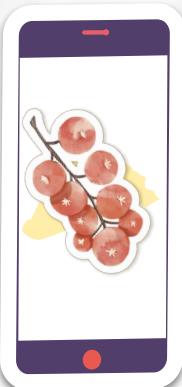




DÉPLOYEZ UN MODÈLE DANS LE CLOUD

Soutenance Projet 8
OpenClassrooms
Formation Data Scientist
13 Août 2023

PROBLÉMATIQUE :



Data Scientist



Application small Data

Passage
à
l'échelle



Data Architect



MISSION :

Construire et tester



Architecture
Big Data



DÉPLOYEZ UN MODÈLE - DANS LE CLOUD - DÉPLOYEZ UN MODÈLE - DANS LE CLOUD



DONNÉES D'ENTRÉE :



Kaggle

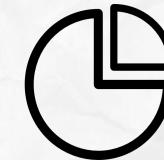
- 90k images
- 131 classes

Test

- 22k images

Training

- 68k images



Échantillon

- **60 images**
- **3 classes**



PLAN DE LA SOUTENANCE



01

Environnement
Big Data

02

Chaîne de
traitement des
images

03

Démonstration

04

Synthèse et
conclusion

01.

ENVIRONNEMENT BIG DATA



BRIQUES DE L'ENVIRONNEMENT :



Cluster de machines

Scalable



Lieux de stockage

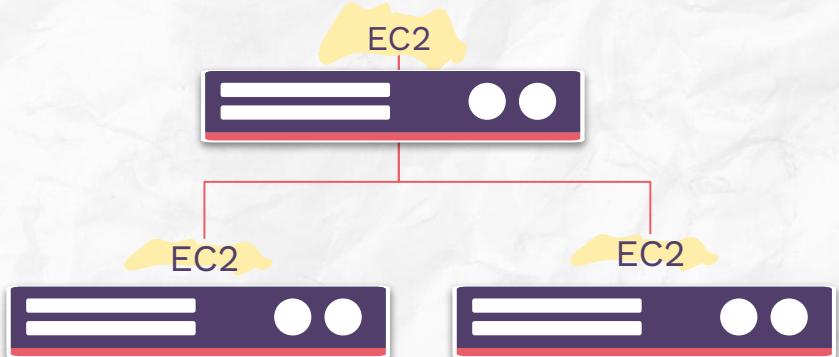
Scalables

FOURNISSEUR DE CLOUD



- Loue la puissance de calcul
- Délègue la maintenance

EMR - ELASTIC MAP REDUCE



Data centers à Paris

- Serveurs préconfigurés
- Installation et mise à jour des packages
- Passage à l'échelle

S3 - SIMPLE STORAGE SERVICE



- Stockage illimité
- Équilibre coût/temps de transfert
- Persistance des données
- Facilité d'accès

TARIFICATION À LA DEMANDE

- Aucun engagement
- Réduction des coûts
- Configuration de budgets



SÉCURITÉ



Utilisateur IAM

- Console AWS
- Ligne de commande AWS CLI

Accès Serveur

- SSH : clés privée/publique
- Ouverture vers mon IP uniquement

Bucket S3

- Non ouvert au public



02.

