

Préparation DP100

30 septembre 2020

SIMPLON
.CO

Entreprise agréée
solidaire et labellisée :



DataDock



FFE
Fondation
la France s'engage



Salut :)



Josselin Tobelem
Formateur

- développeur java, prof de maths
- formateur référent promos dev web
- les maths, les logiciels libres... > geek #<
- depuis 2019, missions data
- passé la DP100 en juin
- et vous?

Groupes Azure, workspaces

jt-dp100-europe-west :

jt-dp100-usa-est :

jt-dp100-usa-west :

jt-dp100-europe-nord :

I - Présentation d'Azure

la plateforme de cloud microsoft

1. portail
2. studio AML
3. sdk

II - Présentation de la certification

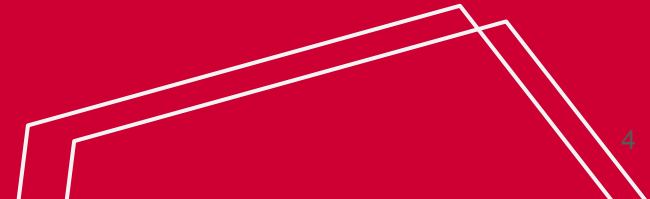
dp100

III - Pratique

1. environnement Azure
2. environnement local
3. retour sur deux notions importantes

IV - Bilan

de cette introduction



Présentation d'Azure et AzureML



Plateforme Azure

- **plateforme de cloud**, comme AWS ou GCP
- cloud computing (application serverless)
- data storage
- services IA (cognitives services, automl, ...)
- devops
- dashboards d'expériences
- ...

Azure ML

Domain specific pretrained models

To simplify solution development



Vision



Speech



Language



Search

Familiar Data Science tools

To simplify model development



Visual Studio Code



Azure Notebooks



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

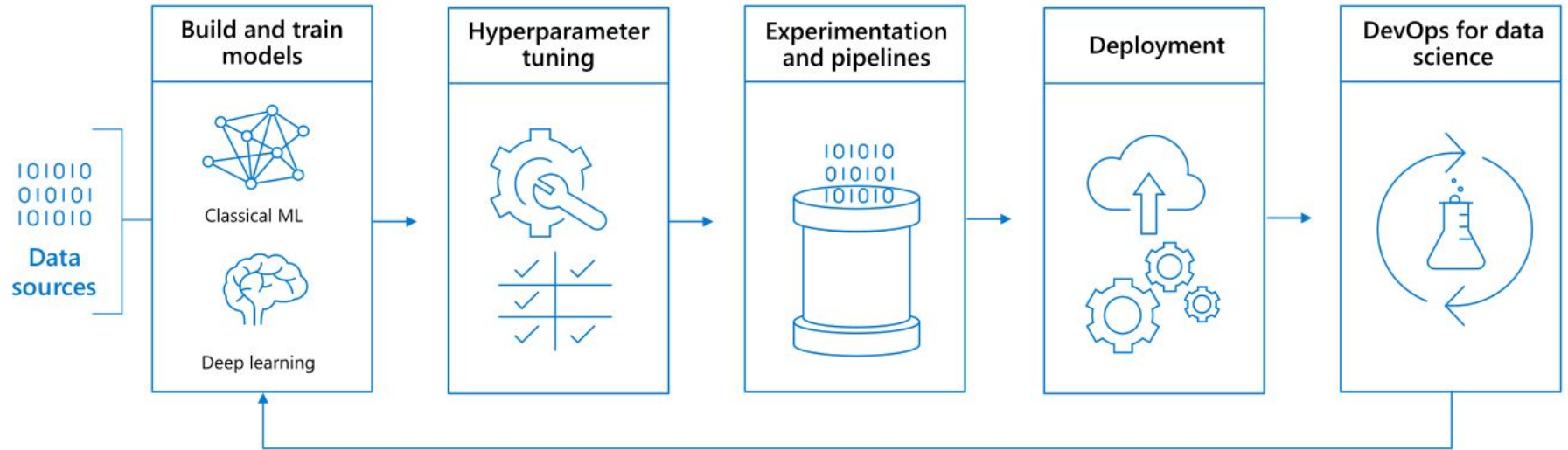


GPU

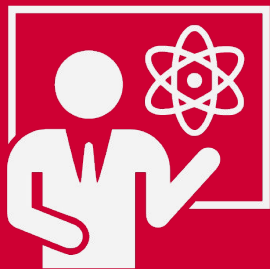


FPGA

Les étapes d'un projet de data science



Notions essentielles



Présentation d'Azure

- 1 Portail**
Gérer les souscriptions, espaces de travail
- 2 Studio**
La partie ML
- 3 Sdk**
Tout faire depuis des notebooks

1. Portail

<https://portal.azure.com/>

- Création de groupe de ressources
- Gestion des espaces de travail
- Gestion de la souscription

Azure services



Create a resource



Virtual networks



Subscriptions



Groups



Cost Management...



All resources



Scheduler Job Collections



Monitor



CloudSimple Virtual...



