

1. Создать Flume поток используя Flume сервис

Сперва в hbase создадим таблицу Student4_10 с одним семейством колонок Message.

Запустим HBASE SHELL

```
[student4_10@manager ~]$ hbase shell
```

Создадим таблицу

```
hbase(main):001:0> create 'Student4_10', 'Message'
```

```
0 row(s) in 8.6700 seconds
```

Посмотрим таблицу через DESCRIBE

```
hbase(main):001:0> describe 'Student4_10'
```

```
Table Student4_10 is ENABLED
```

```
Student4_10
```

```
COLUMN FAMILIES DESCRIPTION
```

```
{NAME => 'Message', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
```

```
1 row(s) in 0.5900 seconds
```

```
student4_10@manager:~$ hbase shell
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be removed in a future release
20/06/27 18:30:57 INFO Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.16.2, rUnknown, Mon Jun 3 03:50:03 PDT 2019

[student4_10@manager ~]$ create 'Student4_10', 'Message'
-bash: create: command not found
[student4_10@manager ~]$ hbase shell
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be removed in a future release
20/06/27 18:33:15 INFO Configuration.deprecation: hadoop.native.lib is deprecated.
Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.16.2, rUnknown, Mon Jun 3 03:50:03 PDT 2019

hbase(main):001:0> create 'Student4_10', 'Message'
0 row(s) in 8.6700 seconds

=> Hbase::Table - Student4_10
hbase(main):002:0> [student4_10@manager ~]$ describe 'Student4_10'
-bash: describe: command not found
[student4_10@manager ~]$ hbase shell
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be removed in a future release
20/06/27 18:35:31 INFO Configuration.deprecation: hadoop.native.lib is deprecated.
Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.16.2, rUnknown, Mon Jun 3 03:50:03 PDT 2019

hbase(main):001:0> describe 'Student4_10'
Table Student4_10 is ENABLED
Student4_10
COLUMN FAMILIES DESCRIPTION
(NAME => 'Message', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', K
EEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COM
PRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536
', REPLICATION_SCOPE => '0')
1 row(s) in 0.5900 seconds

[student4_10@manager ~]$
```

Для Source path берём файл /var/log/cron

```
[student4_10@manager ~]$ ls /var/log/cron
```

/var/log/cron

Создадим папку для sink в локальной папке

```
[student4_10@manager ~]$ hdfs dfs -mkdir /home/student4_10/hbase
```

```
[student4_10@manager ~]$ hdfs dfs -ls /home/student4_10/
```

Found 3 items

drwxrwxrwx	- flume supergroup	0	2020-06-27 18:03	/home/student4_10/20-06-27
drwxr-xr-x	- flume supergroup	0	2020-06-27 18:03	/home/student4_10/flume
drwxr-xr-x	- flume supergroup	0	2020-06-27 18:57	/home/student4_10/hbase

