

Etape n°1 : Choix des données structurées et non structurées

Etape n°2 : Intégration des données structuré dans le Cloud Amazon

Etape n°3 : Intégration des données non structuré dans le Cloud Amazon

Etape n°4 : Réalisation des tableaux de bord

Rendu Final

*03/02/2023
Amazon Web Service S3*

*B.K.Nadim, R.Princy,
R.Krishna, D.Nadejda*

INTRODUCTION



Ce rapport présente l'application des connaissances acquises en matière de développement d'environnements AWS pour intégrer le flux de données structurées et non structurées vers le cloud AWS S3, ainsi que la création d'un tableau de bord pour visualiser nos jeux de données.

Nous avons utilisé l'interface de ligne de commande AWS pour configurer les clés d'accès AWS et les informations d'identification nécessaires pour accéder à AWS. Nous avons également créé un compartiment (« bucket ») sur S3 pour stocker des données, organiser les fichiers et les dossiers, et enfin transférer les données vers AWS S3.

Ce rapport décrit les étapes de manière détaillée et inclut les captures d'écran appropriées pour illustrer chaque étape.

On a appliqué le cours de l'environnement de développement AWS pour intégrer le flux de données structurées vers le cloud AWS S3 comme suit :

- On a installé l'interface de ligne de commande AWS (AWS CLI) d'abord en suivant les instructions fournies dans la documentation AWS.
<https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html> pour windows.
- On a fait la configuration de AWS CLI localement sur cmd avec nos clés d'accès AWS et les informations d'identification et la région choisie.

```
nadim@Nadim:~$ aws configure
AWS Access Key ID [*****SMIZ]:
AWS Secret Access Key [*****w0al]:
Default region name [us-east-1]:
Default output format [json]:
```

- Il y a deux façons pour créer le compartiment « Bucket » sur S3 : soit via un script lié avec AWS CLI avec un langage choisi comme Bash ou Python, soit via le plateforme S3 sur AWS directement comme on faisait en cours à l'école. Ici on choisit la méthode de création directe qui est illustré ci-après : [BUCKET_NAME] : bucket-aws-iabd-npnk

Create bucket Info

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name

Bucket name must be globally unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

AWS Region

US East (N. Virginia) us-east-1

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Choose bucket

- On a bien vérifié la création du compartiment (Bucket) via local c'est-à-dire en avec la commande linux pour rassurer le fonctionnement de AWS CLI et via plateforme, comme suit :

Via terminal :

```
nadim@Nadim:~$ aws s3api list-buckets
{
  "Buckets": [
    {
      "Name": "bucket-aws-iabd-npnk",
      "CreationDate": "2023-01-26T20:06:26.000Z"
    },
    {
      "Name": "bucket-spark-transformation",
      "CreationDate": "2023-01-21T00:21:47.000Z"
    }
  ],
  "Owner": {
    "DisplayName": "nadim.bk66",
    "ID": "db895acb03ffab046a150857d266252aecfbc15918f6a1a618a048c5b3ba02ef"
  }
}
```

On applique la commande `aws s3 cp` pour uploader les données dans notre Bucket. On peut aussi la faire via le site web AWS, mais on a choisi la commande ci-dessous :

