

## INSTALLATION DE APACHE SPARK ET DE MONGODB DANS UN OS LINUX

### Prérequis :

- Installer Java version 8

*\$ sudo apt install openjdk-8-jdk*

*/\*\* Cette commande permet d'installer java 8 dans le but d'installer \*/*

- Installer Anaconda 3 ou Anaconda 4

### INSTALLATION DE APACHE SPARK

#### 1<sup>ère</sup> Étape : Installation via un nouveau utilisateur

- ▶ Création d'un utilisateur Spark (Au choix dans le cas contraire ignorer cette étape)

*\$ sudo adduser spark*

*\$ sudo usermod -aG sudo spark*

- ▶ Se connecter au compte d'utilisateur spark

Télécharger la version 3.0.1 de spark compilée pour Hadoop 3.2

*\$ wget https://downloads.apache.org/spark/spark-3.0.1/spark-3.0.1-bin-hadoop3.2.tgz*

- ▶ Extraire les fichier téléchargé

*\$ tar xzf spark-3.0.0-bin-hadoop3.2.tgz*

- ▶ Déplacer le dossier dans le dossier où vous souhaitez installer spark

*\$ mv spark-3.0.1-bin-hadoop3.2 /chemin\_du\_dossier\_d\_installation*

## 2<sup>ème</sup> Étape : Configuration de Spark pour exécution en mode cluster

Pour cela nous allons passer à la configuration (machine master – machines slaves (les workers))

### ► Générer une clé ssh et la copier sur le master

```
$ ssh-keygen -t rsa -b 4096
```

```
$ ssh-copy-id user@ip-master
```

### ► Exécuter spark

En local : **\$ spark-submit --master local[\*] wordcount.py test100.txt**

### ► En mode cluster

Lancer le master

```
$ start-master.sh --host 192.168.43.2 --port 7077
```

Arrêter le master

```
$ stop-master.sh --host 192.168.43.2 --port 7077
```

Lancer les workers

- ✓ Exécuter sur chaque worker

```
$ start-slave.sh spark://192.168.43.2:7077
```

- ✓ Arrêter les workers

```
$ stop-slave.sh spark://192.168.43.2:7077
```

- ✓ Lancer le traitement

```
$ spark-submit --master spark://192.168.43.2:7077 wordcount.py test100.txt
```

- ✓ Visualisation sur navigateur

**adresse : <adresse\_ip\_du\_master>:8080**

## INSTALLATION DE MONGODB

- ▶ Importez la clé publique utilisée par le système de gestion des packages

***\$ wget -qO - https://www.mongodb.org/static/pgp/server-4.4.asc | sudo apt-key add -***

- ▶ Rechargez la base de données des packages locaux

***\$ sudo apt update***

- ▶ Installez les packages MongoDB

***\$ sudo apt install -y mongodb-org***

Vérifier le statuts de MongoDB une fois installer via la commande suivante :

***\$ sudo service mongod status***

Lorsque le serveur MongoDB n'est pas activer redemarrer le via la commande

***\$ sudo service mongod restart***