



Déployer un modèle dans le cloud

Aicha HACHEMI

OpenClassrooms





Sommaire

- **Présentation du projet**
 - **Environnement BigData**
 - **Stockage des données**
 - **Loi rgpd**
 - **Cloud computing**
 - **Pre-processing et réduction de dimension**
 - **Conclusion**
- 

Présentation du Projet

- **Objectif** La jeune start-up **Fruits** souhaite mettre à disposition du grand public une application mobile qui permettrait aux utilisateurs de prendre en photo un fruit et d'obtenir des informations sur ce fruit.
- **Mission** Mettre en place les premières briques de la chaîne de traitement.

Apple Red 2



Apple Granny Smith

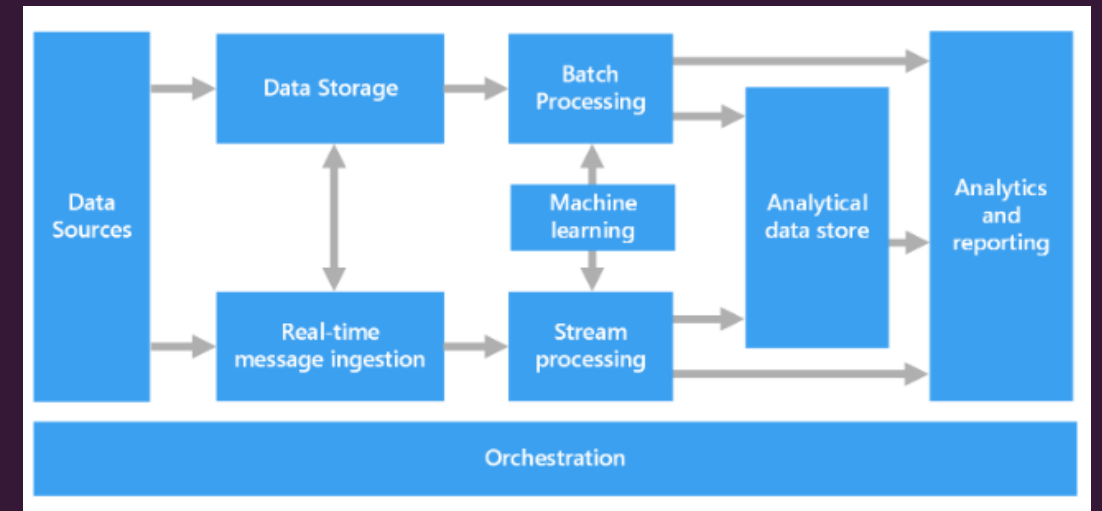
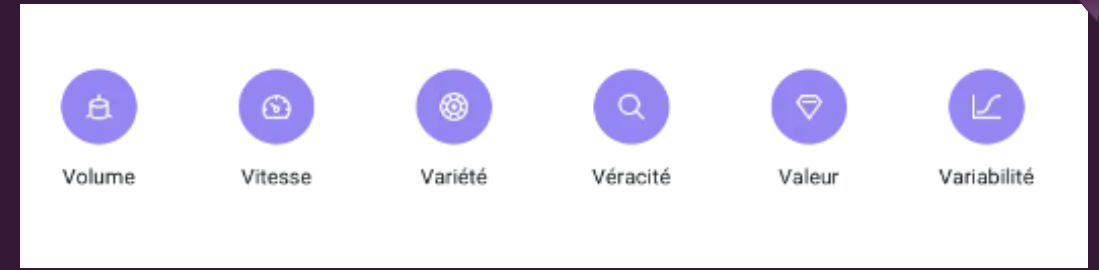


