

Titre – Installer un environnement de BIG DATA SUR AWS

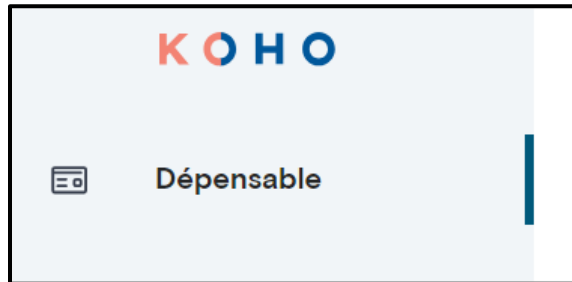
Table des matières

| | |
|--|----|
| Étape 1/6 – Créer un compte sur koho.ca et vous faire un transfert Interac de 10\$. | 2 |
| Étape 2/6 – Créer un compte sur AWS management console en utilisant votre carte de crédit KOHO | 3 |
| Étape 3/6 – Créer une instance UBUNTU sur AWS..... | 4 |
| Étape 4/6 – Installer DOCKER sur votre machine virtuelle. | 8 |
| Étape 5/6 – Exécutez les scripts d’installation de HADOOP..... | 11 |
| Étape 6/6 – Testez..... | 12 |

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/>



- 2) 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.





Explore Free Tier products with a new AWS account.

To learn more, visit aws.amazon.com/free.



Sign up for AWS

Root user email address
Used for account recovery and some administrative functions

AWS account name
Choose a name for your account. You can change this name in your account settings after you sign up.

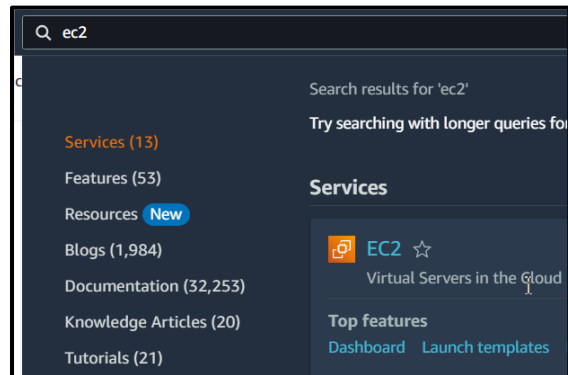
Verify email address

OR

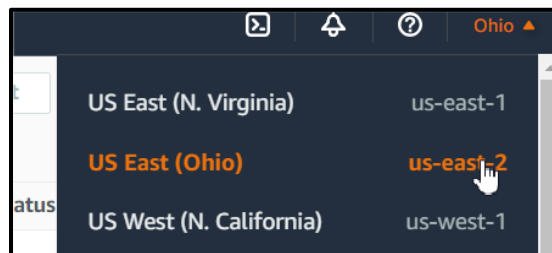
Sign in to an existing AWS account

Étape 3/6 – Créer une instance UBUNTU sur AWS

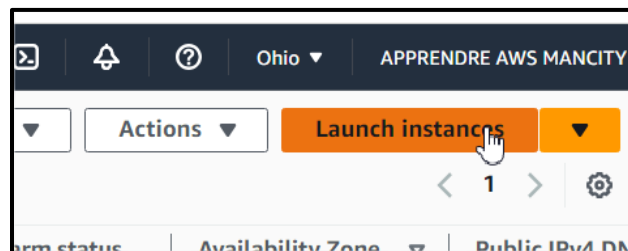
- 1- Connectez vous à AWS management console.
- 2- Cherchez EC2



