**PROGRAMA INGENIERÍA DE SISTEMAS**

**GUIA DE LABORATORIO**

**NOMBRE DE LA PRÁCTICA:** Infraestructura distribuida con Hadoop y Spark

**OBJETIVO:** Crear una Infraestructura Big Data para aplicar conceptos de sistemas distribuidos entre Windows y Linux utilizando Hadoop y Spark

**OBJETIVOS ESPECÍFICOS:**

* Instalar Spark y Hadoop
* Crear un nodo master y un nodo esclavo
* Establecer comunicación master-esclavo entre nodos.

**TEMÁTICAS:**

**1.** Sistemas distribuidos

**2.** nodos

**3.** cluster

**COMPETENCIAS A DESARROLLAR:**

**1.** Conocimiento de distribución en Linux

**2.**  Creación y desarrollo de cluster

**3.** Configuración de nodos en sistemas distribuidos

**CONOCIMIENTOS PREVIOS REQUERIDOS:**

**1.**  Conceptos básicos sistemas distribuidos

**2.** Manejo de Linux

**RECURSOS A NIVEL HARDWARE Y SOFTWARE:**

**1.** Equipo pc con disco duro de mas de 250gb de capacidad

**2.**  Equipo pc con memoria ram 8gb o superior

**3.** Equipo pc con maquinas virtuales instaldas.

**DESCRIPCIÓN DE LA PRÁCTICA (PROCEDIMIENTOS, EJERCICIOS Y METODOLOGÍA):**

1. La practica inicia con la creación de dos maquinas virtuales, para este caso se utilizó el sistema operativo Linux – UBUNTU
2. A cada maquina virtual se le asigna una ip diferente.
3. Por medio de comandos se realiza la instalación de spark en las diferentes máquinas virtuales
4. Se realiza la configuración del spark en una maquina como nodo master y en las restantes se configuran los nodos como slaves.
5. Se realiza la conexión entre nodos, añadiéndose los nodos slaves al nodo master y creando asi el cluster distribuido.

**PROCEDIMIENTO O DESARROLLO INGENIERÍL (Agregar imágenes de los pantallazos de funcionalidad, numerarlos y explicar cada imagen:**

Descarga e instalación.

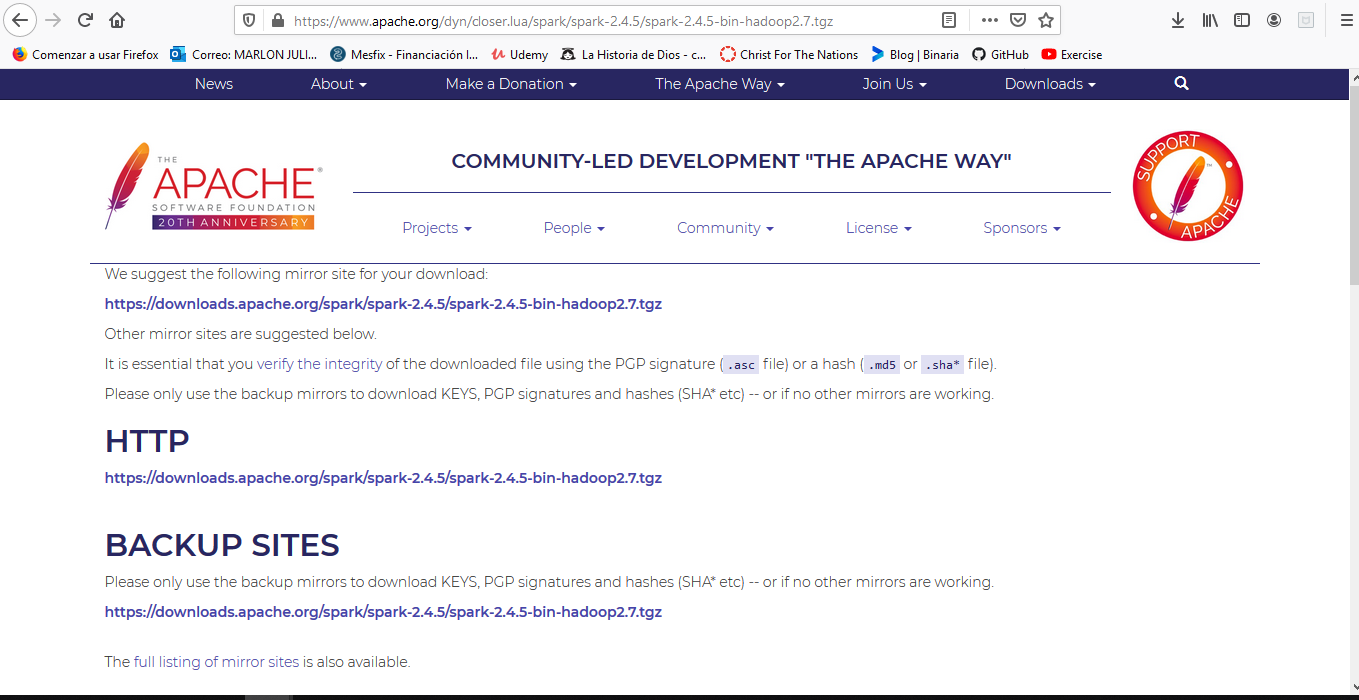
Se debe inicialmente descargar java para las máquinas virtuales

