



## Dúvidas - Aula 11

6 47

[Todas as turmas](#)

Sem mensagem

Não lido

✓ **Assinar**[← Responder](#)[delanosilva\\_182@hotmail.com](mailto:delanosilva_182@hotmail.com) (<https://academy.semantix.com.br/courses/108/users/5706>)

Terça-feira



Professor boa tarde! Estou com um problema na URL nesse comando: `curl -o https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/parquet-hadoop-bundle-1.6.0.jar` `_(https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/parquet-hadoop-bundle-1.6.0.jar)`

Fica aparecendo: `curl: no URL specified!`

Se alguém puder ajudar fico agradecido.

[← Responder](#) (1 curtida)[mauriciochiarante@yahoo.com.br](mailto:mauriciochiarante@yahoo.com.br) (<https://academy.semantix.com.br/courses/108/users/2013>)

Terça-feira



Tenta utilizar "-O" em maiúsculo, tem comandos que são sensitive case.

Eu estava com outro problema, mas acabei fazendo o download do arquivo e jogando no linux.

Editado por [mauriciochiarante@yahoo.com.br](mailto:mauriciochiarante@yahoo.com.br) (<https://academy.semantix.com.br/courses/108/users/2013>) em 15 Fev em 17:31

[← Responder](#) (2 curtidas)



**Rodrigo Augusto Rebouças** (<https://academy.semantix.com.br/courses/108/users/5>)

Terça-feira



O link está funcionando, tente baixar manualmente.

← [Responder](#) 👍 (1 curtida)



**delanosilva\_182@hotmail.com** (<https://academy.semantix.com.br/courses/108/users/5706>)

Terça-feira



Obrigado Rodrigo!

← [Responder](#) 👍 (1 curtida)



**gabrielgustavo96@gmail.com** (<https://academy.semantix.com.br/courses/108/users/5648>)

Terça-feira



Em qual diretório se joga o arquivo baixado manualmente ? o link ta dando connection timed out pra mim

← [Responder](#) 👍 (1 curtida)



**dianam071@gmail.com** (<https://academy.semantix.com.br/courses/108/users/4542>)

Quarta-feira



Gabriel entra no ubuntu e coloca os seguintes comandos

cd docker-bigdata #entra na pasta onde vc tem se docker

explorer.exe #este comando vai abrir uma pasta no windows onde esta seu cluster, dai copia e pega o arquivo nessa pasta.

se nao der certo vai no explorador de pastas do windows e procura algo do tipo wsl entra nele e procura o ubuntu/root/docker-bigdata.

Me fala se funciona.

← [Responder](#) 👍 (2 curtidas)



**Rodrigo Augusto Rebouças** (<https://academy.semantix.com.br/courses/108/users/5>),

Quarta-feira



Obrigado pela ajuda

← [Responder](#) 👍



**delanosilva\_182@hotmail.com** (<https://academy.semantix.com.br/courses/108/users/5706>),

Terça-feira



Muito obrigado colega, deu certo aqui.

Voltamos ao treinamento!

← [Responder](#) 👍 (2 curtidas)



**souza.chagas.felipe@gmail.com** (<https://academy.semantix.com.br/courses/108/users/5584>),

Terça-feira



Olá Professor e pessoal, boa noite!

Consegui resolver todos os exercícios porém enfrentei algumas vezes o problema:

```
/opt/spark/bin/spark-shell: line 47: 391 Killed                  "${SPARK_HOME}"/bin/spark-submit --class org.apache.spark.repl.Main --name "Spark shell" "$@"
```

Precisei rodar o novamente spark-shell e refazer alguns passos.

O que eu fiz para tentar evitar o erro:

- Abri o arquivo docker-compose.yml
  - Fui ao spark e aloquei 5g de memória (Meu Cluster tem 32Gb de memória ram)

Mesmo feito isso ainda tenho o kill do processo.

Também dei uma olhada no log do spark através do docker e não identifiquei nenhuma mensagem de erro.

O que mais posso fazer para evitar este problema?

Desde já agradeço.

← [Responder](#)  (1 curtida)



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>)

Terça-feira



Opa,

Consegui resolver:

- Precisei alocar mais memória para o Spark
- rodar o comando ***docker-compose up -d***

Após isso não tive mais problemas, apaguei a estrutura criada e recriei tudo para confirmar se estava ok e fiz tudo na mesma sessão do spark sem ter problema de queda do serviço.

