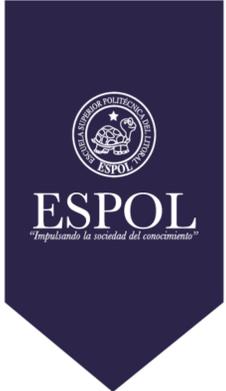


Sistemas de archivos por capas en Hadoop HDFS

En este proyecto, presentamos el desarrollo e implementación de un Sistema de Archivos por Capas automatizado, implementado en Hadoop File System. Este sistema se implementó para poder organizar eficientemente y automáticamente ficheros dentro de un sistema de archivos multicapas en HDFS, a cualquier usuario (ya sea para uso personal, como para uso en un sistema empresarial). El sistema trabaja de manera autónoma y es de uso relativamente sencillo para el usuario, para que pueda ser utilizado sin necesidad de conocer por completo el funcionamiento del sistema en cuestión.



Caviedes, L.

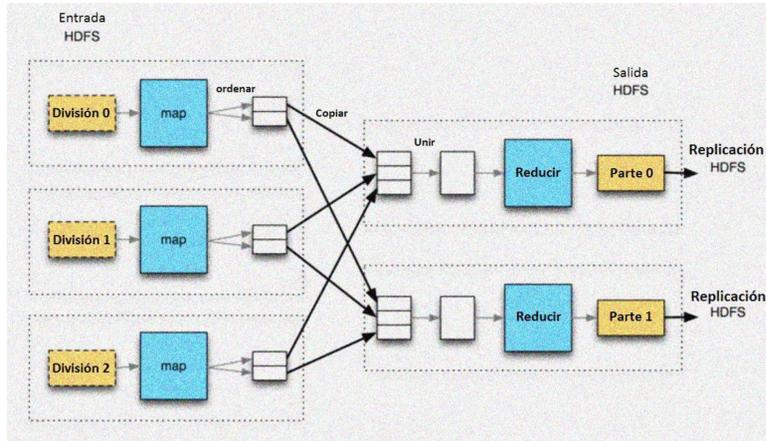
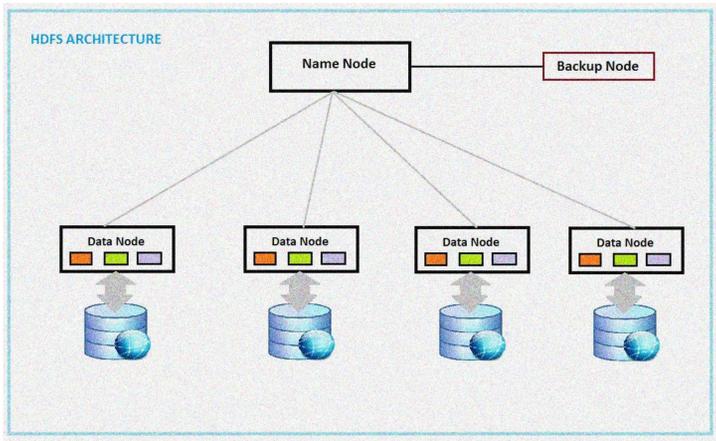
lcaviede@espol.edu.ec



Vasquez, L.

luifevas@espol.edu.ec

Introducción



Métodos y Motivación

Las tecnologías de Big Data están siendo adoptadas cada vez más en el mundo. Un elemento clave son los Sistemas de Archivos para grandes datos, los cuales son implementados de manera distribuida. Dichos sistemas de archivos proporcionan acceso rápido a los datos.

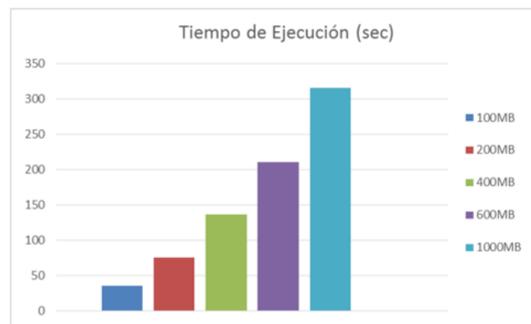
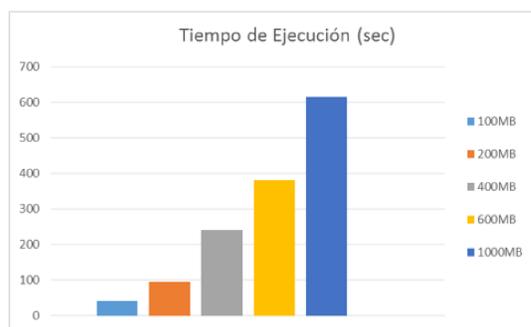
Podemos dividir el diseño de la solución en 2 partes: Primero, levantamiento de el entorno en el cual se va a ejecutar las pruebas y la implementación del programa. Esto se hará usando máquinas virtuales en Amazon Web Services

La segunda parte, es la implementación de la solución del proyecto en sí, la cual se la programará en Java, y en la cual no solo se utilizará conocimiento de programación en Java, si no conocimiento del HDFS como tal.

Resultados

Diseñamos las pruebas en base de uno de los Benchmarks de HDFS, DFSIO el cual esta incluido dentro de las librerías de Hadoop.

Luego de llevar a cabo varias pruebas, a pesar de las limitaciones del producto, se pudo obtener los siguientes resultados



Conclusiones y Referencias

Después de que las pruebas fueran llevadas a cabo, se llegó a la conclusión que si hubo una alza de rendimiento considerable al usar nuestro programa.

Sin embargo, se recomienda seguir ejecutando pruebas con diferentes parámetros.

En un futuro, al llegar a una conclusión más concreta, se desea llevar el proyecto a ser parte del HDFS.

Referencias

Abad, Cristina et al. "A Storage-Centric Analysis Of Mapreduce Workloads: File Popularity, Temporal Locality And Arrival Patterns". PH.D. University of Illinois at Urbana-Champaign, 2016.
Hadoop.apache.org,. "Hadoop – Apache Hadoop 2.7.2". Web. 11 Feb. 2016.