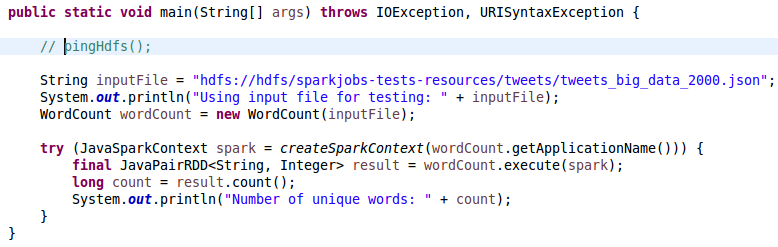
# Strange behavior when accessing hdfs-Resources from Spark-Jobs

A „java.net.UnknownHostException: hdfs“ Exception is raised when Spark tries to read files from HDFS.

Interesting: If there is an access to hdfs fore spark execution the error goes away

## This does NOT work

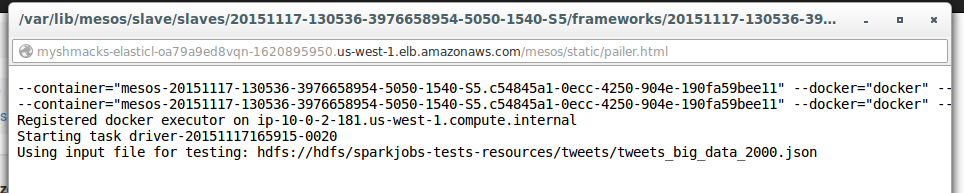


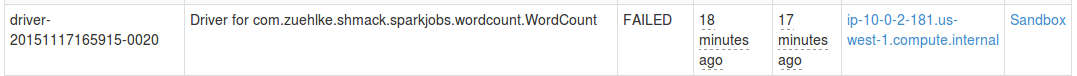
Resulting error:

Caused by: java.net.UnknownHostException: hdfs

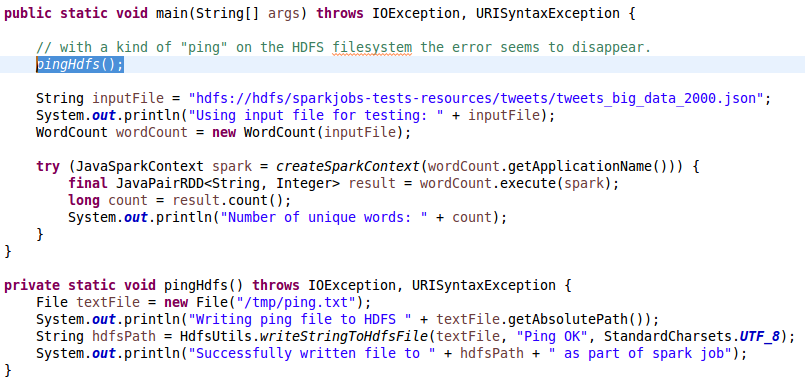
Full stacktrace see 3.1

Standard Output:

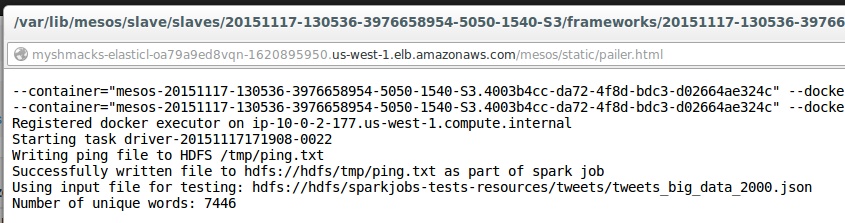




## This work-around seems to “fix” the error



Standard Output:





## Additional Informations

### How is the job executed?

cd ~/shmack/repo/04\_implementation/sparkjobs ; gradle clean fatJarWithTests && sync-to-hdfs.sh ~/shmack/repo/04\_implementation/sparkjobs/build/libs /sparkjobs-tests-libs && submit-spark-job.sh -Dspark.mesos.coarse=true --driver-cores 1 --driver-memory 1024M --class com.zuehlke.shmack.sparkjobs.wordcount.WordCount hdfs://hdfs/sparkjobs-tests-libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar 30



### How is the dummy file read from hdfs?

rm -f /tmp/foo && hadoop fs -copyToLocal hdfs://hdfs/tmp/WordCountRemoteTest/HdfsWriteTest.txt /tmp/foo && more /tmp/foo

# Why?

Is there a solution?