Wa

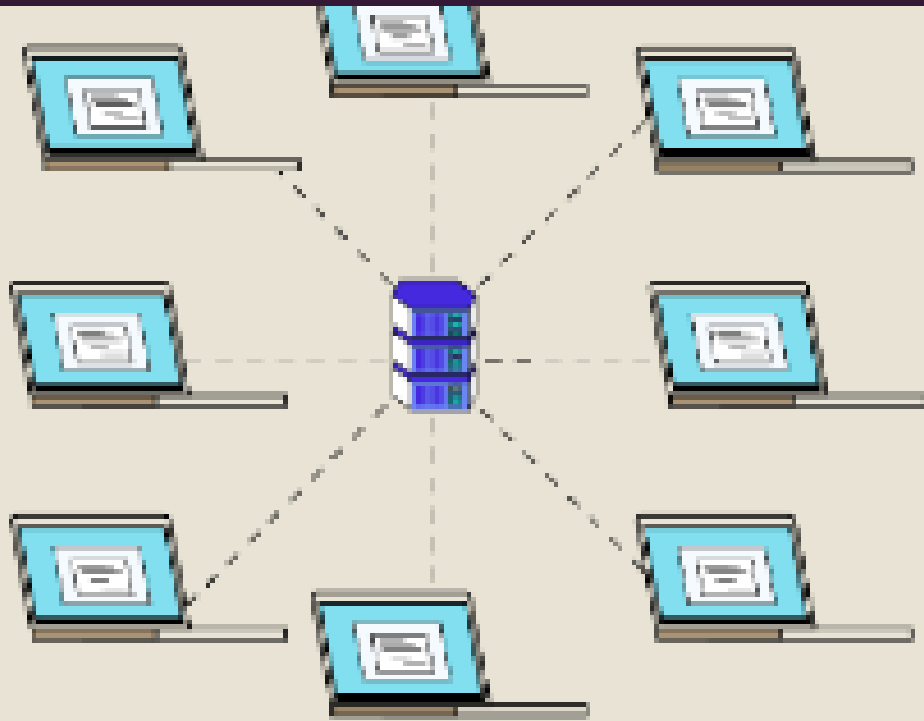


Fruits!

