

# Déployez un modèle dans le cloud

Soutenance du 6 novembre 2023

Yann Pham-Van

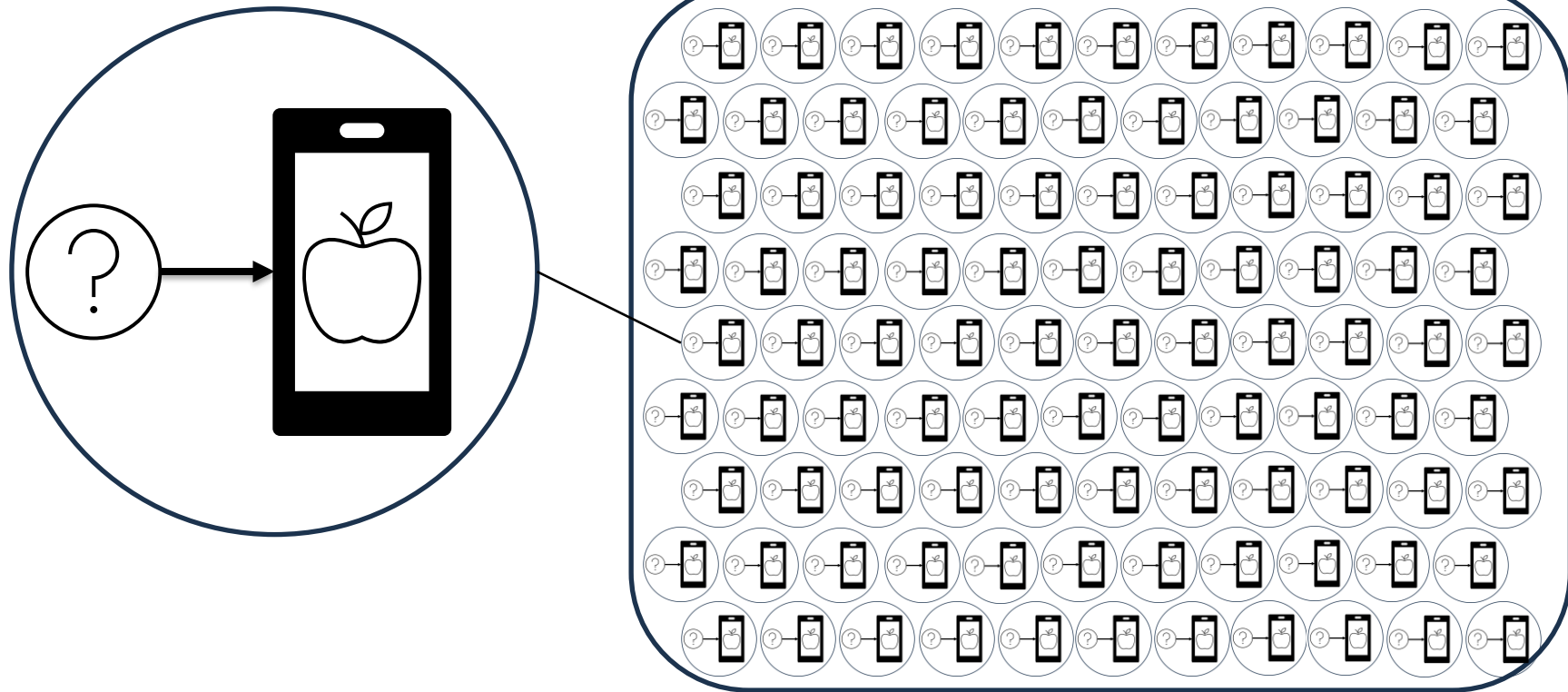
# Déployez un modèle dans le cloud

- Problématique
- Environnement Big Data
- Traitement dans le cloud
- PySpark
- Synthèse et conclusion

# Déployez un modèle dans le cloud

- **Problématique**
- Environnement Big Data
- Traitement dans le cloud
- PySpark
- Synthèse et conclusion

