

Projet 8 : Déployer un modèle dans le cloud

Jean Eudes des Graviers

Executive summary

1. La littérature montre qu'il y a différents modèles de transfert learning qui performant sur la classification d'image de fruit. AlexNet présente une architecture peu profonde et donc moins coûteuse en temps de calcul. L'approche de réseau de neurone from scratch présente aussi de bonnes performances.

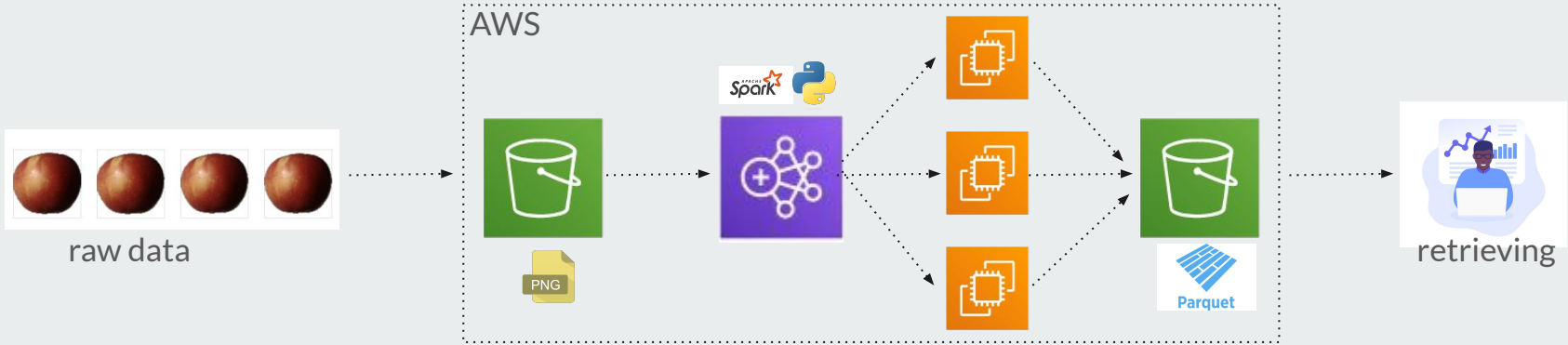
Table 5. Comparison between the proposed example and pre-trained models for fruit classification on Fruit-360 dataset [70].

CNN Model	Depth	Training		Validation		Testing	
		Loss	Accuracy	Loss	Accuracy	Accuracy	F1-Score
Proposed example	6	0.0167	100%	0.0165	100%	95.45%	0.96
AlexNet	8	1.2e-5	100%	4.9e-5	100%	100%	1
VGG16	16	0.2067	96.32%	0.2070	95.94%	91.32%	0.89
MobileNet	88	0.0799	97.45%	0.7201	73.86%	70.02%	0.67
InceptionV3	159	0.5592	80.49%	1.1755	62.82%	54.49%	0.49
ResNet50	168	0.1436	99.17%	0.9102	66.69%	57.74%	0.48

2. La librairie pyspark permet un traitement rapide et évolutif des données avec une exécution en mémoire et une évolutivité horizontale. Il permet aussi une gestion facilitée grâce à son interface UI pour le tracking des expériences ainsi que son intégration avec AWS et EMR.

Job group for statement 15 parquet at NativeMethodAccessorImpl.java:0	2024/03/05 10:20:57	21 s
Job group for statement 15 parquet at NativeMethodAccessorImpl.java:0	2024/03/05 10:12:35	8.3 min
Job group for statement 15 parquet at NativeMethodAccessorImpl.java:0	2024/03/05 10:00:41	12 min
Job group for statement 6 showString at NativeMethodAccessorImpl.java:0	2024/03/05 09:58:55	0.2 s
Job group for statement 5 showString at NativeMethodAccessorImpl.java:0	2024/03/05 09:58:51	2 s
Listing leaf files and directories for 131 paths: s3://p8-data-fruits/Test/Apple Braeburn, ... load at NativeMethodAccessorImpl.java:0	2024/03/05 09:58:34	13 s

3. Nos données seront stockés dans le S3 d'AWS au format PNG puis au format Parquet après le processing effectué. La puissance de calcul sera répartie sur des machines EC2 à travers un EMR et actionnée grâce à un script python avec la librairies PySpark.



