https://github.com/EthanVieville/spark\_grp15.git

Exo 4 :

Comparaison du temps d’exécution.

Voir script :   
spark-handson/src/fr/hymaia/exo4/compare\_job.py

Première exécution des trois, problème de temps entre les différentes méthodes. Et exécution bien trop courte pour les transformations concernées.

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Notre python ne passait pas par un UDF, nous avons ajouté l’UDF et modifié le show () par un write comme action à la suite de recommandations.  
  
Une image contenant texte, capture d’écran, Police

Description générée automatiquement

**Window Function**   
Voir seconde partie de code :   
spark-handson/src/fr/hymaia/exo4/no\_udf.py  
  
Problème de mémoire souvent rencontrés. Diverses corrections ont été tentés, partition, persist etc… mais sans résultats probants.

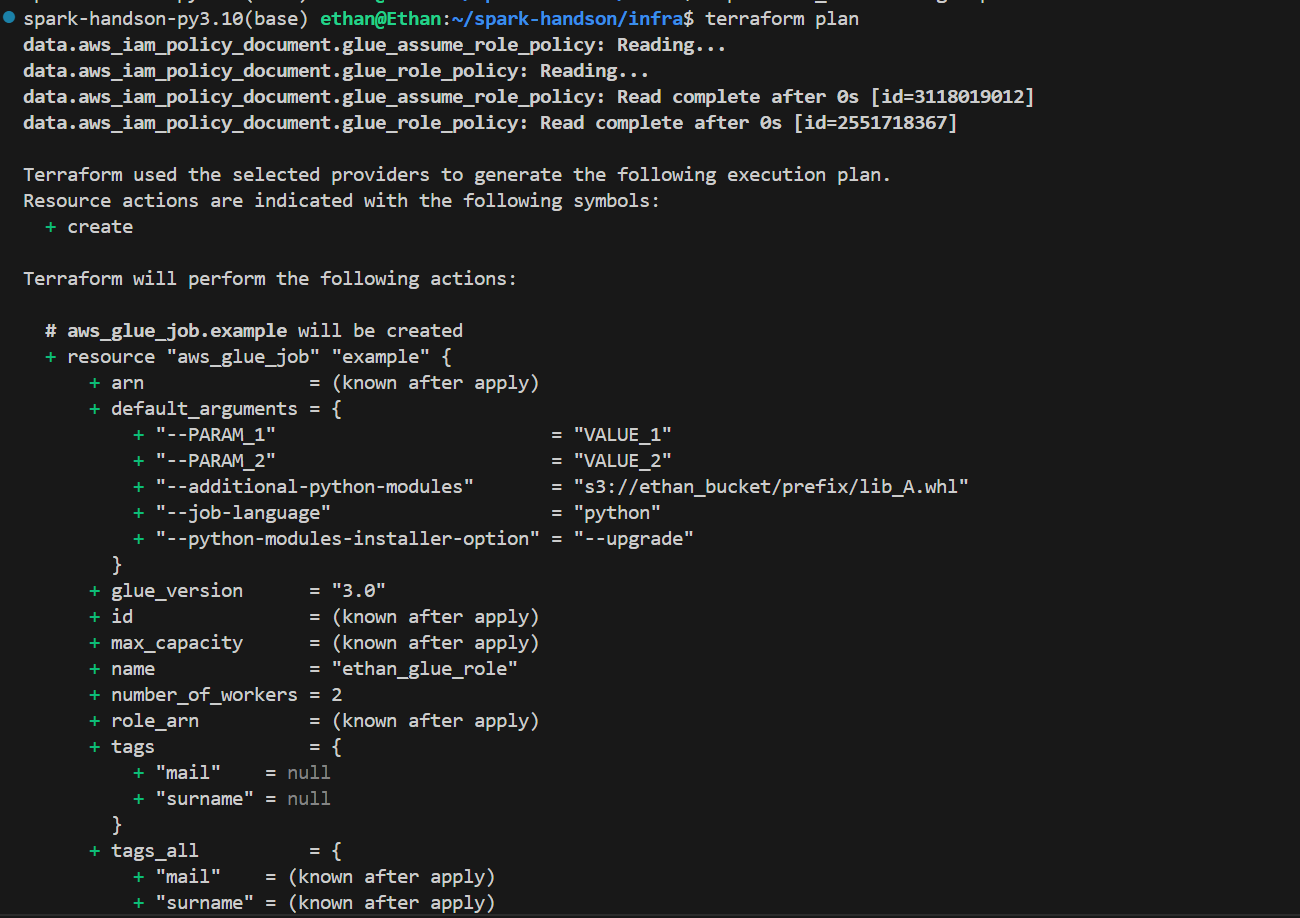
Exo 5 :   
Packaging :

Une image contenant texte, capture d’écran, Police

Description générée automatiquementUne image contenant texte, capture d’écran, Police

Description générée automatiquement  
  
Terraform init

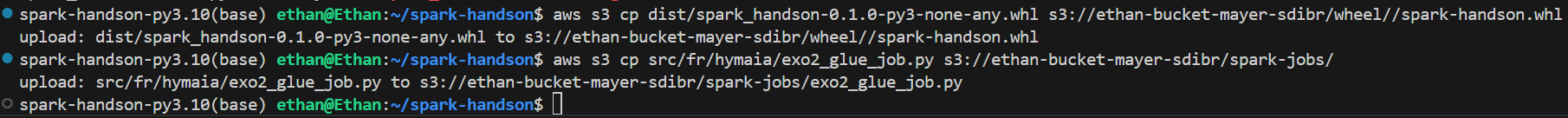
Une image contenant texte, capture d’écran, Police, logiciel

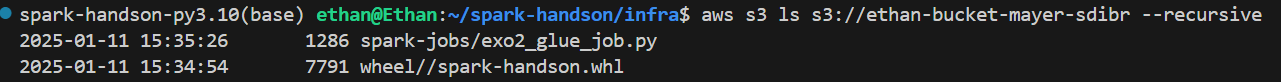
Description générée automatiquementPlan

APPLY

Une image contenant texte, capture d’écran, Police, conception

Description générée automatiquement  
  
UPLOAD S3:





Confirmation sur AWS :

(rôle)

Une image contenant texte, Police, ligne, nombre

Description générée automatiquement

Le job apparait bien :  
  
Une image contenant texte, capture d’écran, logiciel, Icône d’ordinateur

Description générée automatiquement