

Déployez un modèle dans le cloud



Fruits!

*Presented by Maria BOUCHEHBOUN
For OpenClassrooms
Mentor Nicolas Rangeon*

Contents

- 01 ● Mission
- 02 ● Dispositif et architecture
- 03 ● Local & Cloud
- 04 ● Conclusion
- 05 ● Annexes



Mission

Fruits est une startup qui souhaite dans un premier temps se faire connaître en mettant 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**.

Pour la start-up, cette application permettrait de sensibiliser le grand public à la biodiversité des fruits et de mettre en place une première version du **moteur de classification des images de fruits**. De plus, le développement de l'application mobile permettra de construire une première version de **l'architecture Big Data nécessaire**.

Roadmap : Quoi?

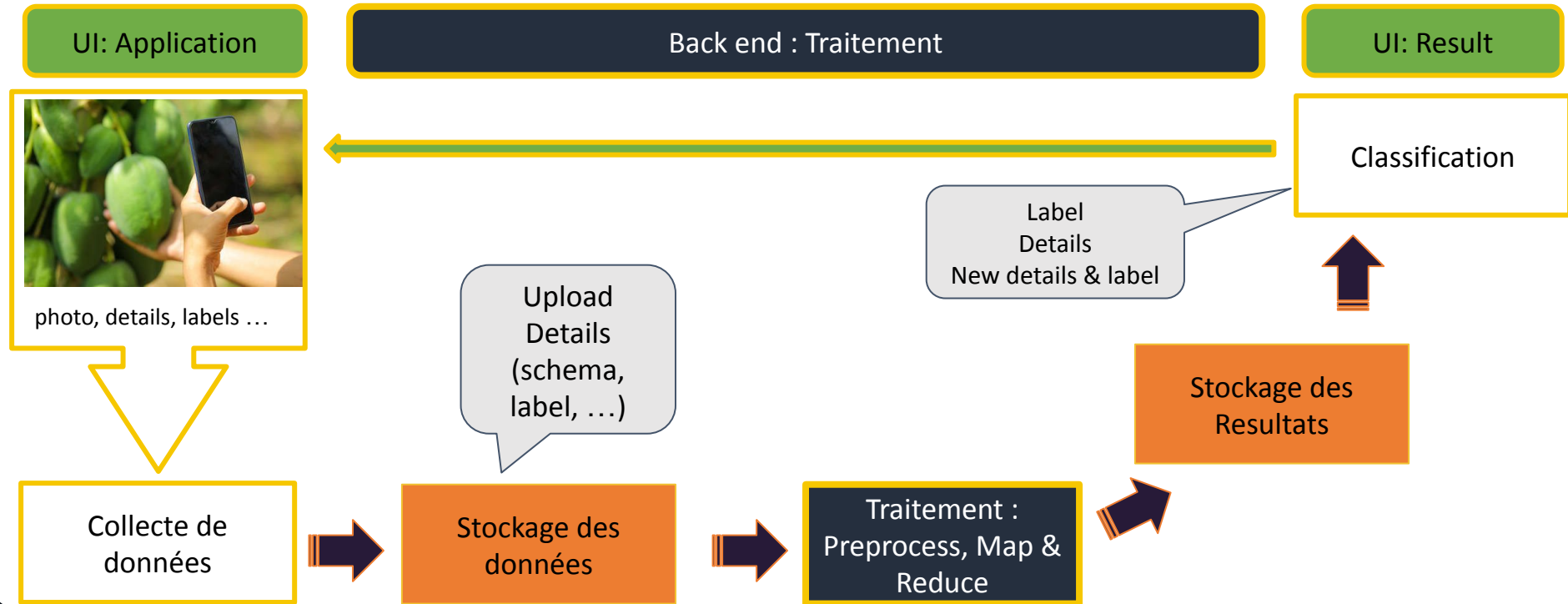
Faire tourner le modele en local

Bascule sur AWS:
Utilisation de EMR, S3, EMR
Studio, EC2

Faire tourner le code Pyspark
sur AWS

Passage à l'échelle Big Data

Roadmap : Comment?





Avec Quoi?