More services

Recent resources

Name	Type	Last Viewed
Microsoft Azure Sponsorship (9114a63e-9210-4e32-97ca-b7d9e8ac403d)	Subscription	2 weeks ago
jt-dp100	Machine Learning	3 weeks ago
emotion-resource	Cognitive Services	4 months ago
jt-dp100-resources	Resource group	4 months ago
cognitive-service-resources	Resource group	4 months ago
aibs-vm	Virtual machine	4 months ago
AI School Azure	Subscription	4 months ago

Navigate



Subscriptions



Resource groups



All resources



Dashboard

Tools



Microsoft Learn

Learn Azure with free online training from Microsoft



Azure Monitor

Monitor your apps and infrastructure



Security Center

Secure your apps and infrastructure



Cost Management

Analyze and optimize your cloud spend for free

2. Studio


<https://ml.azure.com/>


- Regroupe les services liés au ML
- Notebooks
- Instances de calcul
- Datasets, datastores
- Modèles
- ...


Microsoft Azure Machine Learning


jt-dp100 > Accueil

Azure Machine Learning Studio


Créer nouveau ▾


Notebooks
Programmez avec le SDK Python et exécutez des exemples d'expériences.
Démarrer maintenant

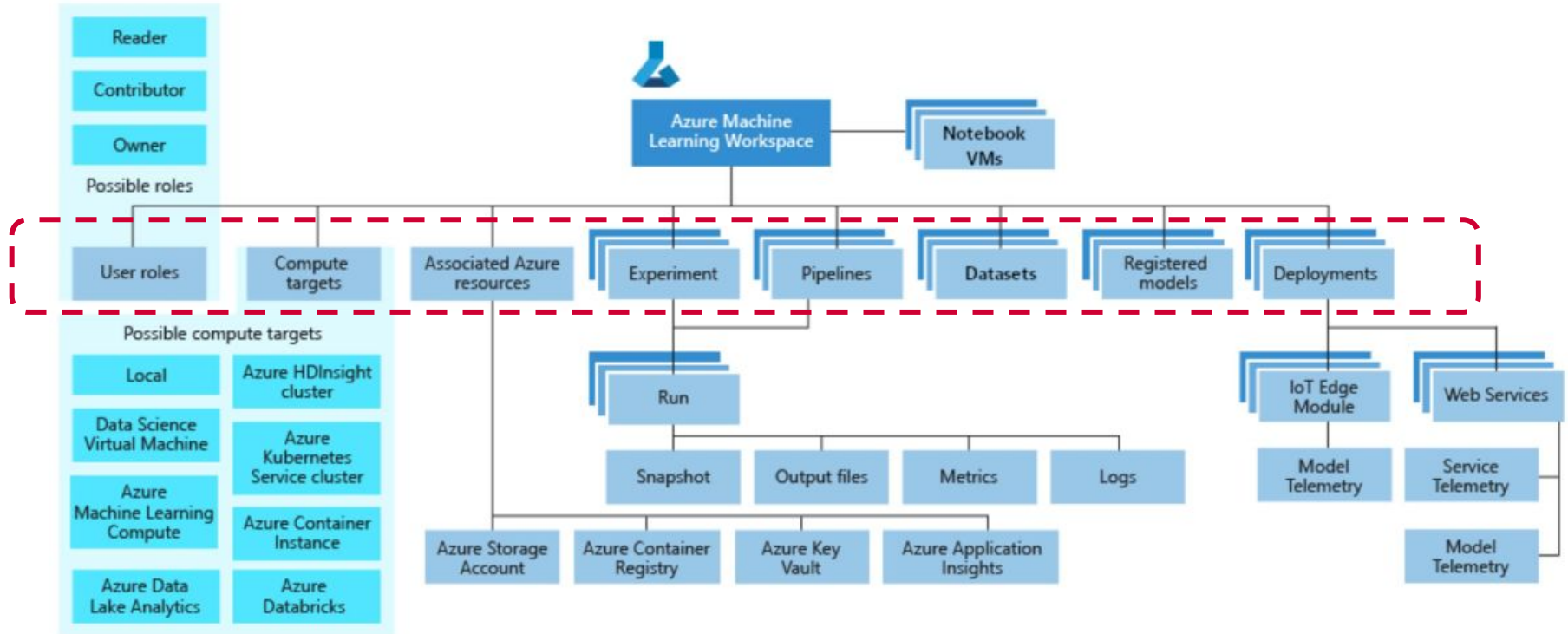

ML automatisé
Entraînez et réglez automatiquement un modèle à l'aide d'une métrique cible.
Démarrer maintenant


Concepteur
Interface de glisser-déposer à la fois pour préparer les données et déployer des modèles.
Démarrer maintenant

Mes ressources récentes

Exécutions							Calcul
Exécu...	ID de série	Expérience	État	Heure d'envoi	Envoyé par	Type de sé...	Nom
Exécut...	titanic-training-exp...	titanic-tra...	Terminé	28 sept. 2020 13:56	Lina Fahsi	Script	aml-cluster
Exécut...	AutoML_f57f610b-7...	titanic_au...	Terminé	28 sept. 2020 11:37	Lina Fahsi	ML automa...	vm-ds3-v2
Exécut...	AutoML_80221d23-...	titanic_au...	Terminé	28 sept. 2020 09:39	Lina Fahsi	ML automa...	
Exécut...	AutoML_baa68c85-...	titanic_au...	Terminé	18 sept. 2020 15:43	Lina Fahsi	ML automa...	
Exécut...	0ea4d434-d499-45...	regressio...	Terminé	16 sept. 2020 17:27	Lina Fahsi	Pipeline	
Exécut...	AutoML_534f109f-3...	titanic_au...	Terminé	16 sept. 2020 17:19	Lina Fahsi	ML automa...	

