Redjechta

Mohamed Yacine

ING3 FISA

CC1 Pipeline

**Partie 1 :**

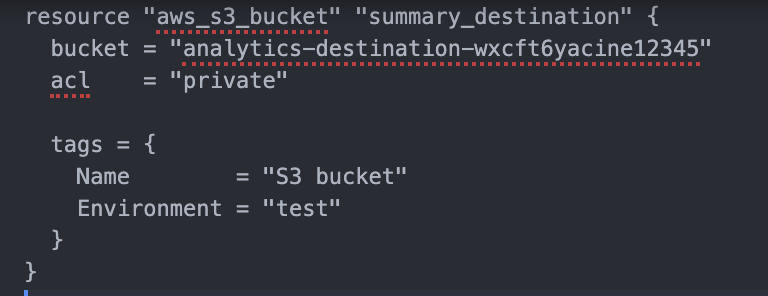
Les ressources manquantes :

destination\_s3\_bucket.tf : Contient les informations sur la Bucket S3 à créer.

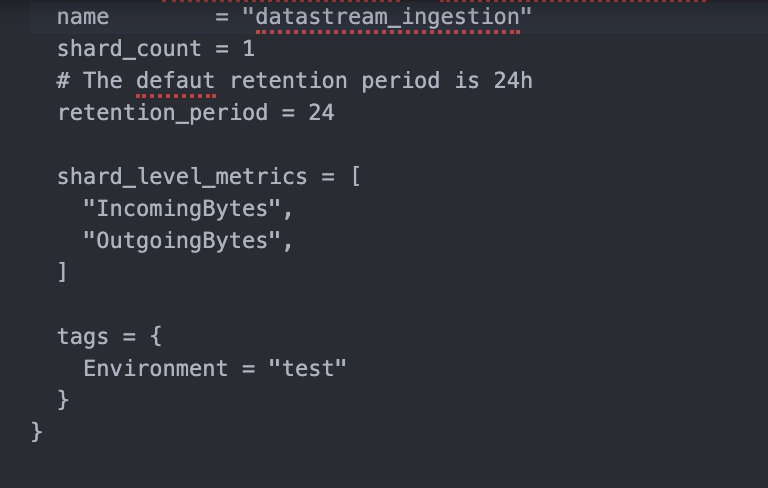
kinesis\_datastream.tf : Contient les informations sur l’input data stream.

Correction du code :

**destination\_s3\_bucket.**tf : bien mettre le « Summary\_destination » car c’est comme ça qu’elle est défini dans **kinesis\_delivery\_stream\_summary.tf**

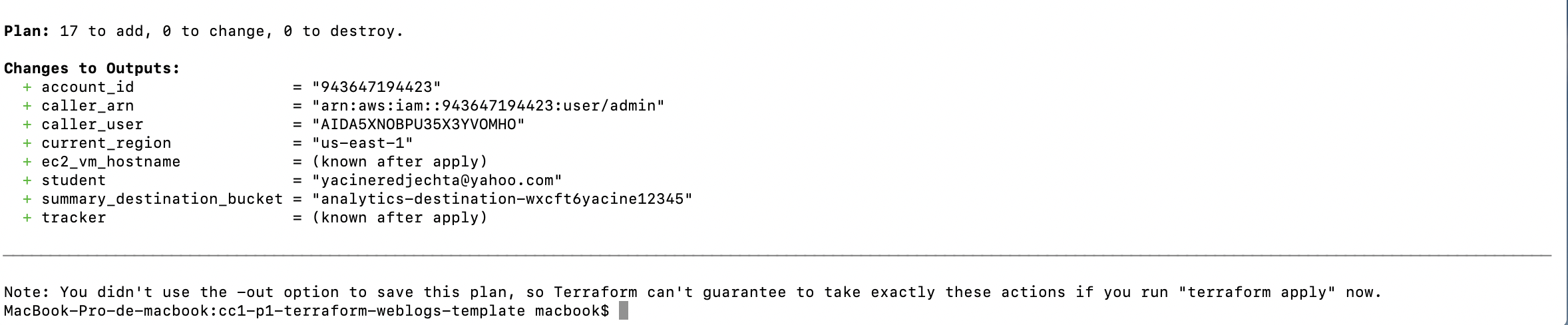


**Kinesis\_datastream.tf :** bien mettre « datastream\_ingestion » car c’est comme ça qu’elle est défini dans **kinesis\_analytics\_application.tf**

