

G2 Virtualization

Planificación y Gestión de Infraestructuras TIC



ESCUELA SUPERIOR DE
INFORMÁTICA

15 MAYO

Máster en Ingeniería Informática

Creado por: **Daniel Ballesteros Almazán**

Luis María Campos de la Morena

María Teresa Merchán Quintana

Sergio Cámara Sevilla



Contenido

Intro	1
Arquitectura	1
Ejecución	2
Generación de imagen	2
Creación de red.....	2
Iniciar contenedores del cluster	2
Iniciar Hadoop	3
Ejecutar HLCA	3
Enlace Repositorio GITHUB	3
Enlace Youtube	3

Intro

En este documento se tratará el desarrollo planteado para la dockerización del ejercicio Hyperespectral Imagen realizado por el **equipo G2** en el entregable anterior, de esta forma, partiendo de la tecnología utilizada en el desarrollo del mismo, apache Pig se tratará de desplegar su ejecución en un cluster de Hadoop.

Arquitectura

Para el desarrollo del mismo se ha utilizado Docker con el cual se realiza el despliegue de 4 contenedores, formando un **clúster de Hadoop YARN**. El clúster se dispone de la siguiente forma:

- **Master**, contenedor/nodo, en el que se encuentra **ResourceManager** que reserva los recursos necesarios para administrar las aplicaciones ejecutadas en el sistema.
- **Workers**, 3 contenedores/nodos, en los que se encuentran los **Node Manager** que se encargan de gestionar los trabajos a partir de las instrucciones de del gestor de recursos del nodo maestro
- **Network**, se crea una red que hemos llamado Hadoop a través de la cual se conectan los nodos del cluster.

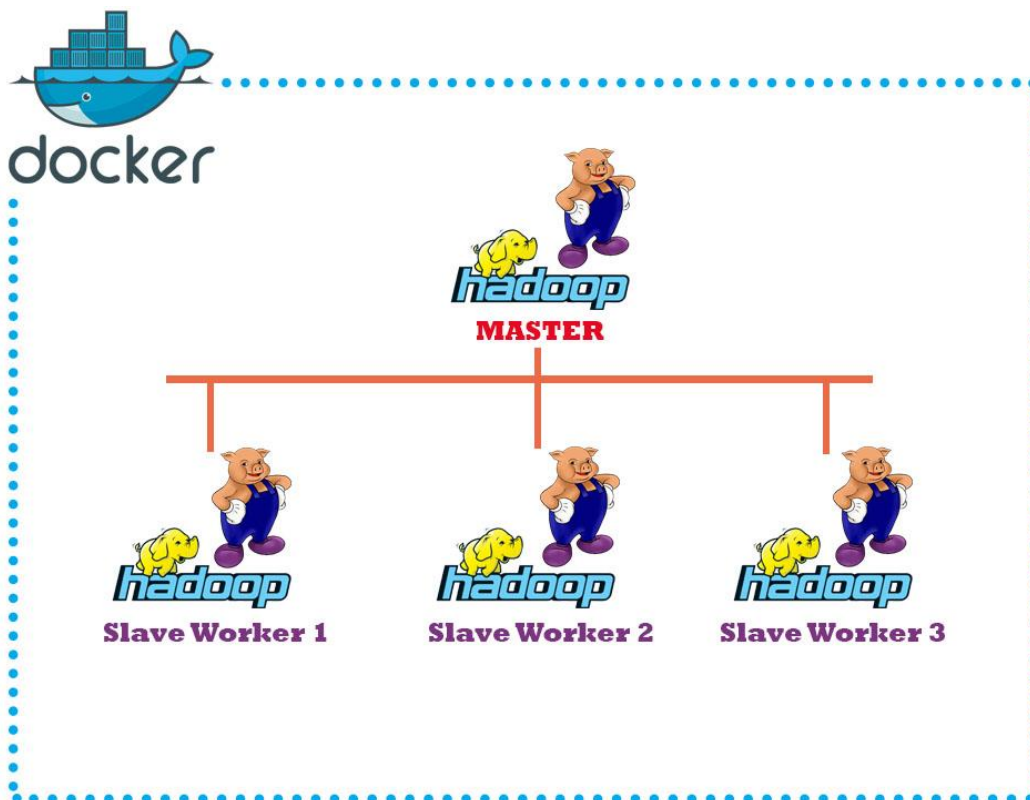


Figura 1 - Arquitectura desarrollada de virtualización de cluster hadoop + pig con Docker

Ejecución

Generación de imagen

Se ha configurado un archivo **Dockerfile** con una imagen de **Ubuntu** y los siguientes paquetes:

- OpenJDK
- OpenSSH-Server
- Hadoop
- Pig

Adicionalmente se incluye las siguientes configuraciones:

- Variables de entorno requeridas
- Generación de clave RSA para el inicio de sesión por **SSH**
- Creación del sistema **HDFS**
- Carga de los diferentes archivos de configuración, script y ficheros para su posterior uso.
- Generación de permisos de ejecución para los diferentes scripts
- Formateamos el sistema de archivos.
- Arrancamos el servicio SSH.

Para generar la imagen del archivo Dockerfile anteriormente especificado. Se debe ejecutar el siguiente script de Bash **build-image.sh** que contiene los comandos de Docker para la creación.

Creación de red

Para la creación de la red, ejecutamos el siguiente comando en el host, es muy importante que el nombre de la red sea **hadoop**, ya que los archivos de configuración están preparados para comunicarse a través de esta red.

```
sudo docker network create --driver=bridge hadoop
```

Figura 2 - comando de ejecución de creación de red clúster

Iniciar contenedores del cluster

Para arrancar los contenedores de Docker, se ha generado un script de Bash **start-container.sh**, en el que se crea un nodo denominado **hadoop-master** y 3 nodos **hadoop-slave** cada uno con su identificador. Y por último se entra dentro del contenedor de master

Iniciar Hadoop

Dentro del contenedor de master ejecutamos el script de Python **start-hadoop.sh**, en el que arrancaremos el **sistema de ficheros de Hadoop y Yarn**.

Ejecutar HLCA

Para ejecutar el algoritmo de HLCA, ejecutamos el script de Python **run-hlca.sh**, en este archivo se realizan las siguientes acciones:

- ejecutamos el script de **Python** que fue implementado en la anterior entrega para tratar la imagen
- Guardamos el archivo que contiene la imagen ya convertida en el sistema de fichero **HDFS**.
- Arrancamos **historyserver**
- **Ejecutamos el script de pig** con el algoritmo de HLCA

Enlace Repositorio GITHUB

El siguiente enlace contiene el código para la realización del trabajo.

https://github.com/sergiocamara/g2_virtualization

Figura 3 - Enlace GITHUB

Enlace Youtube

El siguiente enlace es un video explicativo de la ejecución:

<https://youtu.be/EmdKFAn84-A>

Figura 4 - Enlace Youtube