

Due Date: 25/11/2024

```
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker-compose
up -d
time="2024-11-25T14:58:15+05:30" level=warning msg="D:\\3.2
lectures\\parallel\\Practicel\\Distributed-programming\\docker-compose.yml:
the attribute `version` is obsolete, it will be ignored, please remove it to
avoid potential confusion"
[+] Running 4/4
✓ Network distributed-programming_default Created
0.1s

√ Container namenode

                                             Started
1.6s

√ Container datanode

                                             Started
1.6s

√ Container historyserver
                                             Started
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
datanode jps
>>
Error response from daemon: container
70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c7107a398376d62 is not
running
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker inspect
datanode
>>
        "Id":
"70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c7107a398376d62",
        "Created": "2024-11-25T09:28:15.752295939Z",
        "Path": "/entrypoint.sh",
        "Args": [
            "/run.sh"
        ],
        "State": {
            "Status": "exited".
            "Running": false,
            "Paused": false,
            "Restarting": false,
            "OOMKilled": false,
            "Dead": false,
            "Pid": 0,
            "ExitCode": 0,
            "Error": "",
            "StartedAt": "2024-11-25T09:28:17.092829443Z",
            "FinishedAt": "2024-11-25T09:28:24.607471599Z",
            "Health": {
                "Status": "unhealthy",
                "FailingStreak": 0,
```

```
"Log": []
        "Image":
sha256:35f899bcbe9f983825a8a3bdc135ed0e8e0eaf3b58f9b08bf257b5e86bae3b47",
        "ResolvConfPath":
"/var/lib/docker/containers/70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c
7107a398376d62/resolv.conf",
        "HostnamePath":
"/var/lib/docker/containers/70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c
7107a398376d62/hostname",
        "HostsPath":
"/var/lib/docker/containers/70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c"
7107a398376d62/hosts",
        "LogPath":
"/var/lib/docker/containers/70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c
7107a398376d62/70d5baa87889d70c5915f21fcfe91f5282aac3c7229338803c7107a398376d6
2-json.log",
        "Name": "/datanode",
        "RestartCount": 0,
        "Driver": "overlayfs",
        "Platform": "linux",
        "MountLabel": "",
        "ProcessLabel": "",
        "AppArmorProfile": "",
        "ExecIDs": null,
        "HostConfig": {
            "Binds": null,
            "ContainerIDFile": "",
            "LogConfig": {
                "Type": "json-file",
                "Config": {}
            "NetworkMode": "distributed-programming_default",
            "PortBindings": {},
            "RestartPolicy": {
                "Name": "no",
                "MaximumRetryCount": 0
            "AutoRemove": false,
            "VolumeDriver": "",
            "VolumesFrom": null,
            "ConsoleSize": [
                0,
                0
            "CapAdd": null,
            "CapDrop": null,
```

```
"CgroupnsMode": "host",
"Dns": null,
"DnsOptions": null,
"DnsSearch": null,
"ExtraHosts": [],
"GroupAdd": null,
"IpcMode": "private",
"Cgroup": "",
"Links": null,
"OomScoreAdj": 0,
"PidMode": "",
"Privileged": false,
"PublishAllPorts": false,
"ReadonlyRootfs": false,
"SecurityOpt": null,
"UTSMode": "",
"UsernsMode": "",
"ShmSize": 67108864,
"Runtime": "runc",
"Isolation": "",
"CpuShares": 0,
"Memory": 0,
"NanoCpus": 0,
"CgroupParent": "",
"BlkioWeight": 0,
"BlkioWeightDevice": null,
"BlkioDeviceReadBps": null,
"BlkioDeviceWriteBps": null,
"BlkioDeviceReadIOps": null,
"BlkioDeviceWriteIOps": null,
"CpuPeriod": 0,
"CpuQuota": 0,
"CpuRealtimePeriod": 0,
"CpuRealtimeRuntime": 0,
"CpusetCpus": "",
"CpusetMems": "",
"Devices": null,
"DeviceCgroupRules": null,
"DeviceRequests": null,
"MemoryReservation": 0,
"MemorySwap": 0,
"MemorySwappiness": null,
"OomKillDisable": false,
"PidsLimit": null,
"Ulimits": null,
"CpuCount": 0,
"CpuPercent": 0,
"IOMaximumIOps": 0,
```

```
"IOMaximumBandwidth": 0,
            "Mounts": [
                     "Type": "volume",
                    "Source": "distributed-programming datanode-data",
                     "Target": "/hadoop/dfs/data",
                    "VolumeOptions": {}
            ],
            "MaskedPaths": [
                "/proc/asound",
                "/proc/acpi",
                "/proc/kcore",
                "/proc/keys",
                "/proc/latency stats",
                "/proc/timer list",
                "/proc/timer stats",
                "/proc/sched_debug",
                "/proc/scsi",
                "/sys/firmware",
                "/sys/devices/virtual/powercap"
            "ReadonlyPaths": [
                "/proc/bus",
                "/proc/fs",
                "/proc/irq",
                "/proc/sys",
                "/proc/sysrq-trigger"
        },
        "GraphDriver": {
            "Data": null,
