

MapReduce, Pig, HCatalog and Oozie:

Una guía práctica

Luis F. Rivera
 Departamento Académico de Tecnologías de Información y Comunicaciones (TICs)
 Universidad Icesi
 Cali, Colombia
 Email: lfrivera@icesi.edu.co

Resumen

My abstract.

Keywords

map reduce, pig, hive, hcatalog, oozie.

I. INTRODUCTION

My introduction.

II. OBJETIVOS DEL PROYECTO

My objectives.

III. MAPREDUCE, PIG Y HIVE

En esta sección se presenta la comparación de los tiempos de ejecución de MapReduce, Apache Pig, y Apache Hive para el cálculo de la temperatura máxima registrada por año.

III-A. Ejecución con MapReduce

1) *Preparación:* Inicialmente, es necesario compilar el código fuente del programa *MaxTemperature* en su versión Java. Para hacer esto, es necesario clonar el repositorio del libro *Hadoop: The Definitive Guide*¹. Una vez hecho lo anterior, se debe proceder a compilar los archivos fuente necesarios mediante *Maven*. Finalmente, la ruta del archivo .jar compilado deberá establecerse en una variable de entorno llamada *HADOOP_CLASSPATH*.

```
1 //Clonación del repositorio.
2 git clone https://github.com/tomwhite/hadoop-book.git
3
4 // Compilación del código fuente.
5 mvn package -DskipTests
6
7 // Definición del classpath de Hadoop.
8 export HADOOP_CLASSPATH=/home/sas6/Oozie-Pig-HCatalog-Demos/assets/hadoop-examples.jar
```

2) Ejecución del programa *MaxTemperature*:

```
1 hadoop MaxTemperature /user/hive/warehouse/weather_external/full_data.txt out_mr_300GB
```

3) Seguimiento a la ejecución del programa:

III-B. Running Pig

III-C. Running Hive

IV. PIG AND HCATALOG

Pig and HCatalog.

¹Repositorio provisto por Tom White en <https://github.com/tomwhite/hadoop-book>

V. OOZIE

Oozie.

VI. RESULTS

Results.

VII. CONCLUSIONS

Conclusions.

VIII. FUTURE WORK

Future work.

REFERENCIAS

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.