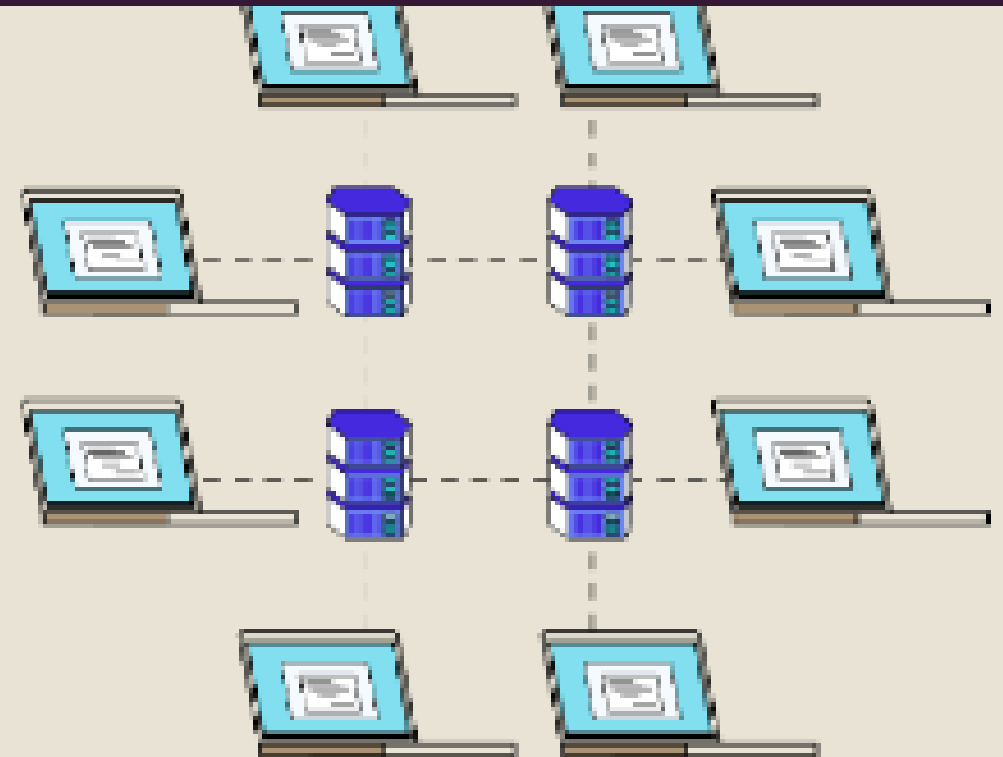
Environnement Big Data



Environnement Big Data



Systeme Centralisé



Systeme Distribué

Environnement Big Data



S3



Bucket



Bucket with
Objects



Stockage des données



Amazon S3

Data storage
for unstructured
data



Amazon EBS

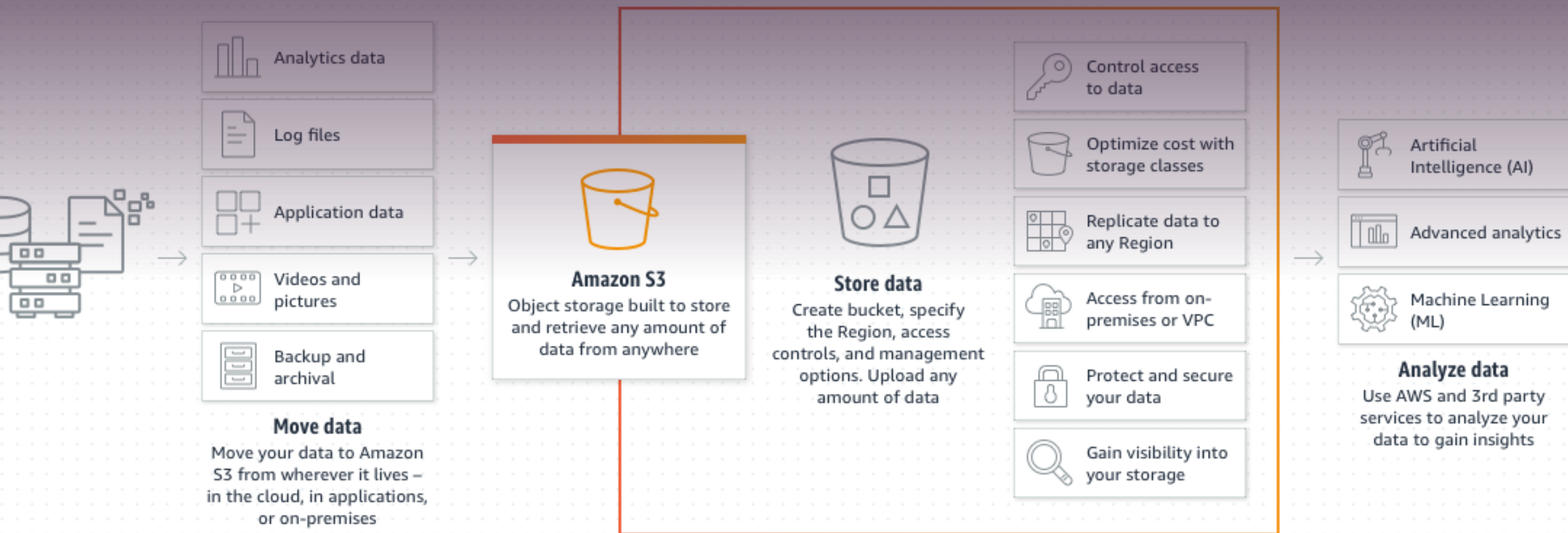
System storage
for Amazon EC2
VMs



Amazon EFS

Scalable data
storage for
Amazon EC2 VMs

Stockage des données




Stockage des données

Règles RGPD : Serveurs européens

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 p

	Nom ▲	Région AWS ▼	Accéder ▼
<input type="radio"/>	aws-logs-989339500288-eu-central-1	Europe (Francfort) eu-central-1	<u>Compartiment et objets non publics</u>
<input type="radio"/>	p8-oc-data-test	Europe (Francfort) eu-central-1	<u>Compartiment et objets non publics</u>

Stockage des données

Services

Rechercher

[Alt+S]

Global

A-H

Objets (7)

Les objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez utiliser l'[inventaire Amazon S3](#) pour obtenir une liste de tous les objets de votre compartiment. Pour que d'autres personnes puissent accéder à vos objets, vous devez leur accorder explicitement des autorisations. [En savoir plus](#)