            "Name": "overlayfs"
        },
        "Mounts": [
                "Type": "volume",
                "Name": "distributed-programming_datanode-data",
                "Source": "/var/lib/docker/volumes/distributed-
programming_datanode-data/_data",
                "Destination": "/hadoop/dfs/data",
                "Driver": "local",
                "Mode": "z",
                "RW": true,
                "Propagation": ""
        ],
        "Config": {
```

```
"Hostname": "70d5baa87889",
            "Domainname": "",
            "User": "",
            "AttachStdin": false,
            "AttachStdout": true,
            "AttachStderr": true,
            "ExposedPorts": {
                "9864/tcp": {}
            "Tty": false,
            "OpenStdin": false,
            "StdinOnce": false,
            "Env": [
                "CORE_CONF_fs_defaultFS=hdfs://namenode:8020",
                "PATH=/opt/hadoop-
3.2.1/bin/:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin",
                "JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/",
                "HADOOP VERSION=3.2.1",
                "HADOOP URL=https://www.apache.org/dist/hadoop/common/hadoop-
3.2.1/hadoop-3.2.1.tar.gz",
                "HADOOP_HOME=/opt/hadoop-3.2.1",
                "HADOOP CONF DIR=/etc/hadoop",
                "MULTIHOMED NETWORK=1",
                "USER=root",
                "HDFS CONF dfs datanode data dir=file:///hadoop/dfs/data"
            ],
            "Cmd": [
                "/run.sh"
            ],
            "Healthcheck": {
                "Test": [
                    "CMD-SHELL",
                    "curl -f http://localhost:9864/ || exit 1"
            },
            "Image": "bde2020/hadoop-datanode:latest",
            "Volumes": {
                "/hadoop/dfs/data": {}
            },
            "WorkingDir": "",
            "Entrypoint": [
                "/entrypoint.sh"
            ],
            "OnBuild": null,
            "Labels": {
                "com.docker.compose.config-hash":
"568a1045cb0015bd223e5ea3416238424b7152b2ecbc163a8461da4ef6324158",
                "com.docker.compose.container-number": "1",
```

```
"com.docker.compose.depends_on":
"namenode:service started:false",
                "com.docker.compose.image":
"sha256:35f899bcbe9f983825a8a3bdc135ed0e8e0eaf3b58f9b08bf257b5e86bae3b47",
                "com.docker.compose.oneoff": "False",
                "com.docker.compose.project": "distributed-programming",
                "com.docker.compose.project.config files": "D:\\3.2
lectures\\parallel\\Practicel\\Distributed-programming\\docker-compose.yml",
                "com.docker.compose.project.working dir": "D:\\3.2
lectures\\parallel\\Practicel\\Distributed-programming",
                "com.docker.compose.service": "datanode",
                "com.docker.compose.version": "2.29.7"
        },
        "NetworkSettings": {
            "Bridge": "",
            "SandboxID": ""
            "SandboxKey": "",
            "Ports": {},
            "HairpinMode": false,
            "LinkLocalIPv6Address": "",
            "LinkLocalIPv6PrefixLen": 0,
            "SecondaryIPAddresses": null,
            "SecondaryIPv6Addresses": null,
            "EndpointID": "",
            "Gateway": "",
            "GlobalIPv6Address": "",
            "GlobalIPv6PrefixLen": 0,
            "IPAddress": "",
            "IPPrefixLen": 0,
            "IPv6Gateway": "",
            "MacAddress": "",
            "Networks": {
                "distributed-programming_default": {
                    "IPAMConfig": null,
                    "Links": null,
                    "Aliases": [
                        "datanode",
                        "datanode"
                    ],
                    "MacAddress": "",
                    "DriverOpts": null,
                    "NetworkID":
"5914f90a19bca14f3f462747e15cfe46b108f2a275c9a5b2b28023d21f7ba931",
                    "EndpointID": "",
                    "Gateway": "",
                    "IPAddress": "",
                    "IPPrefixLen": 0,
```

```
"IPv6Gateway": "",
                    "GlobalIPv6Address": "",
                    "GlobalIPv6PrefixLen": 0,
                    "DNSNames": [
                        "datanode",
                        "70d5baa87889"
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker-compose
down -v
>>
time="2024-11-25T15:00:03+05:30" level=warning msg="D:\\3.2
lectures\\parallel\\Practicel\\Distributed-programming\\docker-compose.yml:
the attribute `version` is obsolete, it will be ignored, please remove it to
avoid potential confusion"
[+] Running 6/6

✓ Container historyserver
                                                  Removed
10.5s

√ Container datanode

                                                  Removed
0.1s

√ Container namenode

                                                  Removed
10.5s
✓ Volume distributed-programming_namenode-data Removed
✓ Volume distributed-programming datanode-data Removed
0.0s
✓ Network distributed-programming_default
                                                  Removed