Is there a relation to this:

<http://apache-spark-user-list.1001560.n3.nabble.com/Inconsistent-behavior-when-running-spark-on-top-of-tachyon-on-top-of-HDFS-HA-td1544.html>

# Appendix

## Full stacktrace of error

I1117 16:59:16.415534 7812 logging.cpp:172] INFO level logging started!

I1117 16:59:16.415745 7812 fetcher.cpp:414] Fetcher Info: {"cache\_directory":"\/tmp\/mesos\/fetch\/slaves\/20151117-130536-3976658954-5050-1540-S5","items":[{"action":"BYPASS\_CACHE","uri":{"extract":true,"value":"hdfs:\/\/hdfs\/sparkjobs-tests-libs\/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar"}}],"sandbox\_directory":"\/var\/lib\/mesos\/slave\/slaves\/20151117-130536-3976658954-5050-1540-S5\/frameworks\/20151117-130536-3976658954-5050-1540-0005\/executors\/driver-20151117165915-0020\/runs\/c54845a1-0ecc-4250-904e-190fa59bee11"}

I1117 16:59:16.417106 7812 fetcher.cpp:369] Fetching URI 'hdfs://hdfs/sparkjobs-tests-libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar'

I1117 16:59:16.417130 7812 fetcher.cpp:243] Fetching directly into the sandbox directory

I1117 16:59:16.417153 7812 fetcher.cpp:180] Fetching URI 'hdfs://hdfs/sparkjobs-tests-libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar'

I1117 16:59:16.676235 7812 fetcher.cpp:105] Downloading resource with Hadoop client from 'hdfs://hdfs/sparkjobs-tests-libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar' to '/var/lib/mesos/slave/slaves/20151117-130536-3976658954-5050-1540-S5/frameworks/20151117-130536-3976658954-5050-1540-0005/executors/driver-20151117165915-0020/runs/c54845a1-0ecc-4250-904e-190fa59bee11/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar'

15/11/17 16:59:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

15/11/17 16:59:17 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.

W1117 16:59:18.263938 7812 fetcher.cpp:265] Copying instead of extracting resource from URI with 'extract' flag, because it does not seem to be an archive: hdfs://hdfs/sparkjobs-tests-libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar

I1117 16:59:18.264170 7812 fetcher.cpp:446] Fetched 'hdfs://hdfs/sparkjobs-tests-libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar' to '/var/lib/mesos/slave/slaves/20151117-130536-3976658954-5050-1540-S5/frameworks/20151117-130536-3976658954-5050-1540-0005/executors/driver-20151117165915-0020/runs/c54845a1-0ecc-4250-904e-190fa59bee11/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar'

mesos-docker-executor: /lib64/libcurl.so.4: no version information available (required by /opt/mesosphere/packages/mesos--8467f0ef9a5aa54a502e0e0ab0f6515d8aecf00f/lib/libmesos-0.24.1.so)

I1117 16:59:18.461035 7885 exec.cpp:133] Version: 0.24.1

I1117 16:59:18.463685 7886 exec.cpp:207] Executor registered on slave 20151117-130536-3976658954-5050-1540-S5

Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties

15/11/17 16:59:19 INFO SparkContext: Running Spark version 1.5.0

15/11/17 16:59:19 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

15/11/17 16:59:19 WARN SparkConf:

SPARK\_JAVA\_OPTS was detected (set to '-Dspark.mesos.coarse=true -Dspark.mesos.executor.docker.image=mesosphere/spark:1.5.0-multi-roles-v2-bin-2.4.0').

This is deprecated in Spark 1.0+.

Please instead use:

- ./spark-submit with conf/spark-defaults.conf to set defaults for an application

- ./spark-submit with --driver-java-options to set -X options for a driver

- spark.executor.extraJavaOptions to set -X options for executors

- SPARK\_DAEMON\_JAVA\_OPTS to set java options for standalone daemons (master or worker)

15/11/17 16:59:19 WARN SparkConf: Setting 'spark.executor.extraJavaOptions' to '-Dspark.mesos.coarse=true -Dspark.mesos.executor.docker.image=mesosphere/spark:1.5.0-multi-roles-v2-bin-2.4.0' as a work-around.

15/11/17 16:59:19 WARN SparkConf: Setting 'spark.driver.extraJavaOptions' to '-Dspark.mesos.coarse=true -Dspark.mesos.executor.docker.image=mesosphere/spark:1.5.0-multi-roles-v2-bin-2.4.0' as a work-around.