# Problématique



# Problématique > dataset

**A disposition : 131 fruits et légumes en  
90380 images de 100 \* 100 px**

**Training set  
67 692 images**

**Test set  
22 688 images**



**Utilisation pour la preuve de concept du  
projet**

**10**



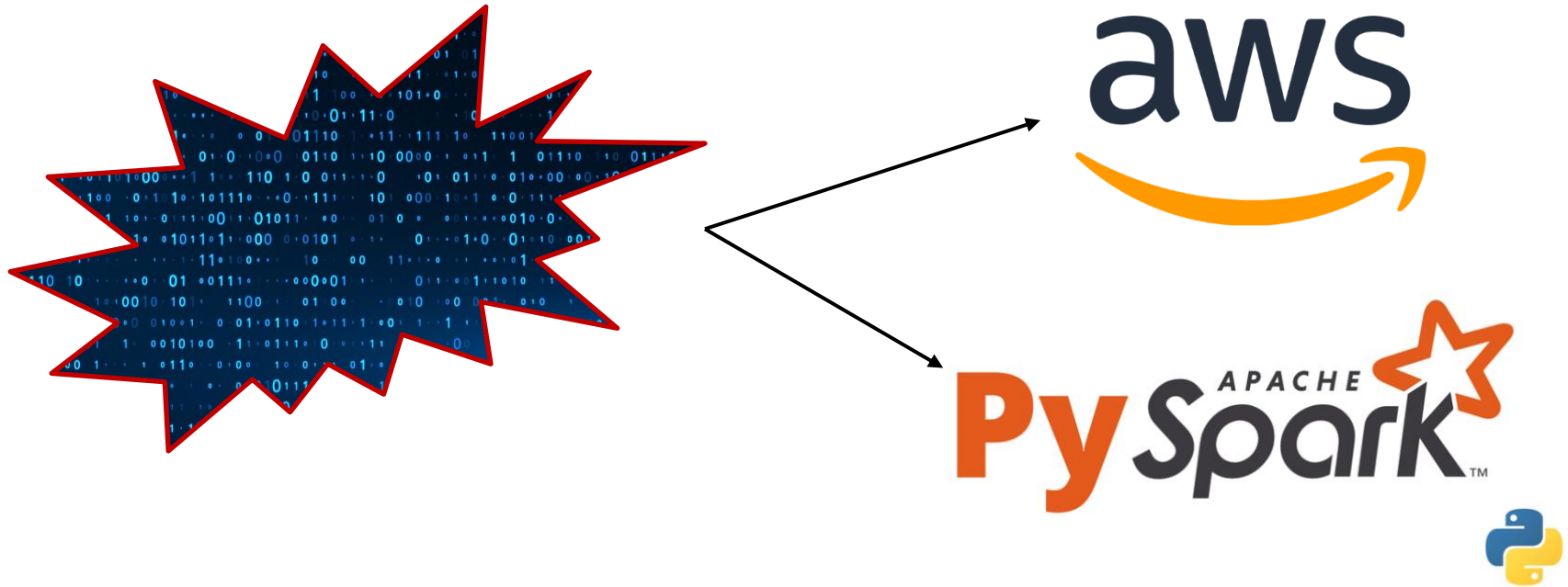
**10**



# Déployez un modèle dans le cloud

- Problématique
- **Environnement Big Data**
- Traitement dans le cloud
- PySpark
- Synthèse et conclusion

# Environnement Big Data



# Environnement Big Data > AWS





# Environnement Big Data > S3

## Amazon S3

### ► Instantané de compte

Storage Lens offre une visibilité sur l'utilisation du stockage et les tendances d'activité. [En savoir plus](#)

Afficher le tableau de bord de Storage Lens

### Compartiments (2) [Info](#)

Les compartiments sont des conteneurs pour les données stockées dans S3. [En savoir plus](#)



Copier l'ARN

Vider

Supprimer

Créer un compartiment

Rechercher des compartiments par nom

< 1 > ⚙

	Nom ▲	Région AWS ▼	Accéder ▼	Date de création ▼
<input type="radio"/>	aws-logs-297696004494-eu-west-1	Europe (Irlande) eu-west-1	<a href="#">Compartiment et objets non publics</a>	29 Oct 2023 04:41:08 PM CET
<input type="radio"/>	ypv-p8-data	Europe (Irlande) eu-west-1	<a href="#">Compartiment et objets non publics</a>	26 Oct 2023 01:36:35 PM CEST

