

Randal Scott King

Video 3.1

Manual Import



In this Video, we are going to take a look at...

- Importing data using Hue
- Importing data using the CLI

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
master/Customers.csv
Resolving raw.github.com... 23.235.46.133
Connecting to raw.github.com|23.235.46.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.github.com".
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://raw.githubusercontent.com/randalscottking/LearningHadoop2/master/Customers.csv [following]
--2015-09-05 11:01:38-- https://raw.githubusercontent.com/randalscottking/LearningHadoop2/master/Customers.csv
Resolving raw.githubusercontent.com... 23.235.39.133
Connecting to raw.githubusercontent.com|23.235.39.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 200 OK
Length: 600 [text/plain]
Saving to: "Customers.csv"

100%[=====>] 600          --.-K/s   in 0s

2015-09-05 11:01:46 (88.9 MB/s) - "Customers.csv" saved [600/600]

[cloudera@quickstart Downloads]$ wget https://raw.githubusercontent.com/randalscottking/LearningHadoop2/master/Customers.csv --no-check-certificate
```

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
--2015-09-05 11:02:00-- https://raw.githubusercontent.com/randalstotting/LearningHadoop2/master/Transactions.csv
Resolving raw.githubusercontent.com... 23.235.39.133
Connecting to raw.githubusercontent.com|23.235.39.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://raw.githubusercontent.com/randalstotting/LearningHadoop2/master/Transactions.csv [following]
--2015-09-05 11:02:00-- https://raw.githubusercontent.com/randalstotting/LearningHadoop2/master/Transactions.csv
Resolving raw.githubusercontent.com... 199.27.76.133
Connecting to raw.githubusercontent.com|199.27.76.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 200 OK
Length: 981 [text/plain]
Saving to: "Transactions.csv"

100%[=====>] 981      --.-K/s   in 0s

2015-09-05 11:02:01 (134 MB/s) - "Transactions.csv" saved [981/981]

[cloudera@quickstart Downloads]$
```



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
--2015-09-05 11:02:08-- https://raw.githubusercontent.com/randalstottking/LearningHadoop2/master/MoreTransactions.csv
Resolving raw.githubusercontent.com... 199.27.76.133
Connecting to raw.githubusercontent.com|199.27.76.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://raw.githubusercontent.com/randalstottking/LearningHadoop2/master/MoreTransactions.csv [following]
--2015-09-05 11:02:08-- https://raw.githubusercontent.com/randalstottking/LearningHadoop2/master/MoreTransactions.csv
Resolving raw.githubusercontent.com... 199.27.76.133
Connecting to raw.githubusercontent.com|199.27.76.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 200 OK
Length: 272 [text/plain]
Saving to: "MoreTransactions.csv"

100%[=====>] 272      --.-K/s   in 0s

2015-09-05 11:02:11 (35.2 MB/s) - "MoreTransactions.csv" saved [272/272]

[cloudera@quickstart Downloads]$
```

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
--2015-09-05 11:02:27-- https://raw.githubusercontent.com/randalscottking/LearningHadoop2/master/Names.txt
Resolving raw.githubusercontent.com... 23.235.39.133
Connecting to raw.githubusercontent.com|23.235.39.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://raw.githubusercontent.com/randalscottking/LearningHadoop2/master/Names.txt [following]
--2015-09-05 11:02:27-- https://raw.githubusercontent.com/randalscottking/LearningHadoop2/master/Names.txt
Resolving raw.githubusercontent.com... 23.235.39.133
Connecting to raw.githubusercontent.com|23.235.39.133|:443... connected.
WARNING: certificate common name "www.github.com" doesn't match requested host name "raw.githubusercontent.com".
HTTP request sent, awaiting response... 200 OK
Length: 380 [text/plain]
Saving to: "Names.txt"

100%[=====>] 380          --.-K/s   in 0s

2015-09-05 11:02:29 (37.1 MB/s) - "Names.txt" saved [380/380]

[cloudera@quickstart Downloads]$
```

My documents

ACTIONS

+ New document

trash 0

MY PROJECTS +

+ You currently own no projects. Click here to add one now!

SHARED WITH ME

+ There are currently no projects shared with you.

Search for name, description, etc...

There are currently no documents in this project or tag.

Hue - Welcome Home - Mozilla Firefox

Connecting...

quickstart.cloudera:8888/home

Search

☆

↓

↑

☺

☰

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE

Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

My documents

+

 Edit Profile

+

 Manage Users

ACTIONS

+

 New document

trash

0

MY PROJECTS

+

 You currently own no projects. Click here to add one now!

SHARED WITH ME

+

 There are currently no projects shared with you.

Search for name, description, etc...

There are currently no documents in this project or tag.

quickstart.cloudera:8888/useradmin/users

Hue - Welcome Home ...

PACKT

VIDEO

Hue Users

Search for name, group, etc...

Delete

Add user

LDAP

<input type="checkbox"/>	Username	First Name	Last Name	E-mail	Groups	Last Login
<input type="checkbox"/>	cloudera			noreply@cloudera.com	default	Aug. 28, 2015 9:07 AM
<input type="checkbox"/>	hdfs				hadoop	June 9, 2015 8:58 AM
<input type="checkbox"/>	llama					June 9, 2015 8:58 AM
<input type="checkbox"/>	yarn				hadoop	June 9, 2015 8:58 AM
<input type="checkbox"/>	mapred				hadoop	June 9, 2015 8:58 AM
<input type="checkbox"/>	scottking				default	Aug. 9, 2015 8:54 AM

