

# Hadoop Lab - CSE 328

## Assignment - 1

20 Jan 2022

Gyanendra Shukla  
191112040  
CSE 1

I've installed hadoop on WSL by following the steps. The screenshots below show the installation -

```
(sok@languor)~/hadoop-3.3.1
$ bin/hadoop
Usage: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
or      hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
       where CLASSNAME is a user-provided Java class

  OPTIONS is none or any of:

buildpaths          attempt to add class files from build tree
--config dir        Hadoop config directory
--debug             turn on shell script debug mode
--help              usage information
hostnames list[,of,host,names] hosts to use in slave mode
hosts filename       list of hosts to use in slave mode
loglevel level       set the log4j level for this command
workers             turn on worker mode

  SUBCOMMAND is one of:

  Admin Commands:

daemonlog           get/set the log level for each daemon

  Client Commands:

archive             create a Hadoop archive
checknative          check native Hadoop and compression libraries availability
classpath            prints the class path needed to get the Hadoop jar and the required libraries
confest             validate configuration XML files
```

Here, I'm using hadoop to find an input pattern. This shows that it is working properly.

```
(sok@languor)~/hadoop-3.3.1
$ mkdir input
(sok@languor)~/hadoop-3.3.1
$ cp etc/hadoop/*.xml input
(sok@languor)~/hadoop-3.3.1
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep input output 'dfs
2022-01-14 15:34:31,910 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-01-14 15:34:32,379 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s)
2022-01-14 15:34:32,380 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-01-14 15:34:32,718 INFO input.FileInputFormat: Total input files to process : 10
2022-01-14 15:34:32,739 INFO mapreduce.JobSubmitter: number of splits:10
2022-01-14 15:34:33,146 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1463104196_
2022-01-14 15:34:33,146 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-01-14 15:34:33,323 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-01-14 15:34:33,327 INFO mapreduce.Job: Running job: job_local1463104196_0001
2022-01-14 15:34:33,334 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-01-14 15:34:33,346 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
2022-01-14 15:34:33,346 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary
t directory:false, ignore cleanup failures: false
2022-01-14 15:34:33,348 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.l
Committer
2022-01-14 15:34:33,506 INFO mapred.LocalJobRunner: Waiting for map tasks
2022-01-14 15:34:33,507 INFO mapred.LocalJobRunner: Starting task: attempt_local1463104196_0001_m_00
2022-01-14 15:34:33,541 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
2022-01-14 15:34:33,542 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary
```