Sommaire

1. Contexte du projet
2. Présentation du jeu de donnée
3. Approche de modélisation
4. Déploiement sur le cloud
5. Conclusions et axes d'amélioration

I. Contexte du projet

"Fruits", une très jeune start-up de l'AgriTech, cherche à proposer des solutions innovantes pour la récolte des fruits.

Votre start-up 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 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.



II. Présentation du jeu de donnée et prétraitement

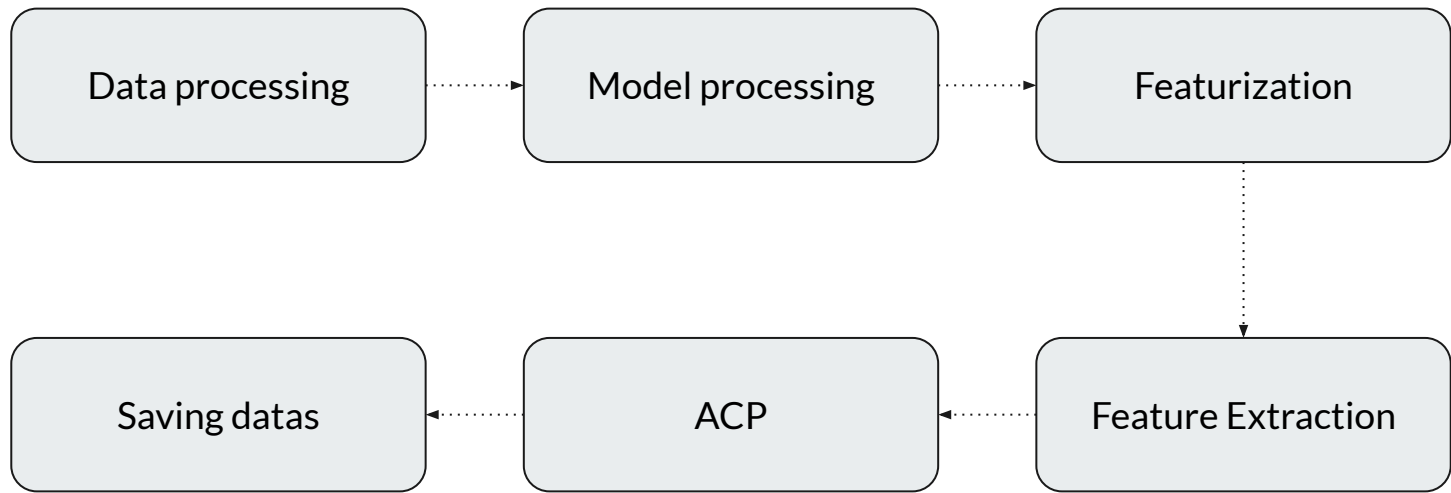
90 000 + images de fruits, 131 catégories de fruits. Point de vue 360°.

L'entraînement se fera en prenant aléatoirement 10 images par catégorie.

```
The folder '/content/drive/MyDrive/fruits/fruits-360_dataset/fruits-360/Training/Plum 2' contains 420 files.  
The folder '/content/drive/MyDrive/fruits/fruits-360_dataset/fruits-360/Training/Lemon Meyer' contains 490 files.  
The folder '/content/drive/MyDrive/fruits/fruits-360_dataset/fruits-360/Training/Grape Blue' contains 984 files.  
The folder '/content/drive/MyDrive/fruits/fruits-360_dataset/fruits-360/Training/Apple Red Yellow 2' contains 672 files.  
...
```



