

20CYS402 — *Distributed Systems and Cloud Computing*

Lab Exercise – 5

Name: R Subramanian

Roll No: CH.EN.U4CYS22043

Objective of the Exercise

To understand and implement Hadoop service management and MapReduce programming by performing:

1. Hadoop Service Startup and HDFS Interaction
2. Word Count using MapReduce
3. Line Count using MapReduce

Task 1: Hadoop Service Startup and HDFS Interaction

Commands Used:

- `start-dfs.sh` → Starts **NameNode**, **DataNode**, and **SecondaryNameNode** servers.
- `start-yarn.sh` → Starts **ResourceManager** and **NodeManager** servers.
- `mapred --daemon start historyserver` → Starts the **JobHistory Server**.

Verification:

- Used the `jps` command to list and verify that all core Hadoop services are running.

Hadoop Daemons and Their Roles:

Daemon	Role
NameNode	Manages HDFS namespace and metadata (directories, files, permissions)
DataNode	Stores actual blocks of data in HDFS
SecondaryNameNode	Periodically merges NameNode metadata (fsimage + edit logs) to prevent bloat
ResourceManager	Master daemon of YARN; manages resource allocation and job scheduling
NodeManager	Manages application containers and reports resource usage to ResourceManager

Output:

```
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:~/Downloads$ hadoop fs -mkdir -p /user/asecys
2025-08-20 09:17:57,670 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:~/Downloads$ hadoop fs -put hello.txt /user/asecys/
2025-08-20 09:18:01,155 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:~/Downloads$ hadoop fs -ls /
2025-08-20 09:18:09,129 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
Found 3 items
drwxr-xr-x - asecys supergroup 0 2025-08-20 09:14 /testdir
drwxrwx--- - asecys supergroup 0 2025-08-20 08:55 /tmp
drwxr-xr-x - asecys supergroup 0 2025-08-20 09:17 /user

asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:~/Downloads$ hadoop fs -mkdir /testdir
2025-08-20 09:14:08,286 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
```

hadoop

ResourceManager

NodeManager

Node Information

List of Applications

List of Containers

Tools

NodeManager information

NodeManager information

Total Vmem allocated for Containers

Vmem enforcement enabled

Total Pmem allocated for Container

Pmem enforcement enabled

Total VCores allocated for Containers

Resource types

NodeHealthyStatus

LastNodeHealthTime

NodeHealthReport

NodeManager started on

NodeManager Version:

Hadoop Version:

hadoop

Application

About Jobs

Tools

JobHistory

Retired Jobs

Show 20 ▼ entries

Submit Time

Start Time

Finish Time

Job ID ▼

Name

User

Queue

State

Maps Total

Maps Completed

Reduces Total

Reduces Completed

Elapsed Time

No data available in table

Showing 0 to 0 of 0 entries

First Previous Next Last

hadoop

Cluster

About Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

All Applications

Cluster Metrics

Apps Submitted

Apps Pending

Apps Running

Apps Completed

Containers Running

Used Resources

Total Resources

Reserved Res

Cluster Nodes Metrics

Active Nodes

Decommissioning Nodes

Decommissioned Nodes

Lost Nodes

Unhealthy Nodes

Scheduler Metrics

Scheduler Type

Scheduling Resource Type

Minimum Allocation

Maximum Allocation

Maximum Cluster Application Priority

Scheduler Busy %

Show 20 ▼ entries

ID ▼

User

Name

Application Type

Application Tags

Queue

Application Priority

StartTime

LaunchTime

FinishTime

State

FinalStatus

Running Containers

Allocated CPU VCores

Allocated Memory MB

Allocated GPUs

Reserved CPU VCores

R

No data available in table

Showing 0 to 0 of 0 entries

```
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:~/Downloads$ jps
7569 Jps
7185 NodeManager
6130 NameNode
6825 ResourceManager
6570 SecondaryNameNode
6333 DataNode
```

```
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: ~/Downloads$ mr-jobhistory-daemon.sh start historyserver
WARNING: Use of this script to start the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon start" instead.
```

```
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: ~/Downloads$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC]
2025-08-20 09:48:47,197 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: ~/Downloads$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

Task 2: Implement Word Count Using MapReduce

Procedure:

1. Create the input directory in HDFS:

```
hdfs dfs -mkdir /input
```

2. Upload sample text file (`textdata.txt`) to HDFS:

```
hdfs dfs -put textdata.txt /input
```

3. Run the MapReduce Job:

```
hadoop jar /home/asecys/Downloads/mapreduce/wordcount.jar com.jordiburgos.WordCount /input/textdata.txt /output/wordcount
```

4. Retrieve the output from HDFS:

```
hdfs dfs -cat /output/wordcount/part-r-00000
```

Output:

```
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:~/Downloads$ hdfs dfs -ls /output/wordcount
2025-08-20 09:38:09,407 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
Found 2 items
-rw-r--r-- 1 asecys supergroup 0 2025-08-20 09:34 /output/wordcount/_SUCCESS
-rw-r--r-- 1 asecys supergroup 4233 2025-08-20 09:34 /output/wordcount/part-r-00000
```

```

asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ hadoop jar /home/asecys/Downloads/mapreduce/wordcount.jar com.jordiburgos.WordCount /input/textdata.txt /output/wordcount
2025-08-20 09:34:32,968 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2025-08-20 09:34:33,244 INFO Client.DefaultHadoopRemoteProxyProvider: Connecting to ResourceManager at /s.d.0.0:8032
2025-08-20 09:34:33,419 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool Interface and execute your application with ToolRunner to remedy this.
2025-08-20 09:34:33,427 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/asecys/-staging/job_1755662662736_0001
2025-08-20 09:34:33,557 INFO Input.FileInputFormat: Total input files to process : 1
2025-08-20 09:34:33,584 INFO mapreduce.JobSubmitter: number of splits:1
2025-08-20 09:34:33,664 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1755662662736_0001
2025-08-20 09:34:33,664 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-08-20 09:34:33,749 INFO conf.Configuration: resource-types.xml not found
2025-08-20 09:34:33,749 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'
2025-08-20 09:34:34,607 INFO Impl.YarnClientImpl: Submitted application application_1755662662736_0001
2025-08-20 09:34:34,688 INFO mapreduce.Job: The url to track the job: http://asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:8088/proxy/application_1755662662736_0001/
2025-08-20 09:34:34,689 INFO mapreduce.Job: Running job: job_1755662662736_0001
2025-08-20 09:34:38,143 INFO mapreduce.Job: Job job_1755662662736_0001 running in uber mode : false
2025-08-20 09:34:38,144 INFO mapreduce.Job:  map 0% reduce 0%
2025-08-20 09:34:41,172 INFO mapreduce.Job:  map 100% reduce 0%
2025-08-20 09:34:45,188 INFO mapreduce.Job:  map 100% reduce 100%
2025-08-20 09:34:45,194 INFO mapreduce.Job: Job job_1755662662736_0001 completed successfully
2025-08-20 09:34:45,233 INFO mapreduce.Job: Counters: 54
File System Counters
    FILE: Number of bytes read=7869
    FILE: Number of bytes written=34439
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4220
    HDFS: Number of bytes written=4233
    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)=1047
    Total time spent by all reduces in occupied slots (ms)=1131
    Total time spent by all map tasks (ms)=1047
    Total time spent by all reduce tasks (ms)=1131
    Total vcore-milliseconds taken by all map tasks=1047
    Total vcore-milliseconds taken by all reduce tasks=1131
    Total megabyte-milliseconds taken by all map tasks=1072128
    Total megabyte-milliseconds taken by all reduce tasks=1158144
Map-Reduce Framework
    Map input records=20
    Map output records=628
    Map output bytes=6687
    Map output materialized bytes=7869
    Input split bytes=105
    Combine input records=0
    Combine output records=0
    Reduce input groups=400
    Reduce shuffle bytes=7869
    Reduce input records=0

```

```

asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ hdfs dfs -mkdir -p /input
2025-08-20 09:26:21,458 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ pwd
/home/asecys/Downloads
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ hdfs dfs -put /home/asecys/Downloads/mapreduce/textdata.txt /input/
2025-08-20 09:26:58,023 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

```

Task 3: Implement Line Count Using MapReduce

Code:

```
package com.mapreduce.linecount;
```

```
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```
public class LineCount {
```

```
    public static class LineMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private final static Text lineKey = new Text("TotalLines");
```

```
        @Override
```

```
        public void map(LongWritable key, Text value, Context context)
            throws IOException, InterruptedException {
            context.write(lineKey, one);
        }
    }
}
```

```

    }

    public static class LineReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
        @Override
        public void reduce(Text key, Iterable<IntWritable> values, Context context)
            throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            context.write(key, new IntWritable(sum));
        }
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "line count");

    job.setJarByClass(LineCount.class);
    job.setMapperClass(LineMapper.class);
    job.setReducerClass(LineReducer.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);
}
}

```

Explanation:

1. Compile the code and create the `.jar` file.
2. Create the input directory in HDFS:

```
hdfs dfs -mkdir /input
```

3. Upload the input file (`textdata.txt`) to HDFS:

```
hdfs dfs -put textdata.txt /input
```

4. Execute the MapReduce Job:

```
hadoop jar /home/asecys/Downloads/mapreduce/linecount.jar com.mapreduce.linecount.LineCount /input/textdata.txt /output/linecount
```

5. Retrieve the output from HDFS:

```
hdfs dfs -cat /output/linecount/part-r-00000
```

Output:

```
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ hdfs dfs -cat /output/linecount/part-r-00000
2025-08-20 09:54:39,160 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
TotalLines      20
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$

asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ hadoop jar /home/asecys/Downloads/linecount.jar com.mapreduce.linecount.LineCount /input/textdata
2025-08-20 09:54:09,990 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2025-08-20 09:54:10,269 INFO client.DefaultHARMFalloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-08-20 09:54:10,419 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your ap
2025-08-20 09:54:10,442 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/asecys/.staging/job_1755662662736_0002
2025-08-20 09:54:10,566 INFO Input.FileInputFormat: Total input files to process : 1
2025-08-20 09:54:10,590 INFO mapreduce.JobSubmitter: number of splits:1
2025-08-20 09:54:10,672 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1755662662736_0002
2025-08-20 09:54:10,672 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-08-20 09:54:10,763 INFO conf.Configuration: resource-types.xml not found
2025-08-20 09:54:10,763 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-08-20 09:54:10,802 INFO Impl.YarnClientImpl: Submitted application application_1755662662736_0002
2025-08-20 09:54:10,815 INFO mapreduce.Job: The url to track the job: http://asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC:8088/proxy/application_1755662662736_0002
2025-08-20 09:54:10,815 INFO mapreduce.Job: Running job: job_1755662662736_0002
2025-08-20 09:54:14,863 INFO mapreduce.Job: Job job_1755662662736_0002 running in uber mode : false
2025-08-20 09:54:14,864 INFO mapreduce.Job:  map 0% reduce 0%
2025-08-20 09:54:17,886 INFO mapreduce.Job:  map 100% reduce 0%
2025-08-20 09:54:21,901 INFO mapreduce.Job:  map 100% reduce 100%
2025-08-20 09:54:21,907 INFO mapreduce.Job: Job job_1755662662736_0002 completed successfully
2025-08-20 09:54:21,946 INFO mapreduce.Job: Counters: 54
    File System Counters
      FILE: Number of bytes read=346
      FILE: Number of bytes written=618593
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=4220
      HDFS: Number of bytes written=14
      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)=1041
      Total time spent by all reduces in occupied slots (ms)=1046
      Total time spent by all map tasks (ms)=1041
      Total time spent by all reduce tasks (ms)=1046
      Total vcore-milliseconds taken by all map tasks=1041
      Total vcore-milliseconds taken by all reduce tasks=1046
      Total megabyte-milliseconds taken by all map tasks=1065984
      Total megabyte-milliseconds taken by all reduce tasks=1071104
    Map-Reduce Framework
      Map input records=20
      Map output records=20
      Map output bytes=300

asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ nano LineCount.java
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ mkdir -p com/mapreduce/linecount
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ # You can use any editor, e.g. nano, vim, or a GUI editor
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ nano com/mapreduce/linecount/LineCount.java
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ nano com/mapreduce/linecount/LineCount.java
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ mkdir -p linecount_classes
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ javac -classpath $(hadoop classpath) -d linecount_classes com/mapreduce/linecount/LineCount.java
asecys@asecys-HP-Pro-Tower-400-G9-PCI-Desktop-PC: /Downloads$ jar -cvf linecount.jar -C linecount_classes/ .
added manifest
adding: com/(ln = 0) (out= 0)(stored 0%)
adding: com/mapreduce/(ln = 0) (out= 0)(stored 0%)
adding: com/mapreduce/linecount/(ln = 0) (out= 0)(stored 0%)
adding: com/mapreduce/linecount/LineCount$LineMapper.class(ln = 1690) (out= 649)(deflated 61%)
adding: com/mapreduce/linecount/LineCount$LineReducer.class(ln = 1701) (out= 709)(deflated 58%)
adding: com/mapreduce/linecount/LineCount.class(ln = 1530) (out= 810)(deflated 47%)
```