- Training set : 67692 images (one fruit or vegetable per image)
- Test set : 22688 images (one fruit or vegetable per image)
- Number of classes: 131 (fruits and vegetables)
- Image size: 100x100 pixels

Source : <https://www.kaggle.com/datasets/moltean/fruits/data>

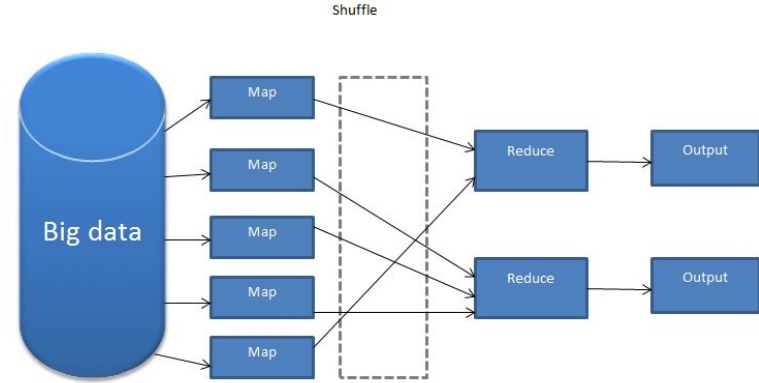
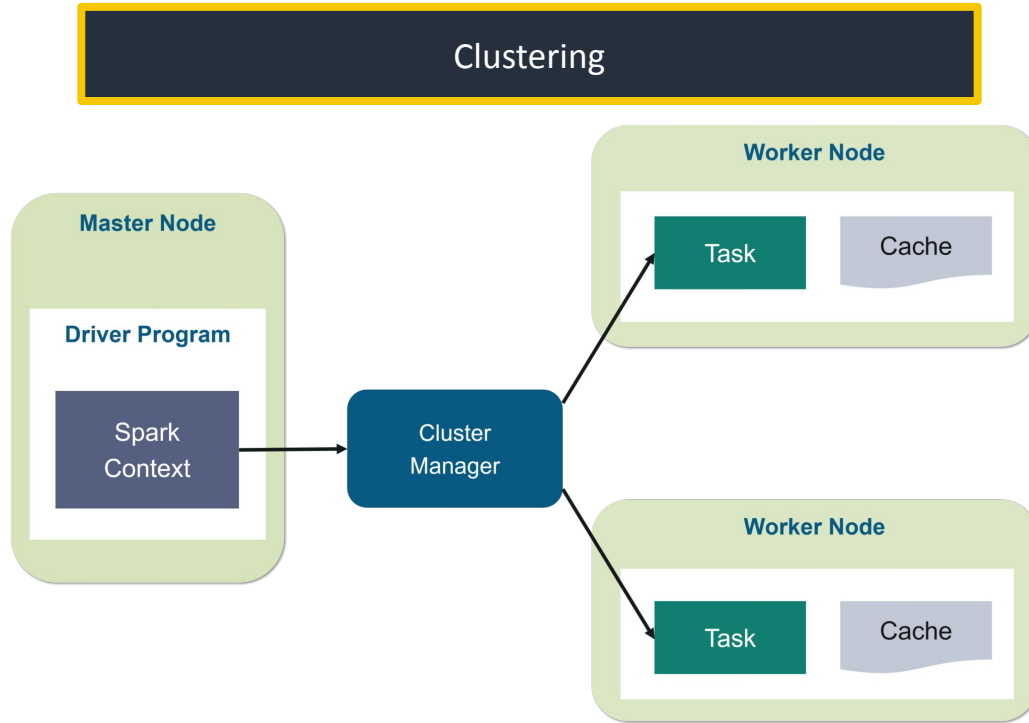
```
None
+-----+-----+
|path                                         |label|
+-----+-----+
|file:/content/fruits/fruits-360_dataset/fruits-360/Training/Raspberry/176_100.jpg|Raspberry|
|file:/content/fruits/fruits-360_dataset/fruits-360/Training/Raspberry/179_100.jpg|Raspberry|
|file:/content/fruits/fruits-360_dataset/fruits-360/Training/Pineapple Mini/170_100.jpg|Pineapple Mini|
|file:/content/fruits/fruits-360_dataset/fruits-360/Training/Raspberry/157_100.jpg|Raspberry|
|file:/content/fruits/fruits-360_dataset/fruits-360/Training/Raspberry/131_100.jpg|Raspberry|
+-----+-----+
only showing top 5 rows
```



