**Atelier 3 : Comprendre le MapReduce pour Hive**

**Etape 1** :**Utiliser la commande describe**

* 1. Exécuter la commande **describe** pour **wh\_visits**

hive> describe wh\_visits;

hive> describe wh\_visits;

OK

lname string

fname string

time\_of\_arrival string

appt\_scheduled\_time string

meeting\_location string

info\_comment string

Time taken: 2.977 seconds, Fetched: 6 row(s)

* 1. Est-ce cette requête a nécessité un MapReduce ?

Non

* 1. Quel est le nom de la ressource Hive qui a permis de récupérer ces Informations sur le schéma?

Le métastore Hive contient les informations du schéma de toutes les tables.

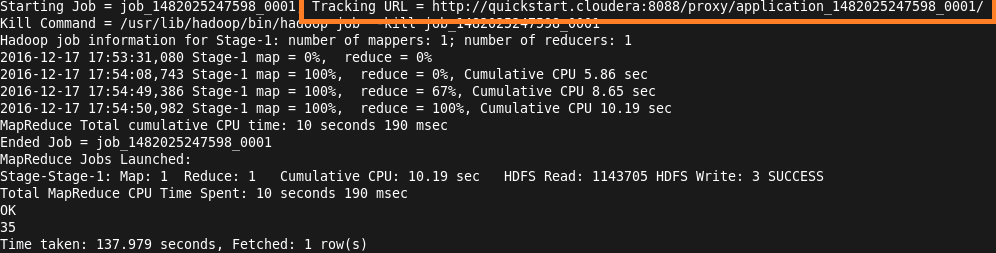
**Etape 2** : **Requête avec count(\*)**

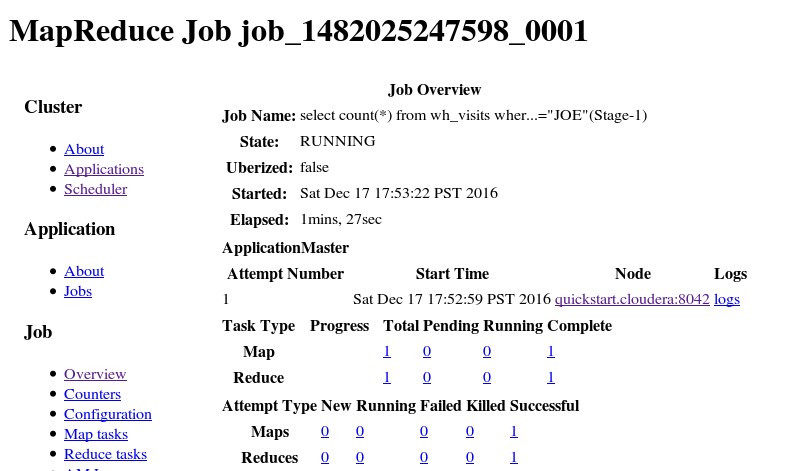
* 1. Lancer la requête suivant :

Select count(\*) from wh\_visits where fname = "JOE";

* 1. Est-ce que Hive exécute un job MapReduce pour générer le résultat?

Oui

* 1. Lorsque la tâche MapReduce est terminée, recherchez la page de détails de log sur le navigateur  
     

**Résultat** :

Ou ouvrez votre navigateur et pointez-le vers l'interface utilisateur JobHistory

**Cloudera** :

<http://quickstart.cloudera:19888/jobhistory/>

<http://quickstart.cloudera:8088/cluster>

**Hortonworks** :

* 1. Combien de tâches Mapper ont été utilisées pour exécuter cette requête?

1 Map

* 1. Combien de tâches de Reducer ont été utilisées pour exécuter cette requête?

1 Réduce

* 1. Combien de temps a t-il fallu pour que cette requête s'exécute

On peut le voir soit sur le terminal : 137,979 secondes ou Elapsed time ( Temps épuisé ) 1m 27secondes sur l’interface de log tracking ci-dessus



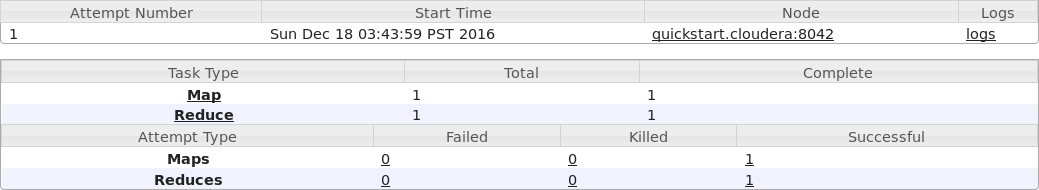
**Etape 3** : **Requête avec sort by**

* 1. Lancer la requête ci-dessous :

hive> select \* from wh\_visits where fname = "JOE" sort by lname;

Lorsque la tâche MapReduce est terminée, recherchez la page de détails de log sur le navigateur





* 1. Combien de tâches Mapper ont été utilisées pour exécuter cette requête?

1 Map

* 1. Combien de tâches de Reducer ont été utilisées pour exécuter cette requête?

1 Réduce

**Etape 4**: **Utiliser la commande EXPLAIN**

* 1. La commande **EXPLAIN** affiche le plan d'exécution d'une requête sans exécuter cette dernière.

Pour démontrer, ajoutez EXPLAIN au Début de la requête suivante que vous avez exécutée plus tôt :

hive>**explain** select \* from wh\_visits where fname = "JOE" sort by lname;

* 1. Que constatez-vous ?

On voit que le plan d’exécution montre bien les deux phases :

La phase de Maper et la Phase de réducer

**Etape 5**: **Utiliser la commande EXPLAIN EXTENDED**

* 1. Exécutez la requête à nouveau le EXPLAIN précédent, sauf que cette fois ajoutez le EXTENDED

hive> explain **extended** select \* from wh\_visits where fname= "JOE" sort by lname;

* 1. Comparer les deux sorties. Que remarquez-vous ?

Vous Remarquez que la commande EXTENDED ajoute informations supplémentaires sur le plan d'exécution.

**Etape 6** :**Requête avec select**

* 1. Lancer la requête suivante :

hive> select \* from wh\_visits limit 5;

* 1. Est-ce que Hive exécute un job MapReduce pour générer les résultats?

Non

* 1. Vérifier que le MapReduce n’a pas été utilisé en utilisant la commande explain

Explain select \* from wh\_visits limit 5;

* 1. Expliquez ?

Hive lit simplement les données directement à partir du fichier visits.txt dans HDFS. Cela ne nécessite pas un MapReduce