← [Responder](#)  (1 curtida)



[Rodrigo Augusto Rebouças](https://academy.semantix.com.br/courses/108/users/5) (<https://academy.semantix.com.br/courses/108/users/5>)

Quarta-feira



Nossa, com 5 GB estava caindo? Para quanto você aumentou?

← [Responder](#) 



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>)

Quarta-feira



Tive que aumentar para 6GB, fiquei chocado, e quando eu subo o cluster e entro no spark e rodo algo, o consumo dele não chega nem perto do limite que coloquei.

O mais complicado foi não encontrar nada nos logs

No dia fiz testes com memória swap

← [Responder](#)  (1 curtida)



[rafaelborges.engcomp@gmail.com](mailto:rafaelborges.engcomp@gmail.com)

<https://academy.semantix.com.br/courses/108/users/5667>

Quinta-feira



Você rodou o comando ***docker-compose up -d*** quando alocou os 5GB?  
Ou rodou somente quando alocou os 6GB?

← [Responder](#) 👍 (1 curtida)



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com)

<https://academy.semantix.com.br/courses/108/users/5584>

Sábado



Rodei nos dois casos

← [Responder](#) 👍



[alexrvale@gmail.com](mailto:alexrvale@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5928>)

Terça-feira



quando utilizo o `DF.write.parquet("caminho")`, por padrão ele irá sobrescrever o arquivo que eu tiver no diretório ou ele faz um "append" e cria um novo arquivo?

Editado por [alexrvale@gmail.com](mailto:alexrvale@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5928>) em 15 Fev em 23:04

← [Responder](#) 👍 (1 curtida)



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>)

Quarta-feira



Olá Alex, boa tarde!

Pelo que entendi (salvo engano) ele sobrescreve o arquivo sim, caso seja o mesmo nome e local.

Consultando a documentação, para realizar o append é necessário utilizar o recurdo `".mode()"`, segue link com exemplo:

**Spark By Examples:**

- <https://sparkbyexamples.com/spark/spark-read-write-dataframe-parquet-example/> [\(https://sparkbyexamples.com/spark/spark-read-write-dataframe-parquet-example/\)](https://sparkbyexamples.com/spark/spark-read-write-dataframe-parquet-example/)
  - Busque pela sessão "Append to existing Parquet file"

### Documentação Spark - Save Modes:

- <https://spark.apache.org/docs/latest/sql-data-sources-load-save-functions.html#save-modes> [\(https://spark.apache.org/docs/latest/sql-data-sources-load-save-functions.html#save-modes\)](https://spark.apache.org/docs/latest/sql-data-sources-load-save-functions.html#save-modes)

Espero ter ajudado.

Abs

Editado por [souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) <https://academy.semantix.com.br/courses/108/users/5584> em 16 Feb em 13:44

← [Responder](#)  (1 curtida)

○



[Rodrigo Augusto Rebouças](https://academy.semantix.com.br/courses/108/users/5) [\(https://academy.semantix.com.br/courses/108/users/5\)](https://academy.semantix.com.br/courses/108/users/5)

Quarta-feira



Obrigado pela ajuda e referências.

← [Responder](#)  (1 curtida)

○



[carlospaulo4@gmail.com](mailto:carlospaulo4@gmail.com) <https://academy.semantix.com.br/courses/108/users/5924>

Terça-feira



Alguém poderia me informar onde fica o formulário de Conclusão do módulo para preenchimento?

← [Responder](#)  (2 curtidas)

○



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) <https://academy.semantix.com.br/courses/108/users/5584>

Quarta-feira



Olá Carlos

Pelo que entendi através da explicação do professor @Rodrigo, o formulário de conclusão é destinado para os alunos que já concluíram em turmas anteriores a atual.

O link se encontra na Aula 0 - Apresentação > Conclusão do Módulo - Big Data Foundations (conforme anexo)

Caso queira acessar diretamente, segue o link:

<https://academy.semantix.com.br/courses/108/modules/items/4864>  
(<https://academy.semantix.com.br/courses/108/modules/items/4864>)

Lembrando que o formulário deveria ser preenchido até ontem (16/02)

Espero ter ajudado.

Abs.

[Captura.png](#) ([https://academy.semantix.com.br/files/6685/download?download\\_frd=1&verifier=JrLMyLAlhpxju1dpCXLZ34bJ41RR2DaUlaaP8oWW](https://academy.semantix.com.br/files/6685/download?download_frd=1&verifier=JrLMyLAlhpxju1dpCXLZ34bJ41RR2DaUlaaP8oWW))

← [Responder](#) 🍌 (1 curtida)

○



[carlospaulo4@gmail.com](mailto:carlospaulo4@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5924>)

Quarta-feira



Muito obrigado pela explica e ajuda.

