

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 : Ambre, Arnaud, Céline François

jt-dp100-usa-est : Francis, Guole, Nohossat, Oorvasi

jt-dp100-usa-west : Pierre-Marie, Rodrigo, Salim, Shadi

jt-dp100-europe-nord : Valérie

I - Présentation d'Azure

la plateforme de cloud microsoft

1. portail
2. studio AML
3. sdk

II - Présentation de la certification

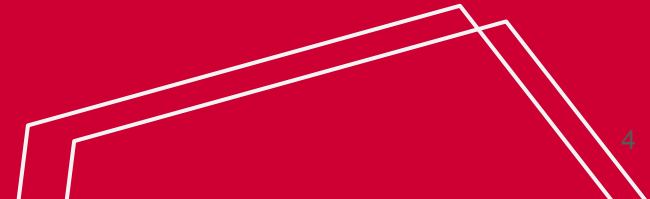
dp100

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

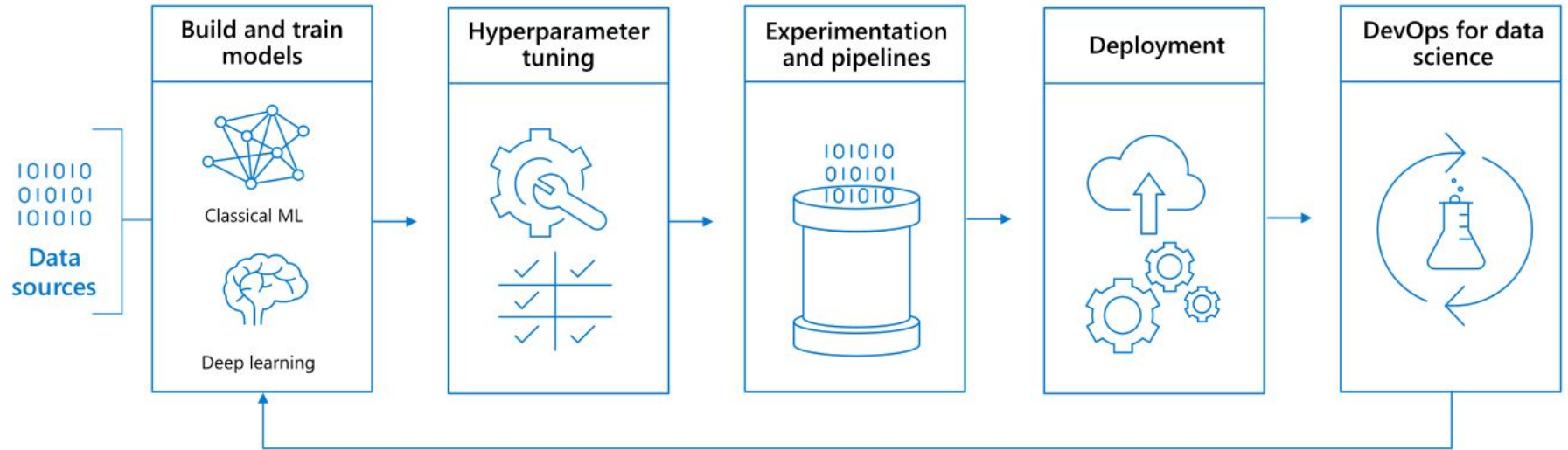


GPU

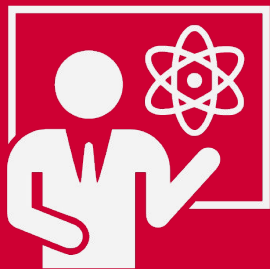


FPGA

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

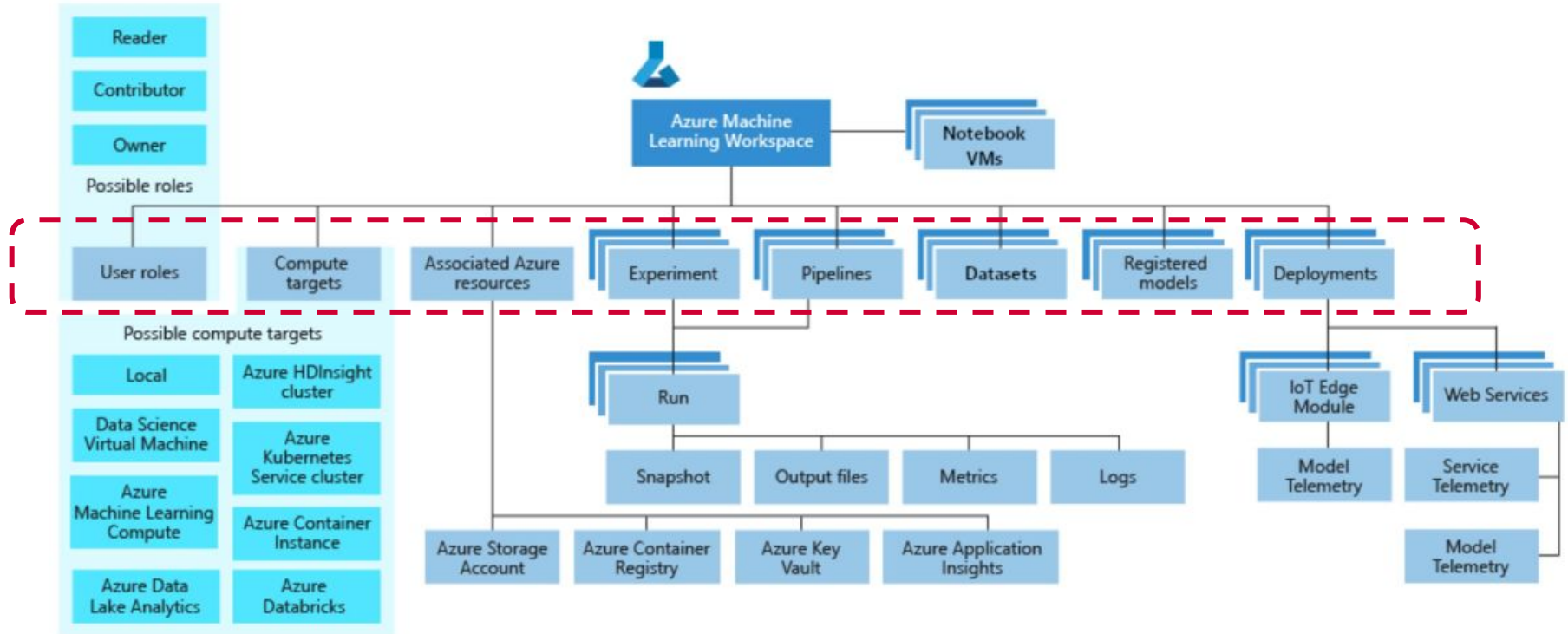
Exécutions						
Exécu...	ID de série	Expérience	État	Heure d'envoi	Envoyé par	Type de sé...
