



## Dúvidas - Aula 7

15

[Todas as turmas](#)

Sem mensagem

[https://](https://academy.semantix.com.br/courses/108/users/5659)[saulodeboni@gmail.com](mailto:saulodeboni@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5659>)

9 de fev de 2022



Conseguí fazer os exercícios, mas fiquei com duas dúvidas:

1. Qual a diferença entre enviar arquivos usando *docker cp* e *hdfs dfs put*?
2. Quando devo usar o *sqoop*? Porque achei as operações parecidas com o *hive*.

 (5 curtidas)[http](http://)[gregorypierroti@hotmail.com](mailto:gregorypierroti@hotmail.com) (<https://academy.semantix.com.br/courses/108/users/5759>)

9 de fev de 2022



Por coincidencia tenho as mesmas duvidas aqui rsrs.

 (1 curtida)[http](http://)[pierre\\_e6@hotmail.com](mailto:pierre_e6@hotmail.com) (<https://academy.semantix.com.br/courses/108/users/5713>)

10 de fev de 2022



Boa noite Saulo,

o comando *docker cp* não tem relação com o *hdfs*, a operação citada na aula, foi apenas para enviar arquivos de um local para o outro (container database). A grosso modo seria um copiar/colar e depois esses arquivos foram usados para criar a estrutura lógica da base sakila e carregar os dados.

O sqoop é usado para fazer o 'meio de campo' entre um banco de dados relacional (mysql, postgresql, oracle, etc) e o Hadoop. Por exemplo, supondo que na sua empresa tem um sistema web que salva os dados dos clientes em uma base relacional (mysql) e você precisa que esses dados sejam carregados no Hadoop. Nesse caso para levar os dados para o Hadoop, você irá usar o sqoop, ele fará a conexão no mysql, pegar os dados que precisa e carregar no Hadoop

← [Responder](#)  (7 curtidas)



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

10 de fev de 2022



Olá Saulo!

Pelo que entendi (caso eu esteja errado, por favor me corrija):

**Pergunta 1** (*Qual a diferença entre enviar arquivos usando docker cp e hdfs dfs put?*):

- **docker cp:**
  - O comando **\$docker cp** possibilita transferências dentro e fora do contêiner (arquivos/pastas da máquina local <-> arquivos/pastas dentro do contêiner)
- **hdfs dfs -put:**
  - O comando **hdfs dfs -put** possibilita transferências entre arquivos/pastas do contêiner <-> arquivos/pastas do serviço/aplicação que está rodando no contêiner

**Pergunta 2** (*Quando devo usar o sqoop? Porque achei as operações parecidas com o hive*):

- O Hive e Sqoop possuem propostas diferentes:
  - **Hive** - Mais voltada para análise dos dados que estão no HDFS, ou seja, você consegue consultar e manipular com linguagem SQL o conteúdo do HDFS (Tabelas, partições, buckets e etc)
  - **Sqoop** - Tem como proposta fazer importação de bancos de dados relacionais (MySQL, SQL Server, Oracle) para dentro do HDFS, bem como a exportação da base do HDFS para um banco de dados relacional

Então quando você quer analisar e manipular dados dentro do HDFS você usa o Hive.

Quando você tem um banco de dados relacional fora do Hadoop (HDFS) e quer que essa base esteja no seu Cluster Hadoop, para realizar a importação desses bancos de dados relacionais, você utiliza o Sqoop para realizar tanto a importação para o cluster Hadoop, quando a exportação.

Lembrando que conforme explicação do professor, o comum é primeiro importar as bases de dados relacionais para o HDFS para depois realizar o processamento e armazenamento no Hive ou Hbase pois quando os dados são transformados em tabelas no Hive ou Hbase, o inverso, (exportar do Hive, Hbase para a base relacional ) não é possível.

É possível fazer exportação do cluster hadoop para uma base de dados relacional a partir do HDFS.

Espero ter ajudado, abs.

← [Responder](#) 👍 (17 curtidas)



[eduarau73@gmail.com](mailto:eduarau73@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5711>)

10 de fev de 2022

Adorei sua explicação Saulo.

← [Responder](#) 👍 (1 curtida)



[eduarau73@gmail.com](mailto:eduarau73@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5711>)

10 de fev de 2022

Corrigindo: Adorei sua explicação Felipe!

← [Responder](#) 👍 (2 curtidas)



[elnataoliveira@gmail.com](mailto:elnataoliveira@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5697>)

12 de fev de 2022

só completando o que os colegas responderam, são sistemas de arquivos distintos entre o docker(sistema de arquivos linux) e hdfs(sistema de arquivos hadoop). Se a importação for feita pelo docker os arquivos não ficam visíveis na aplicação hadoop e vice-versa.

Editado por [elnataoliveira@gmail.com](mailto:elnataoliveira@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5697>) em 12 Fev em 10:32

← [Responder](#) 👍 (2 curtidas)



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

12 de fev de 2022

Muito obrigado pelas respostas Pierre, Felipe e oliveira.