15/11/17 16:59:19 INFO SecurityManager: Changing view acls to: root

15/11/17 16:59:19 INFO SecurityManager: Changing modify acls to: root

15/11/17 16:59:19 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)

15/11/17 16:59:20 INFO Slf4jLogger: Slf4jLogger started

15/11/17 16:59:20 INFO Remoting: Starting remoting

15/11/17 16:59:20 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriver@10.0.2.181:60978]

15/11/17 16:59:20 INFO Utils: Successfully started service 'sparkDriver' on port 60978.

15/11/17 16:59:20 INFO SparkEnv: Registering MapOutputTracker

15/11/17 16:59:20 INFO SparkEnv: Registering BlockManagerMaster

15/11/17 16:59:20 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-39da5787-6174-4cbc-bf60-0cc5fd20cc78

15/11/17 16:59:20 INFO MemoryStore: MemoryStore started with capacity 530.3 MB

15/11/17 16:59:20 INFO HttpFileServer: HTTP File server directory is /tmp/spark-45b60b79-ff85-4559-aca6-3c8563ee9dfe/httpd-cf0e710e-3a0a-4dbf-8fcf-14e030e7a130

15/11/17 16:59:20 INFO HttpServer: Starting HTTP Server

15/11/17 16:59:20 INFO Utils: Successfully started service 'HTTP file server' on port 45750.

15/11/17 16:59:20 INFO SparkEnv: Registering OutputCommitCoordinator

15/11/17 16:59:20 INFO Utils: Successfully started service 'SparkUI' on port 4040.

15/11/17 16:59:20 INFO SparkUI: Started SparkUI at http://10.0.2.181:4040

15/11/17 16:59:21 INFO SparkContext: Added JAR file:/mnt/mesos/sandbox/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar at http://10.0.2.181:45750/jars/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar with timestamp 1447779561067

15/11/17 16:59:21 WARN MetricsSystem: Using default name DAGScheduler for source because spark.app.id is not set.

2015-11-17 16:59:21,243:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@712: Client environment:zookeeper.version=zookeeper C client 3.4.5

2015-11-17 16:59:21,243:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@716: Client environment:host.name=ip-10-0-2-181.us-west-1.compute.internal

2015-11-17 16:59:21,243:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@723: Client environment:os.name=Linux

2015-11-17 16:59:21,244:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@724: Client environment:os.arch=4.1.7-coreos

2015-11-17 16:59:21,244:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@725: Client environment:os.version=#2 SMP Wed Sep 16 22:54:37 UTC 2015

2015-11-17 16:59:21,244:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@733: Client environment:user.name=(null)

2015-11-17 16:59:21,244:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@741: Client environment:user.home=/root

2015-11-17 16:59:21,244:6(0x7f8c6a50a700):ZOO\_INFO@log\_env@753: Client environment:user.dir=/opt/spark/dist

2015-11-17 16:59:21,244:6(0x7f8c6a50a700):ZOO\_INFO@zookeeper\_init@786: Initiating client connection, host=master.mesos:2181 sessionTimeout=10000 watcher=0x7f8cb4b36340 sessionId=0 sessionPasswd=<null> context=0x7f8ccc28d860 flags=0

I1117 16:59:21.245659 99 sched.cpp:157] Version: 0.23.0

2015-11-17 16:59:21,251:6(0x7f8c66401700):ZOO\_INFO@check\_events@1703: initiated connection to server [10.0.7.237:2181]

2015-11-17 16:59:21,252:6(0x7f8c66401700):ZOO\_INFO@check\_events@1750: session establishment complete on server [10.0.7.237:2181], sessionId=0x151158d374e002c, negotiated timeout=10000

I1117 16:59:21.253231 93 group.cpp:313] Group process (group(1)@10.0.2.181:38080) connected to ZooKeeper

I1117 16:59:21.253398 93 group.cpp:787] Syncing group operations: queue size (joins, cancels, datas) = (0, 0, 0)

I1117 16:59:21.253597 93 group.cpp:385] Trying to create path '/mesos' in ZooKeeper

I1117 16:59:21.256209 93 detector.cpp:138] Detected a new leader: (id='1')

I1117 16:59:21.256333 93 group.cpp:656] Trying to get '/mesos/json.info\_0000000001' in ZooKeeper

I1117 16:59:21.259670 90 detector.cpp:481] A new leading master (UPID=master@10.0.7.237:5050) is detected

I1117 16:59:21.259748 90 sched.cpp:254] New master detected at master@10.0.7.237:5050

I1117 16:59:21.259924 90 sched.cpp:264] No credentials provided. Attempting to register without authentication

I1117 16:59:21.261229 90 sched.cpp:448] Framework registered with 20151117-130536-3976658954-5050-1540-0023

15/11/17 16:59:21 INFO CoarseMesosSchedulerBackend: Registered as framework ID 20151117-130536-3976658954-5050-1540-0023

15/11/17 16:59:21 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 46019.