- 3- Choisir la région



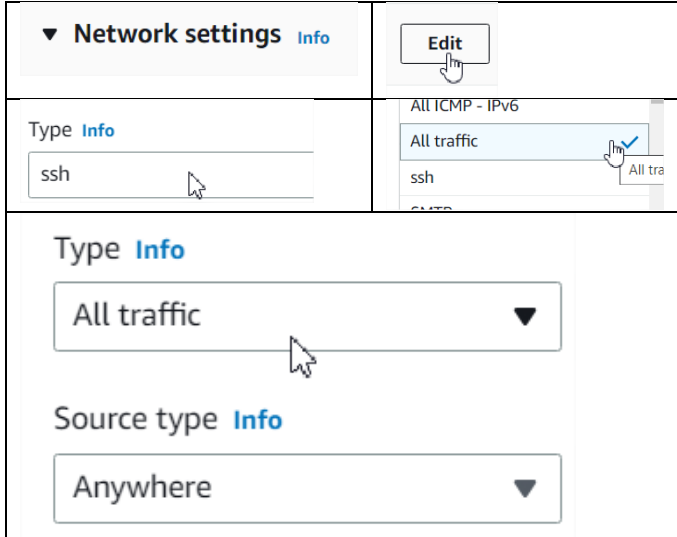

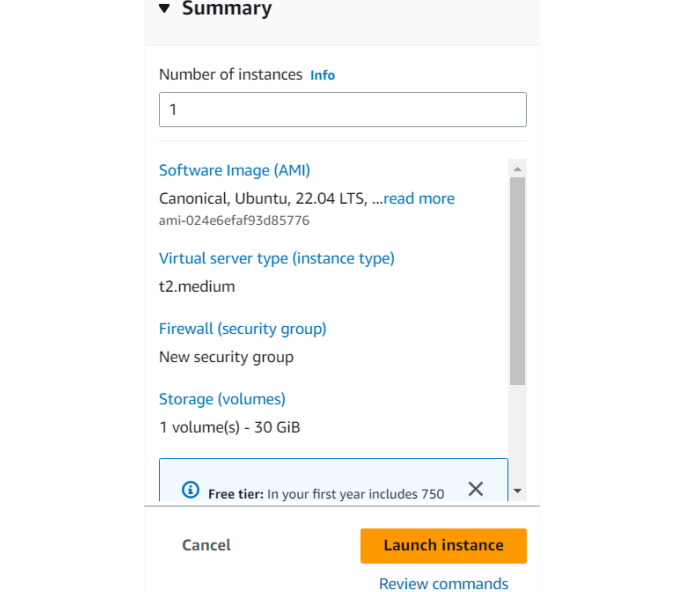
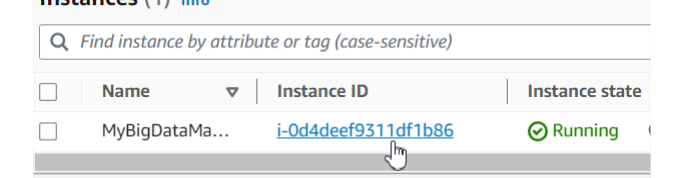
- 4- Choisir Launch instance



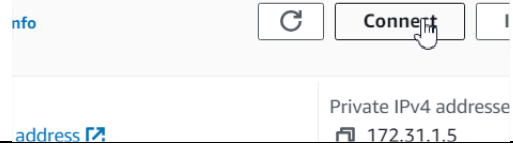
- 5- Configurer votre instance comme illustré dans la table ci-bas :

| Description | Imprime écran |
|------------------------------|--|
| 5-1 - Nom : MyBigDataMachine | A screenshot of the AWS Management Console 'Name and tags' section. It shows a form with a 'Name' field containing the text 'MyBigDataMachine'. There is an 'Info' link next to the section title. |

