**NASA Web Hit Count in Pig**

data = load 'nasa.log';

b = foreach data generate FLATTEN(STRSPLIT($0,':',2));

d = foreach data generate SUBSTRING($0, LAST\_INDEX\_OF($0,'[')+1, LAST\_INDEX\_OF($0,'[')+12) as day;

Groups = GROUP d BY day;

Counts = FOREACH Groups GENERATE group, COUNT(d);

dump Counts;

**Map Reduce Chaining**

<http://sadrzyatarkika.blogspot.ca/2014/10/hadoop-chain-mapper-example.html>

**How to override HDFS block size while writing a file into HDFS:**

1. Command line

--Override the default block size with 256 MB

hadoop fs -D dfs.blocksize=268435456 -copyFromLocal <source> <target>

1. JAVA API

Configuration conf = new Configuration();

FileSystem fs = FileSystem.get(URI.create(dst), conf);

boolean overwrite = false;

int bufferSize = 65536; //64 KB

short replication = 1;

long blockSize = 51200; //50 KB

OutputStream out = fs.create(new Path(dst), overwrite, bufferSize,

replication, blockSize, new Progressable() {

public void progress() {

System.out.print(".");

}

})

**Check HDFS block locations:**

hdfs fsck /myfile.txt -files -blocks -locations

**HDFS REST API:**

<https://hadoop.apache.org/docs/r1.0.4/webhdfs.html#Document+Conventions>

**YARN Tuning**

<https://www.cloudera.com/documentation/enterprise/5-8-x/topics/cdh_ig_yarn_tuning.html>