III. Approche de modélisation

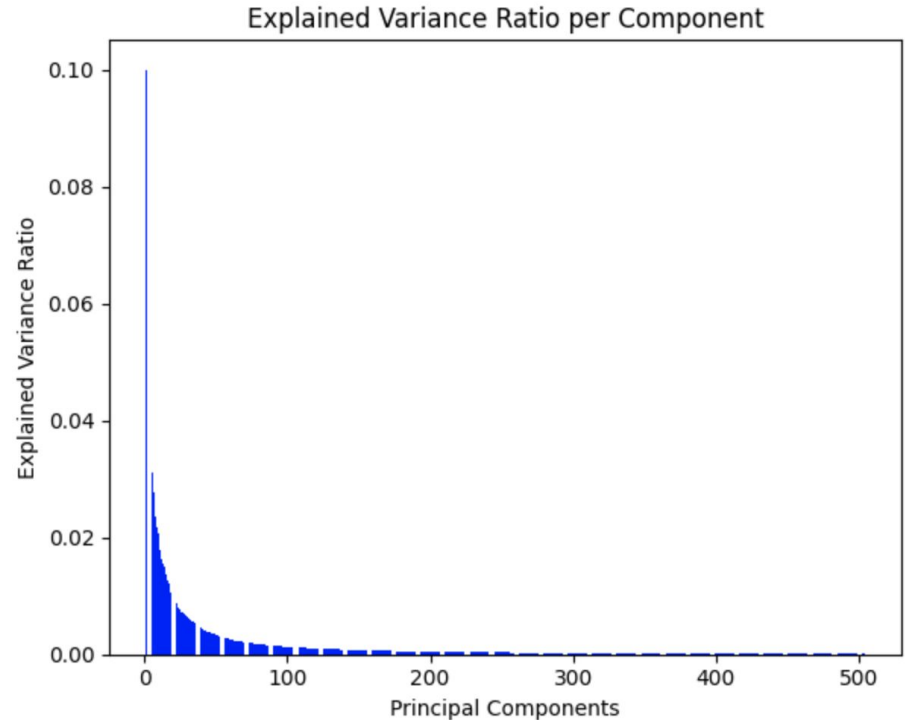


	path	label	features
0	file:/content/drive/MyDrive/fruits/fruits-360_...	Raspberry	[0.0013951088, 0.19356993, 0.30124167, 0.0, 0....
1	file:/content/drive/MyDrive/fruits/fruits-360_...	Pineapple Mini	[0.024196839, 4.4734335, 0.25085992, 0.0, 0.00...
2	file:/content/drive/MyDrive/fruits/fruits-360_...	Rambutan	[0.0, 3.937865, 0.0, 1.4568828, 0.26866707, 0....
3	file:/content/drive/MyDrive/fruits/fruits-360_...	Limes	[0.5705491, 0.0, 0.08695382, 0.0, 0.0, 0.23462...
4	file:/content/drive/MyDrive/fruits/fruits-360_...	Limes	[0.0019624282, 0.0, 0.5263183, 0.0, 0.0, 0.0, ...



ACP réduction de dimension

N composants	variance expliquée
1000	99.6%
800	99.1%
600	98%
400	95%
200	90.7%
100	83.1%



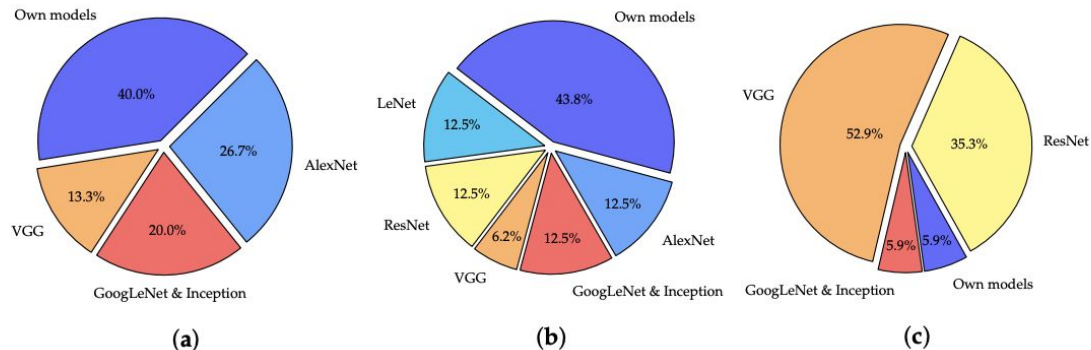


Figure 8. Distribution of CNN architectures used for: (a) fruit classification, (b) fruit quality control, and (c) fruit detection for automatic harvest.

Table 5. Comparison between the proposed example and pre-trained models for fruit classification on Fruit-360 dataset [70].