| | |
|--|--|
| 5-2 - Application and OS Images : Ubuntu | <div> <div>▼ Application and OS Images (Amazon M</div> <div>An AMI is a template that contains the software configurati launch your instance. Search or Browse for AMIs if you don't</div> <div> <div>Q Search our full catalog including 1000s of appl</div> <div>Quick Start</div> <div> <div>Amazon Linux</div> <div>macOS</div> <div>Ubuntu</div> <div>W</div> </div> <div> <div>aws</div> <div>Mac</div> <div>ubuntu</div> <div></div> </div> <div>Amazon Machine Image (AMI)</div> </div> </div> |
| 5-3- Choisir Ubuntu Server 22.04 LTS (HVM) , SSD Volume type | <div> <div>Amazon Machine Image (AMI)</div> <div> <div>Ubuntu Server 22.04 LTS (HVM), SSD Volume Type</div> <div>ami-024e6efaf93d85776 (64-bit (x86)) / ami-08fdd91d87f63bb09 (64-bit (Arm))</div> <div>Virtualization: hvm ENA enabled: true Root device type: ebs</div> <div>Fr</div> </div> <div>Ubuntu Server 22.04 LTS (HVM)</div> </div> |
| 5-4- Architecture : 64 bits | <div> <div>Architecture</div> <div>64-bit (x86) ▼</div> </div> |
| 5-5- Instance type : t2.medium | <div> <div>▼ Instance type Info</div> <div> <div>Instance type</div> <div>t2.medium</div> <div>Family: t2 2 vCPU 4 GiB Memory Current gene</div> <div>On-Demand Windows pricing: 0.0644 USD per Hour</div> <div>On-Demand SUSE pricing: 0.1464 USD per Hour</div> <div>On-Demand RHEL pricing: 0.1064 USD per Hour</div> <div>On-Demand Linux pricing: 0.0464 USD per Hour</div> </div> </div> |
| 5-6- Créez une paire de clé | <div> <div> <div>↻ Create new key pair</div> </div> <div> <div>Create key pair</div> <div> <div>Key pair name</div> <div>Key pairs allow you to connect to your instance securely.</div> <div>bigdata</div> <div>The name can include upto 255 ASCII characters. It can't include leading or trailing spaces.</div> </div> <div> <div>Key pair type</div> <div> <div><input checked="" type="radio"/> RSA</div> <div>RSA encrypted private and public key pair</div> <div><input type="radio"/> ED25519</div> <div>ED25519 encrypted private and public key pair</div> </div> </div> <div> <div>Private key file format</div> <div> <div><input checked="" type="radio"/> .pem</div> <div>For use with OpenSSH</div> <div><input type="radio"/> .ppk</div> <div>For use with PuTTY</div> </div> </div> <div> <div>⚠ When prompted, store the private key in a secure and accessible location on your computer. You will need it later to connect to your instance. Learn more</div> </div> <div> <div>Cancel</div> <div>Create key pair</div> </div> </div> </div> |

| | |
|--|--|
| 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



EC2 Instance Connect | Session Manager | SSH client | EC2 serial console

Instance ID
i-Od4deef9311df1b86 (MyBigDataMachine)

Connection Type

☒ Connect using EC2 Instance Connect
Connect using the EC2 Instance Connect browser-based client, with a public IPv4 address.

☐ Connect using EC2 Instance Connect Endpoint
Connect using the EC2 Instance Connect browser-based client, with a private IPv4 address and a VPC endpoint.

Public IP address
3.137.220.110

User name
Enter the user name defined in the AMI used to launch the instance. If you didn't define a custom user name, use the default user name, ubuntu.
ubuntu

Note: In most cases, the default user name, ubuntu, is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.

Cancel Connect

```
aws | Services | Search [Alt+S]
Swap usage: 0%
Expanded Security Maintenance for Applications is not enabled.
0 updates can be applied immediately.
Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

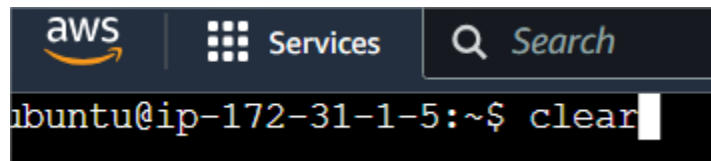
ubuntu@ip-172-31-1-5:~$
```

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

Exécutez les 13 commandes suivantes une par une :

Commande 1 –

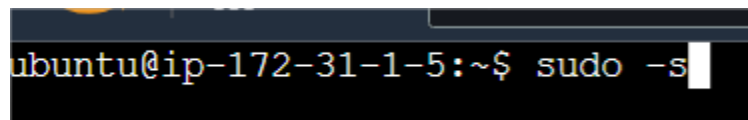
Clear



```
aws Services Search  
ubuntu@ip-172-31-1-5:~$ clear
```