$ sudo apt-get install openjdk-8-jdk

Luego comprobamos la version

$ java -version

Descargamos spark de la pagina oficial



Utilizamos el comando

$ curl -O <https://downloads.apache.org/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz>

Descomprimimos e instalamos

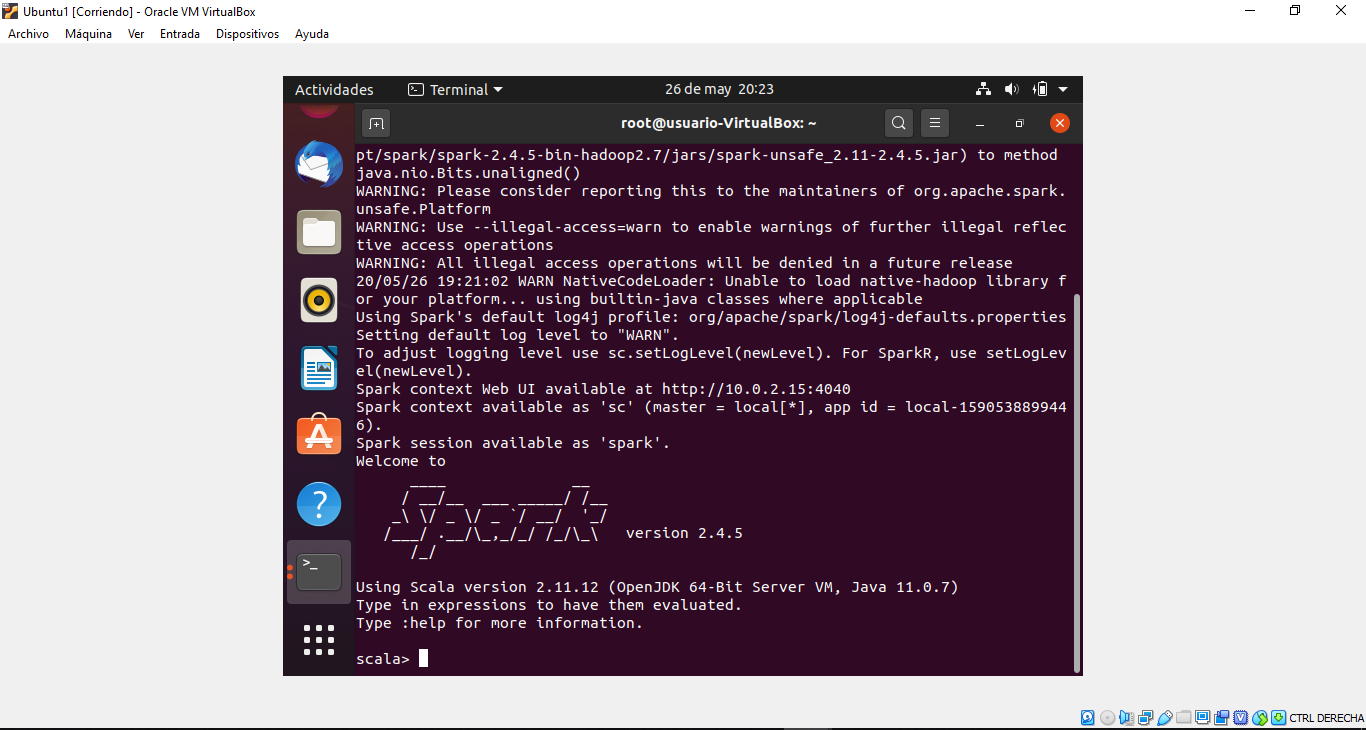
$ tar zxvf spark-2.2.0-bin-hadoop2.7.tgz