CNN Model	Depth	Training		Validation		Testing	
		Loss	Accuracy	Loss	Accuracy	Accuracy	F1-Score
Proposed example	6	0.0167	100%	0.0165	100%	95.45%	0.96
AlexNet	8	1.2e-5	100%	4.9e-5	100%	100%	1
VGG16	16	0.2067	96.32%	0.2070	95.94%	91.32%	0.89
MobileNet	88	0.0799	97.45%	0.7201	73.86%	70.02%	0.67
InceptionV3	159	0.5592	80.49%	1.1755	62.82%	54.49%	0.49
ResNet50	168	0.1436	99.17%	0.9102	66.69%	57.74%	0.48

IV. Déploiement sur le cloud

Parallélisation des calculs : PySpark



Apache Spark est un moteur de traitement de données rapide et polyvalent.

- **Traitement Rapide et Évolutif** des données avec une exécution en mémoire et une évolutivité horizontale.
- **Traitement Polyvalent** : Prise en charge du traitement de données par lots, du traitement en temps réel, et du machine learning.
- **Integration et Gestion Simplifiée** : Utilisez des clusters managés comme Amazon EMR pour simplifier la gestion opérationnelle.
- **Interface UI** pour le tracking des traitement de données

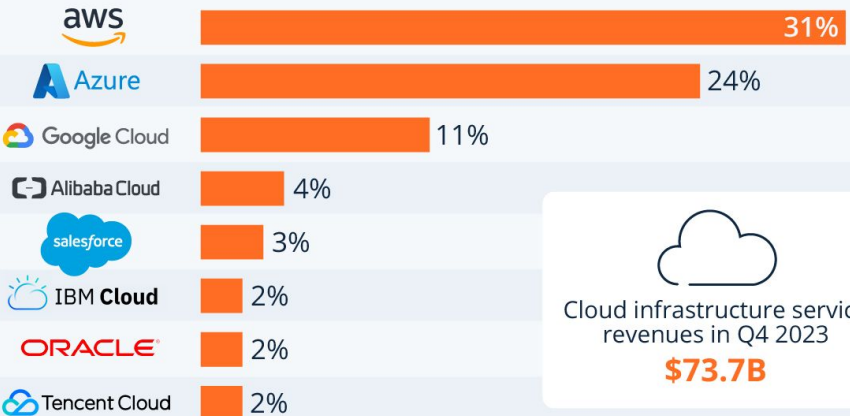
3 (15)	Job group for statement 15 parquet at NativeMethodAccessorImpl.java:0	2024/03/05 10:00:41	12 min	1/1	709/709
2 (6)	Job group for statement 6 showString at NativeMethodAccessorImpl.java:0	2024/03/05 09:58:55	0.2 s	1/1	1/1
1 (5)	Job group for statement 5 showString at NativeMethodAccessorImpl.java:0	2024/03/05 09:58:51	2 s	1/1	1/1
0 (4)	Listing leaf files and directories for 131 paths: s3://p8-data-fruits/Test/Apple Braeburn, ... load at NativeMethodAccessorImpl.java:0	2024/03/05 09:58:34	13 s	1/1	131/131

Prestataire : AWS



Amazon Maintains Cloud Lead as Microsoft Edges Closer

Worldwide market share of leading cloud infrastructure service providers in Q4 2023*



Cloud infrastructure service
revenues in Q4 2023

\$73.7B

Solution technique : EMR - EC2



Amazon
EMR



Amazon
EC2

AWS EMR (Elastic MapReduce) est un service géré pour le traitement de données distribuées.

EC2 (Elastic Compute Cloud) offre des ressources de calcul virtuelles évolutives.







- Adaptez rapidement la capacité de calcul en fonction du volume de données.
- Intégration avec l'Écosystème Big Data (Apache Spark, Hadoop).
- Sécurité et Conformité fonctionnalités de sécurité d'AWS et des options de chiffrement.
- Coût Optimisé pour les ressources utilisées, sans frais initiaux.
- Se connecte facilement avec des services comme S3.

Solution de stockage : S3



AWS S3 (Simple Storage Service) est un service de stockage d'objets évolutif et sécurisé dans le cloud.

- **Stockez un nombre illimité d'objets, Stockage Polyvalent**
- **Sécurisé** : option de chiffrement, politiques d'accès
- Payez uniquement pour l'espace de stockage réellement utilisé.
- **Intégration** avec d'autres services tels que EC2, EMR, et Lambda.