0.3s
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker-compose
up -d
>>
time="2024-11-25T15:00:26+05:30" level=warning msg="D:\\3.2
lectures\\parallel\\Practicel\\Distributed-programming\\docker-compose.yml:
the attribute `version` is obsolete, it will be ignored, please remove it to
avoid potential confusion"
[+] Running 6/6
✓ Network distributed-programming_default
                                                   Created
✓ Volume "distributed-programming_namenode-data" Created
✓ Volume "distributed-programming datanode-data" Created
0.0s

√ Container namenode

                                                    Started
1.7s
```

```
✓ Container datanode
                                                   Started
1.7s

✓ Container historyserver
                                                   Started
1.8s
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -put /sample.txt /input
>>
put: `/sample.txt': No such file or directory
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -mkdir -p /input
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker
cp ./sample.txt namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -put /sample.txt /input
2024-11-25 09:31:24,788 INFO sasl.SaslDataTransferClient: SASL encryption
trust check: localHostTrusted = false, remoteHostTrusted = false
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 3 root supergroup
                                       81 2024-11-25 09:31 /input/sample.txt
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-
mapreducer-examples-3.2.1.jar wordcount/input/output
JAR does not exist or is not a normal file: /opt/hadoop-
3.2.1/share/hadoop/mapreduce/hadoop-mapreducer-examples-3.2.1.jar
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode ls /opt/hadoop-3.2.1/share/hadoop/mapreduce/
hadoop-mapreduce-client-app-3.2.1.jar
                                          hadoop-mapreduce-client-hs-plugins-
                hadoop-mapreduce-client-shuffle-3.2.1.jar
3.2.1.jar
hadoop-mapreduce-client-common-3.2.1.jar hadoop-mapreduce-client-jobclient-
3.2.1-tests.jar hadoop-mapreduce-client-uploader-3.2.1.jar lib-examples
hadoop-mapreduce-client-core-3.2.1.jar
                                          hadoop-mapreduce-client-jobclient-
3.2.1.jar
                hadoop-mapreduce-examples-3.2.1.jar
                                                             sources
hadoop-mapreduce-client-hs-3.2.1.jar
                                          hadoop-mapreduce-client-nativetask-
                jdiff
3.2.1.jar
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jar wordcount /input /output
2024-11-25 09:38:50,567 INFO impl.MetricsConfig: Loaded properties from
hadoop-metrics2.properties
2024-11-25 09:38:50,621 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot
period at 10 second(s).
2024-11-25 09:38:50,621 INFO impl.MetricsSystemImpl: JobTracker metrics system
started
2024-11-25 09:38:50,869 INFO input.FileInputFormat: Total input files to
process : 1
2024-11-25 09:38:50,895 INFO mapreduce.JobSubmitter: number of splits:1
```

```
2024-11-25 09:38:51,066 INFO mapreduce.JobSubmitter: Submitting tokens for job
iob local1344267800 0001
2024-11-25 09:38:51,067 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-25 09:38:51,189 INFO mapreduce.Job: The url to track the job:
http://localhost:8080/
2024-11-25 09:38:51,190 INFO mapreduce.Job: Running job:
job local1344267800 0001
2024-11-25 09:38:51,192 INFO mapred.LocalJobRunner: OutputCommitter set in
config null
2024-11-25 09:38:51,200 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 2
2024-11-25 09:38:51,201 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup temporary folders under output directory:false, ignore cleanup
failures: false
2024-11-25 09:38:51,201 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-25 09:38:51,239 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-25 09:38:51,240 INFO mapred.LocalJobRunner: Starting task:
attempt local1344267800 0001 m 000000 0
2024-11-25 09:38:51,260 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 2
2024-11-25 09:38:51,261 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore cleanup
failures: false
2024-11-25 09:38:51,281 INFO mapred.Task: Using
ResourceCalculatorProcessTree : [ ]
2024-11-25 09:38:51,285 INFO mapred.MapTask: Processing split:
hdfs://namenode:8020/input/sample.txt:0+81
2024-11-25 09:38:51,341 INFO mapred.MapTask: (EQUATOR) 0 kvi
26214396(104857584)
2024-11-25 09:38:51,341 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-25 09:38:51,341 INFO mapred.MapTask: soft limit at 83886080
6553600
2024-11-25 09:38:51,346 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-25 09:38:51,378 INFO sasl.SaslDataTransferClient: SASL encryption
trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-25 09:38:51,484 INFO mapred.LocalJobRunner:
2024-11-25 09:38:51,486 INFO mapred.MapTask: Starting flush of map output
2024-11-25 09:38:51,486 INFO mapred.MapTask: Spilling map output
bufvoid = 104857600
kvend = 26214348(104857392); length = 49/6553600
2024-11-25 09:38:51,496 INFO mapred.MapTask: Finished spill 0
