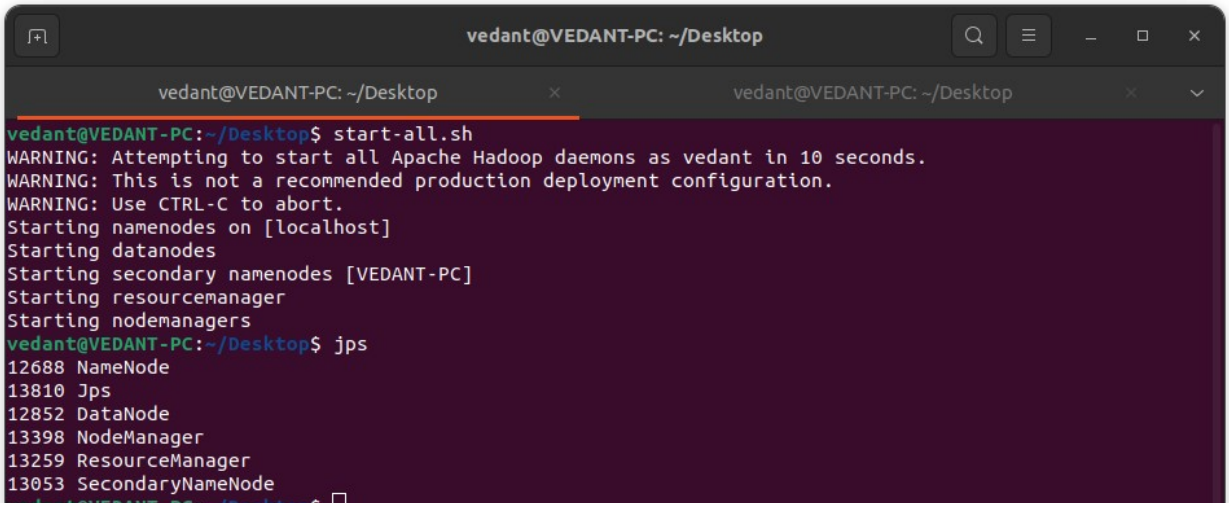
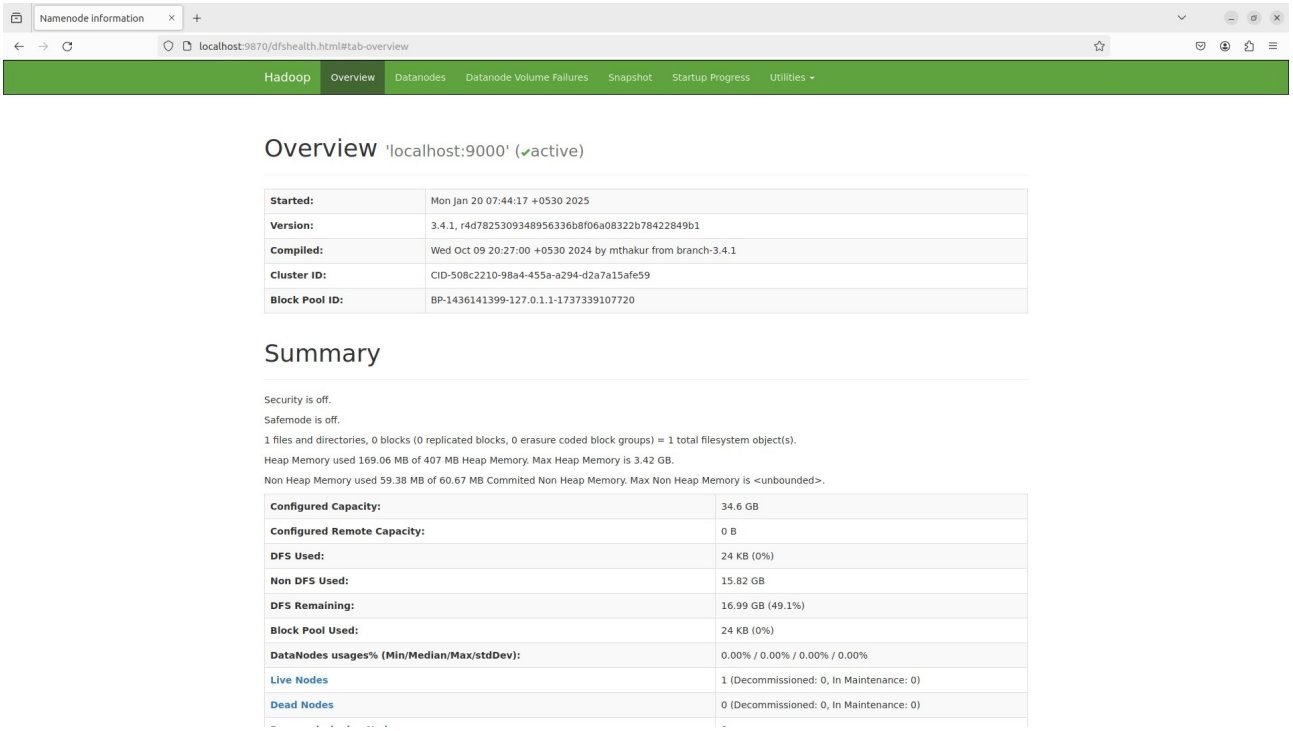
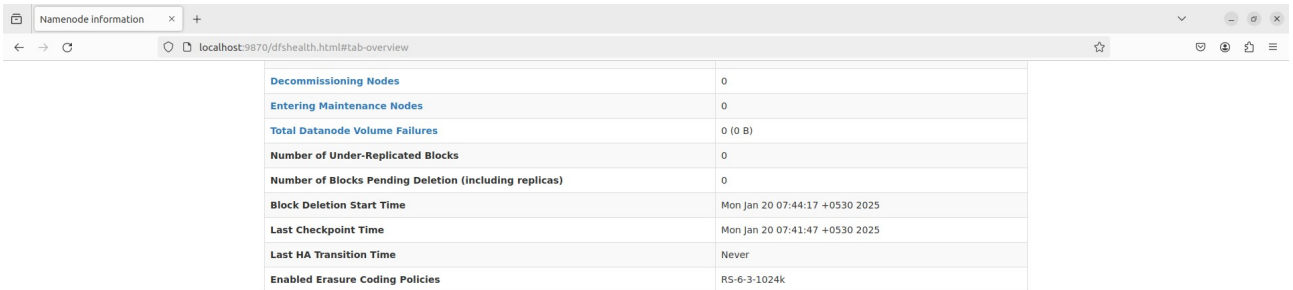
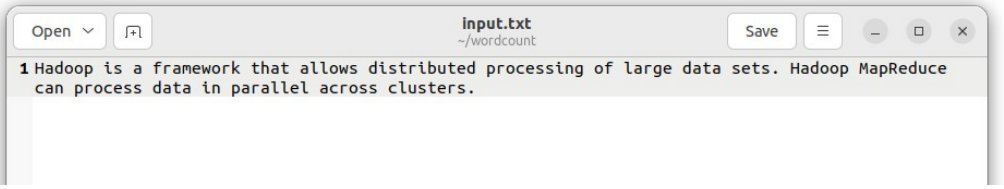

vedant@VEDANT-PC: ~/Desktop
vedant@VEDANT-PC: ~/Desktop
vedant@VEDANT-PC: ~/Desktop
vedant@VEDANT-PC:~/Desktop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as vedant in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [VEDANT-PC]
