



Customer

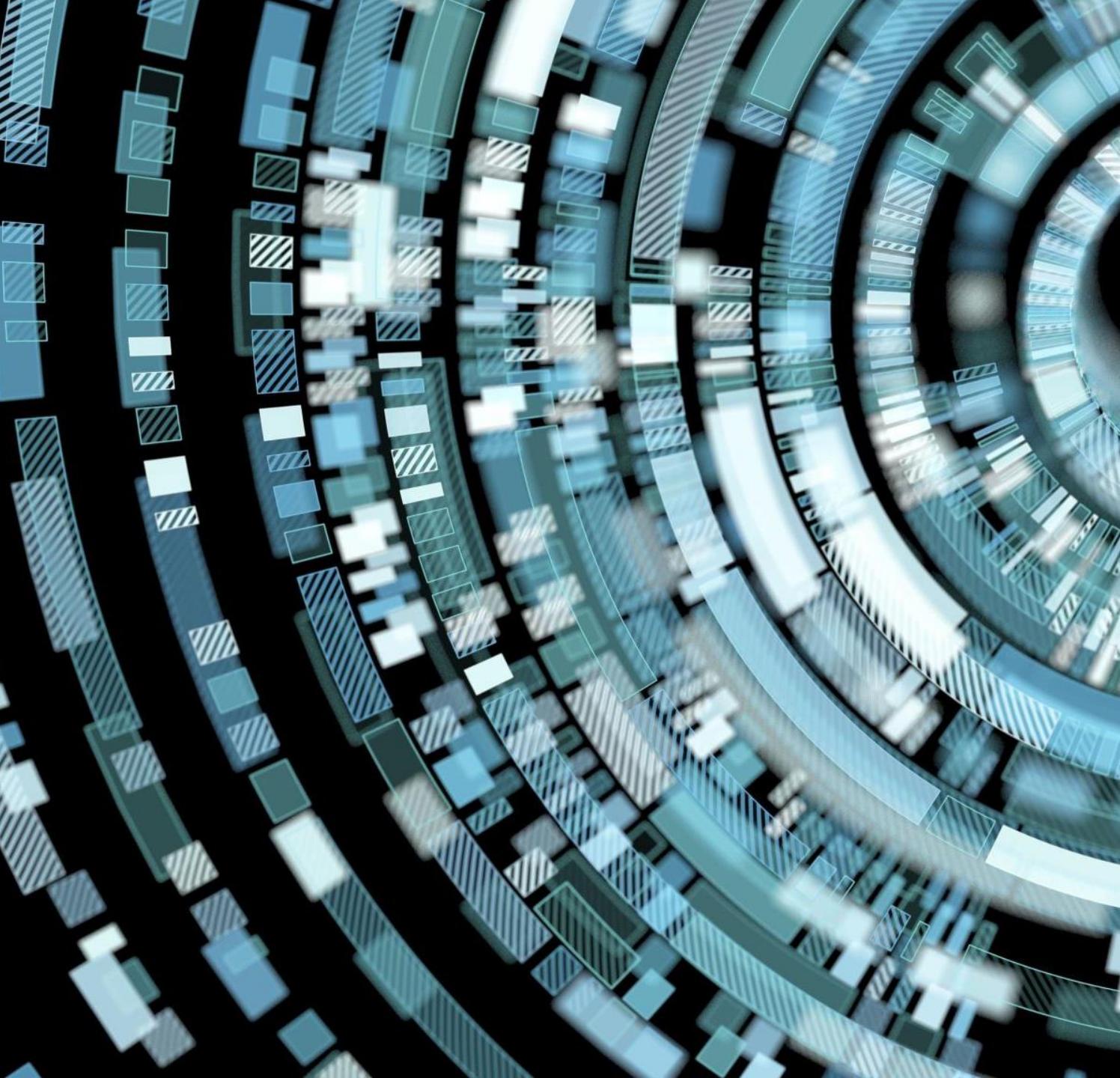
Azure Databricks workshop

Date



Contacts

- Serge Retkowsky
serge.retkowsky@microsoft.com





Program overview

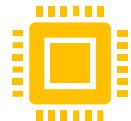
Workshop Prerequisites



Activate and Access an
[Azure Subscription](#)



Create an Azure Machine
Learning Workspace in the
Azure Portal : [LINK HERE](#)



Account for Azure DevOps
: [LINK HERE](#)



Prerequisites by Role



Recommended Azure ML
training on Microsoft
Learn

App Developers : VS Code, Python
ML Engineers : No prerequisites
Citizen Data Scientists : Python,
Jupyter Notebooks
Business Data Scientists : No
prerequisites
Enterprise Architects : Cloud
Policies and Security

ML with Code:
<https://docs.microsoft.com/en-us/learn/patterns/build-ai-solutions-with-azure-ml-service>

ML with No Code / Low Code:
<https://docs.microsoft.com/en-us/learn/patterns/create-no-code-predictive-models-azure-machine-learning>

ML at the Edge for IoT:
<https://docs.microsoft.com/en-us/learn/patterns/ai-edge-engineer>



Workshop 1

Azure ML fundamentals

- Presentation and demo of Azure ML.
- Presentation of Azure documentation resources & certifications path.



Workshop 2

Azure ML fundamentals

- Hands-on labs:
 - Azure ML experimentations
 - AutoML with Azure ML Python SDK
 - Estimators with Azure ML Python SDK
 - Interpretation & Fairness of ML models
 - Hyperparameter tuning with Azure ML
 - Model deployment



Workshop 3

No Code with Azure ML

- Hands-on labs:
 - AutoML graphical user interface
 - Azure ML Designer interface for building no code pipelines
 - Use case: Anomaly detection with Azure ML Designer
 - PowerBI Integration



Workshop 4

Azure Computer Vision

- Introduction to Azure Cognitive Services.
- Deep dive on Azure Computer Vision presentation.
- Hands-on lab:
 - Training of a custom vision model
 - Validation and deployment of a custom vision model



Workshop 5

MLOps

- Introduction to MLOps
- Hands-on lab:
 - Implementing CI/CD pipeline using GitHub Action & Azure DevOps



Workshop 6

Azure Databricks

- Azure Databricks presentation
- Hands-on lab:
 - Data preparation
 - ML
 - Model deployment
 - Azure ML integration



Agenda

1. Introduction
2. Présentation Azure Databricks
3. Atelier pratique Azure Databricks



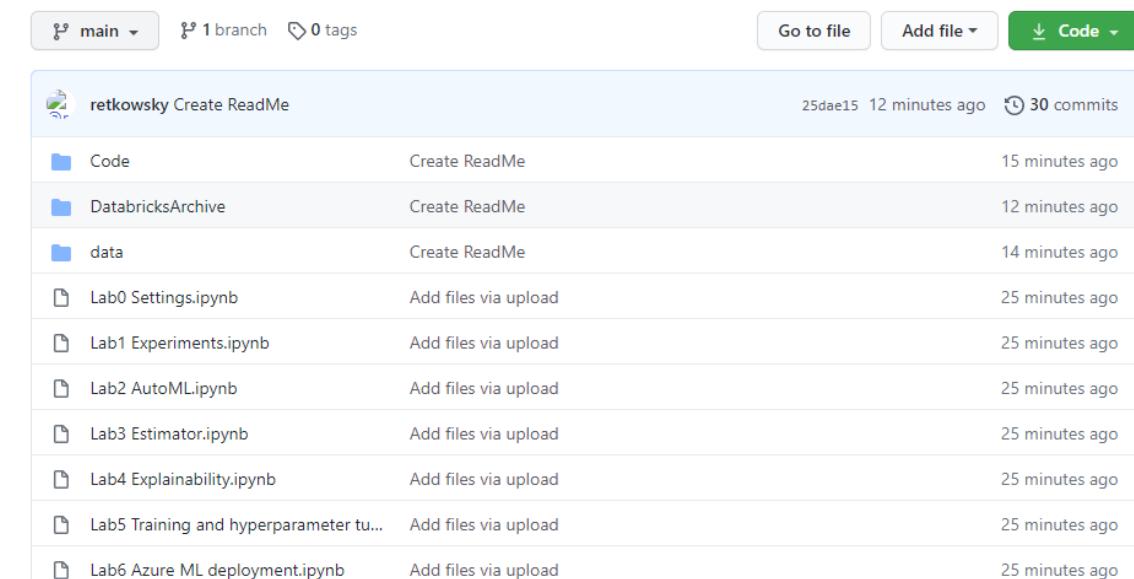
Workshop Azure Databricks

1. Présentation des principes du cloud Azure et des ressources Azure (documentation, parcours formation, certifications...).
2. Présentation détaillée du service Azure Databricks.
3. Atelier pratique Azure Databricks :
 - o Environnement Azure Databricks
 - o Préparation des données
 - o Machine Learning
 - o Déploiement des modèles
 - o Intégration Azure Machine Learning

Documents workshops Azure Machine Learning et Azure Databricks

Les notebooks Python et jeux de données des workshops sont disponibles ici:

<https://aka.ms/AAaf9y4>



A screenshot of a GitHub repository interface. At the top, there are buttons for 'main' (with a dropdown arrow), '1 branch', '0 tags', 'Go to file', 'Add file', and a green 'Code' button with a dropdown arrow. Below this is a table listing files and their details.

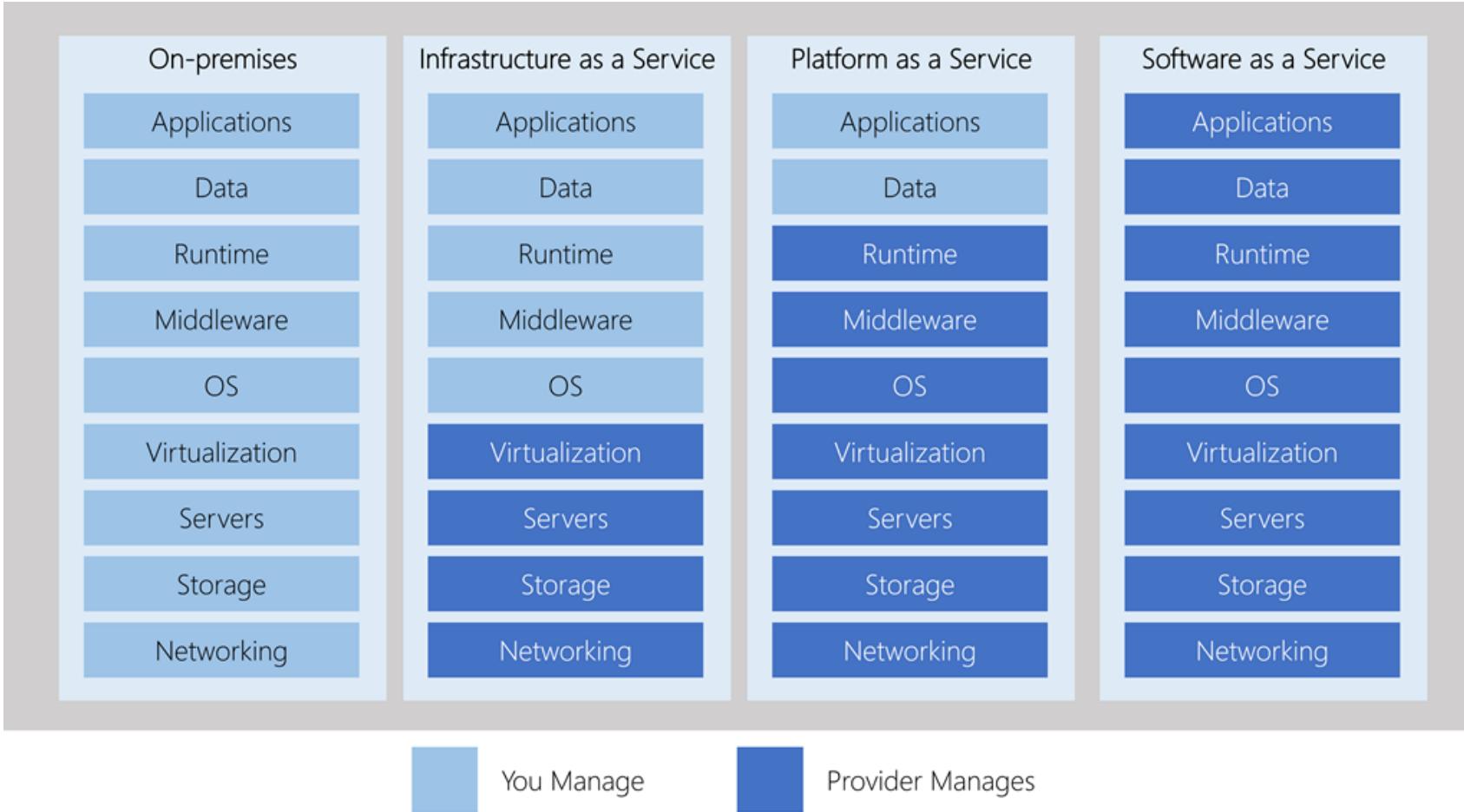
 retkowsky	Create ReadMe	25dae15 12 minutes ago	30 commits
Code	Create ReadMe	15 minutes ago	
DatabricksArchive	Create ReadMe	12 minutes ago	
data	Create ReadMe	14 minutes ago	
Lab0 Settings.ipynb	Add files via upload	25 minutes ago	
Lab1 Experiments.ipynb	Add files via upload	25 minutes ago	
Lab2 AutoMLipynb	Add files via upload	25 minutes ago	
Lab3 Estimator.ipynb	Add files via upload	25 minutes ago	
Lab4 Explainability.ipynb	Add files via upload	25 minutes ago	
Lab5 Training and hyperparameter tu...	Add files via upload	25 minutes ago	
Lab6 Azure ML deployment.ipynb	Add files via upload	25 minutes ago	