```

```
2024-11-25 09:38:51,508 INFO mapred.Task:
Task:attempt local1344267800 0001 m 000000 0 is done. And is in the process of
committing
2024-11-25 09:38:51,512 INFO mapred.Task: Task
attempt local1344267800 0001 m 000000 0' done.
File System Counters
              FILE: Number of bytes read=316693
              FILE: Number of bytes written=842115
              FILE: Number of read operations=0
              FILE: Number of large read operations=0
              FILE: Number of write operations=0
              HDFS: Number of bytes read=81
              HDFS: Number of bytes written=0
              HDFS: Number of read operations=5
              HDFS: Number of large read operations=0
              HDFS: Number of write operations=1
              HDFS: Number of bytes read erasure-coded=0
       Map-Reduce Framework
              Map input records=1
              Map output records=13
              Map output bytes=134
              Map output materialized bytes=166
              Input split bytes=102
              Combine input records=13
              Combine output records=13
              Spilled Records=13
              Failed Shuffles=0
              Merged Map outputs=0
              GC time elapsed (ms)=0
              Total committed heap usage (bytes)=273154048
       File Input Format Counters
              Bvtes Read=81
2024-11-25 09:38:51,521 INFO mapred.LocalJobRunner: Finishing task:
attempt_local1344267800_0001_m_000000_0
2024-11-25 09:38:51,521 INFO mapred.LocalJobRunner: map task executor complete
2024-11-25 09:38:51,524 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2024-11-25 09:38:51,525 INFO mapred.LocalJobRunner: Starting task:
attempt_local1344267800_0001_r_000000_0
2024-11-25 09:38:51,532 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 2
2024-11-25 09:38:51,532 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore cleanup
failures: false
2024-11-25 09:38:51,533 INFO mapred.Task: Using
ResourceCalculatorProcessTree : [ ]
```

```
2024-11-25 09:38:51,538 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@4e1f484
2024-11-25 09:38:51,541 WARN impl.MetricsSystemImpl: JobTracker metrics system
already initialized!
2024-11-25 09:38:51,559 INFO reduce.MergeManagerImpl: MergerManager:
memoryLimit=627939712, maxSingleShuffleLimit=156984928,
mergeThreshold=414440224, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2024-11-25 09:38:51,562 INFO reduce.EventFetcher:
attempt local1344267800 0001 r 000000 0 Thread started: EventFetcher for
fetching Map Completion Events
2024-11-25 09:38:51,590 INFO reduce.LocalFetcher: localfetcher#1 about to
shuffle output of map attempt local1344267800 0001 m 000000 0 decomp: 162 len:
166 to MEMORY
2024-11-25 09:38:51,595 INFO reduce.InMemoryMapOutput: Read 162 bytes from
map-output for attempt local1344267800 0001 m 000000 0
2024-11-25 09:38:51,598 INFO reduce.MergeManagerImpl: closeInMemoryFile ->
map-output of size: 162, inMemoryMapOutputs.size() -> 1, commitMemory -> 0,
usedMemory ->162
2024-11-25 09:38:51,600 INFO reduce.EventFetcher: EventFetcher is interrupted.
2024-11-25 09:38:51,601 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:38:51,602 INFO reduce.MergeManagerImpl: finalMerge called with 1
in-memory map-outputs and 0 on-disk map-outputs
2024-11-25 09:38:51,610 INFO mapred.Merger: Merging 1 sorted segments
2024-11-25 09:38:51,610 INFO mapred.Merger: Down to the last merge-pass, with
1 segments left of total size: 148 bytes
2024-11-25 09:38:51,611 INFO reduce.MergeManagerImpl: Merged 1 segments, 162
bytes to disk to satisfy reduce memory limit
2024-11-25 09:38:51,612 INFO reduce.MergeManagerImpl: Merging 1 files, 166
bytes from disk
2024-11-25 09:38:51,614 INFO reduce.MergeManagerImpl: Merging 0 segments, 0
bytes from memory into reduce
2024-11-25 09:38:51,614 INFO mapred.Merger: Merging 1 sorted segments
2024-11-25 09:38:51,615 INFO mapred.Merger: Down to the last merge-pass, with
1 segments left of total size: 148 bytes
2024-11-25 09:38:51,646 INFO Configuration.deprecation: mapred.skip.on is
deprecated. Instead, use mapreduce.job.skiprecords
2024-11-25 09:38:51,672 INFO sasl.SaslDataTransferClient: SASL encryption
trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-25 09:38:51,715 INFO mapred.Task:
Task:attempt local1344267800 0001_r_0000000_0 is done. And is in the process of
committing
2024-11-25 09:38:51,718 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:38:51,718 INFO mapred.Task: Task
attempt_local1344267800_0001_r_000000_0 is allowed to commit now
2024-11-25 09:38:51,733 INFO output.FileOutputCommitter: Saved output of task
'attempt local1344267800 0001 r 000000 0' to hdfs://namenode:8020/output
```

```
2024-11-25 09:38:51,734 INFO mapred.LocalJobRunner: reduce > reduce
2024-11-25 09:38:51,735 INFO mapred.Task: Task
'attempt local1344267800 0001 r 000000 0' done.
2024-11-25 09:38:51,735 INFO mapred.Task: Final Counters for
attempt_local1344267800 0001 r 000000 0: Counters: 30
        File System Counters
                FILE: Number of bytes read=317057
                FILE: Number of bytes written=842281