Commande 2 –

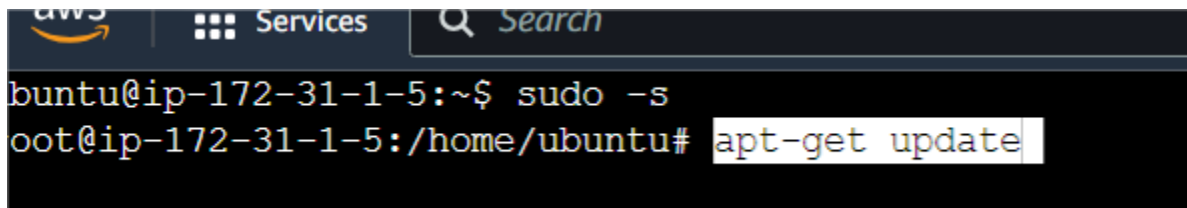
sudo -s



```
aws Services Search  
ubuntu@ip-172-31-1-5:~$ sudo -s
```

Commande 3 –

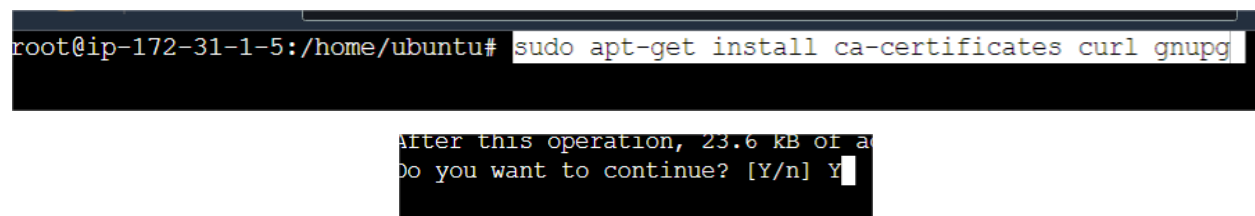
apt-get update



```
aws Services Search  
ubuntu@ip-172-31-1-5:~$ sudo -s  
root@ip-172-31-1-5:/home/ubuntu# apt-get update
```

Commande 4 –

sudo apt-get install ca-certificates curl gnupg



```
aws Services Search  
root@ip-172-31-1-5:/home/ubuntu# sudo apt-get install ca-certificates curl gnupg  
  
After this operation, 23.6 kB of additional disk space will be used.  
Do you want to continue? [Y/n] Y
```

Commande 5 –

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


```
root@ip-172-31-1-5:/home/ubuntu# sudo install -m 0755 -d /etc/apt/keyrings
root@ip-172-31-1-5:/home/ubuntu#
```

Commande 6 –

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

```
root@ip-172-31-1-5:/home/ubuntu# curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o /etc/apt/keyrings/docker.gpg
root@ip-172-31-1-5:/home/ubuntu#
```

Commande 7 –

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

```
root@ip-172-31-1-5:/home/ubuntu# sudo chmod a+r /etc/apt/keyrings/docker.gpg
```

Commande 8 – (copier la commande 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
```

```
root@ip-172-31-1-5:/home/ubuntu# 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
```

Commande 9 –

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

```
root@ip-172-31-1-5:/home/ubuntu# sudo apt-get install docker-ce docker-ce-cli containerd.io docker-buildx-plugin docker-compose-plugin
```

Commande 10 –

```
sudo apt-get update
```

```
oot@ip-172-31-1-5:/home/ubuntu# sudo apt-get update
```

Commande 11 –

```
docker version
```

```
root@ip-172-31-1-5:/home/ubuntu# docker version
```

Commande 12 –

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

```
clonal versions.  
# sudo apt-get install docker-ce docker-ce-cli containerd.io docker-buildx-plugin docker-compose-plugin  
After this operation, 402 MB of additional space will be used.  
Do you want to continue? [Y/n] Y
```

Commande 13 –