TRAITEMENT DES IMAGES

APPLICATION SPARK

Gérer et coordonner
le travail des
machines entre elles



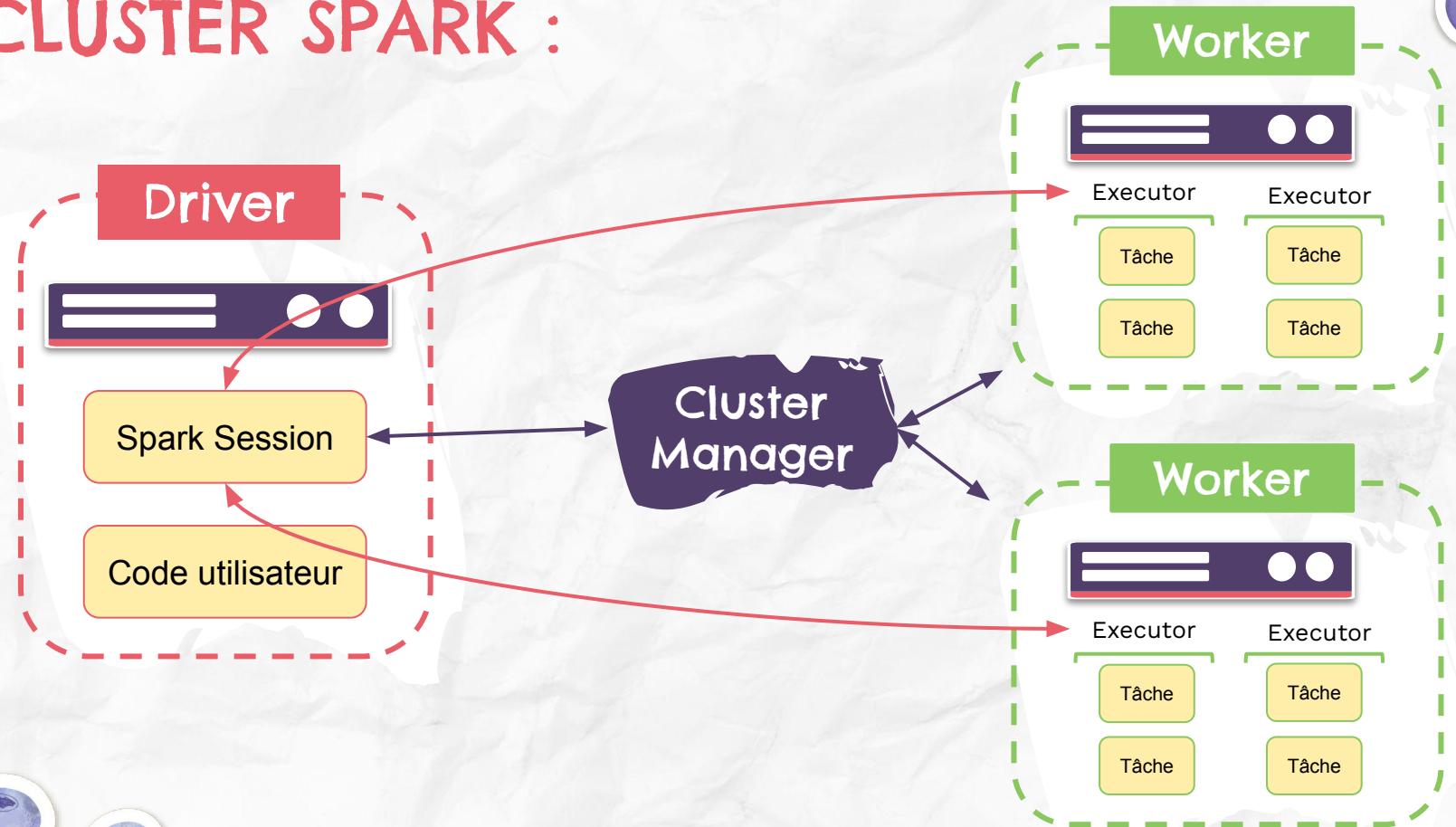
- Divers langages de programmation
- Simulation d'un cluster en local



CLUSTER SPARK :