Modificamos la variable de entorno SPARK\_HOME

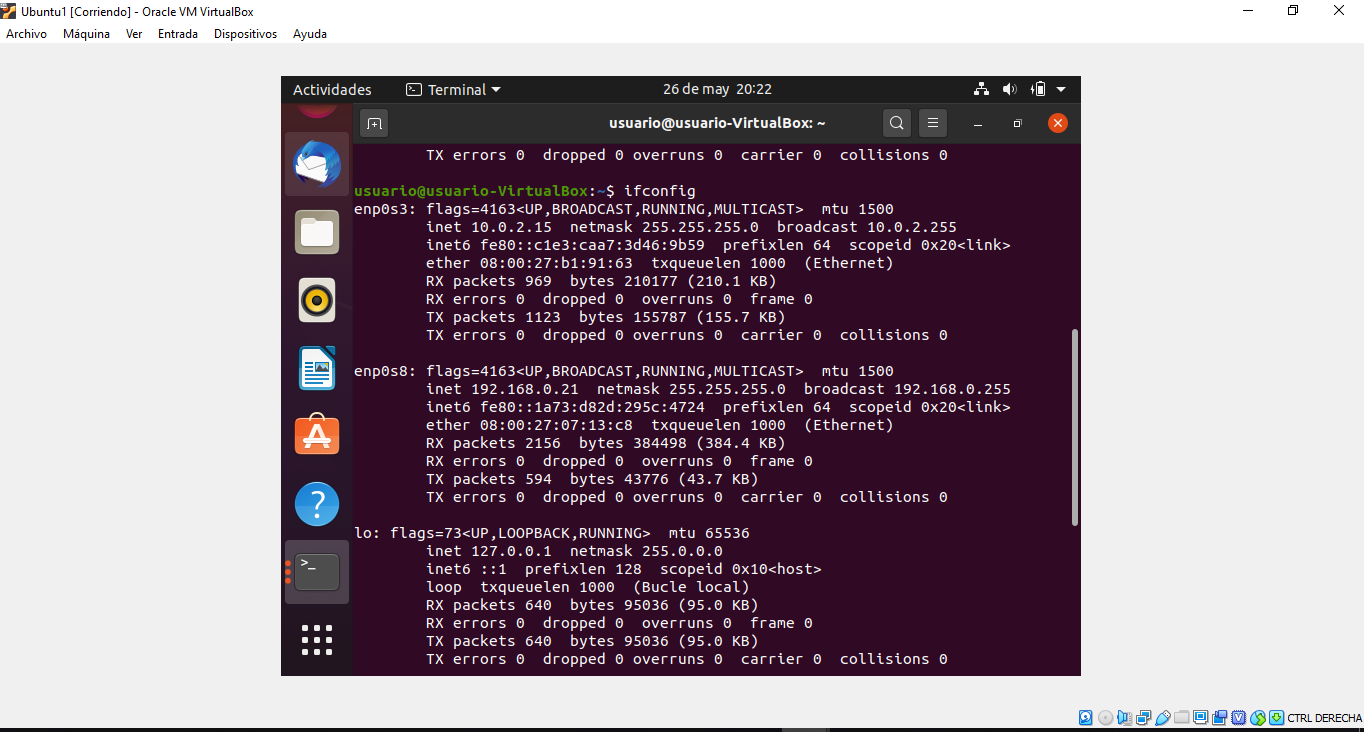
export HADOOP\_HOME="/home/<your\_username>/hadoop-2.8.0"