Hue Users - Edit user: cloudera

Step 1: Credentials Step 2: Names and Groups Step 3: Advanced

Username cloudera

Password

Password confirmation

Previous Password

Create home directory ☐

Back Next Update user

File Browser

Actions Move to trash

Upload New

Files Zip/Tgz/Bz2 file Trash

Home / user / cloudera

	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM

File Browser

Actions

Move to trash

UploadNew

Home / user / cloudera

HistoryTrash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>			hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:10 AM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
<input type="checkbox"/>	MoreTransactions.csv	28.8 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM
<input type="checkbox"/>	Names.txt	28.2 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM
<input type="checkbox"/>	Transactions.csv	34.4 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM

ApplicationsPlacesSystem

Section 3, Lecture 9

Hue - User Admin - Hue Users - Mozilla Firefox

Fri Sep 4, 7:09 AMcloudera

Connecting...

quickstart.cloudera:8888/useradmin/users/edit/cloudera

Search

☆📄⬇️🏠💬☰

ClouderaHueHadoopHBaseImpalaSparkSolrOozieCloudera ManagerGetting Started

HUE🏠Query EditorsData BrowsersWorkflowsSearchSecurityFile BrowserJob Browsercloudera🔔🚩➡️

User AdminUsersGroupsPermissionsManage HDFS

Hue Users - Edit user: cloudera

Step 1: CredentialsStep 2: Names and GroupsStep 3: Advanced

Usernamecloudera

Password*****

Password confirmation

Previous Password

Create home directory☐

BackNextUpdate user

quickstart.cloudera:8888/filebrowser

Hue - User Admin - Hu...

Browse Q&AAdd BookmarkContinue

2:34 / 4:33

File Browser

Actions Move to trash

Upload New

Files Zip/Tgz/Bz2 file Trash

Home / user / cloudera

	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM

File Browser

Actions Move to trash Upload New

Home / user / cloudera History Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	/		hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM

Uploading to: /user/cloudera

Select files

Search for file name

Home / user / cloudera

Upload New

History Trash

Name	Size	User	Group	Permissions	Date
.		hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM

Hue - File Browser

quickstart.cloudera:8888/filebrowser/view/user/cloudera

Cloudera Hue Hadoop

HUE

Query Editor

File Browser

Search for file name

Home / user / c

Name

1

.

.Trash

Show 45 of 1 items

File Upload

cloudera Downloads

Places

Search

Recently Used

cloudera

Desktop

File System

Documents

Music

Pictures

Videos

Downloads

Name	Size	Modified
Customers.csv	30.3 KB	07:06
MoreTransactions.csv	28.8 KB	07:07
Names.txt	28.2 KB	07:07
Transactions.csv	34.4 KB	07:07

AddRemove

All Files

CancelOpen

cloudera

Upload New

History Trash

Date

August 09, 2015 08:54 AM

September 04, 2015 07:09 AM

September 04, 2015 07:09 AM

Hue - File Browser

quickstart.cloudera:8888/filebrowser/view/user/cloudera

Cloudera Hue Hadoop

File Browser

Search for file name

Home / user / c

Name

1

.

.Trash

Show 45 of 1 items

File Upload

cloudera Downloads

Places

Search

Recently Used

cloudera

Desktop

File System

Documents

Music

Pictures

Videos

Downloads

Name	Size	Modified
Customers.csv	30.3 KB	07:06
MoreTransactions.csv	28.8 KB	07:07
Names.txt	28.2 KB	07:07
Transactions.csv	34.4 KB	07:07

Add

Remove

All Files

Cancel

Open

cloudera

Upload

New

History

Trash

Date

August 09, 2015 08:54 AM

September 04, 2015 07:09 AM

September 04, 2015 07:09 AM

/user/cloudera/Transactions.csv upload succeeded

File Browser

Search for file name Actions Move to trash Upload New

Home / user / cloudera History Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:10 AM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
<input type="checkbox"/>	MoreTransactions.csv	28.8 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM
<input type="checkbox"/>	Names.txt	28.2 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM
<input type="checkbox"/>	Transactions.csv	34.4 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM

File Browser

Actions

Move to trash

Upload

New

Home / user / cloudera

History Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	August 09, 2015 08:54 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:10 AM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 04, 2015 07:09 AM
<input type="checkbox"/>	MoreTransactions.csv	28.8 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM
<input checked="" type="checkbox"/>	Names.txt	28.2 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM
<input type="checkbox"/>	Transactions.csv	34.4 KB	cloudera	cloudera	-rw-r--r--	September 04, 2015 07:10 AM

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ hdfs dfs -ls
[cloudera@quickstart Downloads]$
```

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ hdfs dfs -ls
[cloudera@quickstart Downloads]$ hdfs dfs -copyFromLocal Transactions.csv
```



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ hdfs dfs -ls
[cloudera@quickstart Downloads]$ hdfs dfs -copyFromLocal Transactions.csv
[cloudera@quickstart Downloads]$ hdfs dfs -copyFromLocal MoreTransactions.csv
[cloudera@quickstart Downloads]$ hdfs dfs -ls
Found 2 items
-rw-r--r--  1 cloudera cloudera      272 2015-09-05 11:04 MoreTransactions.csv
-rw-r--r--  1 cloudera cloudera      981 2015-09-05 11:04 Transactions.csv
[cloudera@quickstart Downloads]$
```



Next Video

Importing from Databases Using Sqoop



1x

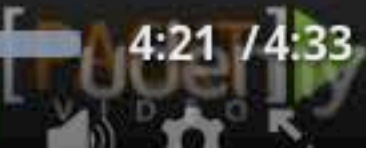


[Browse Q&A](#)

[Add Bookmark](#)

[Continue](#)

4:21 / 4:33



Randal Scott King

Video 3.2

Importing from Databases Using Sqoop



In this Video, we are going to take a look at...

- Creating a database in MySQL
- Using Sqoop to import data from MySQL

My documents

ACTIONS

+ New document

trash 0

MY PROJECTS +

+ You currently own no projects. Click here to add one now!

SHARED WITH ME

+ There are currently no projects shared with you.

Search for name, description, etc...

There are currently no documents in this project or tag.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ sudo service mysqld status  
mysqld (pid 1614) is running...  
[cloudera@quickstart ~]$
```

Search: sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

Hue - Welcome Home - Mozilla Firefox

sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

[cloudera@quickstart ~]\$ sudo service mysqld status
mysqld (pid 1614) is running...
[cloudera@quickstart ~]\$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 1154
Server version: 5.1.66 Source distribution

Copyright (c) 2000, 2012, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
mysql> SHOW

SHARED WITH ME

There are currently no projects shared with you.