Les composants AzureML dans le studio



3. SDKs

<https://docs.microsoft.com/en-us/azure/developer/python/azure-sdk-overview>

- Permet d'utiliser toutes les fonctions d'azure (donc du portail) depuis python
- exemple : [créer un workspace depuis la sdk](#)
- [l'api complète d'azure](#)

- installer la SDK aml (les fonctions du studio) :
<https://docs.microsoft.com/fr-fr/python/api/overview/azure/ml/install?view=azure-ml-py>



pour installer toutes
les dépendances :
environnement
virtuel ou [conda](#)

Présentation de la certification



Exam DP-100
Designing and Implementing a Data Science
Solution on Azure

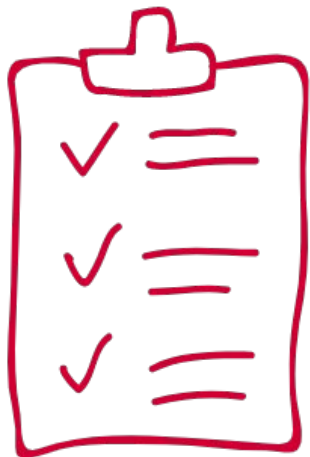
Certification DP100

[Page de la certification](#)

[Contenu détaillé de l'examen](#)

- Le contenu de cet examen a été mis à jour 22 mai 2020.
- Mettre en place un espace de travail Azure Machine Learning (30-35%)
- Exécuter des expériences et former les modèles (25-30%)
- Optimiser et gérer les modèles (20-25%)
- Déployer et consommer les modèles (20-25%)

Programme de la préparation



3 jours en présentiel

introduction à Azure, sdk (experiments, models, data, remote compute, automl)

- mercredi 30 septembre 2020
- jeudi 1er octobre
- vendredi 2 octobre

2 jours en distanciel

pipeline, deploy + modalités exam

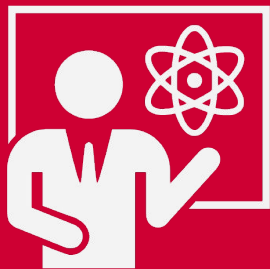
- lundi 23 novembre 2020
- jeudi 26 novembre 2020

Ressources

- [Parcours d'apprentissage microsoft : Créer des solutions d'IA avec AML](#)
(tout ce qu'il faut pour préparer la dp100)
- [Parcours d'apprentissage microsoft : Créer des modèles Machine Learning](#)
(des notebooks pour revoir la base sur du machine learning sans azure + librairies pytorch, tf, etc ...)
- [Repository microsoft](#)
- [API azure](#)
- [Getting Started with AML](#)
- [Mon repo prepa-dp100](#)






Notions essentielles



Partie pratique



-  **1** **Prise en main de la plateforme**
Portail et studio + exercices du parcours
-  **2** **Installer son environnement en local**
Configuration d'azureML
-  **3** **Focus sur deux notions clés**
Datastores/datasets et compute targets

Prise en main de la plateforme

mon repo :

<https://github.com/jtobelem-simplon/dp100-brief-init>

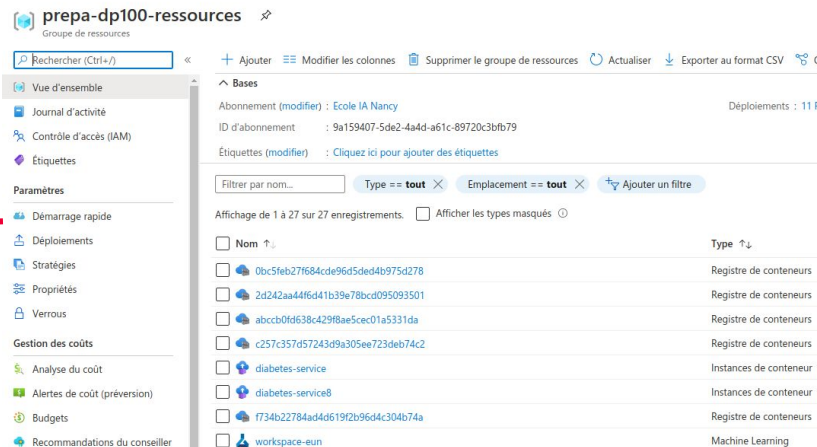
1

Prise en main de la plateforme

<https://portal.azure.com/>

➡ Allez sur le groupe de ressource :
jt-dp100-resources

➡ Sélectionnez l'espace de travail :
jt-dp100
(cela ouvre le studio)