← [Responder](#) 🍌 (1 curtida)

○



[julianabicalhodev@gmail.com](mailto:julianabicalhodev@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5832>)

Quarta-feira



Não consigo baixar o endereço, dá tempo esgotado. Não sei o que fazer.

← [Responder](#) 🍌 (2 curtidas)

○



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>)

Quarta-feira



Olá Juliana,

Tentou baixar manualmente?

Copie o link e tente acessar pelo seu browser de internet (Google Chrome, Firefox, Edge e tals)

<https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/parquet-hadoop-bundle-1.6.0.jar>

(<https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/parquet-hadoop-bundle-1.6.0.jar>)

Você também pode acessar <https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/> (<https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/>)

e clicar no link do:

[parquet-hadoop-bundle-1.6.0.jar](#)

(<https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/parquet-hadoop-bundle-1.6.0.jar>) 2015-04-25 00:42 2796935

Depois que conseguir baixar, pode copiar o arquivo para o spark a partir da pasta onde consta o

[parquet-hadoop-bundle-1.6.0.jar](#)

(<https://repo1.maven.org/maven2/com/twitter/parquet-hadoop-bundle/1.6.0/parquet-hadoop-bundle-1.6.0.jar>)

Espero ter ajudado

↳ [Responder](#) 🍌 (1 curtida)

○



[renanfsvieira@gmail.com](mailto:renanfsvieira@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5866>)

Quarta-feira



boa tarde, estou com o mesmo erro da Juliana. Como copio o arquivo baixado no Windows para a pasta docker-bigdata no Ubuntu?

↳ [Responder](#) 🍌 (1 curtida)

○



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>)

Quarta-feira





Opa, utilize o comando:

Vamos supor que o arquivo foi baixado em sua pasta "Downloads" no windows:

- `docker cp /mnt/c/users/<seu_nome_de_usuario>/downloads/parquet-hadoop-bundle-1.6.0.jar <caminho_destino_docker-bigdata>`

← [Responder](#) 🍌 (3 curtidas)



**Rodrigo Augusto Rebouças** (<https://academy.semantix.com.br/courses/108/users/5>).

Quarta-feira



Muito obrigado pela ajuda Felipe, está contribuindo bastante no treinamento.

← [Responder](#) 🍌 (2 curtidas)



**renanfsvieira@gmail.com** (<https://academy.semantix.com.br/courses/108/users/5866>)

Quarta-feira



Boa tarde pessoal, estou com o mesmo erro relatado por algumas pessoas no fórum , Não consigo baixar o Jar pelo endereço, dá tempo esgotado na porta 443. como faço o procedimento de baixar e inserir manualmente no Windows? tentei buscar a pasta Wsl ou Ubuntu, porém não encontrei. Alguém poderia ensinar o passo a passo? ou como resolver o erro no ubuntu do comando com curl.

← [Responder](#) 🍌 (1 curtida)



**souza.chagas.felipe@gmail.com** (<https://academy.semantix.com.br/courses/108/users/5584>)

Quarta-feira



Boa tarde, Renan!

Conforme resolvemos pelo telegram, seria legal postar como foi solucionado para quem estiver enfrentando o mesmo problema, saber como resolver.

Acredito que quem está com o ambiente Windows + WSL2 e realizou o download manualmente pelo windows, pode copiar o arquivo pelo docker através do caminho completo do windows:

Exemplo:

- Dentro do Ubuntu (WSL2):

```
docker cp /mnt/c/users/<seu_nome_de_usuario>/downloads/parquet-hadoop-bundle-1.6.0.jar <caminho_destino>
```

Caso tenha realizado algum passo a mais, poderia informar os passos que fez para resolver?

Espero ter ajudado.

Editado por [souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>) em 16 Feb em 15:59

← [Responder](#) 🍌 (2 curtidas)



[renanfsvieira@gmail.com](mailto:renanfsvieira@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5866>)

Quarta-feira



Boa tarde, tinha postado a dúvida antes de realizarmos esse procedimento. Galera quem estiver recebendo a mensagem de erro ao executar o comando Curl, pode seguir esse tutorial do Felipe. Baixa o jar pelo navegador e copia o caminho do arquivo no Windows e copia no wsl2: `docker cp /mnt/c/users/<seu_nome_de_usuario>/downloads/parquet-hadoop-bundle-1.6.0.jar spark:/opt/spark/jars`

Editado por [renanfsvieira@gmail.com](mailto:renanfsvieira@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5866>) em 16 Feb em 16:41

← [Responder](#) 🍌 (2 curtidas)



[Rodrigo Augusto Rebouças](mailto:Rodrigo Augusto Rebouças) (<https://academy.semantix.com.br/courses/108/users/5>)

Quarta-feira



Obrigado pela ajuda.