Starting resourcemanager
Starting nodemanagers
vedant@VEDANT-PC:~/Desktop$ jps
12688 NameNode
13810 Jps
12852 DataNode
13398 NodeManager
13259 ResourceManager
13053 SecondaryNameNode

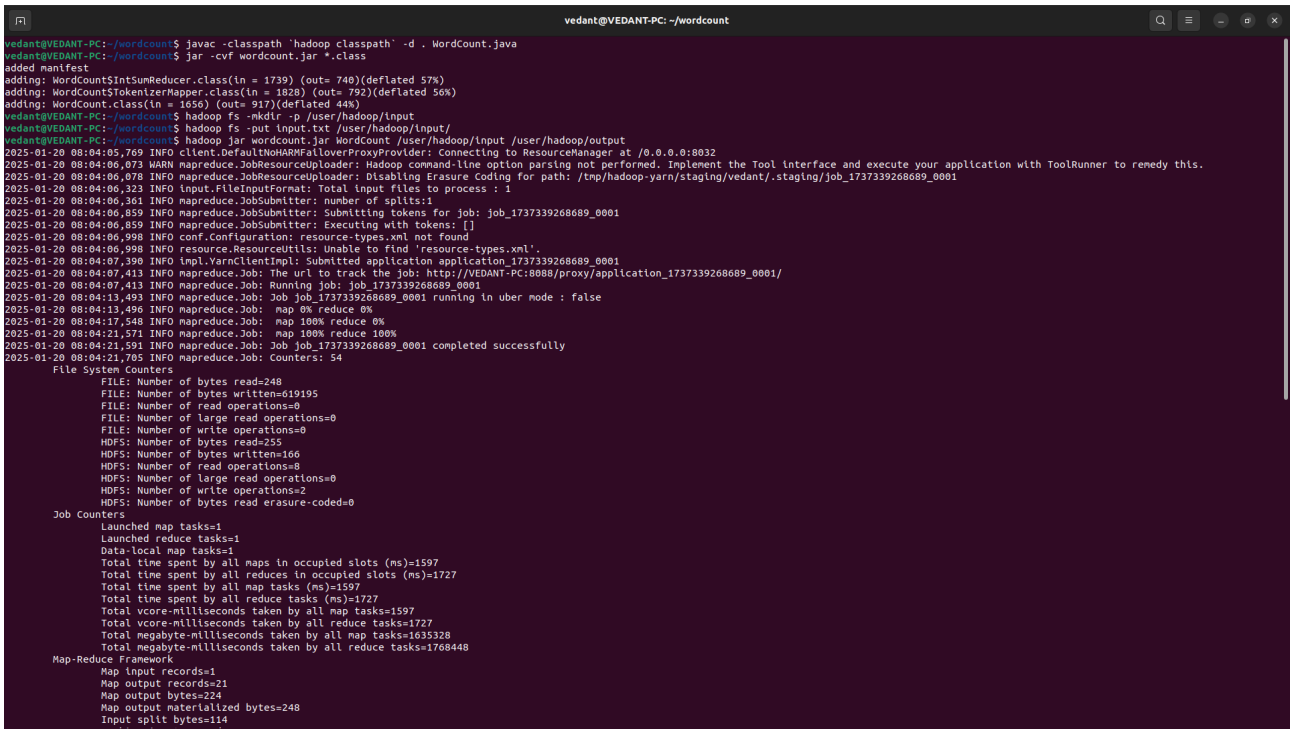

Namenode information
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:9000' (active)
Started: Mon Jan 20 07:44:17 +0530 2025
Version: 3.4.1, r4d7825309348956336b8f06a08322b78422849b1
Compiled: Wed Oct 09 20:27:00 +0530 2024 by mthakur from branch-3.4.1
Cluster ID: CID-508c2210-98a4-455a-a294-d2a7a15afe59
Block Pool ID: BP-1436141399-127.0.1.1-1737339107720
Summary
Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 169.06 MB of 407 MB Heap Memory. Max Heap Memory is 3.42 GB.
Non Heap Memory used 59.38 MB of 60.67 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.