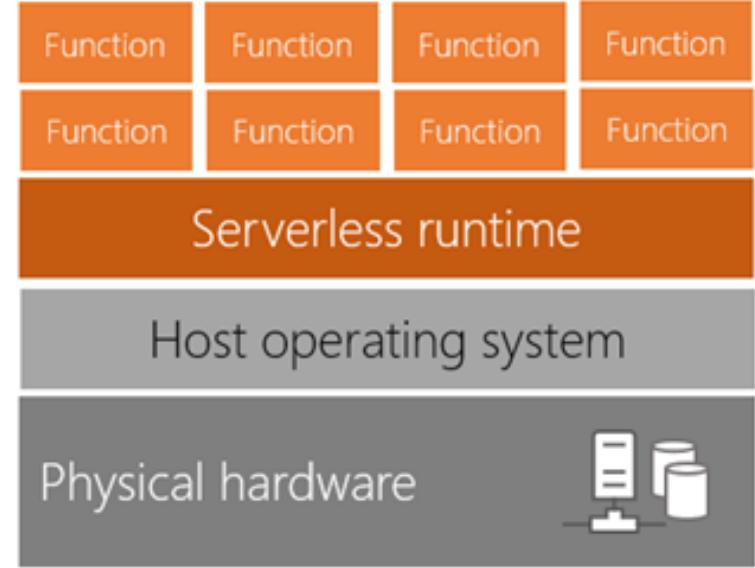
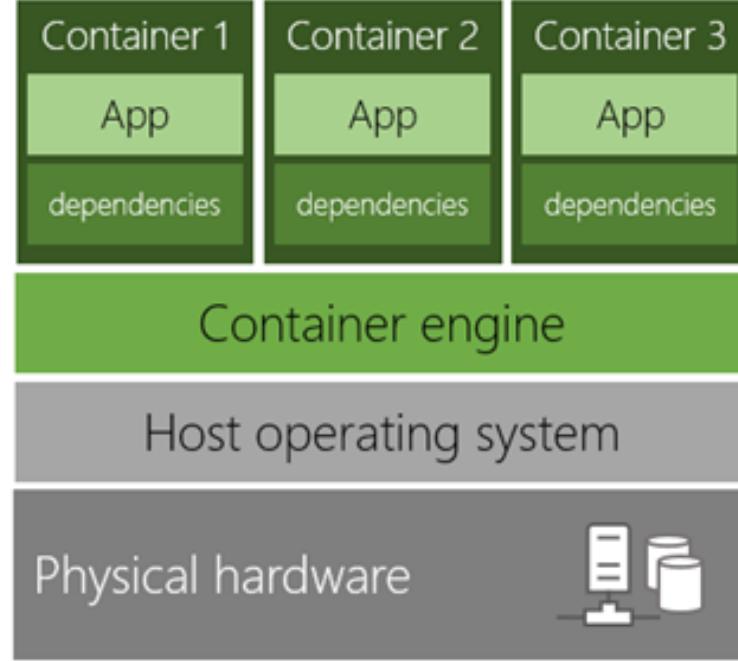
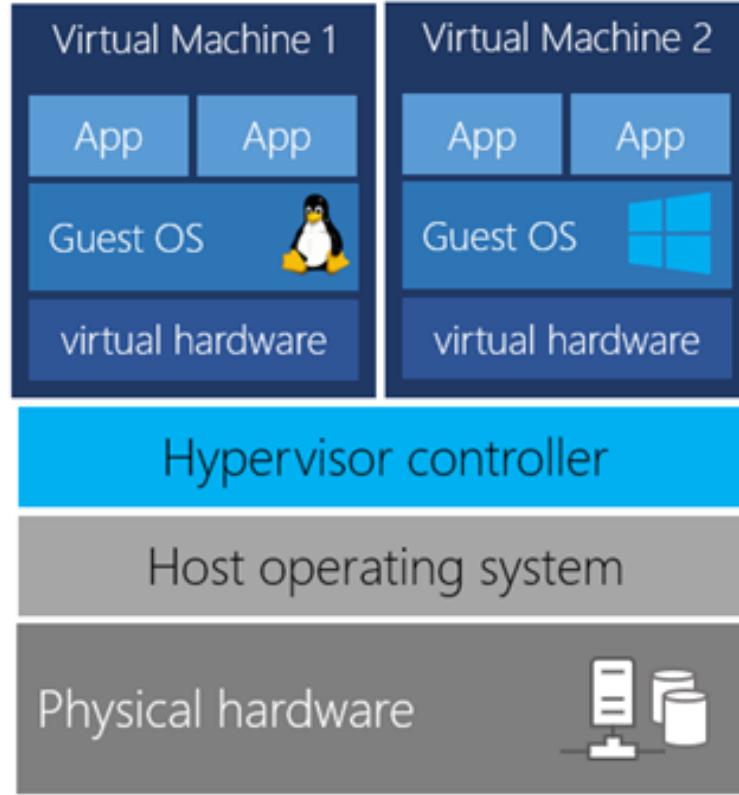
Agenda

1. Introduction
2. Présentation Azure Databricks
3. Atelier pratique Azure Databricks

Azure



Management responsibilities



Virtual machines

Containers

Serverless

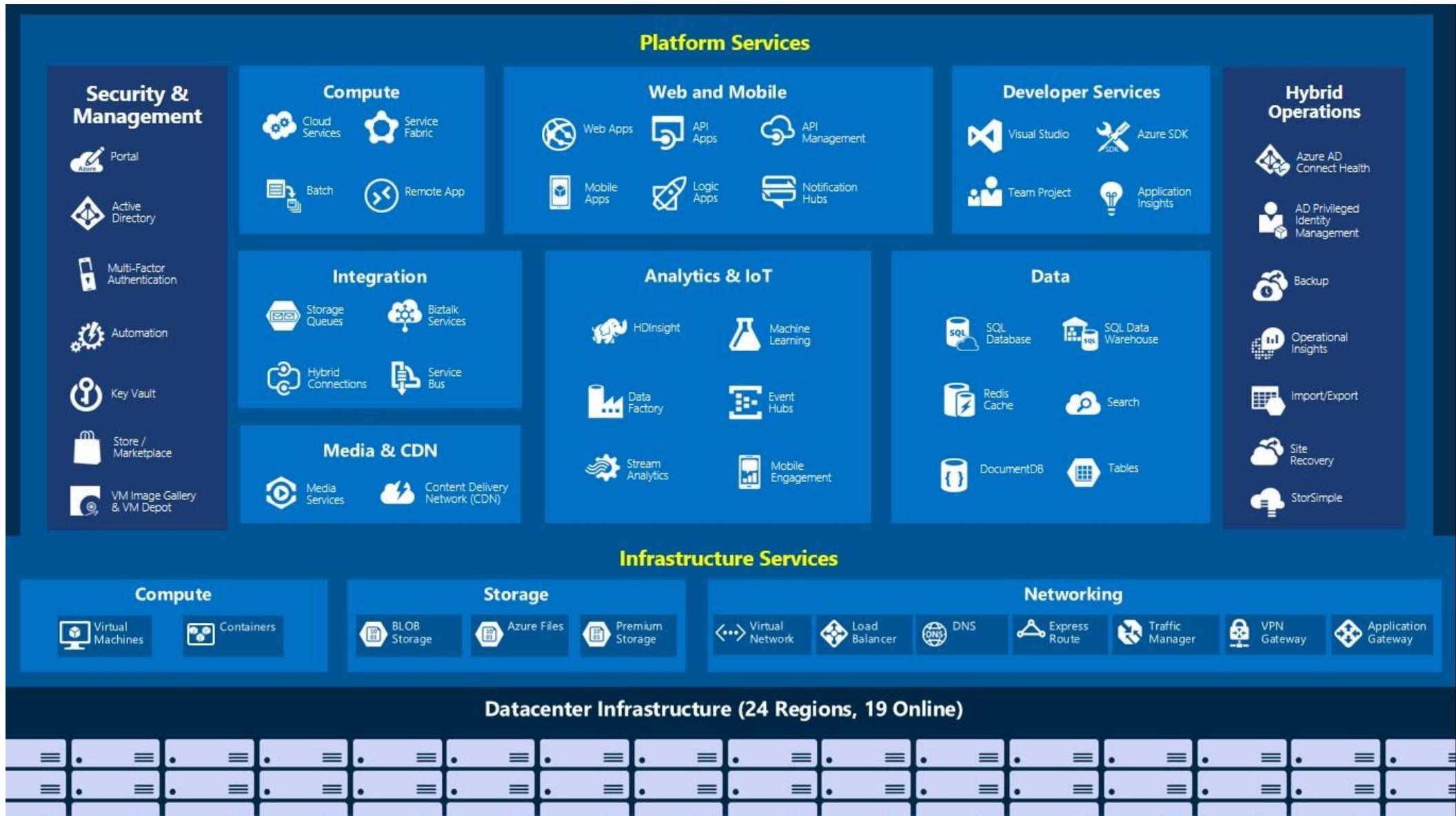
VM, Container, Serverless? |



Régions Azure |

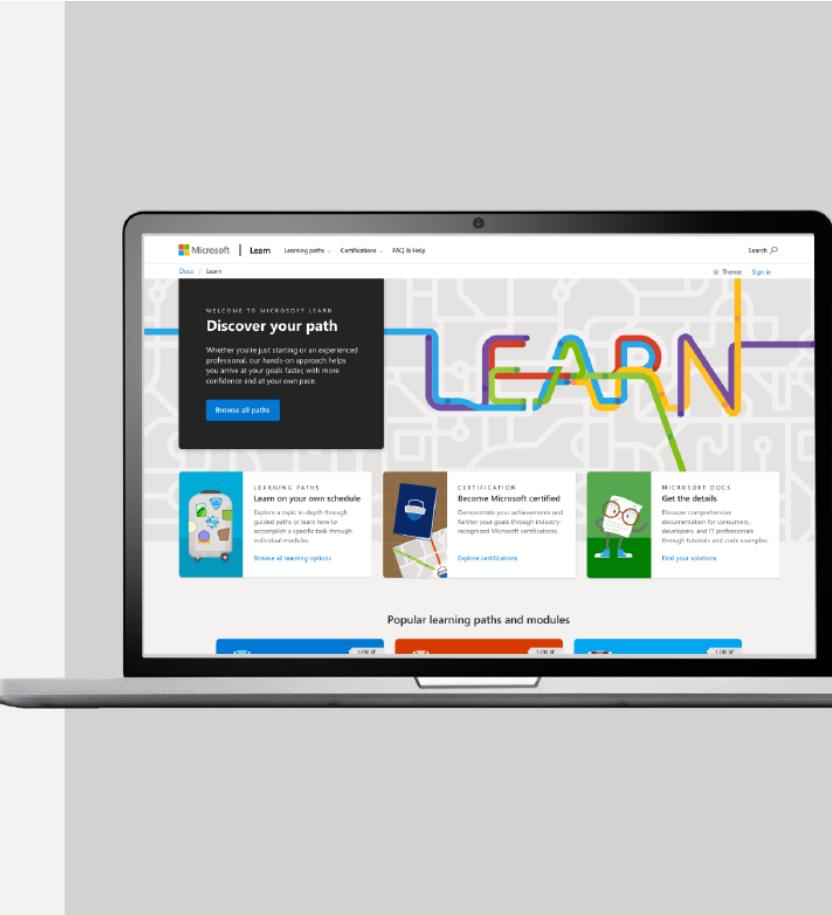
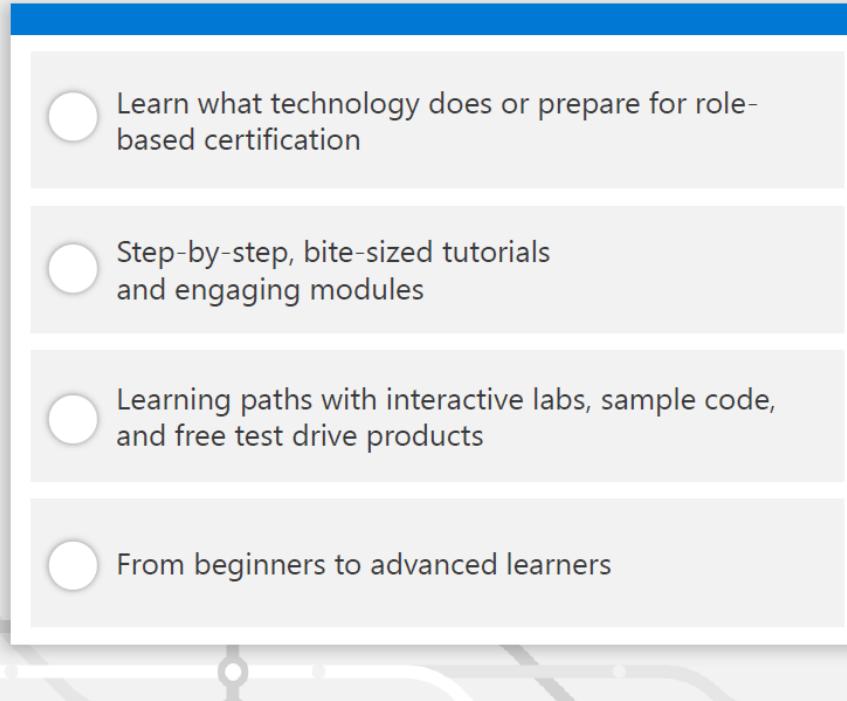
Azure services

<https://azure.microsoft.com/en-us/services/>



Microsoft Learn

Microsoft.com/Learn



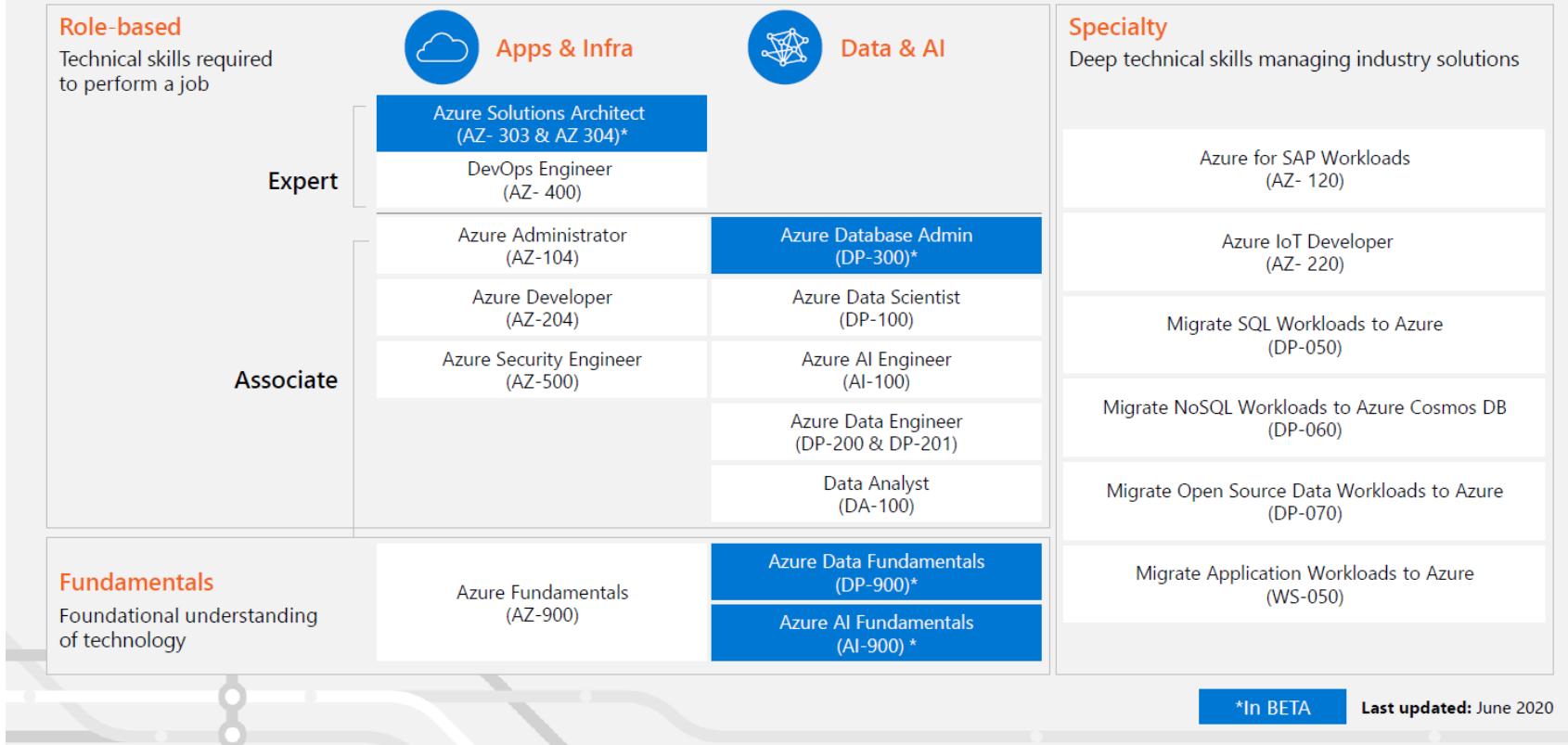
Microsoft Learn |
Microsoft Learn | Microsoft Docs