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

```
mysql> SHOW DATABASES;
```

Database
information_schema
cm
firehose
hue
metastore
mysql
nav
navms
oozie
retail_db
rman
sentry

12 rows in set (0.00 sec)

```
mysql>
```

sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
cm  
firehose  
hue  
metastore  
mysql  
nav  
navms  
oozie  
retail_db  
rman  
sentry  
+-----+  
12 rows in set (0.00 sec)  
  
mysql> CREATE DATABASE xyzcorp;  
Query OK, 1 row affected (0.00 sec)  
  
mysql> USE xyzcorp;  
Database changed  
mysql> CREATE TABLE Customers  
-> (Acct_no TEXT,Acct_name TEXT,Address TEXT,City TEXT,State TEXT,ZIP TEXT);  
Query OK, 0 rows affected (0.00 sec)  
  
mysql> des
```

Search: sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

```
mysql> CREATE DATABASE xyzcorp;
Query OK, 1 row affected (0.00 sec)

mysql> USE xyzcorp;
Database changed

mysql> CREATE TABLE Customers
  -> (Acct_no TEXT,Acct_name TEXT,Address TEXT,City TEXT,State TEXT,ZIP TEXT);
Query OK, 0 rows affected (0.00 sec)

mysql> describe Customers;
+-----+-----+-----+-----+-----+-----+
| Field      | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Acct_no    | text | YES  |     | NULL    |       |
| Acct_name  | text | YES  |     | NULL    |       |
| Address    | text | YES  |     | NULL    |       |
| City       | text | YES  |     | NULL    |       |
| State      | text | YES  |     | NULL    |       |
| ZIP        | text | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
6 rows in set (0.00 sec)

mysql> 
```

sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

PACKT

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

Query OK, 0 rows affected (0.00 sec)

mysql> describe Customers;

Field	Type	Null	Key	Default	Extra
Acct_no	text	YES		NULL	
Acct_name	text	YES		NULL	
Address	text	YES		NULL	
City	text	YES		NULL	
State	text	YES		NULL	
ZIP	text	YES		NULL	

6 rows in set (0.00 sec)

mysql> LOAD DATA LOCAL INFILE

-> '/home/cloudera/Downloads/Customers.csv'

-> INTO TABLE Customers

-> FIELDS TERMINATED BY ','

-> LINES TERMINATED BY '\r';

Query OK, 10 rows affected (0.00 sec)

Records: 10 Deleted: 0 Skipped: 0 Warnings: 0

mysql>

sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

28201						
1000356		Contruction Depot		102 Peachtree Road		Atlanta
30350						GA
1000567		Jupiter Networks		203 Packet Road		Sunnyvale
94085						CA
1000890		Scottish Petroleum		304 Scottish Blvd		Houston
77079						TX
1000446		Yummy Restaurants		405 Cardinal Avenue		Louisville
40203						KY
1000386		Andicle		506 Database Drive		Redwood City
94061						CA
1000336		Banana Computer		607 Infinity Road		Cupertino
94024						CA
1000998		McDougals		708 Burgerton Avenue		Oak Brook
60523						IL
1000326		Jimsung		809 Electronic Lane		Ridgefield Park
07660						NJ
1000269		Dash Cellular		900 Dash Drive		Overland Park
66204						KS

+-----+

10 rows in set (0.00 sec)

mysql>

sqoop import hue

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

Hue - Welcome Home ...

cloudera@quickstart:~

PACKT

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

```
| 07660 |
| 1000269 | Dash Cellular | 900 Dash Drive | Overland Park | KS
| 66204 |
+-----+
10 rows in set (0.00 sec)

mysql> SELECT * FROM Customers;
+-----+-----+-----+-----+-----+-----+
| Acct_no | Acct_name | Address | City | State | ZIP |
+-----+-----+-----+-----+-----+-----+
| 1000245 | Toms Produce | 100 Toms Lane | Charlotte | NC | 28201 |
| 1000356 | Contruction Depot | 102 Peachtree Road | Atlanta | GA | 30350 |
| 1000567 | Jupiter Networks | 203 Packet Road | Sunnyvale | CA | 94085 |
| 1000890 | Scottish Petroleum | 304 Scottish Blvd | Houston | TX | 77079 |
| 1000446 | Yummy Restaurants | 405 Cardinal Avenue | Louisville | KY | 40203 |
| 1000386 | Andicle | 506 Database Drive | Redwood City | CA | 94061 |
| 1000336 | Banana Computer | 607 Infinity Road | Cupertino | CA | 94024 |
| 1000998 | McDougals | 708 Burgerton Avenue | Oak Brook | IL | 60523 |
| 1000326 | Jimsung | 809 Electronic Lane | Ridgefield Park | NJ | 07660 |
| 1000269 | Dash Cellular | 900 Dash Drive | Overland Park | KS | 66204 |
+-----+-----+-----+-----+-----+-----+
10 rows in set (0.00 sec)

mysql>
```

goop import hue

Started

Browser Job Browser cloudera

There are currently no projects shared with you.

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