← [Responder](#) 🍌 (1 curtida)



[rodsund2013@gmail.com](mailto:rodsund2013@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5574>)



Quarta-feira

Professor onde está a avaliação ? não encontrei no site

← [Responder](#) 🍌 (2 curtidas)



[Rodrigo Augusto Rebouças \(https://academy.semantix.com.br/courses/108/users/5\)](https://academy.semantix.com.br/courses/108/users/5)

Quarta-feira

Está abaixo do módulo 12.

← [Responder](#) 🍌



[lucascesar\\_35@hotmail.com \(https://academy.semantix.com.br/courses/108/users/5762\)](https://academy.semantix.com.br/courses/108/users/5762)

Quarta-feira

Preenchi sem querer o formulário pra quem já tinha feito o treinamento, não li direito. Tem algum email que eu possa enviar pra cancelar? Não sei se tem algum problema, etc.

← [Responder](#) 🍌 (1 curtida)



[Rodrigo Augusto Rebouças \(https://academy.semantix.com.br/courses/108/users/5\)](https://academy.semantix.com.br/courses/108/users/5)

Quarta-feira

Não precisa cancelar, só seguir o treinamento.

← [Responder](#) 🍌



[dianam071@gmail.com \(https://academy.semantix.com.br/courses/108/users/4542\)](https://academy.semantix.com.br/courses/108/users/4542)

Quarta-feira

Boa Tarde,

Alguem sabe como recuperar um file do hdfs, tente procurar minha trash com este comando

/user/<user-name>/.Trash mas nao deu certo

← [Responder](#) 🍌 (1 curtida)



[souza.chagas.felipe@gmail.com \(https://academy.semantix.com.br/courses/108/users/5584\)](https://academy.semantix.com.br/courses/108/users/5584)

Quarta-feira



Olá, boa noite Diana

Veja se algum dos artigos de ajuda:

Neste contém os passos para encontrar sua trash e como copiar

<https://stackoverflow.com/questions/51267089/how-to-restore-a-deleted-folder-from-hdfs> [\\_\(https://stackoverflow.com/questions/51267089/how-to-restore-a-deleted-folder-from-hdfs\)](https://stackoverflow.com/questions/51267089/how-to-restore-a-deleted-folder-from-hdfs)

Este possui uma explicação mais ampla, inclusive consta que tem um tempo configurado para o hadoop fazer expurgo do conteúdo da trash:

<https://www.linkedin.com/pulse/recovering-deleted-hdfs-files-cloudera-certified-developer-hadoop-> [\\_\(https://www.linkedin.com/pulse/recovering-deleted-hdfs-files-cloudera-certified-developer-hadoop-\)](https://www.linkedin.com/pulse/recovering-deleted-hdfs-files-cloudera-certified-developer-hadoop-)

Espero ter ajudado.

← [Responder](#)  (1 curtida)



[Rodrigo Augusto Rebouças](https://academy.semantix.com.br/courses/108/users/5) [\\_\(https://academy.semantix.com.br/courses/108/users/5\)](https://academy.semantix.com.br/courses/108/users/5)

Sexta-feira



Obrigado pela ajuda Felipe, em produção é recomendável sempre ter um tempo para a lixeira.

← [Responder](#)  (1 curtida)



[ronnan\\_ok@hotmail.com](https://academy.semantix.com.br/courses/108/users/5610) [\\_\(https://academy.semantix.com.br/courses/108/users/5610\)](https://academy.semantix.com.br/courses/108/users/5610)

Quarta-feira



Não consigo copiar o arquivo para a pasta

```
ronnan@ronnan-virtual-machine:~/docker-bigdata$ docker exec -it namenode bash
root@namenode:/# hdfs dfs -put /input/exercises-data/juros_selic/ /user/aluno/ronnan/data
put: Cannot create file/user/aluno/ronnan/data/juros_selic/juros_selic._COPYING_. Name
node is in safe mode.
put: Cannot create file/user/aluno/ronnan/data/juros_selic/juros_selic.json._COPYING_.
Name node is in safe mode.
```

put: Cannot create file/user/aluno/ronnan/data/juros\_selic/juros\_selic.wsdl.\_COPYING\_.  
Name node is in safe mode.

