

Installer Rstudio-server sur votre cluster et vous y connecter avec PuTTY

Installer FoxyProxy

► Pour google chrome : [lien](#)

► Pour firefox: [lien](#)

Une fois FoxyProxy installé, ouvrez le plugin et importer le fichier se trouvant dans :

- <https://github.com/katossky/panorama-bigdata/blob/master/settings/foxyproxy-settings.json> pour firefox
- <https://github.com/katossky/panorama-bigdata/blob/master/settings/foxyproxy-settings.xml> pour chrome

Etablir une connexion SSH avec votre cluster

► La marche à suivre est également disponible si vous cliquez sur "activez la connexion web" depuis la page de votre cluster

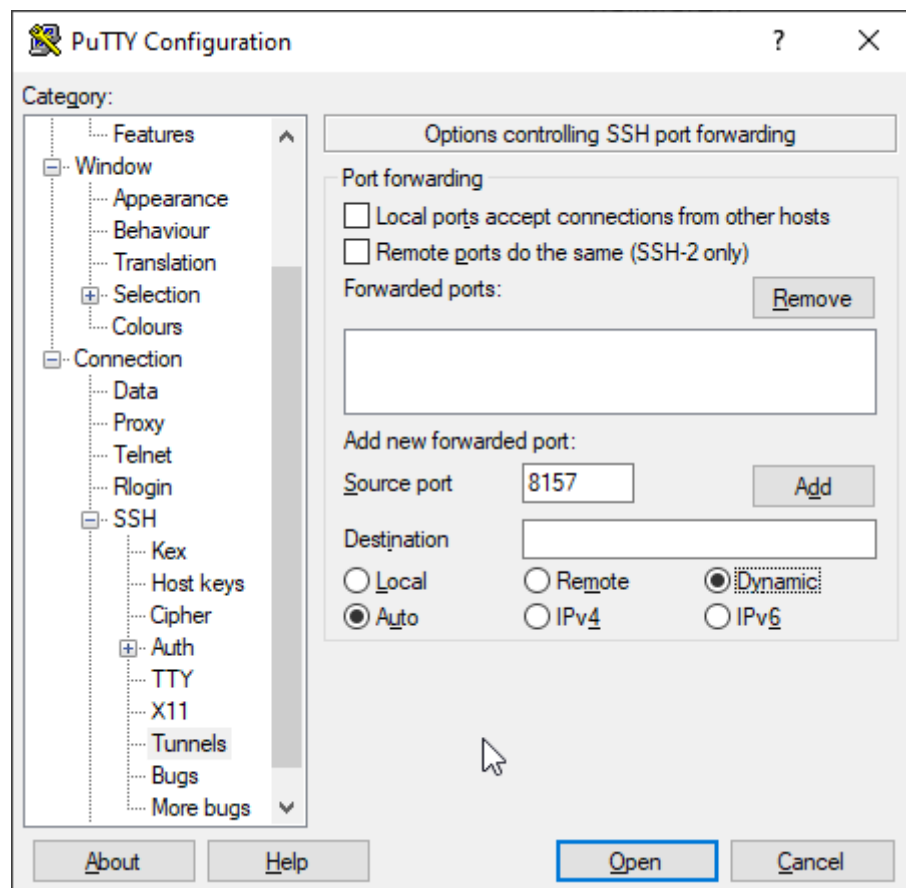
► Vous pouvez vous connecter même si votre cluster est en "Démarrage en cours"

¶ La session SSH que vous allez ouvrir devra rester ouvertes tant que vous utiliser Rstudio-server

- ☐ Lancez PuTTY
- ☐ Dans la liste Category, cliquez sur Session
- ☐ Dans le champ Host Name, tapez **hadoop@[DNS public]** avec **[DNS public]** le DNS public principal de votre cluster (vous le trouverez dans les informations de votre cluster sur l'interface aws). Vous pouvez noter le **[DNS public]** dans un bloc note pour le retrouver plus facilement.