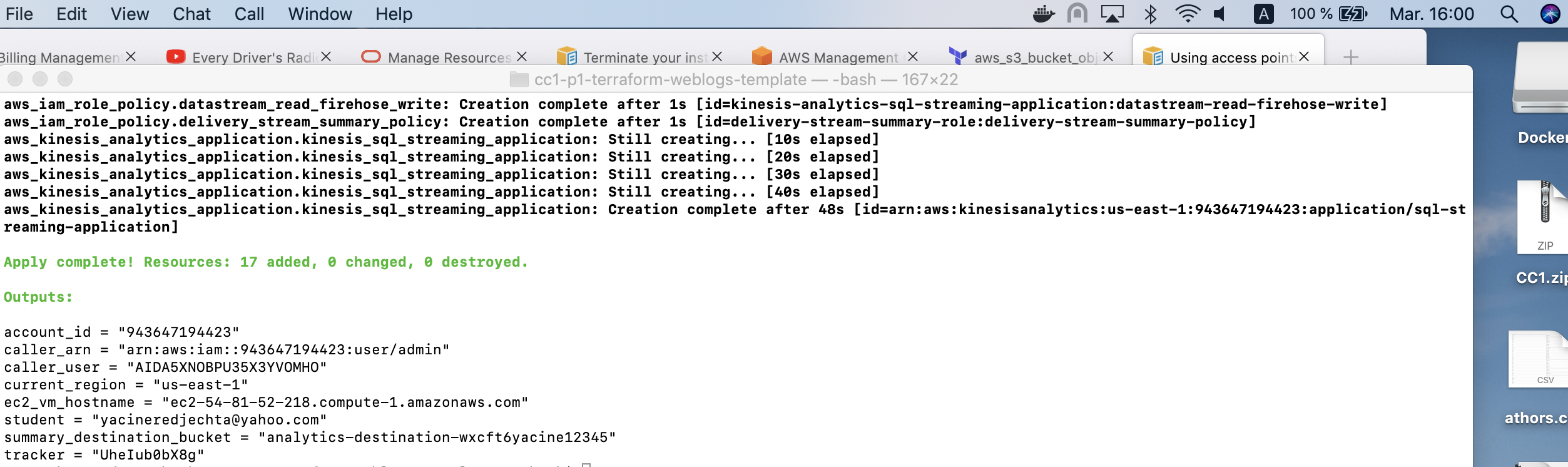


Test de la Data Pipeline :

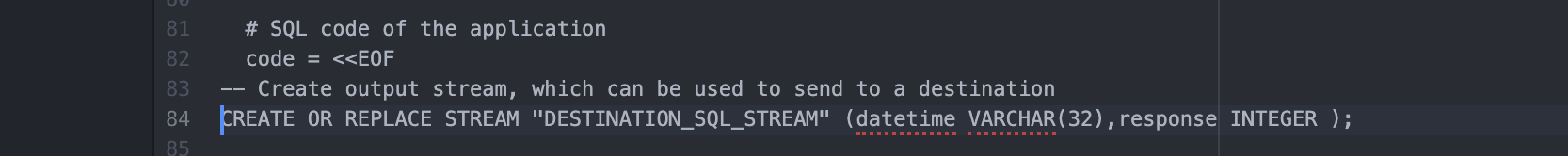
En faisant un terraform init ensuite terraform plan on obtient donc ça.



En faisant maintenant terraform apply, on crée bien nos ressources



Le code n’est pas bon car sur l’application analytics il faut créer l’output stream en ajoutant ce code :



Explication de ce que fait l’application :

* Création de bucket S3 dans aws
* Création de VM EC2 et la mettre dans S3
* Alimenter des des flux de Amazon Kinesis Data Streams à l'aide de Kinesis Agent (Logs) et les traiter via Kinesis analytics

A quoi sert la partie « user\_data » dans le fichier « ec2.tf » :

Cette partie sert à installer les différentes composantes dans le VM EC2 nécéssaires à l’exécution de l’application et de donner les autorisatons à accéder au fichier des logs qui sont principalement pip et des bibliothèques de pip et aws-kinesis-agent.

Quel est le role de l’agent Kinesis sur la VM EC2 :

LE role de l'agent Kinesis sur la VM EC2 est de collecter et d'envoyer les données générées par le script python (producer) dans le Kinesis Data Streams .