```
docker version
```

```
root@ip-172-31-1-5:/home/ubuntu# docker version
```

RÉSULTAT FINAL

```
root@ip-172-31-1-5:/home/ubuntu# docker version  
Client: Docker Engine - Community  
Version: 24.0.2  
API version: 1.43  
Go version: go1.20.4  
Git commit: cb74dfc  
Built: Thu May 25 21:51:00 2023  
OS/Arch: linux/amd64  
Context: default  
  
Server: Docker Engine - Community  
Engine:  
Version: 24.0.2  
API version: 1.43 (minimum version 1.12)  
Go version: go1.20.4  
Git commit: 659604f  
Built: Thu May 25 21:51:00 2023  
OS/Arch: linux/amd64  
Experimental: false  
containerd:  
Version: 1.6.21  
GitCommit: 3dce8eb055cbb6872793272b4f20ed16117344f8  
runc:  
Version: 1.1.7  
GitCommit: v1.1.7-0-g860f061
```

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`

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

EC2 Instance Connect | Session Manager | SSH client | EC2 serial console

Instance ID
i-0d4deef9311df1b86 (MyBigDataMachine)

Connection Type

☒ Connect using EC2 Instance Connect
Connect using the EC2 Instance Connect browser-based client, with a public IPv4 address.

☐ Connect using EC2 Instance Connect Endpoint
Connect using the EC2 Instance Connect browser-based client, with a private IPv4 address and a VPC endpoint.

Public IP address
3.137.220.110

User name
Enter the user name defined in the AMI used to launch the instance. If you didn't define a custom user name, use the default user name, ubuntu.
ubuntu

Note: In most cases, the default user name, ubuntu, is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.


Cancel Connect

3.138.220.110:*50070*

3.138.220.110:*8888*

3.138.220.110:*8080*

3.138.220.110:50070

 3.137.220.110:50070/

Hadoop

Overview

Datanodes

Datanode Volume Failures


Snapshot


Startup Progress

Overview 'namenode:8020' (active)

| | |
|----------------|--|
| Started: | Fri Jun 23 06:02:26 UTC 2023 |
| Version: | 2.7.4, rcd915e1e8d9d0131462a0b7301586c175728a282 |
| Compiled: | 2017-08-01T00:29Z by kshvachk from branch-2.7.4 |
| Cluster ID: | CID-7109fdf3-c917-45a5-8681-7b1d1624d00a |
| Block Pool ID: | BP-1652321417-172.27.1.5-1687500135975 |

3.138.220.110:8888

 3.137.220.110:8888/



Query. Explore. Repeat.

Since this is your first time logging in, pick any username and password. Be sure to remember these, as they will become your Hue superuser credentials.

Create Account

3.138.220.110:8080

← → ↻ ⚠ Not secure 3.137.220.110:8080

Spark Master at spark://ee7795cea865:7077

URL: spark://ee7795cea865:7077
Alive Workers: 1
Cores in use: 2 Total, 0 Used
Memory in use: 2.8 GB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

▼ Workers (1)

| Worker Id | Address | State |
|---|-------------------|-------|
| worker-20230623060220-172.27.1.11-34403 | 172.27.1.11:34403 | ALIVE |

▼ Running Applications (0)

| Application ID | Name | Cores | Memory per Executor | Submitted Time |
|----------------|------|-------|---------------------|----------------|
|----------------|------|-------|---------------------|----------------|

▼ Completed Applications (0)

| Application ID | Name | Cores | Memory per Executor | Submitted Time |
|----------------|------|-------|---------------------|----------------|
|----------------|------|-------|---------------------|----------------|

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 !

EC2 > Instances > i-0d4deef9311df1b86 > Connect to instance

Connect to instance

Instances (1 / 1) Info

Find instance by attribute or tag (case-sensitive)

| <input checked="" type="checkbox"/> | Name | Instance ID |
|-------------------------------------|----------------|---------------------|
| <input checked="" type="checkbox"/> | MyBigDataMa... | i-0d4deef9311df1b86 |

Refresh Connect Instance state ▲

Stop instance

Start instance

Reboot instance

Instance type ▼ Status