

Déployez un modèle dans le cloud

Ilaria Mereu



Canestra di Frutta - Caravaggio

Mission

Notre client: Fruits! Start-up de l'AgriTech

Nos objectifs:

Réaliser un moteur de classification d'images de fruits entièrement déployé dans le cloud.

Nous développerons un prototype de ce moteur:

- structuré selon une architecture Big Data
- RGPD compliant
- Utilisant Pyspark (adapté à un passage à l'échelle)



Fruits!

Sommaire

1. Description de la problématique
2. La plateforme et ses configurations
3. Dataset
4. Résultats
5. Conclusions



Fruits!

I. Description de la problématique

Description de la problématique

Nous disposons d'un jeu d'images de 131 fruits et légumes classifiées sous le nom du fruit même. Exemple:

Apple Braeburn	Grape Blue	Pear Monster
Apple Crimson Snow	Grape Pink	Pear Red
Apple Golden 1	Grape White	Pear Stone
Apple Golden 2	Grape White 2	Pear Williams
Apple Golden 3	Grape White 3	Pepino

Chaque dossier ainsi nommé contient $\sim 10^2$ images.

Par exemple, dans le dossier “Pomegranate” (grenade) nous trouvons ~ 500 images de ce type:



À partir de ce jeu de données, notre objectif est:

- prédisposer un modèle de type CNN afin qu'il reconnaisse et classifie d'autres images de fruits.
- déployer ce modèle dans une infrastructure Big Data.



Fruits!

2. La plateforme et ses configurations

Déployer dans le cloud

Notre infrastructure Big Data doit se composer de :

1. une architecture de ressources de calcul (EMR)
2. un système d'accès et de exploitation de ces ressources (CLI ou UI)
3. un système de stockage à accès contrôlé (S3)
4. des politiques d'accès pour les utilisateurs (IAM)

Nous avons utilisé la plateforme Amazon Web Services.



Fruits!

Scalabilité horizontale

Déployer dans le cloud

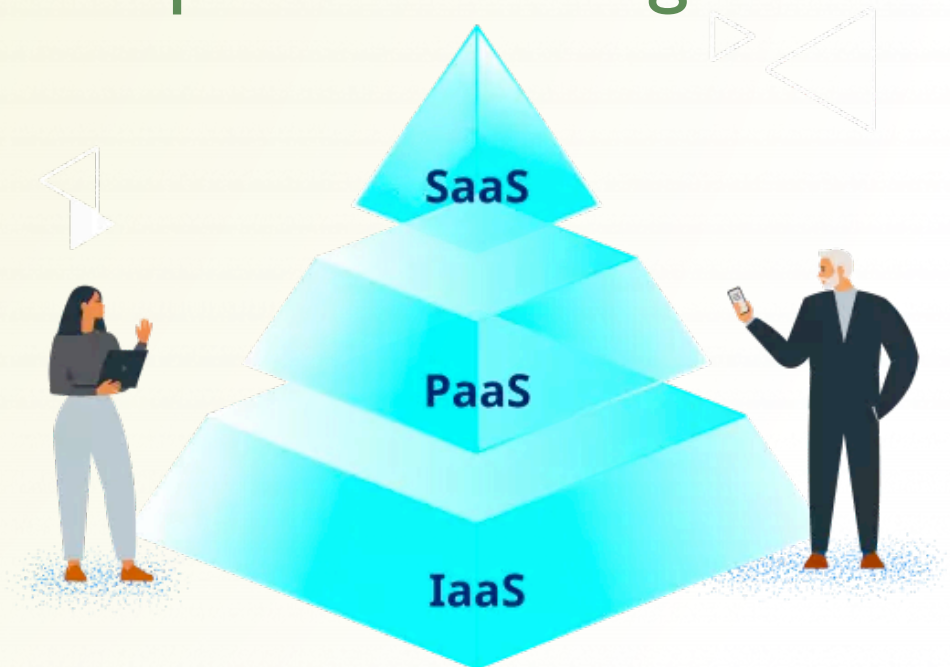
AWS et les autres plateformes concurrentes proposent différentes solutions qui se différencient pour:

- une liberté variable de configurer les ressources allouées
- une offre de solutions déjà configurées et prêtes à l'usage.

1. IaaS

2. PaaS

3. SaaS



Au sein de AWS, nous avons développé notre projet sous forme d'instance notebook ml.t3.medium sur AWS Sagemaker.



Fruits!