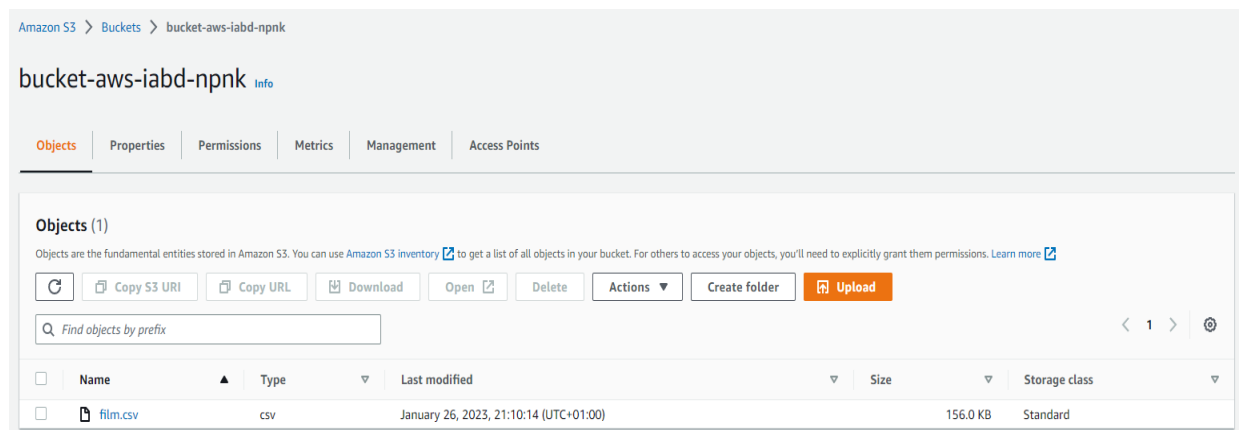
Description de la commande : 'aws s3api list-buckets'

```
nadim@Nadim:~$ aws s3 cp /mnt/c/Users/nadim/Downloads/film.csv s3://bucket-aws-iabd-npnk/
upload: ../../mnt/c/Users/nadim/Downloads/film.csv to s3://bucket-aws-iabd-npnk/film.csv
```

L'option 's3api' dans la commande indique qu'on voulait les API S3 qui fournissent une interface de programmation pour gérer des objets et des compartiments sur Amazon S3.

L'utilisation de l'option 's3api' signifie qu'on ne voulait pas utiliser les fonctionnalités de ligne de commande simplifiées offertes par la CLI AWS (par exemple, 'aws s3 ls'), mais qu'on voulait plutôt utiliser les fonctionnalités plus avancées ou détaillées et plus flexibles des API S3.

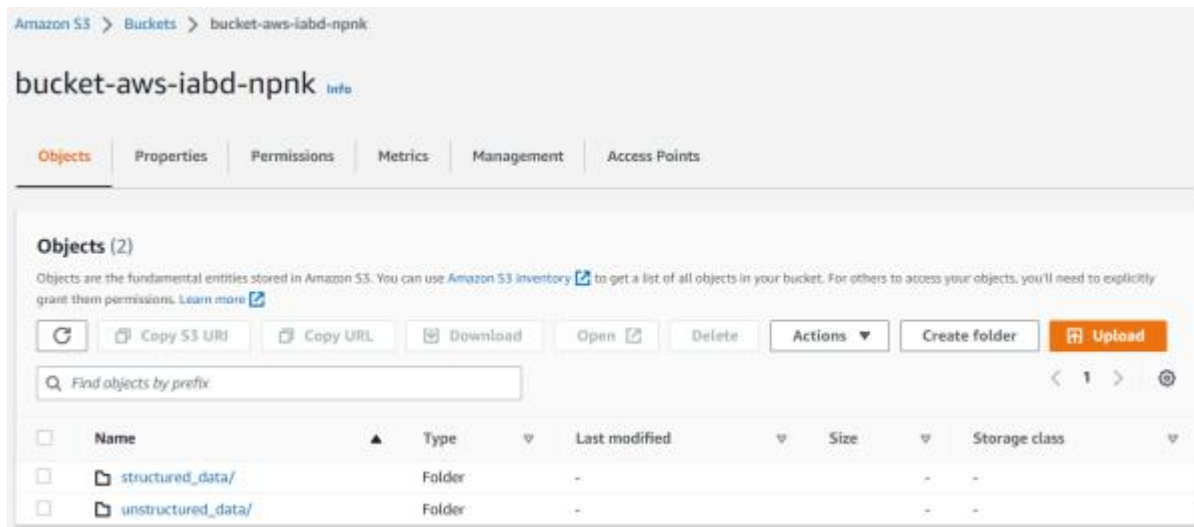
Via plateforme :



[Suite pour l'étape 3] :

Vue qu'on a déjà créer le compartiment (Bucket) dans l'étape 2, on va directement passer dans la réalisation du flux de données non structurées vers le cloud AWS.