Exécut...	titanic-training-exp...	titanic-tra...	Terminé	28 sept. 2020 13:56	Lina Fahsi	Script
Exécut...	AutoML_f57f610b-7...	titanic_au...	Terminé	28 sept. 2020 11:37	Lina Fahsi	ML automa...
Exécut...	AutoML_80221d23-...	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...

Calcul

Nom

- aml-cluster
- vm-ds3-v2

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

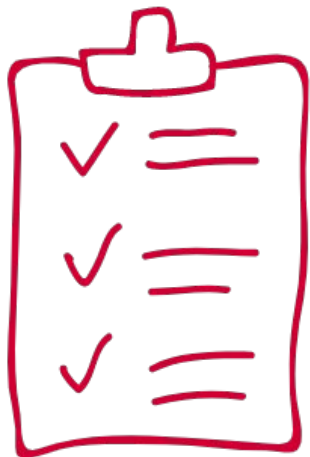
Certification DP100

[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

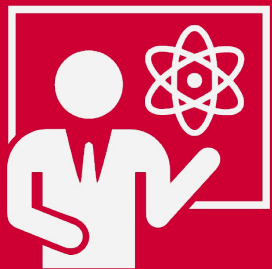
- lundi 23 novembre 2020
- jeudi 26 novembre 2020

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>

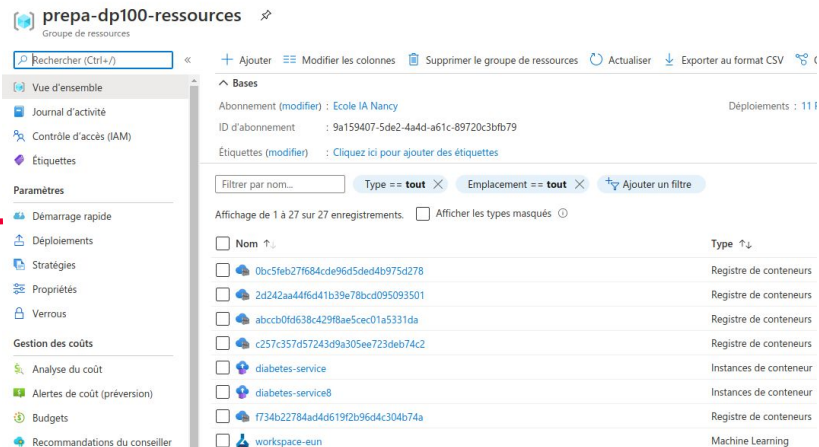
1

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)