Copier l'URI S3

Copier l'URL

Télécharger

Ouvrir

Supprimer

Actions





Créer un dossier

Charger

Rechercher des objets en fonction du préfixe

☐ Afficher les versions

< 1 >

<input type="checkbox"/>	Nom	Type	Dernière modification	Taille	Classe de stockage
<input type="checkbox"/>	 bootstrap-emr.sh	sh	PM CEST	327.0 o	Standard
<input type="checkbox"/>	 config.json	json	18 Oct 2023 08:53:14 AM CEST	128.0 o	Standard
<input type="checkbox"/>	 jupyter/	Dossier	-	-	-
<input type="checkbox"/>	 output.csv/	Dossier	-	-	-

Cloud Computing

Gmail, Trello, Salesforce CRM,
EventPro, Office 365,
Google Docs

end users

SAAS

Heroku, AWS Elastic Beanstalk,
Google App Engine

software
developers

PAAS

AWS Elastic Compute
Cloud (EC2),
Microsoft Azure,
Google Compute
Engine

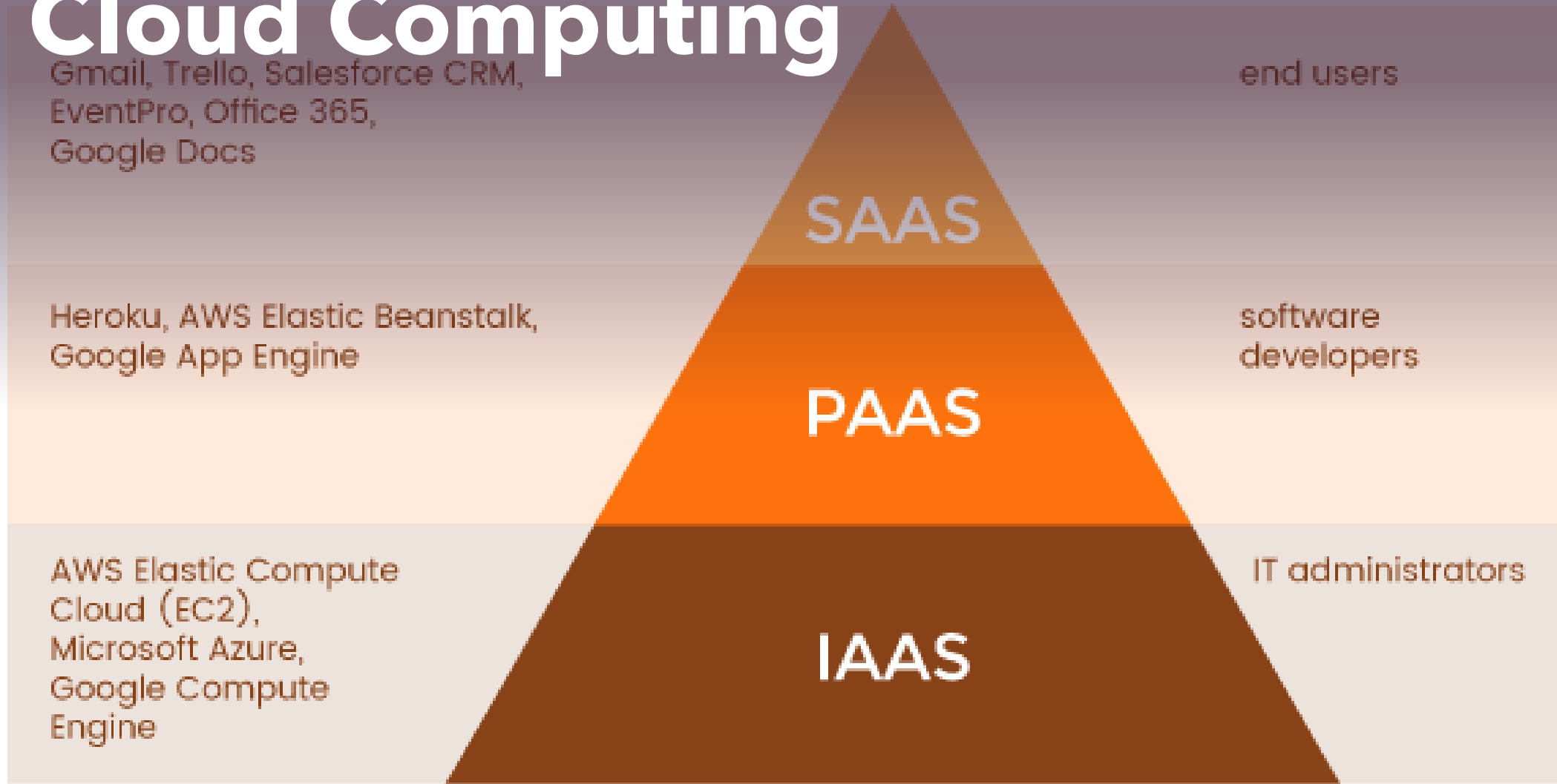
IT administrators

IAAS

less

control

more

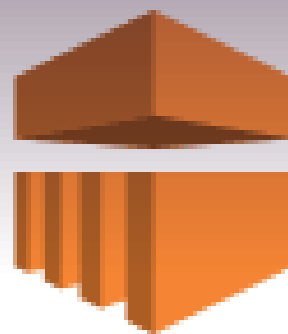
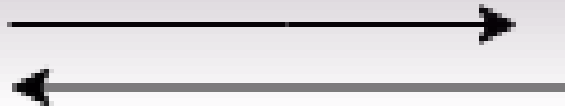


Cloud Computing

Solution Paas : EMR



Amazon S3



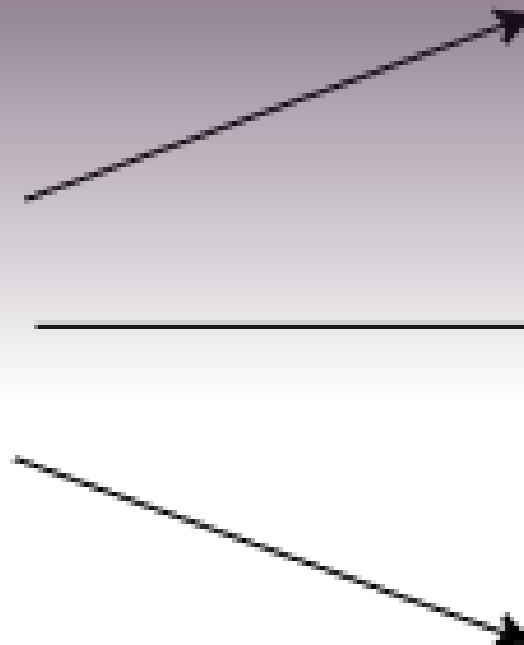
EMR



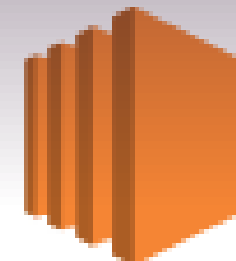
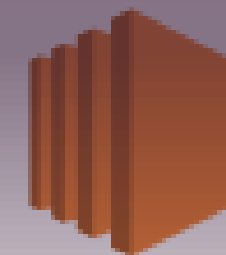
EMR
Cluster