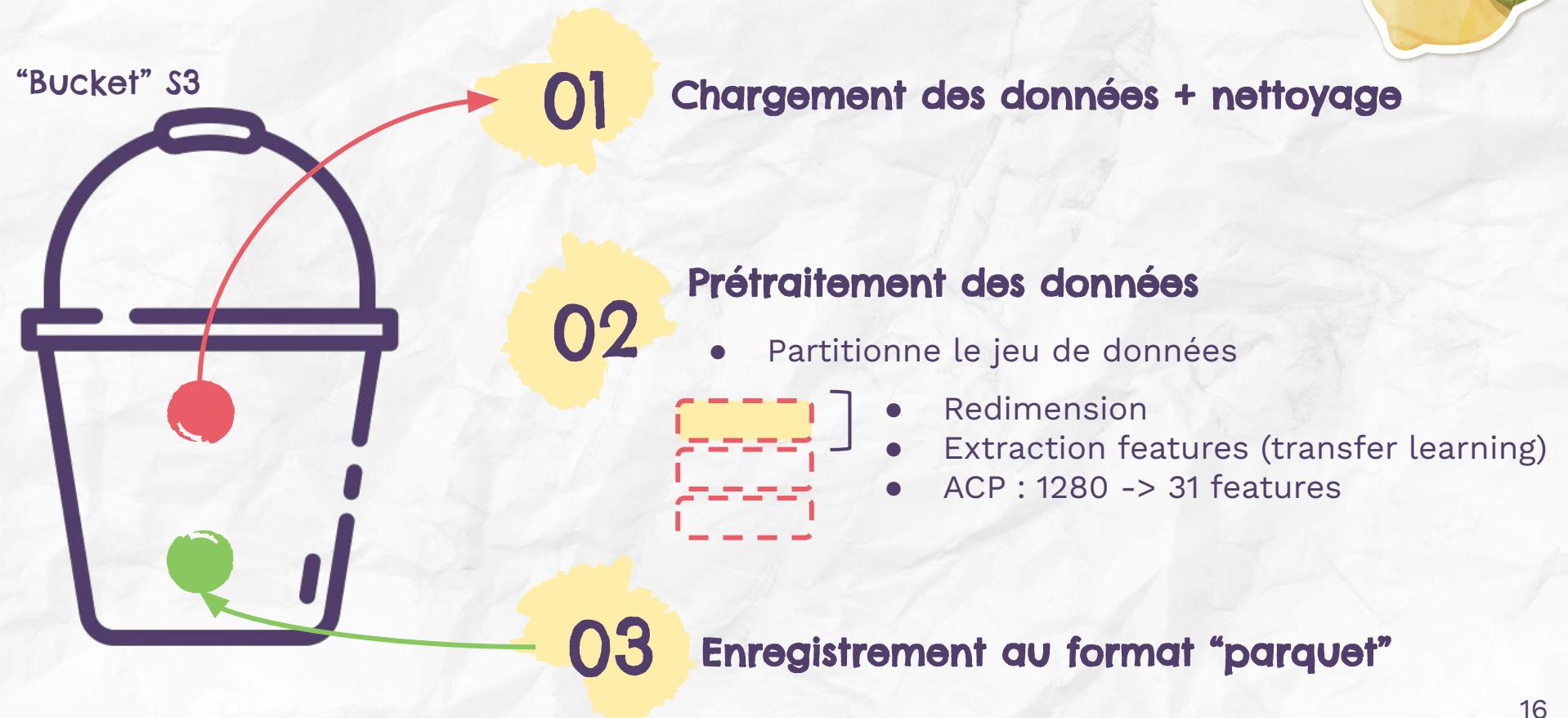


CLUSTER SPARK - CLUSTER SPARK - CLUSTER SPARK - CLUSTER SPARK - CLUSTER SPARK





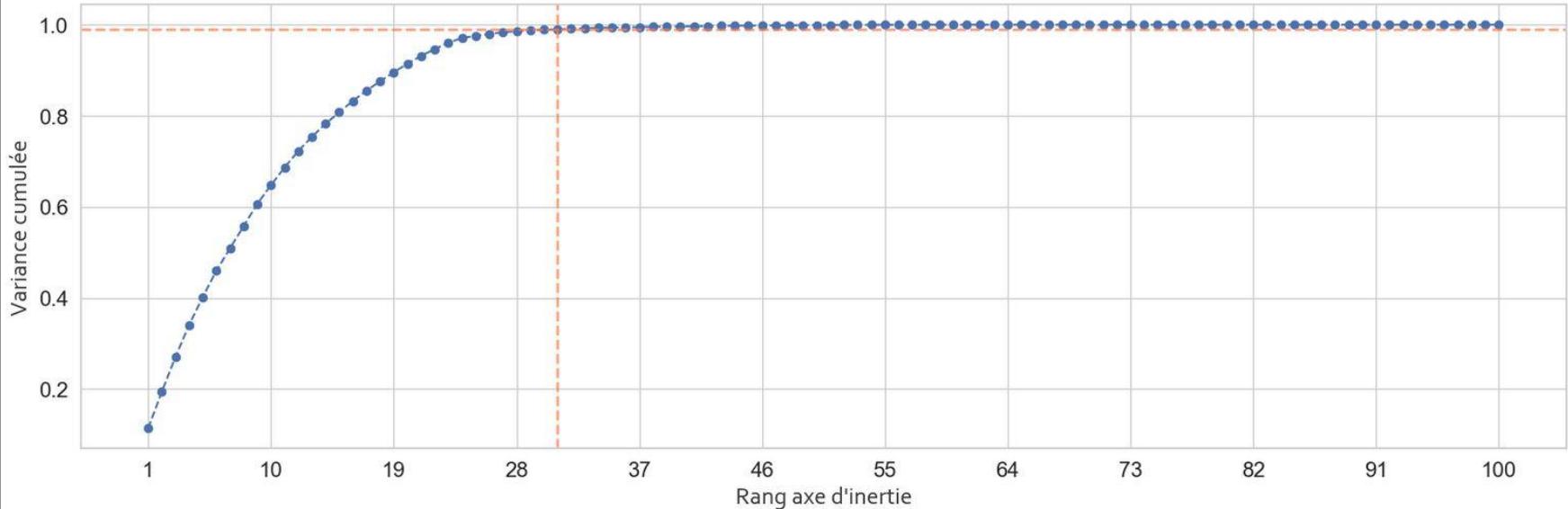
CHAÎNE DE TRAITEMENT





Les 31 premières composantes expliquent plus de 99% de la variance

Eboulis des valeurs propres cumulées



A decorative illustration in the bottom-left corner features a macadamia nut split open, showing its two halves, surrounded by several green, serrated leaves.