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=81
                HDFS: Number of bytes written=108
                HDFS: Number of read operations=10
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=3
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Combine input records=0
                Combine output records=0
                Reduce input groups=13
                Reduce shuffle bytes=166
                Reduce input records=13
                Reduce output records=13
                Spilled Records=13
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=8
                Total committed heap usage (bytes)=274202624
        Shuffle Errors
                BAD ID=0
                CONNECTION=0
                IO ERROR=0
                WRONG LENGTH=0
                WRONG MAP=0
                WRONG_REDUCE=0
        File Output Format Counters
                Bytes Written=108
2024-11-25 09:38:51,736 INFO mapred.LocalJobRunner: Finishing task:
attempt_local1344267800_0001_r_000000_0
2024-11-25 09:38:51,736 INFO mapred.LocalJobRunner: reduce task executor
complete.
2024-11-25 09:38:52,196    INFO mapreduce.Job: Job job_local1344267800_0001
running in uber mode : false
2024-11-25 09:38:52,198 INFO mapreduce.Job: map 100% reduce 100%
2024-11-25 09:38:52,200 INFO mapreduce.Job: Job job_local1344267800_0001
completed successfully
```

```
2024-11-25 09:38:52,207 INFO mapreduce.Job: Counters: 36
        File System Counters
                FILE: Number of bytes read=633750
                FILE: Number of bytes written=1684396
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=162
                HDFS: Number of bytes written=108
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=1
                Map output records=13
                Map output bytes=134
                Map output materialized bytes=166
                Input split bytes=102
                Combine input records=13
                Combine output records=13
                Reduce input groups=13
                Reduce shuffle bytes=166
                Reduce input records=13
                Reduce output records=13
                Spilled Records=26
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=8
                Total committed heap usage (bytes)=547356672
        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=81
        File Output Format Counters
                Bytes Written=108
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jarwordcount /input/output
JAR does not exist or is not a normal file: /opt/hadoop-
3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jarwordcount
```

```
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jarwordcount/input/output
JAR does not exist or is not a normal file: /opt/hadoop-
3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-
3.2.1.jarwordcount/input/output
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jarwordcount/input/output
JAR does not exist or is not a normal file: /opt/hadoop-
3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-
3.2.1.jarwordcount/input/output
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jar wordcount /input /output
2024-11-25 09:42:23,521 INFO impl.MetricsConfig: Loaded properties from
hadoop-metrics2.properties
2024-11-25 09:42:23,619 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot
period at 10 second(s).
2024-11-25 09:42:23,619 INFO impl.MetricsSystemImpl: JobTracker metrics system
started
org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory
hdfs://namenode:8020/output already exists
        at
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileO
utputFormat.java:164)
        at
org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:277)
org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:1
43)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java
1730)
        at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
        at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1588)
        at org.apache.hadoop.examples.WordCount.main(WordCount.java:87)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.j
ava:43)
       at java.lang.reflect.Method.invoke(Method.java:498)
```

```
at
org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.j
        at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
        at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.j
ava:43)
       at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode ls /opt/hadoop-3.2.1/share/hadoop/mapreduce/
hadoop-mapreduce-client-app-3.2.1.jar
                                          hadoop-mapreduce-client-hs-plugins-
3.2.1.jar
                hadoop-mapreduce-client-shuffle-3.2.1.jar
hadoop-mapreduce-client-common-3.2.1.jar hadoop-mapreduce-client-jobclient-
3.2.1-tests.jar hadoop-mapreduce-client-uploader-3.2.1.jar lib-examples
hadoop-mapreduce-client-core-3.2.1.jar
                                          hadoop-mapreduce-client-jobclient-
3.2.1.jar
                 hadoop-mapreduce-examples-3.2.1.jar
                                                             sources