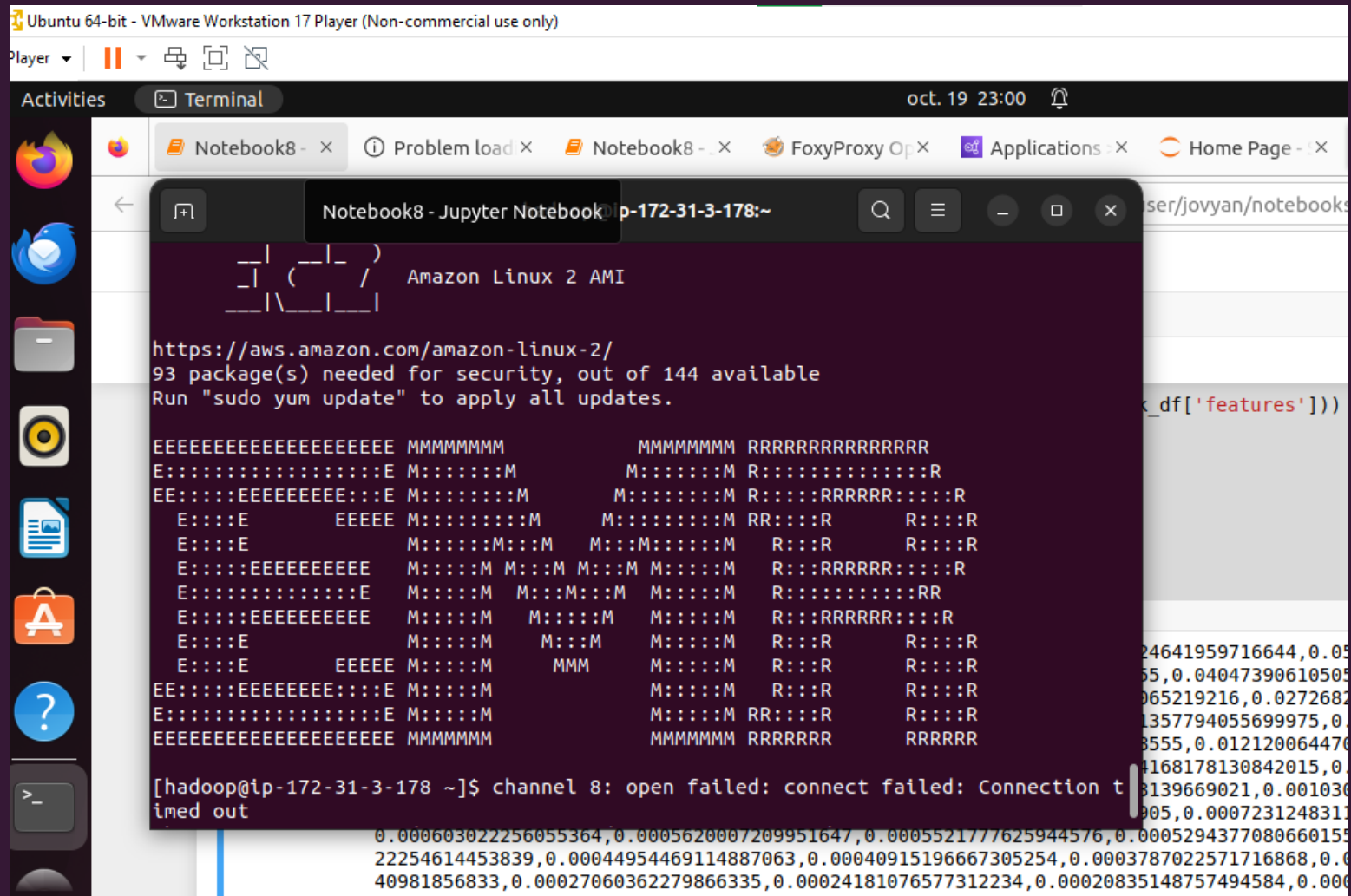


PySpark



EC2's





Cloud Computing

Solution Paas : EMR

Interfaces utilisateur d'application en direct

Ces interfaces utilisateur d'application sur cluster sont disponibles sans tunneling SSH.

Interfaces utilisateur d'application [↗](#)

[Interface utilisateur du serveur d'historique Spark](#)

Interfaces utilisateur d'application sur le nœud primaire

Activer une c

Celles-ci nécessitent l'activation du tunneling SSH.

Application	URL de l'interface utilisateur ↗
Gestionnaire de ressources	http://ec2-18-156-198-3.eu-central-1.compute.amazonaws.com:8088/
JupyterHub	https://ec2-18-156-198-3.eu-central-1.compute.amazonaws.com:9443/
Nom du nœud HDFS	http://ec2-18-156-198-3.eu-central-1.compute.amazonaws.com:9870/
Serveur d'historique Spark	http://ec2-18-156-198-3.eu-central-1.compute.amazonaws.com:18080/

Interface utilisateur d'application sur les nœuds principaux et nœuds de tâches

Cloud Computing

```
ot
-- path: string (nullable = true)
-- modificationTime: timestamp (nullable =
-- length: long (nullable = true)
-- content: binary (nullable = true)
-- label: string (nullable = true)
```