Certification	Role Description	Exam	Skills Measured	Microsoft Official Course	MOC duration
Azure Administrator Associate	Azure Administrators implement, manage, and monitor an organization's Microsoft Azure environment.	AZ-104: Microsoft Azure Administrator	https://aka.ms/az-104examguide	AZ-104T00 Microsoft Azure Administrator	4 days
				AZ-010T00 Azure Administrator for AWS SysOps	2 days
Azure AI Engineer Associate	Azure AI Engineers use cognitive services, machine learning, and knowledge mining to architect and implement Microsoft AI solutions involving natural language processing, speech, computer vision, and conversational AI.	AI-100: Designing and Implementing an Azure AI Solution	http://aka.ms/ai-100examguide	AI-100T01 Designing and Implementing an Azure AI Solution	3 days
Azure Data Engineer Associate	Azure Data Engineers design and implement the management, monitoring, security, and privacy of data solutions using the full stack of Microsoft Azure data services.	DP-200: Implementing an Azure Data Solution	http://aka.ms/dp-200examguide	DP-200T01 Implementing an Azure Data Solution	3 days
		DP-201: Designing an Azure Data Solution	http://aka.ms/d-201examguide	DP-201T01 Designing an Azure Data Solution	2 days
Azure Data Scientist Associate	Azure Data Scientists implement and run machine learning workloads on Microsoft Azure Machine Learning.	DP-100: Designing and Implementing a Data Science Solution on Azure	http://aka.ms/dp-100examguide	DP-100T01 Designing and Implementing a Data Science solution on Azure	3 days
Azure Database Administrator Associate	Azure Database Administrators implement and manage cloud and on-premises relational databases built on top of Microsoft SQL Server and Azure data services.	DP-300: Administering Relational Databases on Microsoft Azure	http://aka.ms/dp-300examguide	DP-300T00 Administering Relational Databases on Microsoft Azure	4 days
Azure Developer Associate	Azure Developers design, build, test, and maintain cloud applications and services.	AZ-204: Developing Solutions for Microsoft Azure	http://aka.ms/az-204examguide	AZ-204T00 Developing Solutions for Microsoft Azure	5 days
				AZ-020T00 Microsoft Azure solutions for AWS Developers	3 days

Certifications

[Parcourez les certifications et les examens | Microsoft Docs](#)

Azure training and certifications



Azure training & certifications |

Azure AI

AI on Azure

Domain specific pretrained models

To simplify solution development



Vision



Speech



Language



Search

Familiar Data Science tools

To simplify model development



Visual Studio Code



Jupyter



Command line

Popular frameworks

To build advanced deep learning solutions



PyTorch



TensorFlow



Scikit-Learn



ONNX

Productive services

To empower data science and development teams



Azure
Databricks



Azure Machine
Learning



Machine
Learning VMs

Powerful infrastructure

To accelerate deep learning



CPU



GPU



FPGA

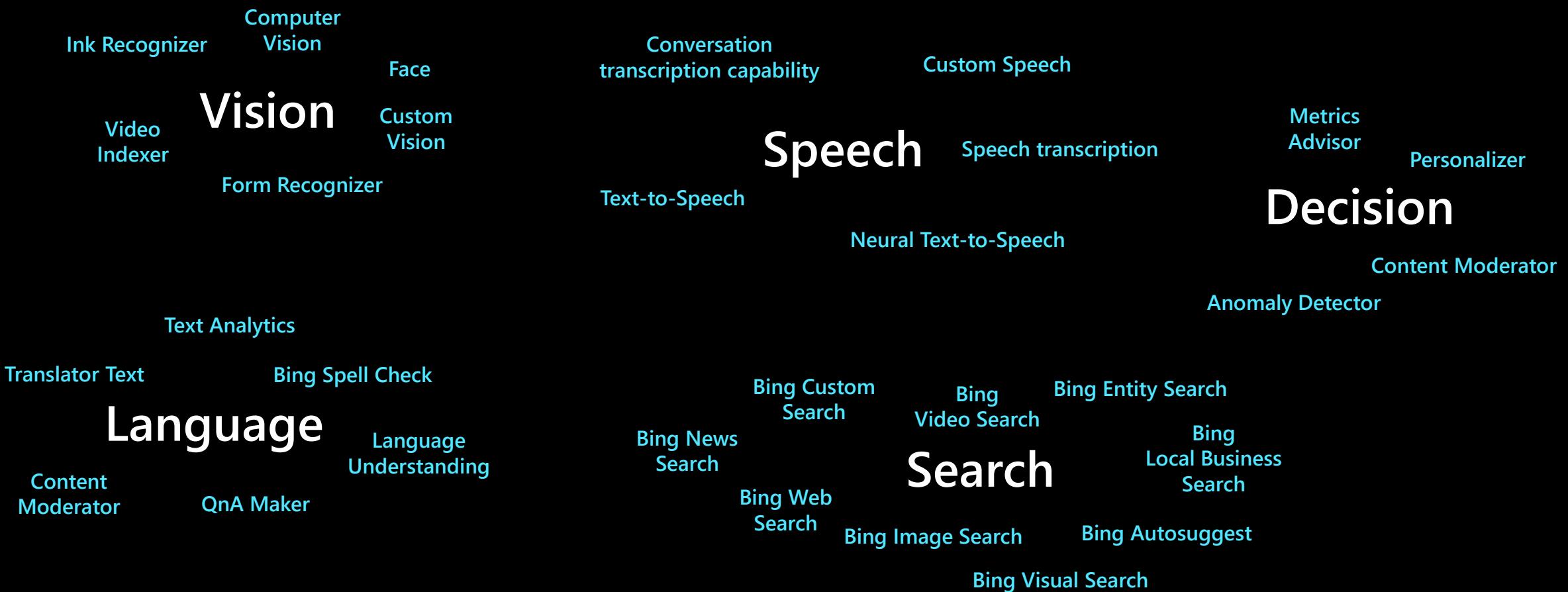


From the Intelligent Cloud to the Intelligent Edge



Azure Cognitive Services

The most comprehensive pre-trained AI



AI on Azure

Domain specific pretrained models

To simplify solution development



Vision



Speech



Language



Search

Familiar Data Science tools

To simplify model development



Visual Studio Code



Jupyter



Command line

Popular frameworks

To build advanced deep learning solutions



PyTorch



TensorFlow



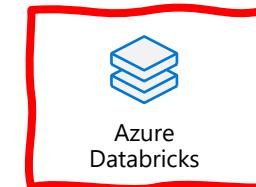
Scikit-Learn



ONNX

Productive services

To empower data science and development teams



Azure Machine Learning



Machine Learning VMs

Powerful infrastructure

To accelerate deep learning



CPU



GPU



FPGA



From the Intelligent Cloud to the Intelligent Edge

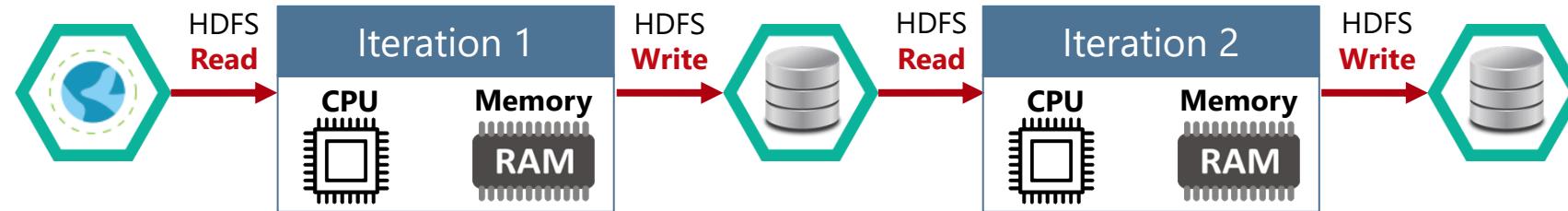




Spark ?

Motivation for Apache Spark

Traditional Approach: MapReduce jobs for complex jobs, interactive query, and online event-hub processing involves lots of **(slow) disk I/O**

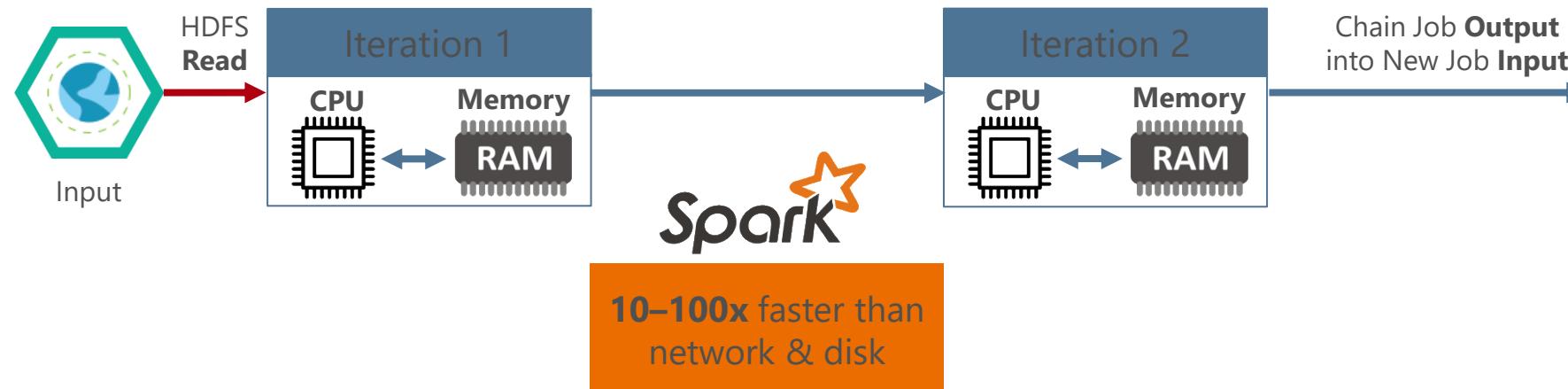


Motivation for Apache Spark

Traditional Approach: MapReduce jobs for complex jobs, interactive query, and online event-hub processing involves lots of **(slow) disk I/O**



Solution: Keep data **in-memory** with a new distributed execution engine



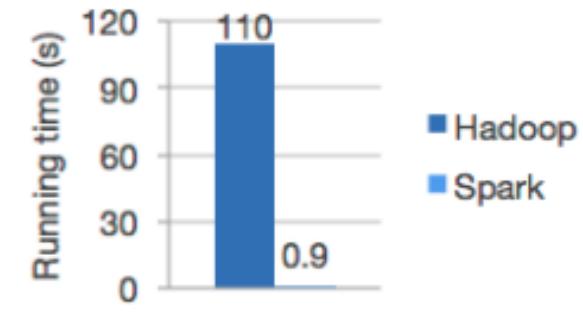
Apache Spark



Performance

High-quality algorithms, 100x faster than MapReduce.

Spark excels at iterative computation, enabling MLlib to run fast. At the same time, we care about algorithmic performance: MLlib contains high-quality algorithms that leverage iteration, and can yield better results than the one-pass approximations sometimes used on MapReduce.