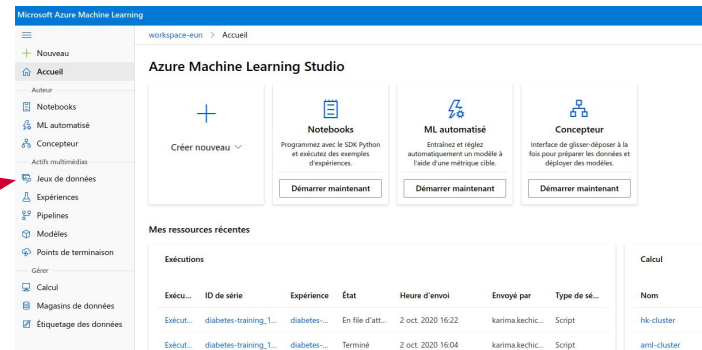


Azure Machine Learning Studio

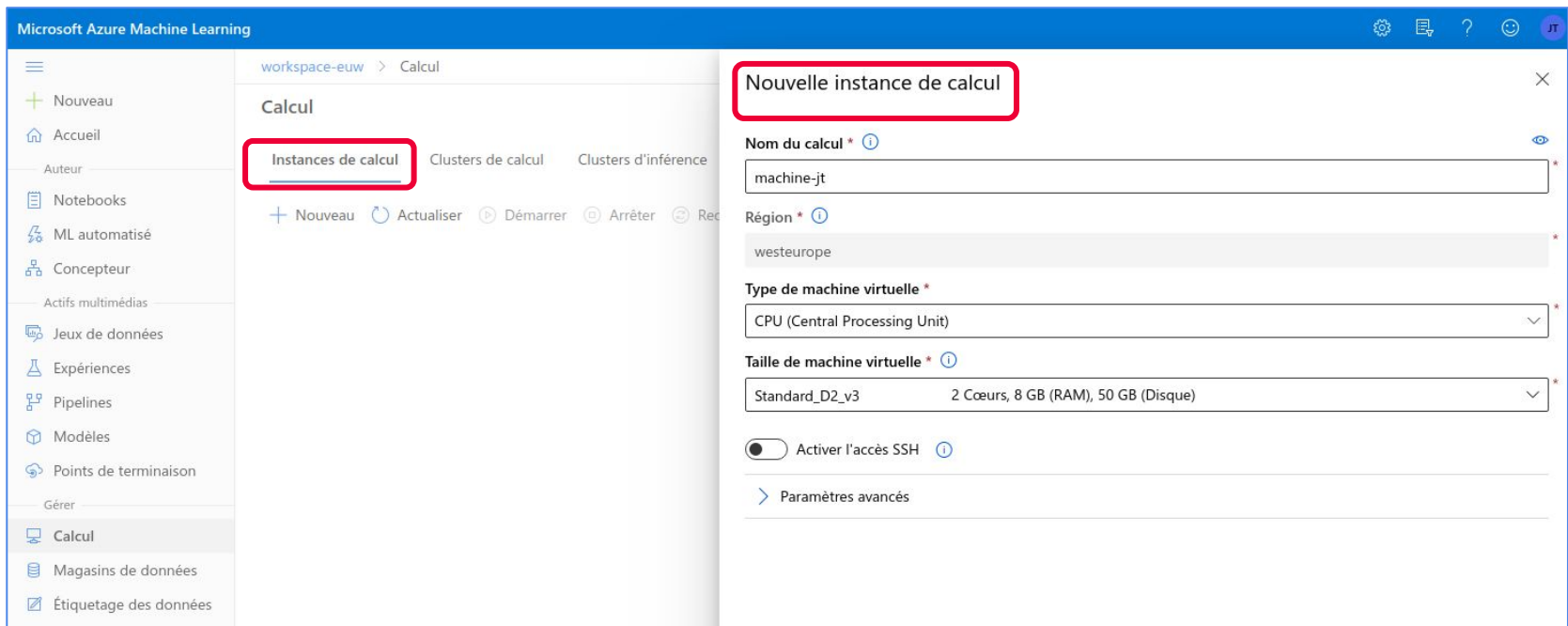
Expérience immersive pour gérer de bout en bout le cycle de vie du Machine Learning.

Lancer maintenant En savoir plus

Localisation : choisissez
europe occidentale (et pas
france)



Création d'une instance de calcul



Créez une machine de type ~~standard d2 v3~~ standard ds11 v2:
machine-xx (remplacez xx par vos initiales)

La vm a été créée avec succès

Microsoft Azure Machine Learning

workspace-euw > Calcul

Calcul

Instances de calcul Clusters de calcul Clusters d'inférence Calcul attaché

+ Nouveau Actualiser Démarrer Arrêter Redémarrer Supprimer Afficher le quota ... Rechercher pour filtrer les éléments...

Nom	État	URI d'application	Taille de machine virtu...	Créé le ↓
machine-jt	En cours d'exécution	JupyterLab Jupyter RStudio SSH	STANDARD_D2_V3	6 oct. 2020 12:12

Navigation:

- Nouveau
- Accueil
- Auteur
 - Notebooks
 - ML automatisé
 - Concepteur
- Actifs multimédias
 - Jeux de données
 - Expériences
 - Pipelines
 - Modèles
 - Points de terminaison
- Gérer
 - Calcul**
 - Magasins de données
 - Étiquetage des données

On accède à jupyter notebook



Quit

Files

Running

Clusters

Conda

AzureML Samples

Select items to perform actions on them.

Upload

New ▾



0 ▾ /

Name ▾

Last Modified

File size

Users

il y a 5 jours

Accès au terminal



Quit

Files

Running

Clusters

Conda

AzureML Samples

Select items to perform actions on them.