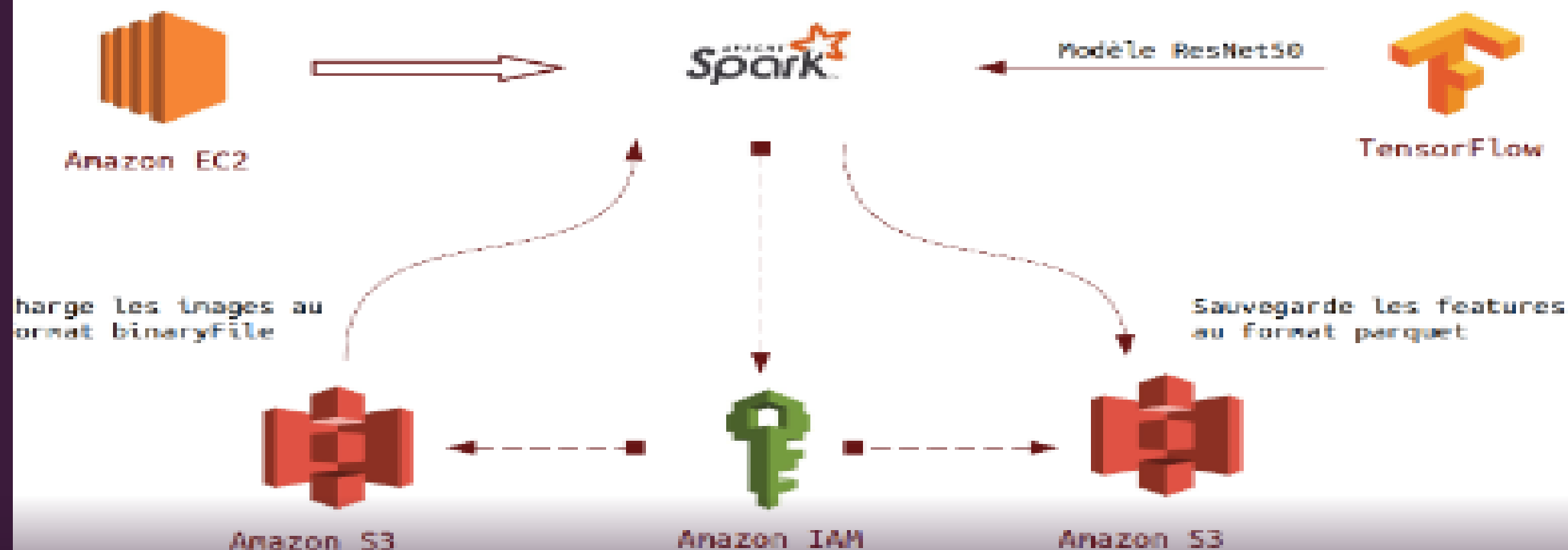
ne

ath

```
3://p8-oc-data-test/Test1/Orange/3_100.jpg
3://p8-oc-data-test/Test1/Orange/4_100.jpg
3://p8-oc-data-test/Test1/Orange/7_100.jpg
3://p8-oc-data-test/Test1/Orange/8_100.jpg
3://p8-oc-data-test/Test1/Orange/6_100.jpg
```

