

Tipo : Guía de laboratorio
Capítulo : Arquitectura de Hadoop
Duración : 40 minutos

I. OBJETIVO

Ejecutando procesos de Map Reduce.

II. REQUISITOS

Los siguientes elementos de software son necesarios para la realización del laboratorio:

- Virtual Box
- Hortonworks Sandbox
- Putty
- Winscp

III. EJECUCIÓN DEL LABORATORIO

1. Ingresar al servidor Hortonworks, desde putty, con el usuario maria_dev.
2. Ejecutar el siguiente programa en Java que genera los procesos de MapReduce (/usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples-2.7.3.2.6.4.0-91.jar), el programa lee el contenido de los archivos que se encuentran en la carpeta /user/maria_dev/input, y devuelve un resultado en la carpeta /user/maria_dev/salida1, buscando las palabras que inician con el nombre de 'yarn[a-z.]+'
3. Crear el programa WordCount.java, el cual realiza la función de contar palabra de un directorio de archivos de texto, utilizando los procesos MapReduce.
4. Ir a a Hortonworks, a la opción YARN, seleccionar Quick Links, luego a la opción ResourceManager UI
5. Verifica que el programa ha ejecutado varios procesos en MapReduce, validar la opción para ver los procesos de tipo MapReduce.