hadoop-mapreduce-client-hs-3.2.1.jar
                                          hadoop-mapreduce-client-nativetask-
3.2.1.jar
                jdiff
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -rm -r /output
Deleted /output
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.2.1.jar wordcount /input /output
>>
2024-11-25 09:44:29,540 INFO impl.MetricsConfig: Loaded properties from
hadoop-metrics2.properties
2024-11-25 09:44:29,741 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot
period at 10 second(s).
2024-11-25 09:44:29,742 INFO impl.MetricsSystemImpl: JobTracker metrics system
2024-11-25 09:44:30,002 INFO input.FileInputFormat: Total input files to
process : 1
2024-11-25 09:44:30,027 INFO mapreduce.JobSubmitter: number of splits:1
2024-11-25 09:44:30,173 INFO mapreduce.JobSubmitter: Submitting tokens for job
job_local756343983_0001
2024-11-25 09:44:30,173 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-25 09:44:30,277 INFO mapreduce.Job: The url to track the job:
http://localhost:8080/
```

```
2024-11-25 09:44:30,279 INFO mapreduce.Job: Running job:
iob local756343983 0001
2024-11-25 09:44:30,279 INFO mapred.LocalJobRunner: OutputCommitter set in
config null
2024-11-25 09:44:30,288 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 2
2024-11-25 09:44:30,288 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore cleanup
failures: false
2024-11-25 09:44:30,288 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-25 09:44:30,331 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-25 09:44:30,331 INFO mapred.LocalJobRunner: Starting task:
attempt local756343983 0001 m 000000 0
2024-11-25 09:44:30,365 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 2
2024-11-25 09:44:30,366 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore cleanup
failures: false
2024-11-25 09:44:30,386 INFO mapred.Task: Using
ResourceCalculatorProcessTree : [ ]
hdfs://namenode:8020/input/sample.txt:0+81
2024-11-25 09:44:30,448 INFO mapred.MapTask: (EQUATOR) 0 kvi
26214396(104857584)
2024-11-25 09:44:30,449 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-25 09:44:30,449 INFO mapred.MapTask: soft limit at 83886080
6553600
2024-11-25 09:44:30,455 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-25 09:44:30,477 INFO sasl.SaslDataTransferClient: SASL encryption
trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-25 09:44:30,547 INFO mapred.MapTask: Starting flush of map output
2024-11-25 09:44:30,547 INFO mapred.MapTask: Spilling map output
bufvoid = 104857600
kvend = 26214348(104857392); length = 49/6553600
2024-11-25 09:44:30,558 INFO mapred.MapTask: Finished spill 0
2024-11-25 09:44:30,568 INFO mapred.Task:
Task:attempt_local756343983_0001_m_000000_0 is done. And is in the process of
committing
2024-11-25 09:44:30,575 INFO mapred.LocalJobRunner: map
2024-11-25 09:44:30,575 INFO mapred.Task: Task
'attempt local756343983 0001 m 000000 0' done.
```

```
2024-11-25 09:44:30,582 INFO mapred.Task: Final Counters for
attempt local756343983 0001 m 000000 0: Counters: 24
        File System Counters
                FILE: Number of bytes read=316693
                FILE: Number of bytes written=839567
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=81
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=1
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=1
                Map output records=13
                Map output bytes=134
                Map output materialized bytes=166
                Input split bytes=102
                Combine input records=13
                Combine output records=13
                Spilled Records=13
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=273154048
        File Input Format Counters
                Bytes Read=81
2024-11-25 09:44:30,582 INFO mapred.LocalJobRunner: Finishing task:
attempt local756343983 0001 m 000000 0
2024-11-25 09:44:30,582 INFO mapred.LocalJobRunner: map task executor complete
2024-11-25 09:44:30,586 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2024-11-25 09:44:30,588 INFO mapred.LocalJobRunner: Starting task:
attempt local756343983 0001 r 000000 0
2024-11-25 09:44:30,595 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 2
2024-11-25 09:44:30,595 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore cleanup
failures: false
2024-11-25 09:44:30,597 INFO mapred.Task: Using
ResourceCalculatorProcessTree : [ ]
2024-11-25 09:44:30,600 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@20567a1d
2024-11-25 09:44:30,602 WARN impl.MetricsSystemImpl: JobTracker metrics system
already initialized!
```

```
2024-11-25 09:44:30,620 INFO reduce.MergeManagerImpl: MergerManager:
memoryLimit=627939712, maxSingleShuffleLimit=156984928,
mergeThreshold=414440224, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2024-11-25 09:44:30,624 INFO reduce.EventFetcher:
attempt local756343983 0001 r 000000 0 Thread started: EventFetcher for
fetching Map Completion Events
2024-11-25 09:44:30,644 INFO reduce.LocalFetcher: localfetcher#1 about to
shuffle output of map attempt_local756343983_0001_m_000000_0 decomp: 162 len:
2024-11-25 09:44:30,647 INFO reduce.InMemoryMapOutput: Read 162 bytes from
map-output for attempt_local756343983_0001_m_000000_0
2024-11-25 09:44:30,650 INFO reduce.MergeManagerImpl: closeInMemoryFile ->
map-output of size: 162, inMemoryMapOutputs.size() -> 1, commitMemory -> 0,
usedMemory ->162
2024-11-25 09:44:30,651 INFO reduce.EventFetcher: EventFetcher is interrupted.
Returning
2024-11-25 09:44:30,652 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:44:30,653 INFO reduce.MergeManagerImpl: finalMerge called with 1
in-memory map-outputs and 0 on-disk map-outputs
2024-11-25 09:44:30,660 INFO mapred.Merger: Merging 1 sorted segments
2024-11-25 09:44:30,660 INFO mapred.Merger: Down to the last merge-pass, with
1 segments left of total size: 148 bytes
2024-11-25 09:44:30,661 INFO reduce.MergeManagerImpl: Merged 1 segments, 162
bytes to disk to satisfy reduce memory limit
2024-11-25 09:44:30,662 INFO reduce.MergeManagerImpl: Merging 1 files, 166
bytes from disk
2024-11-25 09:44:30,662 INFO reduce.MergeManagerImpl: Merging 0 segments, 0
bytes from memory into reduce
2024-11-25 09:44:30,662 INFO mapred.Merger: Merging 1 sorted segments
2024-11-25 09:44:30,663 INFO mapred.Merger: Down to the last merge-pass, with
1 segments left of total size: 148 bytes
2024-11-25 09:44:30,664 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:44:30,693 INFO Configuration.deprecation: mapred.skip.on is
deprecated. Instead, use mapreduce.job.skiprecords
2024-11-25 09:44:30,724 INFO sasl.SaslDataTransferClient: SASL encryption
trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-25 09:44:30,769 INFO mapred.Task:
Task:attempt local756343983_0001_r_000000_0 is done. And is in the process of
committing
2024-11-25 09:44:30,773 INFO mapred.Task: Task
attempt local756343983 0001 r 000000 0 is allowed to commit now
2024-11-25 09:44:30,789 INFO output.FileOutputCommitter: Saved output of task
'attempt_local756343983_0001_r_000000_0' to hdfs://namenode:8020/output
2024-11-25 09:44:30,790 INFO mapred.Task: Task
'attempt_local756343983_0001_r_000000_0' done.
```

```
2024-11-25 09:44:30,790 INFO mapred.Task: Final Counters for
attempt local756343983 0001 r 000000 0: Counters: 30
        File System Counters
                FILE: Number of bytes read=317057
                FILE: Number of bytes written=839733
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=81
                HDFS: Number of bytes written=108
                HDFS: Number of read operations=10
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=3
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Combine input records=0
                Combine output records=0
                Reduce input groups=13
                Reduce shuffle bytes=166
                Reduce input records=13
                Reduce output records=13
                Spilled Records=13
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=10
                Total committed heap usage (bytes)=274202624
        Shuffle Errors
                BAD ID=0
                CONNECTION=0
                IO ERROR=0
                WRONG LENGTH=0
                WRONG MAP=0
                WRONG REDUCE=0
        File Output Format Counters
                Bytes Written=108
2024-11-25 09:44:30,791 INFO mapred.LocalJobRunner: Finishing task:
attempt local756343983 0001 r 000000 0
2024-11-25 09:44:30,791 INFO mapred.LocalJobRunner: reduce task executor
complete.
2024-11-25 09:44:31,286 INFO mapreduce.Job: Job job_local756343983_0001
running in uber mode : false
2024-11-25 09:44:31,287 INFO mapreduce.Job: map 100% reduce 100%
2024-11-25 09:44:31,288 INFO mapreduce.Job: Job job_local756343983_0001
completed successfully
2024-11-25 09:44:31,294 INFO mapreduce.Job: Counters: 36
        File System Counters
               FILE: Number of bytes read=633750
```

```
FILE: Number of bytes written=1679300
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=162
                HDFS: Number of bytes written=108
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=1
                Map output records=13
                Map output bytes=134
                Map output materialized bytes=166
                Input split bytes=102
                Combine input records=13
                Combine output records=13
                Reduce input groups=13
                Reduce shuffle bytes=166
                Reduce input records=13
                Reduce output records=13
                Spilled Records=26
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=10
                Total committed heap usage (bytes)=547356672
        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=81
        File Output Format Counters
                Bytes Written=108
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs -ls /output
opt/hadoop-3.2.1/libexec/hadoop-functions.sh: line 2401: HDFS -LS USER: bad
substitution
/opt/hadoop-3.2.1/libexec/hadoop-functions.sh: line 2366: HDFS_-LS_USER: bad
substitution
ERROR: -ls is not COMMAND nor fully qualified CLASSNAME.
Usage: hdfs [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
```