export LD\_LIBRARY\_PATH=$HADOOP\_HOME/lib/native:$LD\_LIBRARY\_PATH

E iniciamos spark

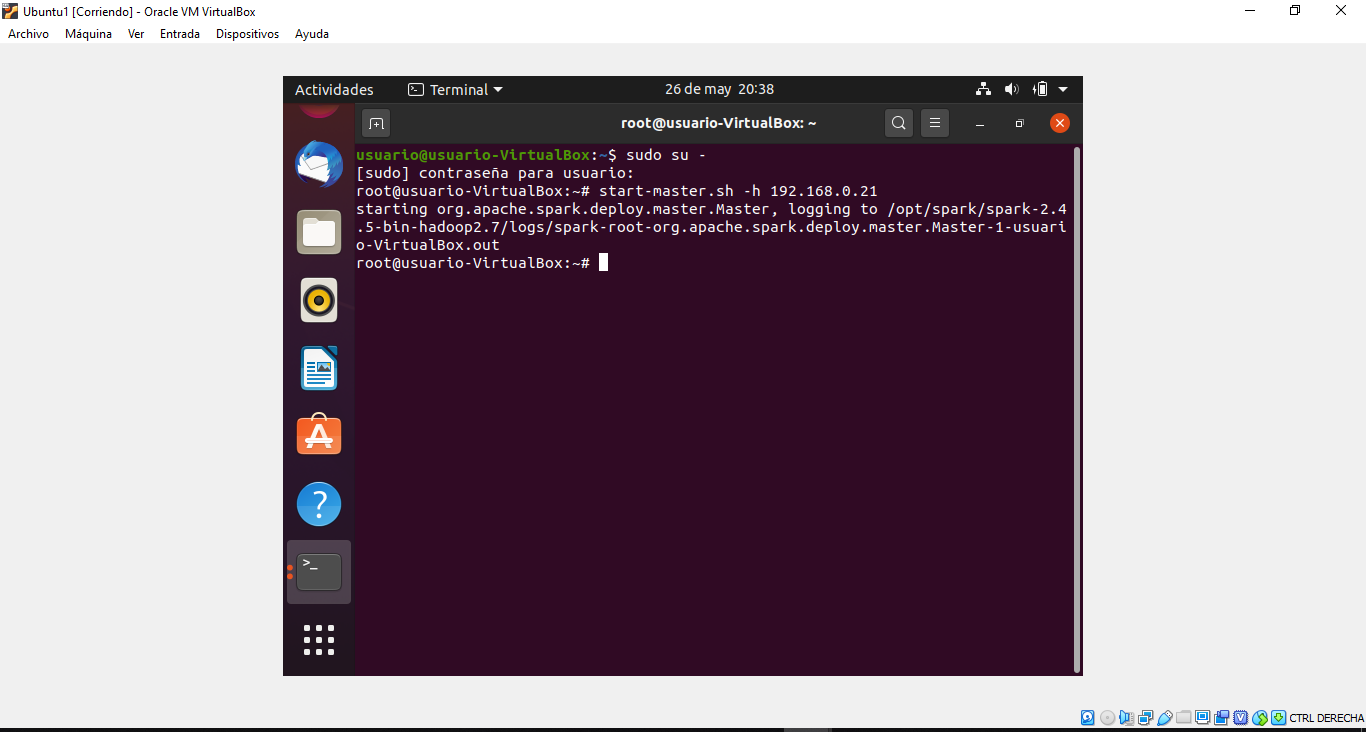


Verificamos la ip

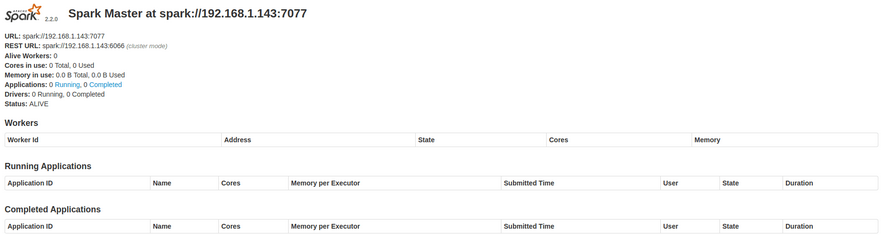


Inicamos como master en la primera maquina el nodo con el comando

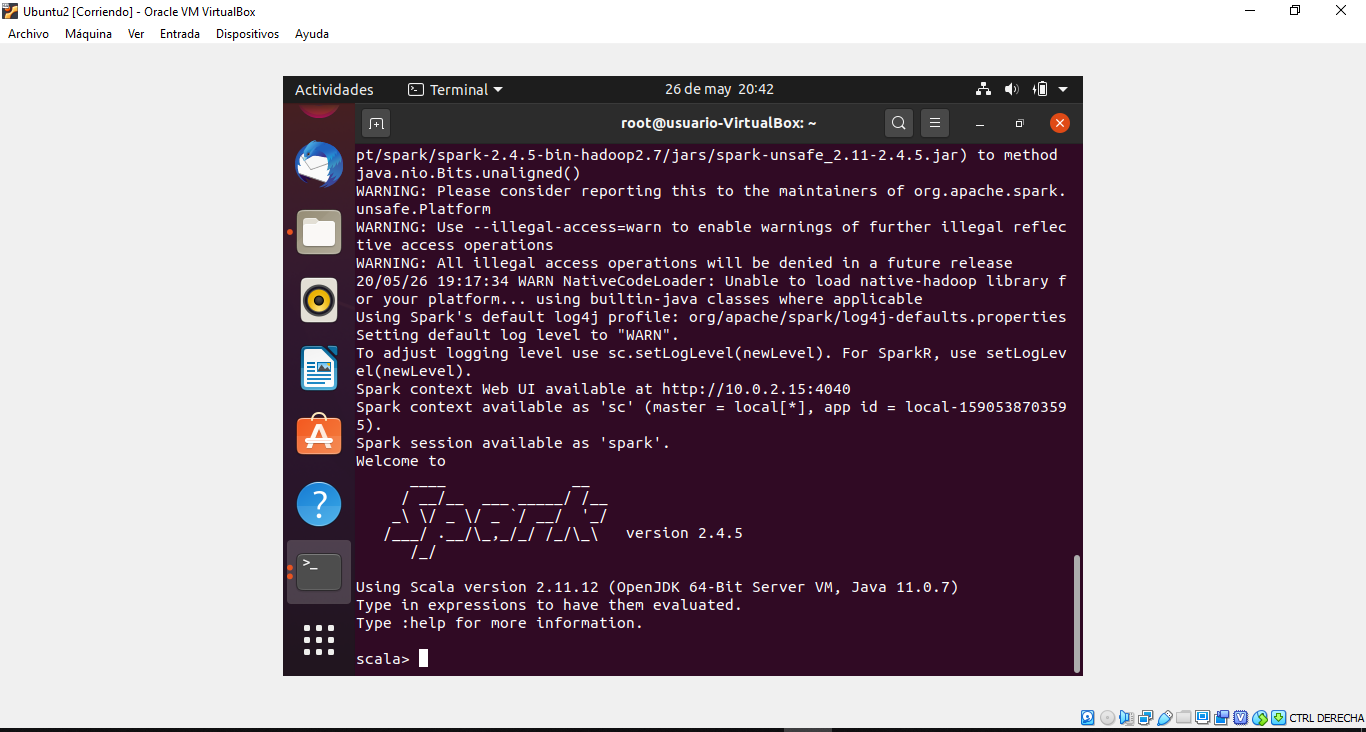
$ ./sbin/start-master.sh -h 192.168.0.21(Debes colocar tu ip)