15/11/17 16:59:21 INFO NettyBlockTransferService: Server created on 46019

15/11/17 16:59:21 INFO BlockManagerMaster: Trying to register BlockManager

15/11/17 16:59:21 INFO BlockManagerMasterEndpoint: Registering block manager 10.0.2.181:46019 with 530.3 MB RAM, BlockManagerId(driver, 10.0.2.181, 46019)

15/11/17 16:59:21 INFO BlockManagerMaster: Registered BlockManager

15/11/17 16:59:21 INFO CoarseMesosSchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0

15/11/17 16:59:22 INFO MemoryStore: ensureFreeSpace(146080) called with curMem=0, maxMem=556038881

15/11/17 16:59:22 INFO MemoryStore: Block broadcast\_0 stored as values in memory (estimated size 142.7 KB, free 530.1 MB)

15/11/17 16:59:22 INFO CoarseMesosSchedulerBackend: Mesos task 2 is now TASK\_RUNNING

15/11/17 16:59:22 INFO MemoryStore: ensureFreeSpace(12866) called with curMem=146080, maxMem=556038881

15/11/17 16:59:22 INFO MemoryStore: Block broadcast\_0\_piece0 stored as bytes in memory (estimated size 12.6 KB, free 530.1 MB)

15/11/17 16:59:22 INFO BlockManagerInfo: Added broadcast\_0\_piece0 in memory on 10.0.2.181:46019 (size: 12.6 KB, free: 530.3 MB)

15/11/17 16:59:22 INFO SparkContext: Created broadcast 0 from textFile at WordCount.java:66

15/11/17 16:59:22 INFO CoarseMesosSchedulerBackend: Mesos task 0 is now TASK\_RUNNING

15/11/17 16:59:22 INFO CoarseMesosSchedulerBackend: Mesos task 1 is now TASK\_RUNNING

15/11/17 16:59:22 INFO CoarseMesosSchedulerBackend: Mesos task 4 is now TASK\_RUNNING

15/11/17 16:59:22 INFO CoarseMesosSchedulerBackend: Mesos task 3 is now TASK\_RUNNING

15/11/17 16:59:22 INFO FileInputFormat: Total input paths to process : 1

15/11/17 16:59:22 INFO SparkContext: Starting job: count at WordCount.java:52

15/11/17 16:59:22 INFO DAGScheduler: Registering RDD 3 (mapToPair at WordCount.java:75)

15/11/17 16:59:22 INFO DAGScheduler: Got job 0 (count at WordCount.java:52) with 2 output partitions

15/11/17 16:59:22 INFO DAGScheduler: Final stage: ResultStage 1(count at WordCount.java:52)

15/11/17 16:59:22 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 0)

15/11/17 16:59:22 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 0)

15/11/17 16:59:22 INFO DAGScheduler: Submitting ShuffleMapStage 0 (MapPartitionsRDD[3] at mapToPair at WordCount.java:75), which has no missing parents

15/11/17 16:59:22 INFO MemoryStore: ensureFreeSpace(5056) called with curMem=158946, maxMem=556038881

15/11/17 16:59:22 INFO MemoryStore: Block broadcast\_1 stored as values in memory (estimated size 4.9 KB, free 530.1 MB)

15/11/17 16:59:22 INFO MemoryStore: ensureFreeSpace(2813) called with curMem=164002, maxMem=556038881

15/11/17 16:59:22 INFO MemoryStore: Block broadcast\_1\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 530.1 MB)

15/11/17 16:59:22 INFO BlockManagerInfo: Added broadcast\_1\_piece0 in memory on 10.0.2.181:46019 (size: 2.7 KB, free: 530.3 MB)

15/11/17 16:59:22 INFO SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:861

15/11/17 16:59:23 INFO DAGScheduler: Submitting 2 missing tasks from ShuffleMapStage 0 (MapPartitionsRDD[3] at mapToPair at WordCount.java:75)

15/11/17 16:59:23 INFO TaskSchedulerImpl: Adding task set 0.0 with 2 tasks

15/11/17 16:59:24 INFO CoarseMesosSchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@ip-10-0-2-180.us-west-1.compute.internal:37252/user/Executor#172323092]) with ID 20151117-130536-3976658954-5050-1540-S4