- Pour bien organiser les dossiers et les fichiers, on a créé deux répertoires (folder) distincts pour séparer les données structurées et les données non structurées. Le voici, tel que représenté par la capture d'écran suivante :



- On a un dossier local nommé 'cloth_dataset' qui contient des sous-répertoires /jeans/ et /tshirt/ qui contient images (.jpg). On a utilisé la commande \$ aws s3 cp --recursive comme montré sur la capture d'écran suivante :

```
nadim@Nadim:~$ aws s3 cp cloth_dataset s3://bucket-aws-iabd-npnk/unstructured_data/ --recursive
upload: cloth_dataset/jeans/103.jpg to s3://bucket-aws-iabd-npnk/unstructured_data/jeans/103.jpg
upload: cloth_dataset/jeans/107.jpg to s3://bucket-aws-iabd-npnk/unstructured_data/jeans/107.jpg
```

Description de la commande : 'aws s3 cp --recursive'

C'est une commande utilisée dans le CLI AWS pour copier de multiples fichiers et dossiers de manière récursive à partir d'un système de fichiers local à un compartiment S3 ou vis-versa à partir d'un compartiment S3 vers un système de fichiers local.

Ici, on a notre processus c'est de mettre les données en local vers CLOUD AWS S3.

La commande 'cp' signifie « copier » et l'option « --recursive » indique au CLI AWS de copier les fichiers et sous-répertoires du répertoire source spécifiée vers la destination.

- Vérification des données non structurées envoyées sur AWS S3. Comme indiqué sur la capture d'écran ci-dessous :

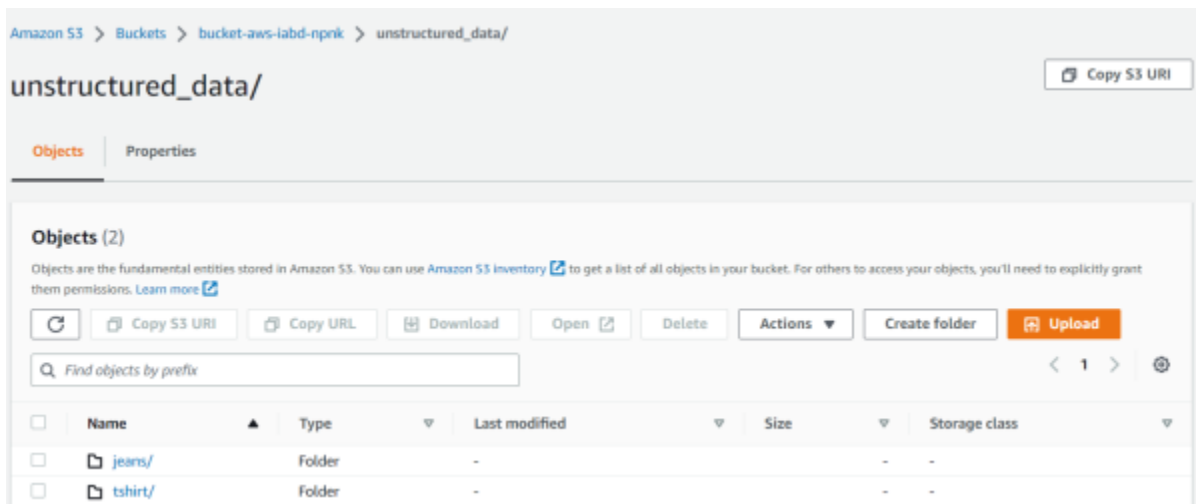
Via terminal :

```
nadim@Nadim:~$ aws s3 ls "bucket-aws-iabd-npnk"
PRE structured_data/
PRE unstructured_data/
```

```
nadim@Nadim:~$ aws s3 ls "bucket-aws-iabd-npnk/unstructured_data/jeans/"
2023-02-02 20:21:02 35952 1.jpg
2023-02-02 20:21:02 13300 10.jpg
2023-02-02 20:21:02 7046 100.jpg
2023-02-02 20:21:02 16004 101.jpg
2023-02-02 20:21:02 20181 102.jpg
2023-02-02 20:21:02 14364 103.jpg
2023-02-02 20:21:02 14364 104.jpg
2023-02-02 20:21:02 9490 105.jpg
2023-02-02 20:21:02 11060 106.jpg
2023-02-02 20:21:02 11702 107.jpg
```

```
nadim@Nadim:~$ aws s3 ls "bucket-aws-iabd-npnk/unstructured_data/tshirt/"
2023-02-02 20:21:06 26432 1.jpg
2023-02-02 20:21:06 31270 10.jpg
2023-02-02 20:21:06 41469 100.jpg
2023-02-02 20:21:06 26611 101.jpg
2023-02-02 20:21:06 15410 102.jpg
2023-02-02 20:21:06 20576 103.jpg
2023-02-02 20:21:06 31967 104.jpg
2023-02-02 20:21:06 24879 105.jpg
2023-02-02 20:21:06 24844 106.jpg
```