Dispositif et architecture



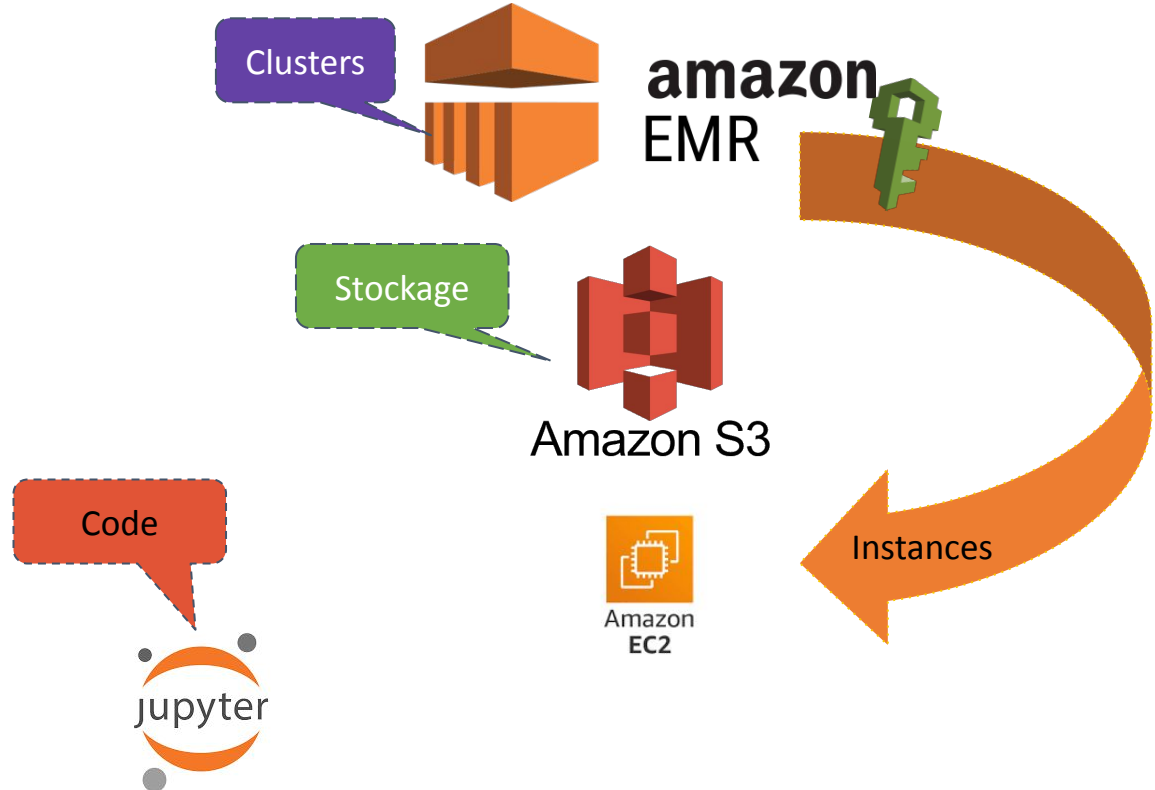
Architecture



- ★ Base architecture : Hadoop + PySpark (python + spark)
- ★ Spark : Big data, Calcul distribué, Parallelisation (UDF), RDD

