

# Configuración de un Edge Node Hadoop

Creado por Eder Villanueva, modificado por última vez en may 23, 2017

Un Edge Node o Cliente node basicamente es una interfaz entre el cluster hadoop y la red externa, tal cual este tipo de nodo: no pertenecen al cluster pero pueden ejecutar binarios o scripts que interactuan con el mismo, tal cual es importantes tener u nodo de este tipo por:

- No permite el acceso directo al cluster
- Interfaz con usuario de forma mas comoda
- Permite interactuar con otras aplicaciones

En esta pagina tomaremos un caso en particular en el cual tendra este nodo edge acceso a HDFS y capacidad de hacer submit de map/reduce operaciones.



La configuración de un Edge node tiene muy baja complejidad, a continuación se explicara a detalle que se necesita para configurarlo, en este punto las pruebas fueron con un cluster hadoop 2.7.3 sin seguridad (kerberos).

Archivos necesarios a modificar

Dentro de el directorio de `hadoop-2.7.3/etc/hadoop/`

- `core-site.xml`
- `yarn-site.xml`
- `mapred-site.xml`
- `hadoop-env.sh`

Variables de entorno de tu SO.

- `/etc/profile`
- `/etc/hosts`

## Configuración Hadoop

Editar archivo `core-site.xml`:

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://MASTERHADOOP:9000</value>
  </property>
```

```
<property>
  <name>fs.default.name</name>
  <value>hdfs://MASTERHADOOP:8020</value>
</property>
</configuration>
```

Editar archivo yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.resourcemanager.address</name>
    <value>MASTERHADOOP:8032</value>
  </property>
</configuration>
```

Editar archivo mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobtracker.address</name>
    <value>MASTERHADOOP:9001</value>
  </property>

</configuration>
```

Agregar o cambiar la variable JAVA\_HOME dentro de hadoop-env.sh

```
# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME='/location/to/your/path7/jdk8/'
```

Con esta configuración en tu hadoop tendras lo basico para usar tu edge node, ahora solo falta las variables de entorno para usar los scripts/binarios para acceder a tu cluster

## Variables de Entorno

Solo hay 2 archivos adicionales que podemos modificar para que tengamos completa la instalación:

Modificar archivo /etc/profile

```
JAVA_HOME='/location/to/your/path7/jdk8/'  
HADOOP_HOME='/home/path/to/hadoop-2.7.3'  
HADOOP_CONF_DIR="${HADOOP_HOME}/etc/hadoop"  
HADOOP_LIBEXEC_DIR="${HADOOP_HOME}/libexec"
```

```
PATH=$PATH:$JAVA_HOME/bin  
PATH=$PATH:$HADOOP_HOME/bin  
PATH=$PATH:$HADOOP_HOME/sbin
```

```
export JAVA_HOME  
export HADOOP_HOME  
export HADOOP_CONF_DIR  
export HADOOP_LIBEXEC_DIR
```

Agregar dentro de tu hosts las siguientes lineas en /etc/hosts:

```
TU-IP-DEL-MASTER-HADOOP    MASTERHADOOP
```

Con esta configuración es todo lo que necesitas para levantar tu nodo edge.