# Workshop : Intégration d'Amazon Athena avec AWS Glue via CLI

## Objectif

Dans ce workshop, vous apprendrez à intégrer Amazon Athena avec AWS Glue en utilisant uniquement l'interface en ligne de commande (CLI) d'AWS. Vous hébergerez des fichiers CSV et JSON sur Amazon S3, configurerez des crawlers Glue pour découvrir et cataloguer les données, créerez un job ETL Glue pour transformer ces données, puis interrogez les données transformées via Athena. Enfin, vous verrez comment surveiller l'exécution via CloudWatch.

## Étape 1 : Hébergement des fichiers CSV et JSON sur Amazon S3

1. Créer un bucket S3 ‘sae-airbnb-data-workshop’.
2. Créer des dossiers ‘csv\_data/’, ‘json\_data/’
3. Chargez vos fichiers CSV et JSON dans les dossiers respectifs

## Étape 2 : Configuration des Crawlers AWS Glue pour découvrir les fichiers

1. Créer une base de données ‘workshop\_db’ dans le Glue Catalog
2. Créer un rôle IAM pour Glue
3. Créer des crawlers pour CSV et JSON
4. Exécuter les crawlers

## Étape 3 : Transformation des données via un job ETL AWS Glue

1. Créer un job ETL Glue
2. Copier le job PySpark sur S3
3. Exécuter le job ETL
4. Surveiller le statut du job exécuté

## Étape 4 : Interroger les données transformées avec Amazon Athena

1. Créer un fichier de requête SQL
2. Exécutez la requête dans Athena
3. Récupérez les résultats

## Étape 5 : Surveiller le pipeline avec Amazon CloudWatch

1. Accéder aux logs de Glue via CloudWatch
2. Recherchez les logs dans CloudWatch pour les crawlers et le job

## Conclusion

Dans ce workshop, vous avez appris à intégrer Amazon Athena avec AWS Glue via AWS CLI. Vous avez :

* Hébergé des fichiers CSV et JSON sur S3.
* Configuré des crawlers Glue pour cataloguer ces fichiers.
* Créé un job ETL Glue pour transformer les données.
* Utilisé Athena pour interroger les données transformées.
* Automatisé le pipeline avec AWS Glue Workflow.
* Surveillé l'exécution via CloudWatch.

Ce processus, entièrement géré via la CLI, offre un moyen flexible et puissant de gérer un pipeline ETL sur AWS.