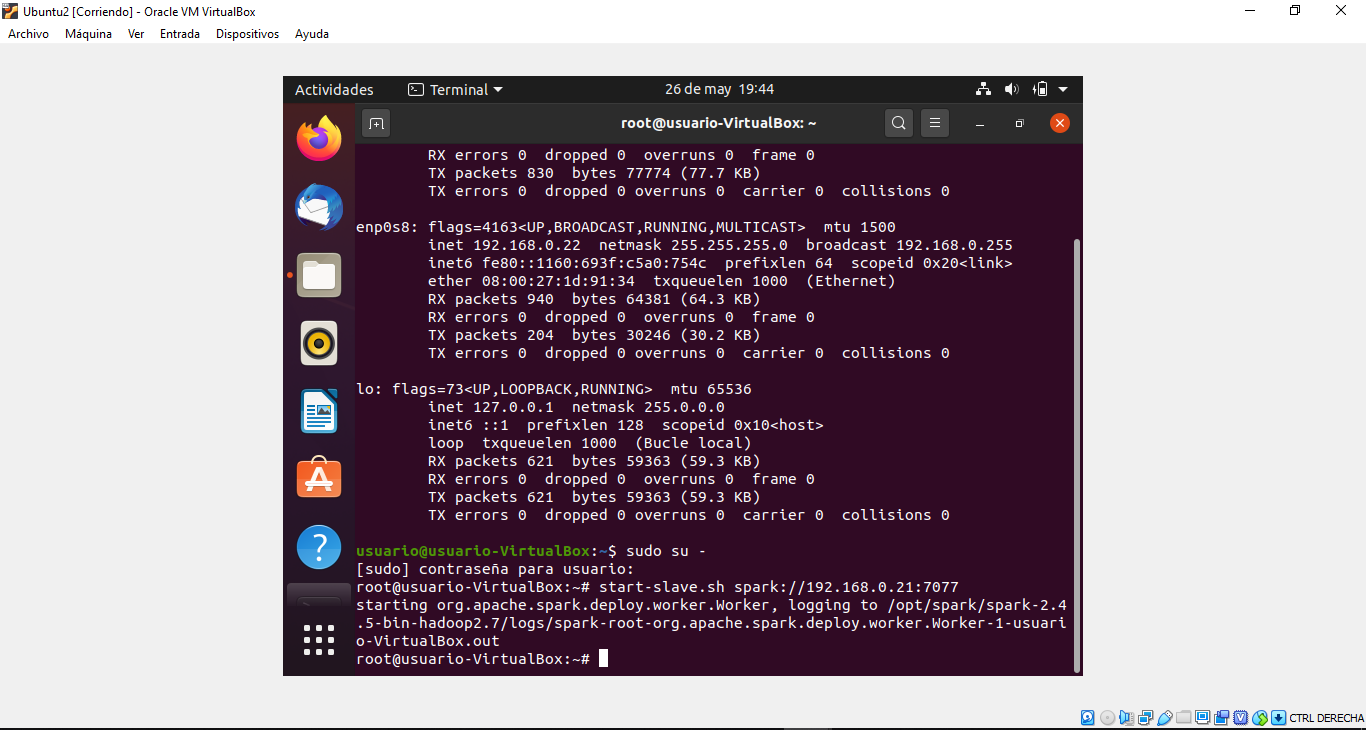
Y verificamos en el localhost que este iniciado.