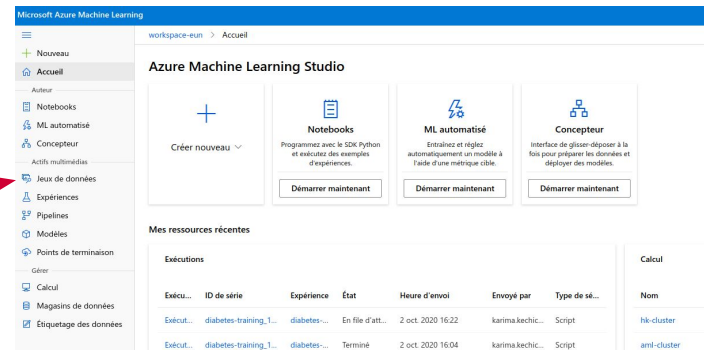


Azure Machine Learning Studio

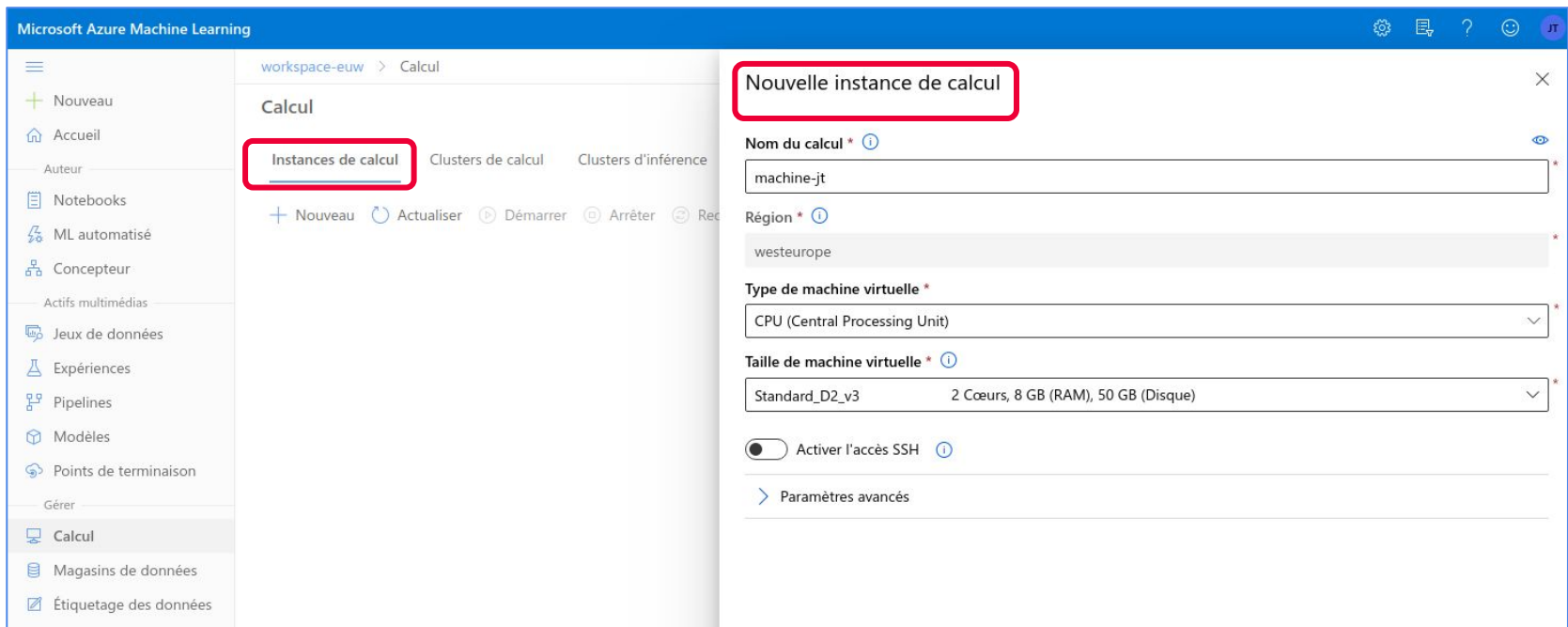
Expérience immersive pour gérer de bout en bout le cycle de vie du Machine Learning.

