CESGA Hadoop 3 User Guide

Release 6.1.1-4

Nov 07, 2023

CONTENTS:

CHAPTER

ONE

OVERVIEW

In this guide we will show you how to use the upgraded BD|CESGA platform based on Hadoop 3.

In the new platform the default version of Spark is also upgraded from 1.6 to 2.4.

There are also new versions of the other main components of the platform like Hive, HBase and HUE. Additionally Impala is now available.

As previously there are two main filesytems:

- HOME: The standard filesystem when you log in
- HDFS: The distributed Hadoop filesystem

See the *Migrating Data* section for more details about how to migrate your data from the previous platform.

If you want just a quick introduction to get you ready to start using the platform you can jump to the *Quickstart* section.

CHAPTER

TWO

QUICKSTART

This section will help you to quickly getting started with the platform. For more details have a look at the rest of this guide, and also check the Tutorials that we have prepared and the *Want to know more* section.

Warning: Before connecting we always recommend that you first start the VPN. If not you will not have access to some services.

If for some reason you are not using the VPN, then one alternative could be to launch a remote desktop from the visualization platform and then connect from there.

By far, the most common way to connect is by establishing an SSH session:

```
ssh username@hadoop3.cesga.es
```

Once connected, you will notice that there are two main filesytems:

- **HOME**: The standard filesystem when you log in
- **HDFS**: The distributed Hadoop filesystem

To migrate your HDFS data from the old platform to the new one, you can use a command similar to the following:

```
hadoop distcp -i -pat -update hdfs://10.121.13.19:8020/user/uscfajlc/wcresult hdfs://
↪nameservice1/user/uscfajlc/wcresult
```

Note: It is recommended to launch the distcp command inside a screen session so it will continue later.

See the *Migrating Data* section for more details about how to migrate your data from the previous platform.

You can then start using the tools you are interested in like *Spark* or *Hive*.

Note: The default version of Spark is 2.4.0. If you plan to use code coming from Spark 1.6 take that into account.

There is also a nice web user interface that you can use to get started with the platform. You can find more information in the *BD|CESGA WebUI* and *HUE: A nice graphical interface to Hadoop* sections.

CHAPTER

THREE

WHAT'S NEW

In comparison with the previous BD|CESGA platform these are the main improvements:

- Hadoop is now upgraded to Hadoop 3.
- Spark 2.4 is now the default version.
- HUE 4.
- HDFS Erasure coding: allows to reduce storage overhead over default 3x replication.
- Impala is now available as an alternative to Hive for interactive SQL queries.
- The HOME system has been migrated from GlusterFS to the new Netapp storage system, this has greatly improved the latency of the HOME filesystem.
- Improved reliability:
  - The HDFS NameNode is now in HA configuration.
  - The YARN ResourceManager is now in HA configuration.
- Improved security:
  - SSL/TLS is now enabled for more secure communications.

CHAPTER

FOUR

KNOWN ISSUES

- r-essentials not installing in Anaconda 2019.04: UnsatisfiableError

CHAPTER

FIVE

HOW TO CONNECT

Warning: The VPN must be active.

Before connecting, remember to start the VPN or connect from a remote desktop of the visualization platform. This will allow you to access the internal addresses of the cluster. If you need to setup the VPN you can check the *ANNEX I: VPN* section.

The most common way to connect is using an SSH client to log in to the cluster edge nodes:

```
ssh hadoop3.cesga.es
```

The cluster also has several web user interfaces available, you can access them through the https://bigdata.cesga.es WebUI Login option. You can find useful for example the link that you have there to the **HUE** portal, that allows you to use the Hadoop cluster from a graphical interface.

CHAPTER

SIX

BD|CESGA WEBUI

The cluster has a Web User Interface that can be accessed through: https://bigdata.cesga.es

From the main BD|CESGA page you can access the WebUI as well as find additional information about the platform including several tutorials that will help you to start using it.

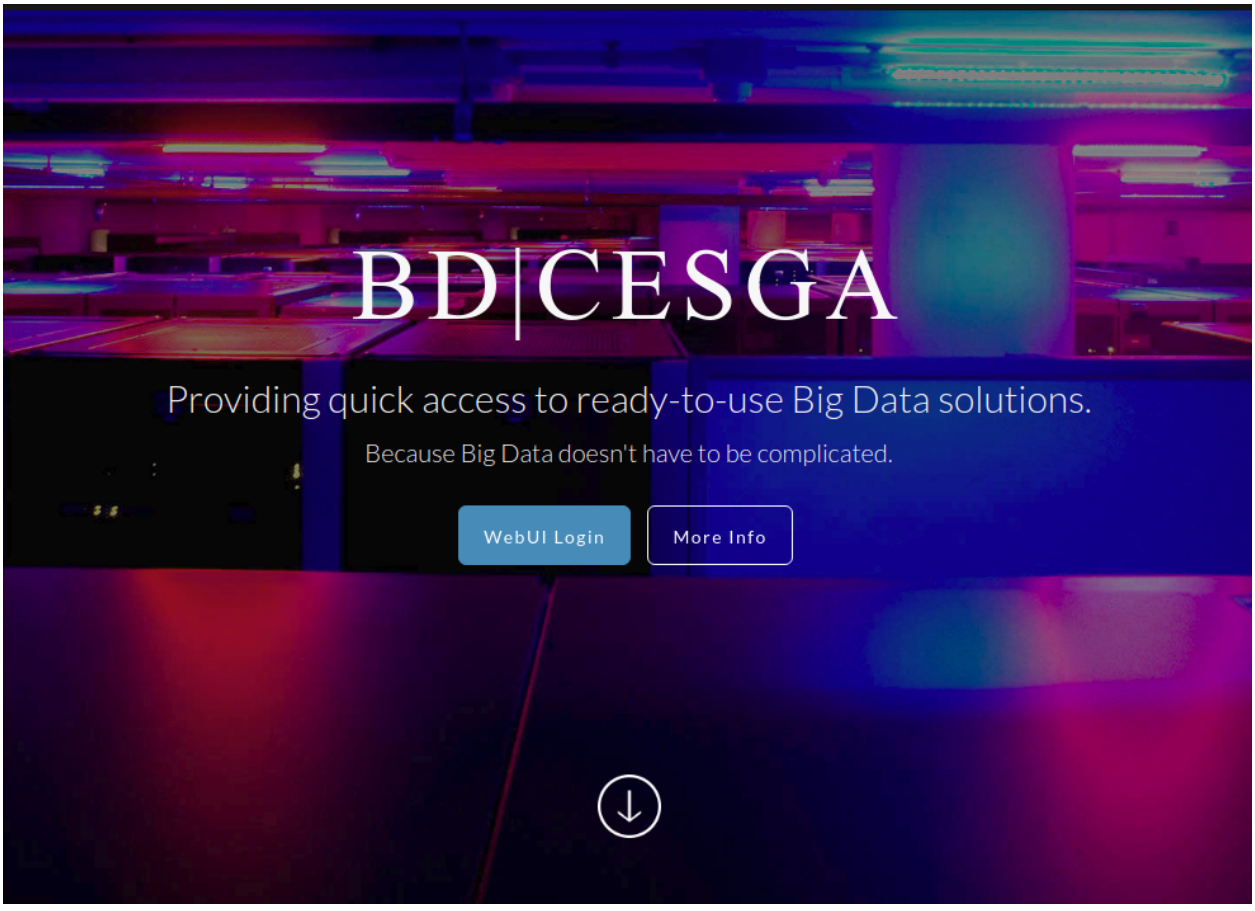


Fig. 1: The main BD|CESGA page.

The BD|CESGA Web User Interface provides you access to useful web interfaces that will allow you to use the platform in a graphical way.

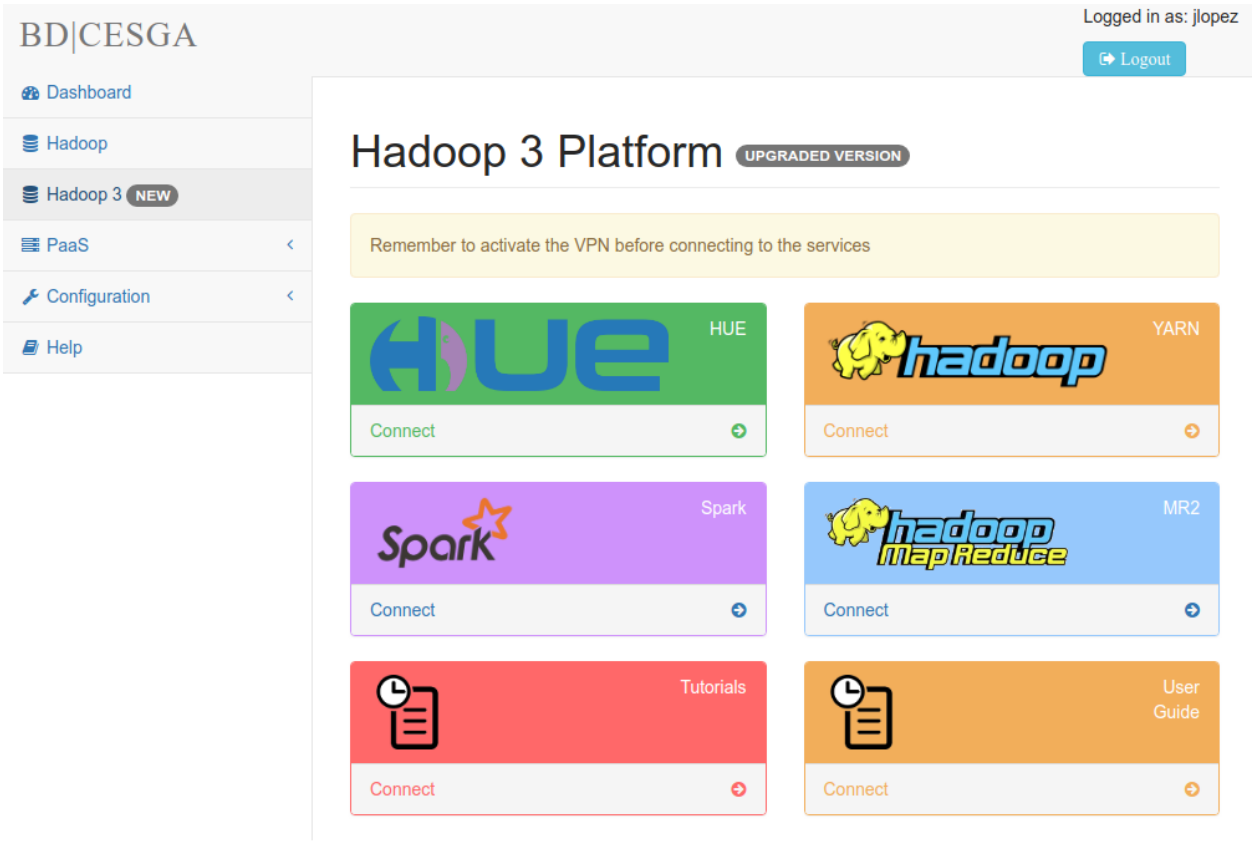


Fig. 2: The BD|CESGA Web User Interface.

CHAPTER

SEVEN

HUE: A NICE GRAPHICAL INTERFACE TO HADOOP

Hadoop User Experience (HUE) allows you to use a web user interface to perform common tasks like submitting new jobs, monitoring existing ones, execute Hive queries or browsing the HDFS filesystem.

The HUE interface can be accessed through the BD|CESGA WebUI.

You just have to follow the link to HUE and then login using your credentials (the same as for FT supercomputer).

Once inside HUE you can use it to launch Hive queries and display the results in a graphical way.

Using HUE you also have a quick Web UI to explore HDFS.