```
File Edit View Search Terminal Help
> --username root -P \
> --direct -m 1;
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
15/09/05 18:02:48 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.2
Enter password:
15/09/05 18:02:54 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
15/09/05 18:02:54 INFO tool.CodeGenTool: Beginning code generation
15/09/05 18:02:55 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Customers` AS t LI
MIT 1
15/09/05 18:02:55 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Customers` AS t LI
MIT 1
15/09/05 18:02:55 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/2edfecb5411580085e9945c0181730cd/Customers.java uses or overrides a d
eprecated API.
Note: Recompile with -Xlint:deprecation for details.
15/09/05 18:02:59 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/2edfecb541
1580085e9945c0181730cd/Customers.jar
15/09/05 18:02:59 INFO manager.DirectMySQLManager: Beginning mysqldump fast path import
15/09/05 18:02:59 INFO mapreduce.ImportJobBase: Beginning import of Customers
15/09/05 18:02:59 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapred
uce.jobtracker.address
15/09/05 18:02:59 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.
jar
```

There are currently
no projects shared
with you.


```

cloudera@quickstart:~
File Edit View Search Terminal Help
Map input records=1
Map output records=10
Input split bytes=87
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=117
CPU time spent (ms)=1120
Physical memory (bytes) snapshot=106545152
Virtual memory (bytes) snapshot=1505992704
Total committed heap usage (bytes)=60882944
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=601
15/09/05 18:03:30 INFO mapreduce.ImportJobBase: Transferred 601 bytes in 29.1969 seconds (20.5844 bytes/sec)
15/09/05 18:03:30 INFO mapreduce.ImportJobBase: Retrieved 10 records.
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 4 items
drwxr-xr-x - cloudera cloudera 0 2015-09-05 17:21 .Trash
drwxr-xr-x - cloudera cloudera 0 2015-09-05 18:03 Customers
-rw-r--r-- 1 cloudera cloudera 272 2015-09-05 11:04 MoreTransactions.csv
-rw-r--r-- 1 cloudera cloudera 981 2015-09-05 11:04 Transactions.csv
[cloudera@quickstart ~]$
    
```

Go to Dashboard

sqoop import hue

Started

Browser Job Browser cloudera

There are currently no projects shared with you.

Hue - Welcome Home - Mozilla Firefox

cloudera@quickstart:~

File Edit View Search Terminal Help

```
mysql> SELECT * FROM Customers;
```

Acct_no	Acct_name	Address	City	State	ZIP
1000245	Toms Produce	100 Toms Lane	Charlotte	NC	28201
1000356	Contruction Depot	102 Peachtree Road	Atlanta	GA	30350
1000567	Jupiter Networks	203 Packet Road	Sunnyvale	CA	94085
1000890	Scottish Petroleum	304 Scottish Blvd	Houston	TX	77079
1000446	Yummy Restaurants	405 Cardinal Avenue	Louisville	KY	40203
1000386	Andicle	506 Database Drive	Redwood City	CA	94061
1000336	Banana Computer	607 Infinity Road	Cupertino	CA	94024
1000998	McDougals	708 Burgerton Avenue	Oak Brook	IL	60523
1000326	Jimsung	809 Electronic Lane	Ridgefield Park	NJ	07660
1000269	Dash Cellular	900 Dash Drive	Overland Park	KS	66204

10 rows in set (0.00 sec)

```
mysql> exit
Bye
[cloudera@quickstart ~]$ sqoop import \
> --connect jdbc:mysql://127.0.0.1:3306/xyzcorp \
> --table Customers \
> --username root -P \
> --direct -m 1;
```

goop import hue

Started

Browser Job Browser cloudera

There are currently no projects shared with you.

Next Video

Using Flume to Import Streaming Data

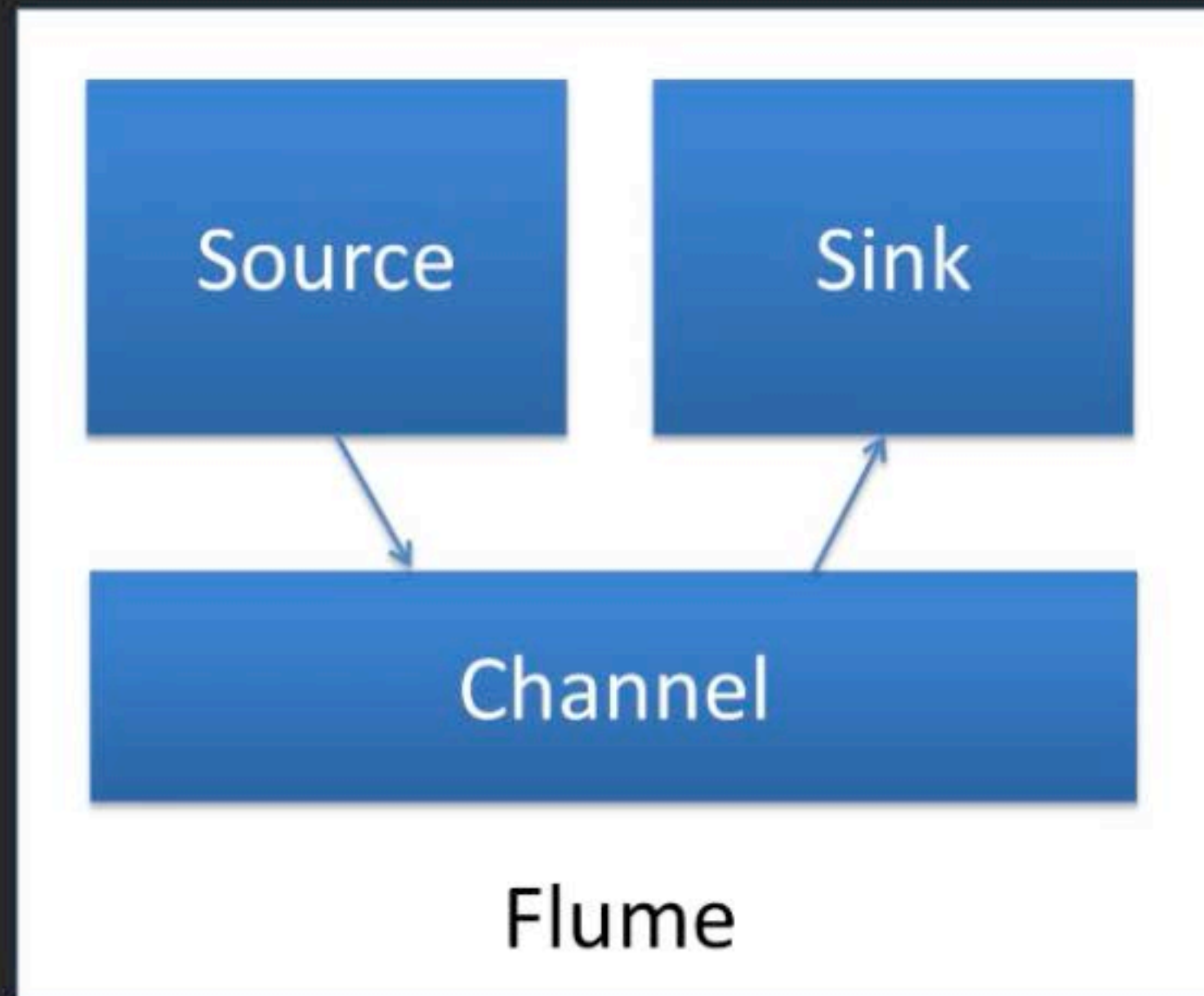
Randal Scott King

Video 3.3

Using Flume to Import Streaming Data



Flume



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ mkdir /tmp/spooldir  
[cloudera@quickstart ~]$ hdfs dfs -mkdir /tm
```