## STATUS DOS CONTAINERS

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS
	PORTS			
	NAMES			
40b0475b2132	fjardim/hive	"entrypoint.sh /bin/..."	11 days ago	Up About a minute
	0.0.0.0:10000->10000/tcp, :::10000->10000/tcp, 10002/tcp	hive-server		
a552c299b8af	fjardim/hive	"entrypoint.sh /opt/..."	11 days ago	Up About a minute
	10000/tcp, 0.0.0.0:9083->9083/tcp, :::9083->9083/tcp, 10002/tcp	hive_metastore		
73467fee4f47	fjardim/hive-metastore	"/docker-entrypoint..."	11 days ago	Up About a minute
	5432/tcp	hive-metastore-postgresql		
871eac2f64a8	fjardim/hbase-master	"/entrypoint.sh /run..."	11 days ago	Up About a minute
	16000/tcp, 0.0.0.0:16010->16010/tcp, :::16010->16010/tcp	hbase-master		
a7f3efaabe18	fjardim/datanode	"/entrypoint.sh /run..."	11 days ago	Up About a minute <b>(unhealthy)</b>
	0.0.0.0:50075->50075/tcp, :::50075->50075/tcp	datanode		
766131bff200	fjardim/jupyter-spark	"/opt/docker/bin/ent..."	11 days ago	Up About a minute
	0.0.0.0:4040-4043->4040-4043/tcp, :::4040-4043->4040-4043/tcp, 0.0.0.0:8889->8889/tcp, :::8889->8889/tcp, 8899/tcp	spark		
9c96d757005e	fjardim/mysql	"docker-entrypoint.s..."	11 days ago	Up About a minute
	33060/tcp, 0.0.0.0:33061->3306/tcp, :::33061->3306/tcp	database		
7c61cf551f69	fjardim/namenode_sqoop	"/entrypoint.sh /run..."	11 days ago	Up About a minute <b>(healthy)</b>
	0.0.0.0:50070->50070/tcp, :::50070->50070/tcp	namenode		
b9844881a867	fjardim/zookeeper	"/bin/sh -c '/usr/sb..."	11 days ago	Up About a minute
	22/tcp, 2888/tcp, 3888/tcp, 0.0.0.0:2181->2181/tcp, :::2181->2181/tcp	zookeeper		

← Responder 🍌 (1 curtida)



[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5584>)

Quinta-feira



Olá Ronan,

Para sair do Safe Mode, utilize o comando:

**`bin/hadoop dfsadmin -safemode leave`**

Caso esteja no namenode, acredito que possa rodar:

**`dfsadmin -safemode leave`**

Referência:

**<https://hadoop.apache.org/docs/r2.4.1/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html#Safemode>**

Espero ter ajudado

← [Responder](#) 🍌 (2 curtidas)



[ronnan\\_ok@hotmail.com](mailto:ronnan_ok@hotmail.com) (<https://academy.semantix.com.br/courses/108/users/5610>)

Quinta-feira



Obrigado, chegando em casa já testo haha!

← [Responder](#) 🍌 (1 curtida)



[ronnan\\_ok@hotmail.com](mailto:ronnan_ok@hotmail.com) (<https://academy.semantix.com.br/courses/108/users/5610>)

Quinta-feira



tentei mas depois de sair do modo safe, dei esse erro:

```
2/02/17 23:25:15 WARN hdfs.DFSCliet: DataStreamer Exception
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File
/user/aluno/ronnan/data/juros_selic/juros_selic._COPYING_ could only be
replicated to 0 nodes instead of minReplication (=1). There are 0 datanode(s)
running and no node(s) are excluded in this operation.
```