Iniciamos en la maquina slave el spark

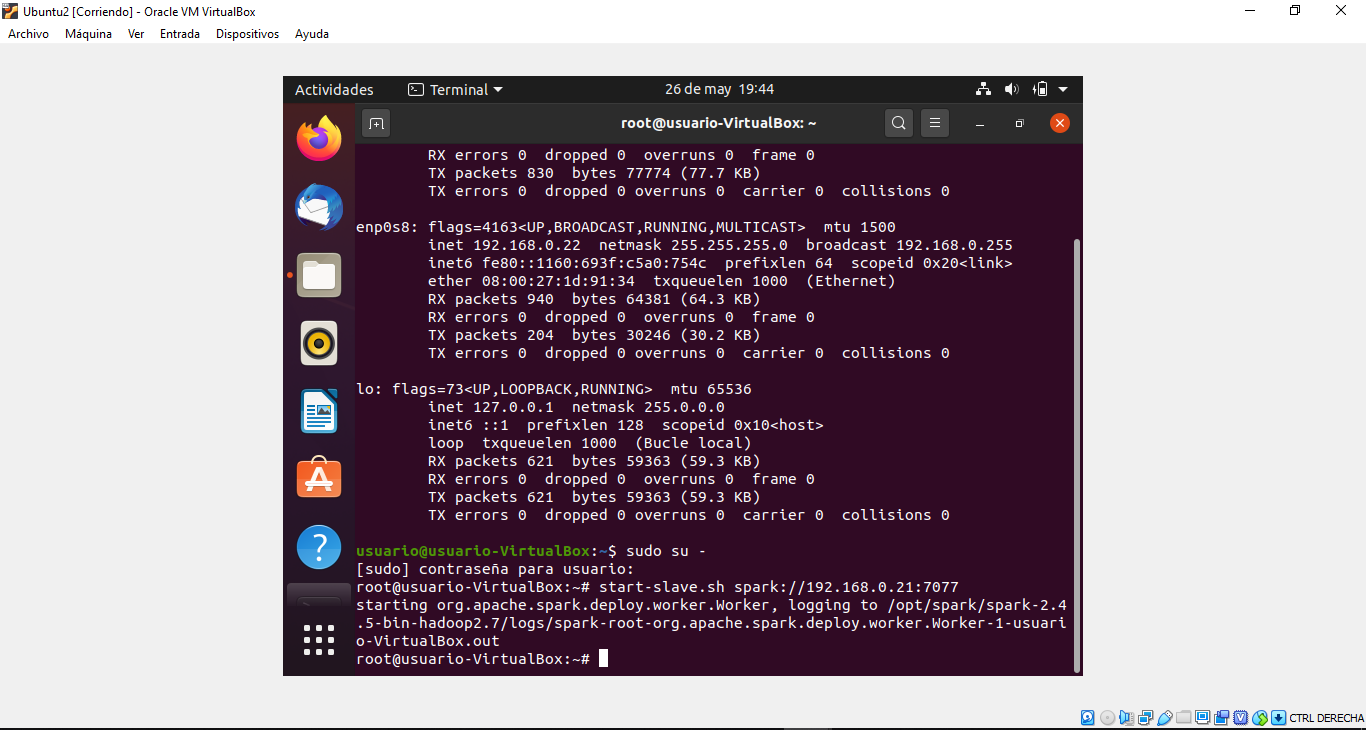


Verificamos la IP

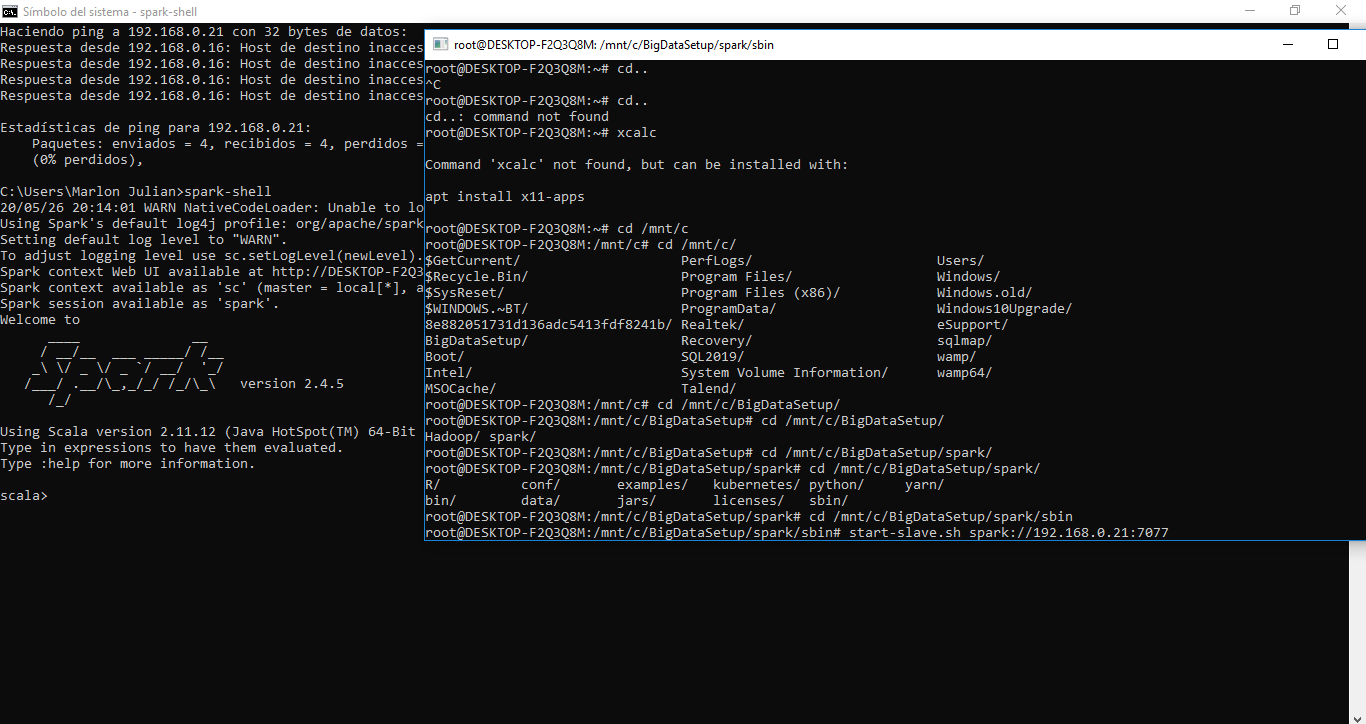


Realizamos la conexión entre el slave y el master

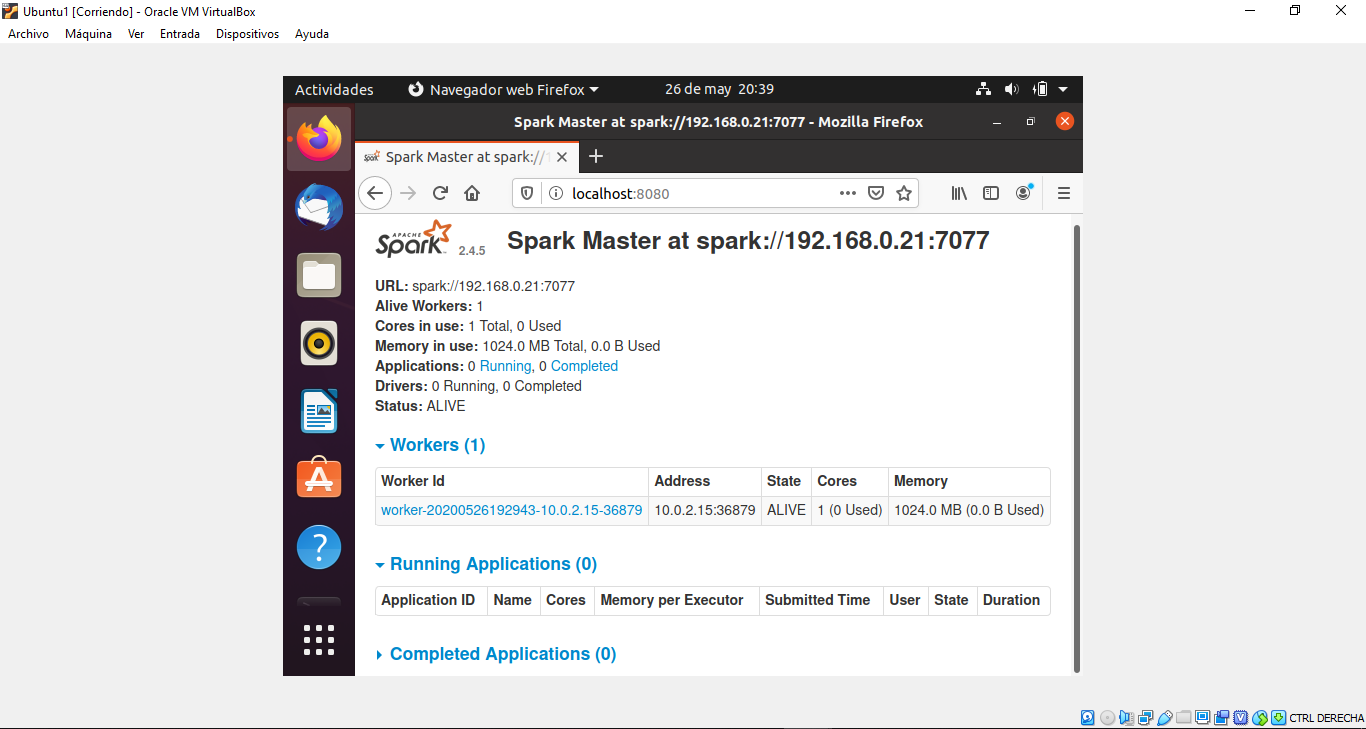