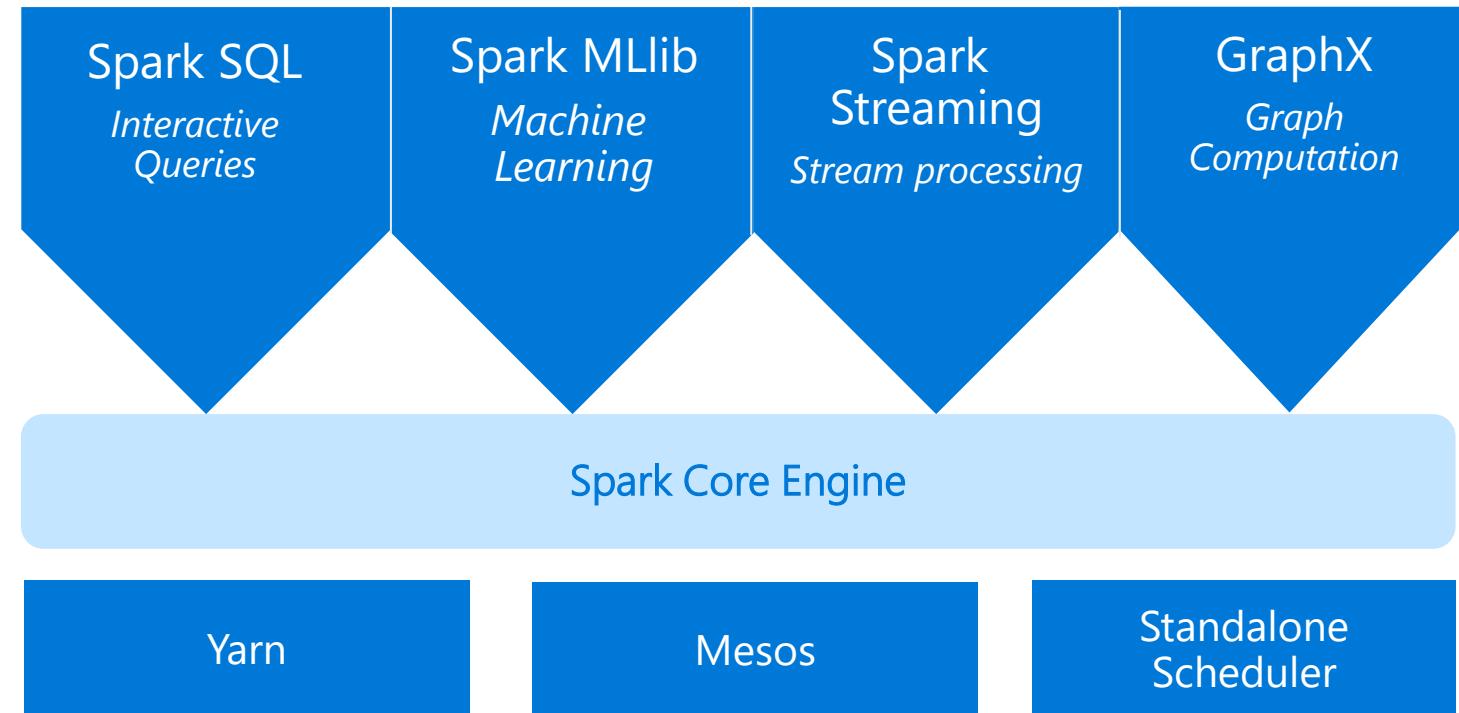
Logistic regression in Hadoop and Spark

Apache Spark

A unified, open source, parallel, data processing framework for Big Data Analytics

Spark Unifies:

- Batch Processing
- Interactive SQL
- Real-time processing
- Machine Learning
- Deep Learning
- Graph Processing



Apache Spark

MLlib



MLlib is **Spark's machine learning (ML) library**.
Its goal is to make practical **machine learning scalable and easy**.

At a high level, it provides tools such as:

- **ML Algorithms:** common learning algorithms such as classification, regression, clustering, and collaborative filtering
- **Featurization:** feature extraction, transformation, dimensionality reduction, and selection
- **Pipelines:** tools for constructing, evaluating, and tuning ML Pipelines
- **Persistence:** saving and load algorithms, models, and Pipelines
- **Utilities:** linear algebra, statistics, data handling, etc.

Algorithms

MLlib contains many algorithms and utilities.

ML algorithms include:

- Classification: logistic regression, naive Bayes,...
- Regression: generalized linear regression, survival regression,...
- Decision trees, random forests, and gradient-boosted trees
- Recommendation: alternating least squares (ALS)
- Clustering: K-means, Gaussian mixtures (GMMs),...
- Topic modeling: latent Dirichlet allocation (LDA)
- Frequent itemsets, association rules, and sequential pattern mining

ML workflow utilities include:

- Feature transformations: standardization, normalization, hashing,...
- ML Pipeline construction
- Model evaluation and hyper-parameter tuning
- ML persistence: saving and loading models and Pipelines

Other utilities include:

- Distributed linear algebra: SVD, PCA,...
- Statistics: summary statistics, hypothesis testing,...



Why Spark? Why Databricks ?

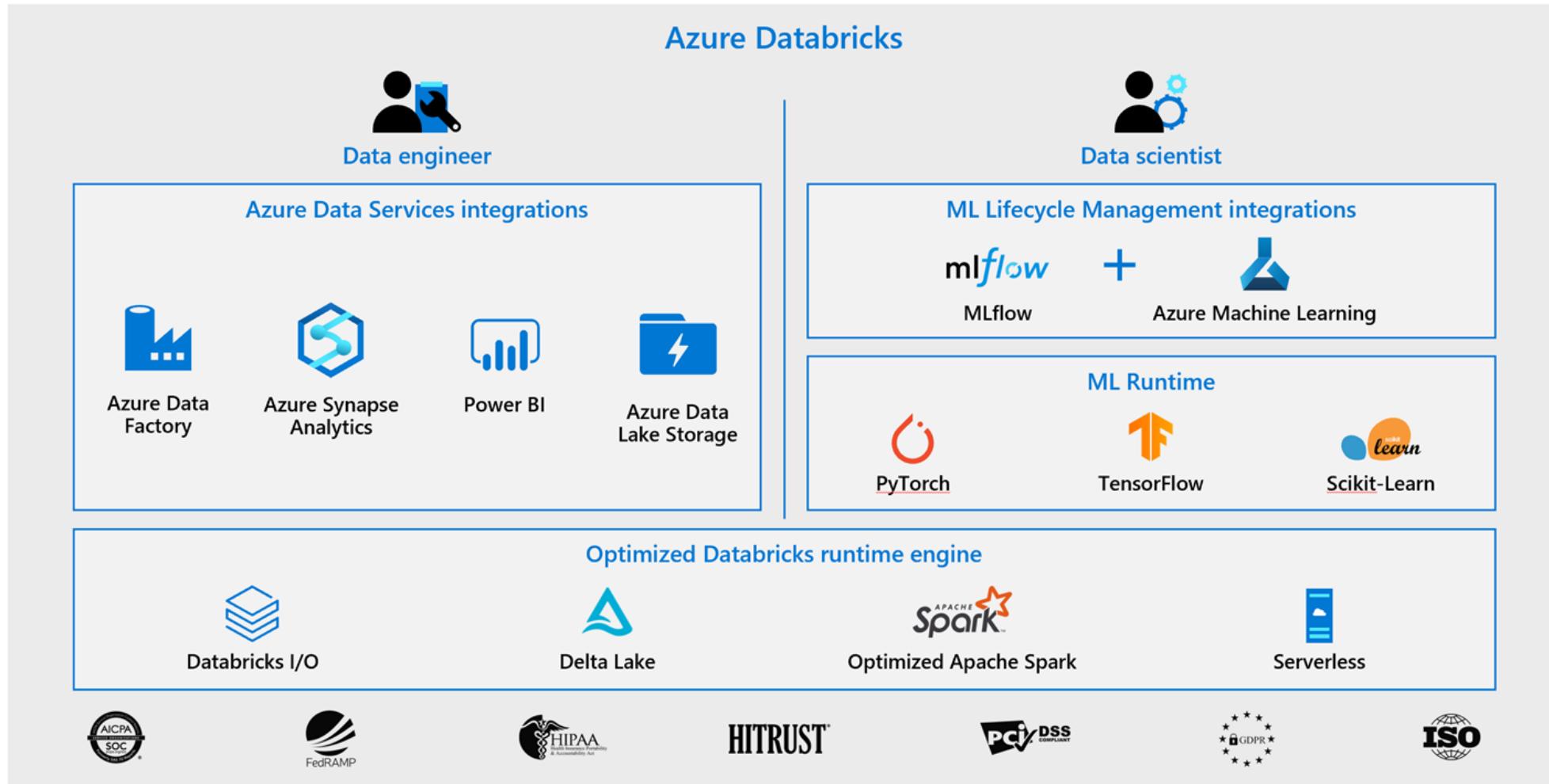
Why Spark?

- Open-source data processing engine built around speed, ease of use, and sophisticated analytics
- In memory engine that is up to 100 times faster than Hadoop
- Largest open-source data project with 1000+ contributors
- Highly extensible with support for Scala, Java and Python alongside Spark SQL, GraphX, Streaming and Machine Learning Library (MLlib)

Why Databricks?

- Databricks commits 75% of the code to Open-Source Spark
- Databricks is the premium version of Spark available in the market
- Spark founders created Databricks. Spark is the dominant workload in Hadoop

What is Azure Databricks?



Azure Databricks

Microsoft Azure | Databricks

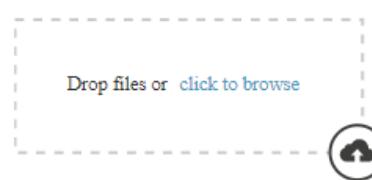


Azure Databricks



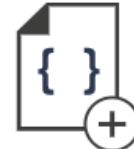
Explore the Quickstart Tutorial

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.



Import & Explore Data

Quickly import data, preview its schema, create a table, and query it in a notebook.



Create a Blank Notebook

Create a notebook to start querying, visualizing, and modeling your data.

Common Tasks

- New Notebook
- Create Table
- New Cluster
- New Job
- New MLflow Experiment
- Import Library
- Read Documentation

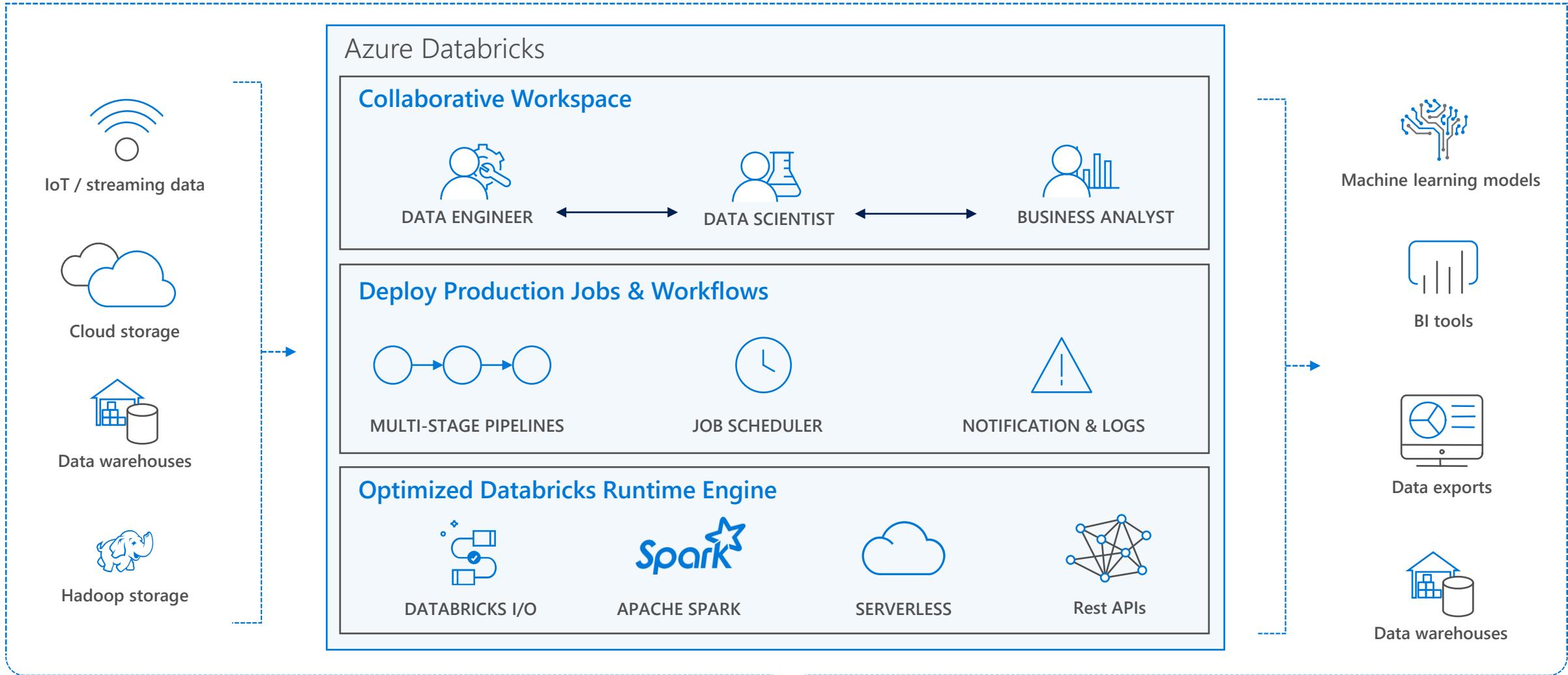
Recents

- Lab09 AutoML with Azure Databricks
- Lab13 MLflow Serving Models with Azure ML service
- Lab12 Azure Databricks et MLFlow - Batch
- Lab11 Azure Databricks et MLFlow - Scikit Learn
- Lab10 Azure Databricks et MLFlow - Introduction
- Lab02 Introduction Apache Spark avec Databricks
- Lab01 Installation et Configuration

Documentation

- Documentation
- Release Notes
- Getting Started

Azure Databricks



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

Collaborative Workspace

GET STARTED IN SECONDS

Single click to launch your new Spark environment

INTERACTIVE EXPLORATION

Explore data using interactive notebooks with support for multiple programming languages including R, Python, Scala, and SQL

COLLABORATION

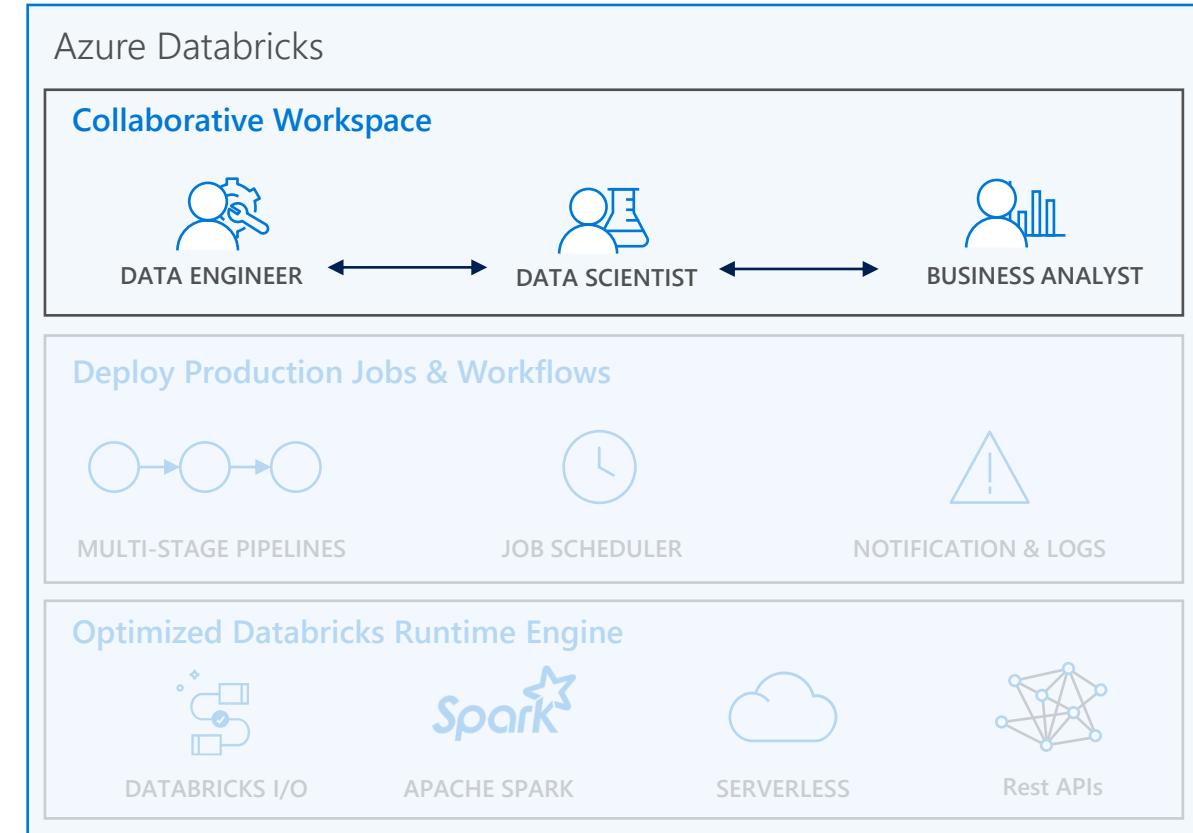
Work on the same notebook in real-time while tracking changes with detailed revision history, GitHub, or Bitbucket