at

```
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarg
et4NewBlock(BlockManager.java:1628)
```

```
at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getNewBlockTargets(FSNamesystem.java:3121)
at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock(FSNamesystem.java:3045)
at
org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(NameNodeRpcServer.java:725)
at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:493)
at
org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)
at
org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:616)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:982)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2217)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2213)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1746)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2213)

    at org.apache.hadoop.ipc.Client.call(Client.java:1476)
    at org.apache.hadoop.ipc.Client.call(Client.java:1413)
    at
org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:229)
    at com.sun.proxy.$Proxy10.addBlock(Unknown Source)
    at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.addBlock(ClientNamenodeProtocolTranslatorPB.java:418)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessor
```

```
Impl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at
org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvoca
tionHandler.java:191)
    at
org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHa
ndler.java:102)
    at com.sun.proxy.$Proxy11.addBlock(Unknown Source)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.locateFollowingBlo
ck(DFSOutputStream.java:1588)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputSt
ream(DFSOutputStream.java:1373)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStr
eam.java:554)
put: File /user/aluno/ronnan/data/juros_selic/juros_selic._COPYING_ could
only be replicated to 0 nodes instead of minReplication (=1). There are 0
datanode(s) running and no node(s) are excluded in this operation.
22/02/17 23:25:15 WARN hdfs.DFSClient: DataStreamer Exception
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File
/user/aluno/ronnan/data/juros_selic/juros_selic.json._COPYING_ could only
be replicated to 0 nodes instead of minReplication (=1). There are 0
datanode(s) running and no node(s) are excluded in this operation.
    at
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarg
et4NewBlock(BlockManager.java:1628)
    at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getNewBlockTarge
ts(FSNamesystem.java:3121)
    at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock
(FSNamesystem.java:3045)
    at
org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(Na
meNodeRpcServer.java:725)
    at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTran
slatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:493)
    at
org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$Client
```



```
NamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)
a)
    at
org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(
ProtobufRpcEngine.java:616)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:982)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2217)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2213)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformatio
n.java:1746)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2213)

    at org.apache.hadoop.ipc.Client.call(Client.java:1476)
    at org.apache.hadoop.ipc.Client.call(Client.java:1413)
    at
org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngin
e.java:229)
    at com.sun.proxy.$Proxy10.addBlock(Unknown Source)
    at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.ad
dBlock(ClientNamenodeProtocolTranslatorPB.java:418)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.jav
a:62)
    at
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessor
Impl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at
org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvoca
tionHandler.java:191)
    at
org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHa
ndler.java:102)
    at com.sun.proxy.$Proxy11.addBlock(Unknown Source)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.locateFollowingBlo
ck(DFSOutputStream.java:1588)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputSt
```

```
ream(DFSOutputStream.java:1373)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStr
eam.java:554)
put: File /user/aluno/ronnan/data/juros_selic/juros_selic.json._COPYING_
could only be replicated to 0 nodes instead of minReplication (=1). There are
0 datanode(s) running and no node(s) are excluded in this operation.
22/02/17 23:25:15 WARN hdfs.DFSClient: DataStreamer Exception
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File
/user/aluno/ronnan/data/juros_selic/juros_selic.wSDL._COPYING_ could only
be replicated to 0 nodes instead of minReplication (=1). There are 0
datanode(s) running and no node(s) are excluded in this operation.
    at
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarg
et4NewBlock(BlockManager.java:1628)
    at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getNewBlockTarge
ts(FSNamesystem.java:3121)
    at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock
(FSNamesystem.java:3045)
    at
org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(Na
meNodeRpcServer.java:725)
    at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTran
slatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:493)
    at
org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$Client
NamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.jav
a)
    at
org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(
ProtobufRpcEngine.java:616)
        at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:982)
        at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2217)
        at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2213)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformatio
n.java:1746)
            at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2213)
```

```

at org.apache.hadoop.ipc.Client.call(Client.java:1476)
at org.apache.hadoop.ipc.Client.call(Client.java:1413)
at
org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:229)
at com.sun.proxy.$Proxy10.addBlock(Unknown Source)
at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.addBlock(ClientNamenodeProtocolTranslatorPB.java:418)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at
org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:191)
at
org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:102)
at com.sun.proxy.$Proxy11.addBlock(Unknown Source)
at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.locateFollowingBlock(DFSOutputStream.java:1588)
at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputStream(DFSOutputStream.java:1373)
at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:554)
put: File /user/aluno/ronnan/data/juros_selic/juros_selic.wSDL._COPYING_
could only be replicated to 0 nodes instead of minReplication (=1). There are
0 datanode(s) running and no node(s) are excluded in this operation.

```

← Responder 🖱️ (1 curta)



[ronnan\\_ok@hotmail.com](mailto:ronnan_ok@hotmail.com) (<https://academy.semantix.com.br/courses/108/users/5610>)

Quinta-feira

Aumentei a RAM e o espaço da VM deu certo, vai entender... Obrigado!

← Responder 👍 (1 curtida)

(http)

[souza.chagas.felipe@gmail.com](mailto:souza.chagas.felipe@gmail.com)

(<https://academy.semantix.com.br/courses/108/users/5584>)

Quinta-feira

Boa!

Que bom que deu certo!

