

Hadoop Distributed File System

Francesco Pugliese, PhD

Italian National Institute of Statistics, Division "Information and Application Architecture", Directorate for methodology and statistical design

Email Francesco Pugliese : francesco.pugliese@istat.it

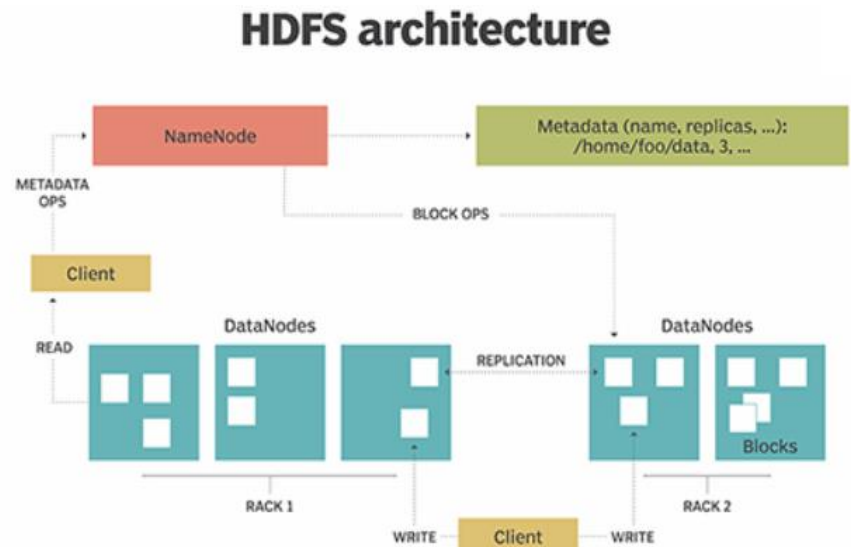
Vamsi Gunturi

Data Scientist and Researcher Intern at Italian National Institute of Statistics (Istat)

Email Vamsi Gunturi :

What is Hadoop Distributed File System (HDFS) ?

- The **Hadoop Distributed File System (HDFS)** is the primary data storage system used by Hadoop applications. It employs a **NameNode** and **DataNode** architecture to implement a distributed file system that provides high-performance access to data across highly scalable Hadoop clusters.



HDFS Commands

- **HDFS** is the primary or major component of the Hadoop ecosystem which is responsible for storing large data sets of structured or unstructured data across various nodes and thereby maintaining the metadata in the form of log files.

```
sbin/start-all.sh
```

→ To start all the services of Hadoop

```
jps
```

```
suraj@suraj:~/hadoop-2.5.0-cdh5.3.2$ jps
2546 SecondaryNameNode
2404 DataNode
2295 NameNode
2760 ResourceManager
2874 NodeManager
4251 Jps
suraj@suraj:~/hadoop-2.5.0-cdh5.3.2$
```

- ***bin/hdfs dfs -ls <path>*** - list all the files
- ***bin/hdfs dfs -mkdir <folder name>*** - create a directory

HDFS Commands

- To create a directory. In Hadoop **dfs** there is no home directory by default. So let's first create it.

```
bin/hdfs dfs -mkdir <folder name>
```

```
creating home directory:
```

```
hdfs/bin -mkdir /user
```

```
hdfs/bin -mkdir /user/username -> write the username of your computer
```

Example:

```
bin/hdfs dfs -mkdir /geeks => '/' means absolute path
```

```
bin/hdfs dfs -mkdir geeks2 => Relative path -> the folder will be  
created relative to the home directory.
```

HDFS Commands

bin/hdfs dfs

```
Usage: hadoop fs [generic options]
[-appendToFile <localsrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]
[-copyFromLocal [-f] [-p] <localsrc> ... <dst>]
[-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-count [-q] <path> ...]
[-cp [-f] [-p | -p[topax]] <src> ... <dst>]
[-createSnapshot <snapshotDir> [<snapshotName>]]
[-deleteSnapshot <snapshotDir> <snapshotName>]
[-df [-h] [<path> ...]]
[-du [-s] [-h] <path> ...]
[-expunge]
[-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-getfacl [-R] <path>]
[-getfattr [-R] {-n name | -d} [-e en] <path>]
[-getmerge [-nl] <src> <localdst>]
[-help [cmd ...]]
[-ls [-d] [-h] [-R] [<path> ...]]
[-mkdir [-p] <path> ...]
[-moveFromLocal <localsrc> ... <dst>]
[-moveToLocal <src> <localdst>]
[-mv <src> ... <dst>]
[-put [-f] [-p] <localsrc> ... <dst>]
[-renameSnapshot <snapshotDir> <oldName> <newName>]
[-rm [-f] [-r|-R] [-skipTrash] <src> ...]
[-rmdir [--ignore-fail-on-non-empty] <dir> ...]
[-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>][[--set <acl_spec> <path>]]
[-setfattr {-n name [-v value] | -x name} <path>]
[-setrep [-R] [-w] <rep> <path> ...]
[-stat [format] <path> ...]
[-tail [-f] <file>]
[-test [-d] [-d] [-s] <path>]
[-text [-ignoreCrc] <src> ...]
```

REFERENCES

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

.

AKNOWLEDGEMENTS



THANK YOU
FOR YOUR ATTENTION
Francesco Pugliese
Vamsi Gunturi