OPTIONS is none or any of:

--buildpaths attempt to add class files from build tree

--config dir Hadoop config directory
--daemon (start|status|stop) operate on a daemon

--debug turn on shell script debug mode

--help usage information

--hostnames list[,of,host,names] hosts to use in worker mode

--hosts filename list of hosts to use in worker mode --loglevel level set the log4j level for this command

--workers turn on worker mode

SUBCOMMAND is one of:

Admin Commands:

cacheadmin configure the HDFS cache

crypto configure HDFS encryption zones

debug run a Debug Admin to execute HDFS debug commands

dfsadmin run a DFS admin client

dfsrouteradmin manage Router-based federation ec run a HDFS ErasureCoding CLI

fsck run a DFS filesystem checking utility

haadmin run a DFS HA admin client

jmxget get JMX exported values from NameNode or DataNode.

oev apply the offline edits viewer to an edits file

oiv apply the offline fsimage viewer to an fsimage

oiv_legacy apply the offline fsimage viewer to a legacy fsimage storagepolicies list/get/set/satisfyStoragePolicy block storage policies

Client Commands:

classpath prints the class path needed to get the hadoop jar and

the required libraries

dfs run a filesystem command on the file system
envvars display computed Hadoop environment variables
fetchdt fetch a delegation token from the NameNode

getconf get config values from configuration groups get the groups which users belong to

lsSnapshottableDir list all snapshottable dirs owned by the current user snapshotDiff diff two snapshots of a directory or diff the current

directory contents with a snapshot version

Daemon Commands:

balancer run a cluster balancing utility

```
datanode
                    run a DFS datanode
dfsrouter
                    run the DFS router
diskbalancer
                    Distributes data evenly among disks on a given node
                    run HttpFS server, the HDFS HTTP Gateway
httpfs
                    run the DFS journalnode
journalnode
                    run a utility to move block replicas across storage types
mover
namenode
                    run the DFS namenode
nfs3
                    run an NFS version 3 gateway
portmap
                    run a portmap service
secondarynamenode
                    run the DFS secondary namenode
sps
                    run external storagepolicysatisfier
zkfc
                    run the ZK Failover Controller daemon
SUBCOMMAND may print help when invoked w/o parameters or with -h.
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 3 root supergroup 0 2024-11-25 09:44 /output/_SUCCESS
-rw-r--r-- 3 root supergroup 108 2024-11-25 09:44 /output/part-r-
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker exec -it
namenode hdfs dfs -cat /output/part-r-00000
2024-11-25 09:55:46,457 INFO sasl.SaslDataTransferClient: SASL encryption
Distributed
Let's
This
      1
       1
explore.
               1
file
       1
for
      1
is
       1
lecture.
programming
simple 1
the
       1
txt
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> docker-compose
time="2024-11-25T15:31:07+05:30" level=warning msg="D:\\3.2
lectures\\parallel\\Practicel\\Distributed-programming\\docker-compose.yml:
the attribute `version` is obsolete, it will be ignored, please remove it to
avoid potential confusion"
[+] Running 4/4

✓ Container historyserver
                                          Removed
10.5s
✓ Container datanode
                                          Removed
10.6s
```

```
√ Container namenode

                                            Removed
10.5s
✓ Network distributed-programming default Removed
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git init
Initialized empty Git repository in D:/3.2
lectures/parallel/Practicel/Distributed-programming/.git/
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git add .
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git commit -m
"Distributed hadoop programming"
[master (root-commit) 3f71ffe] Distributed hadoop programming
2 files changed, 80 insertions(+)
create mode 100644 docker-compose.yml
create mode 100644 sample.txt
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git bramch -M
main
git: 'bramch' is not a git command. See 'git --help'.
The most similar command is
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git branch -M
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git remote add
origin https://github.com/Thars7/Distributed-Hadoop-Programming.git
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming> git push -u
origin main
Enumerating objects: 4, done.
Counting objects: 100% (4/4), done.
Delta compression using up to 8 threads
Compressing objects: 100\% (4/4), done.
Writing objects: 100% (4/4), 774 bytes | 774.00 KiB/s, done.
Total 4 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
To https://github.com/Thars7/Distributed-Hadoop-Programming.git
* [new branch]
                     main -> main
branch 'main' set up to track 'origin/main'.
PS D:\3.2 lectures\parallel\Practicel\Distributed-programming>
```