Big Data

Valeur

Value



Clinically relevant data
Longitudinal studies

Volume

Volume



High-throughput technologies
Continuous monitoring of vital signs

Vitesse

Velocity



High-speed processing for fast clinical decision support
Increasing data generation rate by the health infrastructure

Variété

Variety



Heterogeneous and unstructured data sources
Differences in frequencies and taxonomies

Véridicité

Veracity



Data quality is unreliable
Data coming from uncontrolled environments

Variabilité

Variability



Seasonal health effects and disease evolution
Non-deterministic models of illness and health

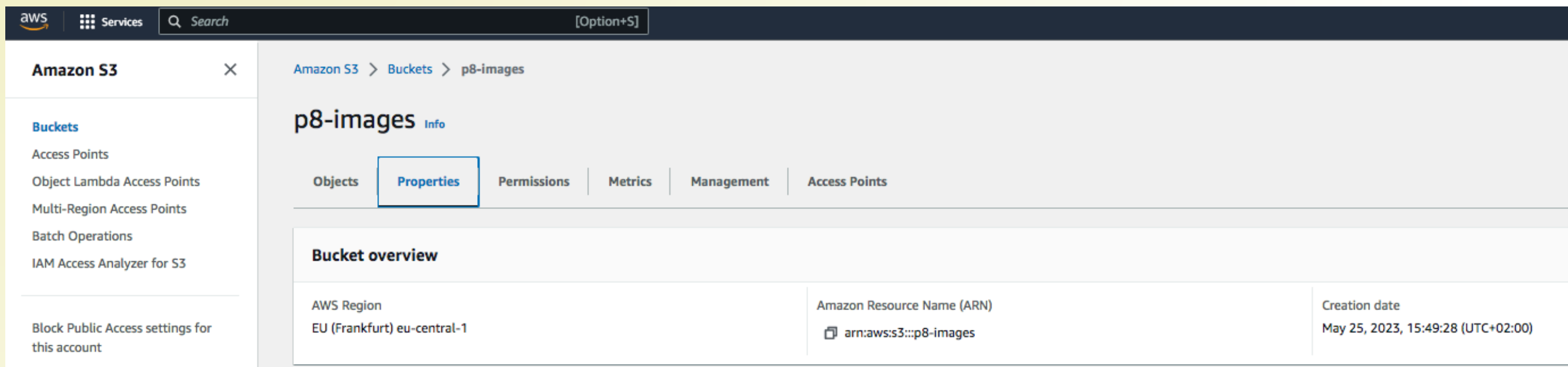


Fruits!

Dataset

Dataset

Nous avons utilisé un dataset réduit à cinq images pour cinq espèces. Il est stocké sur *AWS S3*, région *eu-central-1*.



The screenshot displays the AWS Management Console interface for an Amazon S3 bucket named 'p8-images'. The left sidebar shows the 'Amazon S3' service with a 'Buckets' link. The main content area shows the 'p8-images' bucket with tabs for 'Objects', 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'. The 'Properties' tab is selected, showing a 'Bucket overview' section with the following details:

Property	Value
AWS Region	EU (Frankfurt) eu-central-1
Amazon Resource Name (ARN)	arn:aws:s3:::p8-images
Creation date	May 25, 2023, 15:49:28 (UTC+02:00)



Fruits!

Solution retenue et résultats

Instance Sagemaker

aws

Services

Search

[Option+S]

Domains

SageMaker dashboard

Images

Lifecycle configurations

Search

JumpStart

Amazon SageMaker > Notebook instances

Notebook instances

Info

Search notebook instances

< 1 >

⚙️

	Name	Instance	Creation time	Status	Actions
<input type="radio"/>	p8-medium	ml.t3.medium	6/13/2023, 2:01:04 PM	InService	Open Jupyter Open JupyterLab

⌂

Actions

Create notebook instance

IAM > Roles > SageMaker-DS2

SageMaker-DS2

Delete

Summary

Edit

Creation date
June 06, 2023, 12:47 (UTC+02:00)

Last activity
1 hour ago

ARN
arn:aws:iam::319752300128:role/service-role/SageMaker-DS2

Maximum session duration
1 hour

Permissions | Trust relationships | Tags | Access Advisor | Revoke sessions

Permissions policies (9)

You can attach up to 10 managed policies.

⌂ Simulate Remove Add permissions

Filter policies by property or policy name and press enter.

< 1 > ⚙️

<input type="checkbox"/>	Policy name	Type	Description
<input type="checkbox"/>	AmazonSageMaker-ExecutionPolicy-20230605T231236	Customer managed	
<input type="checkbox"/>	SageMakerS3AllResourcesPolicyTemplate_1686048436957	Customer managed	
<input type="checkbox"/>	SageMakerS3BucketPolicyTemplate_1686048182131	Customer managed	
<input type="checkbox"/>	SM_CommonJobManagement_1686048436957	Customer managed	
<input type="checkbox"/>	SM_ModelManagement_1686048436957	Customer managed	
<input type="checkbox"/>	SM_StudioAppPermissions_1686048436957	Customer managed	
<input type="checkbox"/>	AWSGlueServiceNotebookRole	AWS managed	Policy for AWS Glue service role which allow...
<input type="checkbox"/>	AmazonSageMakerFullAccess	AWS managed	Provides full access to Amazon SageMaker ...
<input type="checkbox"/>	AWSGlueConsoleSageMakerNotebookFullAccess	AWS managed	Provides full access to AWS Glue via the AW...