$ ./sbin/start-slave.sh http://192.168.0.21:7077



Y realizamos el mismo proceso para todos los demas slaves.

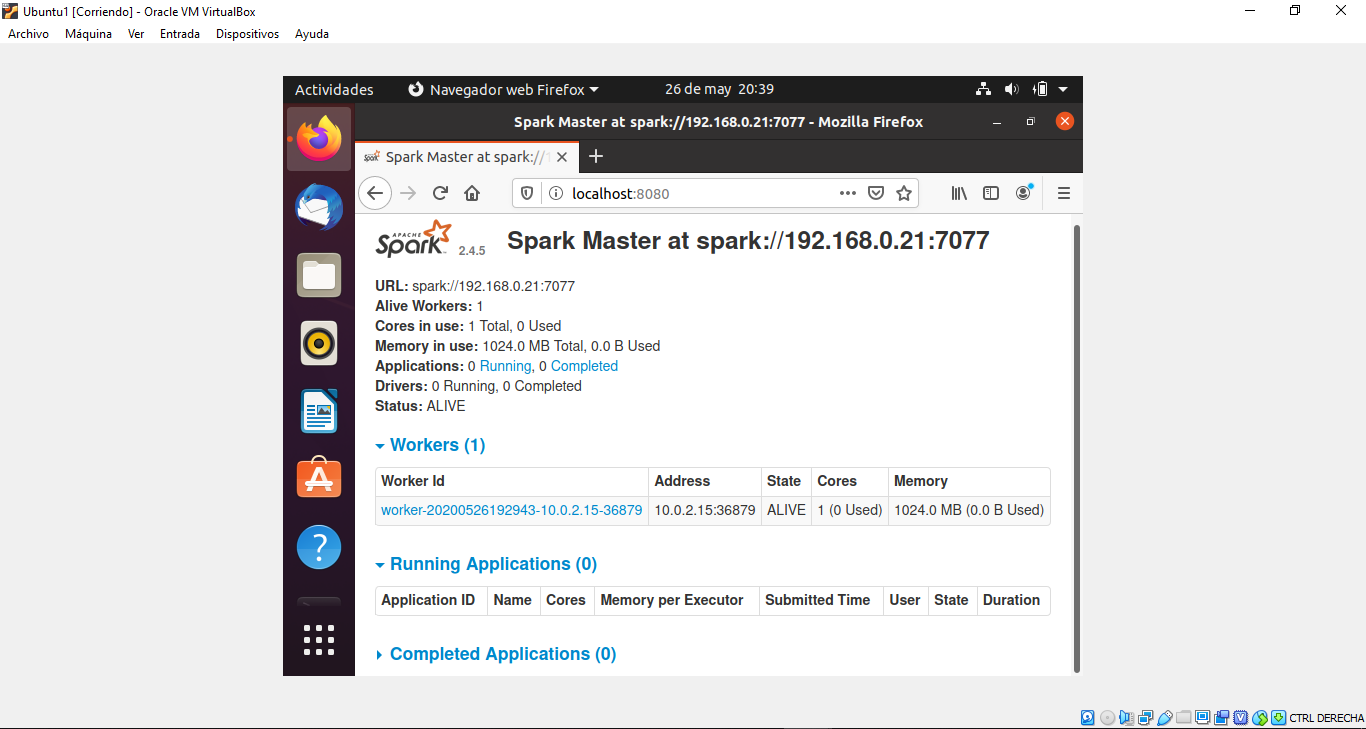


Al ingresar al local host podemos evidenciar como se van añadiendo las maquinas a este.