Configured Capacity: 34.6 GB
Configured Remote Capacity: 0 B
DFS Used: 24 KB (0%)
Non DFS Used: 15.82 GB
DFS Remaining: 16.99 GB (49.1%)
Block Pool Used: 24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev): 0.00% / 0.00% / 0.00% / 0.00%
Live Nodes 1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes 0 (Decommissioned: 0, In Maintenance: 0)

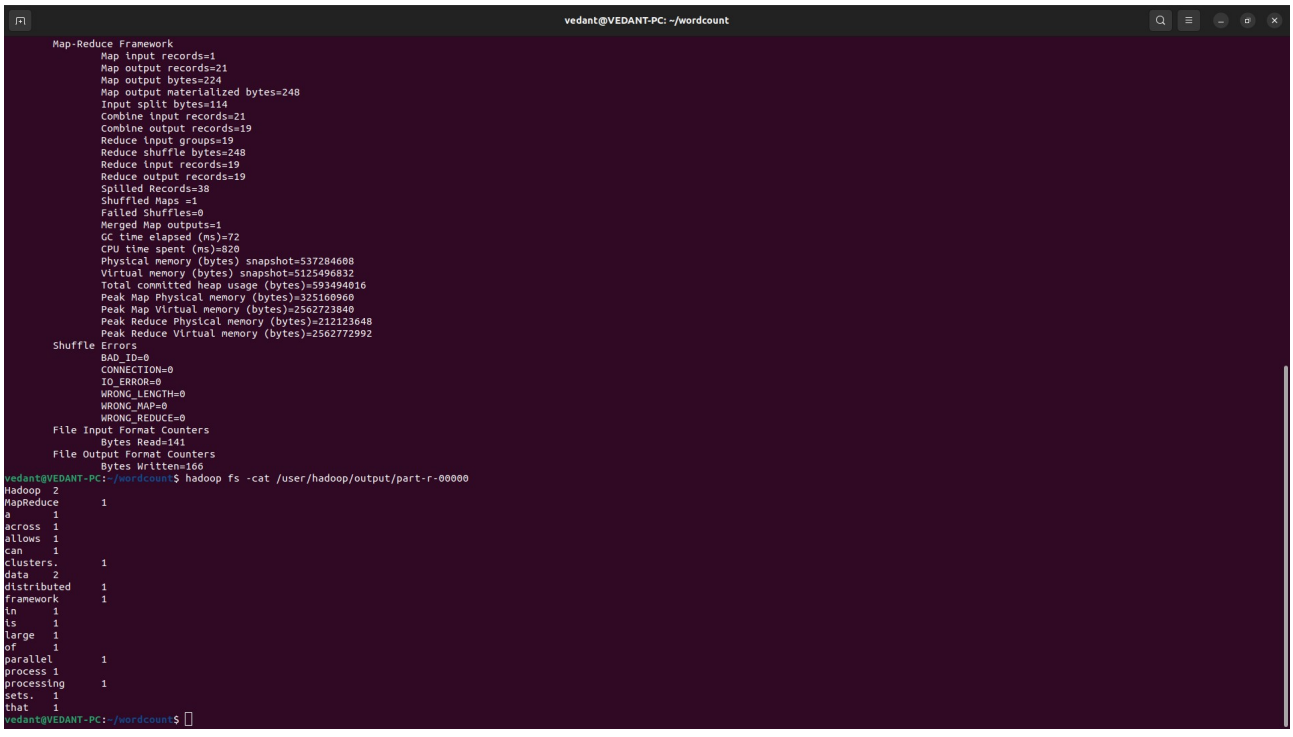

Namenode information
localhost:9870/dfshealth.html#tab-overview
Decommissioning Nodes 0
Entering Maintenance Nodes 0
Total Datanode Volume Failures 0 (0 B)
Number of Under-Replicated Blocks 0
Number of Blocks Pending Deletion (including replicas) 0
Block Deletion Start Time Mon Jan 20 07:44:17 +0530 2025
Last Checkpoint Time Mon Jan 20 07:41:47 +0530 2025
Last HA Transition Time Never
Enabled Erasure Coding Policies RS-6-3-1024k
NameNode Journal Status
Current transaction ID: 2
Journal Manager
State
FileJournalManager(root=/tmp/hadoop-vedant/dfs/name)
EditLogFileOutputStream(/tmp/hadoop-vedant/dfs/name/current/edits_inprogress_0000000000000000002)
NameNode Storage
Storage Directory
Type
State
/tmp/hadoop-vedant/dfs/name
IMAGE_AND_EDITS
Active
DFS Storage Types
Storage Type
Configured Capacity
Capacity Used
Capacity Remaining
Block Pool Used
Nodes In Service
DISK
34.6 GB
24 KB (0%)
16.99 GB (49.1%)
24 KB
1

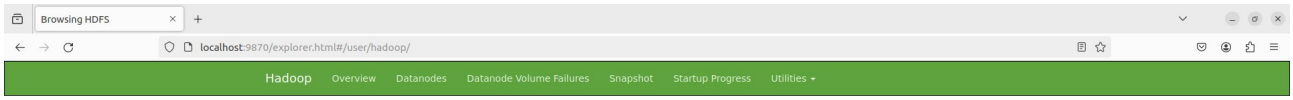
```java
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

import java.io.IOException;

public class WordCount {

    // Mapper class
    public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Context context) throws IOException,
InterruptedException {
            String[] words = value.toString().split("\\s+");
            for (String str : words) {
                word.set(str);
                context.write(word, one);
            }
        }
    }

    // Reducer class
    public static class IntSumReducer extends Reducer<Text, IntWritable, Text,
IntWritable> {
        private IntWritable result = new IntWritable();

        public void reduce(Text key, Iterable<IntWritable> values, Context context)
                throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }

    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Usage: WordCount <input path> <output path>");
            System.exit(-1);
        }

        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "word count");
        job.setJarByClass(WordCount.class);

        job.setMapperClass(TokenizerMapper.class);
        job.setCombinerClass(IntSumReducer.class);