You can also use HUE to monitor your jobs in YARN:

And from the properties tab you can get the link to the tracking URL of the job:

Following that link will give you all the information about your job, for example in case of a Spark job you will access to the Spark UI:

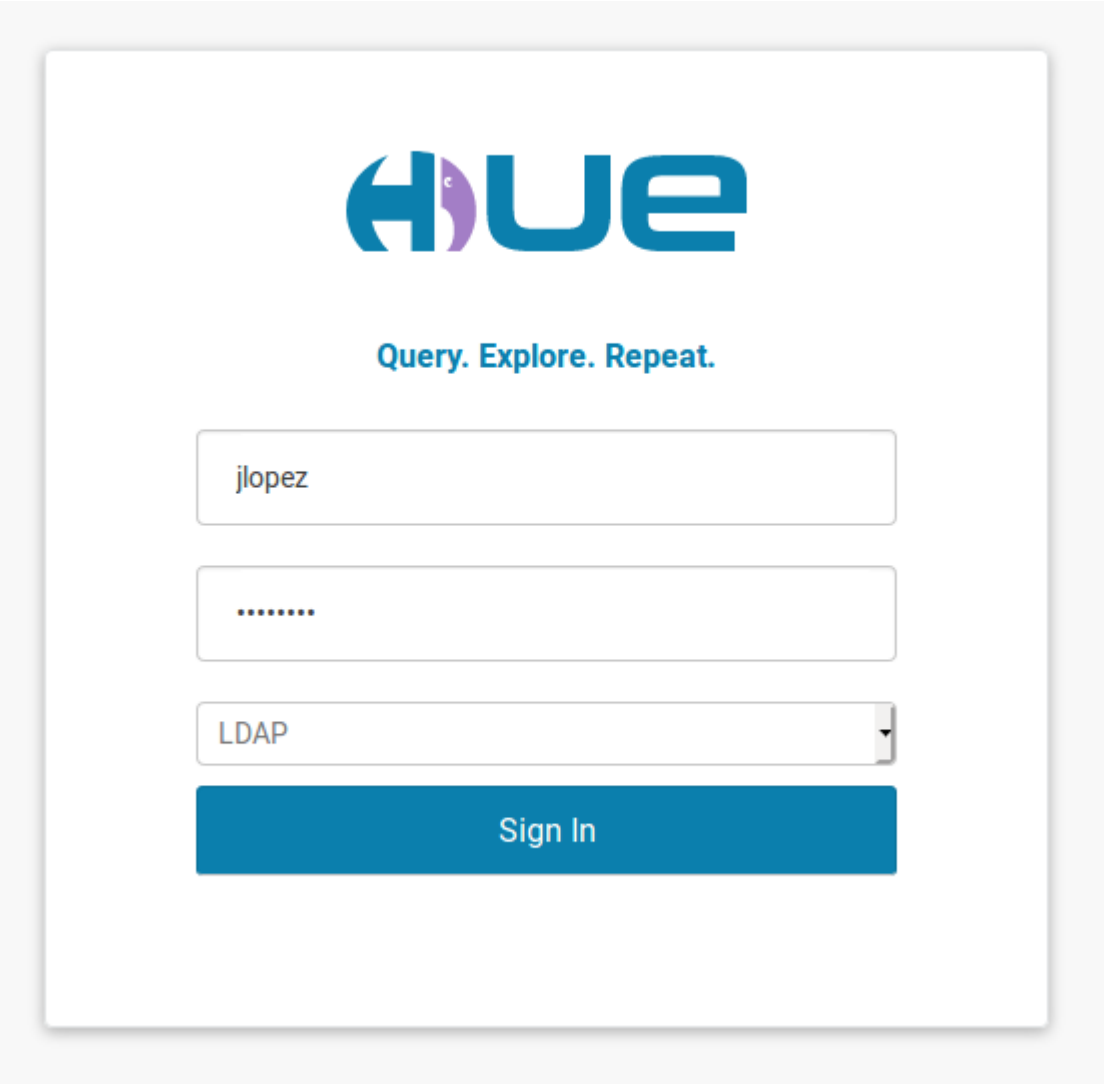


Fig. 1: The HUE login page: select LDAP and enter your credentials.

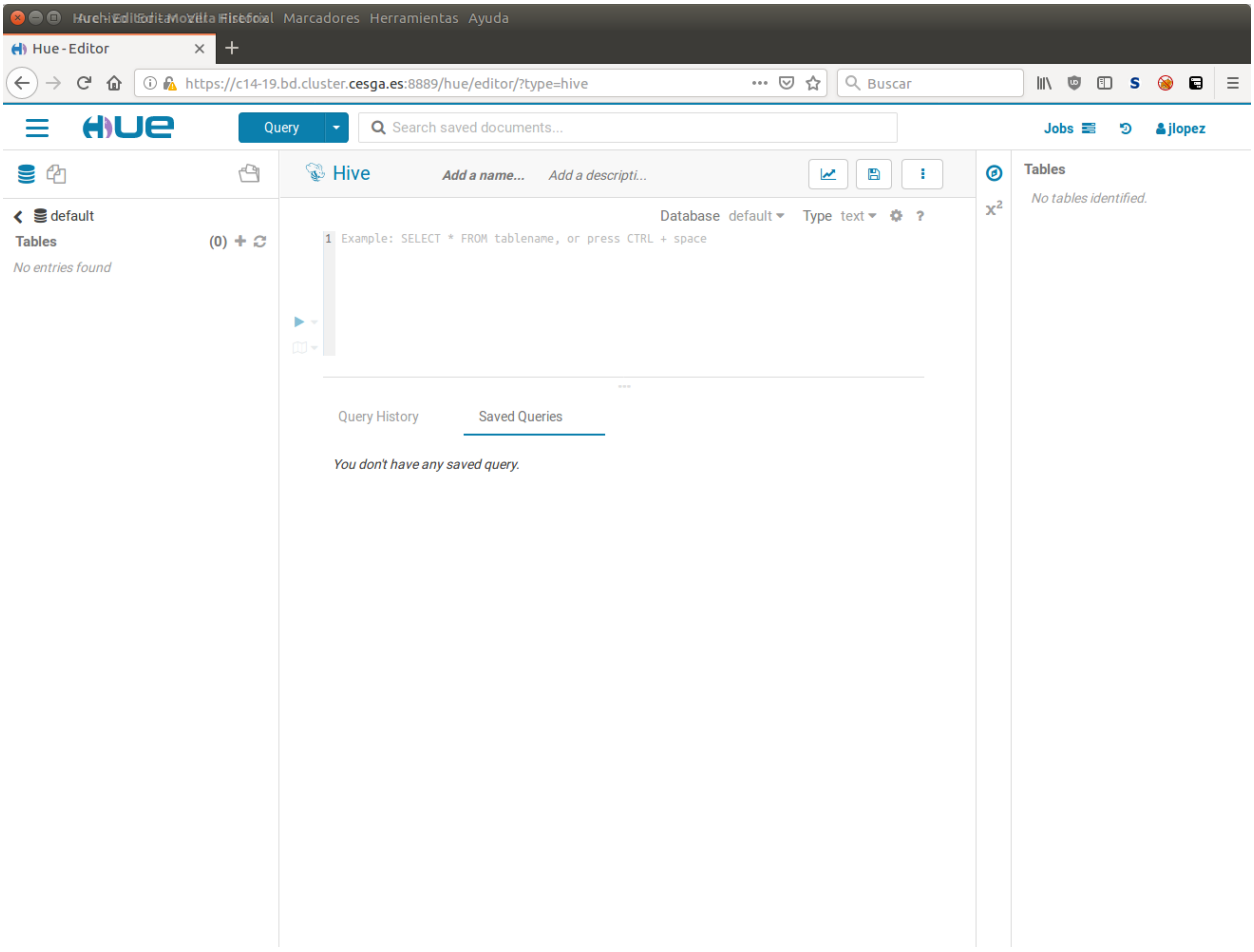


Fig. 2: Launching HiveQL queries from HUE.

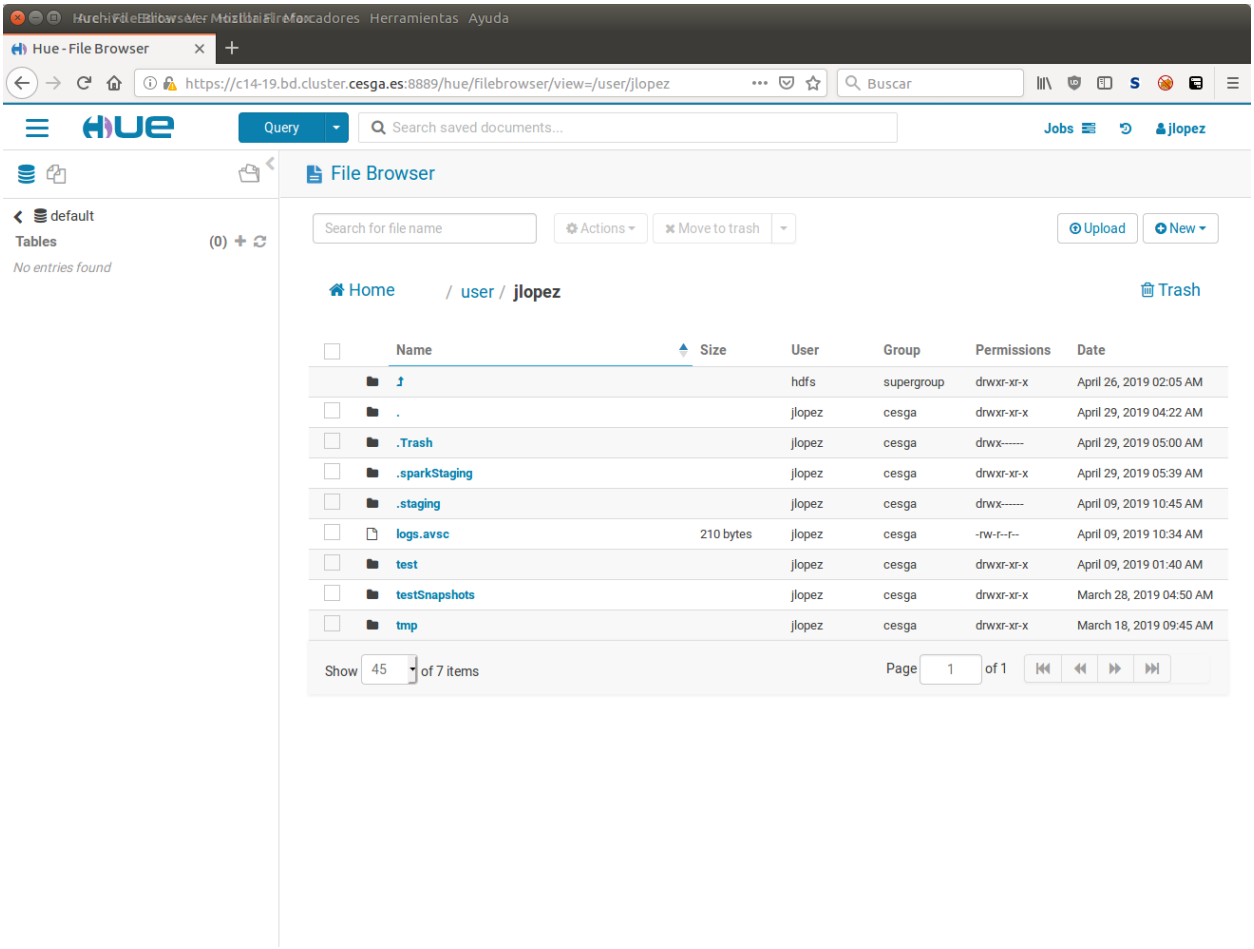


Fig. 3: Exploring HDFS from HUE.