The screenshot shows the Amazon EMR console interface. On the left is a navigation menu with options like 'Clusters', 'Configurations de la sécurité', 'Bloquer l'accès public', 'Sous-réseaux VPC', 'Événements', 'Aide', and 'Nouveautés'. The main area displays details for a cluster named 'Mon cluster' which is in 'En attente' (Pending) state. A red box highlights the 'DNS public principal' field, which contains the value 'ec2-35-180-228-218.eu-west-3.compute.amazonaws.com' with an 'SSH' icon next to it. Below this, there are tabs for 'Résumé', 'Historique de l'application', 'Surveillance', 'Matériel', 'Configurations', 'Événements', 'Étapes', and 'Actions d'amorçage'. The 'Résumé' tab is active, showing a table with columns for 'Résumé', 'Détails de configuration', and 'Réseau et matériel'. The 'Détails de configuration' column lists various settings like 'Étiquette de version', 'Distribution Hadoop', 'Applications', 'URI de connexion', and 'Vue cohérente EMRFS'. The 'Réseau et matériel' column shows 'Zone de disponibilité', 'ID de sous-réseau', 'Maître', 'Principal', and 'Tâche'.

- ☐ Dans la liste Category, développez Connection > SSH > Auth
- ☐ Pour le fichier de clés privées utilisé pour l'authentification, cliquez sur Browse et sélectionnez le fichier de clés privées utilisé pour lancer le cluster.
- ☐ Dans la liste Category, développez Connection > SSH, puis cliquez sur Tunnels.
- ☐ Dans le champ Source port, tapez **8157** (¶ ne changez pas cette valeur)
- ☐ Sélectionnez les options Dynamic et Auto.



- ☐ Laissez le champ Destination vide, puis cliquez sur Add.
- ☐ Cliquez sur Open.
- ☐ Cliquez sur Yes pour ignorer l'alerte de sécurité.

Installer Rstudio server

```
sudo yum install libcurl-devel openssl-devel # used for devtools
```

```
hadoop@ip-172-31-9-22:~
Installing:
libcurl-devel      x86_64      7.61.1-12.93.amzn1    amzn-updates      855 k

Transaction Summary
=====
Install 1 Package

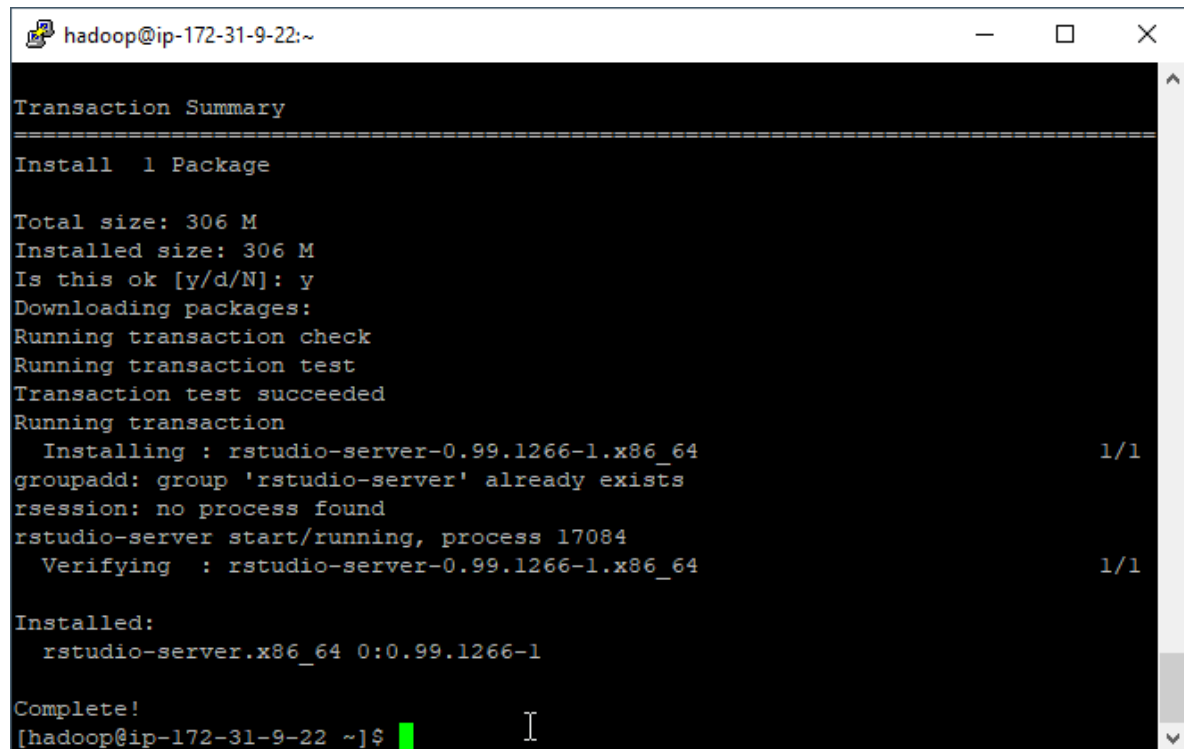
Total download size: 855 k
Installed size: 1.3 M
Is this ok [y/d/N]: y
Downloading packages:
libcurl-devel-7.61.1-12.93.amzn1.x86_64.rpm | 855 kB 00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : libcurl-devel-7.61.1-12.93.amzn1.x86_64      1/1
  Verifying  : libcurl-devel-7.61.1-12.93.amzn1.x86_64      1/1

Installed:
libcurl-devel.x86_64 0:7.61.1-12.93.amzn1

Complete!
[hadoop@ip-172-31-9-22 ~]$
```

```
wget https://download2.rstudio.org/server/centos6/x86_64/rstudio-server-rhel-1.2.5033-x86_64.rpm
sudo yum install rstudio-server-rhel-1.2.5033-x86_64.rpm
```