        job.setReducerClass(IntSumReducer.class);

        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

Input.txt -


Open
input.txt
~/wordcount
Save
1 Hadoop is a framework that allows distributed processing of large data sets. Hadoop MapReduce
can process data in parallel across clusters.


vedant@VEDANT-PC: ~/wordcount
vedant@VEDANT-PC:~/wordcount$ javac -classpath `hadoop classpath` -d . WordCount.java
vedant@VEDANT-PC:~/wordcount$ jar -cvf wordcount.jar *.class
added manifest
adding: WordCount$IntSumReducer.class(in = 1739) (out= 740)(deflated 57%)
adding: WordCount$TokenizerMapper.class(in = 1828) (out= 792)(deflated 56%)
adding: WordCount.class(in = 1656) (out= 917)(deflated 44%)
vedant@VEDANT-PC:~/wordcount$ hadoop fs -mkdir -p /user/hadoop/input
vedant@VEDANT-PC:~/wordcount$ hadoop fs -put input.txt /user/hadoop/input/
vedant@VEDANT-PC:~/wordcount$ hadoop jar wordcount.jar WordCount /user/hadoop/input /user/hadoop/output
2025-01-20 08:04:05,769 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-01-20 08:04:06,073 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-01-20 08:04:06,078 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/vedant/.staging/job_1737339268689_0001
2025-01-20 08:04:06,323 INFO input.FileInputFormat: Total input files to process : 1
2025-01-20 08:04:06,361 INFO mapreduce.JobSubmitter: number of splits:1
2025-01-20 08:04:06,859 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1737339268689_0001
2025-01-20 08:04:06,859 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-01-20 08:04:06,998 INFO conf.Configuration: resource-types.xml not found
2025-01-20 08:04:06,998 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-01-20 08:04:07,390 INFO impl.YarnClientImpl: Submitted application application_1737339268689_0001
2025-01-20 08:04:07,413 INFO mapreduce.Job: The url to track the job: http://VEDANT-PC:8088/proxy/application_1737339268689_0001/
2025-01-20 08:04:07,413 INFO mapreduce.Job: Running job: job_1737339268689_0001
2025-01-20 08:04:13,493 INFO mapreduce.Job: Job job_1737339268689_0001 running in uber mode : false
2025-01-20 08:04:13,496 INFO mapreduce.Job: map 0% reduce 0%
2025-01-20 08:04:17,548 INFO mapreduce.Job: map 100% reduce 0%
2025-01-20 08:04:21,571 INFO mapreduce.Job: map 100% reduce 100%
2025-01-20 08:04:21,591 INFO mapreduce.Job: Job job_1737339268689_0001 completed successfully
2025-01-20 08:04:21,705 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=248
FILE: Number of bytes written=619195
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=255
