Отчёт по заданию №10

По умолчанию команда «sudo apt install default-jdk -y» устанавливает Java 17, которая несовместима с Hadoop – пришлось устанавливать дополнительно Java 11.

Выполнение команд для включения Hadoop:

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

Описанная в задании команда для генерирования классов не работала, поэтому я использовал команду «/usr/lib/jvm/java-11-openjdk-amd64/bin/javac WordCount.java -cp $(hadoop classpath))» и переместил сгенерированные классы в папку tutorial\_classes.

Описанная в задании команда для создания JAR-файл из классов не сработала, поэтому я использовал команду «jar -cvf firstTutorial.jar -C tutorial\_classes .»

Список файлов в папке WordCountTutorial:

Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание

Создание каталогов в Hadoop:



Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание

Проблема при загрузке файла input.txt:

Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание

Через команду также не получается. Вывод в терминале:

shara@shara-VMware-Virtual-Platform:~/Documents/WordCountTutorial$ hadoop fs -put input\_data/**input**.txt /WordCountTutorial/Input  
2023-06-05 03:01:15,888 WARN hdfs.DataStreamer: DataStreamer Exception  
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /WordCountTutorial/Input/**input**.txt.\_COPYING\_ could only be written to 0 of the 1 minReplication nodes. There are 0 datanode(s) running and 0 node(s) are excluded **in** this operation.  
 at org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarget4NewBlock(BlockManager.java:2329)  
 at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.chooseTargetForNewBlock(FSDirWriteFileOp.java:294)  
 at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock(FSNamesystem.java:2942)  
 at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(NameNodeRpcServer.java:915)  
 at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:593)  
 at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:600)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:568)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:552)  
 at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1093)  
 at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:1035)  
 at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:963)  
 at java.base/java.security.AccessController.doPrivileged(Native Method)  
 at java.base/javax.security.auth.Subject.doAs(Subject.java:423)  
 at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)  
 at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2966)  
  
 at org.apache.hadoop.ipc.Client.getRpcResponse(Client.java:1573)  
 at org.apache.hadoop.ipc.Client.call(Client.java:1519)  
 at org.apache.hadoop.ipc.Client.call(Client.java:1416)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:242)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:129)  
 at com.sun.proxy.$Proxy12.addBlock(Unknown Source)  
 at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.addBlock(ClientNamenodeProtocolTranslatorPB.java:530)  
 at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)  
 at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)  
 at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)  
 at java.base/java.lang.reflect.Method.invoke(Method.java:566)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:422)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandler.java:165)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryInvocationHandler.java:157)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(RetryInvocationHandler.java:95)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:359)  
 at com.sun.proxy.$Proxy13.addBlock(Unknown Source)  
 at org.apache.hadoop.hdfs.DFSOutputStream.addBlock(DFSOutputStream.java:1084)  
 at org.apache.hadoop.hdfs.DataStreamer.locateFollowingBlock(DataStreamer.java:1898)  
 at org.apache.hadoop.hdfs.DataStreamer.nextBlockOutputStream(DataStreamer.java:1700)  
 at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:707)  
put: File /WordCountTutorial/Input/**input**.txt.\_COPYING\_ could only be written to 0 of the 1 minReplication nodes. There are 0 datanode(s) running and 0 node(s) are excluded **in** this operation.

Следующая команда не сработала:

hadoop jar '/home/shara/Documents/WordCountTutorial/firstTutorial.jar' WordCount /WordCountTutorial/Input /WordCountTutorial/Output

Результат вывода в терминал:

shara@shara-VMware-Virtual-Platform:~$ hadoop jar '/home/shara/Documents/WordCountTutorial/firstTutorial.jar' WordCount /WordCountTutorial/Input /WordCountTutorial/Output  
2023-06-05 02:32:40,172 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032  
2023-06-05 02:32:40,487 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
2023-06-05 02:32:40,499 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding **for** path: /tmp/hadoop-yarn/staging/shara/.staging/job\_1685921335888\_0001  
2023-06-05 02:32:40,625 WARN hdfs.DataStreamer: DataStreamer Exception  
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /tmp/hadoop-yarn/staging/shara/.staging/job\_1685921335888\_0001/job.jar could only be written to 0 of the 1 minReplication nodes. There are 0 datanode(s) running and 0 node(s) are excluded **in** this operation.  
 at org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarget4NewBlock(BlockManager.java:2329)  
 at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.chooseTargetForNewBlock(FSDirWriteFileOp.java:294)  
 at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock(FSNamesystem.java:2942)  
 at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(NameNodeRpcServer.java:915)  
 at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:593)  
 at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:600)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:568)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:552)  
 at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1093)  
 at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:1035)  
 at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:963)  
 at java.base/java.security.AccessController.doPrivileged(Native Method)  
 at java.base/javax.security.auth.Subject.doAs(Subject.java:423)  
 at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)  
 at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2966)  
  