Local

Cloud : AWS





Local & Cloud

Local

Debut

Spark Session

Fin

collecte

Stockage

Chargement

Featurize

Reduce

Stockage
Resultat

Classification

Load image

Arr to vec

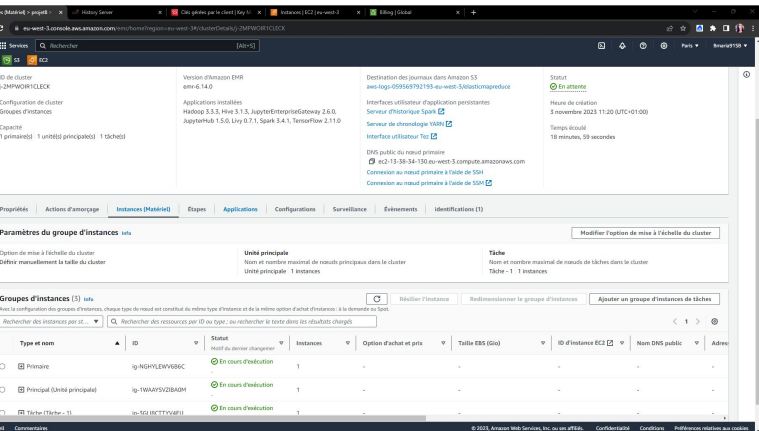
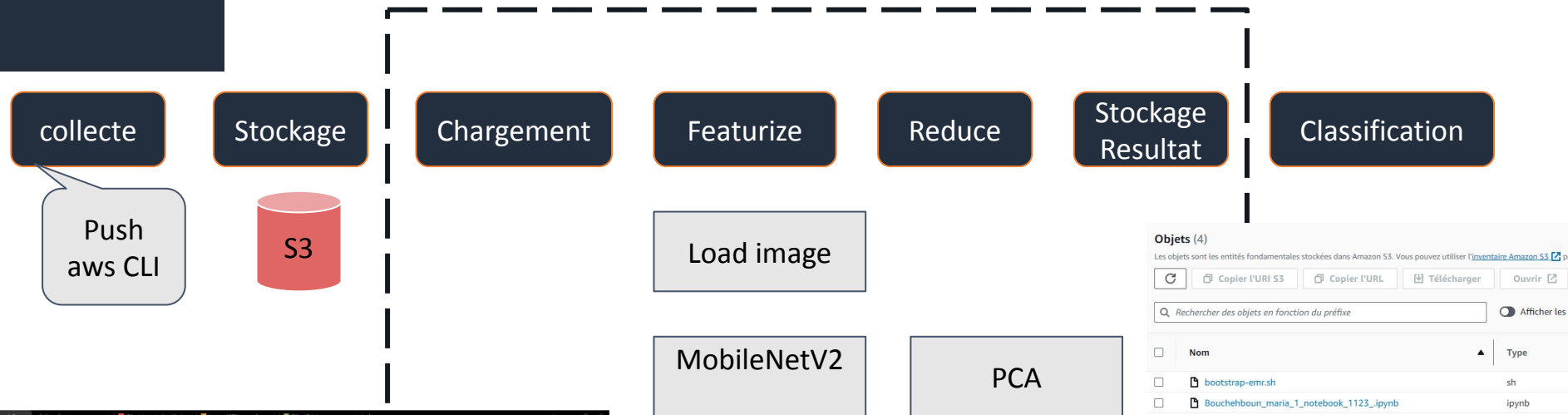
MobileNetV2

PCA

```
spark.read.format().  
option().load(path)
```

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

kaggle



Objets (4)

Les objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez utiliser l'[inventaire Amazon S3](#) pour...

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

Rechercher des objets en fonction du préfixe

Afficher les v

<input type="checkbox"/>	Nom	Type
<input type="checkbox"/>	bootstrap-emr.sh	sh
<input type="checkbox"/>	Bouchehboun_maria_1_notebook_1123_ipynb	ipynb
<input type="checkbox"/>	Resultat/	Dossier
<input type="checkbox"/>	Test/	Dossier

Plus rapide (chargement et stockage)
Volume / Variety/Vitesse



Conclusion

- Projet à haut potentiel mais il faut faire attention à l'aspect économique
- Voir au cas ou on coche lors de la création du cluster , auto scaling, l'impact sur le nombre d'instances, le coût et la rapidité.
- Mieux avancer sur la prise en main de EMR Studio