# Environnement Big Data > EMR

Amazon EMR > EMR sur EC2: Clusters

Clusters (4) <a href="#">Info</a>								
<a href="#">Filtrer les clusters par statut</a> ▼			<input type="text" value="Rechercher des clusters"/>			<a href="#">Filtrer les clusters par date et heure de création</a>		
						< 1 > ⚙		
<input type="checkbox"/>			ID de cluster ▼	Nom du cluster ▼	Statut ▼	Heure de création (UTC+01:00) ▼	Temps écoulé ▼	Heures d'instances normalisées
<input type="checkbox"/>			<a href="#">j-24MOOTRFMG3BP</a>	P8_Fruits_1	⊖ Résilié Demande utilisateur	30 octobre 2023 09:10	40 minutes, 23 secondes	24
<input type="checkbox"/>			<a href="#">j-1Z6KGYA49PMHX</a>	P8_Fruits	⊖ Résilié Demande utilisateur	29 octobre 2023 21:08	2 heures, 22 minutes	72
<input type="checkbox"/>			<a href="#">j-GBYRDK4E4NK5</a>	Mon cluster	⊖ Résilié Arrêt automatique	29 octobre 2023 17:57	1 heure, 15 minutes	48
<input type="checkbox"/>			<a href="#">j-58FEJR1UKYL8</a>	P8_Fruits	⊗ Résilié avec des erreurs Échec d'amorçage	29 octobre 2023 16:41	31 minutes, 37 secondes	0

# Environnement Big Data > IAM

[IAM](#) > Utilisateurs

▼ Prêt à rationaliser l'accès humain à AWS et aux applications cloud ?

[Ignorer](#)

[Gérer les utilisateurs du personnel](#)

Identity Center est activé. Nous vous recommandons de gérer l'accès des utilisateurs de vos employés aux comptes AWS et applications cloud dans Identity Center.

[En savoir plus](#)

[Découvrir son fonctionnement](#)

## Utilisateurs (2) [Infos](#)



[Supprimer](#)

[Créer un utilisateur](#)

Un utilisateur IAM est une identité avec des informations d'identification à long terme utilisées pour interagir avec AWS dans un compte.

< 1 >

<input type="checkbox"/>	Nom d'utilisateur ▲	Âge du mot de passe ▼	Dernière connexion à la console ▼	ARN ▼	Heure de création ▼	Accès par console ▼
<input type="checkbox"/>	<a href="#">Administrator</a>	✔ 5 jours	October 31, 2023, 15:31 (UTC+01:00)	arn:aws:iam::297696004494:user/Administrator	Il y a 5 jours	✔ Activé
<input type="checkbox"/>	<a href="#">Evalueateur</a>	✔ 2 jours	November 01, 2023, 12:59 (UTC+01:00)	arn:aws:iam::297696004494:user/Evalueateur	Il y a 2 jours	✔ Activé

# Déployez un modèle dans le cloud

- Problématique
- Environnement Big Data
- **Traitement dans le cloud**
- PySpark
- Synthèse et conclusion

# Traitements dans le cloud > images

## Chargement au format binaire

```
+-----+-----+-----+-----+
| path | modificationTime | length | content |
+-----+-----+-----+-----+
| s3://ypv-p8-data/... | 2023-10-26 11:50:37 | 4893 | [FF D8 FF E0 00 1... |
| s3://ypv-p8-data/... | 2023-10-26 11:50:37 | 4875 | [FF D8 FF E0 00 1... |
| s3://ypv-p8-data/... | 2023-10-26 11:50:37 | 4841 | [FF D8 FF E0 00 1... |
| s3://ypv-p8-data/... | 2023-10-26 11:50:36 | 4830 | [FF D8 FF E0 00 1... |
| s3://ypv-p8-data/... | 2023-10-26 11:50:36 | 4371 | [FF D8 FF E0 00 1... |
+-----+-----+-----+-----+
only showing top 5 rows
```