Via plateforme :



Amazon S3 > Buckets > bucket-aws-labd-npnk > unstructured_data/ > jeans/

jeans/

Copy S3 URI

Objects | Properties

Objects (199)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	1.jpg	jpg	February 2, 2023, 20:21:02 (UTC+01:00)	35.1 KB	Standard
<input type="checkbox"/>	10.jpg	jpg	February 2, 2023, 20:21:02 (UTC+01:00)	13.0 KB	Standard
<input type="checkbox"/>	100.jpg	jpg	February 2, 2023, 20:21:02 (UTC+01:00)	6.9 KB	Standard
<input type="checkbox"/>	101.jpg	jpg	February 2, 2023, 20:21:02 (UTC+01:00)	15.6 KB	Standard
<input type="checkbox"/>	102.jpg	jpg	February 2, 2023, 20:21:02 (UTC+01:00)	19.7 KB	Standard

Amazon S3 > Buckets > bucket-aws-labd-npnk > unstructured_data/ > tshirt/

tshirt/

Copy S3 URI

Objects | Properties

Objects (199)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	1.jpg	jpg	February 2, 2023, 20:21:06 (UTC+01:00)	25.8 KB	Standard
<input type="checkbox"/>	10.jpg	jpg	February 2, 2023, 20:21:06 (UTC+01:00)	30.5 KB	Standard
<input type="checkbox"/>	100.jpg	jpg	February 2, 2023, 20:21:06 (UTC+01:00)	40.5 KB	Standard
<input type="checkbox"/>	101.jpg	jpg	February 2, 2023, 20:21:06 (UTC+01:00)	26.0 KB	Standard
<input type="checkbox"/>	102.jpg	jpg	February 2, 2023, 20:21:06 (UTC+01:00)	15.0 KB	Standard
<input type="checkbox"/>	103.jpg	jpg	February 2, 2023, 20:21:06 (UTC+01:00)	20.1 KB	Standard

CONCLUSION

En conclusion, l'application des connaissances acquises lors du cours sur l'environnement de développement AWS a permis l'intégration réussie d'un flux de données structurées et non structurées vers le cloud AWS S3. Nous avons décrit les différentes étapes suivies pour parvenir ce résultat, notamment l'installation de l'interface de ligne de commande AWS, la configuration d'AWS CLI localement sur cmd, la création du compartiment sur S3 et l'intégration des données dans le Bucket. En outre, nous avons créé un tableau de bord pour la visualisation du jeu de données, permettant ainsi une meilleure analyse et une prise de décision plus éclairée. L'ensemble de ces étapes a permis une intégration réussie es données dans le cloud AWS S3, démontrant ainsi la pertinence et l'utilité des compétences acquises lors de notre formation.

Krishna Rasamy
Princy Rasoloarivony
Nadim Ben Khalifa
Nadejda Dorosenco

AWS Tableau de bord

Nous avons créé un tableau de bord à l'aide de Power BI sur les données structurées issu de Film.csv.

Nous avons essayé de créer un tableau de bord sur nos données non structurées qui sont des images en jpeg.

Cependant, Power BI est principalement conçu pour travailler avec des données structurées, et n'est pas capable de traiter directement des données non structurées.

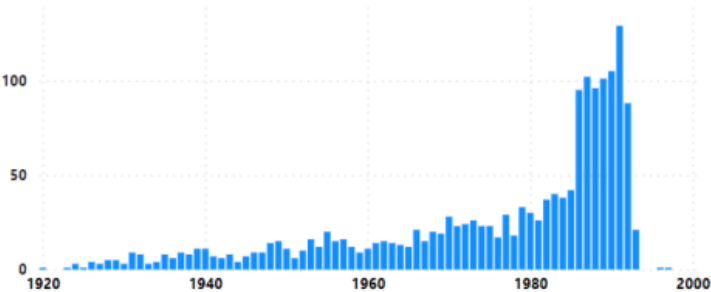
Pour utiliser des données non structurées dans Power BI, nous devons les structurer, par exemple en les convertissant dans un format tabulaire, avant d'effectuer tout tableau de bord

Donc, notre tableau de bord concernant les données non structurées n'est pas très pertinent mais c'est le maximum que l'on peut faire sur Power BI.

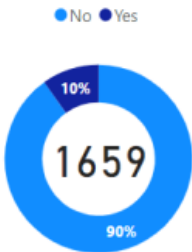
Films de 1920 à 2000

Action	Adventure	Comedy	Crime	Drama	Fantasy	Horror
--------	-----------	--------	-------	-------	---------	--------

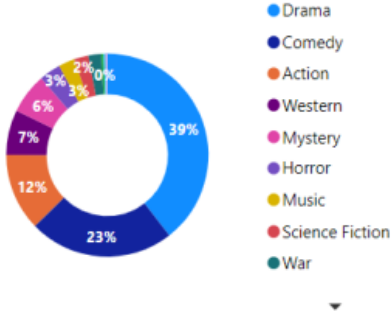
Nombre de titre par Année



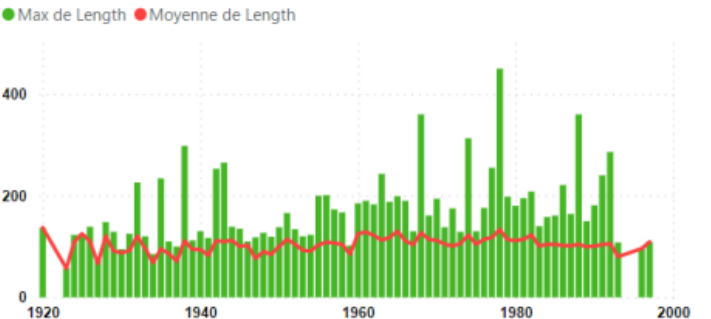
Pourcentage de récompense donné



Répartition des genres



Moyenne et maximum des durées de film en minute



Moyenne de la popularité noté sur 100 par genre

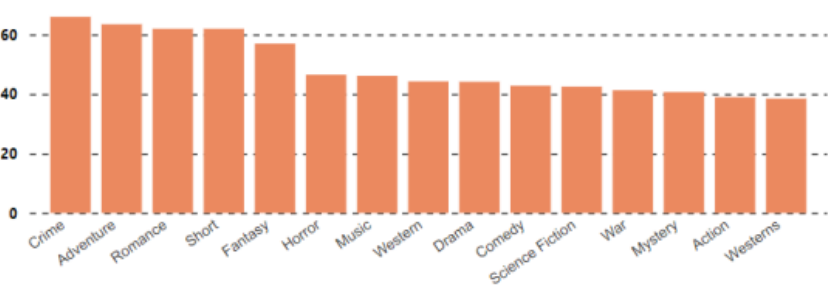


Image de jeans et Tshirts

Nombre de Name



Name	Folder Path	Année	Trimestre	Mois	Jour
114.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
115.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
116.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
117.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
118.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
119.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
12.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
120.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
121.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
122.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
123.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
124.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
125.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
126.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10
127.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\tshirt\	2023	Trim 1	février	10

Jeans

Name	Folder Path	Année	Trimestre	Mois	Jour
1.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
10.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
100.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
101.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
102.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
103.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
104.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
105.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
106.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
107.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
108.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
109.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
11.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
110.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10
111.jpg	C:\Users\jojac\Videos\cloth\cloth_dataset\jeans\	2023	Trim 1	février	10