Search: rmdir

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
#This properties file will create an Agent that will move files from a local spooling directory into HDFS  
#Create the Source, Channel, and Sink  
A1.sources=S1  
A1.sinks=H1  
A1.channels=C1  
  
#Bind the Source and Sink to the Channel  
A1.sources.S1.channels=C1  
A1.sinks.H1.channel=C1  
  
#Specify the Source type and directory  
A1.sources.S1.type=spooldir  
A1.sources.S1.spoolDir=/tmp/spoolDir  
  
#Specify the Sink type, directory, and parameters  
A1.sinks.H1.type=HDFS  
A1.sinks.H1.hdfs.path=/tmp/flume-ng  
A1.sinks.H1.hdfs.filePrefix=events  
A1.sinks.H1.hdfs.fileSuffix=.log  
A1.sinks.H1.hdfs.inUsePrefix=_  
A1.sinks.H1.hdfs.fileType=DataStream  
  
--More-- (90%)
```

Search rmdir

cloudera Manager Getting Started

File Browser Job Browser cloudera

SHARED WITH ME

There are currently no projects shared with you.

cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ flume-ng agent \  
> --conf-file /etc/flume-ng/conf/spoolingToHDFS.properties \  
> --name A1  
Warning: No configuration directory set! Use --conf <dir> to override.  
Info: Including Hadoop libraries found via (/usr/bin/hadoop) for HDFS access  
Info: Excluding /usr/lib/hadoop/lib/slf4j-api-1.7.5.jar from classpath  
Info: Excluding /usr/lib/hadoop/lib/slf4j-log4j12.jar from classpath  
Info: Including HBASE libraries found via (/usr/bin/hbase) for HBASE access  
█
```


File Edit View Search Terminal Help

```
15/09/06 13:59:59 INFO source.DefaultSourceFactory: Creating instance of source S1, type spoolDir
15/09/06 13:59:59 INFO sink.DefaultSinkFactory: Creating instance of sink: H1, type: HDFS
15/09/06 13:59:59 INFO node.AbstractConfigurationProvider: Channel C1 connected to [S1, H1]
15/09/06 13:59:59 INFO node.Application: Starting new configuration: { sourceRunners: { S1=EventDrivenSourceRunner: { source:Spool Directory source S1: { spoolDir: /tmp/spoolDir } } } sinkRunners: { H1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@1703c2db counterGroup: { name:null counters: {} } } } channels: { C1=FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] } } }
15/09/06 13:59:59 INFO node.Application: Starting Channel C1
15/09/06 13:59:59 INFO file.FileChannel: Starting FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] }...
15/09/06 13:59:59 INFO file.Log: Encryption is not enabled
15/09/06 13:59:59 INFO file.Log: Replay started
15/09/06 13:59:59 INFO file.Log: Found NextFileID 0, from []
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFile: Preallocated /home/cloudera/.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFileV3: Starting up with /home/cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume/file-channel/checkpoint/checkpoint.meta
15/09/06 14:00:00 INFO file.FlumeEventQueue: QueueSet population inserting 0 took 0
15/09/06 14:00:00 INFO file.Log: Last Checkpoint Sun Sep 06 13:59:59 PDT 2015, queue depth = 0
15/09/06 14:00:00 INFO file.Log: Replaying logs with v2 replay logic
15/09/06 14:00:00 INFO file.ReplayHandler: Starting replay of []
15/09/06 14:00:00 INFO file.ReplayHandler: read: 0, put: 0, take: 0, rollback: 0, commit: 0, skip: 0, eventCount:0
15/09/06 14:00:00 INFO file.FlumeEventQueue: Search Count = 0, Search Time = 0, Copy Count = 0, Copy Time = 0
15/09/06 14:00:00 INFO file.Log: Rolling /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO file.Log: Roll start /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Unable to get maxDirectMemory from VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Direct Memory Allocation: Allocation = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 20316160
15/09/06 14:00:00 INFO file.LogFile: Opened /home/cloudera/.flume/file-channel/data/log-1
15/09/06 14:00:00 INFO file.Log: Roll end
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0
15/09/06 14:00:00 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
15/09/06 14:00:00 INFO file.FileChannel: Queue Size after replay: 0 [channel=C1]
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: C1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: C1 started
15/09/06 14:00:00 INFO node.Application: Starting Sink H1
15/09/06 14:00:00 INFO node.Application: Starting Source S1
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /tmp/spoolDir
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
```