## Création des labels

```
+-----+-----+-----+-----+
| path | label |
+-----+-----+-----+-----+
| s3://ypv-p8-data/Test/Avocado/r_4_100.jpg | Avocado |
| s3://ypv-p8-data/Test/Avocado/r_5_100.jpg | Avocado |
| s3://ypv-p8-data/Test/Avocado/r_41_100.jpg | Avocado |
| s3://ypv-p8-data/Test/Avocado/r_40_100.jpg | Avocado |
| s3://ypv-p8-data/Test/Avocado/87_100.jpg | Avocado |
+-----+-----+-----+-----+
only showing top 5 rows
```

# Traitements dans le cloud > modèle

```
def model_fn():  
    """  
    Returns a MobileNetV2 model with top layer removed  
    and broadcasted pretrained weights.  
    """  
    model = MobileNetV2(weights='imagenet',  
                        include_top=True,  
                        input_shape=(224, 224, 3))  
    for layer in model.layers:  
        layer.trainable = False  
    new_model = Model(inputs=model.input,  
                     outputs=model.layers[-2].output)  
    new_model.set_weights(broadcast_weights.value)  
    return new_model
```

- Récupération de l'avant-dernière sortie du modèle MobileNetV2  
→ vecteur de dimension 1280
- Diffusion des poids du nouveau modèle sur les workers

# Traitements dans le cloud > features

- Images 100\*100 → 224\*224
- Featurisation des images avec le modèle → array
- Transformation des array en Vector

path	label	features
s3://ypv-p8-data/...	Avocado	[0.43423667550086...
s3://ypv-p8-data/...	Avocado	[0.84571015834808...
s3://ypv-p8-data/...	Avocado	[0.35587260127067...
s3://ypv-p8-data/...	Avocado	[0.08634439110755...
s3://ypv-p8-data/...	Avocado	[1.37034046649932...
s3://ypv-p8-data/...	Avocado	[0.87466561794281...
s3://ypv-p8-data/...	Avocado	[0.63034307956695...
s3://ypv-p8-data/...	Avocado	[0.36890134215354...
s3://ypv-p8-data/...	Apricot	[0.12042748928070...
s3://ypv-p8-data/...	Avocado	[0.83359342813491...
s3://ypv-p8-data/...	Apricot	[0.28019240498542...
s3://ypv-p8-data/...	Avocado	[0.27026844024658...
s3://ypv-p8-data/...	Apricot	[0.22149887681007...
s3://ypv-p8-data/...	Apricot	[0.30879327654838...
s3://ypv-p8-data/...	Apricot	[0.18892598152160...
s3://ypv-p8-data/...	Apricot	[0.22485153377056...
s3://ypv-p8-data/...	Apricot	[0.58618354797363...
s3://ypv-p8-data/...	Apricot	[0.24216072261333...
s3://ypv-p8-data/...	Apricot	[0.27679243683815...
s3://ypv-p8-data/...	Apricot	[0.38278287649154...