ly showing top 5 rows

Cloud Computing



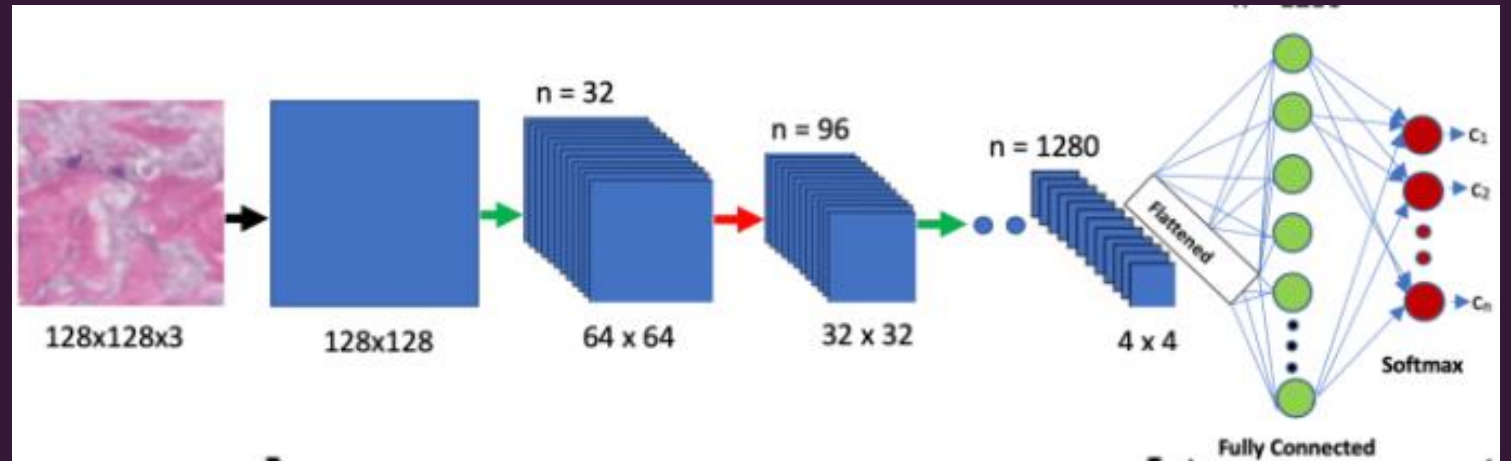
Pre-processing

Modèle de classification CNN

Les dernières layers de prédictions de MobileNetV2 sont supprimées.

Les poids utilisés sont ceux d'Imagenet, et ces mêmes poids seront diffusés à tous les workers.

les images dans notre modèle nous devons transformer les inputs en array, pour cela nous utiliserons une fonctions pandas au sein de PySpark.



Pre-processing



Rechercher [Alt+S]

Objets (25)

Objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez utiliser l'[inventaire Amazon S3](#) pour obtenir une liste de tous les objets de votre compartiment. Pour que d'autres personnes puissent accéder à vos objets, vous devez leur accorder explicitement des autorisations. [En savoir plus](#)

[Copier l'URI S3](#) [Copier l'URL](#) [Télécharger](#) [Ouvrir](#) [Supprimer](#) [Actions](#)

Rechercher des objets en fonction du préfixe ☐ Afficher les versions

Nom	Type	Dernière modification	Taille
66f8-478d-b475-1266cb79fca8-c000.snappy.parquet	parquet	19 Oct 2023 10:55:30 PM CEST	15.0 Ko
part-00002-6691d638-66f8-478d-b475-1266cb79fca8-c000.snappy.parquet	parquet	19 Oct 2023 10:55:30 PM CEST	15.0 Ko
part-00003-6691d638-66f8-478d-b475-1266cb79fca8-c000.snappy.parquet	parquet	19 Oct 2023 10:55:30 PM CEST	10.0 Ko
part-00004-6691d638-66f8-478d-b475-	parquet	19 Oct 2023 10:55:38 PM	10.0 Ko

Services [Alt+S]

Amazon S3

Compartiments

Points d'accès

Points d'accès de l'objet

Lambda

Points d'accès multi-région

Opérations par lot

IAM Access Analyzer pour S3

Paramètres de blocage de l'accès public pour ce compte

Storage Lens

Tableaux de bord

Paramètres AWS Organizations

Fonctionnalité spot

output.csv/

[Copier l'URI S3](#)

Objets Propriétés

Objets (3)

Les objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez utiliser l'[inventaire Amazon S3](#) pour obtenir une liste de tous les objets de votre compartiment. Pour que d'autres personnes puissent accéder à vos objets, vous devez leur accorder explicitement des autorisations. [En savoir plus](#)

[Copier l'URI S3](#) [Copier l'URL](#) [Télécharger](#) [Ouvrir](#) [Supprimer](#) [Actions](#)

[Créer un dossier](#) [Charger](#)

Rechercher des objets en fonction du préfixe ☐ Afficher les versions

<input type="checkbox"/>	Nom	Type	Dernière modification	Taille	Classe de stockage
<input type="checkbox"/>	_SUCCESS	-	20 Oct 2023 12:27:55 AM CEST	0 o	Standard
<input type="checkbox"/>	part-00000-4e9e09fd-45ea-4815-	csv	20 Oct 2023 12:27:55	62.7 Ko	Standard

- Les résultats sont écrits sous forme de parquets au sein de notre bucket S3.
- Une PCA à 100 composantes a été effectuée et les résultats de la réduction sont stockés sous forme csv dans un bucket S3.



Conclusion

- 
- Découverte de l'architecture Big Data
 - Belle expérience dans le cloud computing.
- 