 at org.apache.hadoop.ipc.Client.getRpcResponse(Client.java:1573)  
 at org.apache.hadoop.ipc.Client.call(Client.java:1519)  
 at org.apache.hadoop.ipc.Client.call(Client.java:1416)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:242)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:129)  
 at com.sun.proxy.$Proxy12.addBlock(Unknown Source)  
 at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.addBlock(ClientNamenodeProtocolTranslatorPB.java:530)  
 at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)  
 at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)  
 at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)  
 at java.base/java.lang.reflect.Method.invoke(Method.java:566)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:422)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandler.java:165)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryInvocationHandler.java:157)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(RetryInvocationHandler.java:95)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:359)  
 at com.sun.proxy.$Proxy13.addBlock(Unknown Source)  
 at org.apache.hadoop.hdfs.DFSOutputStream.addBlock(DFSOutputStream.java:1084)  
 at org.apache.hadoop.hdfs.DataStreamer.locateFollowingBlock(DataStreamer.java:1898)  
 at org.apache.hadoop.hdfs.DataStreamer.nextBlockOutputStream(DataStreamer.java:1700)  
 at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:707)  
2023-06-05 02:32:40,630 INFO mapreduce.JobSubmitter: Cleaning up the staging area /tmp/hadoop-yarn/staging/shara/.staging/job\_1685921335888\_0001  
Exception **in** thread "main" org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /tmp/hadoop-yarn/staging/shara/.staging/job\_1685921335888\_0001/job.jar could only be written to 0 of the 1 minReplication nodes. There are 0 datanode(s) running and 0 node(s) are excluded **in** this operation.  
 at org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarget4NewBlock(BlockManager.java:2329)  
 at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.chooseTargetForNewBlock(FSDirWriteFileOp.java:294)  
 at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock(FSNamesystem.java:2942)  
 at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(NameNodeRpcServer.java:915)  
 at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:593)  
 at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:600)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:568)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine2.java:552)  
 at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1093)  
 at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:1035)  
 at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:963)  
 at java.base/java.security.AccessController.doPrivileged(Native Method)  
 at java.base/javax.security.auth.Subject.doAs(Subject.java:423)  
 at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)  
 at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2966)  
  
 at org.apache.hadoop.ipc.Client.getRpcResponse(Client.java:1573)  
 at org.apache.hadoop.ipc.Client.call(Client.java:1519)  
 at org.apache.hadoop.ipc.Client.call(Client.java:1416)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:242)  
 at org.apache.hadoop.ipc.ProtobufRpcEngine2$Invoker.invoke(ProtobufRpcEngine2.java:129)  
 at com.sun.proxy.$Proxy12.addBlock(Unknown Source)  
 at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.addBlock(ClientNamenodeProtocolTranslatorPB.java:530)  
 at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)  
 at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)  
 at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)  
 at java.base/java.lang.reflect.Method.invoke(Method.java:566)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:422)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(RetryInvocationHandler.java:165)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryInvocationHandler.java:157)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(RetryInvocationHandler.java:95)  
 at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:359)  
 at com.sun.proxy.$Proxy13.addBlock(Unknown Source)  
 at org.apache.hadoop.hdfs.DFSOutputStream.addBlock(DFSOutputStream.java:1084)  
 at org.apache.hadoop.hdfs.DataStreamer.locateFollowingBlock(DataStreamer.java:1898)  
 at org.apache.hadoop.hdfs.DataStreamer.nextBlockOutputStream(DataStreamer.java:1700)  
 at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:707)

Папка Output не появилась:

Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание