



Université Paris cité
UFR des Sciences Fondamentales et Biomédicales

Rapport du Projet Cloud AWS

Parcours : Apprentissage Machine pour la Science des Données
(AMSD)

Préparé par :

Mr. MAHMAHI Anis
Mr. HATTABI Ilyes
Mr. BEN TAYEB
Mohamed Amine

Supervisé par :

Mr. OUARET Meriem

Promotion : 2023/2024

Table des matières

1	Création d'un utilisateur	1
2	Création d'un bucket S3	4
3	Création d'un rôle IAM pour la fonction Lambda	7
4	Création d'une fonction Lambda	9
4.1	Code de la fonction	11
4.2	Tester la fonction	12
5	Sauvegarde du fichier csv prétraité	15
6	Amazon SageMaker	17
7	Préparation des données	18
8	Training job	20
9	Hyperparameters tuning job	21
10	Prédiction par lots	24

1 Cration d'un utilisateur

- On commence par se connecter en tant que root. Ensuite, on clique sur IAM, ce qui nous conduit  la page principale d'IAM. Enfin, on accÈde  la section "Utilisateurs" o cette page s'affiche.

The screenshot shows the AWS IAM Users page. The URL in the browser is <https://us-east-1.console.aws.amazon.com/iam/home?region=eu-west-3#/users>. The page title is "Utilisateurs (1) Infos". A table lists the user "Anis" with the following details:

Nom d'utilisateur	Chemin	Groupe	Dernière activité	MFA	Âge du mot de passe	Dernière connexion
Anis	/	-	-	-	43 jours	December 21, 2023, 1:16 AM

The left sidebar shows the navigation menu for IAM, including "Utilisateurs" which is currently selected. The bottom of the screen shows the Windows taskbar with various pinned icons and the system tray.

Table des matières

- On sélectionne "Créer un utilisateur", puis on attribue un nom et un mot de passe à notre utilisateur.

The screenshot shows the 'Créer un utilisateur | IAM' page. The URL is <https://us-east-1.console.aws.amazon.com/iam/home?region=eu-west-3#/users/create>. The left sidebar shows navigation steps: Étape 1 Spécifier les détails de l'utilisateur, Étape 2 Régler les autorisations, Étape 3 Vérifier et créer, and Étape 4 Récupérer le mot de passe. The main content area is titled 'Spécifier les détails de l'utilisateur'. It contains a 'Détails de l'utilisateur' section with a 'Nom d'utilisateur' input field containing 'test-cloud-aws-hattabi'. Below it is a note: 'Le nom d'utilisateur peut comporter jusqu'à 64 caractères. Caractères valides : A-Z, a-z, 0-9 et *=_,@_ - (tiret)'. There is a checked checkbox for 'Fournir aux utilisateurs l'accès à la console de gestion AWS - facultatif' with a note: 'Si vous fournissez à une personne l'accès à la console, c'est aux [bonne pratique](#) de gérer leur accès dans IAM Identity Center.' A blue box highlights the 'Fournissez-vous à une personne un accès à la console ?' question and the 'Je souhaite créer un utilisateur IAM' radio button, which is selected. The note below states: 'La création d'utilisateurs IAM est recommandée uniquement en cas de besoin d'accès par programmation à AWS CodeCommit ou Amazon Keyspaces via des clés d'accès ou des informations d'identification spécifiques à un service, ou en cas de besoin d'un accès d'urgence à un compte via des informations d'identification de secours.'

- Par la suite, nous définissons ses autorisations, nécessitant un accès complet à S3, Lambda, SageMaker et CloudWatch

The screenshot shows the 'Créer un utilisateur | IAM' page. The URL is <https://us-east-1.console.aws.amazon.com/iam/home?region=eu-west-3#/users/create>. The left sidebar shows navigation steps: Étape 1 Spécifier les détails de l'utilisateur, Étape 2 Régler les autorisations, and Étape 3 Vérifier et créer. The main content area is titled 'Vérifier et créer'. It contains a 'Détails de l'utilisateur' section with 'Nom d'utilisateur' set to 'test-cloud-aws-hattabi', 'Type de mot de passe de la console' set to 'None', and 'Demander la réinitialisation du mot de passe' set to 'Non'. Below it is a 'Résumé des autorisations' table:

Nom	Type	Utilisé comme
AmazonS3FullAccess	Gérées par AWS	Stratégie des autorisations
AmazonSageMakerFullAccess	Gérées par AWS	Stratégie des autorisations
AWSLambda_FullAccess	Gérées par AWS	Stratégie des autorisations

At the bottom, there is a 'Balises - facultatif' section.

Table des matières

- À présent, nous procémons à la connexion en utilisant les informations de l'utilisateur.

The screenshot shows a browser window with the title "Amazon Web Services Sign-In". The address bar displays a URL starting with "https://eu-north-1.signin.aws.amazon.com/oauth?client_id=arn%3aws%3asignin%3A%3A%3Aconsole%2fcanvas&code_challenge=lyDBRc8sjXHVotE91UH...". The main content area is titled "Sign in as IAM user". It contains fields for "Account ID (12 digits) or account alias" (with "944395495643" entered), "IAM user name" (with "test-cloud-aws-hattabi" entered), "Password" (with a masked password), and a "Remember this account" checkbox. Below these fields is a blue "Sign in" button. To the right of the sign-in form is a promotional banner for "Amazon Lightsail" with the text "Lightsail is the easiest way to get started on AWS" and a "Learn more »" button. At the bottom of the sign-in form, there are links for "Sign in using root user email" and "Forgot password?". A language selection dropdown shows "English" with a downward arrow. At the very bottom, small links for "Terms of Use" and "Privacy Policy" are visible, along with a copyright notice: "© 1996-2024, Amazon Web Services, Inc. or its affiliates." The browser interface includes tabs for "toulouse 3", "Paris cité", and "sorbonne", and a "Private browsing" indicator.



2 Crédation d'un bucket S3

- Après s'être connecté avec l'utilisateur créé, on accède à la section S3, où la page principale affiche une liste des compartiments existants. Ensuite, en sélectionnant "Créer un compartiment", la page de création apparaît, et on choisit la région (Paris) ainsi qu'un nom pour notre compartiment.

The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with navigation links like 'Compartiments', 'Access Grants', 'Points d'accès', etc. The main area displays 'Instantané de compte' with metrics: Stockage total (26.6 Ko), Nombre d'objets (2), and Taille moyenne d'objet (13.3 Ko). Below this, there are tabs for 'Compartiments à usage général' and 'Compartiments de répertoires'. A prominent orange button labeled 'Créer un compartiment' is located at the top right of the compartment list. The status bar at the bottom indicates it's a private browsing session for user 'test-cloud-aws-hattabi'.

The screenshot shows the 'Créer un compartiment' wizard. Step 1 is 'Configuration générale'. It asks for the 'Région AWS' (set to 'Europe (Paris) eu-west-3') and the 'Nom du compartiment' (set to 'test-bucket-hattabi'). A note says the name must be unique and follow naming rules. It also includes a 'Copier les paramètres depuis un compartiment existant - facultatif' section and a 'Sélectionner un compartiment' button. Step 2, 'Propriété d'objets', is partially visible below.

Table des matières

- Maintenant, avec un compartiment vide, on choisit l'option "Créer un dossier", puis on attribue un nom (raw) à ce dossier.

test-bucket-hattabi [Info](#)

Objets (0) [Info](#)

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](#)

[Créer un dossier](#) [Charger](#)

Aucun objet

Vous n'avez aucun objet dans ce compartiment.

[Charger](#)

Créer un dossier - Compartiment [+](#)

https://s3.console.aws.amazon.com/s3/buckets/test-bucket-hattabi/object/create_folder?region=eu-west-3&bucketType=general

Amazon S3 > Compartiments > test-bucket-hattabi > Créer un dossier

Créer un dossier [Info](#)

Utilisez des dossiers pour regrouper des objets dans des compartiments. Lorsque vous créez un dossier, S3 crée un objet à l'aide du nom que vous spécifiez, suivi d'une barre oblique (/). Cet objet apparaît ensuite en tant que dossier sur la console. [En savoir plus](#)

Votre stratégie de compartiment peut bloquer la création de dossiers.
Si votre stratégie de compartiment empêche le chargement d'objets sans balises, métadonnées ou bénéficiaires de liste de contrôle d'accès (ACL) spécifiques, vous ne pourrez pas créer de dossier à l'aide de cette configuration.
Au lieu de cela, vous pouvez utiliser la [configuration](#) de chargement pour charger un dossier vide et spécifier les paramètres appropriés.

Dossier

Nom du dossier /

Les noms de dossier ne peuvent pas contenir le caractère « / ». [Consulter les règles d'attribution de noms](#)

Chiffrement côté serveur [Info](#)
Le chiffrement côté serveur protège les données au repos.

Table des matières

- Ensuite, on clique sur "Charger", sélectionne le fichier du dataset Titanic, et achève le processus de chargement.

The screenshot shows the AWS S3 console interface. The URL in the address bar is <https://s3.console.aws.amazon.com/s3/buckets/test-bucket-hattabi?region=eu-west-3&bucketType=general&prefix=raw/&showversions=false>. The page displays the contents of the 'raw/' folder under the 'test-bucket-hattabi' bucket. The 'Objets' tab is selected, showing 0 objects. A prominent orange 'Charger' button is visible at the top right of the main content area. The bottom section shows a table header for 'Nom', 'Type', 'Dernière modification', 'Taille', and 'Classe de stockage'. A message indicates 'Aucun objet' (No objects) and a 'Charger' button.

The screenshot shows the AWS CloudShell interface. The URL in the address bar is <https://s3.console.aws.amazon.com/s3/upload/test-bucket-hattabi?region=eu-west-3&bucketType=general&prefix=raw/>. The page title is 'Chargement d'objets : compartiment'. The bottom status bar shows the date and time as 1/13/2024 6:35 AM.

The screenshot shows the AWS S3 console interface during the file upload process. The URL in the address bar is <https://s3.console.aws.amazon.com/s3/upload/test-bucket-hattabi?region=eu-west-3&bucketType=general&prefix=raw/>. The page title is 'Chargement d'objets : compartiment'. The 'Charger' button is highlighted in orange. Below it, a dashed box contains the instruction 'Faites glisser et déposez les fichiers et dossiers que vous souhaitez charger ici, ou sélectionnez Ajouter des fichiers ou Ajouter un dossier.' A table titled 'Fichiers et dossiers' lists one file: 'titanic.csv' (1 Total, 58.9 Ko). The file details show 'application/vnd.ms-' as the type and 58.9 Ko as the size. The bottom status bar shows the date and time as 1/13/2024 6:35 AM.

3 Crédit d'un rôle IAM pour la fonction Lambda

- Nous revenons à la page IAM, puis nous sélectionnons "Rôles" et cliquons sur "Créer un rôle".

The screenshot shows the AWS IAM Roles page. On the left, there's a sidebar with navigation links like 'Tableau de bord', 'Gestion des accès', 'Rôles' (which is selected and highlighted in blue), and 'Rapports d'accès'. The main content area has a heading 'Rôles (20) Infos' with a note: 'Un rôle IAM est une identité que vous pouvez créer et qui dispose d'autorisations spécifiques avec des informations d'identification valides pendant de courtes durées. Les rôles peuvent être endossés par des entités de confiance.' Below this is a search bar and a table with columns: 'Nom du rôle', 'Entités de confiance', and 'Dernière activité'. The table lists 20 roles, each with a small checkbox icon to the left. The roles listed are: Amazon_EventBridge_Scheduler_LAMBDA_be0fbf3f9e, AmazonSageMaker-ExecutionRole-20231221T170677, AmazonSageMaker-ExecutionRole-20240111T144260, AmazonSageMakerServiceCatalogProductsApiGatewayRole, AmazonSageMakerServiceCatalogProductsCloudformationRole, AmazonSageMakerServiceCatalogProductsCodeBuildRole, AmazonSageMakerServiceCatalogProductsCodePipelineRole, AmazonSageMakerServiceCatalogProductsEventsRole, AmazonSageMakerServiceCatalogProductsExecutionRole, and AmazonSageMakerServiceCatalogProductsFirehoseRole. The 'Entités de confiance' column shows service names like 'scheduler', 'sagemaker', 'sagemaker', 'apigateway', 'cloudformation', 'codebuild', 'codepipeline', 'events', 'sagemaker', and 'firehose'. The 'Dernière activité' column shows dates like 'Il y a 22 jours' and '-'.

Table des matières

- On opte pour "Service AWS" en tant que type d'entité, et "Lambda" comme cas d'utilisation lors de la création du rôle.

Créer un rôle | IAM | Global

https://us-east-1.console.aws.amazon.com/iam/home?region=eu-west-3#roles/create

toulouse 3 Paris cité sorbonne

Services Rechercher [Alt+S]

Etape 5 Nommer, vérifier et créer

Autorisez les services AWS tels qu'EC2, Lambda ou autre à effectuer des actions dans ce compte.

Autorisez les entités d'autres comptes AWS qui appartiennent à vous à un tiers à effectuer des actions dans ce compte.

Permet aux utilisateurs fédérés par le fournisseur d'identité web externe spécifié d'assumer ce rôle pour effectuer des actions dans ce compte.

Fédération SAML 2.0 Autoriser les utilisateurs fédérés avec SAML 2.0 à partir d'un répertoire d'entreprise à effectuer des actions dans ce compte.

Stratégie d'approbation personnalisée Créez une stratégie d'approbation personnalisée pour permettre à d'autres utilisateurs d'effectuer des actions dans ce compte.

Cas d'utilisation
Autorisez un service AWS comme EC2, Lambda ou autres à effectuer des actions dans ce compte.

Service ou cas d'utilisation

Lambda

Choisissez un cas d'utilisation pour le service spécifié.

Cas d'utilisation

Lambda Allows Lambda functions to call AWS services on your behalf.

Annuler Suivant

- Nous attachons les politiques nécessaires à la fonction, à savoir l'accès complet à CloudWatch et l'accès complet à S3. Enfin, nous lui attribuons un nom et une description.

Créer un rôle | IAM | Global

https://us-east-1.console.aws.amazon.com/iam/home?region=eu-west-3#roles/create?selectedUseCase=Lambda&trustedEntityType=AWS_SERVICE&select... Private browsing

toulouse 3 Paris cité sorbone

Services Rechercher [Alt+S]

IAM > Rôles > Créer un rôle

Étape 1 Sélectionner une entité de confiance

Étape 2 Ajouter des autorisations

Étape 3 Nommer, vérifier et créer

Ajouter des autorisations Infos

Politiques des autorisations (1/913) Infos

Choisissez une ou plusieurs stratégies à attacher à votre nouveau rôle.

Filtrer par Type

Nom de la politique	Type	Description
<input type="checkbox"/> <input checked="" type="checkbox"/> AmazonDMSRedshiftS3Role	Gérées par AWS	Provides access to manage S3 settings for Redshift e...
<input type="checkbox"/> <input checked="" type="checkbox"/> AmazonS3FullAccess	Gérées par AWS	Provides full access to all buckets via the AWS Mana...
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> AmazonS3ObjectLambdaExecutionRole...	Gérées par AWS	Provides AWS Lambda functions permissions to inte...

AmazonS3ObjectLambdaExecutionRolePolicy

Provides AWS Lambda functions permissions to interact with Amazon S3 Object Lambda. Also grants Lambda permissions to write to CloudWatch Logs.

```
1 ~ [{  
2     "Version": "2012-10-17",  
3     "Statement": [  
4         {  
5             "Effect": "Allow",  
6             "Action": [  
7                 "logs:CreateLogGroup",
```

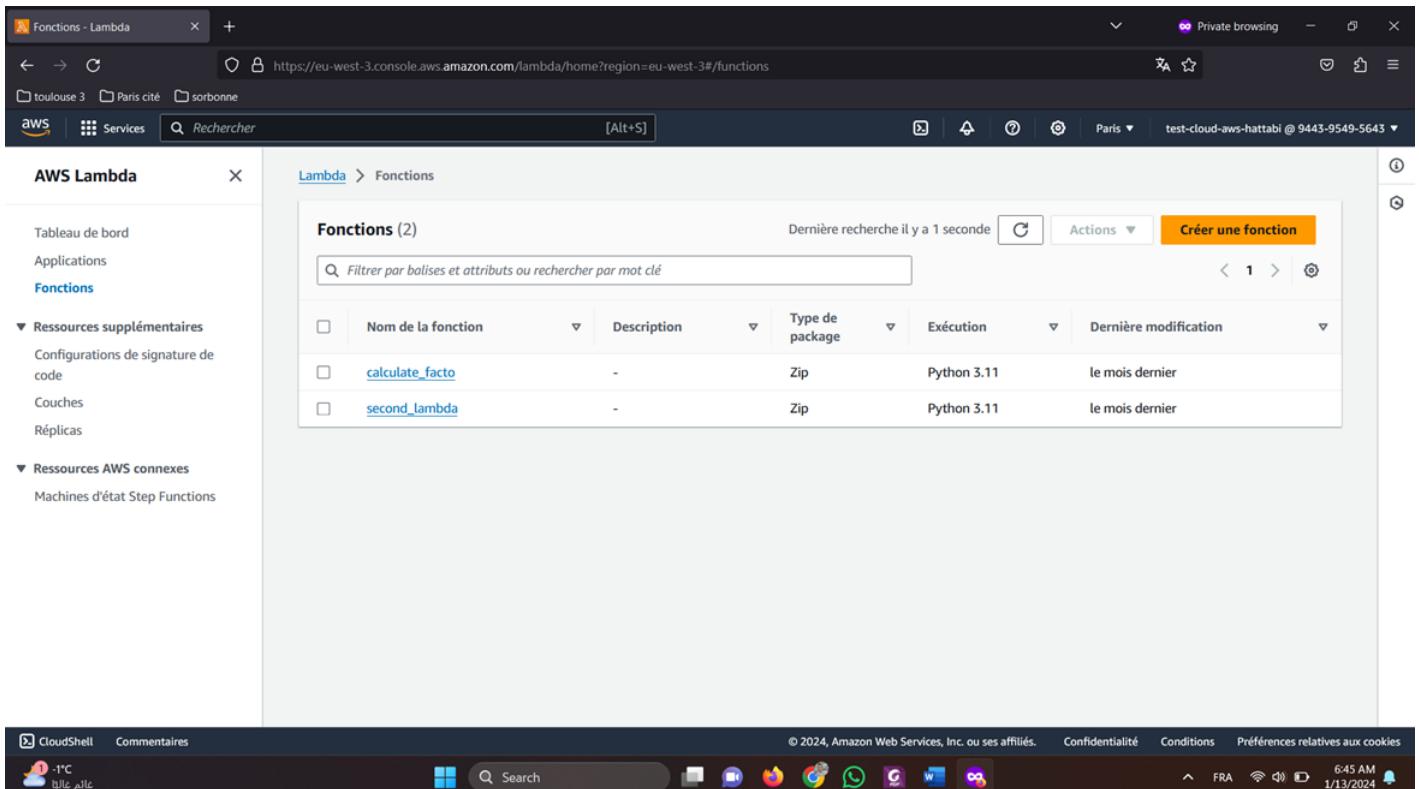
CloudShell Commentaires

© 2024, Amazon Web Services, Inc. ou ses affiliés. Confidentialité Conditions Préférences relatives aux cookies

1°C FRA 6:43 AM 1/13/2024

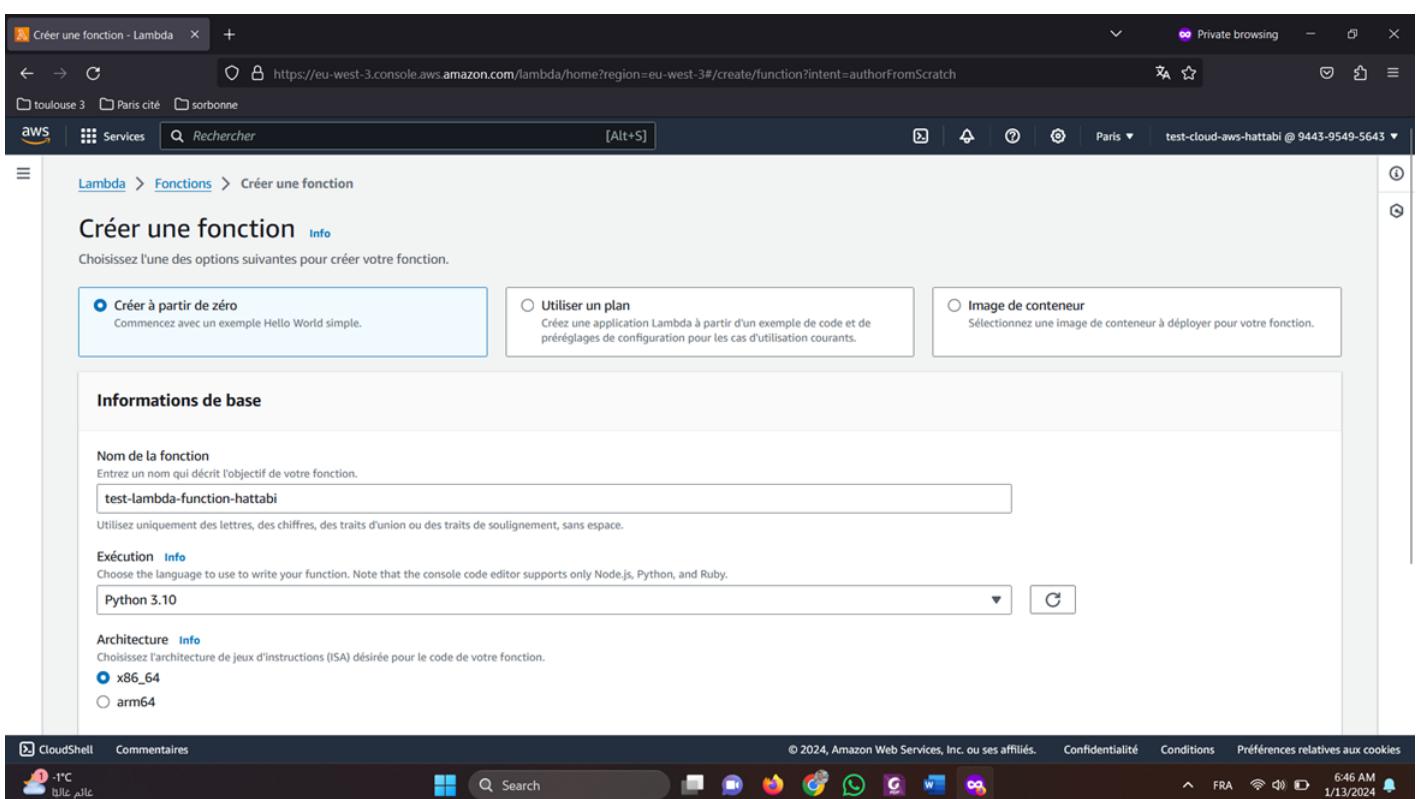
4 Crédation d'une fonction Lambda

- À cette étape, nous ouvrons la page du service Lambda, où nous observons la liste des fonctions existantes. Ensuite, nous cliquons sur "Créer une fonction", lui attribuons un nom, et choisissons le langage de programmation.



The screenshot shows the AWS Lambda console interface. On the left, there is a sidebar with navigation links: Tableau de bord, Applications, Fonctions (selected), Ressources supplémentaires, and Ressources AWS connexes. The main area is titled "Fonctions (2)" and displays a table with two rows:

Nom de la fonction	Description	Type de package	Exécution	Dernière modification
calculate_facto	-	Zip	Python 3.11	le mois dernier
second_lambda	-	Zip	Python 3.11	le mois dernier



The screenshot shows the "Créer une fonction" (Create a function) wizard. The first step, "Choisissez l'une des options suivantes pour créer votre fonction.", is displayed. Three options are available:

- Créer à partir de zéro** (Selected): "Commencez avec un exemple Hello World simple."
- Utiliser un plan**: "Créez une application Lambda à partir d'un exemple de code et de pré-réglages de configuration pour les cas d'utilisation courants."
- Image de conteneur**: "Sélectionnez une image de conteneur à déployer pour votre fonction."

Below these options, the "Informations de base" (Basic information) section is shown. It includes fields for "Nom de la fonction" (Name of the function) set to "test-lambda-function-hattabi", "Exécution" (Execution) set to "Python 3.10", and "Architecture" (Architecture) set to "x86_64".

Table des matières

- Par la suite, nous attachons le rôle Lambda créé à l'étape précédente à la fonction que nous venons de créer.

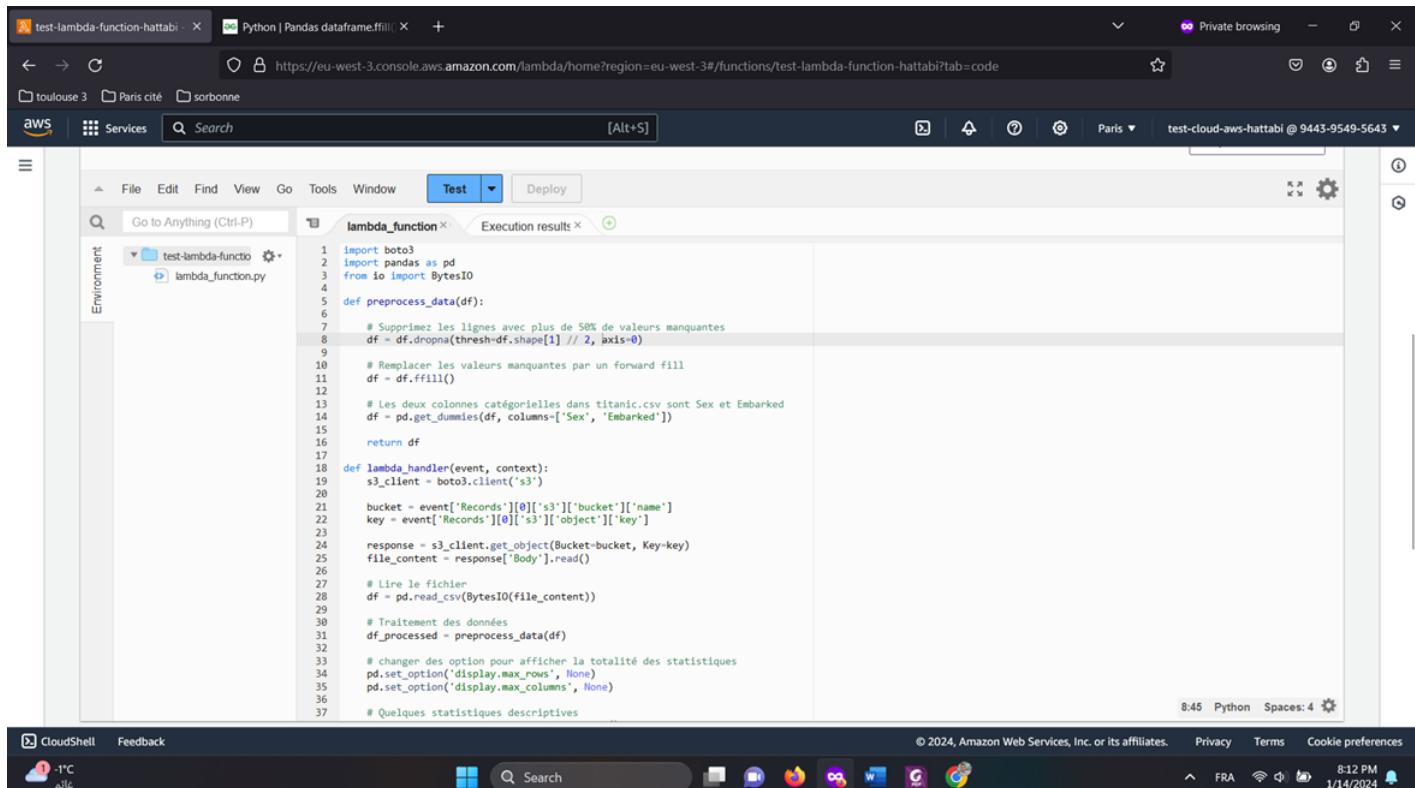
The screenshot shows the 'Créer une fonction - Lambda' (Create Function - Lambda) page. In the 'Architecture' section, 'x86_64' is selected. Under 'Autorisations', the 'Utiliser un rôle existant' option is chosen, with 'test-role-lambda-hattabi' selected from a dropdown. A note states that the role must have the CloudWatch Logs permission. The 'Paramètres avancés' section is collapsed. At the bottom right are 'Annuler' and 'Créer une fonction' buttons.

- Afin que la fonction Lambda s'exécute automatiquement lorsqu'un fichier est chargé dans le dossier "raw" du bucket que nous avons créé, nous accédons à la page de la fonction et cliquons sur "Ajouter un déclencheur". Nous sélectionnons le service (S3), choisissons le bucket approprié, et ajoutons un préfixe (/raw) pour déclencher l'exécution uniquement lorsqu'un fichier est ajouté au dossier "raw".

The screenshot shows the 'Ajouter des déclencheurs - Lambda' (Add Trigger - Lambda) page. Under 'Configuration du déclencheur', 'Bucket' is set to 's3/test-bucket-hattabi'. 'Event types' is set to 'All object create events'. 'Prefix - facultatif' is set to 'raw/'. At the bottom right are 'Annuler' and 'Créer une fonction' buttons.

4.1 Code de la fonction

- On ajoute le code pour la fonction puis on clique sur déployer. Ensuite, on ajoute une couche à la fonction avec les bibliothèques utilisées (dans notre cas, on a seulement pandas).



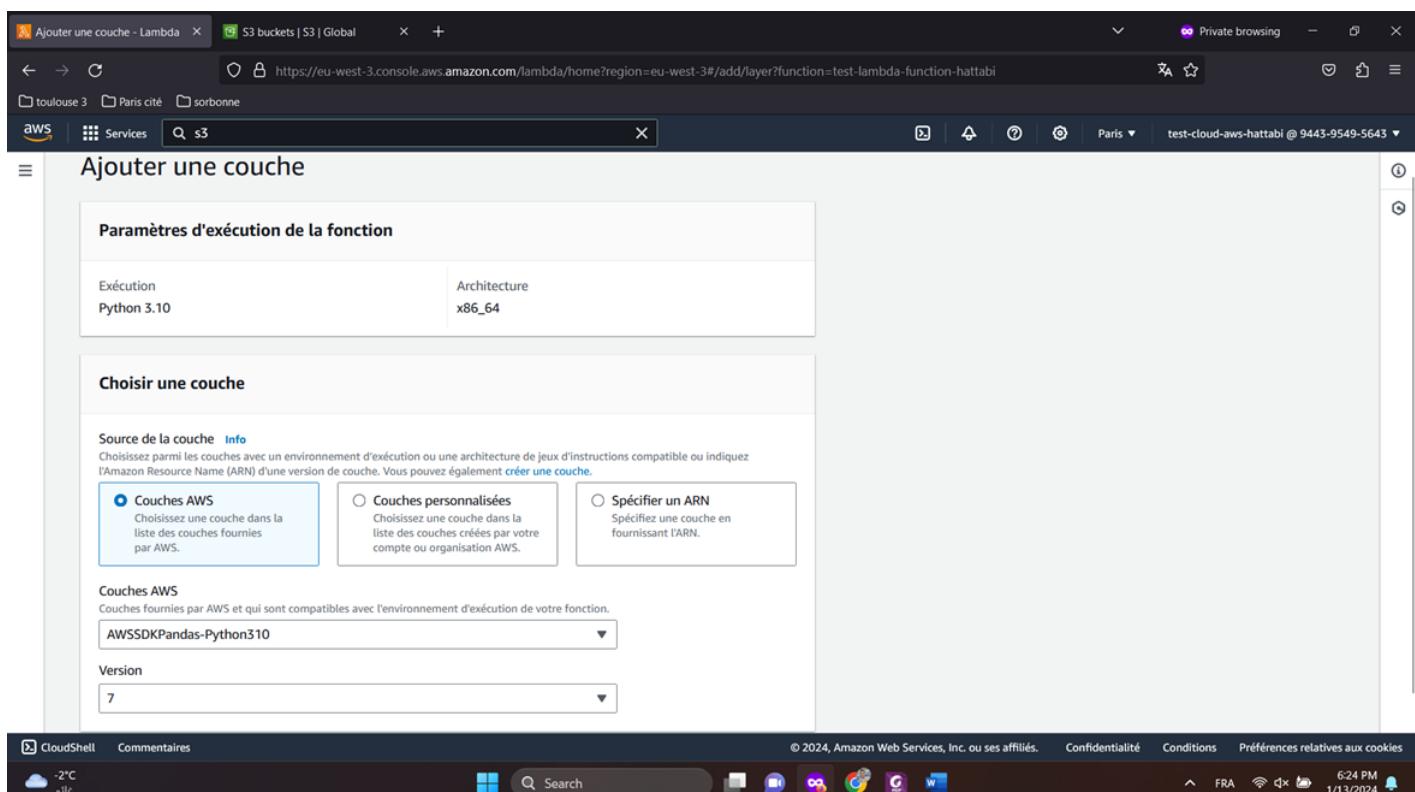
```

File Edit Find View Go Tools Window Test Deploy
lambda_function Execution results

1 import boto3
2 import pandas as pd
3 from io import BytesIO
4
5 def preprocess_data(df):
6
7     # Supprime les lignes avec plus de 50% de valeurs manquantes
8     df = df.dropna(thresh=df.shape[1] // 2, axis=0)
9
10    # Remplacer les valeurs manquantes par un forward fill
11    df = df.fillna()
12
13    # Les deux colonnes catégorielles dans titanic.csv sont Sex et Embarked
14    df = pd.get_dummies(df, columns=['Sex', 'Embarked'])
15
16    return df
17
18 def lambda_handler(event, context):
19     s3_client = boto3.client('s3')
20
21     bucket = event['Records'][0]['s3']['bucket']['name']
22     key = event['Records'][0]['s3']['object']['key']
23
24     response = s3_client.get_object(Bucket=bucket, Key=key)
25     file_content = response['Body'].read()
26
27     # Lire le fichier
28     df = pd.read_csv(BytesIO(file_content))
29
30     # Traitement des données
31     df_processed = preprocess_data(df)
32
33     # changer des option pour afficher la totalité des statistiques
34     pd.set_option('display.max_rows', None)
35     pd.set_option('display.max_columns', None)
36
37     # Quelques statistiques descriptives

```

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 8:45 Python Spaces: 4



Ajouter une couche - Lambda | S3 buckets | S3 Global | Private browsing | FRA | 8:12 PM | 1/14/2024

Ajouter une couche

Paramètres d'exécution de la fonction

Exécution Python 3.10	Architecture x86_64
--------------------------	------------------------

Choisir une couche

Source de la couche [Info](#)
Choisissez parmi les couches avec un environnement d'exécution ou une architecture de jeux d'instructions compatible ou indiquez l'Amazon Resource Name (ARN) d'une version de couche. Vous pouvez également créer une couche.

Couches AWS
Choisissez une couche dans la liste des couches fournies par AWS.

Couches personnalisées
Choisissez une couche dans la liste des couches créées par votre compte ou organisation AWS.

Spécifier un ARN
Spécifiez une couche en fournissant l'ARN.

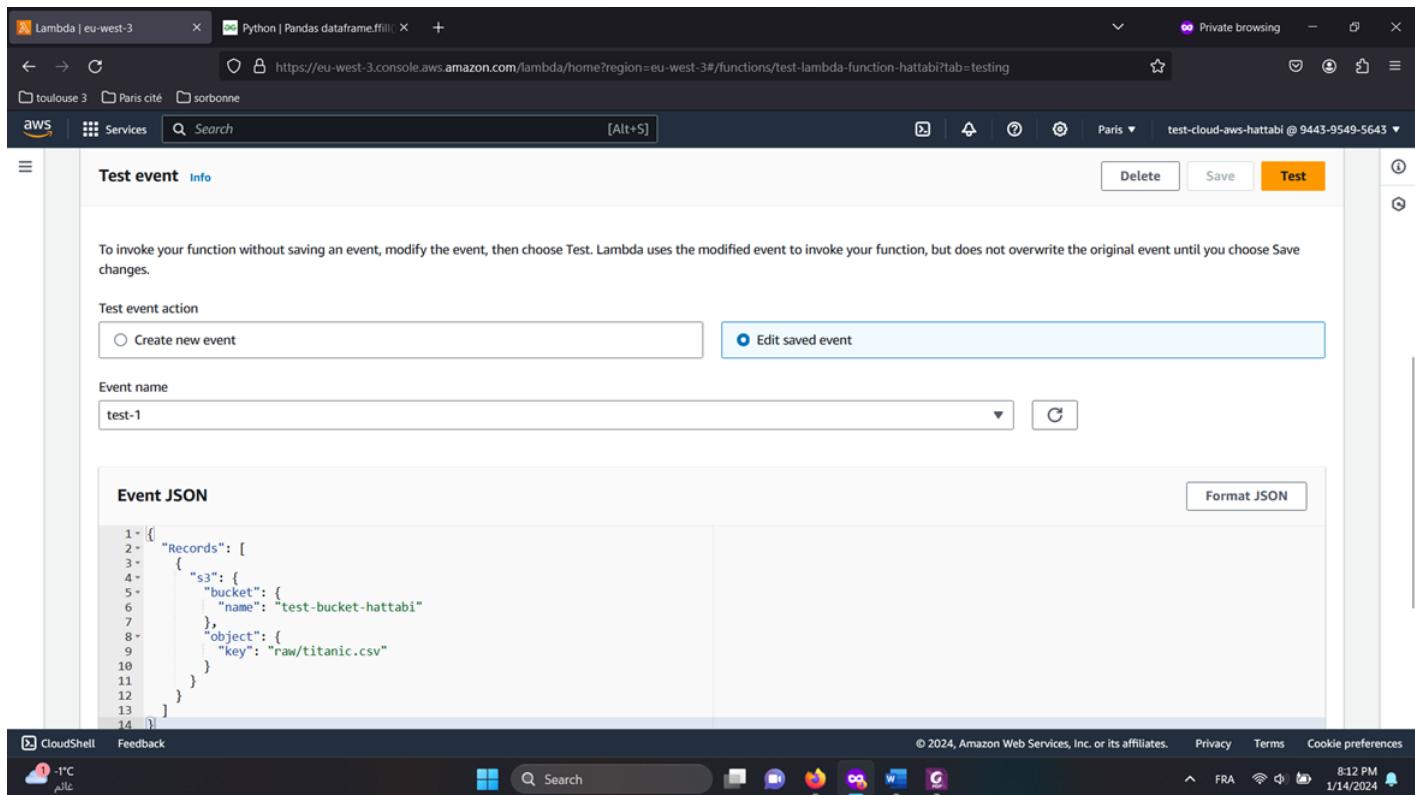
Couches AWS
Couches fournies par AWS et qui sont compatibles avec l'environnement d'exécution de votre fonction.
AWSSDKPandas-Python310

Version
7

CloudShell Commentaires © 2024, Amazon Web Services, Inc. ou ses affiliés. Confidentialité Conditions Préférences relatives aux cookies -2°C pile FRA 6:24 PM 1/13/2024

4.2 Tester la fonction

- Pour tester la fonction, nous ajoutons maintenant un événement de test en utilisant le modèle pour le test du bucket S3. Nous éliminons tous les champs inutiles, conservant uniquement le nom du bucket et la clé du fichier nécessaires..



To invoke your function without saving an event, modify the event, then choose Test. Lambda uses the modified event to invoke your function, but does not overwrite the original event until you choose Save changes.

Test event [Info](#)

Test event action

Create new event [Edit saved event](#)

Event name

test-1

Event JSON

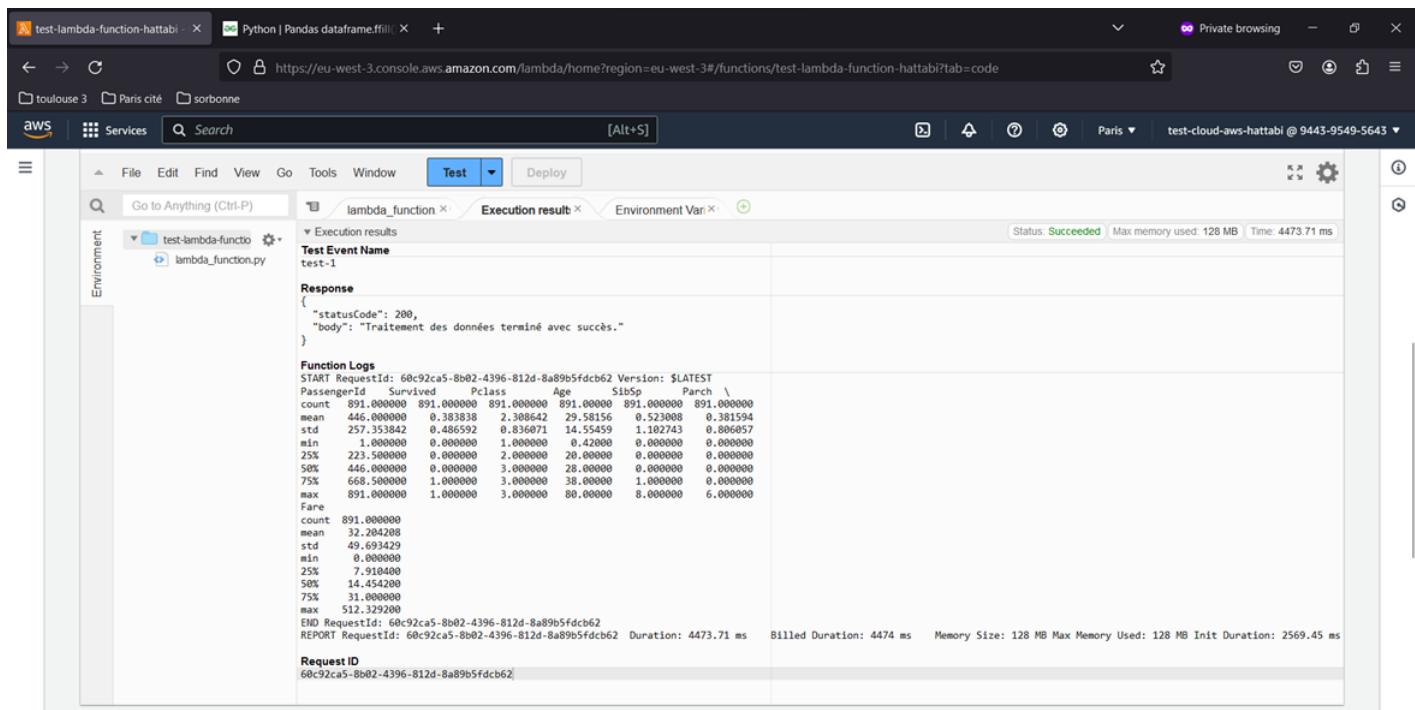
```

1 - [{
2 -   "Records": [
3 -     {
4 -       "s3": {
5 -         "bucket": {
6 -           "name": "test-bucket-hattabi"
7 -         },
8 -         "object": {
9 -           "key": "raw/titanic.csv"
10 -        }
11 -      }
12 -    }
13 -  ]
14 - ]

```

[Format JSON](#)

- On clique sur tester et, comme on peut le voir, les résultats sont corrects.



Execution result

Status: **Succeeded** | Max memory used: 128 MB | Time: 4473.71 ms

Test Event Name

test-1

Response

```
{
  "statusCode": 200,
  "body": "Traitement des données terminé avec succès."
}
```

Function Logs

```

START RequestId: 60c92ca5-8b02-4396-812d-8a89b5fdc62 Version: $LATEST
PassengerId  Survived  Pclass  Age  SibSp  Parch  \
count    891.000000  891.000000  891.000000  891.000000  891.000000
mean     32.042488  0.383818  2.308542  29.58156  0.523008  0.3801594
std      49.907079  0.486592  0.836971  14.55459  1.102743  0.896057
min      1.000000  0.000000  0.42000  0.000000  0.000000
25%    223.500000  0.000000  2.00000  0.000000  0.000000
50%    446.500000  0.000000  3.00000  0.000000  0.000000
75%    668.500000  1.00000  3.00000  38.00000  1.00000  0.000000
max    891.000000  1.00000  3.00000  80.00000  8.00000  6.000000
Fare
count    891.000000
mean     32.042488
std      49.907079
min      0.000000
25%    7.910400
50%    14.454200
75%    31.329200
max    512.329200
END RequestId: 60c92ca5-8b02-4396-812d-8a89b5fdc62
REPORT RequestId: 60c92ca5-8b02-4396-812d-8a89b5fdc62 Duration: 4473.71 ms Billed Duration: 4474 ms Memory Size: 128 MB Max Memory Used: 128 MB Init Duration: 2569.45 ms
Request ID
60c92ca5-8b02-4396-812d-8a89b5fdc62

```

- Nous devons maintenant vérifier si la fonction fonctionne lors du téléchargement

Table des matières

du fichier. Pour ce faire, nous supprimons le fichier qui a été téléchargé et effaçons l'historique de CloudWatch.

The screenshot shows the AWS S3 console in a browser window. The URL is <https://s3.console.aws.amazon.com/s3/buckets/test-bucket-hattabi/object/delete?region=eu-west-3&bucketType=general&prefix=raw&showversions=false>. The page title is "Delete objects".

Specified objects:

Name	Type	Last modified	Size
titanic.csv	csv	January 13, 2024, 06:36:05 (UTC+01:00)	58.9 KB

Permanently delete objects?

To confirm deletion, type *permanently delete* in the text input field.

- Lors du chargement du fichier "titanic.csv" à nouveau, nous constatons que la fonction s'est exécutée simultanément et a généré un journal dans CloudWatch.

The screenshot shows the AWS S3 console in a browser window. The URL is <https://s3.console.aws.amazon.com/s3/upload/test-bucket-hattabi?region=eu-west-3&bucketType=general&prefix=raw/>. The page title is "Upload objects".

Files and folders (1 Total, 58.9 KB)

Name	Type	Size
titanic.csv	application/vnd.ms-...	58.9 KB

Destination: <s3://test-bucket-hattabi/raw/>

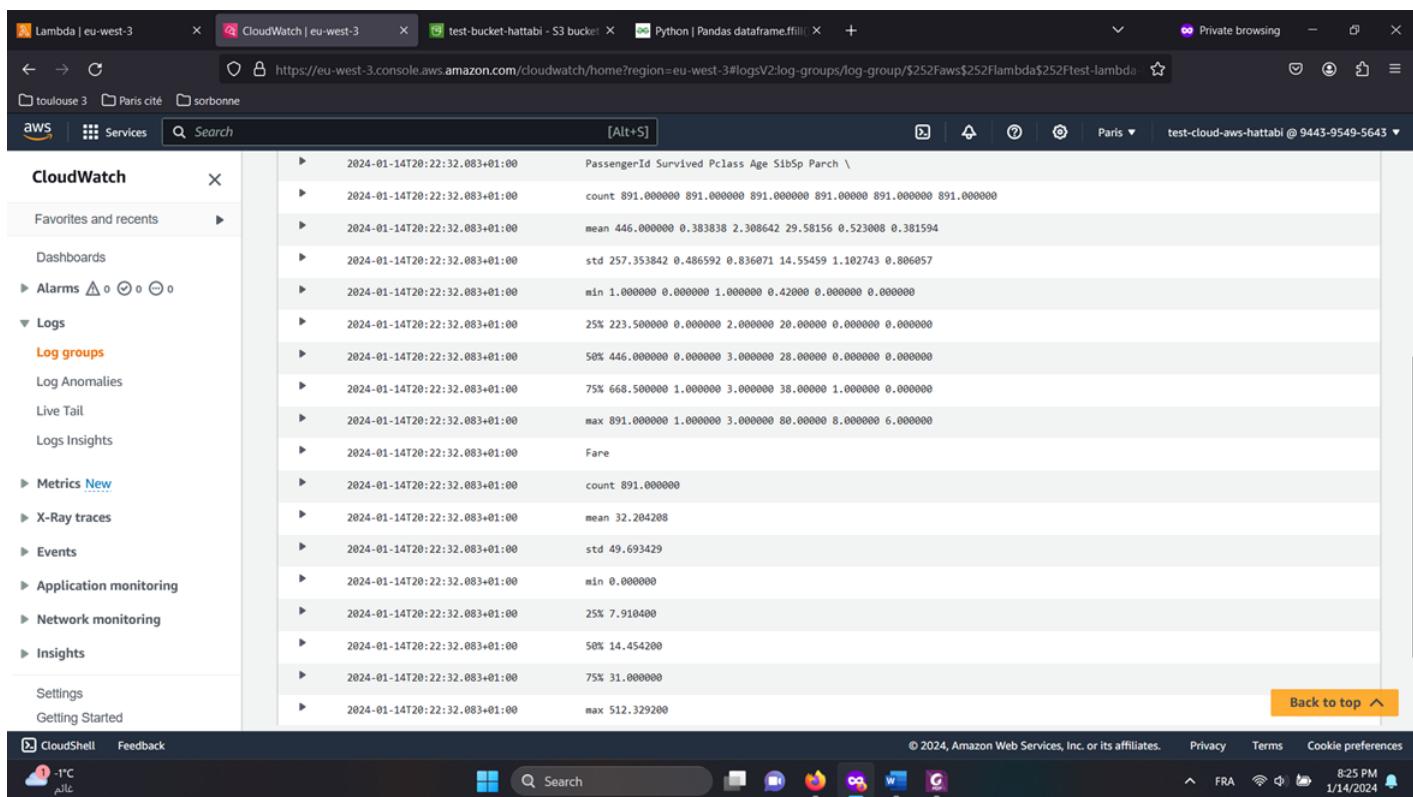
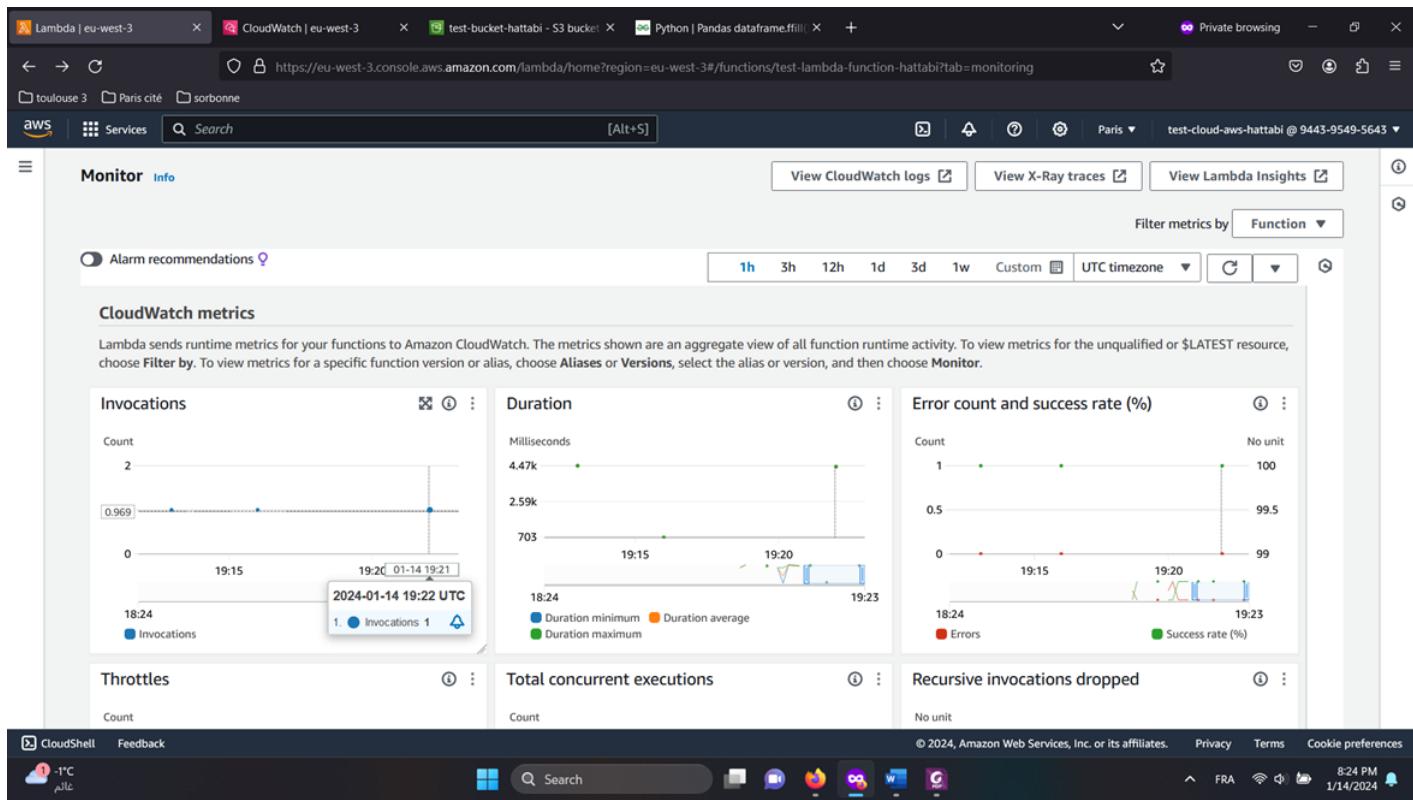
Destination details: Bucket settings that impact new objects stored in the specified destination.

Permissions: Grant public access and access to other AWS accounts.

Properties: Specify storage class, encryption settings, tags, and more.

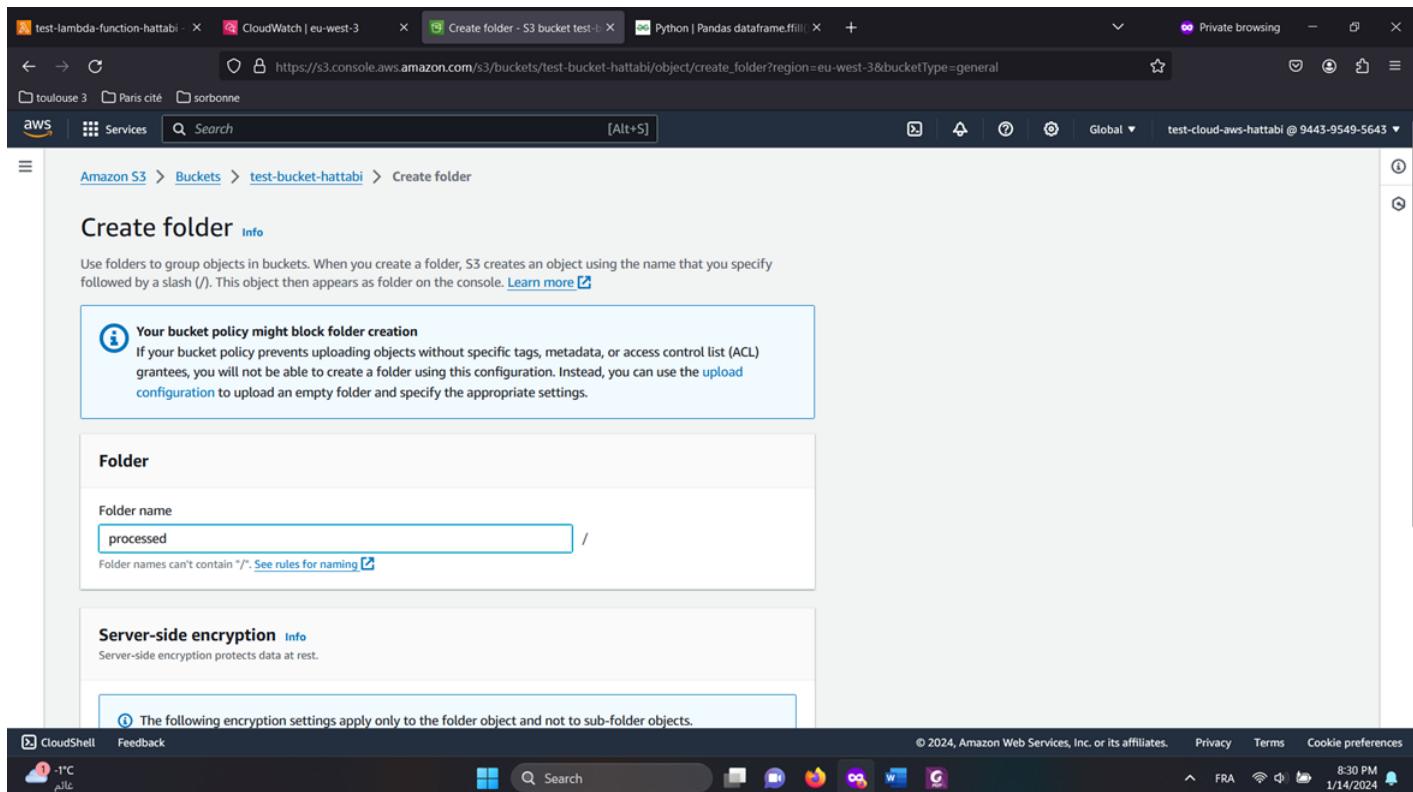
Buttons: Cancel, Upload

Table des matières



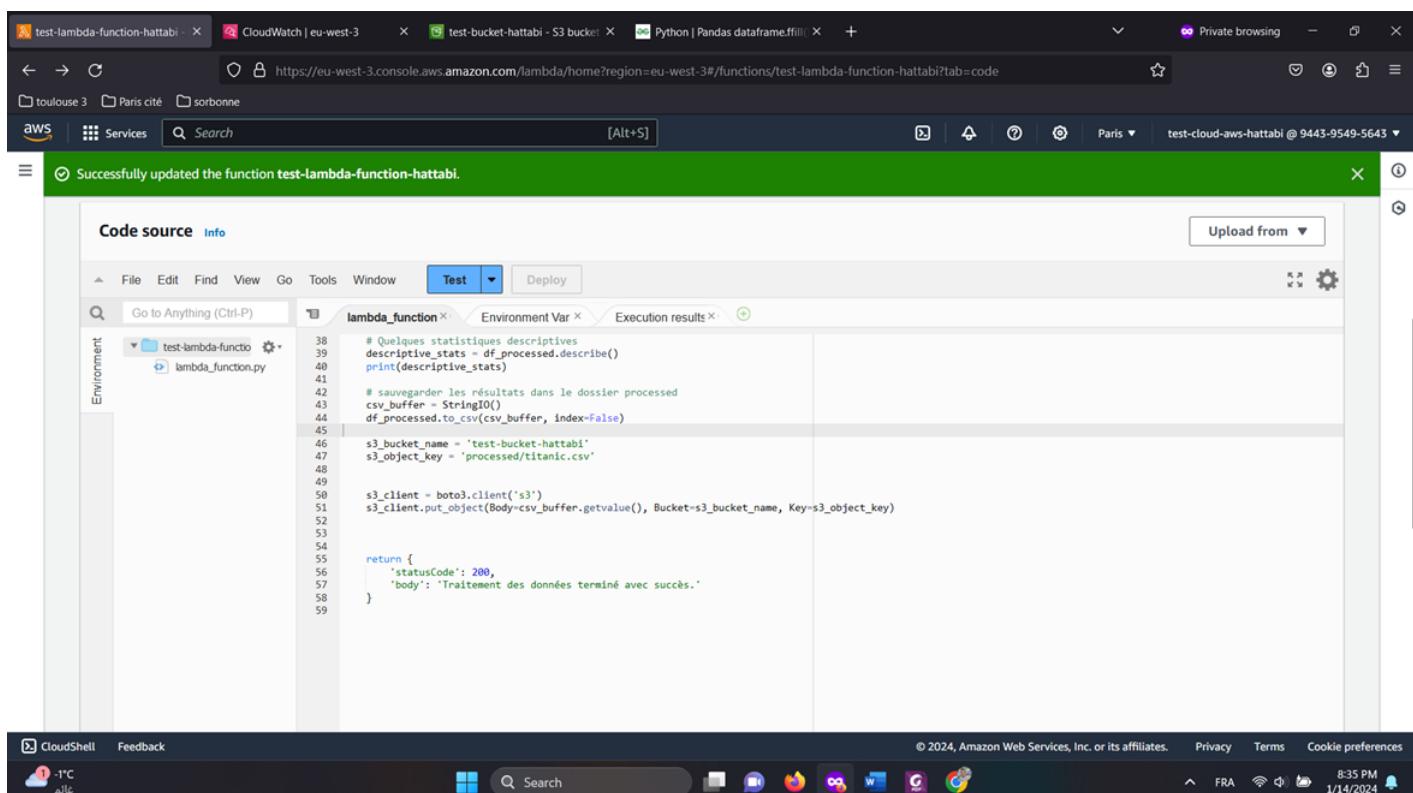
5 Sauvegarde du fichier csv prétraité

- Pour la dernière fonctionnalité, on crée un dossier nommé processed dans le même bucket.



The screenshot shows the AWS S3 'Create folder' interface. The URL in the browser is https://s3.console.aws.amazon.com/s3/buckets/test-bucket-hattabi/object/create_folder?region=eu-west-3&bucketType=general. The page displays a warning about bucket policy restrictions for folder creation. In the 'Folder' section, the 'Folder name' field contains 'processed'. Below it, there's a note about server-side encryption. A message at the bottom states: 'The following encryption settings apply only to the folder object and not to sub-folder objects.'

- Nous intégrons le code permettant d'enregistrer le fichier CSV prétraité dans le répertoire "processed".



The screenshot shows the AWS Lambda function editor for the function 'test-lambda-function-hattabi'. A green success message at the top says: 'Successfully updated the function test-lambda-function-hattabi.' The code source tab is selected, showing the following Python code:

```

# Quelques statistiques descriptives
descriptive_stats = df_processed.describe()
print(descriptive_stats)

# sauvegarder les résultats dans le dossier processed
csv_buffer = StringIO()
df_processed.to_csv(csv_buffer, index=False)

s3_bucket_name = 'test-bucket-hattabi'
s3_object_key = 'processed/titanic.csv'

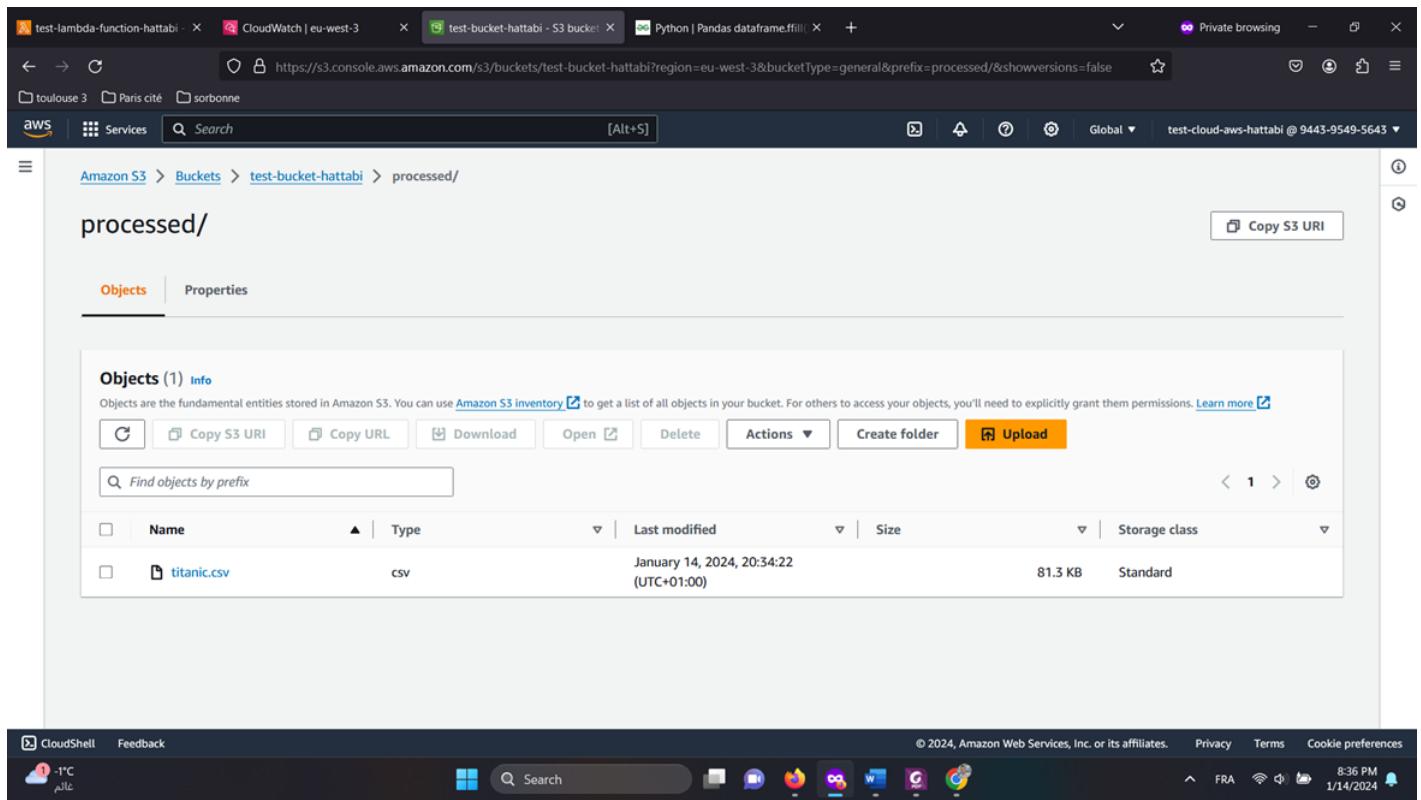
# sauvegarder le CSV dans S3
s3_client = boto3.client('s3')
s3_client.put_object(Body=csv_buffer.getvalue(), Bucket=s3_bucket_name, Key=s3_object_key)

return {
    'statusCode': 200,
    'body': 'Traitement des données terminé avec succès.'
}

```

Table des matières

- Lorsque on charge le fichier dans le dossier raw, on peut voir que le nouveau csv prétraité est sauvegardé dans le dossier souhaité.

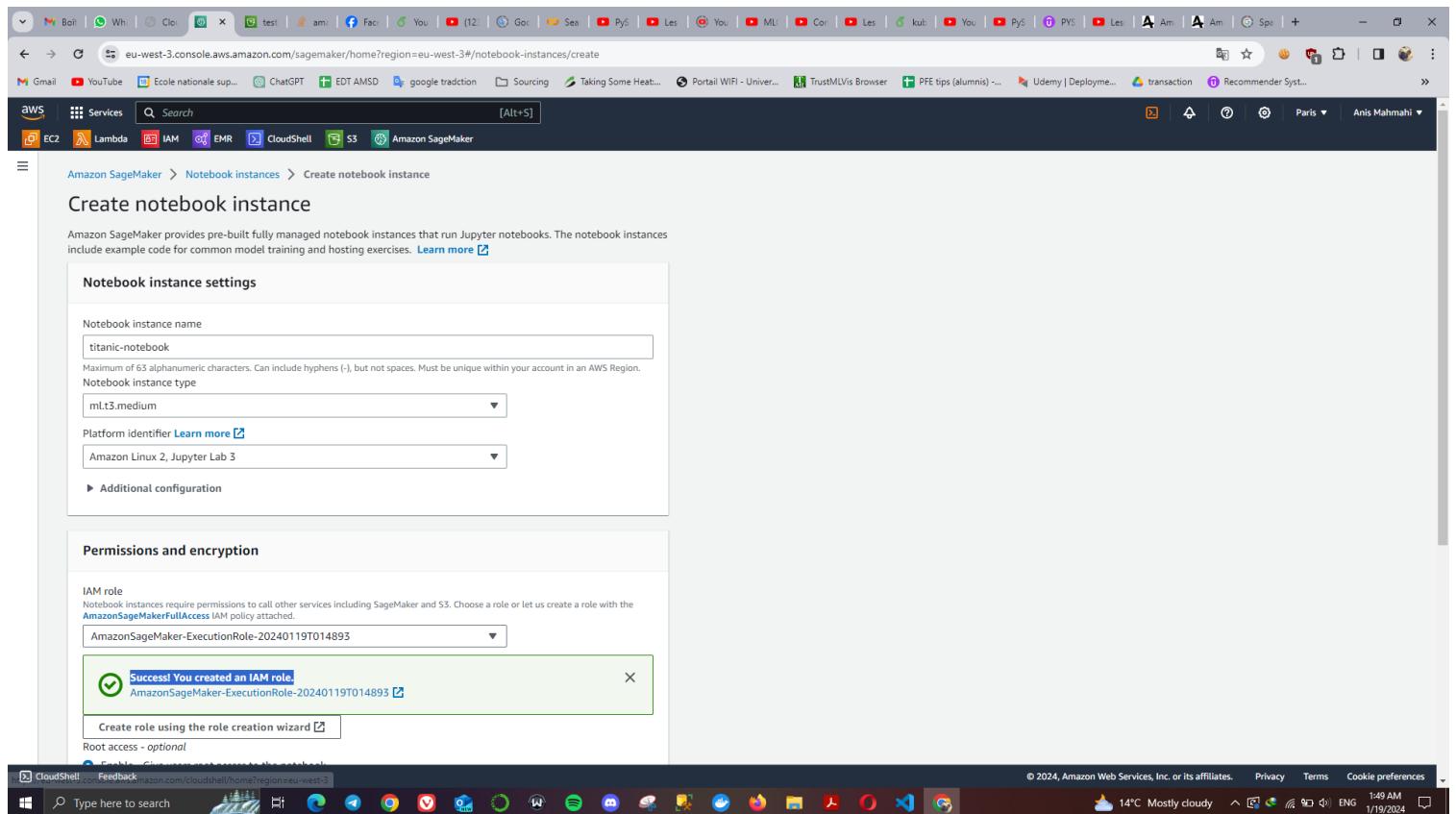


The screenshot shows the AWS S3 console interface. The user is viewing the contents of a folder named 'processed' within the 'test-bucket-hattabi' bucket. The 'Objects' tab is selected, displaying one item: 'titanic.csv'. The object details are as follows:

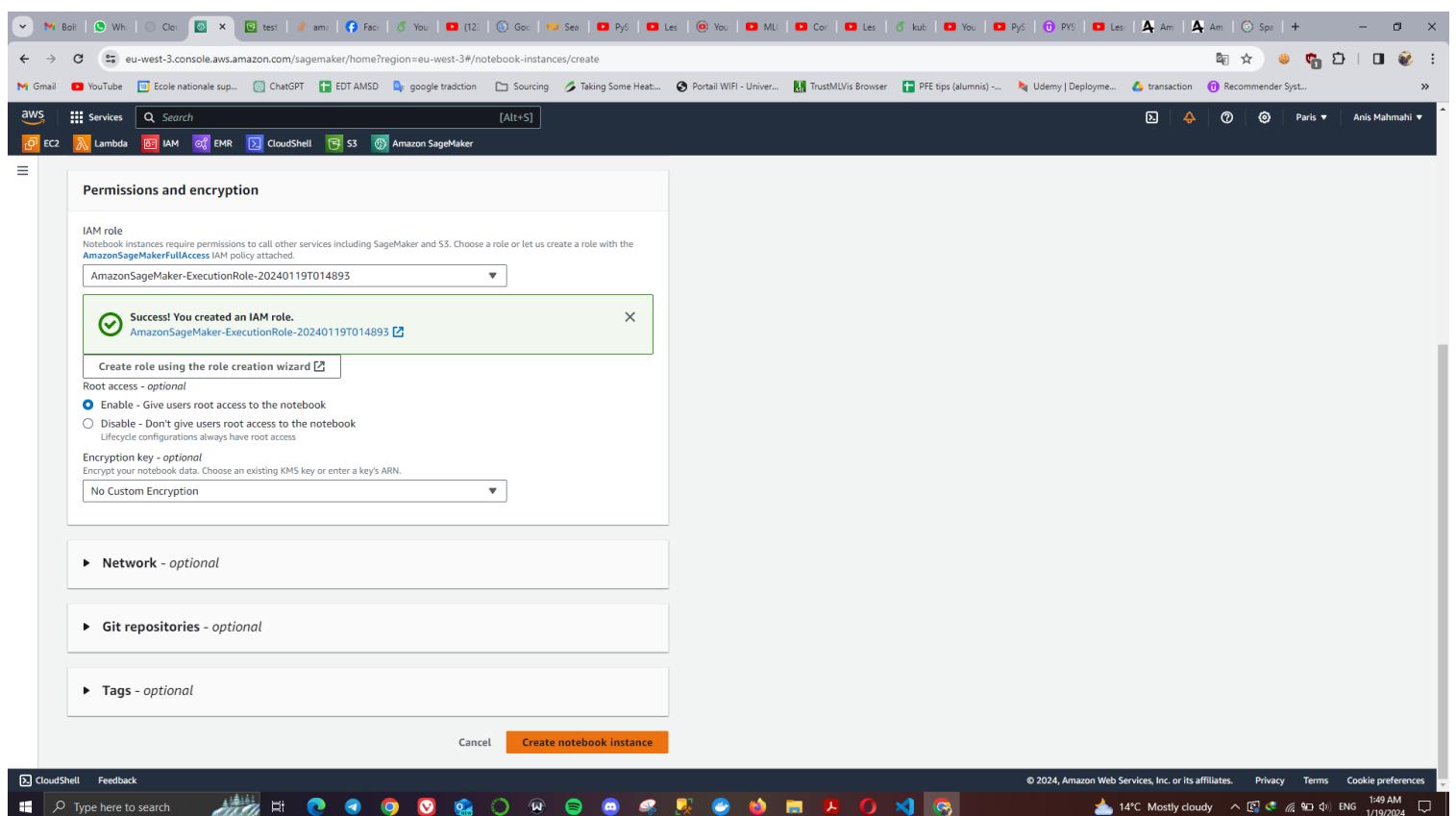
Name	Type	Last modified	Size	Storage class
titanic.csv	csv	January 14, 2024, 20:34:22 (UTC+01:00)	81.3 KB	Standard

6 Amazon SageMaker

On commence par créer un notebook en spécifiant le type d'instance, le système d'opération et le rôle IAM :



The screenshot shows the AWS CloudShell interface with the URL eu-west-3.console.aws.amazon.com/sagemaker/home?region=eu-west-3#/notebook-instances/create. The navigation bar includes links for Gmail, YouTube, Ecole nationale sup..., ChatGPT, EDT AMSD, google traduction, Sourcing, Taking Some Heat..., Portal WiFi - Université de Paris, TrustMLVis Browser, PFE tips (alumni) ~, Udemy | Deployement, transaction, Recommender Syst..., and Paris. The main content area is titled "Create notebook instance". It contains two sections: "Notebook instance settings" and "Permissions and encryption". In the "Notebook instance settings" section, the "Notebook instance name" is set to "titanic-notebook", "Notebook instance type" is "mlt.t3.medium", and "Platform identifier" is "Amazon Linux 2, Jupyter Lab 3". A success message "Success! You created an IAM role." is displayed above the "Create role using the role creation wizard" button. In the "Permissions and encryption" section, the "IAM role" dropdown is set to "AmazonSageMaker-ExecutionRole-20240119T014893". The status bar at the bottom shows "CloudShell Feedback" and the system status "14°C Mostly cloudy".



This screenshot continues the process of creating a new Amazon SageMaker notebook instance. The "Permissions and encryption" section is shown again, with the "IAM role" set to "AmazonSageMaker-ExecutionRole-20240119T014893" and a success message "Success! You created an IAM role." displayed. Below this, there are options for "Root access" (radio buttons for "Enable" and "Disable"), "Encryption key" (dropdown set to "No Custom Encryption"), and "Lifecycle configurations" (checkbox checked). The interface then branches into three optional sections: "Network - optional", "Git repositories - optional", and "Tags - optional". At the bottom right is a large orange "Create notebook instance" button.

FIG. 1 : Crédit : création du notebook

- Le notebook a été bien créé et prêt à être utilisé :

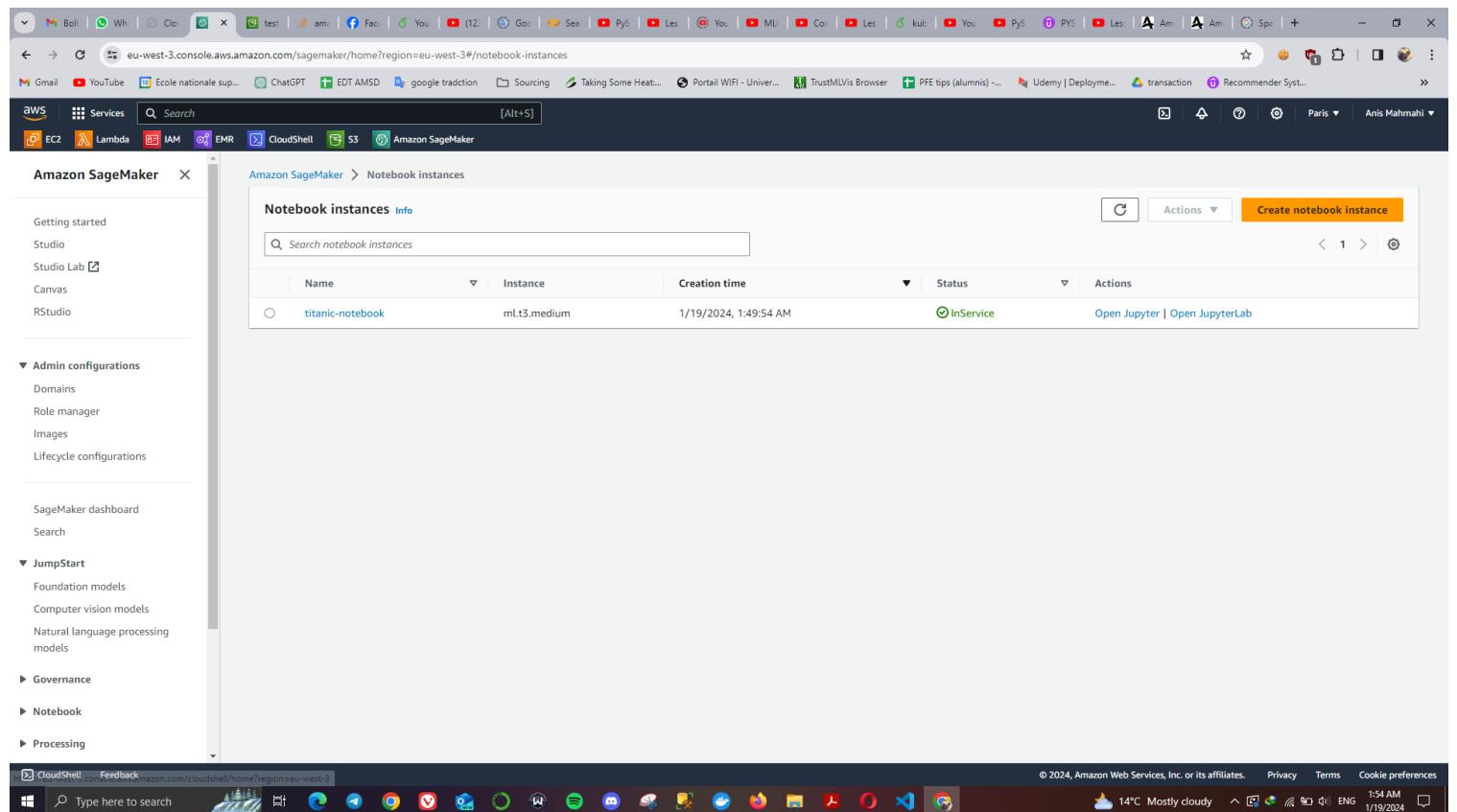


FIG. 2 : Notebook créé

7 Préparation des données

- Dans cette section, on accède au dataset stocké dans le bucket s3 **test-bucket-hattabi** à travers de code.
- On a supprimé la colonne "**Ticket**" en raison de son inutilité, et divisé le dataset en deux parties, soit 80% pour l'entraînement et 20% pour le test.
- On a sauvegardé le dataset d'entraînement et celui de test dans un répertoire **s3://test-bucket-hattabi/train_test**.
- On a également sauvegardé les features du dataset de test, à l'exception de la variable cible, pour les utiliser lors de la phase de prédiction.

Table des matières

The screenshot shows a Jupyter Notebook interface with a code cell containing Python code. The code imports pandas, drops the 'Ticket' column, splits the data into training and test sets, and then saves the datasets as CSV files. It also imports boto3 and uses it to upload the CSV files to an S3 bucket named 'test-bucket-hattabi'. The terminal output at the bottom shows the command 'Upload Successful'.

```

[15]: df.dtypes
[15]: PassengerId      int64
       Survived        int64
       Pclass          int64
       Age             float64
       SibSp          int64
       Parch          int64
       Ticket         object
       Fare            float64
       Sex_male       int64
       Embarked_Q     int64
       Embarked_S     int64
       Title           int64
       dtype: object

[16]: df.drop(['Ticket'], axis=1, inplace=True)

[17]: from sklearn.model_selection import train_test_split
      # Split data into train and test sets
      X_train,X_test,y_train,y_test = train_test_split(df.drop(['Survived'], axis=1),df['Survived'], test_size=0.2,random_state=42)

[18]: import pandas as pd
      # Combine the features and the target variable for train and test sets
      train_data = pd.concat([y_train,X_train],axis=1)
      test_data = pd.concat([y_test,X_test],axis=1)
      test_features = pd.concat([X_test],axis=1)

      # Save the datasets as CSV files
      train_data.to_csv('train.csv', index=False,header=False)
      test_data.to_csv('test.csv', index=False,header=False)
      test_features.to_csv('test_features.csv', index=False,header=False)

[19]: import boto3
      import sagemaker
      # Create a new Sagemaker session
      sagemaker_session = sagemaker.Session()
      # Get the S3 client
      s3 = boto3.client('s3')
      # Upload the train set
      s3.upload_file('train.csv', bucket, f'{prefix}/train.csv')
      # Upload the test set
      s3.upload_file('test.csv', bucket, f'{prefix}/test.csv')
      # Upload the test set features only
      s3.upload_file('test_features.csv', bucket, f'{prefix}/test_features.csv')
      print("Upload Successful")
      Upload Successful
  
```

FIG. 4 : Séparation des données et les sauvegarder dans s3

The screenshot shows the AWS S3 console with the path 'Amazon S3 > Buckets > test-bucket-hattabi > train_test/'. The 'Objects' tab is selected, showing three CSV files: 'test_features.csv', 'test.csv', and 'train.csv'. Each file has a size of approximately 4.4 KB, 4.7 KB, and 18.8 KB respectively, all stored in the 'Standard' storage class. The 'Properties' tab is also visible.