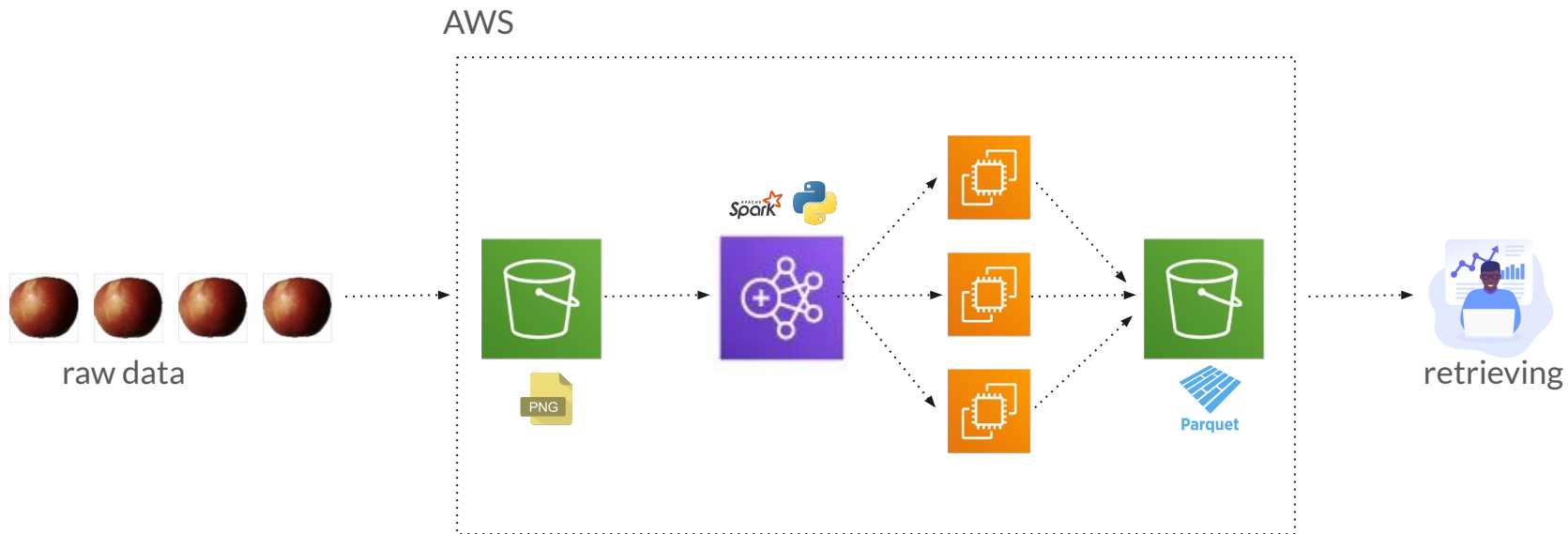
<input type="checkbox"/>	Nom ▲	Type ▼	Dernière modification ▼	Taille ▼	Classe de stockage ▼
<input type="checkbox"/>	 bootstrap-tensorflow.sh	sh	15 Feb 2024 11:29:50 AM CET	371.0 o	Standard
<input type="checkbox"/>	 EMR/	Dossier	-	-	-
<input type="checkbox"/>	 P8_Notebook_Linux_EMR_PySpark_V1.0-Copy1.ipynb	ipynb	06 Feb 2024 05:49:43 PM CET	372.9 Ko	Standard
<input type="checkbox"/>	 Reduced-matrix/	Dossier	-	-	-
<input type="checkbox"/>	 Results/	Dossier	-	-	-
<input type="checkbox"/>	 Test/	Dossier	-	-	-

RGPD

Instance type	Location	coût (/heure)
m5.xlarge	Paris	0.224 usd
m5.xlarge	Irlande	0.214 usd
m5.xlarge	Francfort	0.23 usd
m5.xlarge	Londres	0.222 usd
m5.xlarge	Stockholm	0.204 usd

Europe Paris eu-west-3





IV. Conclusion et axes d'amélioration

- Choix du modèle en fonction de la nature du problème
- Gestion de la puissance de calcul et des coûts
- Travail en collaboration et gestion des risques
- Autres fonctionnalités AWS à tester