**PRODUCTOS Y RESULTADOS ESPERADOS:**

Como resultado se puede visualizar la conexión que se realiza la conexión y se genera un sistema distribuido entre maquinas master y slaves.



**CRITERIOS DE EVALUACIÓN:**

1. **Se tiene en cuenta el conocimiento de las definiciones, desarrollo de la práctica, análisis de los resultados, conclusiones y las habilidades, actitud, motivación para realizar la actividad de acuerdo a la rúbrica (Ver abajo).**
2. **Terminación del laboratorio**
3. **Cumplimiento de entrega**
4. **Cumplimiento de los objetivos del laboratorio**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **RÚBRICAS DE EVALUACIÓN** | | | | | |
| **CRITERIOS DE EVALUACION** | **4.5 - 5.0** | **4.0 - 4.5** | **3.0 - 3.5** | **2.5 - 3.0** | **CALIFICACION** |
| **1.- Definiciones** **Conceptos** | No es copia fiel de los textos consultados sino una síntesis de ideas completas y claras del tema. | Excede a cuatro cuartillas o no alcanza a cubrir una.  Algunos párrafos son copias fieles de los textos consultados.  Algunas ideas del tema están cortadas. | No se presenta de manera clara y completa. La relación con el problema planteado es prácticamente incongruente. El 60% del tema es copiado. | Realiza un 50% ó 60% de los experimentos, mencionando el procedimiento de manera completa incluyendo el material y equipo utilizado. Las respuestas son congruentes con los experimentos realizados. |  |
| **2.- Desarrollo de la Práctica - Dedicación** | Realiza todos los experimentos mencionando el procedimiento de manera completa, incluyendo el material y equipo utilizado. Las respuestas son congruentes con los experimentos realizados. | Realiza un 80% de los experimentos, mencionando el procedimiento de manera completa incluyendo el material y equipo utilizado. Las respuestas son congruentes con los experimentos realizados. | Realiza un 50% ó 60% de los experimentos, mencionando el procedimiento de manera completa incluyendo el material y equipo utilizado. Las respuestas son congruentes con los experimentos realizados. | Realiza un 40% de los experimentos, mencionando el procedimiento de manera completa incluyendo el material y equipo utilizado. Las respuestas son congruentes con los experimentos realizados. |  |
| **3.- Interpretación,**  **Análisis de los**  **Resultados** | Recopila y ordena los datos en relación al procedimiento. Se presentan los datos en tablas, gráficas, dibujos, etc. Claramente identificados. Los datos se interpretan y analizan comparativamente con la información bibliográfica consultada. | Presenta datos ordenados en relación al procedimiento. Se presentan en tablas, gráficas,  dibujos, etc. claramente  identificados, se interpretan y  analizan parcialmente en un  80%. | Tiene datos parcialmente ordenados, presenta algunas tablas o gráficas, los resultados  se interpretan y analizan en 50%  ó 60% | Tiene datos parcialmente ordenados, presenta algunas tablas o gráficas, los resultados se interpretan y analizan en 50%. |  |
| **4.- Conclusión** | Deduce el comportamiento de la(s) variable(s) estudiada(s) a partir del problema planteado. Rechaza o acepta la hipótesis e incluye propuestas de mejora o genera nuevos problemas. | Deduce el comportamiento de la(s) variable(s) estudiada(s) a partir del problema planteado.  Incluye el rechazo o la aceptación de la hipótesis, pero no las propuestas de mejoras. | Deduce el comportamiento de la(s) variable(s) estudiada(s) a partir del problema planteado.  No incluye el rechazo o aceptación de la hipótesis ni propone mejoras. | Deduce el comportamiento de la(s) variable(s) estudiada(s) a partir del problema planteado.  No incluye el rechazo o  aceptación de la hipótesis ni  Propone mejoras. |  |
| **5.- Actitud, Motivación, Disciplina** | Participa propositiva e integralmente en toda la práctica. | Participa ocasionalmente o lo hace constantemente pero sin coordinarse con su compañero. | Es un observador pasivo. | No participa en la realización de la práctica. |  |

**BIBLIOGRAFÍA LIBROS DIGITALES DE LA UNIVERSIDAD:**

**BIBLIOWEB:**

https://www.xataka.com/aplicaciones/asi-es-usar-la-consola-bash-de-ubuntu-en-windows-10

https://www.youtube.com/watch?v=eQ0nPdfVfc0

https://medium.com/beeranddiapers/installing-apache-spark-on-ubuntu-8796bfdd0861

https://thegurus.tech/how-to-spark-cluster/

**AUTOR:**