15/11/17 16:59:24 INFO TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, ip-10-0-2-180.us-west-1.compute.internal, NODE\_LOCAL, 2248 bytes)

15/11/17 16:59:24 INFO TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, ip-10-0-2-180.us-west-1.compute.internal, NODE\_LOCAL, 2248 bytes)

15/11/17 16:59:24 INFO CoarseMesosSchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@ip-10-0-2-178.us-west-1.compute.internal:40538/user/Executor#1067190633]) with ID 20151117-130536-3976658954-5050-1540-S2

15/11/17 16:59:24 INFO CoarseMesosSchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@ip-10-0-2-177.us-west-1.compute.internal:48635/user/Executor#1327093130]) with ID 20151117-130536-3976658954-5050-1540-S3

15/11/17 16:59:24 INFO CoarseMesosSchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@ip-10-0-2-179.us-west-1.compute.internal:57223/user/Executor#1534311043]) with ID 20151117-130536-3976658954-5050-1540-S0

15/11/17 16:59:24 INFO BlockManagerMasterEndpoint: Registering block manager ip-10-0-2-180.us-west-1.compute.internal:56794 with 530.3 MB RAM, BlockManagerId(20151117-130536-3976658954-5050-1540-S4, ip-10-0-2-180.us-west-1.compute.internal, 56794)

15/11/17 16:59:24 INFO BlockManagerMasterEndpoint: Registering block manager ip-10-0-2-178.us-west-1.compute.internal:45107 with 530.3 MB RAM, BlockManagerId(20151117-130536-3976658954-5050-1540-S2, ip-10-0-2-178.us-west-1.compute.internal, 45107)

15/11/17 16:59:24 INFO BlockManagerMasterEndpoint: Registering block manager ip-10-0-2-177.us-west-1.compute.internal:37093 with 530.3 MB RAM, BlockManagerId(20151117-130536-3976658954-5050-1540-S3, ip-10-0-2-177.us-west-1.compute.internal, 37093)

15/11/17 16:59:24 INFO BlockManagerMasterEndpoint: Registering block manager ip-10-0-2-179.us-west-1.compute.internal:43851 with 530.3 MB RAM, BlockManagerId(20151117-130536-3976658954-5050-1540-S0, ip-10-0-2-179.us-west-1.compute.internal, 43851)

15/11/17 16:59:24 INFO CoarseMesosSchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@ip-10-0-2-181.us-west-1.compute.internal:59743/user/Executor#609204125]) with ID 20151117-130536-3976658954-5050-1540-S5

15/11/17 16:59:24 INFO BlockManagerMasterEndpoint: Registering block manager ip-10-0-2-181.us-west-1.compute.internal:45570 with 530.3 MB RAM, BlockManagerId(20151117-130536-3976658954-5050-1540-S5, ip-10-0-2-181.us-west-1.compute.internal, 45570)

15/11/17 16:59:24 INFO BlockManagerInfo: Added broadcast\_1\_piece0 in memory on ip-10-0-2-180.us-west-1.compute.internal:56794 (size: 2.7 KB, free: 530.3 MB)

15/11/17 16:59:24 INFO BlockManagerInfo: Added broadcast\_0\_piece0 in memory on ip-10-0-2-180.us-west-1.compute.internal:56794 (size: 12.6 KB, free: 530.3 MB)

15/11/17 16:59:25 WARN TaskSetManager: Lost task 1.0 in stage 0.0 (TID 1, ip-10-0-2-180.us-west-1.compute.internal): java.lang.IllegalArgumentException: java.net.UnknownHostException: hdfs

at org.apache.hadoop.security.SecurityUtil.buildTokenService(SecurityUtil.java:377)

at org.apache.hadoop.hdfs.NameNodeProxies.createNonHAProxy(NameNodeProxies.java:240)

at org.apache.hadoop.hdfs.NameNodeProxies.createProxy(NameNodeProxies.java:144)

at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:579)

at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:524)

at org.apache.hadoop.hdfs.DistributedFileSystem.initialize(DistributedFileSystem.java:146)

at org.apache.hadoop.fs.FileSystem.createFileSystem(FileSystem.java:2397)

at org.apache.hadoop.fs.FileSystem.access$200(FileSystem.java:89)

at org.apache.hadoop.fs.FileSystem$Cache.getInternal(FileSystem.java:2431)

at org.apache.hadoop.fs.FileSystem$Cache.get(FileSystem.java:2413)

at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:368)

at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:167)

at org.apache.hadoop.mapred.JobConf.getWorkingDirectory(JobConf.java:653)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:427)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:400)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$32.apply(SparkContext.scala:1007)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$32.apply(SparkContext.scala:1007)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:176)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:176)

at scala.Option.map(Option.scala:145)

at org.apache.spark.rdd.HadoopRDD.getJobConf(HadoopRDD.scala:176)

at org.apache.spark.rdd.HadoopRDD$$anon$1.<init>(HadoopRDD.scala:220)

at org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:216)

at org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:101)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:73)

at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:41)

at org.apache.spark.scheduler.Task.run(Task.scala:88)

at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:214)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)

at java.lang.Thread.run(Thread.java:745)