VISUALIZATIONS

Visualize insights through a wide assortment of point-and-click visualizations. Or use powerful scriptable options like matplotlib, ggplot, and D3

DASHBOARDS

Rich integration with PowerBI to discover and share your insights in powerful new ways



Deploy Production Jobs & Workflows

JOB SCHEDULER

Execute jobs for production pipelines on a specific schedule

NOTEBOOK WORKFLOWS

Create multi-stage pipelines with the control structures of the source programming language

RUN NOTEBOOKS AS JOBS

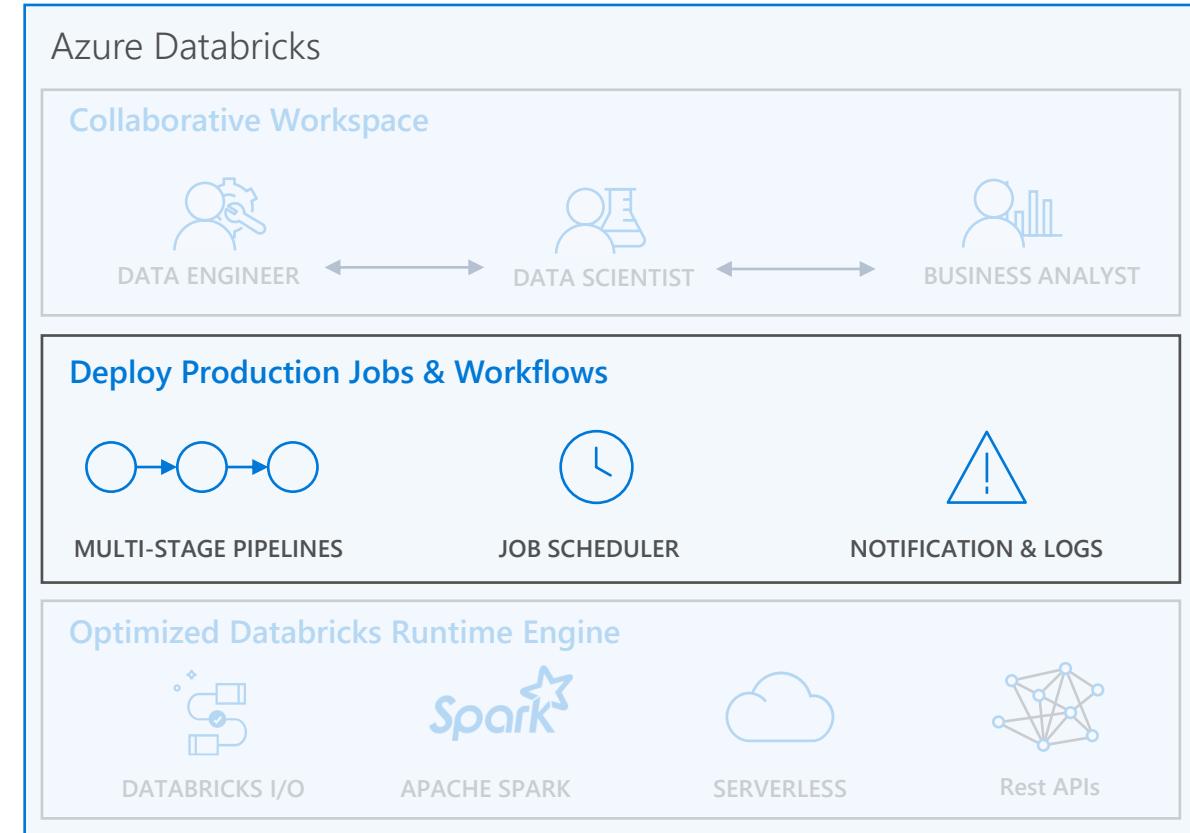
Turn notebooks or JARs into resilient Spark jobs with a click or an API call

NOTIFICATIONS AND LOGS

Set up alerts and quickly access audit logs for easy monitoring and troubleshooting

INTEGRATE NATIVELY WITH AZURE SERVICES

Deep integration with Azure SQL Data Warehouse, Cosmos DB, Azure Data Lake Store, Azure Blob Storage, and Azure Event Hub



Optimized Databricks Runtime Engine

OPTIMIZED I/O PERFORMANCE

The Databricks I/O module (DBIO) takes processing speeds to the next level — significantly improving the performance of Spark in the cloud

FULLY-MANAGED PLATFORM ON AZURE

Reap the benefits of a fully managed service and remove the complexity of big data and machine learning

SERVERLESS INFRASTRUCTURE

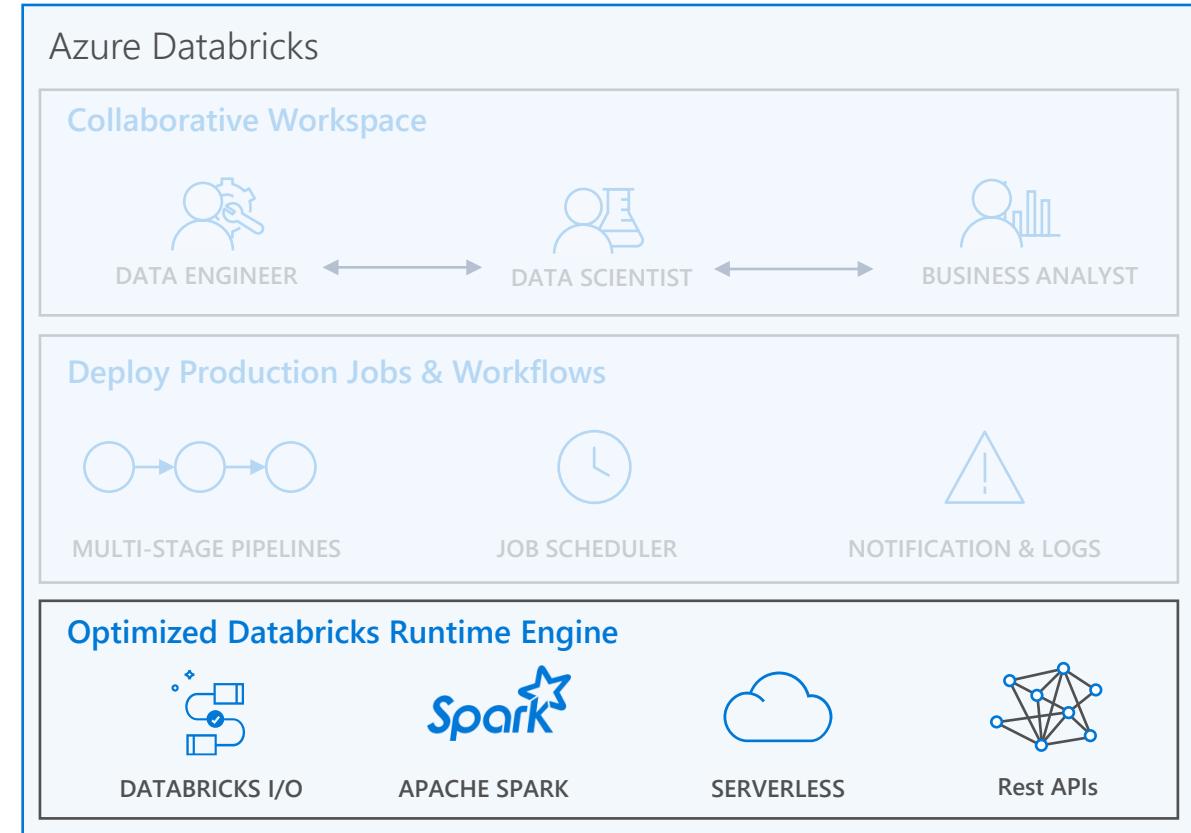
Databricks' serverless and highly elastic cloud service is designed to remove operational complexity while ensuring reliability and cost efficiency at scale

OPERATE AT MASSIVE SCALE

Without limits globally

SUPPORT FOR GPU ENABLED VMS

Specialized compute for your deep learning needs



En résumé



Moteur Spark optimisé

Traitements de données simples sur l'infrastructure à mise à l'échelle automatique, grâce à une instance Apache Spark™ fortement optimisée pour des gains de performances pouvant être multipliés par 50.



MLflow

Suivez et partagez des expériences, reproduisez des exécutions et gérez les modèles de manière collaborative à partir d'un référentiel central.



Blocs-notes collaboratifs

Affichez et explorez rapidement les données, trouvez et partagez de nouveaux insights et créez des modèles de manière collaborative avec les langages et les outils de votre choix.



Intégrations natives avec les services Azure

Enrichissez votre solution d'analytique et Machine Learning de bout en bout avec une forte intégration avec les services Azure, tels que Azure Data Factory, Azure Data Lake Storage, Azure Machine Learning et Power BI.



Sécurité de niveau d'entreprise

La sécurité native et fluide garantit la protection de vos données là où elles résident et crée des espaces de travail analytiques conformes, privés et isolés pour des milliers d'utilisateurs et de jeux de données.



Temps d'exécution Machine Learning

Accès en un clic à des environnements Machine Learning préconfigurés pour un Machine Learning augmenté avec des infrastructures de pointe et populaires telles que PyTorch, TensorFlow et scikit-Learn.



Choix du langage

Recourez à votre langage préféré, notamment Python, Scala, R, Spark SQL et .Net, que vous utilisez des ressources de calcul serverless ou provisionnées.



Delta Lake

Intégrez la fiabilité et la scalabilité des données à votre lac de données actuel avec une couche de stockage transactionnel open source conçue pour le cycle de vie complet des données.



Espaces de travail interactifs

Activez une collaboration fluide entre experts en mégadonnées, ingénieurs de données et analystes d'affaires.

