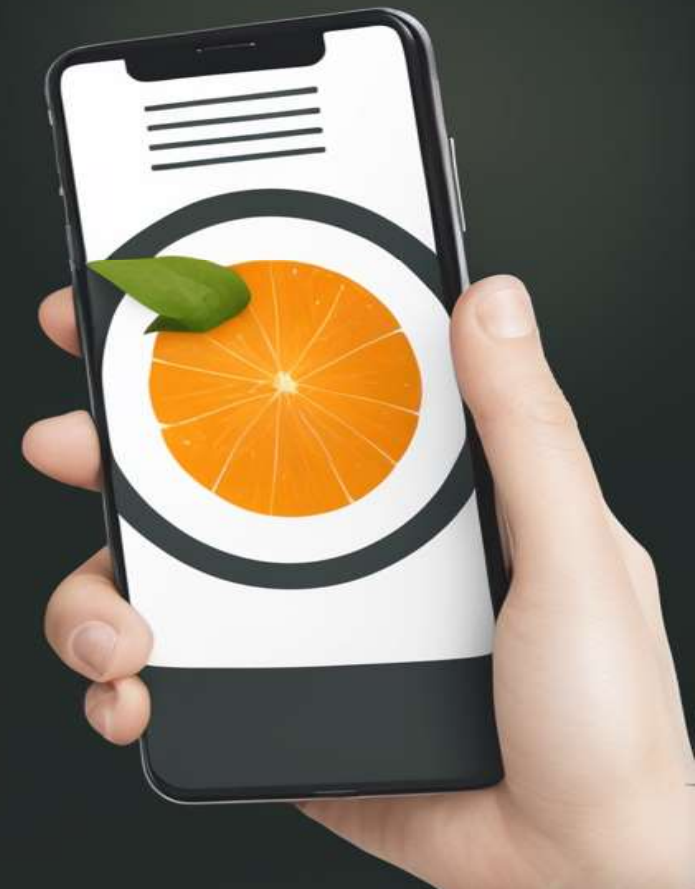


# Fruits!

Réalisation d'un traitement dans un environnement Big Data sur le Cloud



A close-up photograph of a variety of fresh tropical fruits. In the center, two pineapples with their green crowns are prominent. To the left, a bunch of yellow bananas is visible. Surrounding these are several bright orange oranges, yellow lemons, and a few kiwis with their brown, fuzzy skin. A single mango is also visible on the right side. The fruits are piled together, creating a colorful and textured background.

# Problématiques



Récolte de fruits automatisée

Se faire connaître

Application mobile de  
reconnaissance de fruits



Moteur classification fruits

Mise en place architecture Big Data

Chaîne de traitement des données





# Jeu de données



Kernel Kaggle 2017  
94000 images  
140 classes  
2 jeux de données  
75% Train / 25% Test

140 répertoires fruits, légumes  
Photos 360°  
100 x 100 pixels RGB  
Variétés vues comme classes différentes