[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, Auteurs, 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

0



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

Files Running Clusters Conda AzureML Samples

Select items to perform actions on them.

Upload New

	Name	Last Modified	File size
0	Users / jtobelem / mslearn-aml-labs		
..	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		



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

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 https://github.com/jtobelem-simplon/dp100-brief-init-expert
Clonage dans 'dp100-brief-init-expert'...
remote: Enumerating objects: 24, done.
remote: Counting objects: 100% (24/24), done.
remote: Compressing objects: 100% (19/19), done.
remote: Total 24 (delta 8), reused 20 (delta 4), pack-reused 0
Dépaquetage des objets: 100% (24/24), fait.
```

➡ `git clone https://github.com/jtobelem-simplon/dp100-brief-init-expert.git`

Mise à jour de l'environnement avec le repo

```
(base) lab@lab:~$ conda activate azure
(azure) lab@lab:~$ cd dp100-brief-init-expert/
(azure) lab@lab:~/dp100-brief-init-expert$ cd env/
(azure) lab@lab:~/dp100-brief-init-expert/env$ conda env update --file environment.yml
Collecting package metadata (repodata.json): done
Solving environment: /
```

...

Downloading and Extracting Packages

smmap-3.0.4	22 KB	#####	100%
python-editor-1.0.4	9 KB	#####	100%
joblib-0.17.0	204 KB	#####	100%
docker-py-4.3.1	176 KB	#####	100%
urllib3-1.25.11	93 KB	#####	100%
cloudpickle-1.6.0	22 KB	#####	100%
pandas-1.1.4	10.5 MB	#####	100%



conda activate azure



conda env update --file environment.yml

Clone du git qui contient les notebooks mslearn

```
(base) Lab@Lab:~/dp100-brief-init-experts$ git clone https://github.com/MicrosoftDocs/mslearn-aml-labs.git
Clonage dans 'mslearn-aml-labs'...
remote: Enumerating objects: 31, done.
remote: Counting objects: 100% (31/31), done.
remote: Compressing objects: 100% (27/27), done.
remote: Total 309 (delta 12), reused 12 (delta 4), pack-reused 278
Réception d'objets: 100% (309/309), 507.51 KiB | 1.32 MiB/s, fait.
Résolution des deltas: 100% (194/194), fait.
```

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

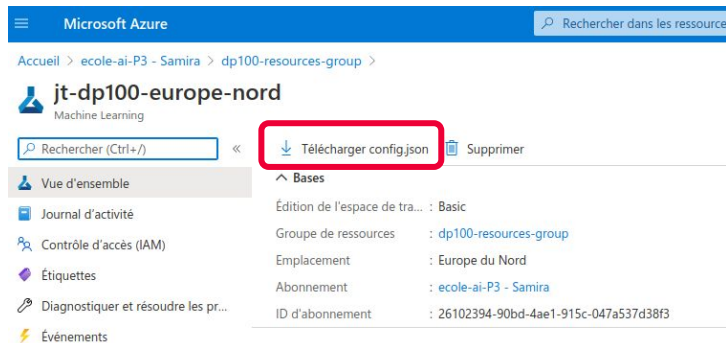
Lancez le serveur : Jupyter notebook