Caused by: java.net.UnknownHostException: hdfs

... 42 more

15/11/17 16:59:25 INFO TaskSetManager: Starting task 1.1 in stage 0.0 (TID 2, ip-10-0-2-177.us-west-1.compute.internal, ANY, 2248 bytes)

15/11/17 16:59:25 INFO TaskSetManager: Lost task 0.0 in stage 0.0 (TID 0) on executor ip-10-0-2-180.us-west-1.compute.internal: java.lang.IllegalArgumentException (java.net.UnknownHostException: hdfs) [duplicate 1]

15/11/17 16:59:25 INFO TaskSetManager: Starting task 0.1 in stage 0.0 (TID 3, ip-10-0-2-177.us-west-1.compute.internal, ANY, 2248 bytes)

15/11/17 16:59:25 INFO BlockManagerInfo: Added broadcast\_1\_piece0 in memory on ip-10-0-2-177.us-west-1.compute.internal:37093 (size: 2.7 KB, free: 530.3 MB)

15/11/17 16:59:25 INFO BlockManagerInfo: Added broadcast\_0\_piece0 in memory on ip-10-0-2-177.us-west-1.compute.internal:37093 (size: 12.6 KB, free: 530.3 MB)

15/11/17 16:59:25 INFO TaskSetManager: Lost task 0.1 in stage 0.0 (TID 3) on executor ip-10-0-2-177.us-west-1.compute.internal: java.lang.IllegalArgumentException (java.net.UnknownHostException: hdfs) [duplicate 2]

15/11/17 16:59:25 INFO TaskSetManager: Lost task 1.1 in stage 0.0 (TID 2) on executor ip-10-0-2-177.us-west-1.compute.internal: java.lang.IllegalArgumentException (java.net.UnknownHostException: hdfs) [duplicate 3]

15/11/17 16:59:25 INFO TaskSetManager: Starting task 1.2 in stage 0.0 (TID 4, ip-10-0-2-178.us-west-1.compute.internal, NODE\_LOCAL, 2248 bytes)

15/11/17 16:59:25 INFO TaskSetManager: Starting task 0.2 in stage 0.0 (TID 5, ip-10-0-2-179.us-west-1.compute.internal, ANY, 2248 bytes)

15/11/17 16:59:26 INFO BlockManagerInfo: Added broadcast\_1\_piece0 in memory on ip-10-0-2-178.us-west-1.compute.internal:45107 (size: 2.7 KB, free: 530.3 MB)

15/11/17 16:59:26 INFO BlockManagerInfo: Added broadcast\_1\_piece0 in memory on ip-10-0-2-179.us-west-1.compute.internal:43851 (size: 2.7 KB, free: 530.3 MB)

15/11/17 16:59:26 INFO BlockManagerInfo: Added broadcast\_0\_piece0 in memory on ip-10-0-2-178.us-west-1.compute.internal:45107 (size: 12.6 KB, free: 530.3 MB)

15/11/17 16:59:26 INFO BlockManagerInfo: Added broadcast\_0\_piece0 in memory on ip-10-0-2-179.us-west-1.compute.internal:43851 (size: 12.6 KB, free: 530.3 MB)

15/11/17 16:59:26 INFO TaskSetManager: Lost task 1.2 in stage 0.0 (TID 4) on executor ip-10-0-2-178.us-west-1.compute.internal: java.lang.IllegalArgumentException (java.net.UnknownHostException: hdfs) [duplicate 4]

15/11/17 16:59:26 INFO TaskSetManager: Starting task 1.3 in stage 0.0 (TID 6, ip-10-0-2-180.us-west-1.compute.internal, NODE\_LOCAL, 2248 bytes)

15/11/17 16:59:26 INFO TaskSetManager: Lost task 0.2 in stage 0.0 (TID 5) on executor ip-10-0-2-179.us-west-1.compute.internal: java.lang.IllegalArgumentException (java.net.UnknownHostException: hdfs) [duplicate 5]