03.

DÉMONSTRATION



CONCLUSION

Une Solution

- Rapide
 - Facile
 - Optimisée
- 



mais pas parfaite



- Bugs techniques
- Machines non disponibles



Annexes



EMR - Console AWS



Services

Rechercher

[Alt+S]



Paris ▾

AdministratorAccess/projet8_user ▾

Amazon EMR X

Changer pour l'ancienne console

EMR Serverless [Nouveau](#)

EMR sur EC2

Clusters

Événements

Bloquer l'accès public

Configurations de sécurité

EMR sur EKS

Clusters virtuels

EMR Studio

Mise en route

Amazon EMR > EMR sur EC2: Clusters

Clusters (15) [Info](#)



Afficher les détails

Résilier

Cloner

Créer un cluster

Filtrer les clusters par statut

Rechercher des clusters

Filtrer les clusters par date et heure de création

< 1 >

	ID de cluster	Nom du cluster	Statut	Détails du statut
<input type="checkbox"/>	j-T5QQ3Z27ZWR7	OCP8_Fruits	Résilié	Demande utilisateur
<input type="checkbox"/>	j-3PDG8CZ6QV15W	OCP8_Fruits	Résilié	Demande utilisateur
<input type="checkbox"/>	j-2OI3S1IXA4Y8S	OCP8_Fruits	Résilié	Demande utilisateur
<input type="checkbox"/>	j-MYVIXG1DZEFV	OCP8_Fruits	Résilié	Demande utilisateur
<input type="checkbox"/>	j-29CHAGJ58ZRO	OCP8_Fruits	Résilié	Demande utilisateur
<input type="checkbox"/>	j-24B1CTD9NMSO3	OCP8_Fruits	Résilié avec des erreurs	Erreur interne
<input type="checkbox"/>	j-2BVJ8637R6OUH	OCP8_Fruits	Résilié avec des erreurs	Erreur interne
<input type="checkbox"/>	j-2424ERNF3K4JD	OCP8_Fruits	Résilié avec des erreurs	Erreur interne
<input type="checkbox"/>	i-1NFI4ZGU58K76	OCP8_Fruits	Résilié	Demande utilisateur



S3 - Compartiments



Compartiments (1) [Infos](#)

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

Copier le contenu Vider Supprimer **Créer un compartiment**

Rechercher des compartimente

< 1 >

Nom	Région AWS	Accéder	Date de création
projet8-oc-082023-data	EU (Paris) eu-west-3	Compartiment et objets non publics	01 Aug 2023 10:25:13 AM CEST

S3 - Contenu du compartiment

Objets (6)

Les objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez accéder à 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](#)

Actions ▾ Copier l'URI S3 Copier l'URL Charger

Rechercher des objets en fonction

Nom	Type	Dernière modification
j-MYVIXG1DZEFV/	Dossier	-
j-T5QQ3Z27ZWR7/	Dossier	-
Results/	Dossier	-
Test/	Dossier	-

Objets (25)

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](#)

C Copier l'URI S3 Copier l'URL Télécharger Ouvrir Supprimer

Actions ▾ Crée un dossier Charger

Rechercher des objets en fonction

Nom	Type	Dernière modification	Taille	Classe de stockage
_SUCCESS	-	03 Aug 2023 04:02:49 PM CEST	0 o	Standard
part-00000-a56ac840-7383-453e-9e09-69d72375baa	parquet	03 Aug 2023 04:02:36 PM CEST	2.2 Ko	Standard

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](#)

C Copier l'URI S3 Copier l'URL Télécharger Ouvrir Supprimer

Actions ▾ Crée un dossier Charger

Rechercher des objets en fonction

Nom	Type	Dernière modification	Taille	Classe de stockage
Apple Braeburn/	Dossier	-	-	-
Mango/	Dossier	-	-	-
Tomato Cherry Red/	Dossier	-	-	-

Thanks!

This presentation template was created by **Slidesgo**, including infographics & images by **Freepik** and icons by **Flaticon** :

- Icônes “Stockage en ligne” créées par Konkapp
- Icônes “Informatique en cluster” créées par Vectors Tank
- Icônes “Seau-deau” créées par Good Ware



