*Titre – Installer un environnement de BIG DATA SUR AWS*

Table des matières

[Étape 1/6 – Créer un compte sur koho.ca et vous faire un transfert Interac de 10$. 1](#_Toc240341486)

[Étape 2/6 – Créer un compte sur AWS management console en utilisant votre carte de crédit KOHO 2](#_Toc849263974)

[Étape 3/6 – Créer une instance UBUNTU sur AWS 3](#_Toc1041833889)

[Étape 4/6 – Installer DOCKER sur votre machine virtuelle. 7](#_Toc2068203305)

[Étape 5/6 – Exécutez les scripts d’installation de HADOOP 10](#_Toc674501340)

[Étape 6/6 – Testez 11](#_Toc435172181)

[Étape 7 – Lancer les applications Hadoop (Ajoutée le 29-06-2023) 14](#_Toc1556291104)

[Étape 8 – Contrôler et vérifier les coûts d’utilisation 15](#_Toc291028620)

# Étape 1/6 – Créer un compte sur koho.ca et vous faire un transfert Interac de 10$.

1. Créez un compte de carte de crédit prépayée. Pour la confirmation, vérifiez vos spams.

<https://www.koho.ca/>

A picture containing text, screenshot, font, logo

Description automatically generated

1. Envoyez-vous 10$ en mode Interac à l’adresse finissant par @kohotransfers.ca.

Il suffit de cliquer sur le bouton ajouter de l’argent pour voir cette adresse. Pour voir les informations de votre carte, cliquez à droite sur l’œil en dessus du cadenas.

A screenshot of a credit card

Description automatically generated with low confidence

# Étape 2/6 – Créer un compte sur AWS management console en utilisant votre carte de crédit KOHO

A picture containing text, screenshot, cartoon, design

Description automatically generated

# Étape 3/6 – Créer une instance UBUNTU sur AWS

1. Connectez vous à AWS management console.
2. Cherchez EC2

A screenshot of a computer

Description automatically generated with medium confidence

1. Choisir la région

A screenshot of a computer

Description automatically generated with medium confidence

1. Choisir Launch instance

A screenshot of a computer

Description automatically generated with medium confidence

1. Configurer votre instance comme illustré dans la table ci-bas :

|  |  |
| --- | --- |
| *Description* | *Imprime écran* |
| 5-1 - Nom : MyBigDataMachine |  |
| 5-2 - Application and OS Images : Ubuntu |  |
| 5-3- Choisir Ubuntu Server 22.04 LTS (HVM) , SSD Volume type |  |
| 5-4- Architecture : 64 bits |  |
| 5-5- Instance type : t2.medium |  |
| 5-6- Créez une paire de clé |  |
| 5-7- Modifier les network settings | |  |  | | --- | --- | |  |  | |  |  | |  | | |
| 5-8- Pour le stockage choisir 30 Go |  |
| 5-9- Lancer l’instance |  |
| 5-10 – attendre que le paramètre instance state change à Running |  |
| 5-11- Cliquez sur connect |  |

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

# Étape 4/6 – Installer DOCKER sur votre machine virtuelle.

Exécutez les 13 commandes suivantes une par une :

Commande 1 –

*clear*

A screenshot of a computer

Description automatically generated with medium confidence

Commande 2 –

*sudo -s*

A picture containing text, screenshot, font, multimedia software

Description automatically generated

Commande 3 –

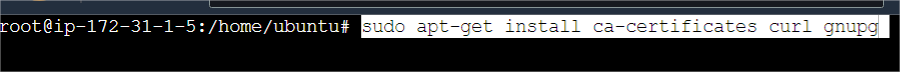
*apt-get update*

A screenshot of a computer

Description automatically generated with low confidence

Commande 4 –

*sudo apt-get install ca-certificates curl gnupg*



A black background with white text

Description automatically generated with low confidence

Commande 5 –

*sudo install -m 0755 -d /etc/apt/keyrings*



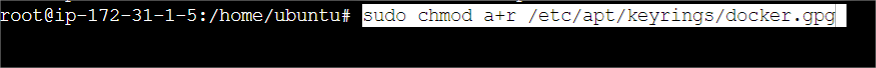
Commande 6 –

*curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o /etc/apt/keyrings/docker.gpg*



