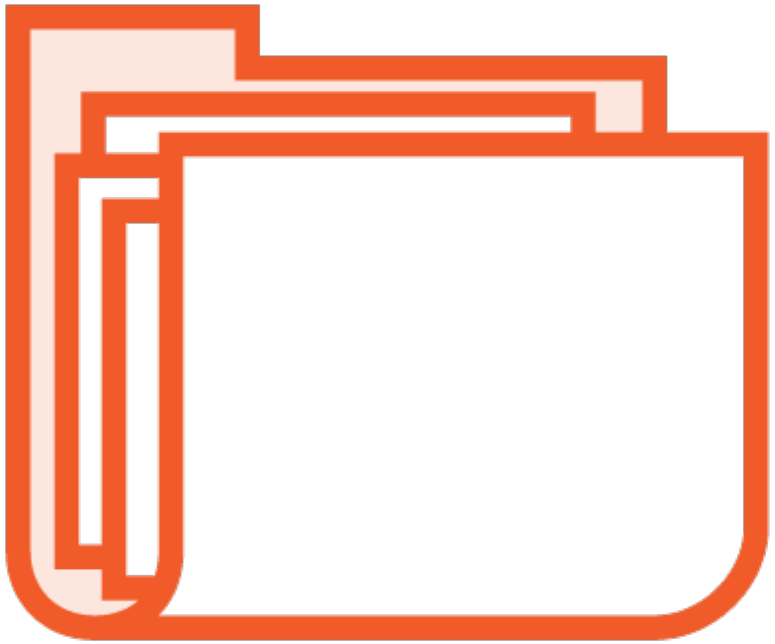
Special folders

# Folders



Storage for all workspaces assets

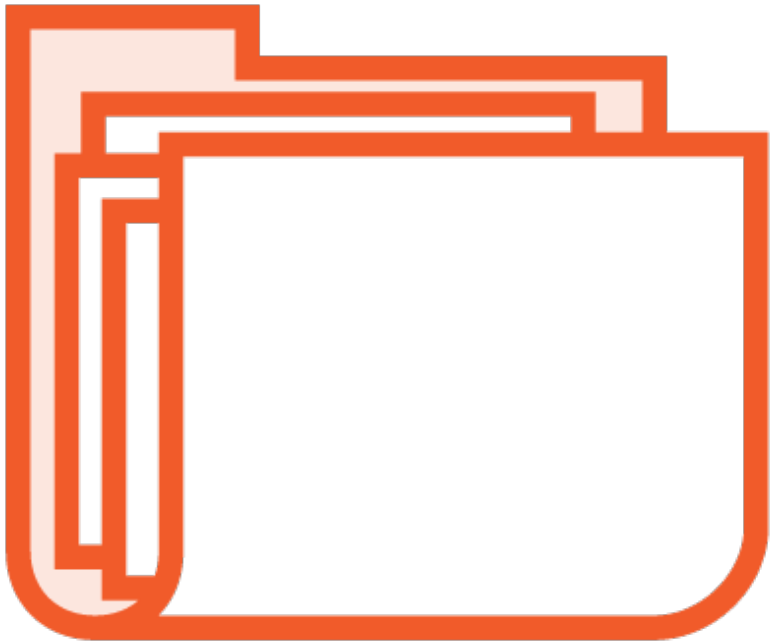
# Folders



Storage for all workspaces assets

Check the icon for object type contained

# Folders

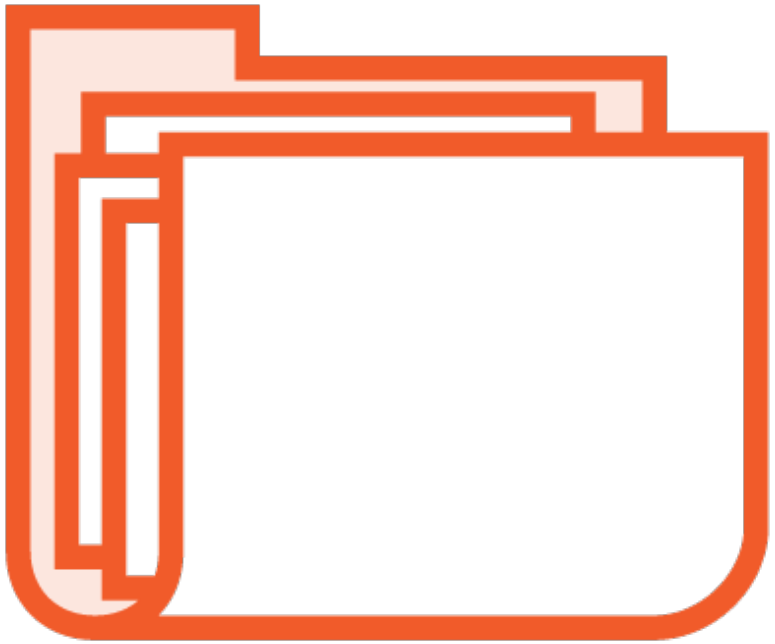


Storage for all workspaces assets

Check the icon for object type contained

Use access control to manage collaboration

# Folders



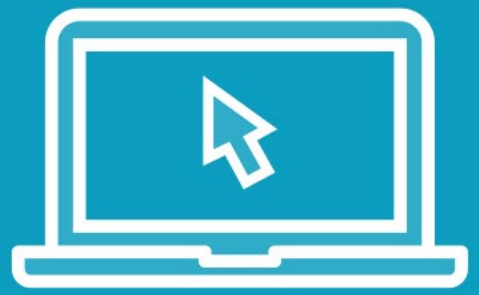
Storage for all workspaces assets

Check the icon for object type contained

Use access control to manage collaboration

Special folders

# Demo

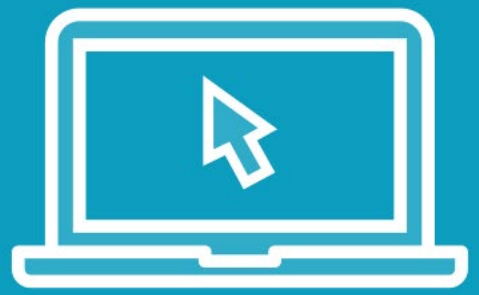


Creating an Azure Databricks workspace

Exploring the Azure Databricks portal

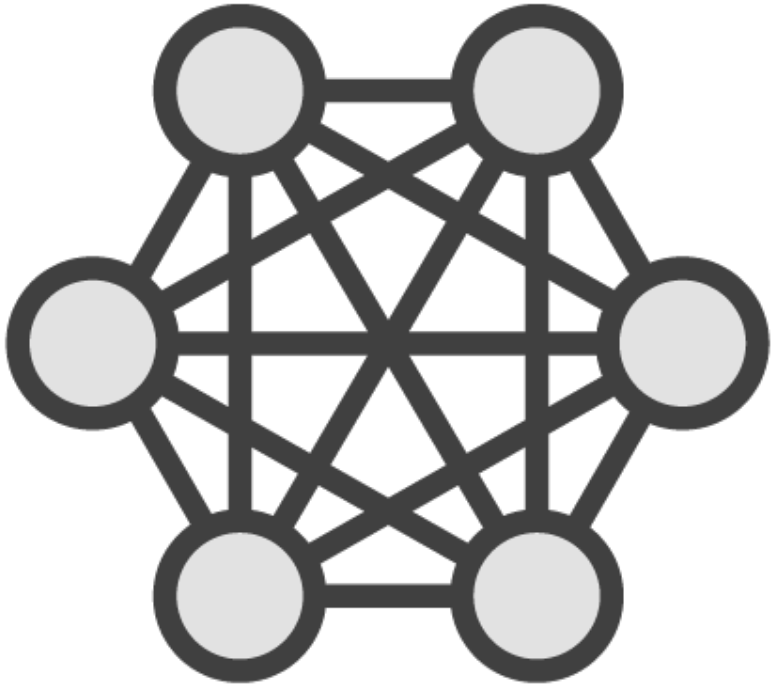
Working with Folders