HDFS: Number of bytes written=166
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=1597
Total time spent by all reduces in occupied slots (ms)=1727
Total time spent by all map tasks (ms)=1597
Total time spent by all reduce tasks (ms)=1727
Total vcore-milliseconds taken by all map tasks=1597
Total vcore-milliseconds taken by all reduce tasks=1727
Total megabyte-milliseconds taken by all map tasks=1635328
Total megabyte-milliseconds taken by all reduce tasks=1768448
Map-Reduce Framework
Map input records=1
Map output records=21
Map output bytes=224
Map output materialized bytes=248
Input split bytes=114


vedant@VEDANT-PC: ~/wordcount
Map-Reduce Framework
Map input records=1
Map output records=21
Map output bytes=224
Map output materialized bytes=248
Input split bytes=114
Combine input records=21
Combine output records=19
Reduce input groups=19
Reduce shuffle bytes=248
Reduce input records=19
Reduce output records=19
Spilled Records=38
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=72
CPU time spent (ms)=820
Physical memory (bytes) snapshot=537284608
Virtual memory (bytes) snapshot=5125496832
Total committed heap usage (bytes)=593494016
Peak Map Physical memory (bytes)=325160960
Peak Map Virtual memory (bytes)=2562723840
Peak Reduce Physical memory (bytes)=212123648
Peak Reduce Virtual memory (bytes)=2562772992
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=141
File Output Format Counters
Bytes Written=166
vedant@VEDANT-PC:~/wordcount$ hadoop fs -cat /user/hadoop/output/part-r-00000
Hadoop 2
MapReduce 1
a 1
across 1
allows 1
can 1
clusters. 1
data 2
distributed 1
framework 1
in 1
is 1
large 1
of 1
parallel 1
process 1
processing 1
sets. 1
that 1
vedant@VEDANT-PC:~/wordcount$



Browse Directory

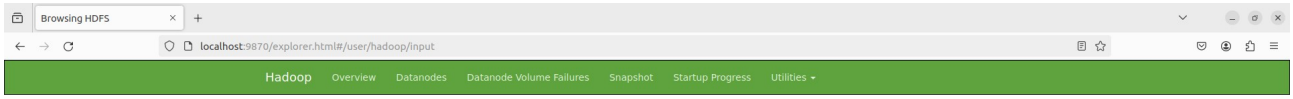
/user/hadoop/ Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	vedant	supergroup	0 B	Jan 20 08:03	0	0 B	input
drwxr-xr-x	vedant	supergroup	0 B	Jan 20 08:04	0	0 B	output

Showing 1 to 2 of 2 entries Previous 1 Next

Hadoop, 2024.



Browse Directory

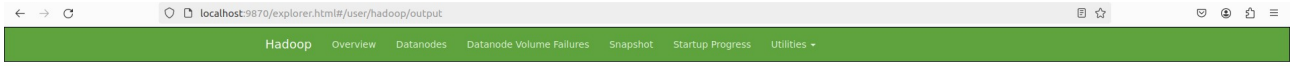
/user/hadoop/input Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	vedant	supergroup	141 B	Jan 20 08:03	1	128 MB	input.txt

Showing 1 to 1 of 1 entries Previous 1 Next

Hadoop, 2024.



Browse Directory

/user/hadoop/output Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	vedant	supergroup	0 B	Jan 20 08:04	1	128 MB	_SUCCESS
-rw-r--r--	vedant	supergroup	166 B	Jan 20 08:04	1	128 MB	part-r-00000

Showing 1 to 2 of 2 entries Previous 1 Next

Hadoop, 2024.