← [Responder](#) 👍 (1 curtida)



(https:)

[arnaldo.carlos.lucas@gmail.com](mailto:arnaldo.carlos.lucas@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5901>)

9 de fev de 2022

Boa noite!

Ao realizar o insert values(null,'food vale'), porque na tabela o cod ao invés de ser NULL, ele vira o valor 1?

Obrigado!

← [Responder](#) 👍 (1 curtida)



(http

[lucasdsfranco@gmail.com](mailto:lucasdsfranco@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5684>)

9 de fev de 2022

Boa noite, Arnaldo.

Isso acontece porque foi especificado no comando de criar a tabela (passo 5 do exercício ) que a chave primária desta tabela tenha esse comportamento, ou seja, a chave primaria será criada pelo próprio banco fazendo um auto incremento de valor. O primeiro registro que inserir a tabela vai colocar com o código 1, se inserir outro terá o código 2, e assim por diante. Este é um conceito comum nos banco de dados relacionais.

← [Responder](#) 👍 (6 curtidas)



(http

[arnaldo.carlos.lucas@gmail.com](mailto:arnaldo.carlos.lucas@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5901>)

10 de fev de 2022

Obrigado Lucas!

← Responder 👍 (1 curtida)



[jonatasgomesdealencar12@gmail.com](mailto:jonatasgomesdealencar12@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5568>)

13 de fev de 2022

não estou conseguindo usar o sqoop pra ver os databases fica dando erro alguém pode me ajudar

```
sqoop list-databases --connect jdbc:mysql://database --username root --password secret
```

Warning: /usr/lib/sqoop/./hbase does not exist! HBase imports will fail.

Please set \$HBASE\_HOME to the root of your HBase installation.

Warning: /usr/lib/sqoop/./hcatalog does not exist! HCatalog jobs will fail.

Please set \$HCAT\_HOME to the root of your HCatalog installation.

Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO\_HOME to the root of your Accumulo installation.

Warning: /usr/lib/sqoop/./zookeeper does not exist! Accumulo imports will fail.

Please set \$ZOOKEEPER\_HOME to the root of your Zookeeper installation.

22/02/13 16:37:42 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Error parsing arguments for list-databases:

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Unrecognized argument: --connect

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Unrecognized argument:

jdbc:mysql://database

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Unrecognized argument: --username

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Unrecognized argument: root

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Unrecognized argument: --password

22/02/13 16:37:42 ERROR tool.BaseSqoopTool: Unrecognized argument: secret

← Responder 👍



[harlisbruno@gmail.com](mailto:harlisbruno@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5724>)

13 de fev de 2022

"connect" está escrito errado. Faltou um "n". Tenta escrever sqoop list-databases --connect jdbc:mysql://database --username root --password secret

 [Responder](#)  (1 curta)[jonatasgomesdealencar12@gmail.com](mailto:jonatasgomesdealencar12@gmail.com)<https://academy.semantix.com.br/courses/108/users/5568>

14 de fev de 2022

corrigi mas ainda assim não consigo ver os databases

```
root@namenode:/# sqoop list-databases --connect jdbc:mysql:\\database --
username root --password secret
```

Warning: /usr/lib/sqoop/./hbase does not exist! HBase imports will fail.

Please set \$HBASE\_HOME to the root of your HBase installation.

Warning: /usr/lib/sqoop/./hcatalog does not exist! HCatalog jobs will fail.

Please set \$HCAT\_HOME to the root of your HCatalog installation.

Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO\_HOME to the root of your Accumulo installation.

Warning: /usr/lib/sqoop/./zookeeper does not exist! Accumulo imports will fail.

Please set \$ZOOKEEPER\_HOME to the root of your Zookeeper installation.

22/02/14 21:58:23 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7

22/02/14 21:58:23 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.

22/02/14 21:58:24 ERROR tool.BaseSqoopTool: Got error creating database manager: java.io.IOException: No manager for connect string:

jdbc:mysql:\\database

at org.apache.sqoop.ConnFactory.getManager(ConnFactory.java:192)

at org.apache.sqoop.tool.BaseSqoopTool.init(BaseSqoopTool.java:272)

at org.apache.sqoop.tool.ListDatabasesTool.run(ListDatabasesTool.java:44)

at org.apache.sqoop.Sqoop.run(Sqoop.java:147)

at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)

at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:183)

at org.apache.sqoop.Sqoop.runTool(Sqoop.java:234)

at org.apache.sqoop.Sqoop.runTool(Sqoop.java:243)

at org.apache.sqoop.Sqoop.main(Sqoop.java:252)

 [Responder](#)  (1 curta)[cdias07@gmail.com](mailto:cdias07@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5904>)

Terça-feira

Oi, jonatas, o seu contrabarra da conexão do databases está contrário, tem que colocar //

Segue o código da questão 1 para copiar e colar:

```
sqoop list-databases --connect jdbc:mysql://database --username root --  
password secret
```

Editado por [cdias07@gmail.com](mailto:cdias07@gmail.com) (<https://academy.semantix.com.br/courses/108/users/5904>) em 15 Feb em 15:33

← **Responder** 👍 (2 curtidas)