Upload

New ▾



0 ▾

/

Name



Users

Notebook:

Python 3

Python 3.6 - AzureML

R

Other:

Text File

Folder

Terminal


Clone du git qui contient les notebooks mslearn



```
azureuser@machine-jt:/mnt/batch/tasks/shared/LS_root/mounts/clusters/machine-jt/code$ ls
Users
azureuser@machine-jt:/mnt/batch/tasks/shared/LS_root/mounts/clusters/machine-jt/code$ cd Users/jtobelem/
azureuser@machine-jt:/mnt/batch/tasks/shared/LS_root/mounts/clusters/machine-jt/code/Users/jtobelem$ git clone https://github.com/MicrosoftDocs/mslearn-aml-labs.git
```

➔ Allez dans votre dossier avec `cd` et clonez le repo des notebooks :
`git clone https://github.com/MicrosoftDocs/mslearn-aml-labs.git`

Liste des notebooks du repo

 jupyter

Quit

FilesRunningClustersCondaAzureML Samples

Select items to perform actions on them.

UploadNew ↻

0 ▾

Users / jtobelem / mslearn-aml-labs

Name ▾Last ModifiedFile size

..

il y a quelques secondes

☐

data

il y a 4 jours

☐

diabetes-training


il y a 4 jours

☐

labdocs

il y a 4 jours


☐

 01-Getting_Started_with_Azure_ML.ipynb

il y a 4 jours

21.9 kB


☐

 02-Training_Models.ipynb

il y a 4 jours

16.5 kB


☐

 03-Working_with_Data.ipynb

il y a 4 jours

28.5 kB


☐

 04-Working_with_Compute.ipynb

il y a 4 jours

17.9 kB


☐

 05-Creating_a_Pipeline.ipynb

il y a 4 jours

23 kB


☐

 06-Deploying_a_model.ipynb

il y a 4 jours

17.7 kB


☐

 07-Creating_a_Batch_Inferencing_Service.ipynb

il y a 4 jours

20.9 kB


☐

 08-Tuning_Hyperparameters.ipynb

il y a 4 jours

15.1 kB


☐

 09-Automated_ML.ipynb

il y a 4 jours

11.3 kB


☐

 10-Interpreting_Models.ipynb

il y a 4 jours

17 kB


☐

 11-FairLearn.ipynb

il y a 4 jours

22.6 kB


☐

 12-Monitoring_a_Model.ipynb

il y a 4 jours

14.8 kB


☐

 13-Monitoring_Data_Drift.ipynb

il y a 4 jours

14.5 kB


☐

 README.md

il y a 4 jours

2.55 kB


☐

 SECURITY.md

il y a 4 jours

2.87 kB


☐

 LICENSE

il y a 4 jours

1.16 kB

A vous de jouer

 jupyter Quit

Files Running Clusters Conda AzureML Samples

Select items to perform actions on them. Upload New ↻

<input type="checkbox"/>	Name	Last Modified	File size
<input type="checkbox"/>	..	il y a quelques secondes	
<input type="checkbox"/>	data	il y a 4 jours	
<input type="checkbox"/>	diabetes-training	il y a 4 jours	
<input type="checkbox"/>	labdocs	il y a 4 jours	
<input checked="" type="checkbox"/>	01-Getting_Started_with_Azure_ML.ipynb	il y a 4 jours	21.9 kB
<input type="checkbox"/>	02-Training_Models.ipynb	il y a 4 jours	16.5 kB
<input type="checkbox"/>	03-Working_with_Data.ipynb	il y a 4 jours	28.5 kB
<input type="checkbox"/>	04-Working_with_Compute.ipynb	il y a 4 jours	17.9 kB
<input type="checkbox"/>	05-Creating_a_Pipeline.ipynb	il y a 4 jours	23 kB
<input type="checkbox"/>	06-Deploying_a_model.ipynb	il y a 4 jours	17.7 kB
<input type="checkbox"/>	07-Creating_a_Batch_Inferencing_Service.ipynb	il y a 4 jours	20.9 kB
<input type="checkbox"/>	08-Tuning_Hyperparameters.ipynb	il y a 4 jours	15.1 kB
<input type="checkbox"/>	09-Automated_ML.ipynb	il y a 4 jours	11.3 kB
<input type="checkbox"/>	10-Interpreting_Models.ipynb	il y a 4 jours	17 kB
<input type="checkbox"/>	11-FairLearn.ipynb	il y a 4 jours	22.6 kB
<input type="checkbox"/>	12-Monitoring_a_Model.ipynb	il y a 4 jours	14.8 kB
<input type="checkbox"/>	13-Monitoring_Data_Drift.ipynb	il y a 4 jours	14.5 kB
<input type="checkbox"/>	README.md	il y a 4 jours	2.55 kB
<input type="checkbox"/>	SECURITY.md	il y a 4 jours	2.87 kB
<input type="checkbox"/>	LICENSE	il y a 4 jours	1.16 kB



Présentation d'Azure Machine Learning

42 min • Module • 6 Unités

★★★★★ 4.6 (2 135)

Attribuer une note

Débutant Data Scientist Étudiant Azure Machine Learning service

Présentation d'Azure Machine Learning

✓ 700 XP

Introduction

2 min



Espaces de travail Azure Machine Learning

5 min



Outils et interfaces Azure Machine Learning

5 min



Expériences Azure Machine Learning

5 min



Exercice - Bien démarrer avec Azure Machine Learning

20 min



Récapitulatif

5 min



Suivez les modules et exécutez les notebooks correspondants :

