

EX:10 WORD COUNT PROGRAM USING MAP REDUCE

Date :14-10-22

Name :M.Divya varshini

Roll no: 19UITE032

The screenshot shows a Windows desktop with a PowerShell window and an Outlook window. The PowerShell window is titled "Administrator: Windows PowerShell" and displays the following commands and output:

```
PS C:\Windows\system32> cd D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> .\start-dfs.cmd
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> .\start-yarn.cmd
starting yarn daemons
7620 NameNode
2920 NodeManager
3756 DataNode
5340 ResourceManager
6844 Jps
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hdfs dfs -put D:\divya_
```

The Outlook window shows an email from "A VARSHINI.M" with the subject "PRINT" and a date of "Thu 11/3/2022 4:24 PM". The email contains a list of attachments:

- 19uite002_Ex10 (1).pdf (430 KB)
- 19UITE032_EX10_Diffie_Hell... (9 KB)
- EX.9_RSA1.pdf (315 KB)
- 19uite002_EX_9.pdf (1 MB)
- CRYPTO_7_19UITE032.pdf (402 KB)
- OpenStack Installation 11.pdf (404 KB)
- OpenStack Installation.pdf (403 KB)

The screenshot shows a Windows desktop with a PowerShell window and an Outlook window. The PowerShell window is titled "Administrator: Windows PowerShell" and displays the following commands and output:

```
3756 DataNode
5340 ResourceManager
6844 Jps
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hdfs dfs -put D:\divya.txt /user
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> ls

Directory: D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin

Mode                LastWriteTime         Length Name
----                -
d----- 03-01-2021 15:24                FederationStateStore
-a----- 03-01-2021 15:02          2756 distribute-exclude.sh
-a----- 03-01-2021 14:58          1983 hadoop-daemon.sh
-a----- 03-01-2021 14:58          2522 hadoop-daemons.sh
-a----- 03-01-2021 15:03          1542 httpfs.sh
-a----- 03-01-2021 14:59          1508 kms.sh
-a----- 03-01-2021 15:27          1841 mr-jobhistory-daemon.sh
-a----- 03-01-2021 15:02          2086 refresh-namenodes.sh
-a----- 03-01-2021 14:58          1770 start-all.cmd
-a----- 03-01-2021 14:58          2221 start-all.sh
-a----- 03-01-2021 15:02          1880 start-balancer.sh
-a----- 03-01-2021 15:02          1401 start-dfs.cmd
-a----- 03-01-2021 15:02          5170 start-dfs.sh
-a----- 03-01-2021 15:02          1793 start-secure-dns.sh
-a----- 03-01-2021 15:24          1571 start-yarn.cmd
-a----- 03-01-2021 15:24          3342 start-yarn.sh
-a----- 03-01-2021 14:58          1770 stop-all.cmd
-a----- 03-01-2021 14:58          2168 stop-all.sh
-a----- 03-01-2021 15:02          1783 stop-balancer.sh
-a----- 03-01-2021 15:02          1455 stop-dfs.cmd
-a----- 03-01-2021 15:02          3898 stop-dfs.sh
-a----- 03-01-2021 15:02          1750 stop-secure-dns.sh
-a----- 03-01-2021 15:24          1642 stop-yarn.cmd
-a----- 03-01-2021 15:24          3883 stop-yarn.sh
-a----- 03-01-2021 14:58          1982 workers.sh
-a----- 03-01-2021 15:24          1814 yarn-daemon.sh
-a----- 03-01-2021 15:24          2320 yarn-daemons.sh

PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin>
```

The Outlook window shows an email from "A VARSHINI.M" with the subject "PRINT" and a date of "Thu 11/3/2022 4:24 PM". The email contains a list of attachments:

- 19uite002_Ex10 (1).pdf (430 KB)
- 19UITE032_EX10_Diffie_Hell... (9 KB)
- EX.9_RSA1.pdf (315 KB)
- 19uite002_EX_9.pdf (1 MB)
- CRYPTO_7_19UITE032.pdf (402 KB)
- OpenStack Installation 11.pdf (404 KB)
- OpenStack Installation.pdf (403 KB)

```
Administrator: Windows PowerShell

JAR does not exist or is not a normal file: D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoopmapreduce-examples-3.2.2.jar
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hadoop jar D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoopmapreduce-examples-3.2.2.jar wordcount /user /out
JAR does not exist or is not a normal file: D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoopmapreduce-examples-3.2.2.jar
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hadoop jar D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoopmapreduce-examples-3.2.2.jar wordcount /user /out
JAR does not exist or is not a normal file: D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoopmapreduce-examples-3.2.2.jar
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hadoop jar D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoopmapreduce-examples-3.2.2.jar wordcount /user /out
2022-11-04 14:12:17,533 INFO client.RMProxy: Connecting to ResourceManager at /6.0.0.0:8032
2022-11-04 14:12:18,548 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/19uite032/.staging/job_1667550281802_0001
2022-11-04 14:12:19,446 INFO input.FileInputFormat: Total input files to process : 1
2022-11-04 14:12:19,822 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-04 14:12:20,169 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667550281802_0001
2022-11-04 14:12:20,172 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-04 14:12:20,584 INFO conf.Configuration: resource-types.xml not found
2022-11-04 14:12:20,585 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-04 14:12:21,249 INFO impl.YarnClientImpl: Submitted application application_1667550281802_0001
2022-11-04 14:12:21,342 INFO mapreduce.Job: The url to track the job: http://a4it64:8088/proxy/application_1667550281802_0001/
2022-11-04 14:12:21,344 INFO mapreduce.Job: Running job: job_1667550281802_0001
```

```
Administrator: Windows PowerShell

[-head <file>]
[-help [cmd ...]]
[-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]
[-mkdir [-p] <path> ...]
[-moveFromLocal [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
[-moveToLocal <src> <localdst>]
[-mv <src> ... <dst>]
[-put [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc> ... <dst>]
[-renameSnapshot <snapshotDir> <oldName> <newName>]
[-rm [-f] [-r] [-R] [-skipTrash] [-safely] <src> ...]
[-rmDir [-ignoreFailOnNonEmpty] <dir> ...]
[-setfacl [-R] [[-b|-k] {-m|-x <acl_spec>} <path>]] [--set <acl_spec> <path>]]
[-setfacl [-n name [-v value] | -x name] <path>]
[-setrep [-R] [-w] <rep> <path> ...]
[-stat [format] <path> ...]
[-tail [-f] [-s <sleep interval>] <file>]
[-test [-defswrz] <path>]
[-text [-ignoreCrc] <src> ...]
[-touch [-a] [-m] [-t TIMESTAMP] [-c] <path> ...]
[-touch <path> ...]
[-truncate [-w] <length> <path> ...]
[-usage [cmd ...]]

Generic options supported are:
-conf <configuration file>      specify an application configuration file
-D <property=value>              define a value for a given property
-fs <file:///hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <local|resource-manager:port> specify a Resource Manager
-files <file1,...>               specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>              specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>        specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin> hadoop fs -cat /out/*
blava 1
iam 1
a 1
day 1
friday 1
hello 1
is 2
today 2
varshini 1
wonderful 1
PS D:\hadoop-env\hadoop-3.2.2\hadoop-3.2.2\sbin>
```