Name	Type	Last modified	Size	Storage class
test_features.csv	csv	January 19, 2024, 13:58:01 (UTC+01:00)	4.4 KB	Standard
test.csv	csv	January 19, 2024, 13:58:01 (UTC+01:00)	4.7 KB	Standard
train.csv	csv	January 19, 2024, 13:58:01 (UTC+01:00)	18.8 KB	Standard

FIG. 5 : Répartition des données stockées dans s3

8 Training job

Ce code configure un modèle XGBoost, spécifie les données d'entraînement et de test, définit les hyperparamètres du modèle, puis lance le processus d'entraînement.

```

[21]: from sagemaker import Image_uris
import sagemaker
from sagemaker import get_execution_role
container = image_uris.retrieve("xgboost", regions='eu-west-3', version='latest')
INFO:sagemaker_image_uris:Ignoring unnecessary instance type: None.

[22]: s3_input_train = sagemaker.inputs.TrainingInput(s3_data="s3://(/)/titanic.csv".format(bucket,prefix),content_type="csv")
s3_input_test = sagemaker.inputs.TrainingInput(s3_data="s3://(/)/test.csv".format(bucket,prefix),content_type="csv")

[23]: sess = sagemaker.Session()
role = get_execution_role()
xgb = sagemaker.estimator.Estimator(container_role,
                                       instance_count=1,
                                       instance_type="ml.m5.large",
                                       output_path="s3://(/)/output".format(bucket),
                                       sageMaker_session=sess)

xgb.set_hyperparameters(max_depth=5,
                       eta=0.2,
                       gamma=4,
                       min_child_weight=6,
                       verbosity=1,
                       objective='binary:logistic',
                       num_rounds=40,
                       eval_metric='auc') # Add this line to monitor AUC

[26]: xgb.fit(['train':s3_input_train,'validation':s3_input_test])
INFO:sagemaker:Creating training job with name: xgboost-2024-01-19-12-41-43-265
2024-01-19 12:41:43 Starting - Starting the training job...
2024-01-19 12:42:11 Starting - Preparing the instances for training.....
2024-01-19 12:43:32 Downloading - Downloading input data.....
2024-01-19 12:44:27 Downloading - Downloading the training image.. Arguments: train
[2024-01-19 12:44:57:INFO] Running standalone xgboost training.
[2024-01-19 12:44:57:INFO] File size need to be processed in the node: 0.02mb. Available memory size in the node: 354.82mb
[2024-01-19 12:44:57:INFO] Determined delimiter of CSV Input is ','
[12:44:57] S3DistributionType set as FullyReplicated
[12:44:57] 7129x matrix with 6408 entries loaded from /opt/ml/input/data/train?format=csv&label_column=0&delimiter=,
[2024-01-19 12:44:57:INFO] Determined delimiter of CSV Input is ','
[12:44:57] S3DistributionType set as FullyReplicated
[12:44:57] 1790x matrix with 1611 entries loaded from /opt/ml/input/data/validation?format=csv&label_column=0&delimiter=,
[12:44:57] src/tree/updater_prune.cc:74: tree pruning end, 1 roots, 10 extra nodes, 8 pruned nodes, max_depth=3
[0]#0@11train-auc:0.858772#0@1validation-auc:0.868983

```

FIG. 6 : Création d'un trainig job

Name	Creation time	Duration	Job status	Warm pool status	Time left
xgboost-2024-01-19-12-41-43-265	1/19/2024, 1:41:43 PM	4 minutes	Completed	-	-
xgboost-2024-01-19-12-27-31-229	1/19/2024, 1:27:31 PM	3 minutes	Completed	-	-
xgboost-2024-01-19-12-19-54-664	1/19/2024, 1:19:54 PM	3 minutes	Failed	-	-
xgboost-2024-01-19-12-16-05-862	1/19/2024, 1:16:06 PM	3 minutes	Failed	-	-
sagemaker-xgboost-2024-01-19-02-06-22-586	1/19/2024, 3:06:23 AM	2 minutes	Stopped	-	-
sagemaker-sklearn-2024-01-11-14-30-29-037	1/11/2024, 3:30:29 PM	3 minutes	Completed	-	-

FIG. 7 : Training job a été bien exécuté

9 Hyperparameters tuning job

Le tuner d'hyperparamètres va explorer différentes combinaisons d'hyperparamètres pour le modèle XGBoost et chercher à trouver la meilleure configuration qui maximise la métrique "validation :auc" tout en respectant les plages spécifiées. Le nombre total de modèles à entraîner lors de la recherche d'hyperparamètres (20 dans notre cas).

```

Start HyperParameters tuning job

[47]: from sagemaker.tuner import IntegerParameter, ContinuousParameter, HyperparameterTuner

# Define the hyperparameter ranges
hyperparameter_ranges = {
    'max_depth': IntegerParameter(3, 10),
    'eta': ContinuousParameter(0.01, 0.2),
    'min_child_weight': IntegerParameter(1, 10),
    'subsample': ContinuousParameter(0.5, 1),
    'gamma': ContinuousParameter(0, 5)
}

# Specify the objective metric that we'd like to tune and its definition
objective_metric_name = 'validation:auc'. # Example metric
objective_type = 'Maximize' # Can be 'Maximize' or 'Minimize' depending on the metric

# Configure the tuner object
tuner = HyperparameterTuner(
    estimator=xgb, # The estimator object to use as the basis for the training job.
    objective_metric_name=objective_metric_name, # The metric used to compare trained models.
    hyperparameter_ranges=hyperparameter_ranges, # The range of hyperparameters to tune.
    metric_definitions=[{'Name': objective_metric_name, 'Regex': 'validation-auc:[0-9.]+'}], # The regex to extract the metric from the logs
    max_jobs=20, # The total number of models to train
    max_parallel_jobs=3, # The number of models to train in parallel
    objective_type=objective_type
)

# Start the hyperparameter tuning job
tuner.fit({'train': s3_input_train, 'validation': s3_input_test})

INFO:sagemaker:Creating hyperparameter tuning job with name: xgboost-240119-1322
.....!

```

Batch Prediction

+ 5 cells hidden

Evaluation

+ 3 cells hidden

FIG. 8 : Code de la recherche d'hyperparamètres

Table des matières

Amazon SageMaker > Hyperparameter tuning jobs

Hyperparameter tuning jobs

Name	Status	Training completed/total	Creation time	Duration
xgboost-240119-1322	Completed	20 / 20	1/19/2024, 2:22:48 PM	9 minutes

Getting started
Studio
Studio Lab
Canvas
RStudio

Admin configurations
Domains
Role manager
Images
Lifecycle configurations

SageMaker dashboard
Search

JumpStart
Foundation models
Computer vision models
Natural language processing models

Governance
Notebook
Processing

https://eu-west-3.console.aws.amazon.com/sagemaker/home?region=eu-west-3#hyper-tuning-jobs/xgboost-240119-1322?tab=trainingJobs

FIG. 9 : 20 modèles différents exécutés avec succès.

Amazon SageMaker > Training job status counter

Completed 20 In Progress 0 Stopped 0 Failed 0 (Retryable: 0, Non-retryable: 0)

Training jobs

Name	Status	Final objective metric value	Creation time	Training Duration
xgboost-240119-1322-020-3dbb8bf	Completed	0.8980050086975098	1/19/2024, 2:31:23 PM	1 minute(s)
xgboost-240119-1322-019-737ae82e	Completed	0.9092659950256348	1/19/2024, 2:30:56 PM	1 minute(s)
xgboost-240119-1322-018-f2eb2f4a	Completed	0.9076579809188843	1/19/2024, 2:30:54 PM	1 minute(s)
xgboost-240119-1322-017-bd0b736f	Completed	0.9099739789962769	1/19/2024, 2:30:25 PM	1 minute(s)
xgboost-240119-1322-016-41c37be1	Completed	0.907014012336731	1/19/2024, 2:29:58 PM	1 minute(s)
xgboost-240119-1322-015-85f0f2ac	Completed	0.9016090035438538	1/19/2024, 2:29:36 PM	1 minute(s)
xgboost-240119-1322-014-7b0c6603	Completed	0.8989059925079346	1/19/2024, 2:29:18 PM	1 minute(s)
xgboost-240119-1322-013-5e0957bc	Completed	0.9036039710044861	1/19/2024, 2:29:05 PM	1 minute(s)
xgboost-240119-1322-012-b7660c97	Completed	0.9028310179710388	1/19/2024, 2:28:28 PM	1 minute(s)
xgboost-240119-1322-011-fae49310	Completed	0.897104024887085	1/19/2024, 2:28:15 PM	1 minute(s)

Getting started
Studio
Studio Lab
Canvas
RStudio

Admin configurations
Domains
Role manager
Images
Lifecycle configurations

SageMaker dashboard
Search

JumpStart
Foundation models
Computer vision models
Natural language processing models

Governance
Notebook
Processing

https://eu-west-3.console.aws.amazon.com/sagemaker/home?region=eu-west-3#hyper-tuning-jobs/xgboost-240119-1322?tab=trainingJobs

FIG. 10 : 10 premiers modèles

Table des matières

Training job status counter

Completed: 20 In Progress: 0 Stopped: 0 Failed: 0 (Retryable: 0, Non-retryable: 0)

Training jobs

Sorting by objective metric value will display only jobs that have metric values.

Name	Status	Final objective metric value	Creation time	Training Duration
xgboost-240119-1322-010-f5023cd3	Completed	0.8853279948234558	1/19/2024, 2:28:05 PM	1 minute(s)
xgboost-240119-1322-009-cea335d5	Completed	0.8839769959449768	1/19/2024, 2:27:29 PM	1 minute(s)
xgboost-240119-1322-008-ba0aa4ed	Completed	0.9068210124969482	1/19/2024, 2:27:20 PM	1 minute(s)
xgboost-240119-1322-007-4644f96c	Completed	0.9022520184516907	1/19/2024, 2:27:10 PM	1 minute(s)
xgboost-240119-1322-006-a2b1def2	Completed	0.9104250073432922	1/19/2024, 2:26:26 PM	1 minute(s)
xgboost-240119-1322-005-dbee95d5	Completed	0.9032819867134094	1/19/2024, 2:26:13 PM	1 minute(s)
xgboost-240119-1322-004-a41f4b8a	Completed	0.8859720230102539	1/19/2024, 2:26:11 PM	1 minute(s)
xgboost-240119-1322-003-e5555726	Completed	0.9024450182914734	1/19/2024, 2:22:56 PM	2 minute(s)
xgboost-240119-1322-002-eab33567	Completed	0.8847489953041077	1/19/2024, 2:22:55 PM	2 minute(s)
xgboost-240119-1322-001-9a4cfdec	Completed	0.913320004940033	1/19/2024, 2:22:53 PM	2 minute(s)

FIG. 11 : les 10 modèles restants

Best training job summary

This training job is the best training job for only this hyperparameter tuning job.

Name	Status	Objective metric	Value
xgboost-240119-1322-001-9a4cfdec	Completed	validation:auc	0.913320004940033

Best training job hyperparameters

Name	Type	Value
_tuning_objective_metric	FreeText	validation:auc
eta	Continuous	0.13077357310460885
eval_metric	FreeText	auc
gamma	Continuous	2.2879899294244583
max_depth	Integer	6
min_child_weight	Integer	3
num_round	FreeText	40
objective	FreeText	binary:logistic
subsample	Continuous	0.9725941126497504
verbosity	FreeText	1

FIG. 12 : les hyperparamètres du meilleur modèle

10 Prédiction par lots

- Télécharger le meilleur model trouvé durant la phase (hyper-parameters tuning).

```

Batch Prediction

[4]: best_training_job_name = "xgboost-240119-1322-001-3edcfdec"
best_estimator = sagemaker.estimator.Estimator.attach(best_training_job_name)

2024-01-19 13:26:31 Starting - Preparing the instances for training
2024-01-19 13:26:31 Downloading - Downloading the training image
2024-01-19 13:26:31 [train] - Training image download completed. Training in progress.
2024-01-19 13:26:31 Uploading - Uploading generated training model
2024-01-19 13:26:31 Completed - Resource reused by training job: xgboost-240119-1322-006-a2b1def2

[5]: best_estimator.hyperparameters()

[5]: {"tuning_objective_metric": "validation:auc",
       "eta": "0.0173573310460885",
       "eval_metric": "auc",
       "gamma": "2.0879899294244583",
       "max_depth": "6",
       "min_child_weight": "3",
       "num_round": "40",
       "objective": "binary:logistic",
       "subsample": "0.9725941126497504",
       "verbosity": "1"}

[6]: best_model_transformer = best_estimator.transformer(
    instance_count=1,
    instance_type='ml.m5.large',
    output_path='s3://test-bucket-hattabi/predictions/' # Specify your output path
)

[10]: best_model_transformer.transform(
    data='s3://test-bucket-hattabi/train_test/test_features.csv', # Specify your input data path
    content_type='text/csv', # The content type of the input data
    split_type='Line', # How the input data is split
)
best_model_transformer.wait()

INFO:sagemaker:Creating transform job with name: xgboost-2024-01-25-08-19-58-117
.....
Arguments: serve
[2024-01-25 08:24:30 +0000] [1] [INFO] Starting gunicorn 19.9.0
[2024-01-25 08:24:30 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 ([1])
[2024-01-25 08:24:30 +0000] [20] [INFO] Worker created.
[2024-01-25 08:24:30 +0000] [20] [INFO] Booting worker with pid: 20
[2024-01-25 08:24:30 +0000] [21] [INFO] Booting worker with pid: 21
/opt/amazon/lib/python3.7/site-packages/gunicorn/workers/ggevent.py:65: MonkeyPatchWarning: Monkey-patching ssl after ssl has already been imported may lead to errors, including RecursionError on Python 3.6. It may also silently lead to incorrect behaviour on Python 3.7. Please monkey-patch earlier. See https://github.com/gevent/gevent/issues/1016. Modules that had direct imports (NOT patched): ['urllib3.util.ssl_'] (/opt/amazon/lib/python3.7/site-packages/urllib3/util/_init__.py')

```

FIG. 13 : Code pour exécuter la prédiction par lots (Batch Transform)

- Même si la machine "ml.m5.large" est inclus dans l'offre gratuit mais l'exécution du code dessus pose une erreur et qui nous indique que il faut contacter AWS et leur demander d'augmenter le quota dédié pour "Transform job" car par default il est 0.