<https://docs.microsoft.com/fr-fr/learn/paths/build-ai-solutions-with-azure-ml-service/>

Installation de l'environnement Azure en local

mon repo :

<https://github.com/jtobelem-simplon/dp100-brief-init-expert>

2

Prérequis

- Python > 3.6
- [Conda](#) (+ prompt pour windows), anaconda ou miniconda (miniconda suffit)
- Git ([git pour windows](#))
- [vscode](#)

Création d'un environnement vide conda

Toutes les commandes suivantes se font dans le terminal linux ou dans le prompt conda

```
(base) lab@lab:~$ conda -V
conda 4.8.5
(base) lab@lab:~$ conda update conda
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

(base) lab@lab:~$
```

➡ Vérifiez la version de conda et updatez conda

```
(base) lab@lab:~$ conda create -n azure python=3.7 anaconda
```

➡ Créez un environnement appelé "azure"

Clonage du repo du cours et du repo microsoft

```
(base) lab@lab:~$ git clone --recurse-submodules https://github.com/jtobelem-simplon/prepa-dp100.git
```

➡ git clone
--recurse-submodules
https://github.com/jtobelem-simplon/prepa-dp100.git

Mise à jour de l'environnement avec le repo

```
(base) lab@lab:~$ cd prepa-dp100/env/  
(base) lab@lab:~/prepa-dp100/env$ conda activate azure  
(azure) lab@lab:~/prepa-dp100/env$ conda env update --file environment.yml
```

➡ conda activate azure

➡ conda env update --file
environment.yml

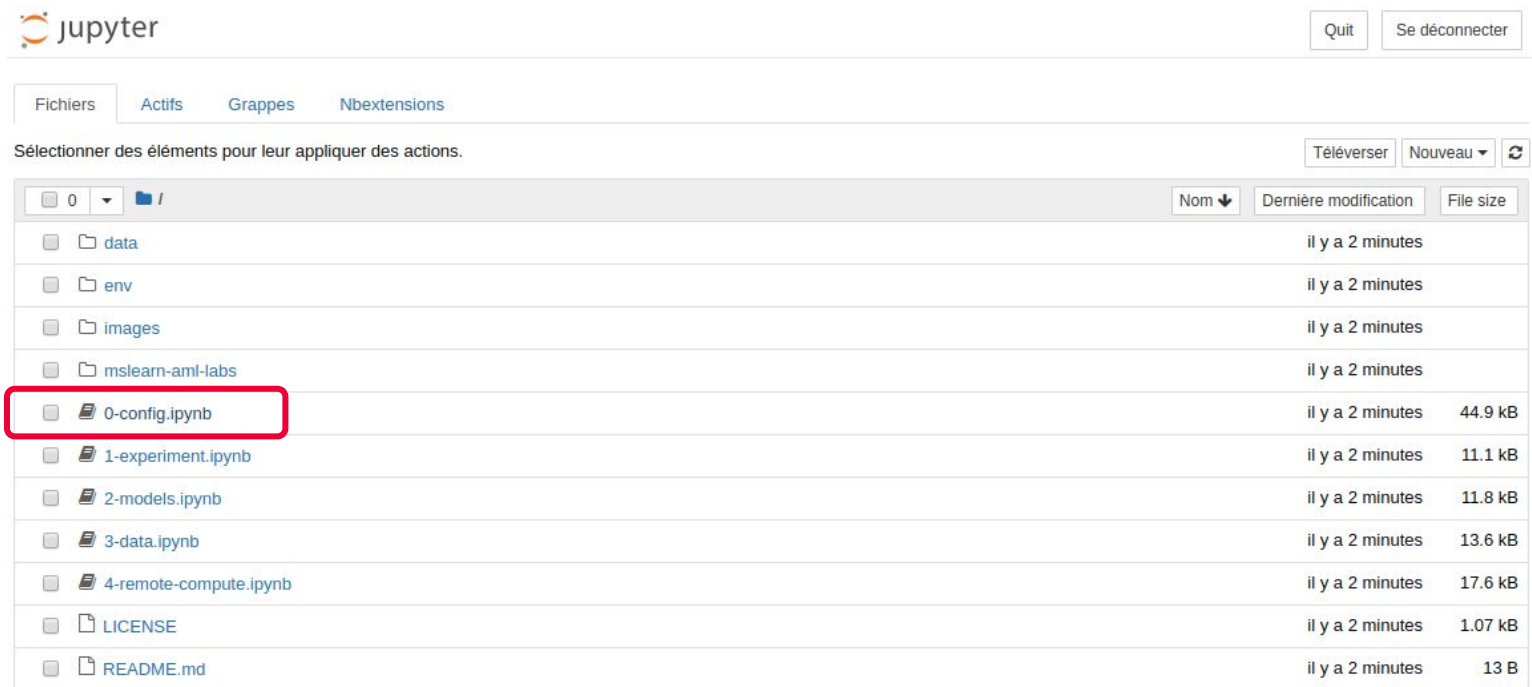
Jupyter notebook

```
(base) Lab@Lab:~$ git clone --recurse-submodules https://github.com/jtobelem-sim  
plon/prepa-dp100.git  
Clonage dans 'prepa-dp100'...  
remote: Enumerating objects: 191, done.  
remote: Counting objects: 100% (191/191), done.  
remote: Compressing objects: 100% (137/137), done.  
remote: Total 191 (delta 90), reused 137 (delta 39), pack-reused 0  
Réception d'objets: 100% (191/191), 1.24 MiB | 460.00 KiB/s, fait.  
Résolution des deltas: 100% (90/90), fait.  
Sous-module 'mslearn-aml-labs' (https://github.com/MicrosoftDocs/mslearn-aml-lab  
s.git) enregistré pour le chemin 'mslearn-aml-labs'  
Clonage dans '/home/lab/prepa-dp100/mslearn-aml-labs'...  
remote: Enumerating objects: 14, done.  
remote: Counting objects: 100% (14/14), done.  
remote: Compressing objects: 100% (12/12), done.  
remote: Total 292 (delta 3), reused 5 (delta 2), pack-reused 278  
Réception d'objets: 100% (292/292), 491.70 KiB | 294.00 KiB/s, fait.  
Résolution des deltas: 100% (185/185), fait.  
Chemin de sous-module 'mslearn-aml-labs' : 'c4671ca81f8cfb1cf1415a43c9d3ee0d506d  
7231' extrait  
(base) Lab@Lab:~$ conda activate azure  
(azure) Lab@Lab:~$ cd prepa-dp100/  
(azure) Lab@Lab:~/prepa-dp100$ jupyter notebook
```



