

# Cloud Workshop « La Data Science avec les technologies Microsoft »

## **Exercice Azure Databricks**



# Vos interlocuteurs Microsoft

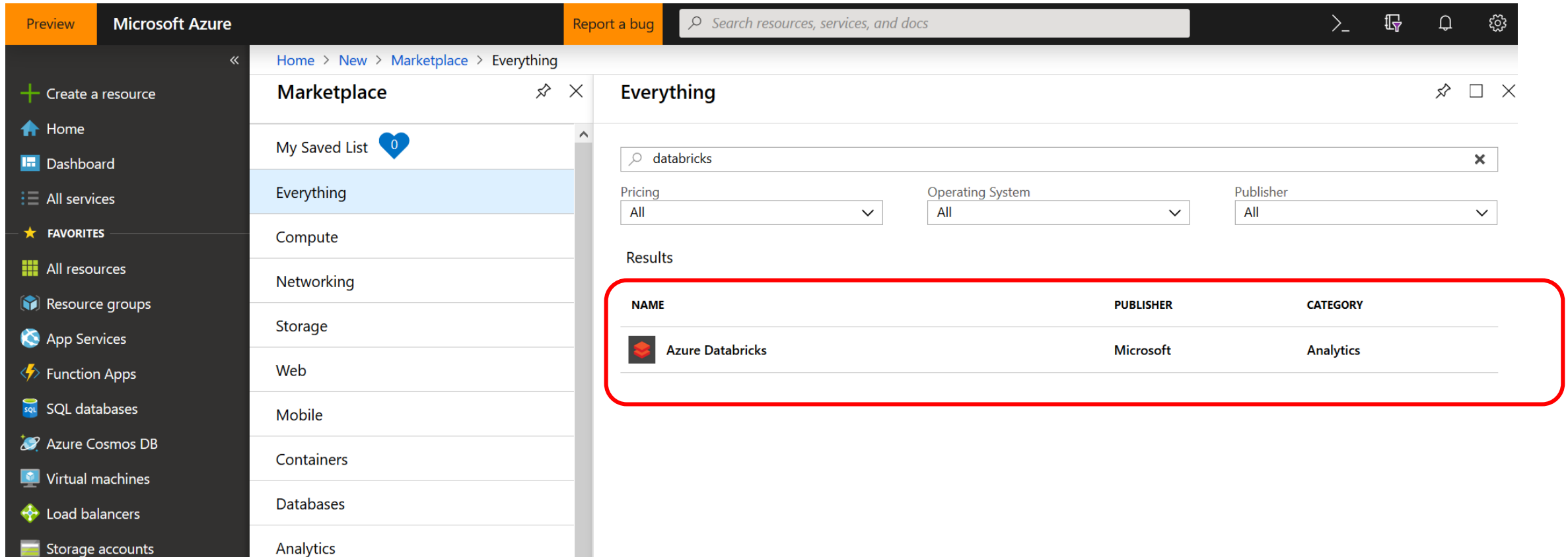
Serge Retkowsky

[serge.retkowsky@microsoft.com](mailto:serge.retkowsky@microsoft.com)


# Agenda Cloud Workshop

Agenda	Contenu
09h00 - 09h30	Accueil et Introduction
09h30 - 11h00	Présentation de la plateforme Azure pour les analyses de Machine Learning
11h00 - 12h30	Hands on Lab « Réalisation et déploiement de modèles prédictifs avec Azure ML Studio »
12h30 - 13h30	Déjeuner
13h30 - 15h00	Hands on lab « Réalisation et déploiement de modèles prédictifs avec le SDK Azure ML Service »
15h00 - 16h30	Hands On Lab « Réalisation et déploiement de modèles prédictifs avec Azure Databricks et Azure ML Services »
16h30 - 17h00	Questions & réponses, conclusion

# 1. Création du service Azure Databricks dans Azure



The screenshot displays the Microsoft Azure portal interface. On the left is a dark sidebar with navigation options: 'Create a resource', 'Home', 'Dashboard', 'All services', and a 'FAVORITES' section containing 'All resources', 'Resource groups', 'App Services', 'Function Apps', 'SQL databases', 'Azure Cosmos DB', 'Virtual machines', 'Load balancers', and 'Storage accounts'. The main area is titled 'Microsoft Azure' and includes a search bar with the text 'Search resources, services, and docs'. Below the search bar, the breadcrumb path 'Home > New > Marketplace > Everything' is visible. The 'Marketplace' section on the left lists categories: 'My Saved List' (0 items), 'Everything' (selected), 'Compute', 'Networking', 'Storage', 'Web', 'Mobile', 'Containers', 'Databases', and 'Analytics'. The 'Everything' view shows search results for 'databricks'. There are filters for 'Pricing' (All), 'Operating System' (All), and 'Publisher' (All). The results table is highlighted with a red border and contains one entry: 'Azure Databricks' by 'Microsoft' in the 'Analytics' category.