← Responder 👍 (2 curtidas)

(http)

[Rodrigo Augusto Rebouças](#)

(<https://academy.semantix.com.br/courses/108/users/5>)

Sexta-feira

Que bom que funcionou, aumentou para quanto?

← Responder 👍 (1 curtida)



[marqueslmp@yahoo.com.br](mailto:marqueslmp@yahoo.com.br) (<https://academy.semantix.com.br/courses/108/users/5934>)

Sexta-feira

Boa tarde.

Obtive o seguinte erro ao tentar usar o .write.saveAsTable....

```
scala> val jurosDF10 = jurosDF.where("valor > 10")
jurosDF10: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [data: string, valor: string]
```

```
scala> jurosDF10.write.saveAsTable("Imp.tab_juros_selic")
java.lang.NoClassDefFoundError: parquet/hadoop/ParquetOutputFormat
```

```
at java.lang.Class.forName0(Native Method)
at java.lang.Class.forName(Class.java:348)
at org.apache.spark.util.Utils$.classForName(Utils.scala:238)
at
org.apache.spark.sql.hive.client.HiveClientImpl$.org$apache$spark$sql$hive$client$HiveC
lientImpl$$toOutputFormat(HiveClientImpl.scala:913)
at
org.apache.spark.sql.hive.client.HiveClientImpl$$anonfun$toHiveTable$6.apply(HiveClientI
mpl.scala:947)
at
org.apache.spark.sql.hive.client.HiveClientImpl$$anonfun$toHiveTable$6.apply(HiveClientI
mpl.scala:947)
at scala.Option.map(Option.scala:146)
at org.apache.spark.sql.hive.client.HiveClientImpl$.toHiveTable(HiveClientImpl.scala:947)
at
org.apache.spark.sql.hive.client.HiveClientImpl$$anonfun$createTable$1.apply$mcV$sp(Hi
veClientImpl.scala:482)
at
org.apache.spark.sql.hive.client.HiveClientImpl$$anonfun$createTable$1.apply(HiveClientI
mpl.scala:480)
at
org.apache.spark.sql.hive.client.HiveClientImpl$$anonfun$createTable$1.apply(HiveClientI
mpl.scala:480)
at
org.apache.spark.sql.hive.client.HiveClientImpl$$anonfun$withHiveState$1.apply(HiveClie
ntImpl.scala:275)
at org.apache.spark.sql.hive.client.HiveClientImpl.liftedTree1$1(HiveClientImpl.scala:213)
at org.apache.spark.sql.hive.client.HiveClientImpl.retryLocked(HiveClientImpl.scala:212)
at
org.apache.spark.sql.hive.client.HiveClientImpl.withHiveState(HiveClientImpl.scala:258)
at org.apache.spark.sql.hive.client.HiveClientImpl.createTable(HiveClientImpl.scala:480)
at
org.apache.spark.sql.hive.HiveExternalCatalog.saveTableIntoHive(HiveExternalCatalog.sc
ala:499)
at
org.apache.spark.sql.hive.HiveExternalCatalog.org$apache$spark$sql$hive$HiveExternal
Catalog$$createDataSourceTable(HiveExternalCatalog.scala:387)
at
org.apache.spark.sql.hive.HiveExternalCatalog$$anonfun$createTable$1.apply$mcV$sp(H
iveExternalCatalog.scala:263)
at
org.apache.spark.sql.hive.HiveExternalCatalog$$anonfun$createTable$1.apply(HiveExtern
alCatalog.scala:236)
```

```
    at
org.apache.spark.sql.hive.HiveExternalCatalog$$anonfun$createTable$1.apply(HiveExtern
alCatalog.scala:236)
    at
org.apache.spark.sql.hive.HiveExternalCatalog.withClient(HiveExternalCatalog.scala:97)
    at
org.apache.spark.sql.hive.HiveExternalCatalog.createTable(HiveExternalCatalog.scala:236
)
    at
org.apache.spark.sql.catalyst.catalog.ExternalCatalogWithListener.createTable(ExternalCat
alogWithListener.scala:94)
    at
org.apache.spark.sql.catalyst.catalog.SessionCatalog.createTable(SessionCatalog.scala:3
24)
    at
org.apache.spark.sql.execution.command.CreateDataSourceTableAsSelectCommand.run(
createDataSourceTables.scala:185)
    at
org.apache.spark.sql.execution.command.DataWritingCommandExec.sideEffectResult$lzy
compute(commands.scala:104)
    at
org.apache.spark.sql.execution.command.DataWritingCommandExec.sideEffectResult(co
mmands.scala:102)
    at
org.apache.spark.sql.execution.command.DataWritingCommandExec.doExecute(comman
ds.scala:122)
    at
org.apache.spark.sql.execution.SparkPlan$$anonfun$execute$1.apply(SparkPlan.scala:13
1)
    at
org.apache.spark.sql.execution.SparkPlan$$anonfun$execute$1.apply(SparkPlan.scala:12
7)
    at
org.apache.spark.sql.execution.SparkPlan$$anonfun$executeQuery$1.apply(SparkPlan.sc
ala:155)
    at
org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:151)
    at org.apache.spark.sql.execution.SparkPlan.executeQuery(SparkPlan.scala:152)
    at org.apache.spark.sql.execution.SparkPlan.execute(SparkPlan.scala:127)
    at
org.apache.spark.sql.execution.QueryExecution.toRdd$lzycompute(QueryExecution.scala:
80)
    at org.apache.spark.sql.execution.QueryExecution.toRdd(QueryExecution.scala:80)
```

```
at
org.apache.spark.sql.DataFrameWriter$$anonfun$runCommand$1.apply(DataFrameWriter
.scala:668)
at
org.apache.spark.sql.DataFrameWriter$$anonfun$runCommand$1.apply(DataFrameWriter
.scala:668)
at
org.apache.spark.sql.execution.SQLExecution$$anonfun$withNewExecutionId$1.apply(SQ
LExecution.scala:78)
at
org.apache.spark.sql.execution.SQLExecution$.withSQLConfPropagated(SQLExecution.sc
ala:125)
at
org.apache.spark.sql.execution.SQLExecution$.withNewExecutionId(SQLExecution.scala:
73)
at org.apache.spark.sql.DataFrameWriter.runCommand(DataFrameWriter.scala:668)
at org.apache.spark.sql.DataFrameWriter.createTable(DataFrameWriter.scala:465)
at org.apache.spark.sql.DataFrameWriter.saveAsTable(DataFrameWriter.scala:444)
at org.apache.spark.sql.DataFrameWriter.saveAsTable(DataFrameWriter.scala:400)
... 49 elided
Caused by: java.lang.ClassNotFoundException: parquet.hadoop.ParquetOutputFormat
at java.net.URLClassLoader.findClass(URLClassLoader.java:382)
at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
at
org.apache.spark.sql.hive.client.IsolatedClientLoader$$anon$1.doLoadClass(IsolatedClient
Loader.scala:226)
at
org.apache.spark.sql.hive.client.IsolatedClientLoader$$anon$1.loadClass(IsolatedClientLo
ader.scala:215)
at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
... 95 more

scala>
```