Instance Sagemaker

Notre notebook:

- accède aux images
- extrait leur classification du nom du dossier et leur features avec un modèle de type VGG16
- réalise une réduction de dimensions
- sauvegarde les résultats sur S3:

```
df_pandas.head()  
write_dataframe_to_csv_on_s3(df_pandas, S3_bucket, f'{OUT_FOLDER}/resultats_pca.csv')
```

Writing 25 records to Results/resultats_pca.csv on bucket [s3a://p8-images](#)



Fruits!

Non retenue: EMR Workspace

aws

Services

Search

[Option+S]

Frankfurt

aws_ilaria

Amazon EMR

EMR Serverless

EMR on EC2

Clusters

Events

Block public access

Security configurations

EMR on EKS

Virtual clusters

EMR Studio

Getting Started

Studios

You're using the new Amazon EMR console

Starting July 10, 2023, we'll bring you to the new Amazon EMR console by default. To use the old console, select [Switch to the old console](#) in the new console sidebar. Amazon EMR will deprecate the old console starting September 30, 2023.

[Learn more](#)

Amazon EMR

EMR on EC2: Clusters

Clusters (29)

Info

Filter clusters by status

Find clusters by ID or name, or search for text within loaded results

Filter clusters by creation date-time

View details

Terminate

Clone

Create cluster

	Cluster Id	Cluster name	Status	Status details	Creation time (UTC+02:00)	Elapsed time	Normalized instance hours
<input type="checkbox"/>	j-1LF1A332P3Z39	My cluster	Terminated	User request	June 23, 2023, 14:21	2 hours, 2 minutes	96
<input type="checkbox"/>	j-3DYSPPQS4CIEQ	My cluster	Terminated	User request	June 21, 2023, 13:33	1 hour, 50 minutes	56
<input type="checkbox"/>	j-164RDL4W8Q22J	My cluster	Terminated	User request	June 12, 2023, 14:46	1 hour, 13 minutes	24
<input type="checkbox"/>	j-43HBEBMO7MVP	My cluster	Terminated		June 12, 2023, 11:14	1 hour, 9 minutes	8
<input type="checkbox"/>	j-1EI7SJA1P41EV	My cluster	Terminated	User request	June 09, 2023, 02:25	4 minutes, 50 seconds	0
<input type="checkbox"/>	j-1N5MC7ZZBETWF	My cluster	Terminated with errors	Validation error	June 09, 2023, 02:23	38 seconds	0
<input type="checkbox"/>	j-180EM8C7CM1UE	My cluster	Terminated	User request	June 08, 2023, 21:27	4 hours, 54 minutes	40

Workspaces (Notebooks) (1)

Info

EMR Notebooks are now EMR Studio Workspaces. You can organize and run interactive notebooks in Workspaces.

All Studios

Find Workspaces by name, Studio, status, or last modified by

Actions

Launch Workspace

Create Workspace

	Workspace name	Studio name	Status	Cluster ID	Creation time (UTC+02:00)	Last modified by	Last modified (UTC+02:00)
<input type="radio"/>	p8-3	p8	Idle	j-1LF1A332P3Z39	June 08, 2023, 22:26	ut-p8-ws	June 23, 2023, 16:22

Non retenue: EMR Workspace

IAM > Users > ut-p8-ws

ut-p8-ws

Delete

Summary

ARN

arn:aws:iam::319752300128:user/ut-p8-ws

Console access

Enabled without MFA

Access key 1

AKIAUU4VY4ZQCKGHYKQ5 - Active

Never used, 14 days old.

Created

June 07, 2023, 16:17 (UTC+02:00)

Last console sign-in

2 days ago

Access key 2

Not enabled

Permissions

Groups (1)

Tags (1)

Security credentials

Access Advisor

Permissions policies (5)

Permissions are defined by policies attached to the user directly or through groups.



Remove

Add permissions

Search

Filter by Type

All types

< 1 > ⚙

<input type="checkbox"/>	Policy name	Type	Attached via
<input type="checkbox"/>	AmazonEMRFullAccessPolicy_v2	AWS managed	Group accesso-p8
<input type="checkbox"/>	AmazonEMRServicePolicy_v2	AWS managed	Group accesso-p8
<input type="checkbox"/>	EMRStudio-user-role-ilaria	Customer managed	Group accesso-p8
<input type="checkbox"/>	policy-hopefully-launch	Customer inline	Inline
<input type="checkbox"/>	PowerUserAccess	AWS managed - job function	Group accesso-p8

Conclusions

Conclusions

Il émerge la nécessité d'expertises spécifiques:

- Côté langage: Pyspark s'appuie sur Spark, qui est écrit en Scala, qui s'appuie sur JVM. Débogage problématique car les erreurs sont difficiles à interpréter.
- Côté plateforme: les solutions AWS sont en mise à jour continue, le support payant semble indispensable pour une mise en production efficace.



Fruits!

Merci