

Tecnológico de Costa Rica

Asignatura: Bases de Datos II

Profesor: Erick Hernández Bonilla

Proyecto: Administrador de Logs Hadoop

Alumnos:

Nelson Abarca Quirós 2013105521

Liza Chaves Carranza 2013016573

Melissa Molina Corrales 2013006074

Amanda Solano Astorga 2013100025

Fecha de Entrega: Martes 17 de Noviembre

Última fecha de modificación: 15 de Noviembre 2015

Arquitectura y Diseño

El Proyecto para administrar logs que contienen errores se basa en Hadoop, y su técnica de MapReduce para poder analizar grandes cantidades de datos. Para mayor facilidad al momento de desarrollar el Proyecto, se optó por utilizar la máquina virtual que provee Cloudera, que contiene los siguientes productos utilizados para el ambiente de desarrollo:

* Linux CentOs
* CDH 5.4
* Hadoop 2.6.1
* Eclipse Juno

La técnica de MapReduce consiste en tres partes, que siempre van a ser necesarias para el desarrollo del análisis, y son estas partes son:

Map: en el *mappeo,* entra una lista con elementos, que los transforma individualmente a un elemento con un formato específico de salida, este formato tendrá forma de key:value.

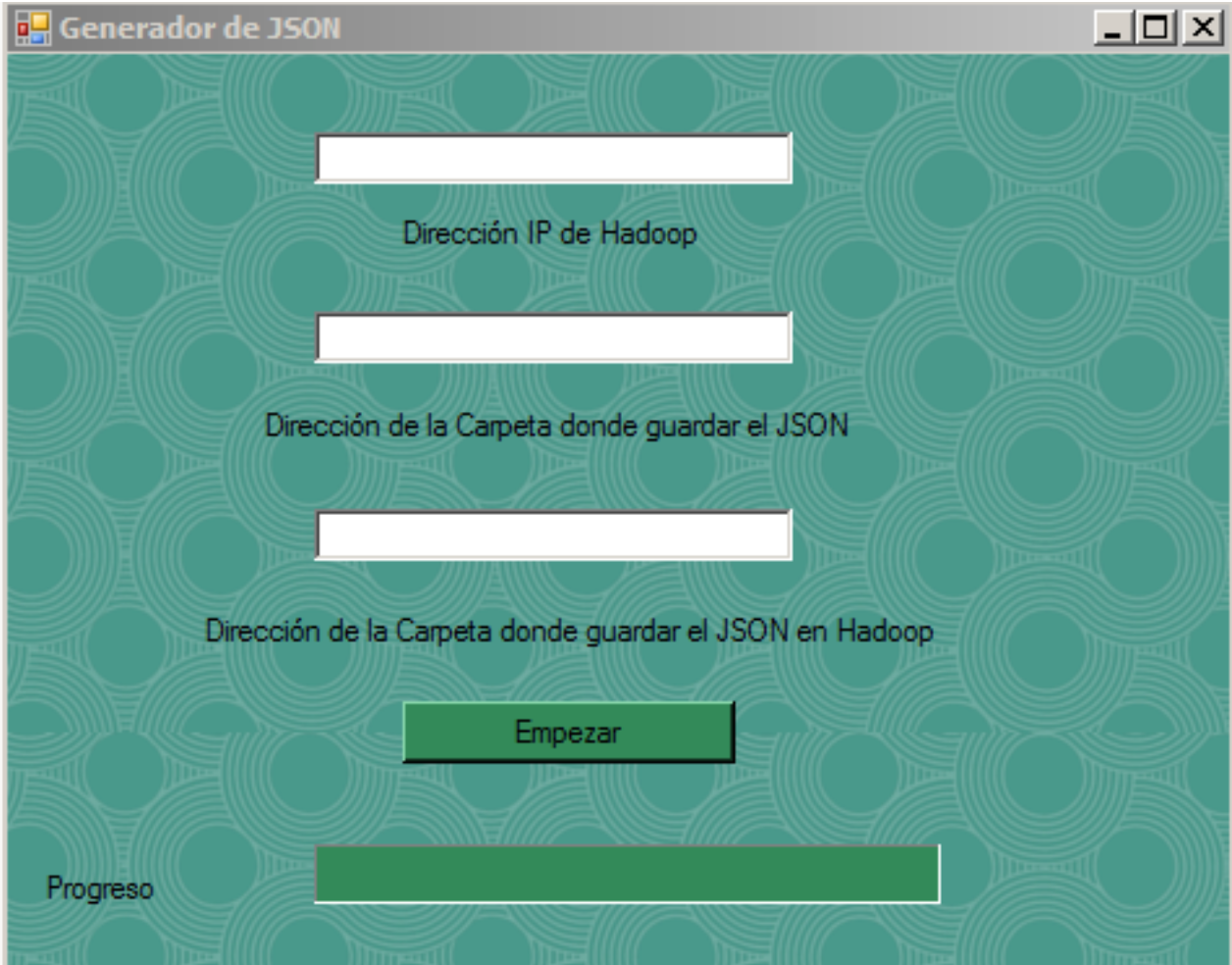
Reduce: en el *reducer,* se reciben los valores que salieron del mappeo, y son procesados de la forma que el desarollador decida. Se utiliza para minimizer la cantidad de datos que salen después de ser procesados, por ejemplo, la suma de los datos puede ser un proceso, y se retornará el resultado.

