

1. Connexion à la machine virtuelle

1.1 Ligne de commande :

Initialisé la connection en entrant la commande.

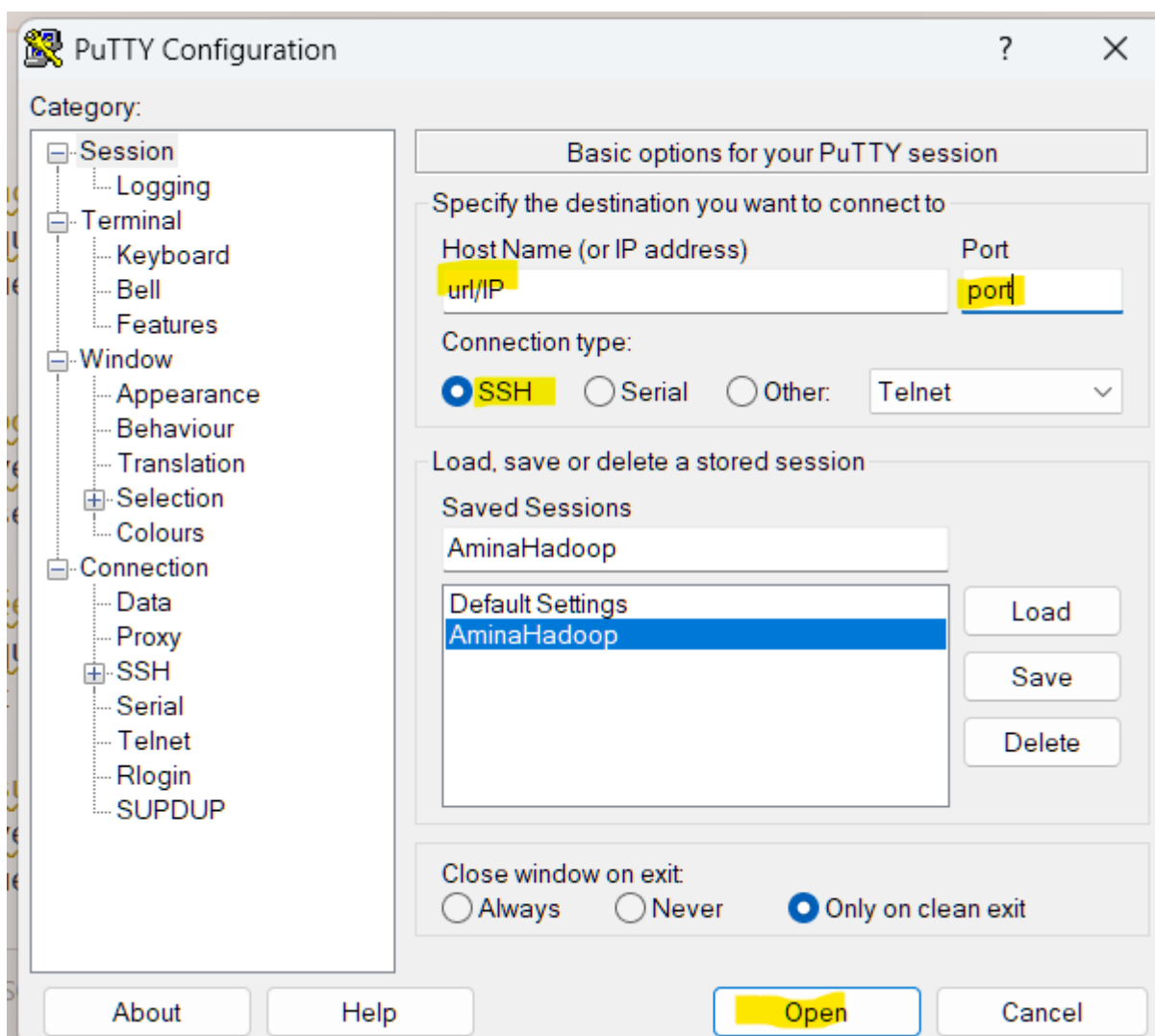
```
ssh [user_name]@[url/IP] -p [port]
```

Puis entrer le mots de pass de l'utilisateur.

```
> ssh [user_name]@[url/IP] -p [port]
[password]'s password: [password]
```

1.2 PuTTY :

Entrer l'url/ip et le port public,
puis appuyez sur "Open".



Ensuite précisez le nom utilisateur et le mots de passe du compte auquel vous vous connecté.



2. Lancer les conteneurs et services Hadoop

- Expliquez les services nécessaires pour Hadoop (HDFS, YARN, Zookeeper, HBase)
- Indiquez les commandes nécessaires pour démarrer ces services sur la machine virtuelle.

3. Importer les données dans HDFS

- Décrivez la manière dont vous allez importer les données pour qu'elles puissent être traitées avec MapReduce.
- Précisez où stocker les fichiers

4. Créer et exécuter un job MapReduce

- Expliquez la structure d'un job MapReduce (mapper, reducer)
- Donnez les commandes à exécuter pour soumettre le job sur Hadoop

5. Visualiser les résultats

- Décrivez comment récupérer les résultats que se soit MapReduce ou Hbase.
- Indiquez comment trouver les données générées comme avec Matplotlib, Pandas...

6. Récupérer les résultats

- Expliquez comment récupérer les fichiers de sortie du job MapReduce depuis HDFS pour le mettre sur la partie linux de votre container hadoop-master.
- Donnez la procédure pour récupérer les données du container jusque son pc local.