15/11/17 16:59:26 INFO TaskSetManager: Starting task 0.3 in stage 0.0 (TID 7, ip-10-0-2-181.us-west-1.compute.internal, ANY, 2248 bytes)

15/11/17 16:59:26 INFO TaskSetManager: Lost task 1.3 in stage 0.0 (TID 6) on executor ip-10-0-2-180.us-west-1.compute.internal: java.lang.IllegalArgumentException (java.net.UnknownHostException: hdfs) [duplicate 6]

15/11/17 16:59:26 ERROR TaskSetManager: Task 1 in stage 0.0 failed 4 times; aborting job

15/11/17 16:59:26 INFO TaskSchedulerImpl: Cancelling stage 0

15/11/17 16:59:26 INFO TaskSchedulerImpl: Stage 0 was cancelled

15/11/17 16:59:26 INFO DAGScheduler: ShuffleMapStage 0 (mapToPair at WordCount.java:75) failed in 3.716 s

15/11/17 16:59:26 INFO DAGScheduler: Job 0 failed: count at WordCount.java:52, took 3.868935 s

15/11/17 16:59:26 INFO SparkUI: Stopped Spark web UI at http://10.0.2.181:4040

15/11/17 16:59:26 INFO DAGScheduler: Stopping DAGScheduler

15/11/17 16:59:26 INFO CoarseMesosSchedulerBackend: Shutting down all executors

15/11/17 16:59:26 INFO CoarseMesosSchedulerBackend: Asking each executor to shut down

I1117 16:59:26.816882 38 sched.cpp:1591] Asked to stop the driver

I1117 16:59:26.817010 90 sched.cpp:835] Stopping framework '20151117-130536-3976658954-5050-1540-0023'

15/11/17 16:59:26 INFO CoarseMesosSchedulerBackend: driver.run() returned with code DRIVER\_STOPPED

15/11/17 16:59:26 WARN ReliableDeliverySupervisor: Association with remote system [akka.tcp://sparkExecutor@ip-10-0-2-178.us-west-1.compute.internal:40538] has failed, address is now gated for [5000] ms. Reason: [Disassociated]

15/11/17 16:59:26 WARN ReliableDeliverySupervisor: Association with remote system [akka.tcp://sparkExecutor@ip-10-0-2-180.us-west-1.compute.internal:37252] has failed, address is now gated for [5000] ms. Reason: [Disassociated]

15/11/17 16:59:26 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

15/11/17 16:59:26 INFO MemoryStore: MemoryStore cleared

15/11/17 16:59:26 INFO BlockManager: BlockManager stopped

15/11/17 16:59:26 INFO BlockManagerMaster: BlockManagerMaster stopped

15/11/17 16:59:26 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

15/11/17 16:59:26 INFO SparkContext: Successfully stopped SparkContext

Exception in thread "main" org.apache.spark.SparkException: Job aborted due to stage failure: Task 1 in stage 0.0 failed 4 times, most recent failure: Lost task 1.3 in stage 0.0 (TID 6, ip-10-0-2-180.us-west-1.compute.internal): java.lang.IllegalArgumentException: java.net.UnknownHostException: hdfs

at org.apache.hadoop.security.SecurityUtil.buildTokenService(SecurityUtil.java:377)

at org.apache.hadoop.hdfs.NameNodeProxies.createNonHAProxy(NameNodeProxies.java:240)

at org.apache.hadoop.hdfs.NameNodeProxies.createProxy(NameNodeProxies.java:144)

at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:579)

at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:524)

at org.apache.hadoop.hdfs.DistributedFileSystem.initialize(DistributedFileSystem.java:146)

at org.apache.hadoop.fs.FileSystem.createFileSystem(FileSystem.java:2397)

at org.apache.hadoop.fs.FileSystem.access$200(FileSystem.java:89)

at org.apache.hadoop.fs.FileSystem$Cache.getInternal(FileSystem.java:2431)

at org.apache.hadoop.fs.FileSystem$Cache.get(FileSystem.java:2413)

at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:368)

at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:167)

at org.apache.hadoop.mapred.JobConf.getWorkingDirectory(JobConf.java:653)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:427)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:400)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$32.apply(SparkContext.scala:1007)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$32.apply(SparkContext.scala:1007)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:176)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:176)

at scala.Option.map(Option.scala:145)

at org.apache.spark.rdd.HadoopRDD.getJobConf(HadoopRDD.scala:176)

at org.apache.spark.rdd.HadoopRDD$$anon$1.<init>(HadoopRDD.scala:220)

at org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:216)

at org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:101)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:73)

at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:41)

at org.apache.spark.scheduler.Task.run(Task.scala:88)

at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:214)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)

at java.lang.Thread.run(Thread.java:745)