⚠️ Faites bien attention, il est possible qu'en copiant/collant les lignes, un saut à la ligne se mette sur la première instruction, ce qui conduira à une erreur d'installation



```
hadoop@ip-172-31-9-22:~
Transaction Summary
=====
Install 1 Package

Total size: 306 M
Installed size: 306 M
Is this ok [y/d/N]: y
Downloading packages:
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : rstudio-server-0.99.1266-1.x86_64      1/1
groupadd: group 'rstudio-server' already exists
rsession: no process found
rstudio-server start/running, process 17084
  Verifying  : rstudio-server-0.99.1266-1.x86_64      1/1

Installed:
  rstudio-server.x86_64 0:0.99.1266-1

Complete!
[hadoop@ip-172-31-9-22 ~]$
```

☐ Créez un user pour Rstudio

```
# Make User
sudo useradd -m rstudio-user
sudo passwd rstudio-user
```

⚠️ Retenez votre mot de passe

☐ Créez un dossier dans HDFS pour votre user

```
# Create new directory in hdfs
hadoop fs -mkdir /user/rstudio-user
hadoop fs -chmod 777 /user/rstudio-user
```

☐ Installez postgresql-devel

```
sudo yum install postgresql-devel
```

☐ Connectez-vous à l'interface web de Rstudio server avec l'adresse suivante
`https://[DNS public]:8787` avec `[DNS public]` le DNS public de votre cluster. Puis connectez vous avec l'utilisation `rstudio-user` et le mot de passe que vous avez choisi.

Se connecter au cluster spark via Rstudio-server

Voici un code minimal pour vous connecter au cluster spark avec Rstudio-server

```
install.packages("sparklyr")
library(sparklyr)
Sys.setenv(SPARK_HOME="/usr/lib/spark")
sc <- spark_connect(master="yarn-client")
```

The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains the R script code from the top block.
- Console:** Displays the output of the script execution. It shows the successful installation of the `sparklyr` package, followed by an error message: `Error in shell_connection(master = master, spark_home = spark_home, app_name = app_name, : Failed to connect to Spark (SPARK_HOME is not set).` The error occurs because the `SPARK_HOME` environment variable is not set correctly.
- Environment Panel:** Shows the current environment with the `yarn-client` connection selected. It indicates that there are no tables in the connection.

⚠⚠⚠ Pensez à sauvegarder votre script sur votre poste de temps en temps via des copier coller.
Surtout avant d'éteindre votre cluster !
⚠⚠⚠⚠⚠⚠⚠⚠ **PENSEZ À ÉTEINDRE VOTRE CLUSTER À LA FIN**