après avoir fait tout ça,
vous pouvez lancer le
server jupyter

Test de l'environnement

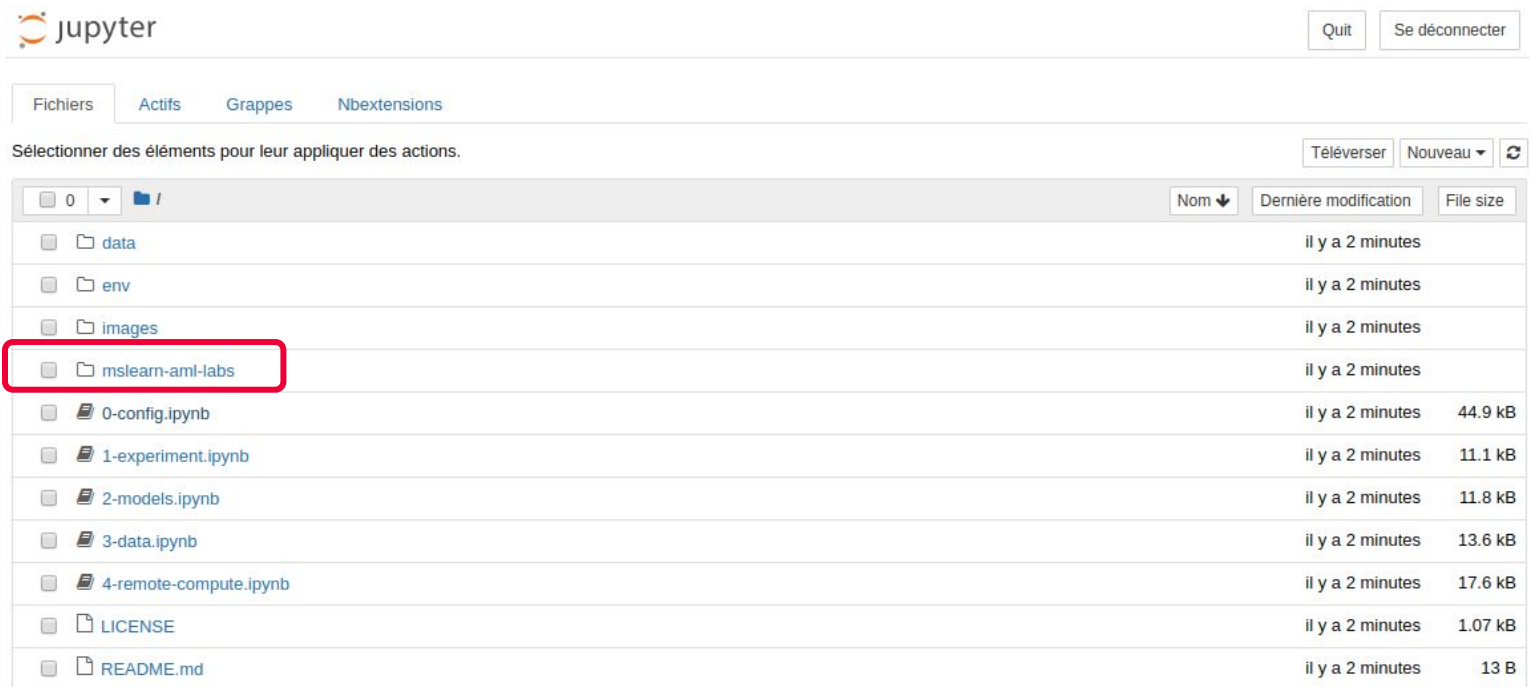


The image shows the JupyterLab interface. At the top left is the Jupyter logo. To the right are buttons for 'Quit' and 'Se déconnecter'. Below the logo is a tab bar with 'Fichiers', 'Actifs', 'Grappes', and 'Nbextensions'. The 'Fichiers' tab is active. Below the tab bar is a text prompt: 'Sélectionner des éléments pour leur appliquer des actions.' To the right of this prompt are buttons for 'Téléverser', 'Nouveau', and a refresh icon. Below this is a file browser table. The table has columns for 'Nom', 'Dernière modification', and 'File size'. The first row is a header row. The subsequent rows list files and folders. The file '0-config.ipynb' is highlighted with a red rectangular box. Other files include '1-experiment.ipynb', '2-models.ipynb', '3-data.ipynb', '4-remote-compute.ipynb', 'LICENSE', and 'README.md'. Folders include 'data', 'env', 'images', and 'mslearn-aml-labs'.