# Création de l'environnement



# Big Data

Gestion volume données important

Variations d'échelles rapides & automatiques

Environnement virtuel





Capacité de calcul

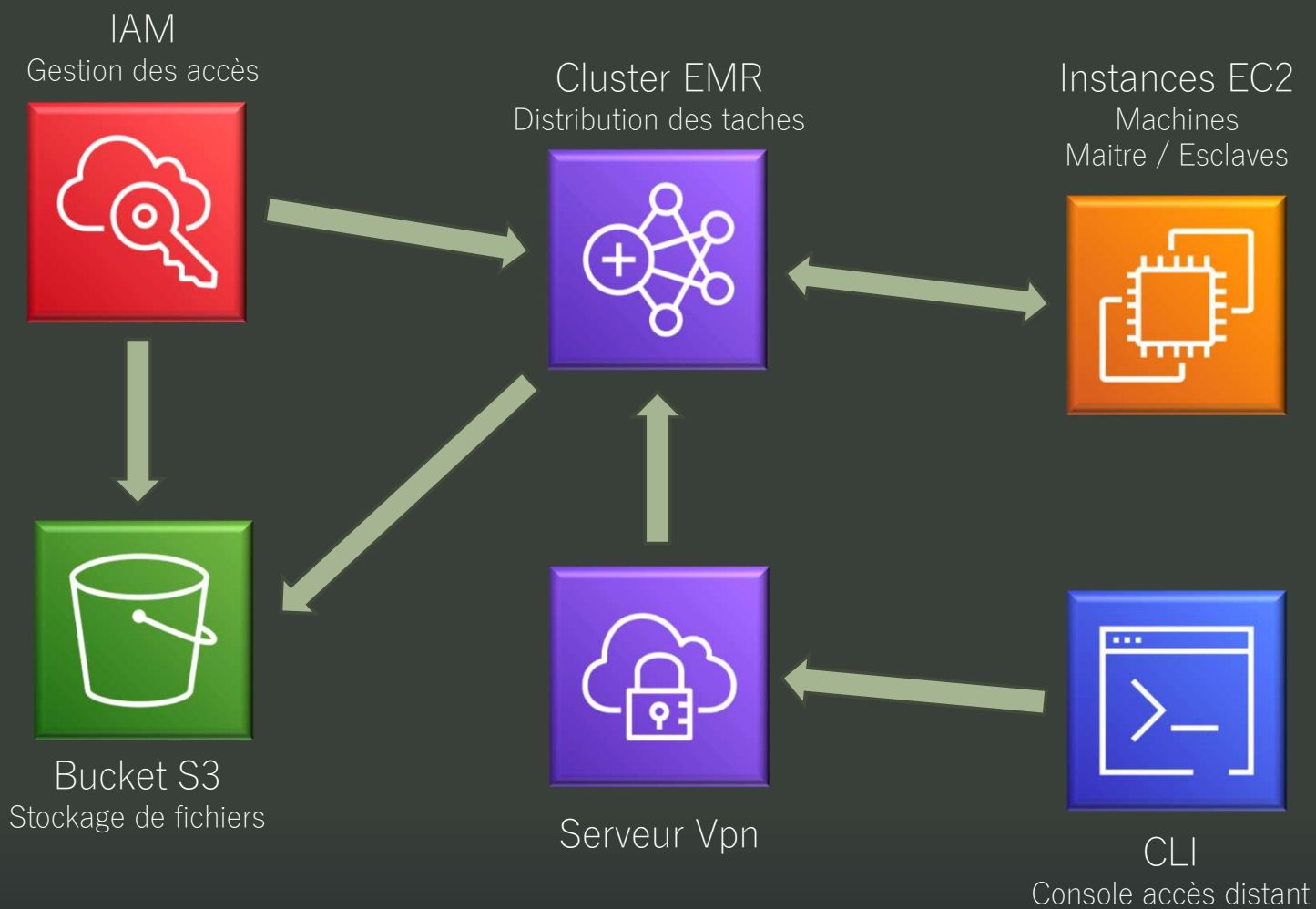
Capacité mémoire

Automatisation



Stockage illimité

Conforme RGPD  
Datacenter en Europe





# Cluster EMR

## Configurations de cluster

Les configurations de cluster sont définies lorsque vous créez un cluster.

[Afficher JSON](#)

Filtrer la classification

Tout classification ▼

< 1 >

Classification ▼	Propriété ▼	Valeur ▼	Source ▼
jupyter-s3-conf	s3.persistence.bucket	ocp9-fr-data	Configurations de cluster
jupyter-s3-conf	s3.persistence.enabled	true	Configurations de cluster

## Configurations de groupe d'instances (6)

Les groupes d'instances héritent des configurations de cluster qui sont spécifiées lors de la création d'un cluster. Avec EMR 5.21.0 et les versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier d'autres classifications de configuration pour chaque groupe d'instances ci-dessous, tant que son statut est **En cours d'exécution**.

[Reconfigurer](#)[Afficher JSON](#)

Filtrer le groupe d'instances

Tout groupe d'instances ▼

Filtrer la classification

Toute classification ▼

< 1 >

	Classification ▼	Statut ▼	Propriété ▼	Valeur ▼	Source ▼
<input type="radio"/>	<input checked="" type="checkbox"/> Groupe d'instances Primary (ig-18HNKTCJODM4B)	✓ En cours d'exécution	-	-	-
	└ jupyter-s3-conf	-	s3.persistence.bucket	ocp9-fr-data	Configurations de cluster
	└ jupyter-s3-conf	-	s3.persistence.enabled	true	Configurations de cluster
<input type="radio"/>	<input checked="" type="checkbox"/> Groupe d'instances Core (ig-1EH7SYPTRHQ2X)	✓ En cours d'exécution	-	-	-
	└ jupyter-s3-conf	-	s3.persistence.bucket	ocp9-fr-data	Configurations de cluster
	└ jupyter-s3-conf	-	s3.persistence.enabled	true	Configurations de cluster



# Instances EC2

**Groupes d'instances (2)** [Info](#)

[Résilier l'instance](#) [Redimensionner le groupe d'instances](#) [Ajouter un groupe d'instances de tâches](#)

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.