File Edit View Search Terminal Help

```
15/09/06 13:59:59 INFO source.DefaultSourceFactory: Creating instance of source S1, type spoolDir
15/09/06 13:59:59 INFO sink.DefaultSinkFactory: Creating instance of sink: H1, type: HDFS
15/09/06 13:59:59 INFO node.AbstractConfigurationProvider: Channel C1 connected to [S1, H1]
15/09/06 13:59:59 INFO node.Application: Starting new configuration: { sourceRunners: { S1=EventDrivenSourceRunner: { source:Spool Directory source S1: { spoolDir: /tmp/spoolDir } } } sinkRunners: { H1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@1703c2db counterGroup: { name:null counters: {} } } } channels: { C1=FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] } } }
15/09/06 13:59:59 INFO node.Application: Starting Channel C1
15/09/06 13:59:59 INFO file.FileChannel: Starting FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] }...
15/09/06 13:59:59 INFO file.Log: Encryption is not enabled
15/09/06 13:59:59 INFO file.Log: Replay started
15/09/06 13:59:59 INFO file.Log: Found NextFileID 0, from []
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFile: Preallocated /home/cloudera/.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFileV3: Starting up with /home/cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume/file-channel/checkpoint/checkpoint.meta
15/09/06 14:00:00 INFO file.FlumeEventQueue: QueueSet population inserting 0 took 0
15/09/06 14:00:00 INFO file.Log: Last Checkpoint Sun Sep 06 13:59:59 PDT 2015, queue depth = 0
15/09/06 14:00:00 INFO file.Log: Replaying logs with v2 replay logic
15/09/06 14:00:00 INFO file.ReplayHandler: Starting replay of []
15/09/06 14:00:00 INFO file.ReplayHandler: read: 0, put: 0, take: 0, rollback: 0, commit: 0, skip: 0, eventCount:0
15/09/06 14:00:00 INFO file.FlumeEventQueue: Search Count = 0, Search Time = 0, Copy Count = 0, Copy Time = 0
15/09/06 14:00:00 INFO file.Log: Rolling /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO file.Log: Roll start /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Unable to get maxDirectMemory from VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Direct Memory Allocation: Allocation = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 20316160
15/09/06 14:00:00 INFO file.LogFile: Opened /home/cloudera/.flume/file-channel/data/log-1
15/09/06 14:00:00 INFO file.Log: Roll end
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0
15/09/06 14:00:00 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
15/09/06 14:00:00 INFO file.FileChannel: Queue Size after replay: 0 [channel=C1]
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: C1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: C1 started
15/09/06 14:00:00 INFO node.Application: Starting Sink H1
15/09/06 14:00:00 INFO node.Application: Starting Source S1
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /tmp/spoolDir
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
```


File Edit View Search Terminal Help

```
15/09/06 13:59:59 INFO source.DefaultSourceFactory: Creating instance of source S1, type spoolDir
15/09/06 13:59:59 INFO sink.DefaultSinkFactory: Creating instance of sink: H1, type: HDFS
15/09/06 13:59:59 INFO node.AbstractConfigurationProvider: Channel C1 connected to [S1, H1]
15/09/06 13:59:59 INFO node.Application: Starting new configuration: { sourceRunners: { S1=EventDrivenSourceRunner: { source:Spool Directory source S1: { spoolDir: /tmp/spoolDir } } } sinkRunners: { H1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@1703c2db counterGroup: { name:null counters: {} } } } channels: { C1=FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] } } }
15/09/06 13:59:59 INFO node.Application: Starting Channel C1
15/09/06 13:59:59 INFO file.FileChannel: Starting FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] }...
15/09/06 13:59:59 INFO file.Log: Encryption is not enabled
15/09/06 13:59:59 INFO file.Log: Replay started
15/09/06 13:59:59 INFO file.Log: Found NextFileID 0, from []
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFile: Preallocated /home/cloudera/.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFileV3: Starting up with /home/cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume/file-channel/checkpoint/checkpoint.meta
15/09/06 14:00:00 INFO file.FlumeEventQueue: QueueSet population inserting 0 took 0
15/09/06 14:00:00 INFO file.Log: Last Checkpoint Sun Sep 06 13:59:59 PDT 2015, queue depth = 0
15/09/06 14:00:00 INFO file.Log: Replaying logs with v2 replay logic
15/09/06 14:00:00 INFO file.ReplayHandler: Starting replay of []
15/09/06 14:00:00 INFO file.ReplayHandler: read: 0, put: 0, take: 0, rollback: 0, commit: 0, skip: 0, eventCount:0
15/09/06 14:00:00 INFO file.FlumeEventQueue: Search Count = 0, Search Time = 0, Copy Count = 0, Copy Time = 0
15/09/06 14:00:00 INFO file.Log: Rolling /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO file.Log: Roll start /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Unable to get maxDirectMemory from VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Direct Memory Allocation: Allocation = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 20316160
15/09/06 14:00:00 INFO file.LogFile: Opened /home/cloudera/.flume/file-channel/data/log-1
15/09/06 14:00:00 INFO file.Log: Roll end
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0
15/09/06 14:00:00 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
15/09/06 14:00:00 INFO file.FileChannel: Queue Size after replay: 0 [channel=C1]
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: C1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: C1 started
15/09/06 14:00:00 INFO node.Application: Starting Sink H1
15/09/06 14:00:00 INFO node.Application: Starting Source S1
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /tmp/spoolDir
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
```


cloudera@quickstart:~

cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hdfs dfs -ls /tmp/flume-ng
```