Driver: es el paso donde se ejecutan las funciones de map y reduce juntas. N cantidad de maps se pueden utilizar con n cantidad de reducer y viceversa.

Generador de Logs

El generador crea 500 000 logs de SQL Server y Oracle DataBase que son almacenados en un archivo JSON, cuyo nombre por defecto es *log.json*.

La interfaz es la siguiente:



Debido a que el programa realiza la transferencia de archivos mediante *Telnet* y *FTP*, se debe de indicar la dirección IP de la máquina donde se encuentra instalado Hadoop.

Para poder obtener dicho dato, debemos de ejecutar el siguiente comando en la terminal de Cloudera:

***ifconfig***

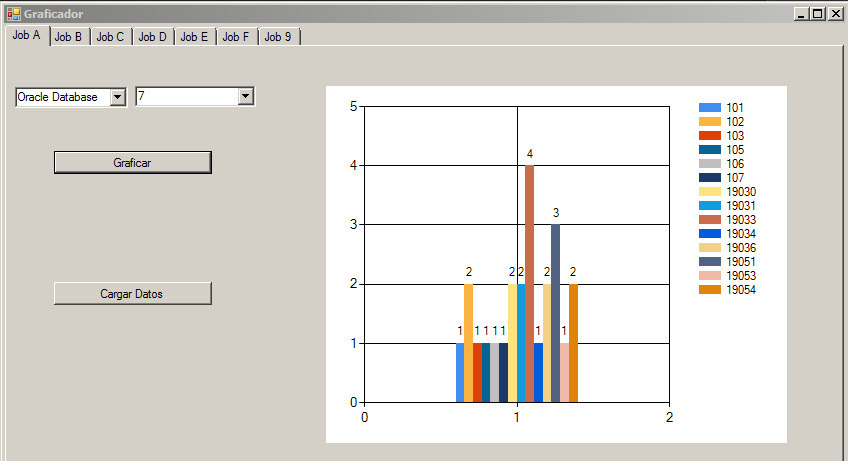
En la dirección del host, se debe de indicar la ruta donde se almacenará el archivo JSON. En el caso de Windows, debe ir con el siguiente formato *[Unidad]:\\* donde *[Unidad]* es el dispositivo donde se almacenará el archive, por ejemplo *C:\\*.

En la dirección de Hadoop se debe indicar la carpeta donde se desea insertar el archivo. De quedar en blanco el sistema insertará en la carpeta raíz del HDFS.

Se cuenta con una barra de progreso que mostrará la cantidad de logs generados. Cuando todo esté terminado se mostrará un mensaje de éxito al usuario, en caso de tener algún inconveniente también se informará.

Manual de Usuario

En el manual de usuario se presentarán imágenes de la interfaz gráfica con las explicaciones de su uso adecuado.

Ventana Principal

En la ventana principal se presentan menús para escoger las opciones específicas para cada job que se desean mostrar en el gráficador. Antes de presionar el botón “Graficar” es necesario cargar los datos, presionando el botón con este mismo nombre. A continuación se detallarán los menús que presenta cada pestaña, con cada job, respectivamente:

* **Job A**: producto y versión.
* **Job B**: producto, versión y mes.
* **Job C:** producto y versión.
* **Job D:** error y mes.
* **Job E:** producto y versión.
* **Job F:** product
* **Job G:**

Una característica que presenta este software es que todas las pestañas tienen la misma estructura, menús para escoger la información a graficar, los botones de cargar información y de graficar, razón por la cual su facilidad de uso aumenta.