NAME	PUBLISHER	CATEGORY
 Azure Databricks	Microsoft	Analytics

# Configuration du service

[Dashboard](#) > [Azure Databricks](#) > Databricks

**Azure Databricks**  
Microsoft

[+ Add](#) [Edit columns](#) [More](#)

**NAME** ↑↓

**Databricks**

**Databricks**  
Azure Databricks Service

[Overview](#)  
[Activity log](#)  
[Access control \(IAM\)](#)  
[Tags](#)

**Settings**  
[Virtual Network Peerings](#)  
[Locks](#)  
[Automation script](#)

**Support + troubleshooting**  
[New support request](#)

[Delete](#)

Resource group [\(change\)](#)  
AzureDB

Subscription [\(change\)](#)  
Microsoft Azure Internal Consumption

Subscription ID

Managed Resource Group

URL  
<https://westeurope.azuredatabricks.net>

Pricing Tier  
premium

[Launch Workspace](#)

Documentation

Getting Started


Import Data from File


Import Data from Azure Storage


Notebook

Admin Guide

## 2. Création du Cluster Spark Databricks

  
Azure  
Databricks




  
Home

  
Workspace

### Clusters

+ Create Cluster

▼ Interactive Clusters

Name	State	Nodes	Driver	Worker	Runtime	Creator		
 Cluster HC 5.2	Running	3	Standard_DS3_v2	Standard_DS3_v2	5.2 (includes Apach...	seretkow@micros...	2	3


## 2. Création du Cluster Spark Databricks


Clusters / Cluster HC 5.2


 Cluster HC 5.2



 Edit

 Clone

 Restart

 Terminate

 Delete

Configuration

Notebooks (2)

Libraries (3)

Event Log

Spark UI

Driver Logs

Spark Cluster UI - Master ▼

Cluster Mode 

Standard


Databricks Runtime Version


5.2 (includes Apache Spark 2.4.0, Scala 2.11)

Python Version 

3

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

Standard\_DS3\_v2








14.0 GB Memory, 4 Cores, 0.75 DBU

► Advanced Options



### 3. Association des librairies AzureML service au cluster Databricks




Microsoft Azure

Clusters / Cluster HC 5.2

**Cluster HC 5.2**    Edit  Clone  Restart  Terminate  Delete

[Configuration](#) [Notebooks \(2\)](#) [Libraries \(3\)](#) [Event Log](#) [Spark UI](#) [Driver Logs](#) [Spark Cluster UI - Master ▼](#)

 Uninstall  Install New

<input type="checkbox"/>	Name	Type	Status	Source
<input type="checkbox"/>	azureml-sdk[automl_databricks]	PyPI	 Installed	
<input type="checkbox"/>	azureml-sdk[databricks]	PyPI	 Installed	
<input type="checkbox"/>	tqdm	PyPI	 Installed	




## 4. Contenu du workshop

retkowsky / [Hands-on-Lab-Azure-ML](#) Unwatch 1 Star 0 Fork 0

[Code](#) [Issues 0](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Insights](#) [Settings](#)

Branch: master [Hands-on-Lab-Azure-ML / 3 Azure Databricks /](#) [Create new file](#) [Upload files](#) [Find file](#) [History](#)

 retkowsky Add files via upload Latest commit f725eb0 10 minutes ago

..

<a href="#">01.Installation_and_Configuration.ipynb</a>	Add files via upload	10 minutes ago
<a href="#">02.Ingest_data.ipynb</a>	Add files via upload	an hour ago
<a href="#">03.Build_model_runHistory.ipynb</a>	Add files via upload	an hour ago
<a href="#">04.Deploy_to_ACI.ipynb</a>	Add files via upload	an hour ago
<a href="#">05.AutoML_Databricks.ipynb</a>	Add files via upload	14 minutes ago

<http://aka.ms/SergeADB>