← [Responder](#) 🍌 (1 curtida)

○



**Rodrigo Augusto Rebouças** (<https://academy.semantix.com.br/courses/108/users/5>)

Sexta-feira



O erro "java.lang.NoClassDefFoundError: parquet/hadoop/ParquetOutputFormat" é porque não está aceitando o formato parquet, você fez a configuração do .jar no spark? Isso eu mostro na aula de inicialização do Spark.

← [Responder](#) 🍌

○

[marqueslmp@yahoo.com.br](mailto:marqueslmp@yahoo.com.br) (<https://academy.semantix.com.br/courses/108/users/5934>)

Sexta-feira



Ok, agora funcionou.

Obrigado!

← [Responder](#) 👍-

○

[izabela.schick@gmail.com](mailto:izabela.schick@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5755>)

Ontem



Gostaria de uma ajuda, estou tendo esse erro (foto anexa) e não estou conseguindo resolver, já verifiquei na pasta Hive e minha tabela está lá!

já desliguei e liguei o cluster, já fiz tudo o que estava no forum e mesmo assim não roda.

Obs. estou refazendo as atividades para fixar e isso aconteceu, da primeira vez deu tudo ok!

Obs2. criei uma nova VM para refazer as atividades justamente para nao dar conflitos.

Obrigada

[erro.jpg](https://academy.semantix.com.br/files/6699/download?erro.jpg) ([https://academy.semantix.com.br/files/6699/download?](https://academy.semantix.com.br/files/6699/download?erro.jpg)

[download\\_frd=1&verifier=r7BNO9OfbZMv3w4ExEkU3C9huz8CjINTt1NQgeIK](https://academy.semantix.com.br/files/6699/download?download_frd=1&verifier=r7BNO9OfbZMv3w4ExEkU3C9huz8CjINTt1NQgeIK))

← [Responder](#) 👍-