	Nom	Dernière modification	File size
<input type="checkbox"/>	0		
<input type="checkbox"/>	data	il y a 2 minutes	
<input type="checkbox"/>	env	il y a 2 minutes	
<input type="checkbox"/>	images	il y a 2 minutes	
<input type="checkbox"/>	mslearn-aml-labs	il y a 2 minutes	
<input type="checkbox"/>	0-config.ipynb	il y a 2 minutes	44.9 kB
<input type="checkbox"/>	1-experiment.ipynb	il y a 2 minutes	11.1 kB
<input type="checkbox"/>	2-models.ipynb	il y a 2 minutes	11.8 kB
<input type="checkbox"/>	3-data.ipynb	il y a 2 minutes	13.6 kB
<input type="checkbox"/>	4-remote-compute.ipynb	il y a 2 minutes	17.6 kB
<input type="checkbox"/>	LICENSE	il y a 2 minutes	1.07 kB
<input type="checkbox"/>	README.md	il y a 2 minutes	13 B

➡ Ouvrez le notebook 0-config pour vérifier votre environnement

MSLearn labs en local



The image shows the JupyterLab web interface. At the top left is the Jupyter logo. At the top right are buttons for 'Quit' and 'Se déconnecter'. Below the logo are tabs for 'Fichiers', 'Actifs', 'Grappes', and 'Nbextensions'. A message says 'Sélectionner des éléments pour leur appliquer des actions.' To the right of this message are buttons for 'Téléverser', 'Nouveau', and a refresh icon. The main area is a file browser showing a directory structure. The 'mslearn-aml-labs' folder is highlighted with a red rectangle. Below the file browser is a table with columns for 'Nom', 'Dernière modification', and 'File size'.

	Nom	Dernière modification	File size
<input type="checkbox"/>	data	il y a 2 minutes	
<input type="checkbox"/>	env	il y a 2 minutes	
<input type="checkbox"/>	images	il y a 2 minutes	
<input type="checkbox"/>	mslearn-aml-labs	il y a 2 minutes	
<input type="checkbox"/>	0-config.ipynb	il y a 2 minutes	44.9 kB
<input type="checkbox"/>	1-experiment.ipynb	il y a 2 minutes	11.1 kB
<input type="checkbox"/>	2-models.ipynb	il y a 2 minutes	11.8 kB
<input type="checkbox"/>	3-data.ipynb	il y a 2 minutes	13.6 kB
<input type="checkbox"/>	4-remote-compute.ipynb	il y a 2 minutes	17.6 kB
<input type="checkbox"/>	LICENSE	il y a 2 minutes	1.07 kB
<input type="checkbox"/>	README.md	il y a 2 minutes	13 B



Vous pouvez maintenant tester le repo MSLearn en local

Problèmes rencontrés

Voici une liste des problèmes (corrigés ou pas) rencontrés :

- dépendance mlflow
- installation de docker sous windows (lors du run d'un estimator)
- 04-Working_with_Compute :
le cluster-name ne peut dépasser 16 caracteres, utilisez le nom 'aml-cluster'

Focus sur deux notions clés

3

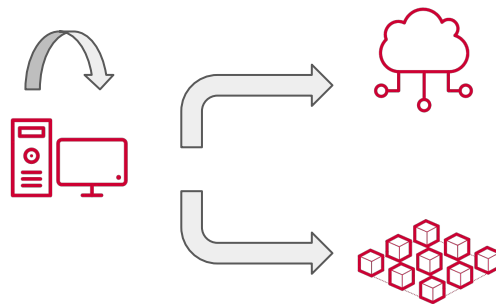
Datastores/dataset

- Magasins de données : [datastores](#)
 - [dateref](#)
 - [datasets](#)
 - tabular dataset
 - file data set

Compute target

Pour exécuter un notebook, il faut une machine :

- [local](#)
- [ComputeInstance](#)



A l'intérieur d'un notebook, on peut réaliser des expériences (par exemple fit un modèle) que l'on va faire tourner sur une computeTarget :

- [local](#)
- [ComputeInstance](#)
- [AmlCompute](#)

Bilan

Retour sur le programme

01-Getting_Started_with_Azure_ML.ipynb

02-Training_Models.ipynb

03-Working_with_Data.ipynb

04-Working_with_Compute.ipynb

05-Creating_a_Pipeline.ipynb

06-Deploying_a_model.ipynb

07-Creating_a_Batch_Inferencing_Serv...

08-Tuning_Hyperparameters.ipynb

09-Automated_ML.ipynb

10-Interpreting_Models.ipynb

11-FairLearn.ipynb

12-Monitoring_a_Model.ipynb

13-Monitoring_Data_Drift.ipynb

Déjà abordé :

- configuration, compute, datasets/stores
- automl

Reste à voir :

- pipeline, deploy
- interprétation des modèles