Caused by: java.net.UnknownHostException: hdfs

... 42 more

Driver stacktrace:

at org.apache.spark.scheduler.DAGScheduler.org$apache$spark$scheduler$DAGScheduler$$failJobAndIndependentStages(DAGScheduler.scala:1280)

at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply(DAGScheduler.scala:1268)

at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply(DAGScheduler.scala:1267)

at scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)

at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:47)

at org.apache.spark.scheduler.DAGScheduler.abortStage(DAGScheduler.scala:1267)

at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFailed$1.apply(DAGScheduler.scala:697)

at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFailed$1.apply(DAGScheduler.scala:697)

at scala.Option.foreach(Option.scala:236)

at org.apache.spark.scheduler.DAGScheduler.handleTaskSetFailed(DAGScheduler.scala:697)

at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOnReceive(DAGScheduler.scala:1493)

at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(DAGScheduler.scala:1455)

at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(DAGScheduler.scala:1444)

at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:48)

at org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.scala:567)

at org.apache.spark.SparkContext.runJob(SparkContext.scala:1813)

at org.apache.spark.SparkContext.runJob(SparkContext.scala:1826)

at org.apache.spark.SparkContext.runJob(SparkContext.scala:1839)

at org.apache.spark.SparkContext.runJob(SparkContext.scala:1910)

at org.apache.spark.rdd.RDD.count(RDD.scala:1121)

at org.apache.spark.api.java.JavaRDDLike$class.count(JavaRDDLike.scala:445)

at org.apache.spark.api.java.AbstractJavaRDDLike.count(JavaRDDLike.scala:47)

at com.zuehlke.shmack.sparkjobs.wordcount.WordCount.main(WordCount.java:52)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:606)

at org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:672)

at org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:180)

at org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:205)

at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:120)

at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)

Caused by: java.lang.IllegalArgumentException: java.net.UnknownHostException: hdfs

at org.apache.hadoop.security.SecurityUtil.buildTokenService(SecurityUtil.java:377)

at org.apache.hadoop.hdfs.NameNodeProxies.createNonHAProxy(NameNodeProxies.java:240)

at org.apache.hadoop.hdfs.NameNodeProxies.createProxy(NameNodeProxies.java:144)

at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:579)

at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:524)

at org.apache.hadoop.hdfs.DistributedFileSystem.initialize(DistributedFileSystem.java:146)

at org.apache.hadoop.fs.FileSystem.createFileSystem(FileSystem.java:2397)

at org.apache.hadoop.fs.FileSystem.access$200(FileSystem.java:89)

at org.apache.hadoop.fs.FileSystem$Cache.getInternal(FileSystem.java:2431)

at org.apache.hadoop.fs.FileSystem$Cache.get(FileSystem.java:2413)

at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:368)

at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:167)

at org.apache.hadoop.mapred.JobConf.getWorkingDirectory(JobConf.java:653)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:427)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:400)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$32.apply(SparkContext.scala:1007)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$32.apply(SparkContext.scala:1007)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:176)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:176)

at scala.Option.map(Option.scala:145)

at org.apache.spark.rdd.HadoopRDD.getJobConf(HadoopRDD.scala:176)

at org.apache.spark.rdd.HadoopRDD$$anon$1.<init>(HadoopRDD.scala:220)

at org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:216)

at org.apache.spark.rdd.HadoopRDD.compute(HadoopRDD.scala:101)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scala:38)

at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:297)

at org.apache.spark.rdd.RDD.iterator(RDD.scala:264)

at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:73)

at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.scala:41)

at org.apache.spark.scheduler.Task.run(Task.scala:88)

at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:214)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)

at java.lang.Thread.run(Thread.java:745)

Caused by: java.net.UnknownHostException: hdfs

... 42 more

15/11/17 16:59:26 INFO RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.

15/11/17 16:59:26 INFO RemoteActorRefProvider$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.

15/11/17 16:59:26 INFO ShutdownHookManager: Shutdown hook called

15/11/17 16:59:26 INFO ShutdownHookManager: Deleting directory /tmp/spark-45b60b79-ff85-4559-aca6-3c8563ee9dfe

15/11/17 16:59:26 INFO RemoteActorRefProvider$RemotingTerminator: Remoting shut down.