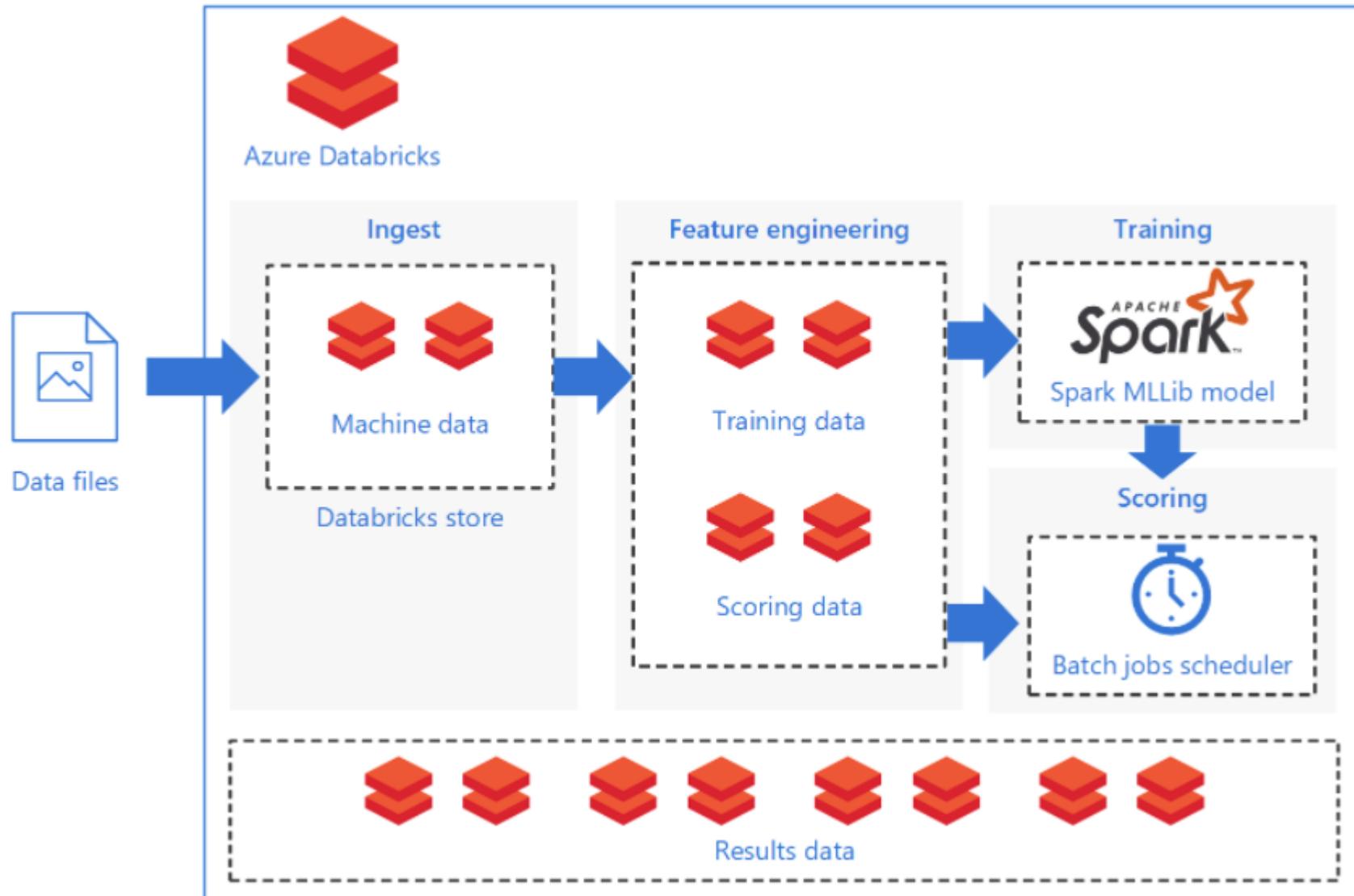


Prêt pour la production

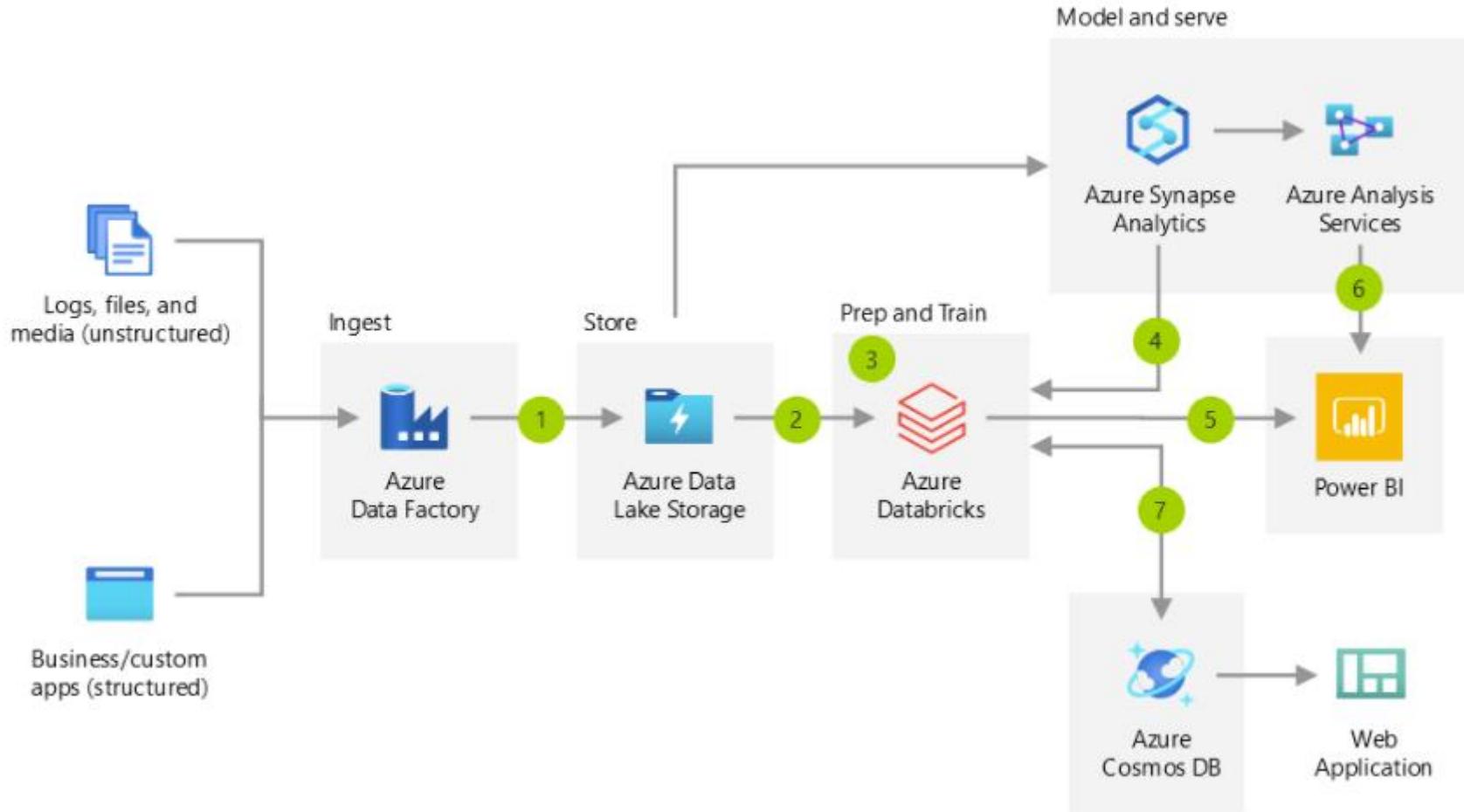
Exécutez et mettez à l'échelle vos charges de travail de données les plus critiques en toute confiance sur une plateforme de données de confiance, avec des intégrations d'écosystème pour CI/CD et la supervision.

Architecture

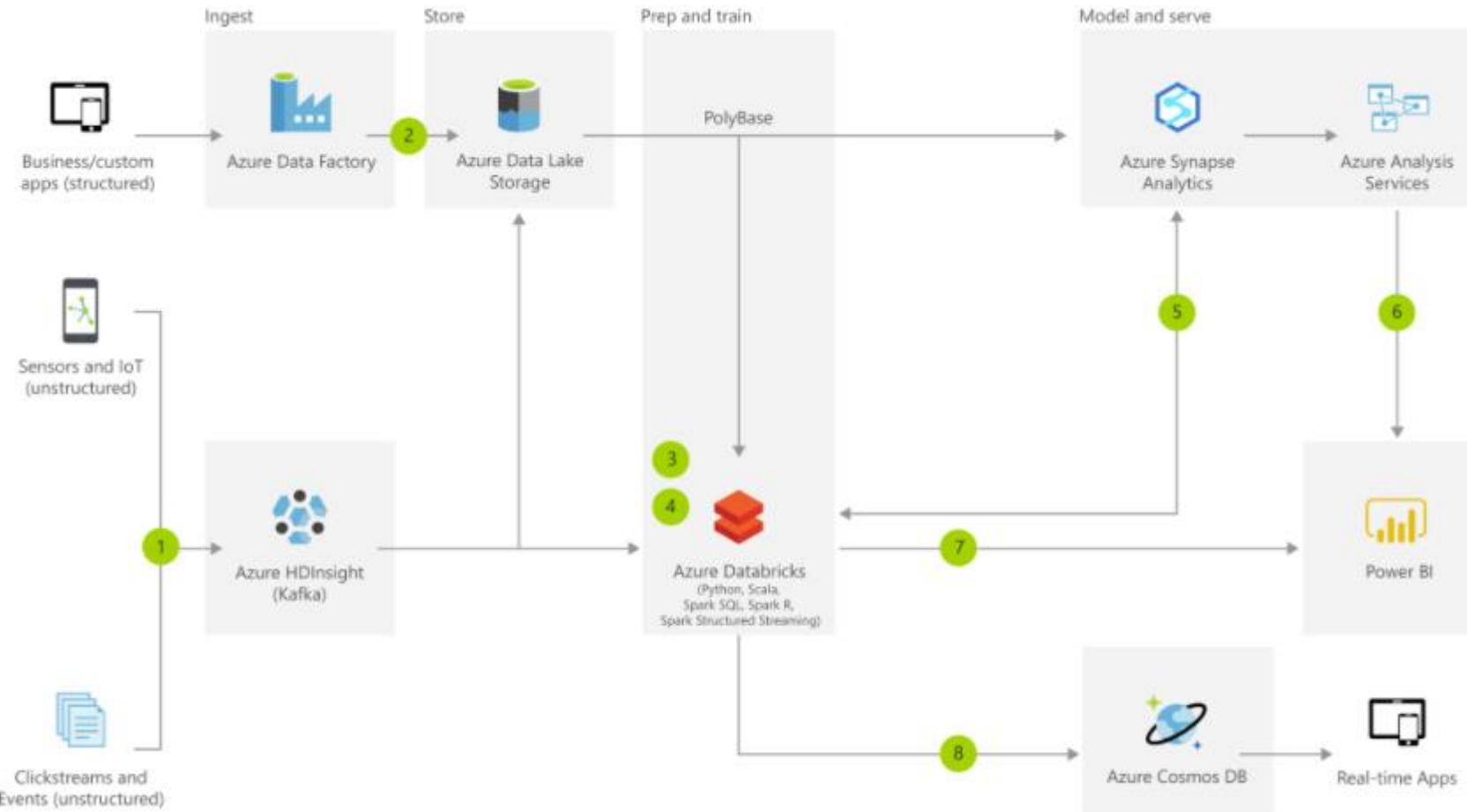
Architecture



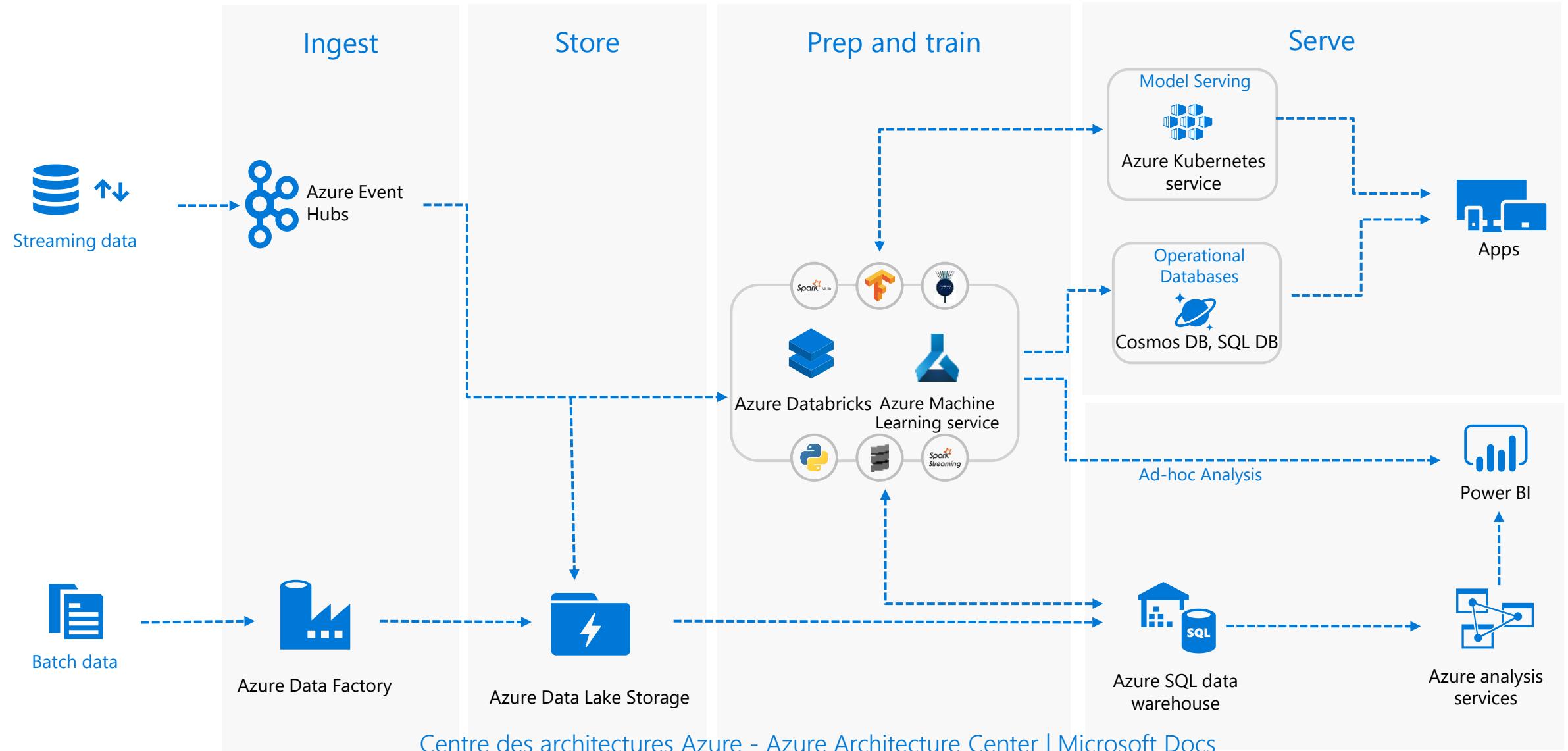
Analytics



Big Data Temps réel



Azure Databricks + Azure ML



Références

Shell invests in safety with Azure, AI, and machine vision to better protect customers and service champions

In the energy industry, Shell manages everything from wells to retail gas stations—44,000 of them. The company works hard to ensure the safety of service champions and customers at its retail sites. **Shell is piloting a new cloud-based, deep learning solution built on Microsoft Azure.** The solution uses closed-circuit camera footage and Internet of Things technology to automatically identify safety hazards and alert service champions so they can quickly respond and eliminate potential problems.

For the full story please see this [video](#)



Products and Services

Microsoft Azure
Azure Databricks
Azure IoT Edge
Azure IoT Hub

Organization Size

86,000 employees

Industry

Mining, Oil and Gas

Country

The Netherlands



Lennox improves service through vastly improved analysis of potential problems

Lennox International, is addressing thousands of consumer devices that are streaming IoT data to identify switch failures. **By implementing Azure Databricks, they have moved from 60% efficiency to 94% efficiency on detecting failures.**

Previously, processing 15 devices, which resulted in 2 million records took six hours. With Azure Databricks, they are now able to process 25,000 devices - about 10 billion records - in under 14 minutes. The real reason they chose Azure Databricks was the way it enables their data science and data engineering teams to collaborate - leading to much faster innovation.

This Power opens the flood gates to all kinds of new use cases at Lennox.



Products and Services
Microsoft Azure
Azure Databricks
Azure Storage

Organization Size
10,000 employees

Industry
Manufacturing/HVAC

Country
United States

Logistics provider accelerates future innovation with Azure Databricks

When LINX Cargo Care Group, a national supply chain and logistics provider headquartered in Sydney, Australia, broke away from its former parent company Asciano and was purchased by an investment firm, it took the opportunity to jettison its legacy IT systems and forge a more innovative future. Partnering with Empired, an IT solutions provider based in Perth, **LINX adopted Microsoft Azure and Azure Databricks** to benefit from a scalable, flexible, and agile cloud and analytics platform, which it uses to quickly develop IT solutions that propel its business forward.



Products and Services

Microsoft Azure
Azure Databricks
Azure Data Factory
Azure Data Lake Storage

Organization Size

3,800 employees

Industry

Travel and transportation

Country

Australia

Partner

Empired



Energy intelligence: renewables.AI uses Azure and Apache Spark to help build a stable and profitable solar energy market

Solar power producers need to predict when they will generate electricity and coordinate distribution with multiple energy markets. That's why **renewables.AI uses Microsoft Azure and Apache Spark** to deliver a data analytics service that helps solar energy providers manage production and identify high-value energy markets. With **Azure Databricks** and other Azure services, renewables.AI can streamline product development, drive solar industry revenue, and promote a stable energy future for everyone.