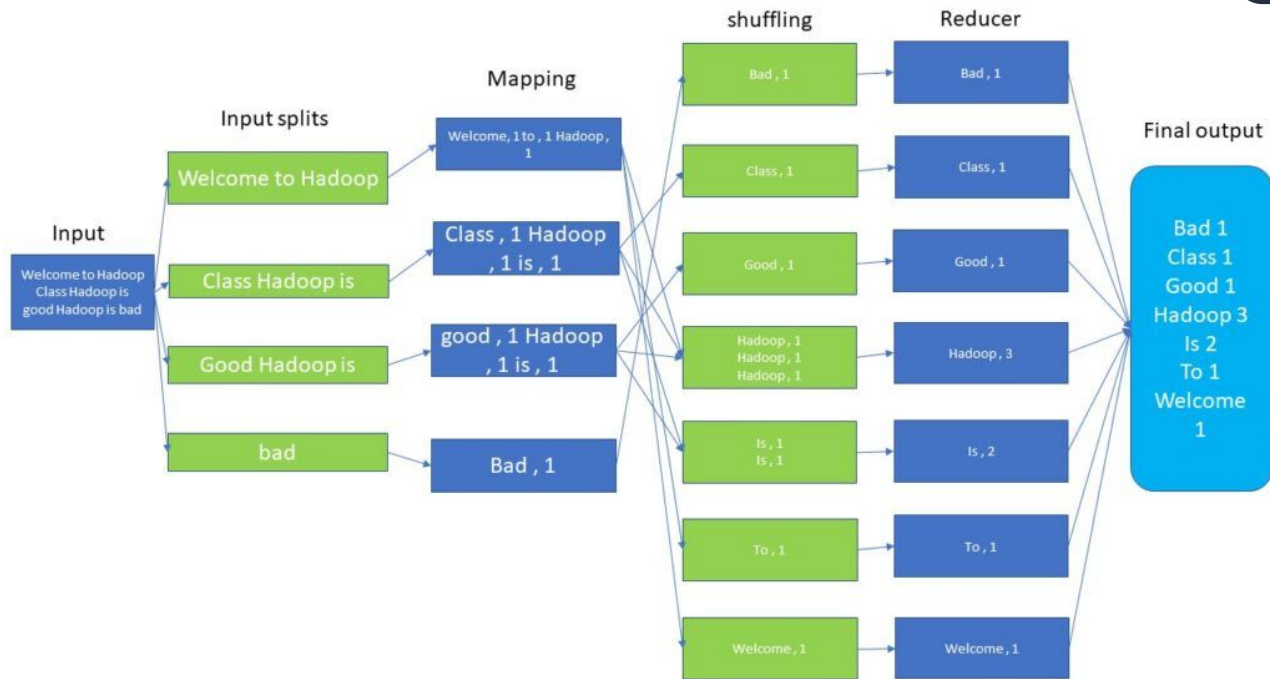


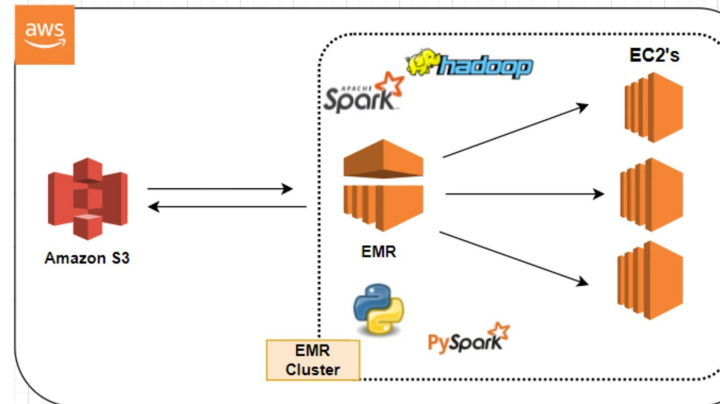
Annexes

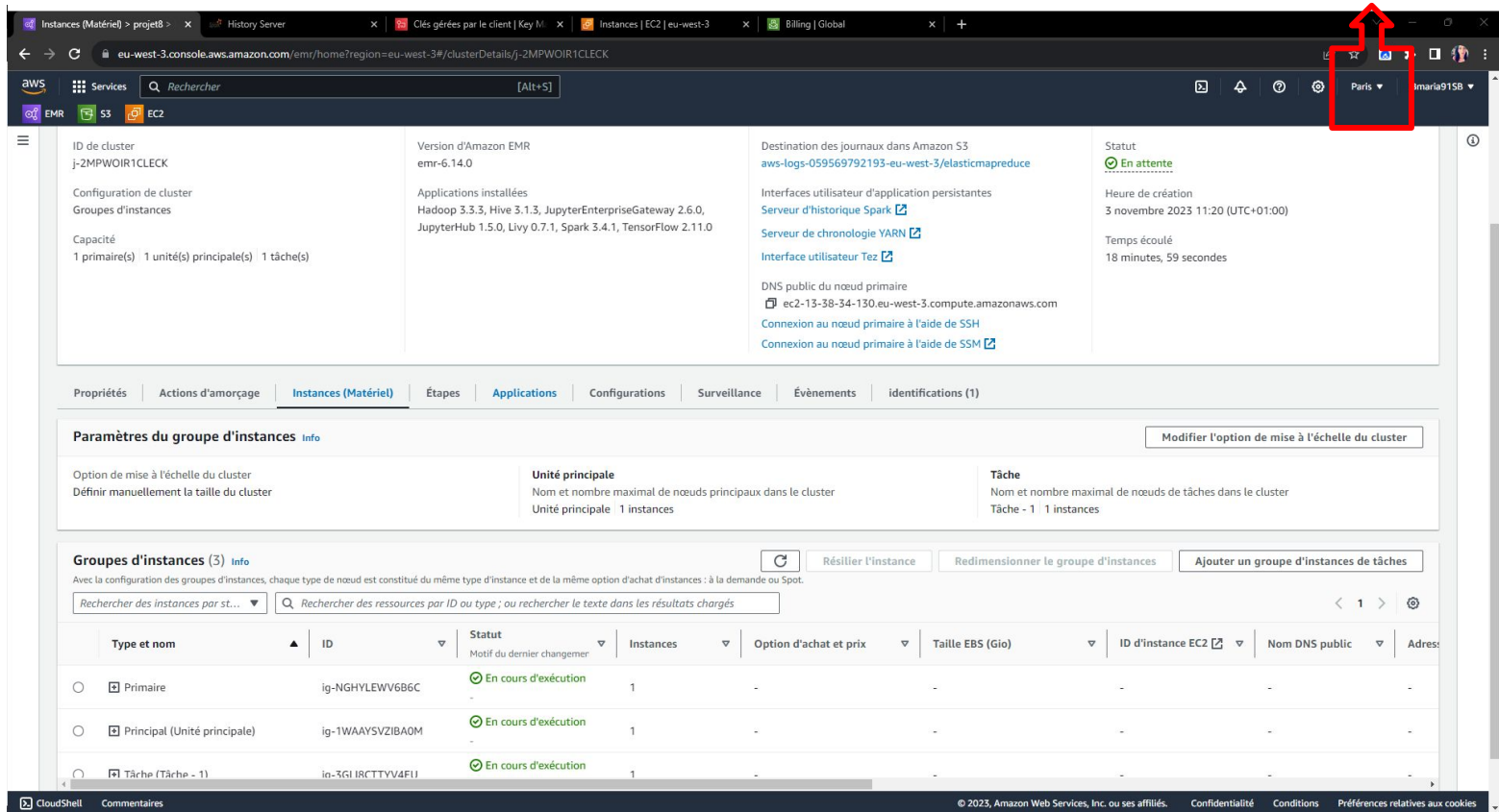


Source :

<https://intellipaat.com/blog/what-is-apache-spark/>







The screenshot displays the AWS Management Console for an Amazon EMR cluster named j-2MPWOIR1CLECK. The interface is in French. The top navigation bar shows the 'Paris' region selected. The main content area provides details about the cluster, including its ID, version, configuration, and status. Below this, there are tabs for 'Propriétés', 'Actions d'amorçage', 'Instances (Matériel)', 'Étapes', 'Applications', 'Configurations', 'Surveillance', 'Événements', and 'Identifications (1)'. The 'Instances (Matériel)' tab is active, showing the 'Paramètres du groupe d'instances' and a table of 'Groupes d'instances'.

Paramètres du groupe d'instances

Option de mise à l'échelle du cluster:
Définir manuellement la taille du cluster

Unité principale
Nom et nombre maximal de nœuds principaux dans le cluster
Unité principale 1 instances

