



**TIJUANA INSTITUTE OF
TECHNOLOGY**

SYSTEMS AND COMPUTER DEPARTMENT

SEMESTER

January - June 2021

CAREER

Computer Systems Engineer

CLASS

Big data / Data mining

STUDENT

Camacho Manabe Juan Daniel # 17210534

Salazar Ibarra Jesus Rodrigo 16212542

TEACHER

Ing. José Christian Romero Hernández

Tijuana, B.C. a 02/03/2021

Before installing spark it is necessary to install the java jdk, there are several versions, for the moment the most stable is version 8, in the Ubuntu 20.04 operating system terminal. we install with the following command.

```
sudo apt-get -y install openjdk-8-jdk-headless
```

Then we must make sure that we have a version of Python installed, otherwise we install with the following command.

```
sudo apt install python3
```

Next, we go to the page <https://spark.apache.org/downloads.html> to download version 2.4.7, which is the one we are going to work with as it is more stable.



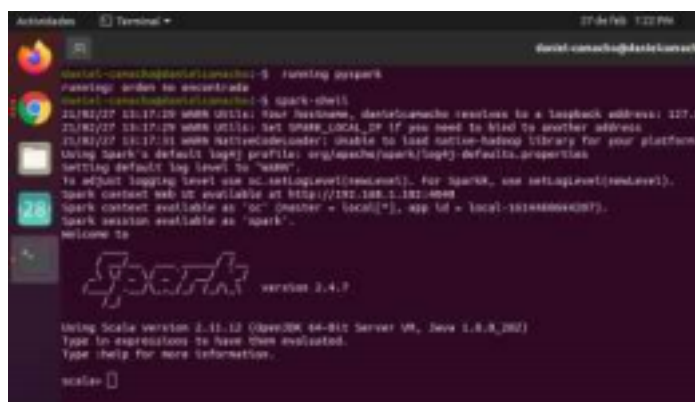
We go to the terminal and we position ourselves in the folder where this is the file, then we unzip the file and with the command "mv" we move the directory to "spark".

```
tar -xzf spark-2.4.7-bin-hadoop2.7.tgz
mv spark-2.4.7-bin-hadoop2.7 spark
TOOD - Add Python environment
```

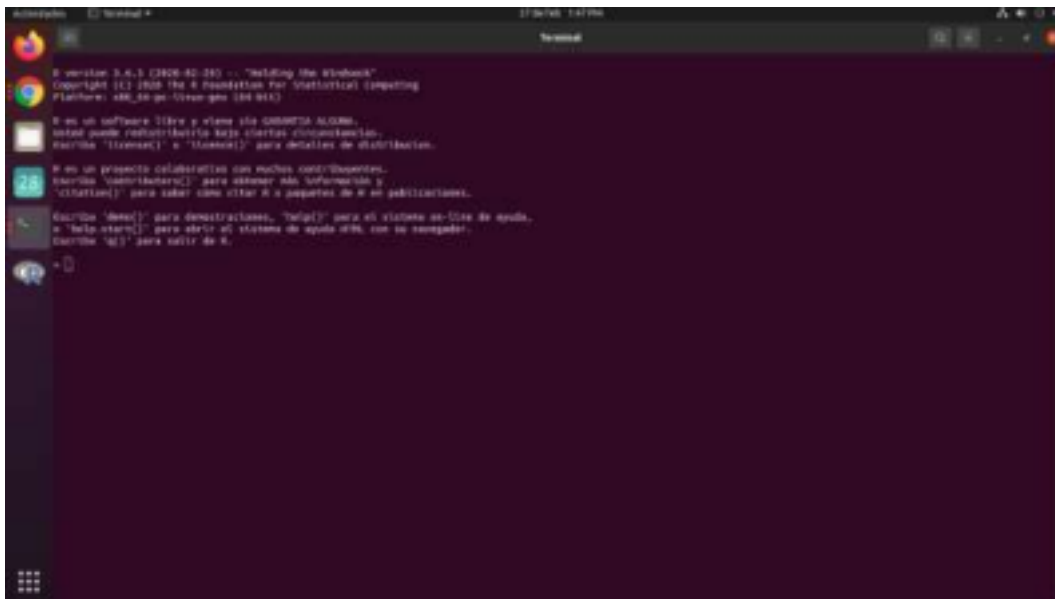
Edit the /.bashrc by adding the two commands called "export" and saving.

```
sparkuser@sparknode:~$ vi ~/.bashrc
# Add below lines at the end of the .bashrc file.
export SPARK_HOME=/home/sparkuser/spark
export PATH=$PATH:$SPARK_HOME/bin
```