# Demo



## Getting started with the Databricks CLI

# Azure Spark Clusters

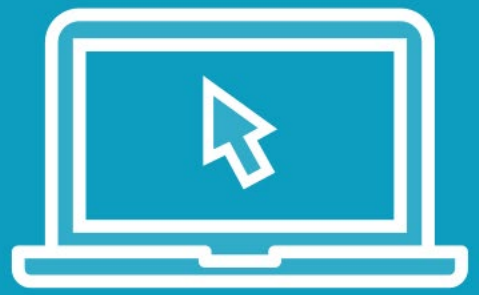


Unified cluster computing platform

Interactive or job

Manage using UI, CLI or Clusters API

# Demo



## Working with Spark clusters

# Notebooks

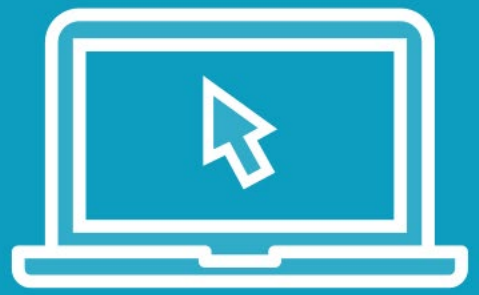
Web-based interface

Combine code, visualizations and text

Used for building pipelines

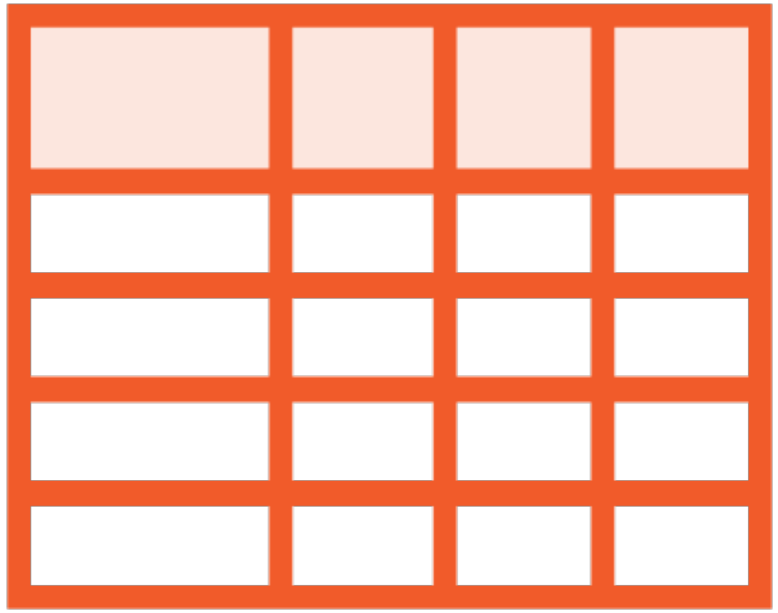
Support formats of Python, Scala, Markdown and more

# Demo



## Working with Notebooks in Azure Databricks

# Azure Databricks Tables



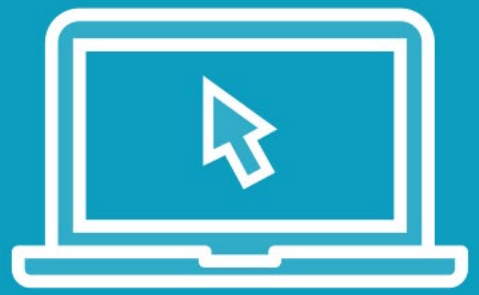
Collection of structured sata

Equivalent to Apache Spark DataFrame

Can be cached, filtered, queried, and more

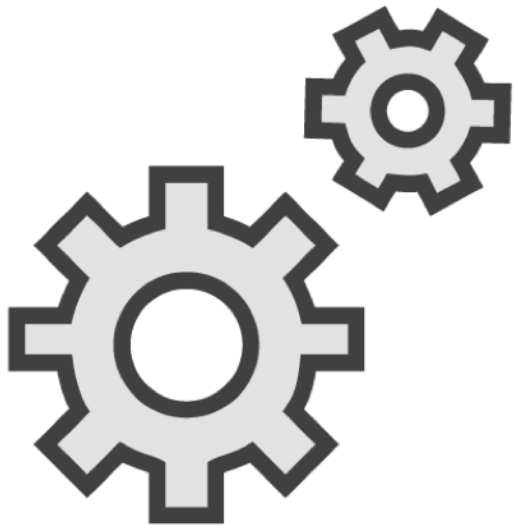
Temporary storage during transformation