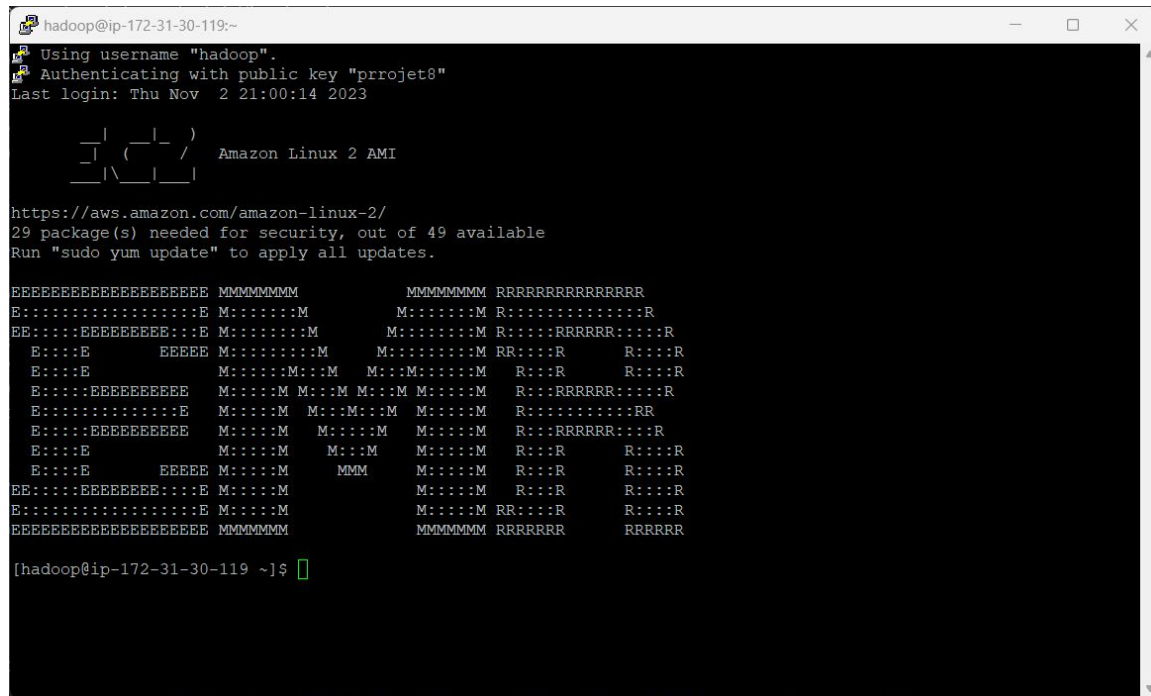
Tâche
Nom et nombre maximal de nœuds de tâches dans le cluster
Tâche - 1 1 instances

Groupes d'instances (3)

Avec la configuration des groupes d'instances, chaque type de nœud est constitué du même type d'instance et de la même option d'achat d'instances : à la demande ou Spot.

Rechercher des instances par st...
Rechercher des ressources par ID ou type ; ou rechercher le texte dans les résultats chargés

Type et nom	ID	Statut	Instances	Option d'achat et prix	Taille EBS (Gio)	ID d'instance EC2	Nom DNS public	Adresse
Primaire	ig-NGHYLEWV6B6C	En cours d'exécution	1	-	-	-	-	-
Principal (Unité principale)	ig-1WAAYSVZIBA0M	En cours d'exécution	1	-	-	-	-	-
Tâche (Tâche - 1)	in-3GI IRCTTVV4FI	En cours d'exécution	1	-	-	-	-	-



EMR

FoxyProxy OptionsApplications > projet8 > EMR

https://eu-west-3.console.aws.amazon.com/emr/home?region=eu-west-3#/clusterDetails/j-2DOLDXU9GN7AS

ServicesRechercher[Alt+S]

EMR S3 EC2

ParisBmaria915B

PropriétésActions d'amorçageInstances (Matériel)ÉtapesApplicationsConfigurationsSurveillanceÉvènementsIdentifications (1)

Interfaces utilisateur d'application Info

Les applications installées sur votre cluster Amazon EMR publient des interfaces utilisateur en tant que sites web. Vous pouvez les utiliser pour surveiller l'activité du cluster.

☒ Interfaces utilisateur d'application sur le cluster

Les interfaces utilisateur sur le cluster sont disponibles uniquement pendant l'exécution de votre cluster. Utilisez les liens suivants pour démarrer. Pour accéder à toutes les interfaces utilisateur d'application, configurez le tunneling SSH.

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 persistantes

Les interfaces utilisateur persistantes ne nécessitent pas de tunneling SSH. Elles sont hébergées hors du cluster et sont disponibles pendant 30 jours après la fin d'une application.

Interfaces utilisateur d'application sur le nœud primaire

Celles-ci nécessitent l'activation du tunneling SSH.