```
[cloudera@quickstart ~]$
```

```
type spoolDir
```

```
type: HDFS
```

```
[S1, H1]
```

```
Source: {S1=EventDrivenSourceRunner: { source:Spool Directory source S1: { spoolDir: /tmp/spoolDir DefaultSinkProcessor@1703c2db counterGroup: { name:null counters:{} } } } }
```

```
...e/cloudera/.flume/file-channel/data] }...
```

```
...a/.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000  
...cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume
```

```
...ue depth = 0
```

```
...commit: 0, skip: 0, eventCount:0  
...Copy Count = 0, Copy Time = 0
```

```
...VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)  
...ion = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 2
```

```
...a/log-1
```

```
15/09/06 14:00:00 INFO file.Log: Roll end
```

```
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
```

```
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0
```

```
15/09/06 14:00:00 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
```

```
15/09/06 14:00:00 INFO file.FileChannel: Queue Size after replay: 0 [channel=C1]
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: C1: Successfully registered new MBean.
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: C1 started
```

```
15/09/06 14:00:00 INFO node.Application: Starting Sink H1
```

```
15/09/06 14:00:00 INFO node.Application: Starting Source S1
```

```
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /tmp/spoolDir
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
```



```
type spooldir
```

File System: HDFS

[S1, H1]

```

s:{S1=EventDrivenSourceRunner: { source:Spool Directory source S1: { spoolD
defaultSinkProcessor@1703c2db counterGroup:{ name:null counters:{} } }} cha

```

```
./cloudera/.flume/file-channel/data] }...
```

```
./flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000
./flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume
```

depth = 0

```
commit: 0, skip: 0, eventCount:0
```

Copy Count = 0, Copy Time = 0

```
VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)
```

```
lon = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 2
```

 $\frac{1}{\log-1}$

```
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
```

```
15/09/06 14:00:00 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: C1: Successfully registered new MBean.
```

```
15/09/06 14:00:00 INFO node.Application: Starting Sink H1
```

```
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDi
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
```


File Edit View Search Terminal Help

```
15/09/06 13:59:59 INFO node.Application: Starting new configuration:{ sourceRunners:{S1=EventDrivenSourceRunner: { source:Spool Directory source S1: { spoolDir: /tmp/spooldir } }} sinkRunners:{H1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@1703c2db counterGroup:{ name:null counters:{} } }} channels:{C1=FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] }} }
15/09/06 13:59:59 INFO node.Application: Starting Channel C1
15/09/06 13:59:59 INFO file.FileChannel: Starting FileChannel C1 { dataDirs: [/home/cloudera/.flume/file-channel/data] }...
15/09/06 13:59:59 INFO file.Log: Encryption is not enabled
15/09/06 13:59:59 INFO file.Log: Replay started
15/09/06 13:59:59 INFO file.Log: Found NextFileID 0, from []
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFile: Preallocated /home/cloudera/.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000
15/09/06 13:59:59 INFO file.EventQueueBackingStoreFileV3: Starting up with /home/cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume/file-channel/checkpoint/checkpoint.meta
15/09/06 14:00:00 INFO file.FlumeEventQueue: QueueSet population inserting 0 took 0
15/09/06 14:00:00 INFO file.Log: Last Checkpoint Sun Sep 06 13:59:59 PDT 2015, queue depth = 0
15/09/06 14:00:00 INFO file.Log: Replaying logs with v2 replay logic
15/09/06 14:00:00 INFO file.ReplayHandler: Starting replay of []
15/09/06 14:00:00 INFO file.ReplayHandler: read: 0, put: 0, take: 0, rollback: 0, commit: 0, skip: 0, eventCount:0
15/09/06 14:00:00 INFO file.FlumeEventQueue: Search Count = 0, Search Time = 0, Copy Count = 0, Copy Time = 0
15/09/06 14:00:00 INFO file.Log: Rolling /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO file.Log: Roll start /home/cloudera/.flume/file-channel/data
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Unable to get maxDirectMemory from VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)
15/09/06 14:00:00 INFO tools.DirectMemoryUtils: Direct Memory Allocation: Allocation = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 20316160
15/09/06 14:00:00 INFO file.LogFile: Opened /home/cloudera/.flume/file-channel/data/log-1
15/09/06 14:00:00 INFO file.Log: Roll end
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
15/09/06 14:00:00 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0
15/09/06 14:00:00 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
15/09/06 14:00:00 INFO file.FileChannel: Queue Size after replay: 0 [channel=C1]
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: C1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: C1 started
15/09/06 14:00:00 INFO node.Application: Starting Sink H1
15/09/06 14:00:00 INFO node.Application: Starting Source S1
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /tmp/spooldir
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
15/09/06 14:01:30 INFO avro.ReliableSpoolingFileEventReader: Preparing to move file /tmp/spooldir/text1.txt to /tmp/spooldir/text1.txt.COMPLETED
15/09/06 14:01:35 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
15/09/06 14:01:35 INFO hdfs.BucketWriter: Creating /tmp/flume-ng/_events.1441573295325.log.tmp
```


cloudera@quickstart:~

cloudera@quickstart:/tmp/spooldir

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hdfs dfs -ls /tmp/flume-ng
```

```
[cloudera@quickstart ~]$ cd /tmp/spooldir
```

```
[cloudera@quickstart spooldir]$ vi text1.txt
```

```
[cloudera@quickstart spooldir]$ hdfs dfs -ls /tmp/flume-ng
```

```
Found 1 items
```

```
-rw-r--r-- 1 cloudera supergroup 18 2015-09-06 14:01 /tmp/flume-ng/_ev
```

```
ents.1441573295325.log.tmp
```

```
[cloudera@quickstart spooldir]$
```

```
.../cloudera/.flume/file-channel/data] }...  
.../cloudera/.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000  
.../cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume  
...  
...ue depth = 0  
...commit: 0, skip: 0, eventCount:0  
...Copy Count = 0, Copy Time = 0  
...  
...VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)  
...ion = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 2  
...a/log-1  
.../cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0  
...ta: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0  
...e/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103  
...up for type: CHANNEL, name: C1: Successfully registered new MBean.  
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: C1 started  
15/09/06 14:00:00 INFO node.Application: Starting Sink H1  
15/09/06 14:00:00 INFO node.Application: Starting Source S1  
15/09/06 14:00:00 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /tmp/spooldir  
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.  
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started  
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.  
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started  
15/09/06 14:01:30 INFO avro.ReliableSpoolingFileEventReader: Preparing to move file /tmp/spooldir/text1.txt to /tmp/spooldir/text1.txt.COMPLETED  
15/09/06 14:01:35 INFO hdfs.HDFSDataStream: Serializer = TEXT, UserRawLocalFileSystem = false  
15/09/06 14:01:35 INFO hdfs.BucketWriter: Creating /tmp/flume-ng/_events.1441573295325.log.tmp  
15/09/06 14:01:59 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 1  
15/09/06 14:01:59 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200108, queueSize: 0, queueHead: 0  
15/09/06 14:01:59 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 174 logWriteOrderID: 1441573200108
```