< 1 >

Type et nom	ID	Statut <small>Motif du dernier changement</small>	Instances	Option d'achat et prix	Taille EBS (Gio)	ID d'instance EC2
<div><input type="radio"/></div> <div><input checked="" type="checkbox"/> <b>Primaire (Primary)</b></div>	ig-18HNKTCJODM4B	<div><input checked="" type="checkbox"/> En cours d'exécution</div>	1	À la demande	-	-
<div><input type="radio"/></div> <div>m5.xlarge <small>4 vCore, 16 Gio de mémoire, Stockage E</small></div>	ci-04633173G69JHGV...	<div><input checked="" type="checkbox"/> En cours d'exécution</div>	-	À la demande (0.224 USD/h)	64 (2 volumes)	<a href="#">i-036f7b79c8ce7b</a>
<div><input type="radio"/></div> <div><input checked="" type="checkbox"/> <b>Principal (Core)</b></div>	ig-1EH7SYPTRHQ2X	<div><input checked="" type="checkbox"/> En cours d'exécution</div>	2	À la demande	-	-
<div><input type="radio"/></div> <div>m5.xlarge <small>4 vCore, 16 Gio de mémoire, Stockage E</small></div>	ci-03009572A9G4J86...	<div><input checked="" type="checkbox"/> En cours d'exécution</div>	-	À la demande (0.224 USD/h)	64 (2 volumes)	<a href="#">i-05f052536527a</a>
<div><input type="radio"/></div> <div>m5.xlarge <small>4 vCore, 16 Gio de mémoire, Stockage E</small></div>	ci-0322686VFP5DPIR...	<div><input checked="" type="checkbox"/> En cours d'exécution</div>	-	À la demande (0.224 USD/h)	64 (2 volumes)	<a href="#">i-033c7c174a8379</a>





# Interfaces & Applications

## 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://ec2-13-38-69-192.eu-west-3.compute.amazonaws.com:8088/>

JupyterHub

<https://ec2-13-38-69-192.eu-west-3.compute.amazonaws.com:9443/>

Livy

<http://ec2-13-38-69-192.eu-west-3.compute.amazonaws.com:8998/>

Nom du nœud HDFS

<http://ec2-13-38-69-192.eu-west-3.compute.amazonaws.com:9870/>

Serveur d'historique Spark

<http://ec2-13-38-69-192.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)