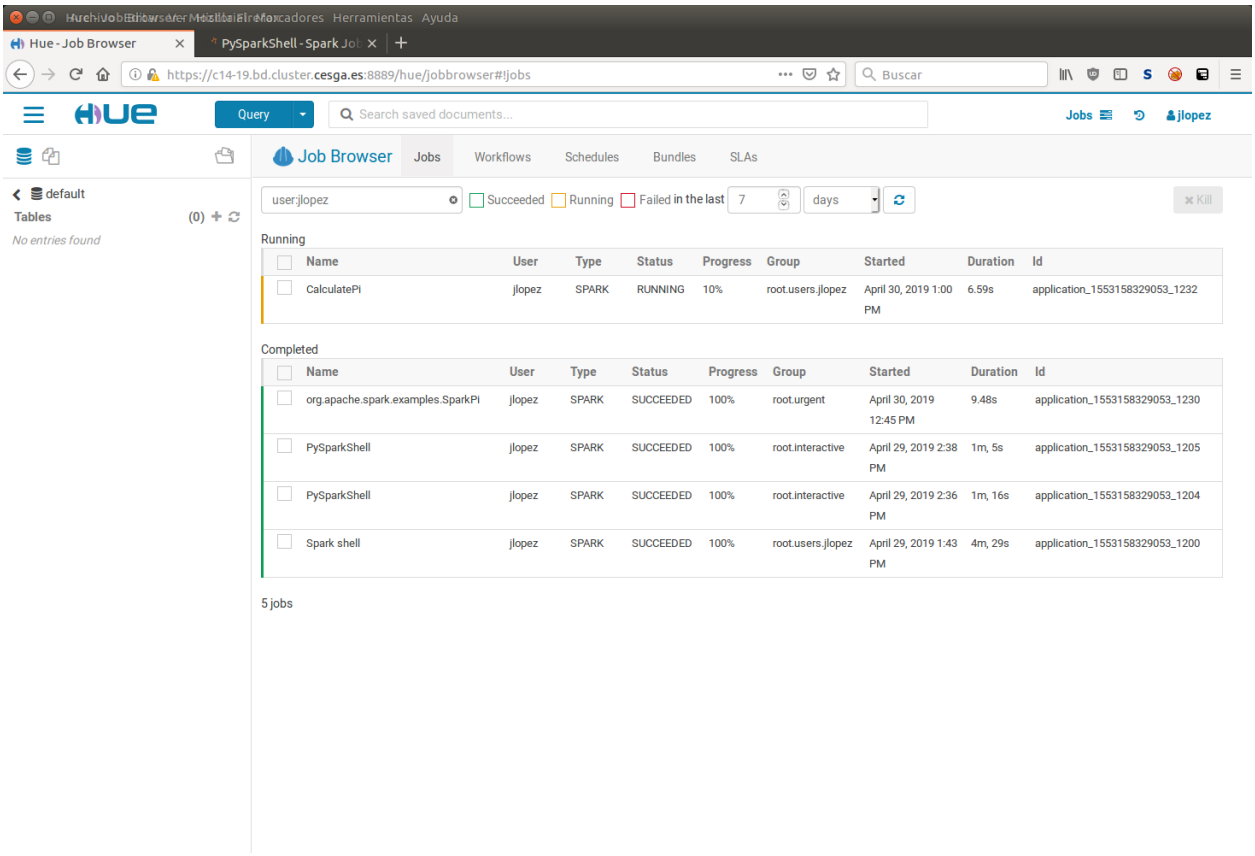


Fig. 4: Monitoring jobs using HUE.

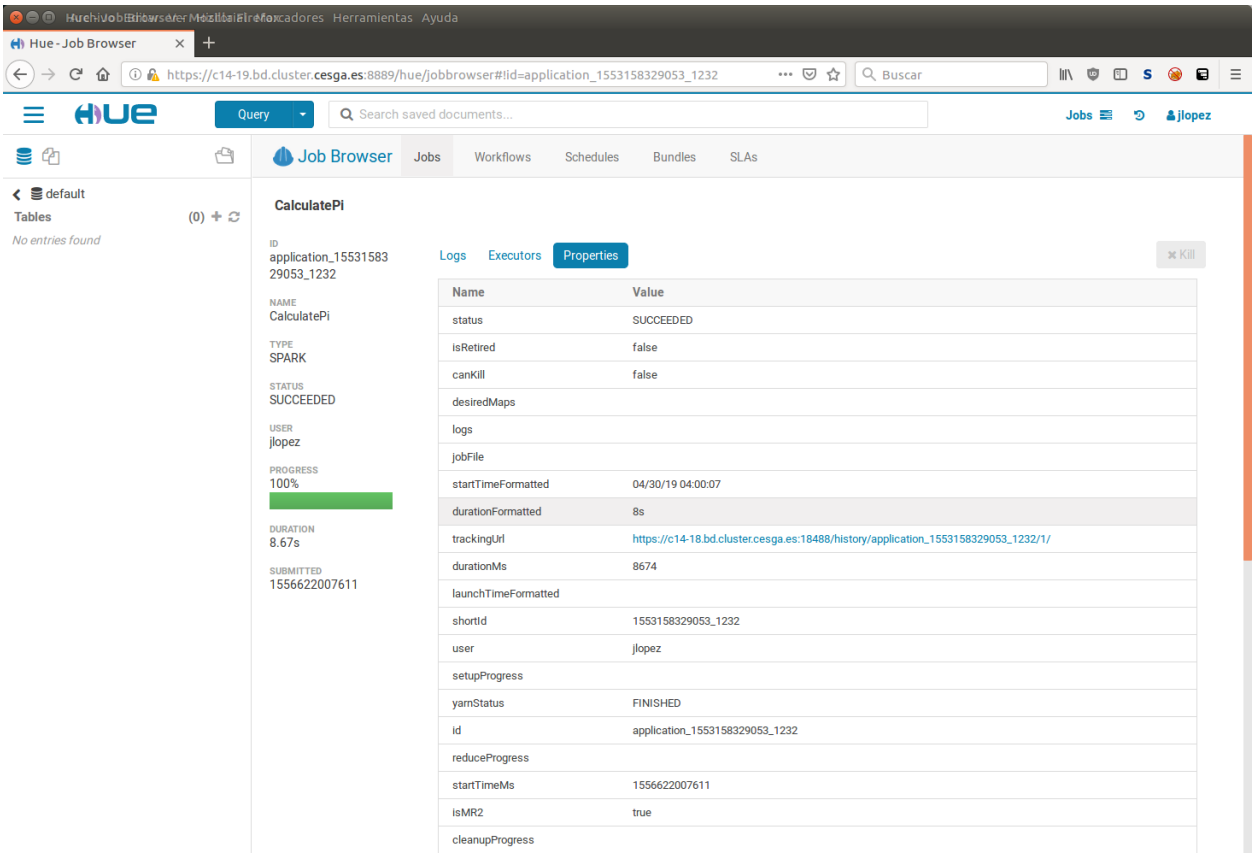


Fig. 5: Getting the tracking URL of a given job in the properties tab.

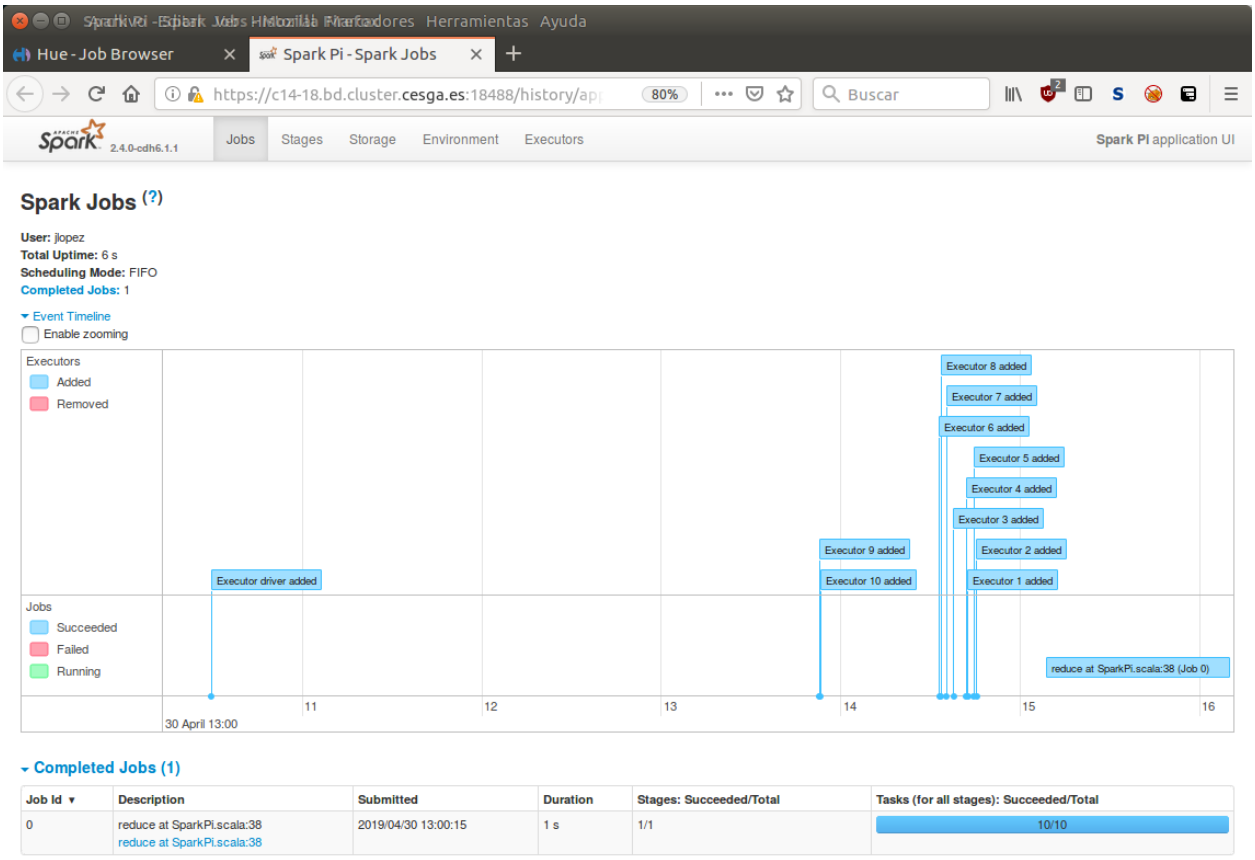


Fig. 6: The Spark UI showing details of a given job.

CHAPTER

EIGHT

MIGRATING DATA

The data that you have in the HOME have been automatically migrated so no need to move data. Now the old platform is using this same HOME.

To migrate the data in HDFS we recommend that you use the **discp** tool.

For example running the following command it will copy the directory *wcresult* from the old platform to the new one creating the target *wcresult* directory:

```
hadoop distcp -i -pat -update hdfs://10.121.13.19:8020/user/uscfajlc/wcresult hdfs://
↪nameservice1/user/uscfajlc/wcresult
```

Warning: Always run the *distcp* command from the new platform so it takes into account HA (nameservice1 is the HA nameservice ID).

Note: It is recommended to launch the distcp command inside a screen session so it will continue later.

Take into account that the HDFS data that is not migrated from the old platform will be lost once the old platform is decommissioned.

CHAPTER

NINE

FILESYSTEM QUOTAS

The filesystems have usage quotas that limit both the maximum allowed number of files and the total space used.

To see your current filesystem quotas and how close you are to reach the limits you can use the **myquota** command:

```
[sistemas@cdh61-login8 ~]$ myquota

 - HOME and Store filesystems:
   -------------------------
Filesystem                   space  quota  limit  files  quota  limit
10.117.49.101:/Home_BD/home  600K   800G   1024G  47     4295m  4295m

 - HDFS filesystem:
   -----------------
 FILES QUOTA        REMAINING    SPACE QUOTA        REMAINING        DIRS        FILES  ␣
↪            SIZE
         39.1 K          39.1 K            18 T            18 T            1           ␣
↪0                 0 /user/sistemas
```

To avoid unexpected failures in your jobs we recommend that you verify that you have enough space for your jobs before submitting them.

If you are close to reach your quota you need to increase the limits you can do an Additional Storage Request.

CHAPTER

TEN

HOW TO UPLOAD DATA

Depending on the size of the data that you want to upload you have different options:

- For small amounts of data (<10GB) you can use scp directly.
- For large amounts of data it is recommended that you use the GridFTP service through Globus Online.

In all cases it is recommended that you do the transfer against our DTN server: **dtn.srv.cesga.es**, this will give you a much better network performance. In Globus Online the endpoint of this server is: **cesga#dtn**.

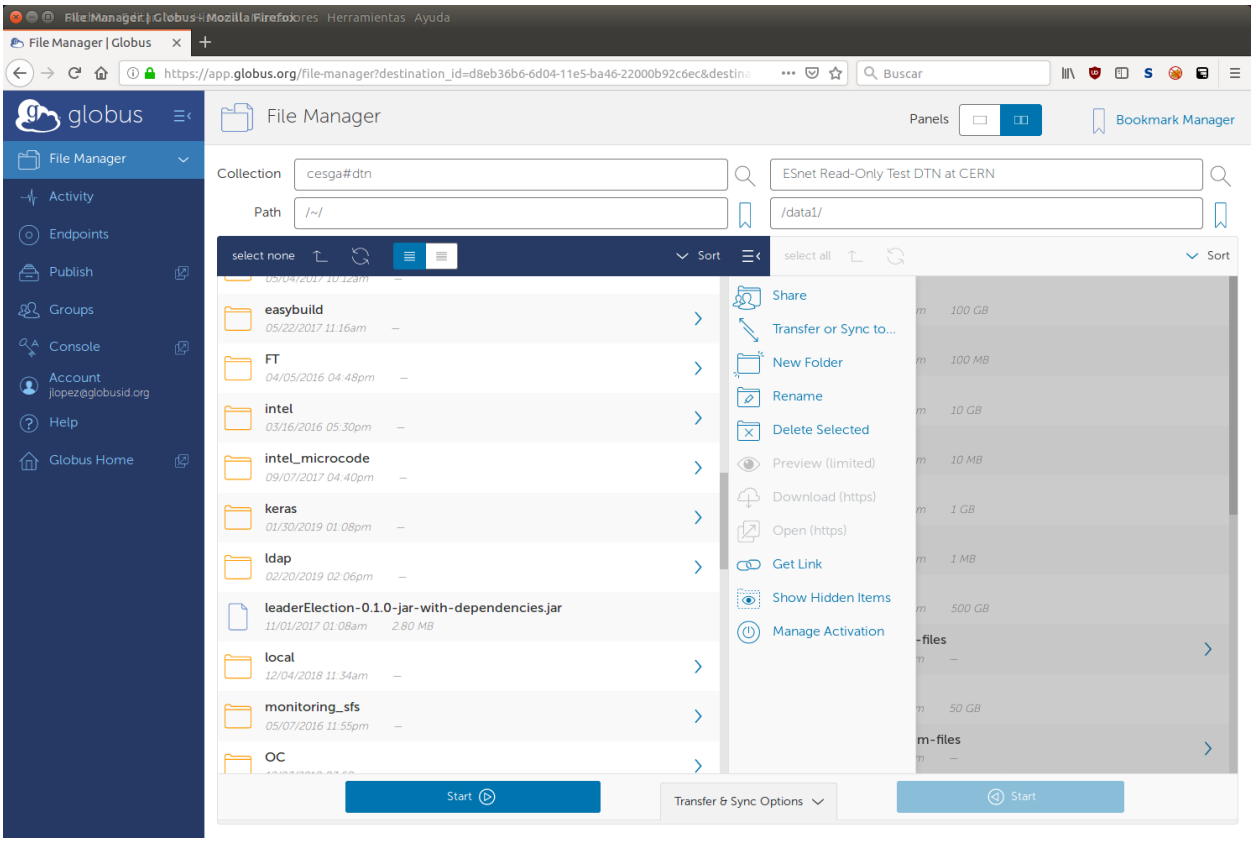


Fig. 1: Using Globus Online to transfer files.

CHAPTER

ELEVEN

YARN: THE HADOOP BATCH SYSTEM

YARN is the batch system in a Hadoop ecosystem, playing the same role as SLURM in the FT supercomputer.

To launch an application usually you just use the corresponing tool commands instead of using directly YARN commands.

For example to submit a Spark job you will use:

```
spark-submit
```

or in the case of Hive or Impala they take care of launching the MapReduce jobs needed to perform your SQL query automatically.

In case of MapReduce jobs you will launch them using the YARN CLI with a command similar to the following:

```
yarn jar application.jar DriverClass input output
```

It is also useful to list the running applications with:

```
yarn application -list
```

And, sometimes it is even easier to can check the overal status of the platform running:

```
yarn top
```

YARN takes care of collecting all the logs generated by your application in all the nodes, to see them you just have to run:

```
yarn logs -applicationId applicationId
```

And finally to kill an application you will use:

```
yarn application -kill applicationId
```

You can also find useful the YARN Web UI that allows to easily track the progress of your application and to get all the details about your job. You can access it through the *HUE: A nice graphical interface to Hadoop*.

Note: We recommend to use HUE to access the tracking URL of your jobs. If you try to do it directly through the YARN UI you will find restrictions due to the fact that the YARN UI runs as the dr.who user.

By default jobs will be submitted to your own queue and resources will be shared with the rest of users using the YARN fair share scheduler and Dominant Resource Fairness (ie. it will take into account both CPU and memory).

Interactive jobs like *Jupyter Notebooks* run in a dedicate queue for interactive jobs.

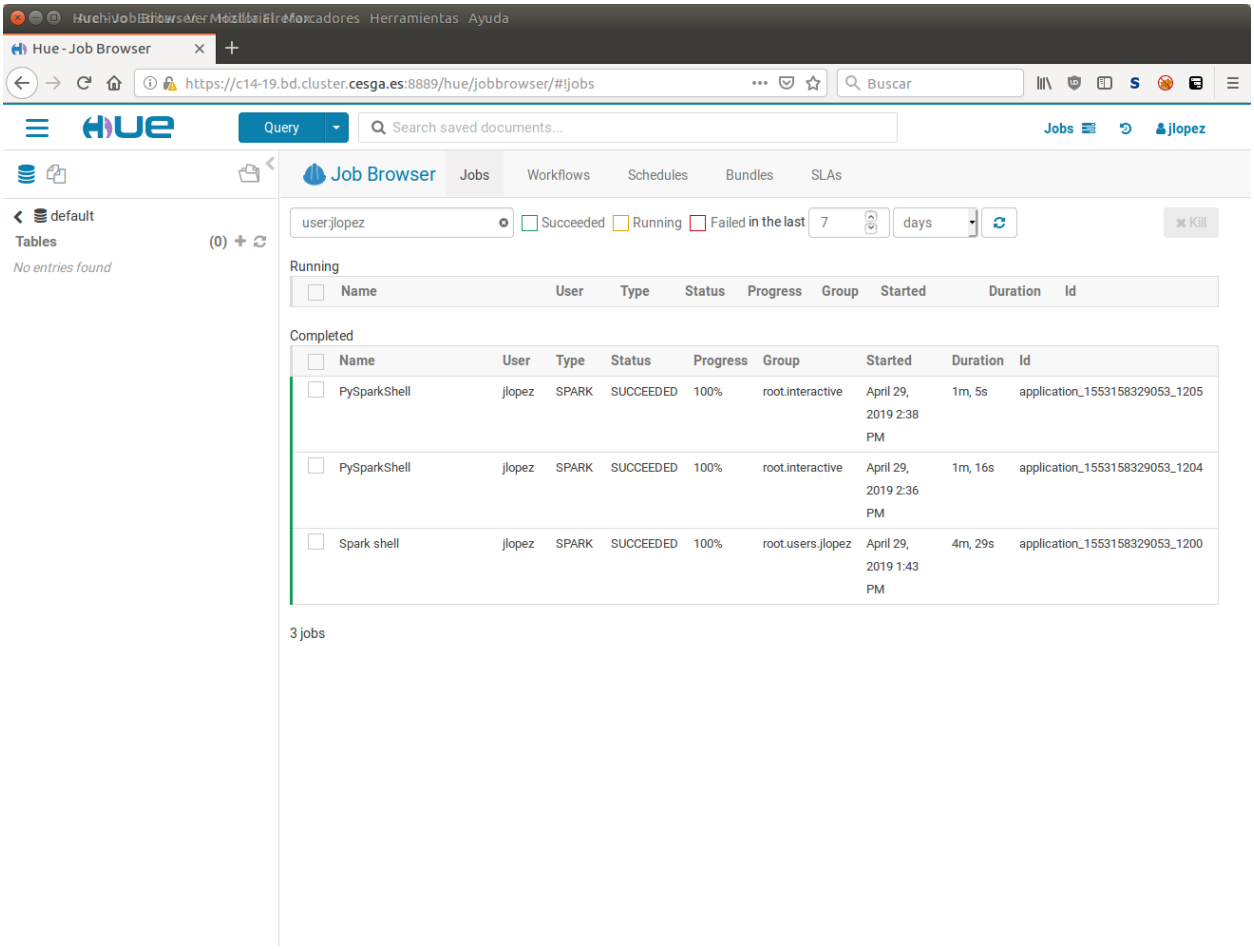


Fig. 1: Monitoring jobs using HUE.

There is also a queue that can be used for small but urgent jobs (urgent queue).

Jobs should be composed of lots of short running tasks so they share resources nicely with other jobs. In case of jobs with long running tasks that monopolize resources during large times the schedule is configured to preempt this tasks so they allow other applications to also run.

The common pattern in MapReduce jobs is to automatically split the job in lots of small tasks, and each of them runs in a small portion of the data. YARN is optimized for this type of jobs, for jobs that do not follow this pattern other platforms like the FT supercomputer with the SLURM batch system SLURM are usually best suited to the task.

In case of doubts or if you have special needs don't hesitate to contact us.

For further information on how to use YARN you can check the YARN Tutorial that we have prepared to get you started and the Hadoop Documentation as reference.

CHAPTER

TWELVE

HDFS: THE HADOOP FILE SYSTEM

HDFS is the underlying distributed filesystem that you will use to run your applications so they will take advantadge of parallel data processing.

HDFS is optimized for large sequencial reads and the best performance is obtained on large files (>1GB).

The files are split in blocks that by default have a block size of 128MB and blocks are replicated across multiple nodes, guaranteeing fault-tolerance in case of a node failure.

By default each block will have **3 replicas** but you can control the amount of replicas when you create the file.

We recommend that you upload the files first to the BD HOME filesystem using the DTN server and then from there you copy them to HDFS. See the *How to upload data* section for more information.

To put a file in HDFS you can run the following command:

```
hdfs dfs -put file.txt file.txt
```

To list files:

```
hdfs dfs -ls
```

To create a directory:

```
hdfs dfs -mkdir mydir
```

To get a file from HDFS to the local filesystem:

```
hdfs dfs -get file.txt
```

Using *HUE: A nice graphical interface to Hadoop* you have a nice Web UI to explore HDFS, you can access it through the *BD|CESGA WebUI*.

For further information on how to use HDFS you can check the HDFS Tutorial that we have prepared to get you started and the Hadoop Documentation as reference.

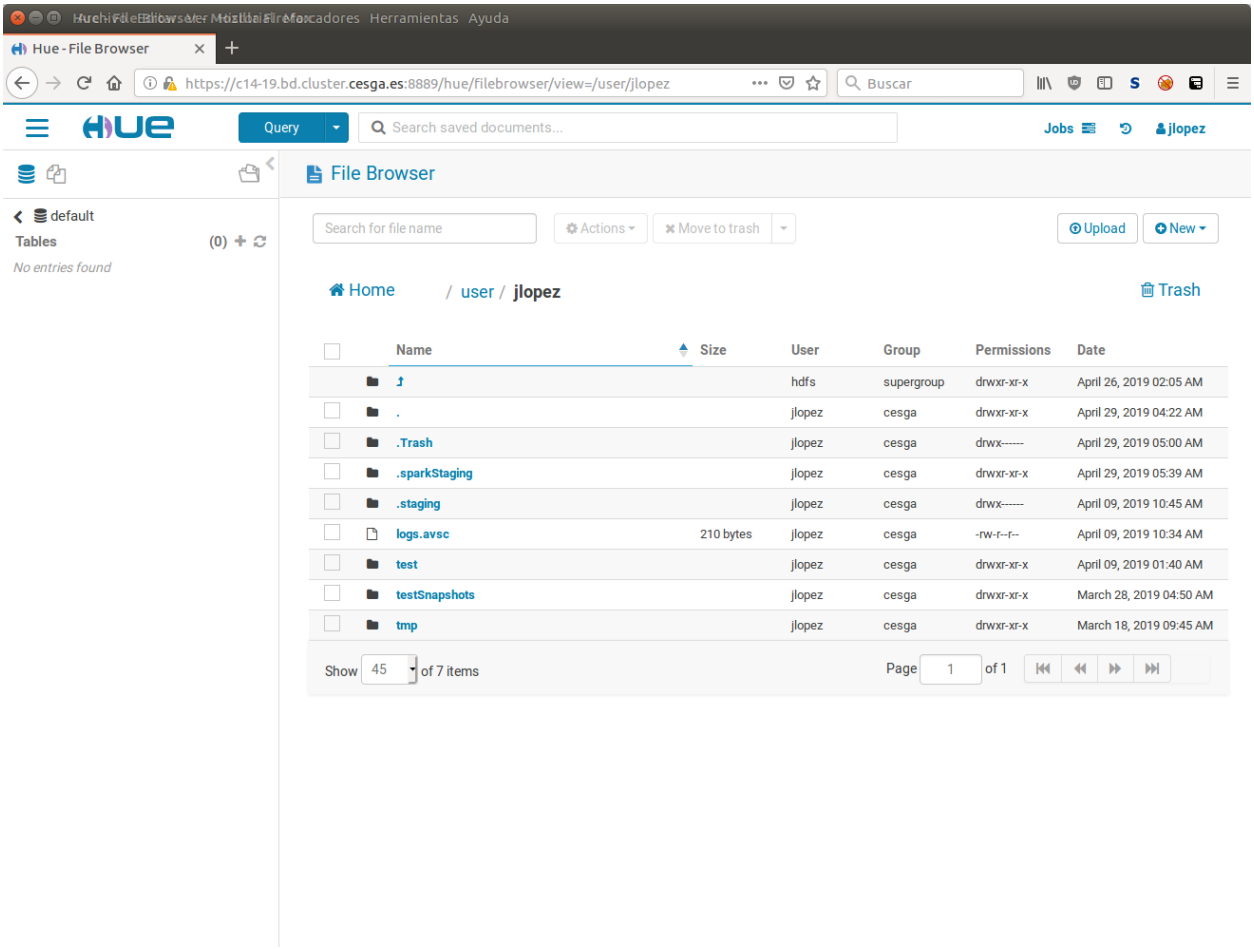


Fig. 1: Exploring HDFS from HUE.

CHAPTER

THIRTEEN

SPARK

Spark is probably the most popular tool in the Hadoop ecosystem nowdays. Apart from the performance improvements it offers over standard MapReduce jobs, it simplifies considerably the process of developing an application because it offers high level programming APIs like the DataFrames API.

The default version included in the platform is Spark 2.4.0, and you can easily start an interactive shell:

```
spark-shell
```

Similarly to start an interactive session using Python:

```
pyspark
```

Note: If using Python we recommend that you use the Anaconda version provided through *Modules: Additional Software*.

To use it with Python from the Anaconda distribution first load the desired anaconda version of the module, for example:

```
module load anaconda2/2018.12
```

If using Anaconda, then you can also use ipython for the interactive pyspark session so you get a nicer CLI:

```
PYSPARK_DRIVER_PYTHON=ipython pyspark
```

To submit a job:

```
spark-submit --name testWC test.py input output
```

The jobs will be submitted to YARN and queued for execution. Depending on the load of the platform the execution will take more or less time.

Note: You can acess the Spark UI through the *BD|CESGA WebUI* and *HUE: A nice graphical interface to Hadoop*.

For further information on how to use Spark you can check the Spark Tutorial that we have prepared to get you started. For more information you can check the PySpark Course Material and the Sparklyr Course Material, these are courses that you can also attend to learn more. Finally, you can also find useful the Spark Guide in the CDH documentation, and of course, the great documentation provided by the Spark project.

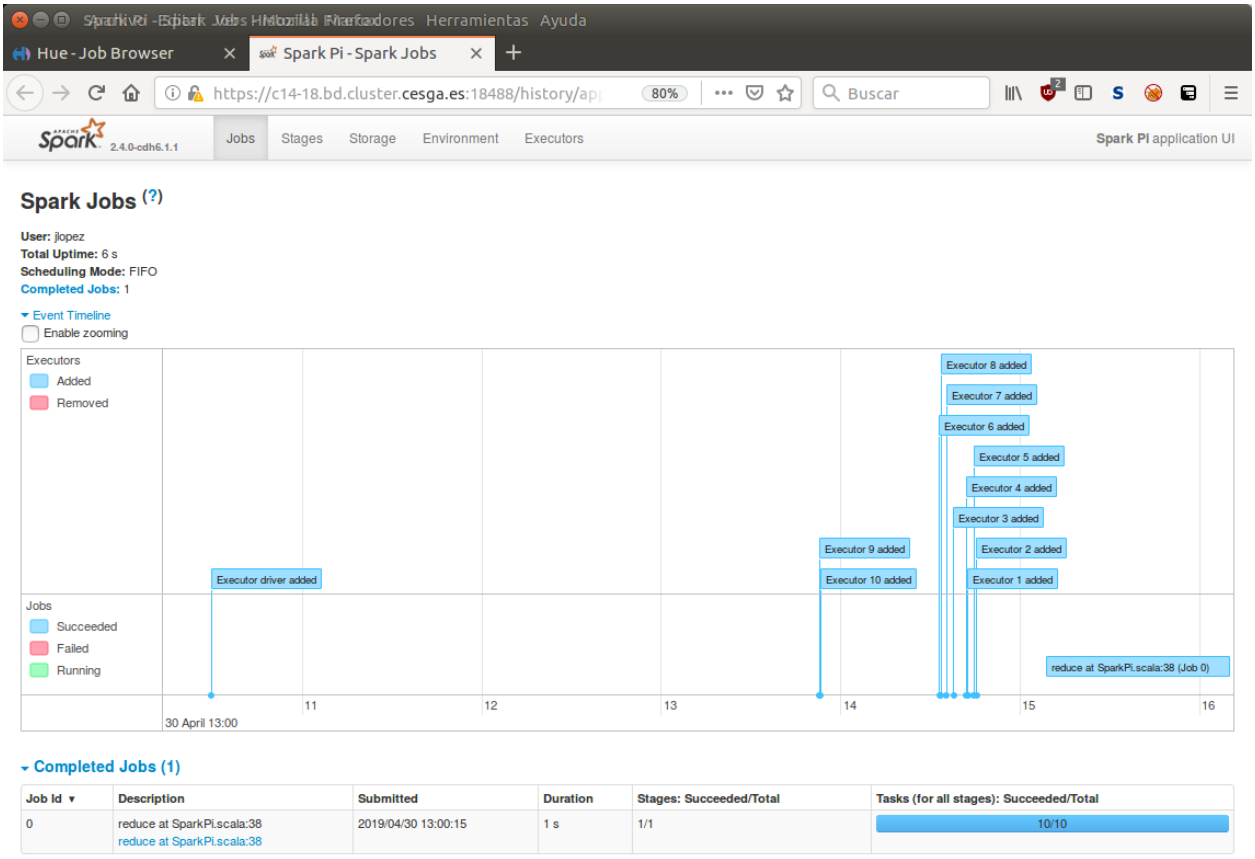


Fig. 1: The Spark UI showing details of a given job.

CHAPTER

FOURTEEN

SPARKLYR

Sparklyr is an R package that interfaces from R to Apache Spark. It was created in 2016 by the Rstudio team and fits in the tydiverse ecosystem providing a complete dplyr backend for Spark.

It makes Spark's APIs accesible from R, including SparkDataFrames and the MLlib machine learning library.

To use sparklyr on the platform you will need to load the sparklyr module (*Modules: Additional Software*).:

```
module load sparklyr
```

This module includes an anaconda installation of python 2.7, R 3.1.5, sparklyr 1.0.5, and all its dependencies, so in order to use it you only need to start R and load the package:

```
R
```

```
library(sparklyr)
```

Note: You can check the list of preinstalled packages by typing *installed.packages()* on the R console.

After that you will need to connect to the spark cluster, this is done using the *spark_connect()* function.:

```
sc <- spark_connect(master = "yarn-client", spark_home = Sys.getenv('SPARK_HOME'))
```

And then use your spark connection *sc* to access any spark tool.

Finally execute::

```
spark_disconnect(sc)
```

to disconnect from spark.

You can also use Sparklyr on a Jupyter Notebook with an R kernel.

Warning: Remember to disconnect from spark and properly shut down the notebook server before logging out.

Sparklyr is not limited to interactive use, you can also use *spark-submit* to launch a script as a job:

```
spark-submit --class sparklyr.Shell '/opt/cesga/anaconda/Anaconda2-2018.12-sparklyr/lib/
↪R/library/sparklyr/java/sparklyr-2.4-2.11.jar' 8880 1234 --batch example_sparklyr_
↪script.R
```

For further information on Sparklyr you can check the getting started Sparklyr Tutorial and take a look at the Sparklyr workshop. There is also the official documentation by the RStudio Team, including this handy cheatsheet.

CHAPTER

FIFTEEN

JUPYTER NOTEBOOKS

The Jupyter Notebooks allow you to have a very nice web UI to develop and run your applications.

Warning: To access the Jupyter web UI first you have to start the VPN.

To start Jupyter you can run:

```
start_jupyter
```

and it will provide you the address you need to point your browser.

Note: Under the hood the Jupyter notebooks runs a Spark session to run your Spark commands in the cluster.

If you want to customize the version of Python used you can do it by loading first the Anaconda module you are interested in, and after that running Jupyter. For example:

```
module load anaconda3/2018.12
start_jupyter
```

You can also try the new Jupyter Lab interface:

```
module load anaconda2/2018.12
start_jupyter-lab
```

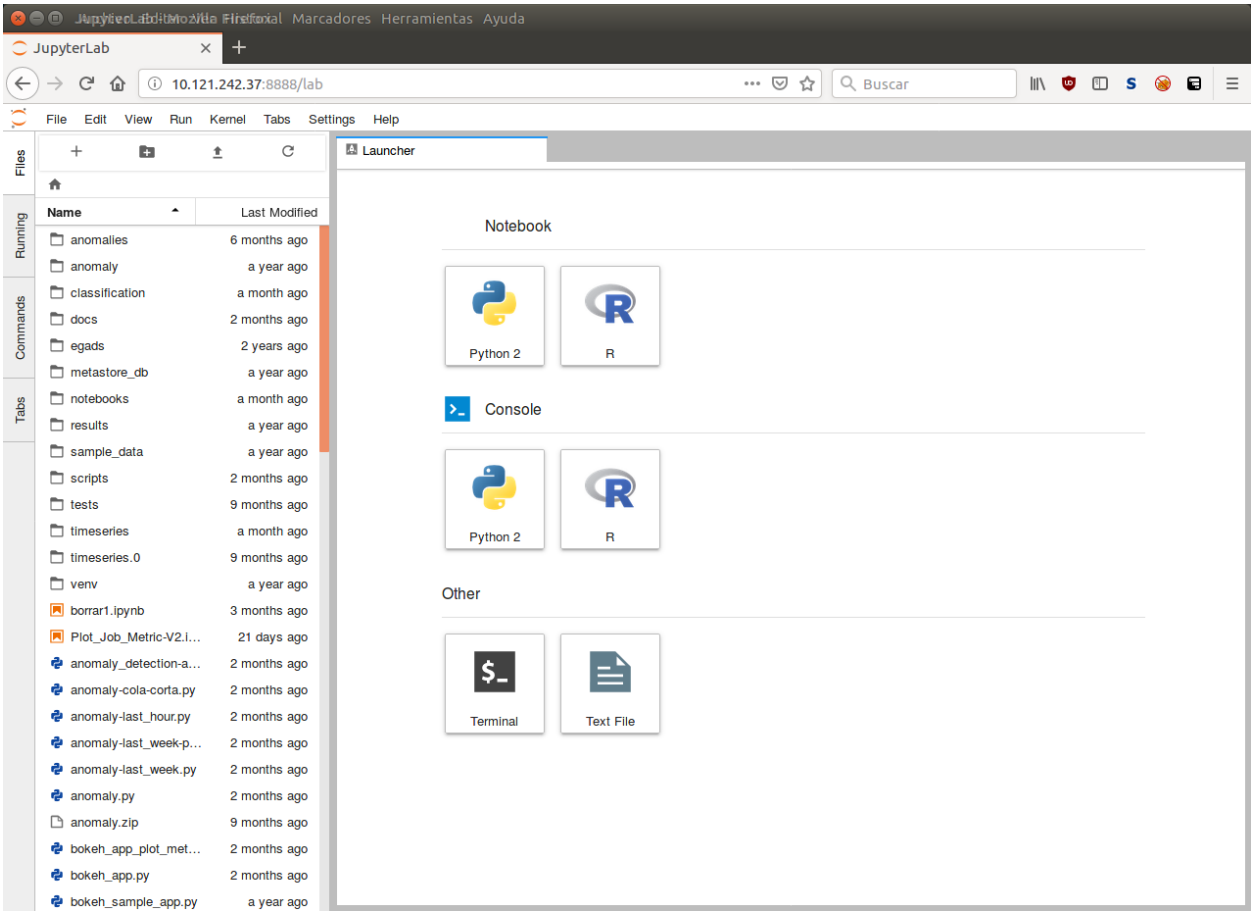


Fig. 1: The new Jupyter Lab User Interface.

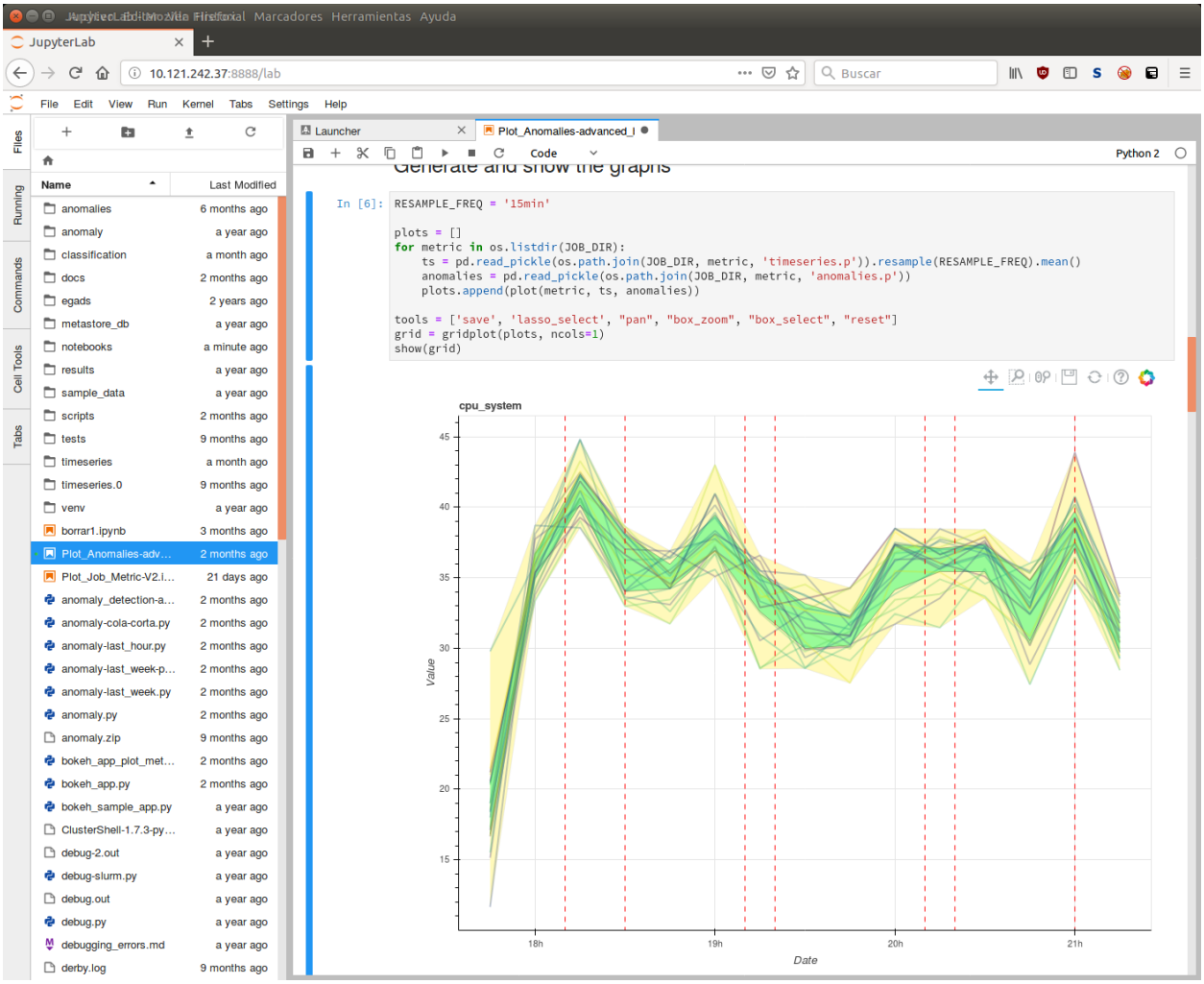


Fig. 2: Showing interactive graphs inside a Jupyter Lab notebook.

CHAPTER

SIXTEEN

HIVE

Hive allows to query data in HDFS using SQL queries, so it is a very useful tool for all those people familiar with SQL.

Under the hood Hive translates the SQL queries into MapReduce jobs that are run using YARN.

There are several options to start executing queries with Hive:

- Using the old and now deprecated hive CLI:

```
hive
```

- Using the new beeline client:

```
beeline> !connect jdbc:hive2://c14-19.bd.cluster.cesga.es:10000/default;ssl=true;
↪sslTrustStore=/opt/cesga/cdh61/hiveserver2.jks;trustStorePassword=notsecret
```

- Using the HUE web interface that you can access through:

```
https://bigdata.cesga.es
```

If you are just testing hive, we recommend that you start using the **testing database** instead of the default one:

```
use testing;
```

We do not recommend to create tables in the default database, instead if you have tables that you want to keep create a database with your username and then create your tables in this database. For example if your username is uscfajlc create a database with that name and then use it to create your tables:

```
create database if not exists uscfajlc;
use uscfajlc;
```

For enhanced privacy, you can restrict access to the data in your database just to your username:

```
hdfs dfs -chmod go-rwx /user/hive/warehouse/uscfajlc.db
```

Of course you can use HDFS ACLs to fine tune the permissions to further fit your needs.

For further information on how to use Hive you can check the Hive Tutorial that we have prepared to get you started and the Hive Guide in the CDH documentation.

CHAPTER

SEVENTEEN

IMPALA

Like Hive, Impala allows to do interactive SQL queries on HDFS data. It is the recommended option to perform fast small interactive queries because you will benefit from the faster start-up time. To run long batch queries we recommend to use Hive.

Since we are using a secure cluster you have to invoke impala-shell with the –ssl option and you also have to indicate the location of one of the impalad daemons. For example you can launch the impala-shell running:

```
impala-shell –ssl –impalad=c14-1
```

Note: In the impalad option you can indicate any of the worker nodes of the cluster: c14-[1-14].

For more information about Impala you can check the Impala Guide.

CHAPTER

EIGHTEEN

SQOOP

Sqoop allows to easily import data from relational databases into HDFS.

We have already deployed the Sqoop connectors for the following databases:

- MySQL / MariaDB
- PostgreSQL
- Microsoft SQL Server
- Oracle 18c

This way, out of the box you can use the Sqoop tool to import data from any of these databases:

```
sqoop import \
    --username ${USER} --password ${PASSWORD} \
    --connect jdbc:postgresql://${SERVER}/${DB} \
    --table mytable \
    --target-dir /user/username/mytable \
    --num-mappers 1
```

Note: We recommend that you use only one mapper process to avoid overloading your database.

If you need to import data from a different database don't hesitate to contact us.

For further information on how to use Sqoop you can check the Sqoop Tutorial that we have prepared to get you started and the Sqoop Guide in the CDH documentation.

CHAPTER

NINETEEN

MODULES: ADDITIONAL SOFTWARE

Similarly to how it is done in the FT supercomputer, the modules allow you to load additional software that is not included by default in the Hadoop distribution that we are using to deploy the platform (in our case CDH 6.1.1).

It also allows to load different versions of the tools than the ones included in the platform.

You can check the available software using:

```
module available
```

For example, using modules you can load Python 3 that is not officially supported neither by Centos 7 nor by CDH 6. To load it you can run:

```
module load anaconda3/2018.12
```

This is the current list of available modules:

- anaconda2
- anaconda3
- maven
- sbt

For further information on how to use modules you can check the Modules Tutorial that we have prepared to get you started, and the official Lmod documentation.

CHAPTER

TWENTY

WANT TO KNOW MORE

For more information we recommend that you start looking at the tutorials that we have prepared to get you started with each of the tools.

After that, if you need additional information you can check CDH documentations with the different component guides as well as the official documentation of each component.

20.1 Tutorials

You can find the complete list of tutorials at https://bigdata.cesga.es/#tutorials

- HDFS Tutorial
- YARN Tutorial
- Spark Tutorial
- Jupyter Tutorial
- Hive Tutorial
- Sqoop Tutorial
- MapReduce Tutorial

Additionally you can find useful the material of the PySpark and Sparklyr courses:

- PySpark Course Material
- Sparklyr Course Material

20.2 CDH Documentation

- CDH Documentation for CDH 6.1
- Spark Guide
- Hive Guide
- Impala Guide
- HUE Guide

20.3 Reference Documentation

In the following page you can find the links to the Reference documentation for each component.

CHAPTER

TWENTYONE

ANNEX I: VPN

The VPN software allows you not only to connect to CESGA in a secure way but also to access internal resources that you can not reach otherwise.

21.1 How to install the VPN software in Windows

Checkpoint VPN is used to establish a secure connection to our services. It enables remote users to securely access our network resources from anywhere in the world using encrypted tunnels, ensuring confidentiality and integrity of data being transmitted over the internet.

To install Checkpoint, you must first download the executable file on this link. Please note that this file is compatible with **Windows 7, 8.1, 10, and 11**. If your Windows version is not one of these, you may encounter some difficulties when installing Checkpoint.

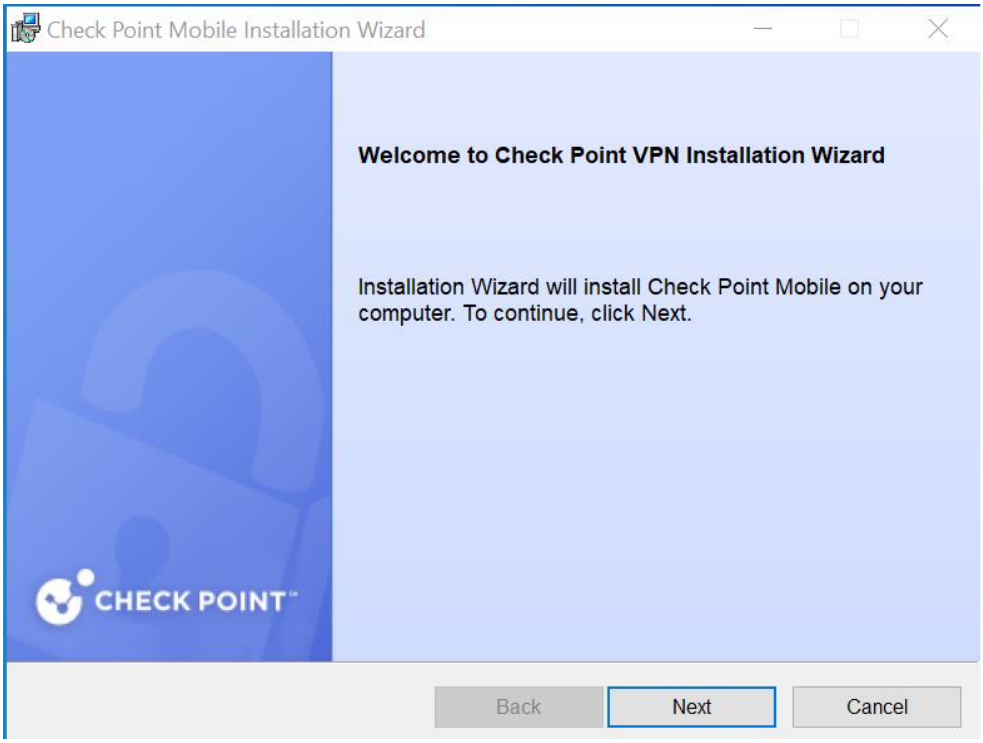
The installation will be carried out with the CheckPointVPN_CESGA_HPC executable file following these steps:

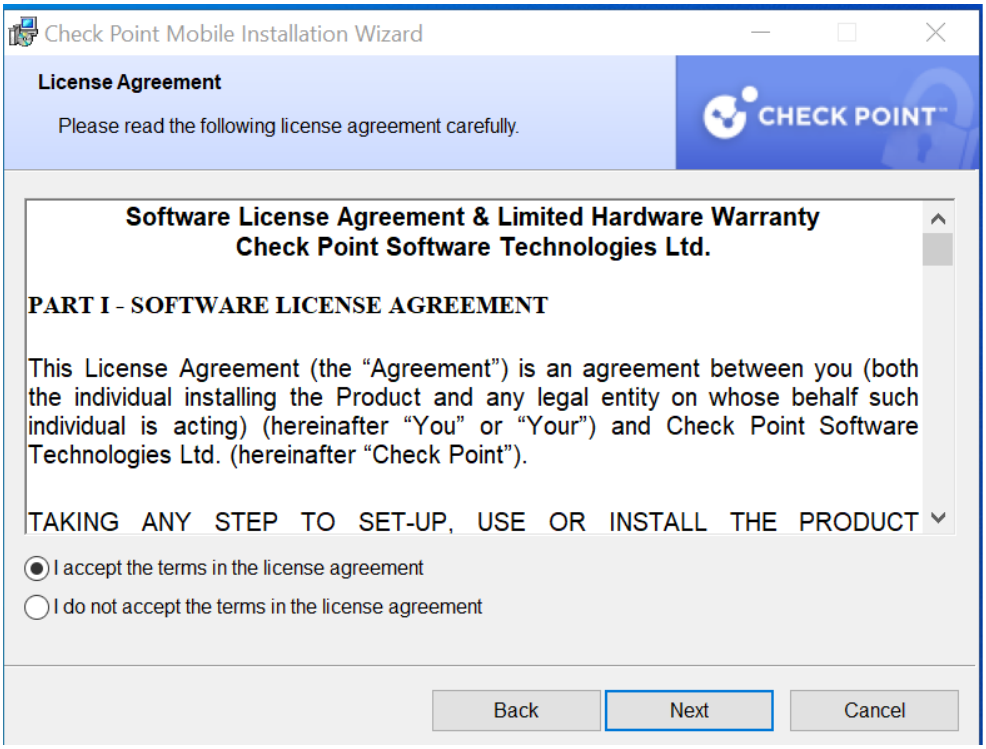
1. Double-click on the executable file. A Windows message will appear indicating that changes will be made to the system. You must accept these changes.
2. Next, the installation wizard will start with a welcome message. Click on Next.
3. Next, the License Agreement will appear, it must be accepted by checking the option "I accept the terms in the license agreement" and clicking on Next.
4. The next screen will display the **default** directory chosen to save the installation files. It is **recommended** to leave the default path as shown. Then, click on Install.
5. When the installation starts, a progress bar will appear which should not take more than 5 minutes. Finally, it will show that the installation has finished. Click on Finish.
6. Automatically after finishing the installation, the **Checkpoint** menu will open:

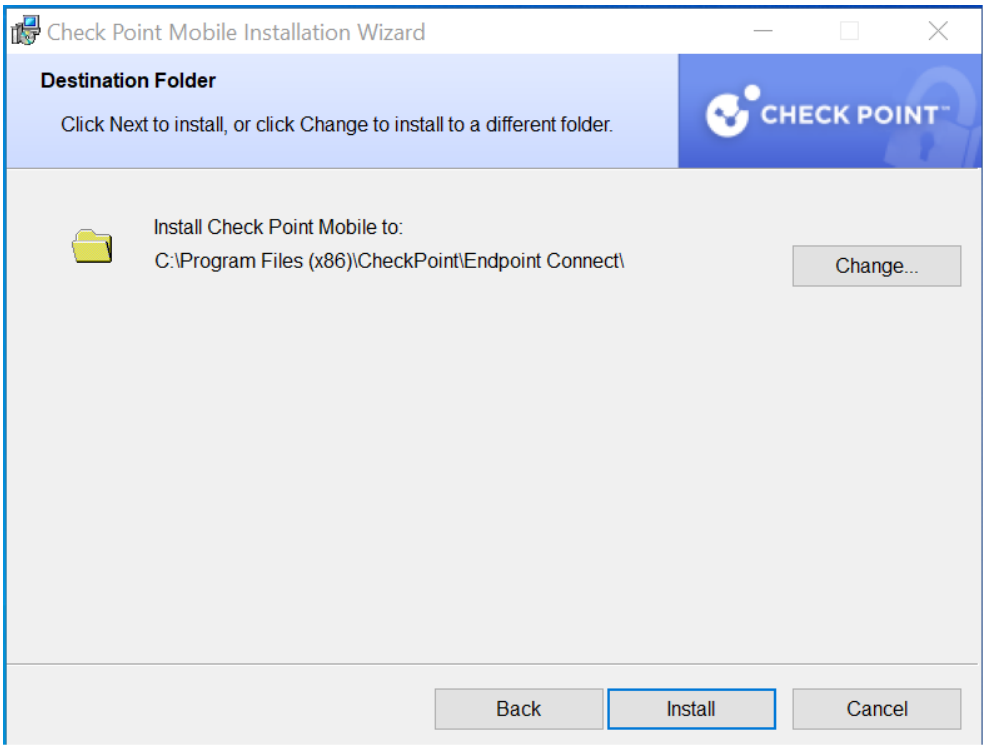
As you can see at the top, the site (CESGA-VPN) is already configured by **default**, so it will only be necessary to enter the username and password and click on Connect. If the hostname/IP address is not set by default, complete the server configuration with **secure.cesga.es** in the blank space labeled "Server address or name". If you check the "Display name" box, it will allow you to enter an alternative name for the connection such as "CESGA-VPN".

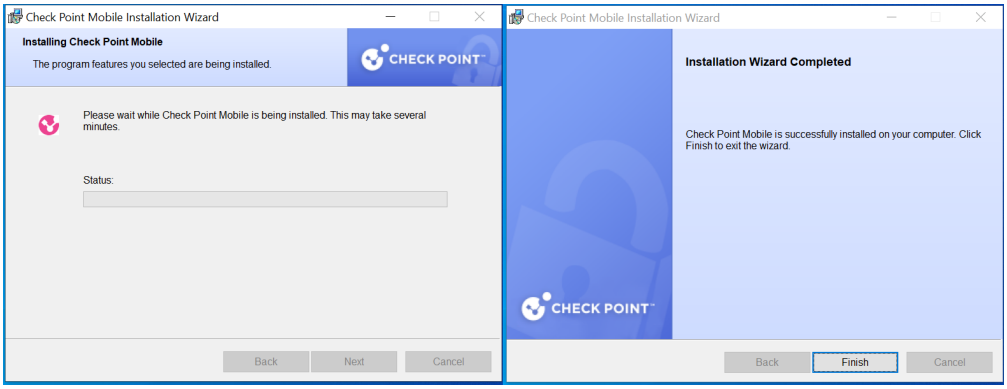
Warning: These credentials are the same ones used to access FinisTerrae III or other services offered by CESGA. That is, it's the username that was granted when registering for CESGA services. **DO NOT ENTER YOUR FULL EMAIL OR DOMAIN @FT3.CESGA.ES**.

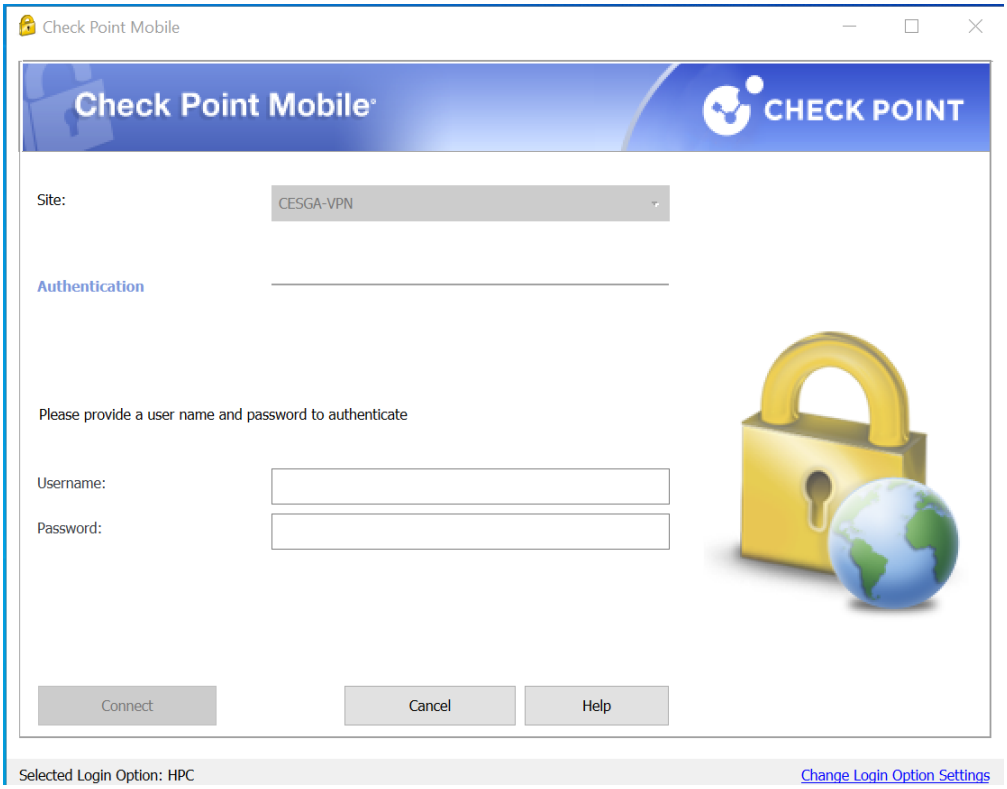
For example, if you use user_cesga@ft3.cesga.es to connect to FinisTerrae III or your mail is user_cesga@dominion.of.your.center.com the username that should be entered in the CheckPoint credentials is just **user_cesga**.


Check Point Mobile Installation Wizard
Welcome to Check Point VPN Installation Wizard
Installation Wizard will install Check Point Mobile on your computer. To continue, click Next.
CHECK POINT
Back
Next
Cancel


Check Point Mobile Installation Wizard
License Agreement
Please read the following license agreement carefully.
CHECK POINT
Software License Agreement & Limited Hardware Warranty
Check Point Software Technologies Ltd.
PART I - SOFTWARE LICENSE AGREEMENT
This License Agreement (the "Agreement") is an agreement between you (both the individual installing the Product and any legal entity on whose behalf such individual is acting) (hereinafter "You" or "Your") and Check Point Software Technologies Ltd. (hereinafter "Check Point").
TAKING ANY STEP TO SET-UP, USE OR INSTALL THE PRODUCT
I accept the terms in the license agreement
I do not accept the terms in the license agreement
Back
Next
Cancel


Check Point Mobile Installation Wizard
Destination Folder
Click Next to install, or click Change to install to a different folder.
CHECK POINT
Install Check Point Mobile to:
C:\Program Files (x86)\CheckPoint\Endpoint Connect\
Change...
Back
Install
Cancel


Check Point Mobile Installation Wizard
Installing Check Point Mobile
The program features you selected are being installed.
CHECK POINT
Please wait while Check Point Mobile is being installed. This may take several minutes.
Status:
Back
Next
Cancel
Check Point Mobile Installation Wizard
Installation Wizard Completed
Check Point Mobile is successfully installed on your computer. Click Finish to exit the wizard.
CHECK POINT
Back
Finish
Cancel



Also, if by any reason you are prompted with the window below, please select the option **HPC (default)**.

When the connection configuration is complete, a window will appear similar to the one shown in step 6. Simply enter your username and password to activate the VPN.

8. Once the credentials are checked it will show that the connection is active.

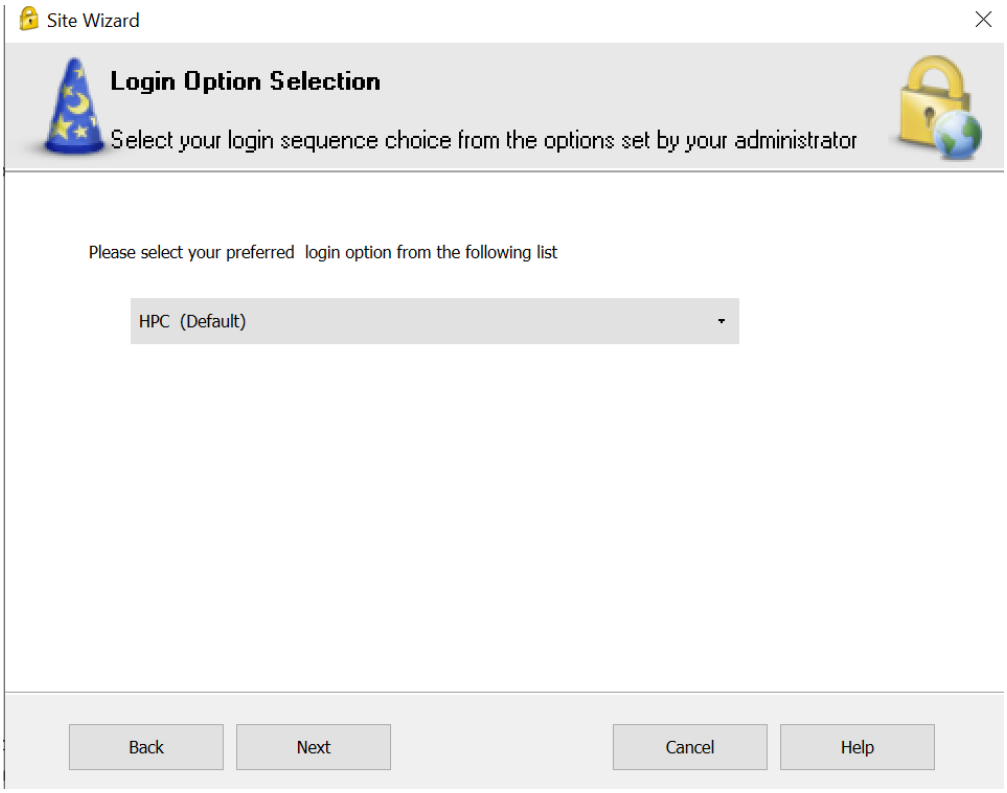
Note: As indicated by the above message, the maximum duration of the VPN connection is **24 hours**. 5 minutes before this time expires, a **notification** will appear to re-enter the password. This will **restart** the connection time counter and allow you to connect for another 24 hours.

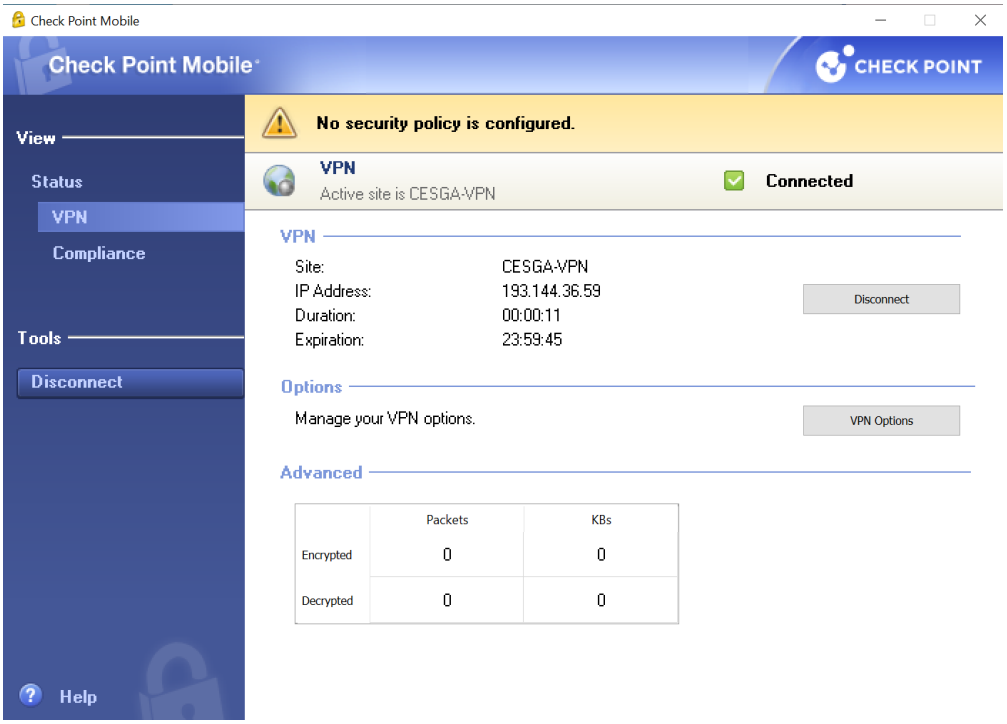
How to log in once Checkpoint is installed?

Once the CheckPoint client has been installed on your computer and to activate the VPN, you should follow these steps:

1. Look for CheckPoint in your installed applications and open it.
2. The login screen shown in the screenshot of section 6 will appear. As indicated in that section, you should enter your credentials and click on Connect.
3. It is very likely that the program will automatically run when you turn on your computer, so you can find the CheckPoint icon (a yellow padlock) on the desktop taskbar. If you right-click on it, the Connect option will appear and will let you to activate the VPN connection.

If you wish to disconnect from the VPN, on the menu shown in the previous screenshot, you can turn it off by clicking on Shutdown Client.


Site Wizard
Login Option Selection
Select your login sequence choice from the options set by your administrator
Please select your preferred login option from the following list
HPC (Default)
Back
Next
Cancel
Help


Check Point Mobile
Check Point Mobile
CHECK POINT
View
Status
VPN
Compliance
Tools
Disconnect
Help
No security policy is configured.
VPN
Active site is CESGA-VPN
Connected
VPN
Site: CESGA-VPN
IP Address: 193.144.36.59
Duration: 00:00:11
Expiration: 23:59:45
Disconnect
Options
Manage your VPN options.
VPN Options
Advanced
Packets
KBs
Encrypted
0
0
Decrypted
0
0

21.2 How to install the VPN sofware in MacOS

Checkpoint VPN is used to connect to our services. To install the Checkpoint software, you must first download the right version depending on your macOS version:

- For macOS 10.14, 10.15, 11 and 12: Download Checkpoint VPN for macOS
- For macOS from 10.11 to 10.13: Use version 80.89
- For older versions try version 80.41 However, we cannot guarantee that it will work on every older version.

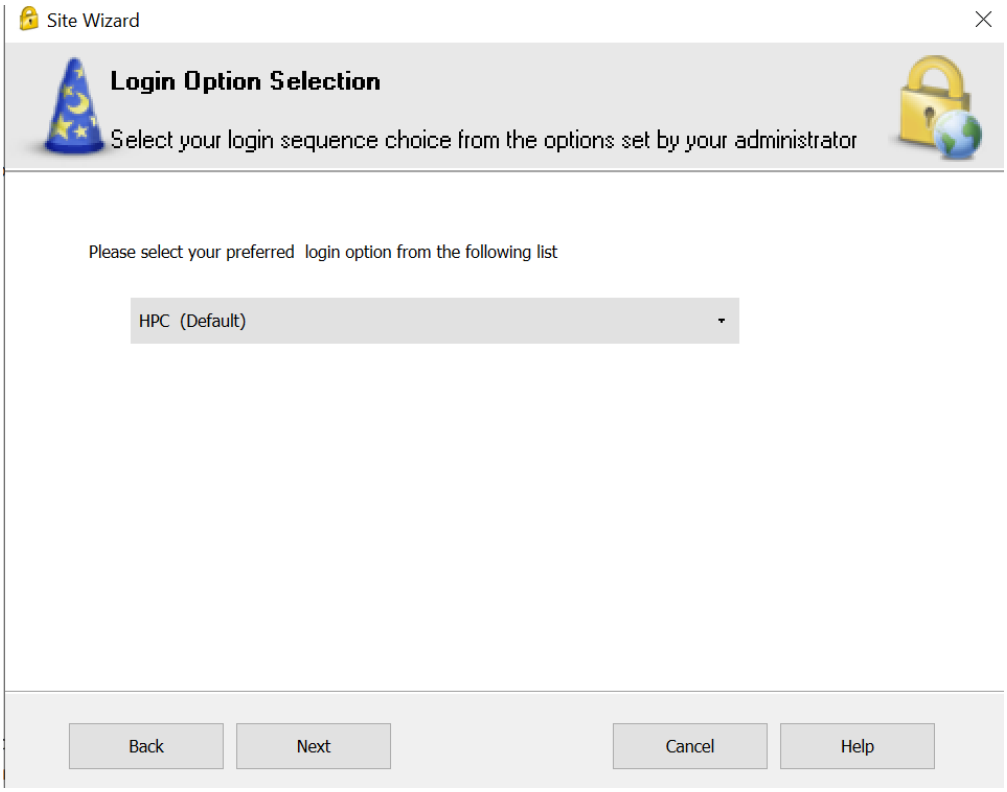
The installation will be carried out with any of the excutable files described above and following the steps of the wizard. Be careful, the server/hostname/IP address **is not set by default** on macOS, so you will have to complete the configuration being the hostname/IP address **secure.cesga.es**.

As shown on the screenshot above, you have to add **secure.cesga.es** on the blank space of "Server address or name". If you check the "Display name" box, It would let you to write and alternative name for the connection, for example "CESGA-VPN".

If, by any reason, you are prompted with the window above please select the option **HPC (default)**.

Once the configuration of the server is made and you connect the VPN, it will prompt you to add your user and password. The credentials to log in have the same warning as in the others OS:

Warning: These credentials are the same ones used to access FinisTerrae III or other services offered by CESGA. That is, it's the username that was granted when registering for CESGA services. **DO NOT ENTER YOUR FULL EMAIL OR DOMAIN @FT3.CESGA.ES**.



For example, if you use user_cesga@ft3.cesga.es to connect to FinisTerrae III or your mail is user_cesga@dominion.of.your.center.com the username that should be entered in the CheckPoint credentials is just **user_cesga**.

21.3 How to install the VPN software in Linux

Checkpoint VPN is used to connect to our services. In Linux we will use the **snx** client to connect. Just follow the steps explained below:

1. From the command line of your computer, download the snx file executing:

```
wget http://bigdata.cesga.es/files/snx
```

2. Change the permissions of the file to make it executable:

```
chmod a+x snx
```

3. Install the **required dependencies**, multiarch must be enable because snx is a i386 binary:

```
sudo dpkg --add-architecture i386
sudo apt update
sudo apt install libaudit1:i386 libbsd0:i386 libc6:i386 libcap-ng0:i386 libgcc-
↪s1:i386 libpam0g:i386 libstdc++5:i386 libx11-6:i386 libxau6:i386 libxcb1:i386␣
↪libxdmcp6:i386
```

4. Once the installation is complete, to start the VPN connection you must execute the following command: `sudo ./snx -s secure.cesga.es -u <username>` You will need to enter your username and password.

> **Warning:** The <username> is your cesga username eg. uscfajlc. Do not confuse with your email address.

5. It will prompt you to enter your password, and once the connection is established, it will display the message:

```
Check Point's Linux SNX
build 800010003
Please enter your password:
NX - connected.
Session parameters:
===================
Office Mode IP      : ...
DNS Server          : ...
Secondary DNS Server: ...
Timeout             : 24 hours
```

As indicated by the above message, the maximum duration of the VPN connection is **24 hours**. 5 minutes before this time expires, a **notification** will appear to re-enter the password. This will **restart** the connection hours counter and allow you to connect for another 24 hours.

6. To disconnect the VPN, use the following command:

```
sudo snx -d
```