# Demo



## Working with Tables

# Apache Spark Jobs

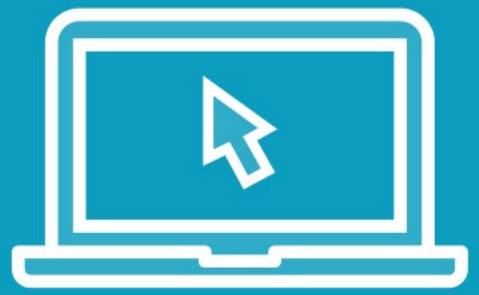


Run workflows interactively or scheduled

Accessible through UI, CLI or API

Limits

# Demo



## Creating Spark Jobs in Azure Databricks

# Summary

Azure Databricks is quick and easy to set up

Don’t forget about auto-scaling for your clusters

Use your favorite Language

Notebooks are key to working in Azure Databricks

# For Further Learning

Azure Databricks documentation at **docs.microsoft.com**
https://docs.microsoft.com/en-us/azure/azure-databricks/

Azure Databricks documentation at
docs.azuredatabricks.net
https://docs.azuredatabricks.net/user-guide/index.html

Remember the module exercise files

Questions? Join on the conversation at pluralsight.com

# Next Up:
# Performing ETL (Extract, Transform, Load) Operations with Azure Databricks