Products and Services

- Azure Databricks
- Azure SQL Database
- Azure Service Fabric
- Azure Event Hubs
- Azure Data Lake

Organization Size
12 Employees

Industry
Energy

Country
United Kingdom



Documentation & resources

Documentation

- Azure Databricks
<https://azure.microsoft.com/fr-fr/services/databricks/Azure>
- Databricks documentation
<https://docs.microsoft.com/fr-fr/azure/databricks/>
- Azure Databricks pricing
<https://azure.microsoft.com/fr-fr/pricing/details/databricks/>
- Release notes
<https://docs.microsoft.com/en-us/azure/databricks/release-notes/product/>
- Getting Started
<https://docs.microsoft.com/en-us/azure/databricks/scenarios/quickstart-create-databricks-workspace-portal?tabs=azure-portal>

Microsoft Learn

Formations gratuites

<https://docs.microsoft.com/en-us/learn/browse/?terms=databricks>

databricks

Search

16 results for "databricks"

 MODULE
Describe Azure Databricks
53 min ★★★★★ 4.5 (502)
Azure Data Engineer Intermediate

 LEARNING PATH
Data engineering with Azure Databricks
10 hr 32 min
Azure Data Engineer Intermediate

 MODULE
Read and write data in Azure Databricks
1 hr ★★★★★ 4.5 (255)
Azure Data Engineer Intermediate

 MODULE
Work with DataFrames in Azure Databricks
46 min ★★★★★ 4.7 (181)
Azure Data Engineer Intermediate

 MODULE
Work with DataFrames columns in Azure Databricks
43 min ★★★★★ 4.6 (114)
Azure Data Engineer Intermediate

 MODULE
Describe lazy evaluation and other performance features in Azure Databricks
38 min ★★★★★ 4.7 (117)
Azure Data Engineer Intermediate

 MODULE
Work with DataFrames advanced methods in Azure Databricks
42 min ★★★★★ 4.6 (107)
Azure Data Engineer Intermediate

 MODULE
Describe platform architecture, security, and data protection in Azure Databricks
1 hr 3 min ★★★★★ 4.6 (83)
Azure Data Engineer Intermediate

 MODULE
Create production workloads on Azure Databricks with Azure Data Factory
33 min ★★★★★ 4.5 (103)
Azure Data Engineer Intermediate



Demo

Pause – 10 Min



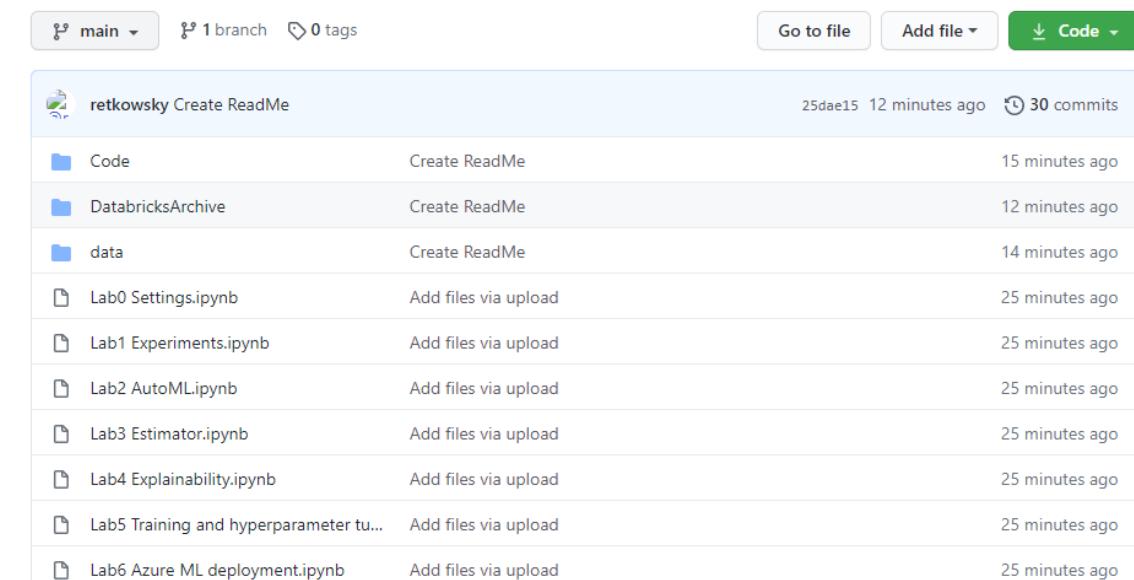
Agenda

- Introduction
- Présentation Azure Databricks
- Atelier pratique Azure Databricks

Documents workshops Azure Machine Learning et Azure Databricks

Les notebooks Python et jeux de données des workshops sont disponibles ici:

<https://aka.ms/AAaf9y4>



A screenshot of a GitHub repository interface. At the top, there are buttons for 'main' (with a dropdown arrow), '1 branch', '0 tags', 'Go to file', 'Add file', and a green 'Code' button with a dropdown arrow. Below this is a table listing files and their details.

retkowsky Create ReadMe		
Code	Create ReadMe	15 minutes ago
DatabricksArchive	Create ReadMe	12 minutes ago
data	Create ReadMe	14 minutes ago
Lab0 Settings.ipynb	Add files via upload	25 minutes ago
Lab1 Experiments.ipynb	Add files via upload	25 minutes ago
Lab2 AutoMLipynb	Add files via upload	25 minutes ago
Lab3 Estimator.ipynb	Add files via upload	25 minutes ago
Lab4 Explainability.ipynb	Add files via upload	25 minutes ago
Lab5 Training and hyperparameter tu...	Add files via upload	25 minutes ago
Lab6 Azure ML deployment.ipynb	Add files via upload	25 minutes ago

1. Crédit workspace Azure Databricks

Création d'un workspace
depuis Azure Portal

Microsoft Azure (Preview) Report a bug Search resources, services, and docs (G+)

Create a resource Home Dashboard All services FAVORITES What's new Recent All resources Help + support Activity log Advisor Alerts App Services Azure Active Directory Azure Data Explorer Clu... Analysis Services Application Insights Azure Databricks Azure Synapse Analytics... Azure Synapse Analytics... Azure DevOps organizat... Bot Services Cognitive Services Container instances Azure Cosmos DB Data factories Data Box

Dashboard > New > Azure Databricks Microsoft

Azure Databricks Microsoft Save for later Azure benefit eligible Create

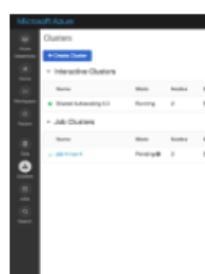
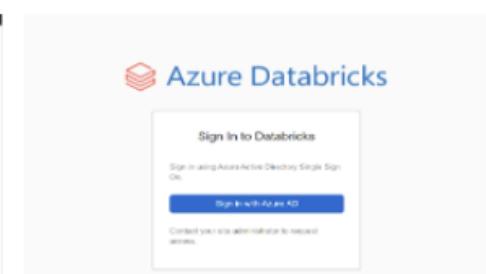
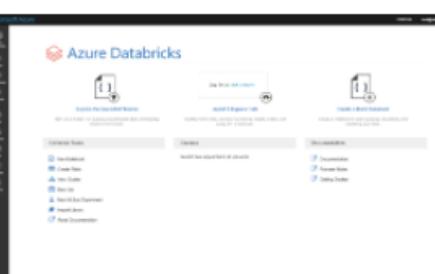
Overview Plans Usage Information + Support Reviews

Fast, easy, and collaborative Apache Spark-based analytics platform Accelerate innovation by enabling data science with a high-performance analytics platform that's optimized for Azure. Drive innovation and increase productivity Bring teams together in an interactive workspace. From data gathering to model creation, use Databricks Notebooks to unify the pro... Launch your new Spark environment with a single click. Integrate effortlessly with a wide variety of data stores and services such as A... DB, Azure Data Lake Store, Azure Blob storage, and Azure Event Hub. Add advanced artificial intelligence (AI) capabilities instantly and integration with PowerBI.

Build on secure, trusted cloud Protect your data and business with Azure Active Directory integration, role-based controls, and enterprise-grade SLAs. Get peace of... enabling secure access to Databricks Notebooks, clusters, jobs, and data.

Scale without limits Globally scale your analytics and data science projects. Build and innovate faster using advanced machine learning capabilities. Add complexity with a fully managed, cloud-native platform. Target any size data or project using a complete set of analytics technologies GraphX.

Media



1. Création workspace Azure Databricks

Création d'un workspace Azure Databricks depuis Azure Portal.

Une version d'évaluation gratuite de 14 jours est disponible.

Il est recommandé d'utiliser la version PREMIUM.

Microsoft Azure (Preview) Report a bug Search resources, services, and docs (G+)

Create a resource Home Dashboard All services FAVORITES What's new Recent All resources Help + support Activity log Advisor Alerts App Services Azure Active Directory Azure Data Explorer Clu... Analysis Services Application Insights Azure Databricks Azure Synapse Analytics... Azure Synapse Analytics... Azure DevOps organizat... Bot Services Cognitive Services Container instances Azure Cosmos DB Data factories Data Box Azure Stack Edge / Data...

Dashboard > New > Azure Databricks > Create an Azure Databricks workspace

Basics Networking Tags Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Microsoft Azure Internal Consumption

Resource group * (New) axadatabricks-rg Create new

Instance Details

Workspace name * axadatabricks

Region * West Europe

Pricing Tier * Standard (Apache Spark, Secure with Azure AD)

Standard (Apache Spark, Secure with Azure AD)

Premium (+ Role-based access controls)

Trial (Premium - 14-Days Free DBUs)

Review + create < Previous Next : Networking >

 **axadatabricks** Azure Databricks Service Search (Ctrl+/<<)Delete[JSON View](#) Overview Activity log Access control (IAM) Tags**Settings** Virtual Network Peerings Encryption Properties Locks**Monitoring** Diagnostic settings**Automation** Tasks (preview) Export template**Support + troubleshooting** New support request^ Essentials

Status : Active

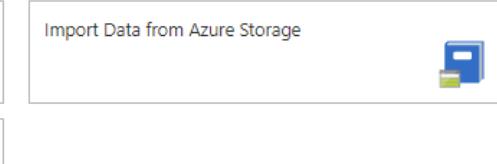
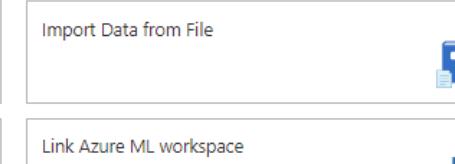
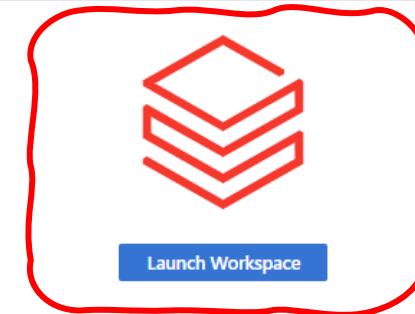
Managed Resource Group : [databricks-rg-axadatabricks-5xsqsdz3h6o5g](#)Resource group : [axadatabricks-rg](#)URL : <https://adb-5071530083052474.14.azure.databricks.net>

Location : West Europe

Pricing Tier : premium

Subscription : Microsoft Azure Internal Consumption

Subscription ID : 70b8f39e-8863-49f7-b6ba-34a80799550c

Tags ([change](#)) : Click here to add tags

1. Création workspace Azure Databricks

Le workspace est maintenant disponible.

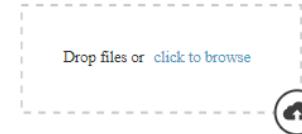


Azure Databricks



Explore the Quickstart Tutorial

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.



Import & Explore Data

Quickly import data, preview its schema, create a table, and query it in a notebook.



Create a Blank Notebook

Create a notebook to start querying, visualizing, and modeling your data.

Common Tasks

-  New Notebook
-  Create Table
-  New Cluster
-  New Job
-  New MLflow Experiment
-  Import Library
-  Read Documentation

Recents

Recent files appear here as you work.

Documentation

-  Documentation
-  Release Notes
-  Getting Started

2. Accès workspace

On accède au workspace.

3. Création cluster Azure Databricks

On va créer un nouveau cluster Databricks.

Sélectionnez un runtime Databricks ML.

Microsoft Azure | Databricks

Create Cluster

New Cluster Cancel Create Cluster

Cluster Name: clusterdb

Cluster Mode: Standard

Pool: None

Databricks Runtime Version: Runtime: 7.4 ML (Scala 2.12, Spark 3.0.1) Learn more

Autopilot Options:

Enable autoscaling

Terminate after 120 minutes of inactivity

Worker Type: Standard_DS3_v2 (14.0 GB Memory, 4 Cores, 0.75 DBU)

Min Workers: 2

Max Workers: 8

Driver Type: Same as worker (14.0 GB Memory, 4 Cores, 0.75 DBU)

▶ Advanced Options

2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores, 1.5-6 DBU
1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU

The screenshot shows the Microsoft Azure Databricks interface for creating a new cluster. The 'Clusters' tab is selected in the sidebar. The main area is titled 'Create Cluster' and 'New Cluster'. A red box highlights the 'New Cluster' tab, the 'Cluster Name' input field ('clusterdb'), and the 'Create Cluster' button. The 'Cluster Mode' is set to 'Standard', 'Pool' is 'None', and the 'Databricks Runtime Version' is 'Runtime: 7.4 ML (Scala 2.12, Spark 3.0.1)'. Under 'Autopilot Options', 'Enable autoscaling' and 'Terminate after 120 minutes of inactivity' are checked. The 'Worker Type' is 'Standard_DS3_v2 (14.0 GB Memory, 4 Cores, 0.75 DBU)', with 'Min Workers' at 2 and 'Max Workers' at 8. The 'Driver Type' is 'Same as worker (14.0 GB Memory, 4 Cores, 0.75 DBU)'. There is also a 'Advanced Options' link.



Clusters

All-Purpose Clusters

Job Clusters

Pools

Cluster Policies

+ Create Cluster

Name	State	Nodes	Runtime
clusterdb	Pending	0	7.4 ML (includes Apache Spark 3.0.1, Scala 2.12)



Home



Workspace



Recents



Data



Clusters

3. Crédit cluster Azure Databricks

Création en cours.



Clusters

All-Purpose Clusters

Job Clusters

Pools

Cluster Policies

+ Create Cluster

	Name	State	Nodes	Runtime
	clusterdb	Running	3	7.4 ML (includes Apache Spark 3.0.1, Scala 2.12)



Home



Workspace



Recents



Clusters

3. Crédit cluster Azure Databricks

Le cluster est créé.

The screenshot shows the Databricks interface. On the left, there's a sidebar with icons for Home, Workspace, Recents, Data, Clusters (which is selected and highlighted in red), Jobs, and Models. The main area is titled 'Clusters /' and shows a cluster named 'clusterdb'. Below the cluster name are buttons for Edit, Clone, Restart, Terminate, and Delete. A sub-menu bar includes Configuration, Notebooks, Libraries (which is also highlighted in red), Event Log, Spark UI, Driver Logs, Metrics, and Admin. Under the Libraries tab, there are 'Uninstall' and 'Install New' buttons. A table lists libraries by Name, Type, and Status. A modal window titled 'Install Library' is open, containing fields for 'Library Source' (with options for Upload, DBFS, PyPI, Maven, CRAN, and Workspace, where PyPI is selected), 'Package' (set to 'azureml-sdk[databricks]'), and 'Repository' (set to 'Optional'). At the bottom of the modal are 'Cancel' and 'Install' buttons.

