

	<b>Training – Practice Guide 2</b>	
	Trainee:	Date:

## OBJETIVOS

- Entender el proceso de map reduce, utilizando código en Python

## GUIA

### 1. Ir a la carpeta de los libros:

```
cd /taller/libros
```

### 2. Crear carpeta en HDFS

```
hdfs dfs -mkdir /user/<tu_nombre>/libros
```

### 3. Subir archivo a HDFS

```
hdfs dfs -put libro* /user/<tu_nombre>/libros
```

### 4. Ir a la carpeta del código

```
cd /taller/mapreduce
```

### 5. Ejecutar el job MapReducer

```
hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-streaming-2.6*.jar -file mapper.py -mapper mapper.py -file reducer.py -reducer reducer.py -input /user/<tu_nombre>/libros -output /user/<tu_nombre>/libros_output
```

### 6. Revisar salida del proceso

```
hdfs dfs -ls /user/<tu_nombre>/libros_output
```

```
hdfs dfs -cat /user/<tu_nombre>/libros_output/<archivo> (elegir un archivo del listado anterior)
```

## REFERENCIA

[https://hadoop.apache.org/docs/r1.2.1/mapred\\_tutorial.html](https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html)