Table des matières

```

(...)

6140:     # pylint: disable=unused-argument
6141:     """
6142:         This function intercepts the create job request.
6143:
6144:         PipelineSession inherits this Session class and will override
6145:             ...
6146:             func_name (str): the name of the function needed intercepting
6147:             ...
6148:             return
6149:     """
6150:     def __init__(self, func_name):
6151:         self._func_name = func_name
6152:         return create(request)

File ~/anaconda3/envs/pytorch_p310/lib/python3.10/site-packages/sagemaker/session.py:3560, in Session.transform.<locals>.submit(request)
3558 logger.info("Creating transform job with name: %s", job_name)
3559 logger.debug("Transform request: %s", json.dumps(request, indent=4))
-> 3560 self.sagemaker_client.create_transform_job(*request)

File ~/anaconda3/envs/pytorch_p310/lib/python3.10/site-packages/botocore/client.py:553, in ClientCreator._create_api_method.<locals>._api_call(self, *args, **kwargs)
549     raise TypeError(
550         f"(py_operation_name)() only accepts keyword arguments."
551     )
552 # The "self" in this scope is referring to the BaseClient.
--> 553 return self._make_api_call(operation_name, kwargs)

File ~/anaconda3/envs/pytorch_p310/lib/python3.10/site-packages/botocore/client.py:1009, in BaseClient._make_api_call(self, operation_name, api_params)
1005     error_code = error_info.get("QueryErrorCode") or error_info.get(
1006         "Code"
1007     )
1008     error_class = self.exceptions.from_code(error_code)
-> 1009     raise error_class(parsed_response, operation_name)
1010 else:
1011     return parsed_response

ResourceLimitExceeded: An error occurred (ResourceLimitExceeded) when calling the CreateTransformJob operation: The account-level service limit 'ml.m5.large for transform job usage' is 0 Instances, with current utilization of 0 Instances and a request delta of 1 Instances. Please use AWS Service Quotas to request an increase for this quota. If AWS Service Quotas is not available, contact AWS support to request an increase for this quota.

```

Starting the HyperParameters tunning

```

[ ]: from sagemaker.tuner import IntegerParameter, ContinuousParameter, HyperparameterTuner
hyperparameter_ranges = {
    'max_depth': IntegerParameter(3, 10),
    'eta': ContinuousParameter(0.05, 0.5),
    'min_child_weight': IntegerParameter(1, 10)
}

```

Would you like to receive official Jupyter news? Please read the privacy policy.

Open privacy policy Yes No

FIG. 14 : Erreur du manque de machine pour exécuter le code

Case ID 170566970201830

Case details	
Subject	Status
Limit Increase: SageMaker Batch Transform	Unassigned
Case ID	Severity
170566970201830	General question
Created	Category
2024-01-19T13:08:22.050Z	Service Limit Increase, SageMaker Batch Transform
Case type	Language
Service limits	English
Opened by	Additional contacts
ja_mahmahi@esi.dz	-

Correspondence

Anis Mahmahi
Fri Jan 19 2024
14:08:22 GMT+0100
(Central European Standard Time)

Limit increase request 1
Region: EU (Paris)
Resource Type: SageMaker Batch Transform
Service: SageMaker Batch Transform
Limit name: ml.m5.large
New limit value: 1

Use case description: Hello,
I'm trying to run batch prediction on ml.m5.large but I got this error:
ResourceLimitExceeded: An error occurred (ResourceLimitExceeded) when calling the CreateTransformJob operation: The account-level service limit 'ml.m5.large for transform job usage' is 0 Instances, with current utilization of 0 Instances and a request delta of 1 Instances. Please use AWS Service Quotas to request an increase for this quota. If AWS Service Quotas is not available, contact AWS support to request an increase for this quota.

FIG. 15 : Contacter AWS concernant le problème

Table des matières

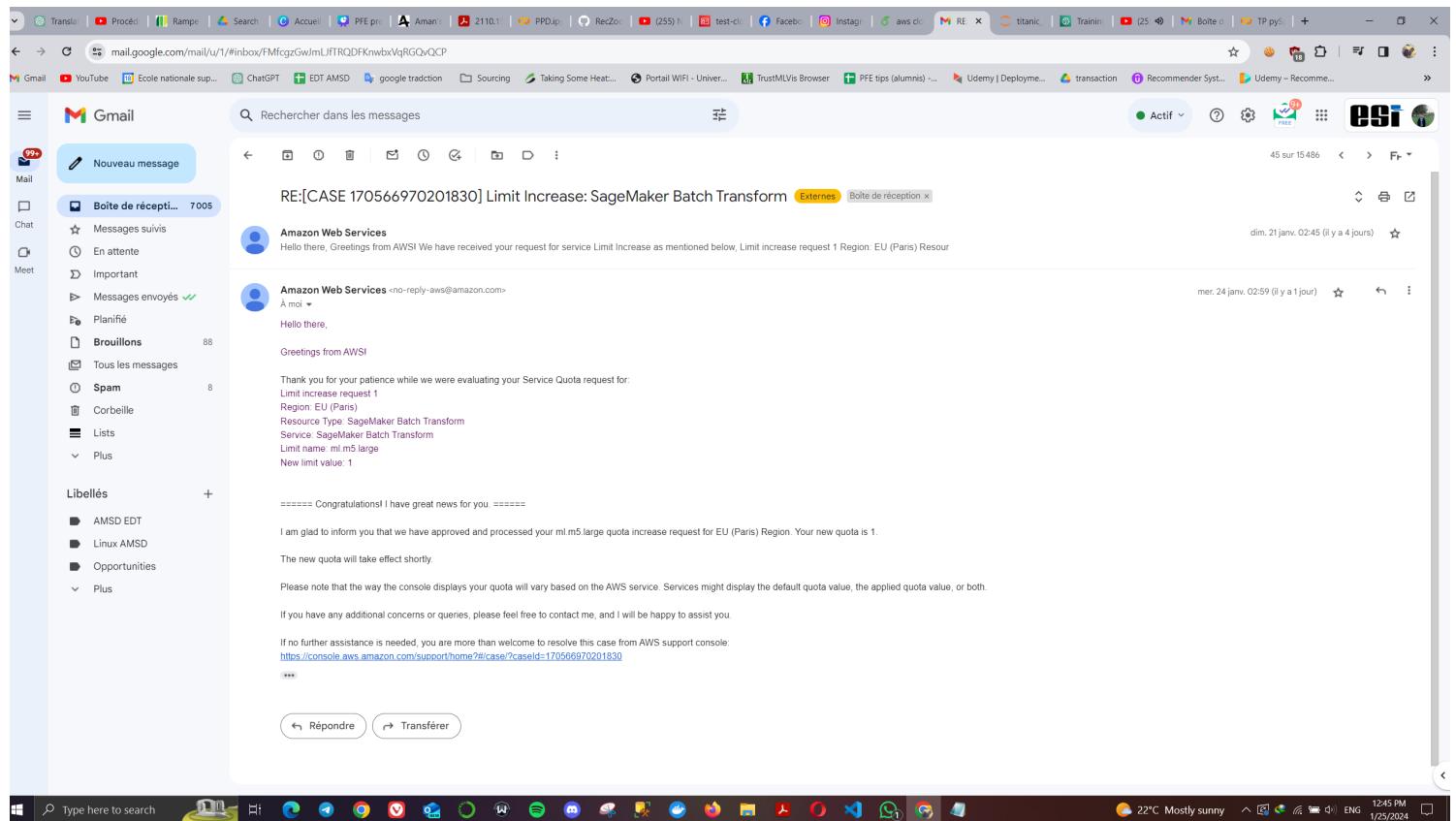


FIG. 16 : Réponse positive de l'apart du support

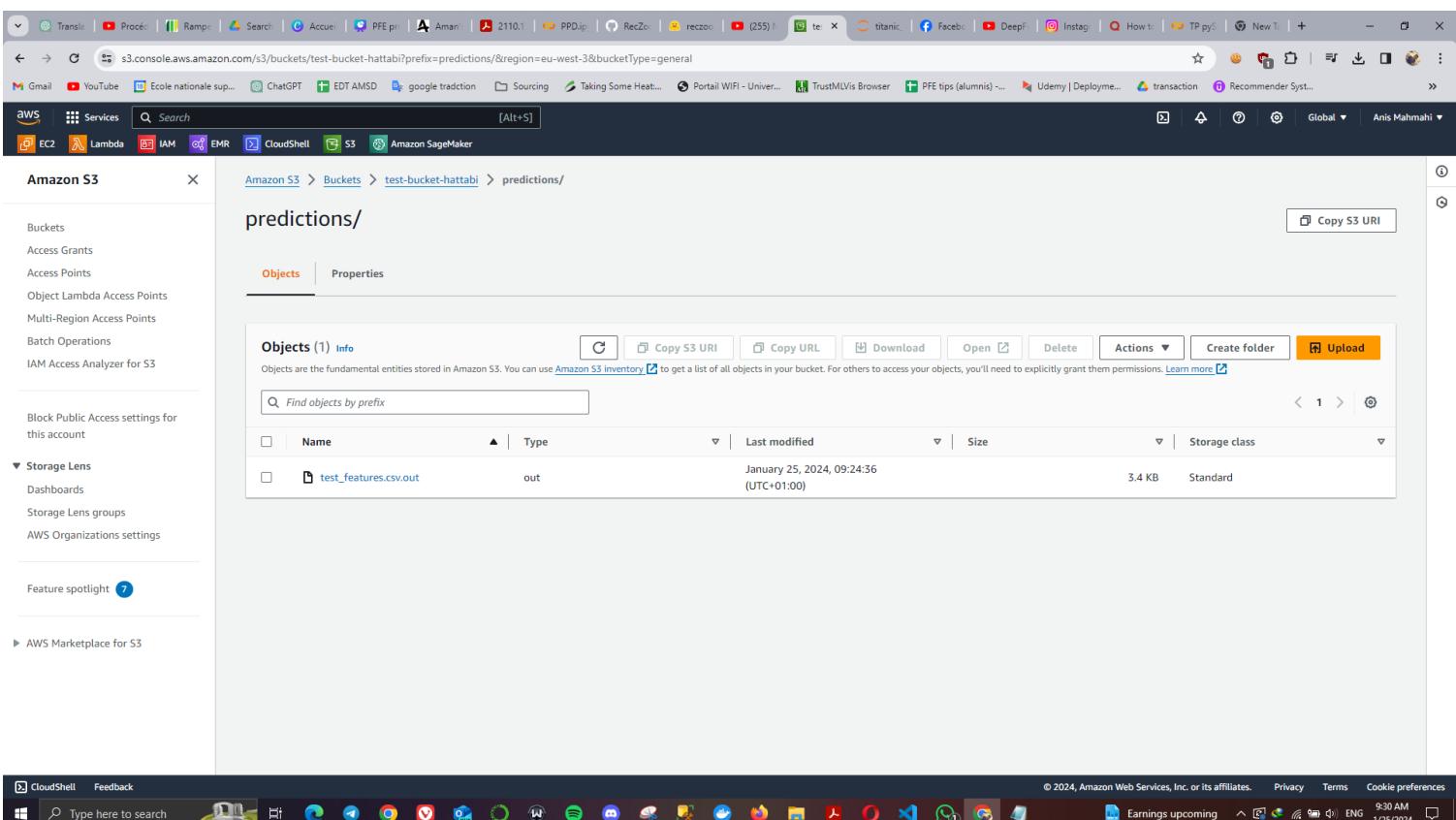


FIG. 17 : Sauvegarde des prédictions dans S3

Table des matières

The screenshot shows a Jupyter Notebook interface within a browser window. The left sidebar displays a file tree with files like 'new-workspace/jupyterlab-workspace', 'test_features.csv', 'test_features.csv.out', 'test.csv', and 'titanic_notebook.ipynb'. The main area contains Python code for evaluation:

```
[25]: import boto3
import os
s3 = boto3.client('s3')
# Bucket Name where the file is located
bucket_name = 'test-bucket-hattabi'
# Path in S3
s3_path = 'predictions/test_features.csv.out'
# Local file name to save the data
local_file_name = 'test_features.csv.out'
# Download the file from S3
s3.download_file(bucket_name, s3_path, local_file_name)

# Threshold
threshold = 0.5

# Read the probabilities from the file (assuming the file is named 'predictions.txt')
with open('./test_features.csv.out', 'r') as file:
    probabilities = file.readlines()

# Convert probabilities to class labels
y_pred = [int(float(prob.strip())) >= threshold for prob in probabilities]

test = pd.read_csv("s3://test-bucket-hattabi/train_test/test.csv", header=None)
y_true = test.iloc[:,0]

[30]: from sklearn.metrics import classification_report
print(classification_report(y_true, y_pred))

precision    recall  f1-score   support

          0       0.82      0.87      0.84     105
          1       0.79      0.73      0.76      74

   accuracy                           0.81      179
  macro avg       0.81      0.80      0.80      179
weighted avg       0.81      0.81      0.81      179
```

FIG. 18 : Evaluation des prédictions obtenus