cloudera@quickstart:~

cloudera@quickstart:/tmp/spooldir

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hdfs dfs -ls /tmp/flume-ng
```

```
[cloudera@quickstart ~]$ cd /tmp/spooldir
```

```
[cloudera@quickstart spooldir]$ vi text1.txt
```

```
[cloudera@quickstart spooldir]$ hdfs dfs -ls /tmp/flume-ng
```

```
Found 1 items
```

```
-rw-r--r-- 1 cloudera supergroup 18 2015-09-06 14:01 /tmp/flume-ng/_events.1441573295325.log.tmp
```

```
[cloudera@quickstart spooldir]$ hdfs dfs -ls /tmp/flume-ng
```

```
Found 1 items
```

```
-rw-r--r-- 1 cloudera supergroup 18 2015-09-06 14:02 /tmp/flume-ng/_events.1441573295325.log
```

```
[cloudera@quickstart spooldir]$
```

```
.../.flume/file-channel/checkpoint/checkpoint to 8008232 for capacity 1000000
```

```
...cloudera/.flume/file-channel/checkpoint/checkpoint and /home/cloudera/.flume
```

```
...ue depth = 0
```

```
...commit: 0, skip: 0, eventCount:0
```

```
...by Count = 0, Copy Time = 0
```

```
...VM: NoSuchMethodException: sun.misc.VM.maxDirectMemory(null)
```

```
...ion = 1048576, Allocated = 0, MaxDirectMemorySize = 20316160, Remaining = 2
```

```
...a/log-1
```

```
.../cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 0
```

```
...ta: logWriteOrderID: 1441573200103, queueSize: 0, queueHead: 0
```

```
...e/file-channel/data/log-1 position: 0 logWriteOrderID: 1441573200103
```

```
...oup for type: CHANNEL, name: C1: Successfully registered new MBean.
```

```
...NEL, name: C1 started
```

```
15/09/06 14:00:00 INFO source.PoolDirectorySource: PoolDirectorySource source starting with directory: /tmp/spooldir
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: H1: Successfully registered new MBean.
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: H1 started
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: S1: Successfully registered new MBean.
```

```
15/09/06 14:00:00 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: S1 started
```

```
15/09/06 14:01:30 INFO avro.ReliableSpoolingFileEventReader: Preparing to move file /tmp/spooldir/text1.txt to /tmp/spooldir/text1.txt.COMPLETED
```

```
15/09/06 14:01:35 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
```

```
15/09/06 14:01:35 INFO hdfs.BucketWriter: Creating /tmp/flume-ng/_events.1441573295325.log.tmp
```

```
15/09/06 14:01:59 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/cloudera/.flume/file-channel/checkpoint/checkpoint, elements to sync = 1
```

```
15/09/06 14:01:59 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1441573200108, queueSize: 0, queueHead: 0
```

```
15/09/06 14:01:59 INFO file.Log: Updated checkpoint for file: /home/cloudera/.flume/file-channel/data/log-1 position: 174 logWriteOrderID: 1441573200108
```

```
15/09/06 14:02:08 INFO hdfs.BucketWriter: Closing /tmp/flume-ng/_events.1441573295325.log.tmp
```

```
15/09/06 14:02:08 INFO hdfs.BucketWriter: Renaming /tmp/flume-ng/_events.1441573295325.log.tmp to /tmp/flume-ng/events.1441573295325.log
```

```
15/09/06 14:02:08 INFO hdfs.HDFSEventSink: Writer callback called.
```


Summary

- Manual import (Hue and CLI)
- Database import with Sqoop
- Streaming import with Flume

Next Section

Using MapReduce and Pig