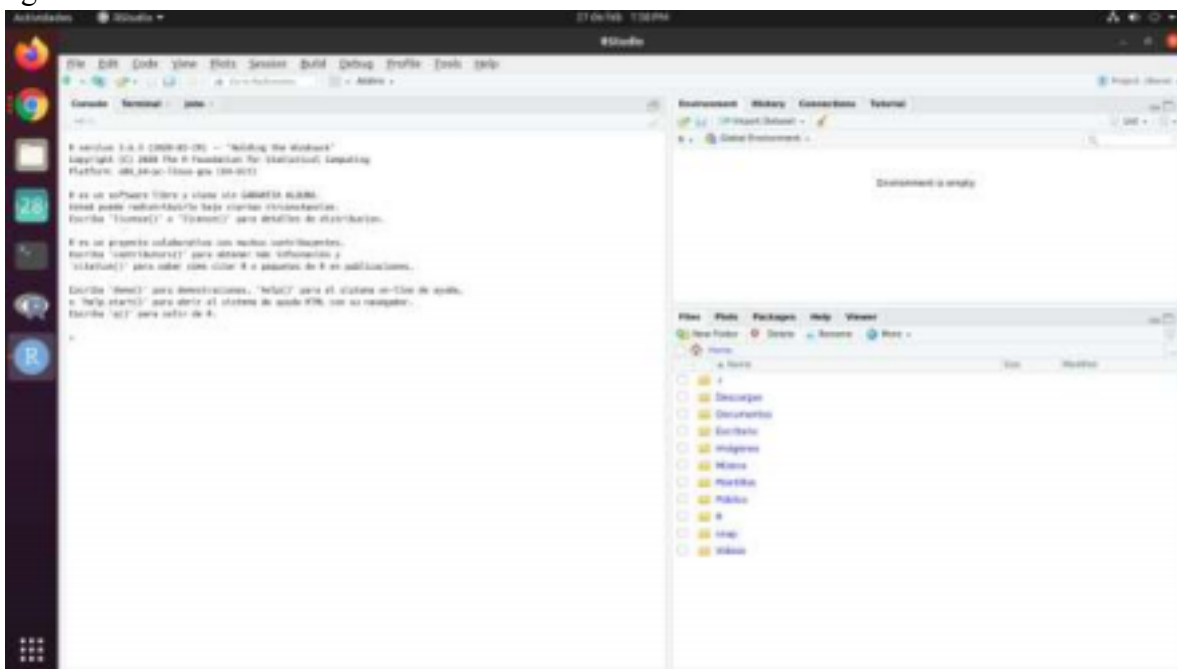
With the command "spark-shell" it is already possible to execute.