Activer une connexion SSH

Application	URL de l'interface utilisateur
Gestionnaire de ressources	http://ec2-15-237-43-41.eu-west-3.compute.amazonaws.com:8088/
JupyterHub	http://ec2-15-237-43-41.eu-west-3.compute.amazonaws.com:9443/
Livy	http://ec2-15-237-43-41.eu-west-3.compute.amazonaws.com:8998/
Nom du nœud HDFS	http://ec2-15-237-43-41.eu-west-3.compute.amazonaws.com:9870/
Serveur d'historique Spark	http://ec2-15-237-43-41.eu-west-3.compute.amazonaws.com:18080/

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

Application	URL de l'interface utilisateur
Gestionnaire de nœuds	http://ec2-000-000-000-000.compute-1.amazonaws.com:8042/
Nœud de données HDFS	http://ec2-000-000-000-000.compute-1.amazonaws.com:9864/

Applications installées (7)

CloudShell Commentaires

© 2023, Amazon Web Services, Inc. ou ses affiliés. Confidentialité Conditions Préférences relatives aux cookies

20

aws Services Rechercher [Alt+S]

EMR S3 EC2 Billing

Amazon S3

Compartmentiments

- 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
- Groupes Storage Lens **Nouveau**
- Paramètres AWS Organizations

Fonctionnalité spot

► AWS Marketplace pour S3

Amazon S3 > Compartiments

► **Instantané de compte**

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

Compartiments (2) [Info](#)

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

Rechercher des compartiments par nom

	Nom	Région AWS	Accéder
<input type="radio"/>	aws-logs-299252609028-eu-west-3	Europe (Paris) eu-west-3	Compartiment et objets non publics
<input type="radio"/>	fruit-project	Europe (Paris) eu-west-3	Compartiment et objets non publics



Objets (4)

Les objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez utiliser l'[inventaire Amazon S3](#) pour obtenir une liste.

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

Rechercher des objets en fonction du préfixe ☐ Afficher les versions

<input type="checkbox"/>	Nom	Type
<input type="checkbox"/>	bootstrap-emr.sh	sh
<input type="checkbox"/>	Bouchehboun_maria_1_notebook_1123_.ipynb	ipynb
<input type="checkbox"/>	Resultat/	Dossier
<input type="checkbox"/>	Test/	Dossier

Objets (79)

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.

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

Rechercher des objets en fonction du préfixe

☐ Afficher les versions

<input type="checkbox"/>	Nom	Type	Dernière mod
<input type="checkbox"/>	Apple Braeburn/	Dossier	-
<input type="checkbox"/>	Apple Crimson Snow/	Dossier	-
<input type="checkbox"/>	Apple Golden 1/	Dossier	-
<input checked="" type="checkbox"/>	Apple Golden 2/	Dossier	-
<input type="checkbox"/>	Apple Golden 3/	Dossier	-
<input type="checkbox"/>	Apple Granny Smith/	Dossier	-
<input type="checkbox"/>	Apple Pink Lady/	Dossier	-
<input type="checkbox"/>	Apple Red 1/	Dossier	-
<input type="checkbox"/>	Apple Red 2/	Dossier	-

Services [Alt+S]

EMR S3 EC2 Billing

Amazon S3 > Compartiments > fruit-project > Test/ > Apple Golden 2/ > 322_100.jpg

322_100.jpg [Info](#)

Propriétés

Autorisations

Versions

Présentation de l'objet

Propriétaire
e9d323801790be9577517262f94c55ad70ff365c8b8bbc598e34e60c18b577de

Région AWS
Europe (Paris) eu-west-3

Dernière modification
21 Nov 2023 09:31:51 PM CET

Taille
4.1 Ko

Type
jpg

Clé
Test/Apple Golden 2/322_100.jpg

URI S3

s3://fruit-project/Test/Apple Golden 2/322_100.jpg

Amazon Resource Name (ARN)

arn:aws:s3:::fruit-project/Test/Apple Golden 2/322_100.jpg

Balise d'entité (Etag)

<94d687ca5ae68da95a2f6d9ce254849e>

URL de l'objet

https://fruit-project.s3.eu-west-3.amazonaws.com/Test/Apple+Golden+2/322_100.jpg

Objets (164)

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.

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

Rechercher des objets en fonction du préfixe

☐ Afficher les versions

<input type="checkbox"/>	Nom	Type
<input type="checkbox"/>	3_100.jpg	jpg
<input type="checkbox"/>	32_100.jpg	jpg
<input type="checkbox"/>	321_100.jpg	jpg
<input type="checkbox"/>	322_100.jpg	jpg
<input type="checkbox"/>	323_100.jpg	jpg
<input type="checkbox"/>	324_100.jpg	jpg
<input type="checkbox"/>	325_100.jpg	jpg





thanks!