**Partie 2 :**

Quelle est l’architecture de données de ce « data pipeline »

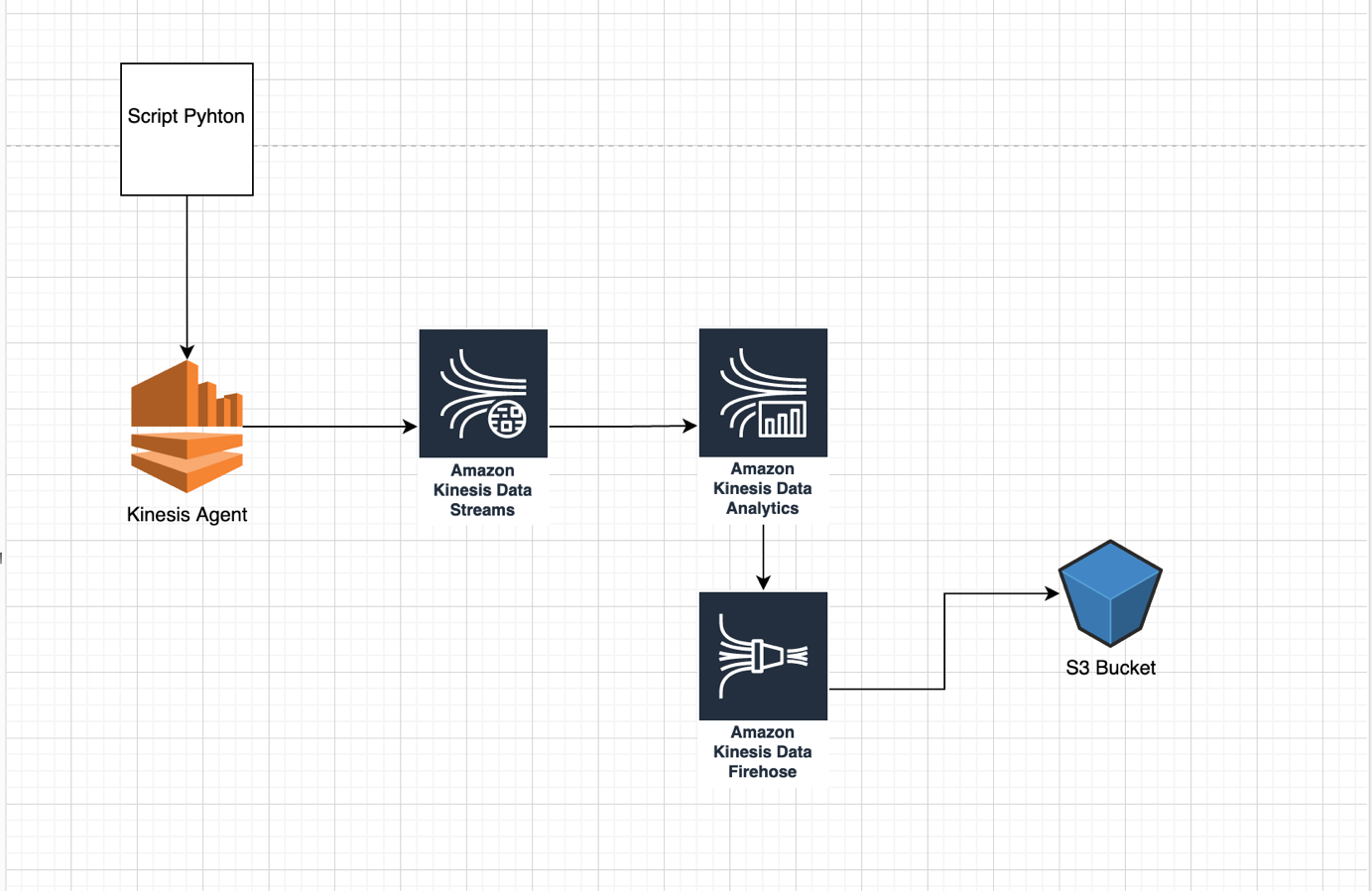
**Ingestion** correspond dans notre application à la récupération des données qui sont ici des logs générés par le script python

Les données sont ensuite chargées grâce kinesis dataStream qui va lire ce que produit le script python ( logs ) et les ingérer dans notre pipline pour qu’on puisse les traiter dans anayltics .

**Stockage** : Les objets de notre pipeline vont être stocké dans une bucket S3 de AWS grace au service kinesis firehose qui va mettre les objets dans la bucket .

**Transformation :** On utilise pour cette partie le service kinesis data analytics qui va utiliser SQL dans notre cas pour faire des requetes sur le stream et fair les transformations nécessaires avant de les mettre dans l'output stream

**Exposition :** Les données traités peuvent etre exposée à différents systèmes externes ou re stockés dans S3 via kinesis firehose.



**Partie 3 :**