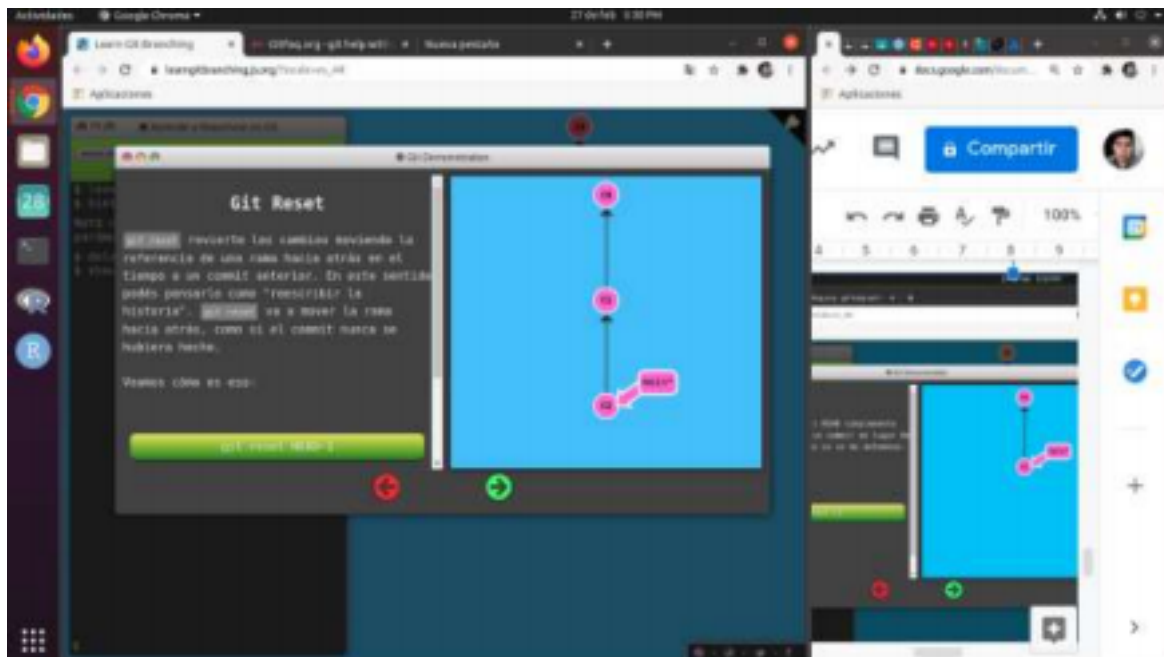
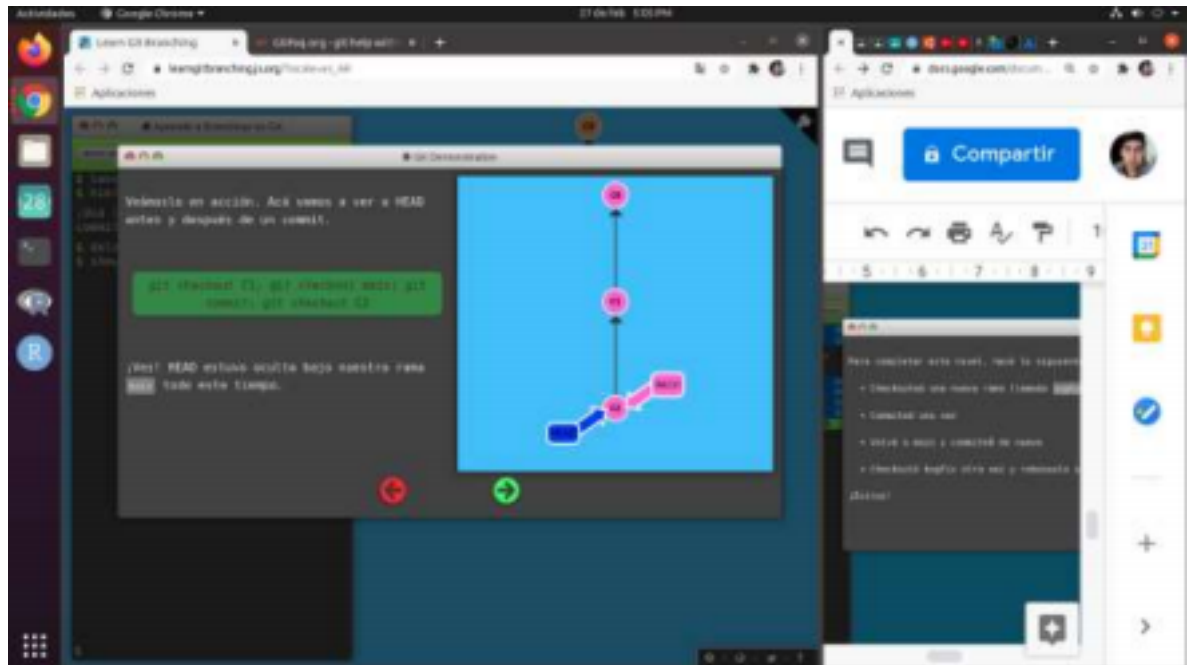
To install the “r” languages in Ubuntu we use the command “sudo apt install r-base” and to execute them we use the command “sudo -i R”.

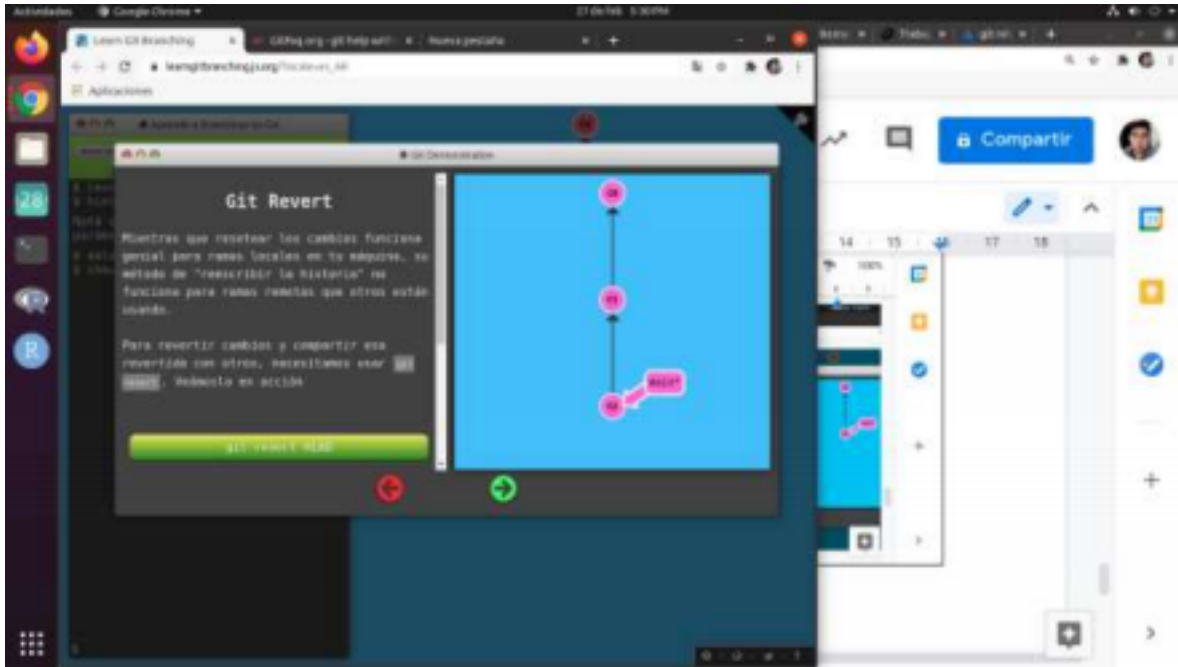


To install Rstudio, we do it with the Ubuntu installer, download the file ending in ".deb" and right click on install as software.



To know the basic git commands and how it works you can practice on the following page:
<https://learnitbranching.js.org/>





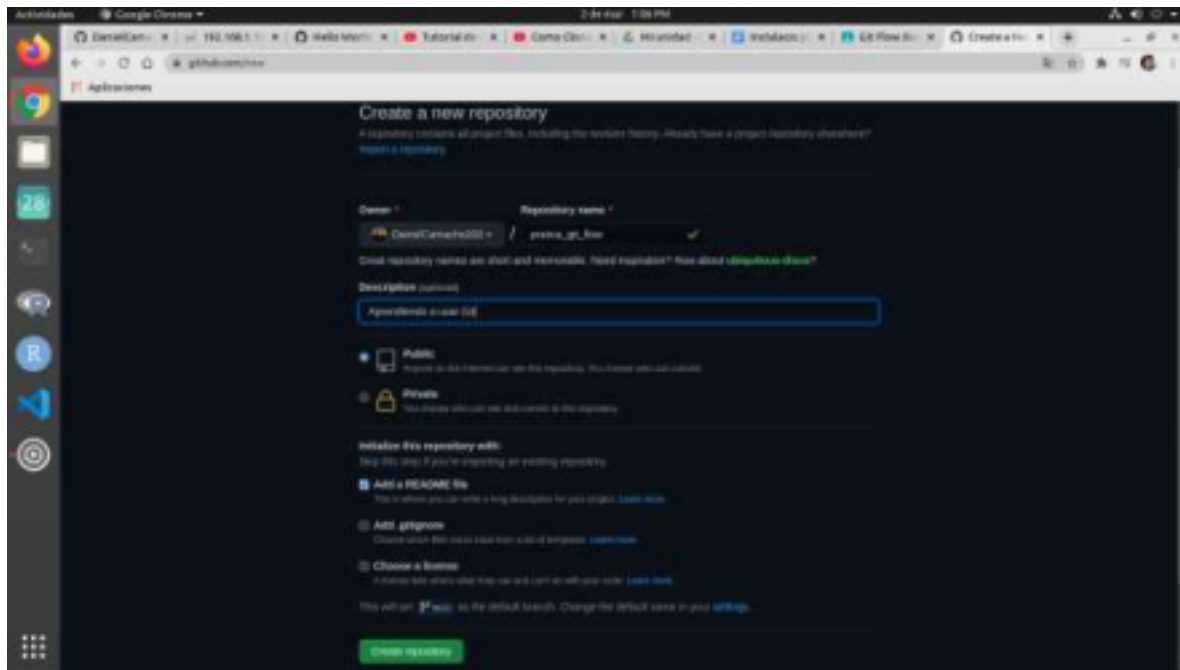
To configure SSH watch the following video:

<https://www.youtube.com/watch?v=G69dfwG2DJ4>

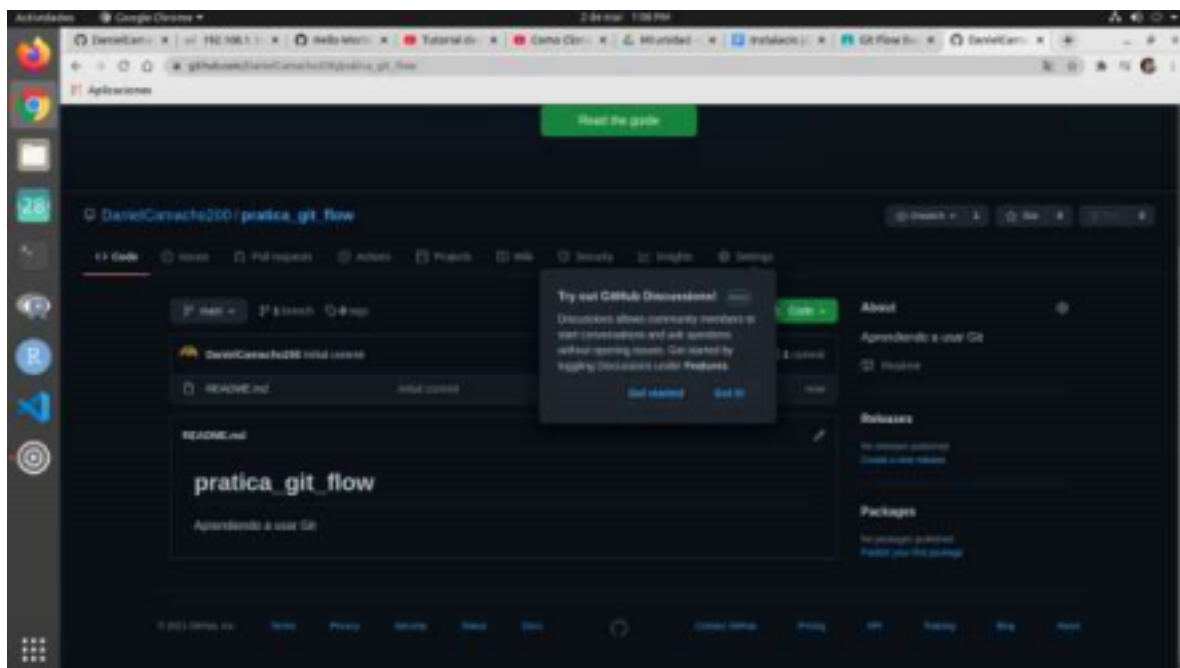
For some reason I also had to declare the Git user in the global settings.



Creating a new repository called "practice git flow". Mark as public and select to add a "README" by default.



It is now possible to clone the repository.



Clone the repository with "SSH"



We must have installed "git", we check the version with the command "git --version" We access the folder where we want to clone the repository, for this case I want it on the desktop.

We clone with the command "git clone" and the address copied from the repository".

We access the folder where the repository is.

We show the contents of the folder with the command "ls".



Since we want to know how many branches exist in the repository, we use the "git branch" command. We create the development branch We go inside, with the command "git checkout development"



We edit the file "README.md" and perform a "git commit".





We comment on the changes.
We do a "git push" to upload the changes.



In the same way we create a new branch, make changes to the "README" file, do a "commit" and do a "push"



We check that there were changes.



Bibliography

1. https://github.com/DanielCamacho200/practica_git_flow
2. https://www.youtube.com/watch?v=t8O7dMDD5nw&ab_channel=makigas%3Atutorialesdeprogramaci%C3%B3n

3. <https://www.youtube.com/watch?v=t8O7dMDD5nw>
4. <http://gitfaq.org/articles/what-is-a-detached-head.html>
5. <https://www.atlassian.com/es/git/tutorials/syncing/git-pull#:~:text=El%20comando%20git%20pull%20se,local%20para%20reflejar%20e se%20contenido.&text=El%20comando%20git%20pull%20es,fetch%20seguido%2 0de%20git%20merge%20.>
6. <https://www.hostinger.mx/tutoriales/comandos-de-git>
7. <https://qastack.mx/programming/1696433/how-to-exit-a-git-status-list-in-a-terminal#:~:text=Ctrl%20%2B%20q%20y%20c%20para%20salir%20de%20la%2 0situaci%C3%B3n%20de%20ejecuci%C3%B3n.&text=q%20o%20SHIFT%20%2 B%20q%20har%C3%A1,show%20HEAD%20%2C%20git%20diff%20etc.>
8. <https://pagure.io/git-branch/status/tree/#:~:text=By%20default%2C%20the%20git%20Dbranch,all%20local%2 0r%20remote%20branches.>
9. <https://git-scm.com/book/es/v2/Ramificaciones-en-Git-Gesti%C3%B3n-de-Ramas>
10. <https://www.youtube.com/watch?v=S1Mr0l87IpE>
11. https://www.youtube.com/watch?v=sZySm2Xh0vU&ab_channel=MitoCode
12. https://www.youtube.com/watch?v=GEcSKzHvAoY&ab_channel=FelixBlanco
13. https://www.youtube.com/watch?v=x0_LOimK7w8&ab_channel=Piensa3D
14. https://www.youtube.com/watch?v=84ishq1-Ybs&ab_channel=AnthonyCode
15. https://www.youtube.com/watch?v=j0U9jBmP3LM&ab_channel=makigas%3Atuto rialesdeprogramaci%C3%B3n
16. <https://rstudio.com/products/rstudio/download/#download>

17. https://learngitbranching.js.org/?locale=es_AR
18. <https://filisantillan.com/blog/como-copiar-un-commit-de-una-rama-a-otra-con-git/>
19. <https://stackoverflow.com/questions/10228760/how-do-i-fix-a-git-detached-head>
20. <https://git-scm.com/docs/git-rebase>
21. <https://git-scm.com/docs/git-checkout>
22. <https://www.youtube.com/watch?v=G69dfwG2DJ4>
23. <https://www.youtube.com/watch?v=HiXLkL42tMU>
24. <https://www.youtube.com/watch?v=3ni-jP2qEWg>
25. <https://www.esic.edu/rethink/tecnologia/apache-spark-introduccion-que-es-y-como funciona>
26. <https://computingforgeeks.com/how-to-install-apache-spark-on-ubuntu-debian/>
27. https://www.youtube.com/watch?v=T3_HdWxd9XM
28. https://www.youtube.com/watch?v=k6j5TMLm_uo
29. <https://www.youtube.com/watch?v=eQ0nPdfVfc0>
30. <https://www.youtube.com/watch?v=PGNfuitspw8>