```
(azure) Lab@Lab:~/dp100-brief-init-experts$ jupyter notebook
[I 09:55:13.957 NotebookApp] [jupyter_nbextensions_configurator] enabled 0.4.1
[W 09:55:13.957 NotebookApp] Error loading server extension jupyterlab
Traceback (most recent call last):
  File "/home/lab/anaconda3/envs/azure/lib/python3.7/site-packages/notebook/notebookapp.py", line 1942, in init_server_extensions
    mod = importlib.import_module(modulename)
  File "/home/lab/anaconda3/envs/azure/lib/python3.7/importlib/__init__.py", line 127, in import_module
    return _bootstrap.gcd_import(name[level:], package, level)
  File "<frozen importlib._bootstrap>", line 1006, in _gcd_import
  File "<frozen importlib._bootstrap>", line 983, in _find_and_load
  File "<frozen importlib._bootstrap>", line 965, in _find_and_load_unlocked
ModuleNotFoundError: No module named 'jupyterlab'
[I 09:55:13.958 NotebookApp] Serving notebooks from local directory: /home/lab/dp100-brief-init-expert
[I 09:55:13.958 NotebookApp] Jupyter Notebook 6.1.4 is running at:
[I 09:55:13.958 NotebookApp] http://localhost:8888/?token=b6cb26f4f516840d9e080c0bd26f9017bc1e4c9461c89080
[I 09:55:13.958 NotebookApp] or http://127.0.0.1:8888/?token=b6cb26f4f516840d9e080c0bd26f9017bc1e4c9461c89080
[I 09:55:13.958 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 09:55:13.977 NotebookApp]
```



jupyter notebook

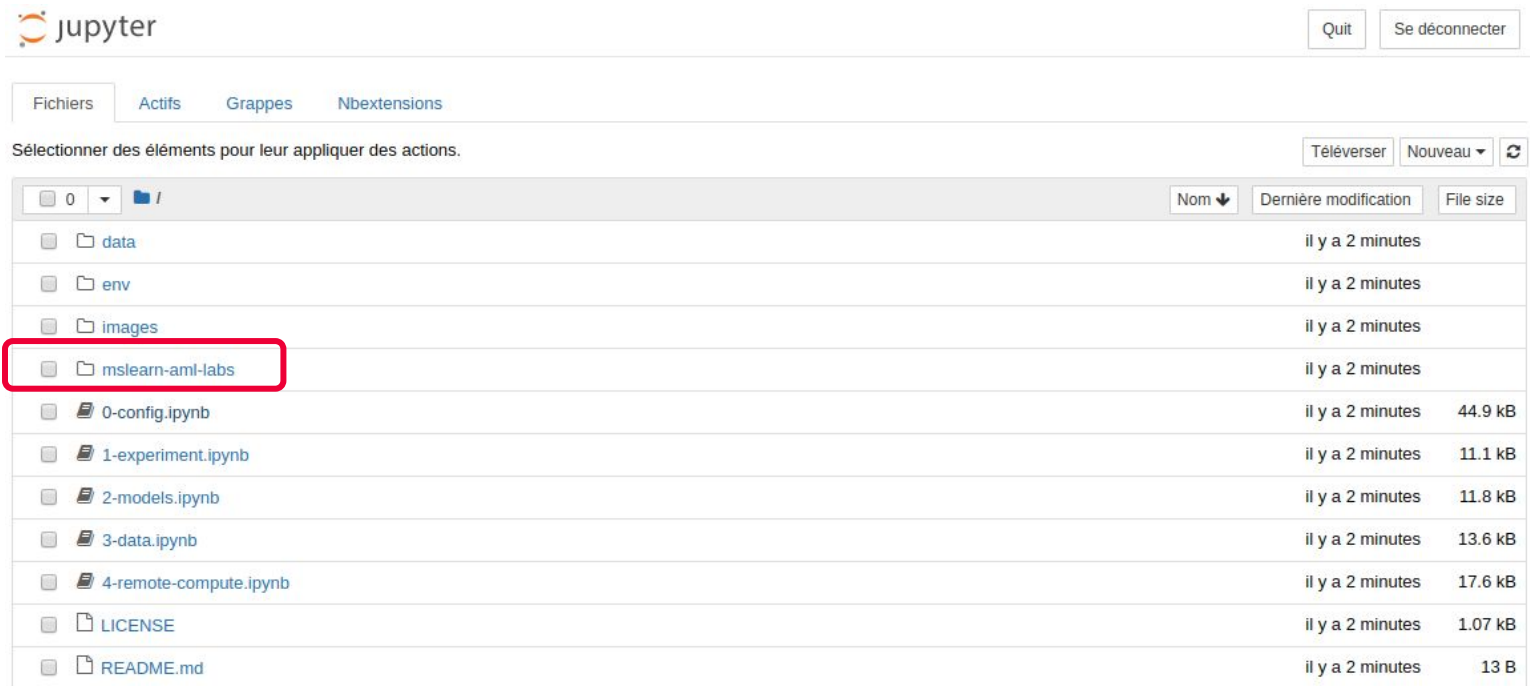
Connexion au workspace depuis le portail



➡ Pour indiquer à votre notebook les références de votre workspace, téléchargez le fichier config.json

➡ Placez ce fichier dans votre dossier local qui contient les notebooks

MSLearn labs en local



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 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"/>	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



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

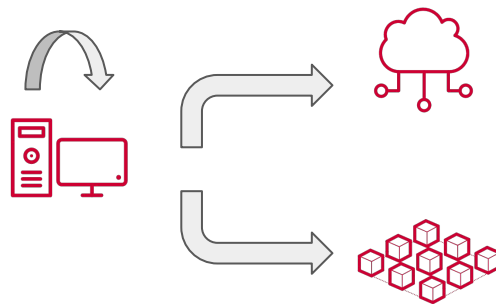
Datastores/dataset

- 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