# Traitements dans le cloud > PCA

## Etapes de la PCA

features

|

StandardScaler

|

scaledFeatures

|

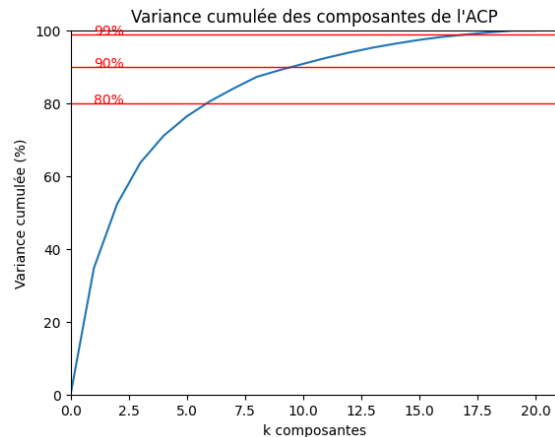
PCA

|

pcaFeatures

path	label	features	scaledFeatures	pcaFeatures
s3://ypv-p8-data/...	Avocado	[0.43423667550086...	-0.0338907992986...	[-12.422524818156...
s3://ypv-p8-data/...	Avocado	[0.84571015834808...	1.24459169052246...	[-23.330534821401...
s3://ypv-p8-data/...	Avocado	[0.35587260127067...	-0.2773745252512...	[-28.621932378135...
s3://ypv-p8-data/...	Avocado	[0.08634439110755...	-1.1148211933135...	[-30.111787551363...
s3://ypv-p8-data/...	Avocado	[1.37034046649932...	2.87466189081741...	[-26.436692392112...
s3://ypv-p8-data/...	Avocado	[0.87466561794281...	1.33455872271230...	[-14.022579001952...
s3://ypv-p8-data/...	Avocado	[0.63034307956695...	0.57542818282728...	[-13.775099661257...
s3://ypv-p8-data/...	Avocado	[0.36890134215354...	-0.2368931387767...	[-14.140202571470...
s3://ypv-p8-data/...	Apricot	[0.12042748928070...	-1.0089221596462...	[18.1141014612940...
s3://ypv-p8-data/...	Avocado	[0.83359342813491...	1.20694399739719...	[-12.797352595869...

only showing top 10 rows





# Déployez un modèle dans le cloud

- Problématique
- Environnement Big Data
- Traitement dans le cloud
- **PySpark**
- Synthèse et conclusion

# PySpark

## Démonstration d'exécution dans le cloud

### Démarrage de la session Spark

Entrée [1]: `# L'exécution de cette cellule démarre l'application Spark`

Starting Spark application

ID	YARN Application ID	Kind	State	Spark UI	Driver log	Current session?
0	application_1698654140027_0001	pyspark	idle	<a href="#">Link</a>	<a href="#">Link</a>	✓

FloatProgress(value=0.0, bar\_style='info', description='Progress:', layout=Layout(height='25px', width='50%'),...

SparkSession available as 'spark'.

FloatProgress(value=0.0, bar\_style='info', description='Progress:', layout=Layout(height='25px', width='50%'),...

### Affichage des informations sur la session en cours et liens vers Spark UI

Entrée [2]: `%info`

Current session configs: {'driverMemory': '1000M', 'executorCores': 2, 'proxyUser': 'jovyan', 'kind': 'pyspark'}

ID	YARN Application ID	Kind	State	Spark UI	Driver log	Current session?
0	application_1698654140027_0001	pyspark	idle	<a href="#">Link</a>	<a href="#">Link</a>	✓

### Import des librairies

Entrée [3]: `import pandas as pd  
from PIL import Image  
import numpy as np  
import io  
import os  
  
import tensorflow as tf  
from tensorflow.keras.applications.mobilenet_v2 import MobileNetV2, preprocess_input  
from tensorflow.keras.preprocessing.image import img_to_array  
from tensorflow.keras import Model  
from pyspark.sql.functions import col, pandas_udf, PandasUDFType, element_at, split, udf  
from pyspark.sql import SparkSession  
from pyspark.ml.feature import PCA, StandardScaler  
from pyspark.ml.linalg import Vectors, VectorUDT`

FloatProgress(value=0.0, bar\_style='info', description='Progress:', layout=Layout(height='25px', width='50%'),...

### Définition des PATH pour charger les images et enregistrer les résultats

# Déployez un modèle dans le cloud

- Problématique
- Environnement Big Data
- Traitement dans le cloud
- PySpark
- **Synthèse et conclusion**

# Synthèse et conclusion

- 2 phases
  - Local

Difficulté d'installation d'un environnement Spark sous Windows → Google Colab
  - Passage à l'échelle
    - EMR, le PaaS d'AWS, pour une installation aisée des librairies et packages nécessaires en prenant garde à l'inter-compatibilité
    - S3 pour un stockage des données à faible coût en illimité
- Conclusion

Accompagnement facilité de la montée en charge avec redimensionnement horizontal (nombre d'instances) et/ou vertical (puissance des clusters)

Avez-vous des questions ?