Commande 7 –

*sudo chmod a+r /etc/apt/keyrings/docker.gpg*



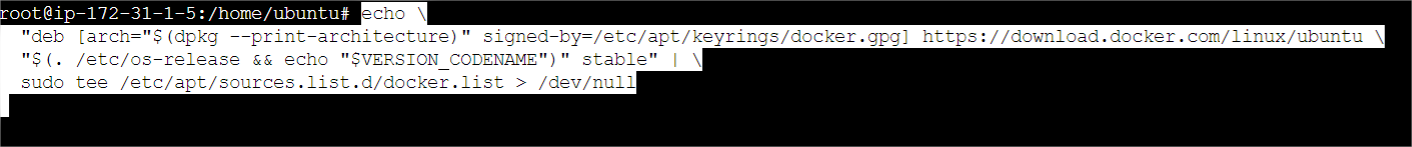
Commande 8 – (copier la commade en totalité et coller la)

*echo \*

*"deb [arch="$(dpkg --print-architecture)" signed-by=/etc/apt/keyrings/docker.gpg] https://download.docker.com/linux/ubuntu \*

*"$(. /etc/os-release && echo "$VERSION\_CODENAME")" stable" | \*

*sudo tee /etc/apt/sources.list.d/docker.list > /dev/null*



Si la commande ci-dessus n’a pas fonctionnée faire Ctrl C et utiliser la commande suivante:

*echo "deb [arch=$(dpkg --print-architecture) signed-by=/etc/apt/keyrings/docker.gpg] https://download.docker.com/linux/ubuntu focal stable" | sudo tee /etc/apt/sources.list.d/docker.list > /dev/null*

Commande 9 –

*sudo apt-get install docker-ce docker-ce-cli containerd.io docker-buildx-plugin docker-compose-plugin*



Commande 10 –

*sudo apt-get update*



Commande 11 –

***docker version***



Commande 12 –

*sudo apt-get install docker-ce docker-ce-cli containerd.io docker-buildx-plugin docker-compose-plugin*



A picture containing text, font, screenshot, line

Description automatically generated

Commande 13 –

***docker version***



Résultat final

A picture containing text, screenshot, software

Description automatically generated

# Étape 5/6 – Exécutez les scripts d’installation de HADOOP

Exécutez les 9 commandes suivantes une par une :

***Commande 1 :*** *docker --version*

***Commande 2 :*** *docker-compose --version (sinon apt install docker-compose)*

***Commande 3 :*** *git --version*

***Commande 4 :*** *mkdir hadoop\_docker*

***Commande 5 :*** *cd hadoop\_docker/*

***Commande 6 :*** *git clone https://github.com/hrhouma/bd-infra.git*

***Commande 7 :*** *cd bd-infra/*

***Commande 8 :*** docker-compose up -d

***Commande 9 :*** docker-compose up -ls

# Étape 6/6 – Testez

Essayez sur votre navigateur

*http://<ip of remote server>:50070*

*http://<ip of remote server>:8888*

*http://<ip of remote server>:8080*

Exemple , dans mon cas, l’adresse IP est ***à***

A screenshot of a computer

Description automatically generated with medium confidence

3.138.220.110*:50070*

3.138.220.110*:8888*

3.138.220.110*:8080*

3.138.220.110*:50070*

A screenshot of a computer

Description automatically generated with medium confidence

3.138.220.110*:8888*

A screenshot of a computer

Description automatically generated with medium confidence

3.138.220.110*:8080*

A screenshot of a computer

Description automatically generated

N’oubliez pas d’arrêter votre instance (et non pas de la résilier , chaque fois, à la fin de vos manipulations !). La prochaine fois, l’adresse IP changera . Il faut toujours vérifier votre adresse IP quand vous vous connectez !

A picture containing text, font, screenshot, line

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

# Étape 7 – Lancer les applications Hadoop (Ajoutée le 29-06-2023)

À chaque nouveau démarrage de l’instance MyBigDataMachine, se connecter et exécutez les commandes suivantes une à une  à :

***Commande 1 : sudo -s***

***Commande 2 :*** *cd hadoop\_docker/bd-infra*

***Commande 3 :*** docker-compose up –d

Lancer sur votre navigateur

**Hadoop**

*http://<ip of remote server>:50070*

**Hue**

*http://<ip of remote server>:8888*

**Spark**

*http://<ip of remote server>:8080*

# Étape 8 – Contrôler et vérifier les coûts d’utilisation

<https://docs.aws.amazon.com/cost-management/latest/userguide/ce-getting-started.html>

Start Cost Explorer by opening the AWS Cost Management console.

###### To open Cost Explorer

* Sign in to the AWS Management Console and open the AWS Cost Management console at <https://console.aws.amazon.com/cost-management/home>.

This opens the Cost dashboard that shows you the following:

* Your estimated costs for the month to date
* Your forecasted costs for the month
* A graph of your daily costs
* Your five top cost trends
* A list of reports that you recently viewed