Hadoop 3.4.0

Hive 3.1.3

JupyterEnterpriseGateway 2.6.0

JupyterHub 1.5.0

Livy 0.8.0

Spark 3.5.3

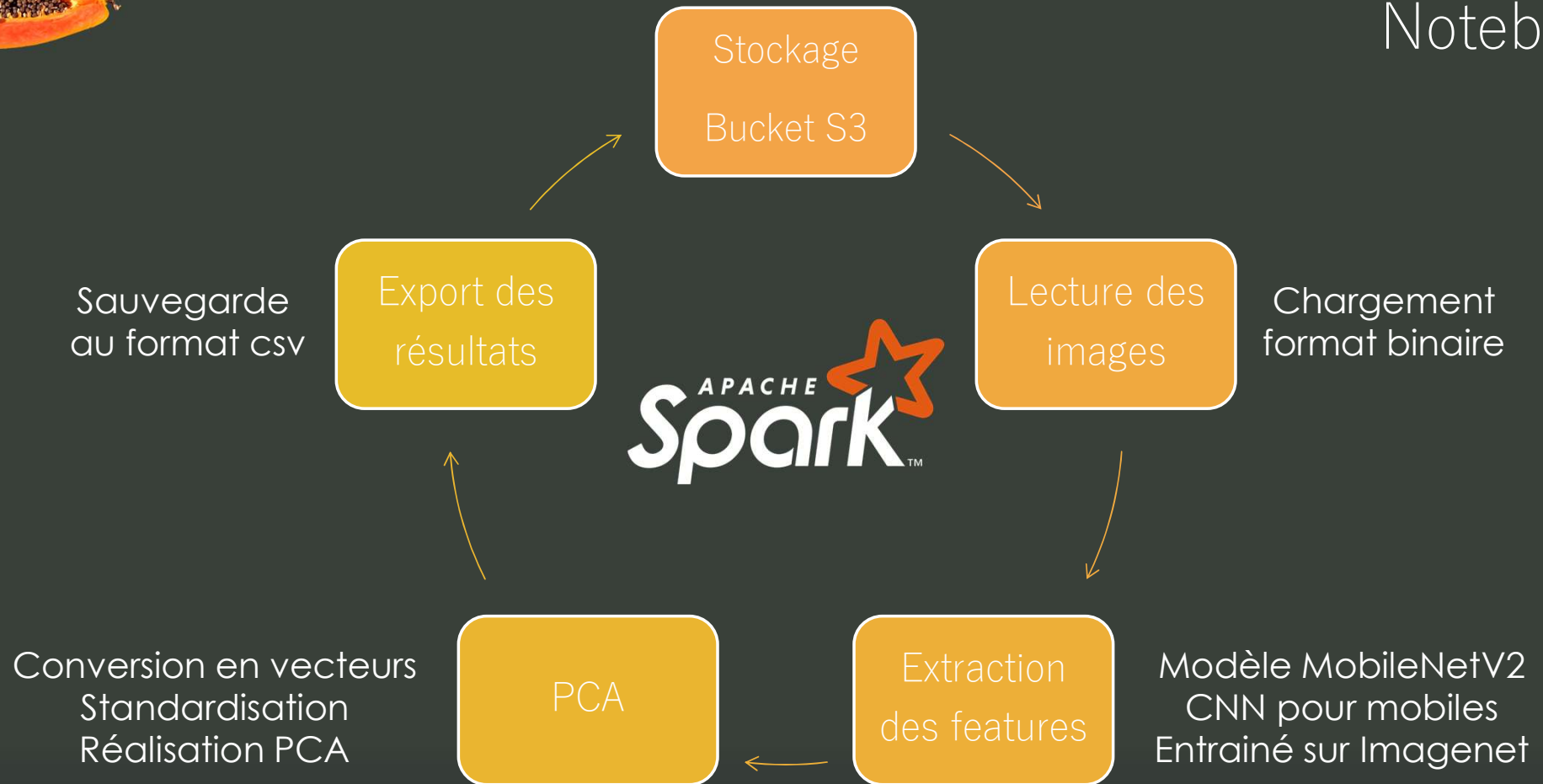
TensorFlow 2.16.1



# Chaine de traitement

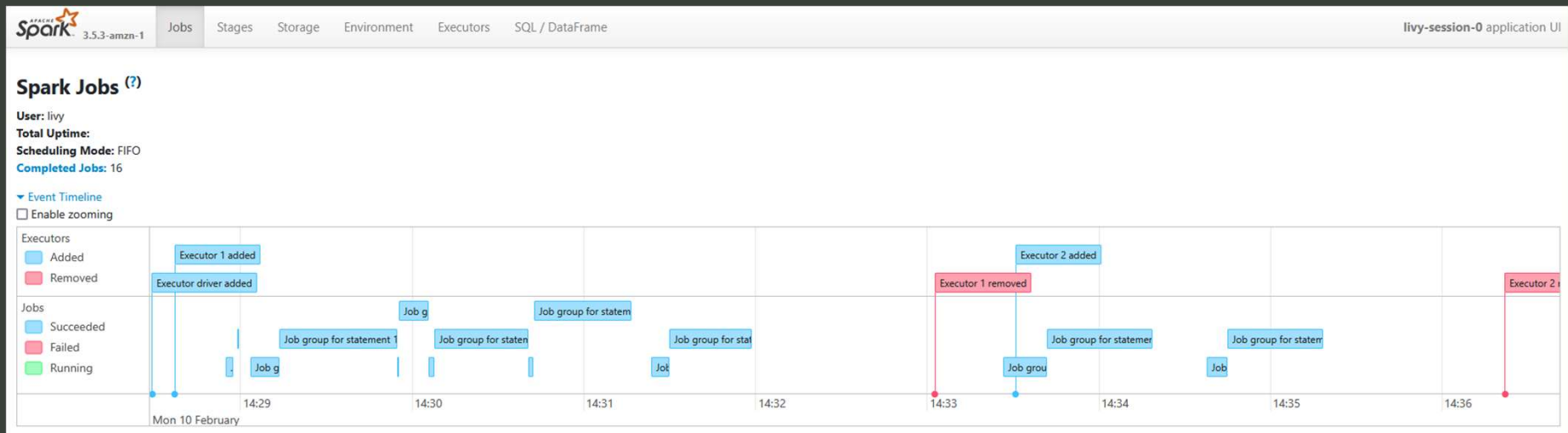


# Cluster EMR Notebook





# Vérification des Jobs Spark







# Démonstration



La démonstration en ligne  
suivra cette présentation

A small white ceramic bowl is filled with bright yellow, cubed fruit, likely mango, which is sprinkled with dark specks of seasoning. A single wedge of a yellow lemon is placed on the left side of the bowl. The bowl sits on a light-colored bamboo mat with a vertical grain pattern. The word "Conclusion" is written in a large, white, sans-serif font across the center of the bowl.

# Conclusion



Mise a l'échelle simplifiée

Rapidité mise en place & évolution

Optimisation des coûts





Pour la suite

Modèle final à choisir

Budget à fixer

Merci pour votre attention