4. Ajout de la librairie Azure ML au cluster

azureml-sdk[databricks]

main ▾

AMLLabs / DatabricksArchive /

Go to file



retkowsky Add files via upload

24dea1e 3 days ago



Labs Azure Databricksdbc

Add files via upload



ReadMe

Create ReadMe

ReadMe

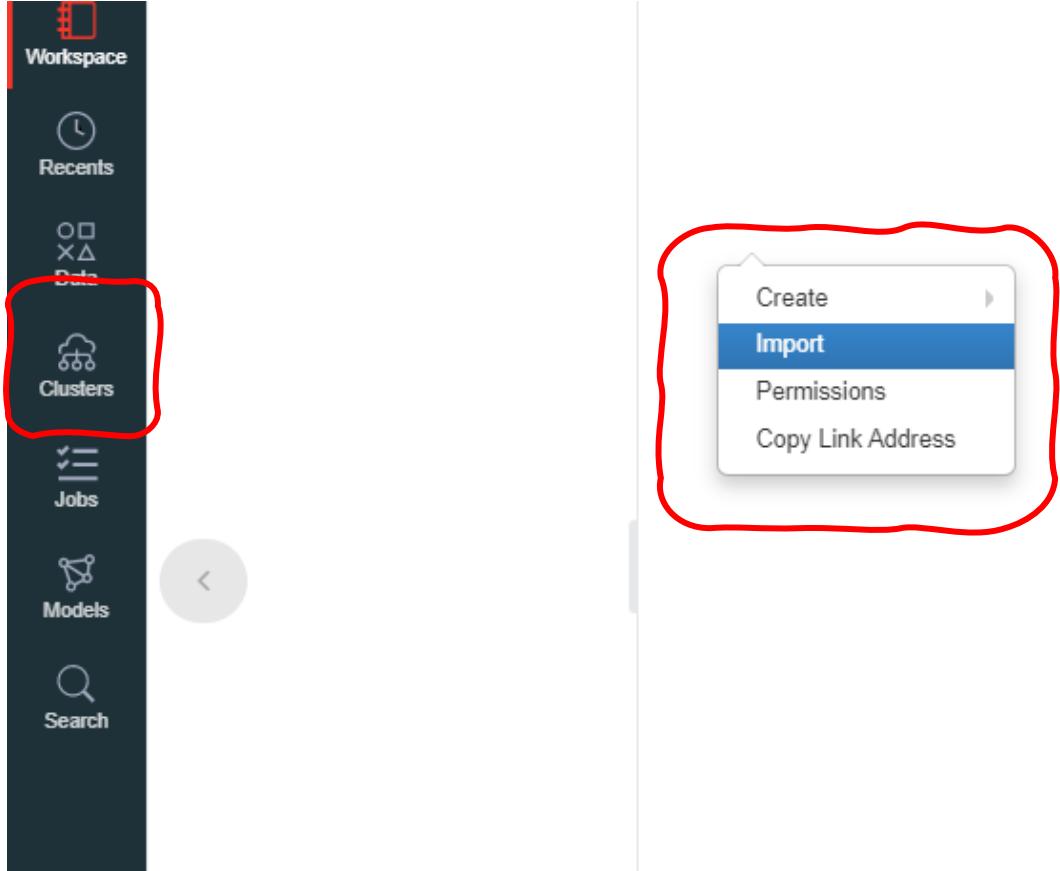
Azure Databricks archive that contains the Databricks notebooks for the labs.

5. Importation des notebooks

Les notebooks du workshop Azure Databricks sont disponibles dans un fichier archive Databricks :

<https://github.com/retkowsky/AMLLabs/blob/main/DatabricksArchive/Labs%20Azure%20Databricksdbc>

Téléchargez le fichier .dbc sur votre poste de travail

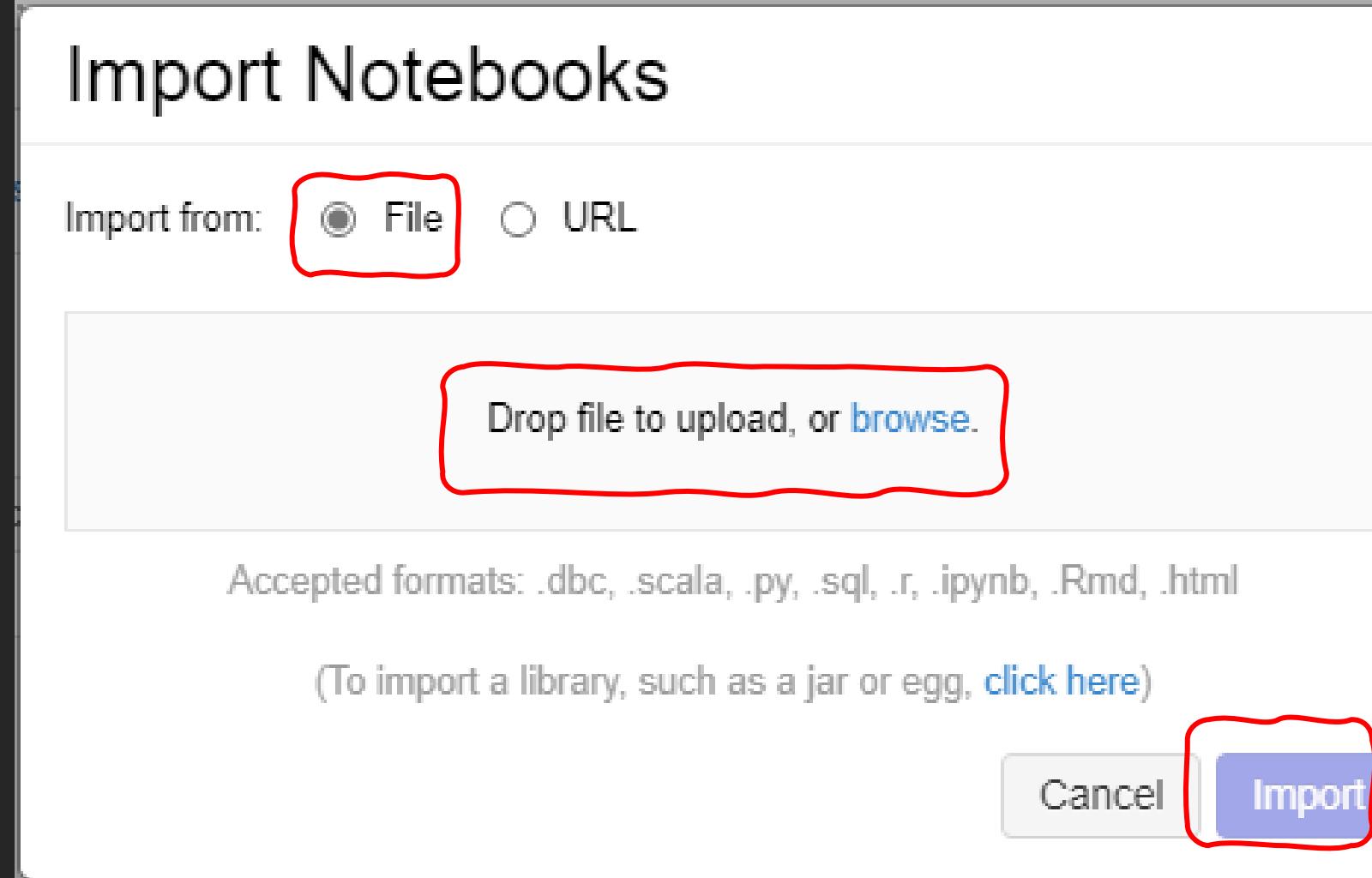


5. Importation des notebooks

Importation du fichier
DBC depuis le menu
Clusters (clic droit)

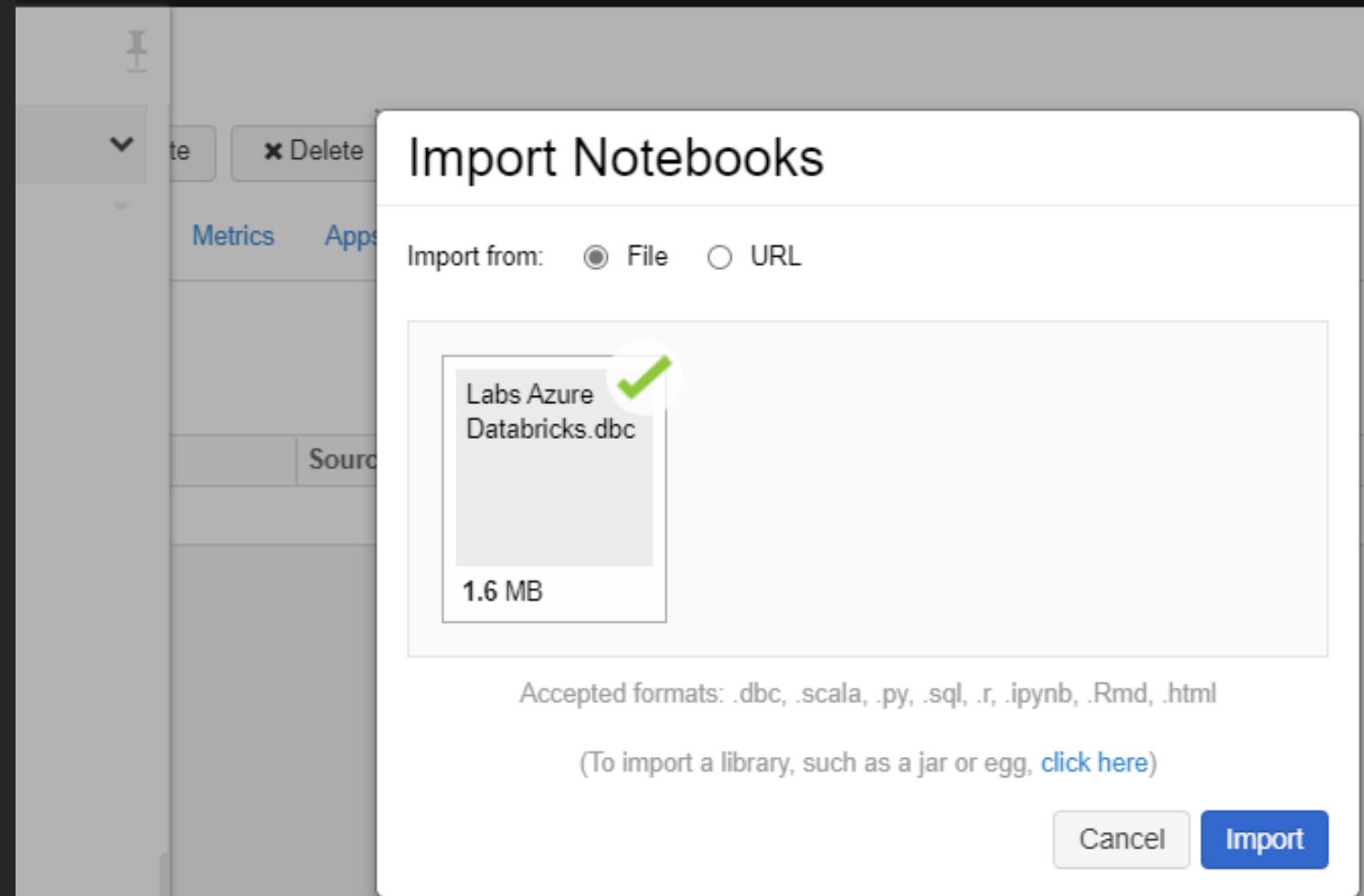
5. Importation des notebooks

Sélectionnez le fichier .dbc téléchargé et cliquez sur import.



5. Importation des notebooks

Fin de l'importation



5. Importation des notebooks

Les notebooks sont disponibles dans votre workspace Azure Databricks.

The screenshot shows the Microsoft Azure Databricks workspace interface. On the left, there is a sidebar with icons for Home, Workspace, Recents, Data, Clusters, Jobs, Models, and Search. The 'Workspace' icon is highlighted with a red border. The main area is titled 'Workspace' and shows a list of items. At the top of the list is 'Trash'. Below it is a folder icon followed by the text 'Labs Azure Databricks', which is also highlighted with a red box. To the right of this, there is a dropdown menu labeled 'Labs Azure Databricks' containing a list of 13 lab notebooks. These notebooks are listed with icons and names: Lab0 Experiments, Lab01 Installation et Configurat..., Lab02 Introduction Apache Spa..., Lab03 Data Exploration, Lab04 Clustering, Lab05 Regression, Lab06 Classification, Lab07 Classification Evaluation..., Lab08 Cross Validation, Lab09 AutoML with Azure Data..., Lab10 Azure Databricks et ML..., Lab11 Azure Databricks et MLF..., Lab12 Azure Databricks et ML..., and Lab13 MLflow Serving Models >. A large red oval highlights the entire list of lab notebooks.

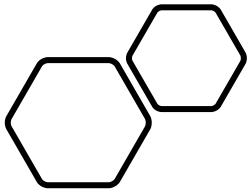
- Labs Azure Databricks
 - Lab0 Experiments
 - Lab01 Installation et Configurat...
 - Lab02 Introduction Apache Spa...
 - Lab03 Data Exploration
 - Lab04 Clustering
 - Lab05 Regression
 - Lab06 Classification
 - Lab07 Classification Evaluation...
 - Lab08 Cross Validation
 - Lab09 AutoML with Azure Data...
 - Lab10 Azure Databricks et ML...
 - Lab11 Azure Databricks et MLF...
 - Lab12 Azure Databricks et ML...
 - Lab13 MLflow Serving Models >

Quiz

Quiz!

[Quiz Link](#)

Call to Action



Call to Action



1. Microsoft Learn

[Browse all - Learn | Microsoft Docs](#)



2. Exemples de notebooks

[nthacker/Azure-Databricks-Labs \(github.com\)](#)

[microsoft/Azure-Databricks-NYC-Taxi-Workshop: An Azure Databricks workshop leveraging the New York Taxi and Limousine Commission Trip Records dataset \(github.com\)](#)



Best Practices

[Azure/AzureDatabricksBestPractices: Version 1 of Technical Best Practices of Azure Databricks based on real world Customer and Technical SME inputs \(github.com\)](#)



Azure