Код для Config File

Naming the components on the current agent

Flume4_10.sources = ExecSource

Flume4_10.channels = MemChannel

Flume4_10.sinks = HdfsSink HbaseSink

Describing/Configuring the source

Flume4_10.sources.ExecSource.type = exec

Flume4_10.sources.ExecSource.command = /bin/tailf /var/log/cron

#Flume4_10.sources.ExecSource.command = /bin/tailf /tmp/myfile

#Flume4_10.sources.ExecSource.command = /bin/tailf /tmp/source4_10

Flume4_10.sources.ExecSource.interceptors = TimestampInterceptor

Flume4_10.sources.ExecSource.interceptors.TimestampInterceptor.type = timestamp

#SINKS

Describing/Configuring the HDFS sink

Flume4_10.sinks.HdfsSink.type = hdfs

#Flume4_10.sinks.HdfsSink.hdfs.path= /flume/Flume4_10/%y-%m-%d/

Flume4_10.sinks.HdfsSink.hdfs.path= /home/student4_10/hbase/%y-%m-%d/

Flume4_10.sinks.HdfsSink.hdfs.filePrefix = events

Describing/Configuring the HBASE sink

Flume4_10.sinks.HbaseSink.type = hbase

Flume4_10.sinks.HbaseSink.table = Student4_10

Flume4_10.sinks.HbaseSink.columnFamily = Message

Describing/Configuring the channel

Flume4_10.channels.MemChannel.type = memory

Flume4_10.channels.MemChannel.capacity = 10000

Flume4_10.channels.MemChannel.transactionCapacity = 10

Bind the source and sink to the channel

Flume4_10.sources.ExecSource.channels = MemChannel

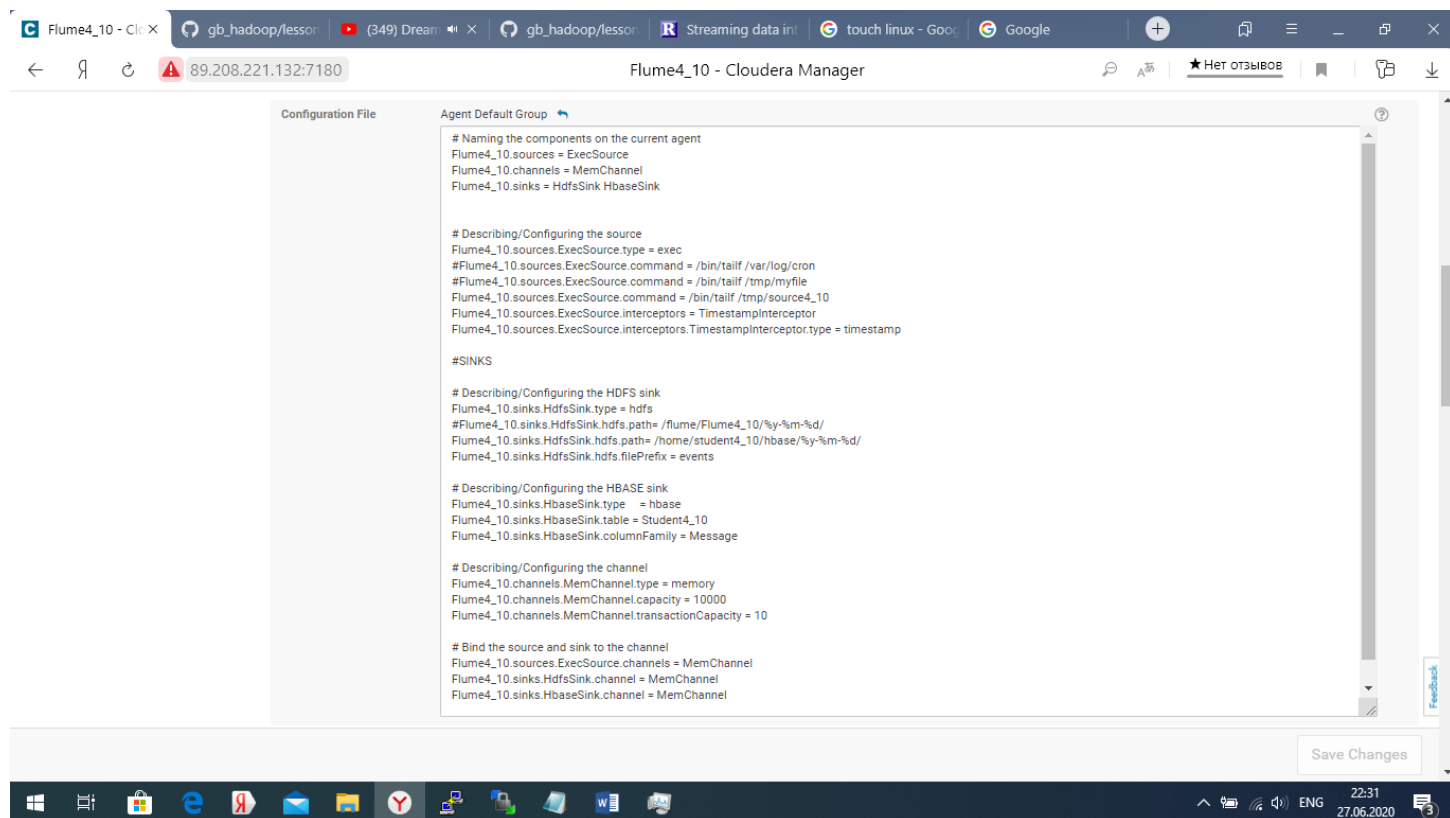
Flume4_10.sinks.HdfsSink.channel = MemChannel

Flume4_10.sinks.HbaseSink.channel = MemChannel

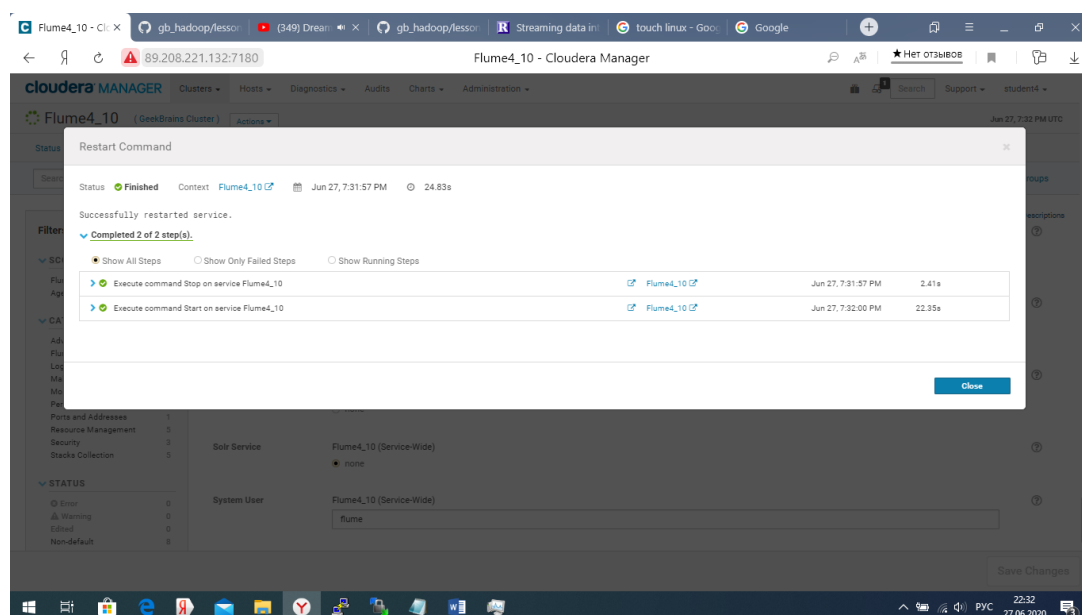
The screenshot displays the Cloudera Manager web interface for the 'Flume4_10' service on the 'GeekBrains Cluster'. The 'Configuration' tab is active, showing a table of service configurations. The table has columns for the service name, the configuration value, and a 'Show All Descriptions' link. The services listed are HDFS Service, HBase Service, Kafka Service, and Solr Service. Each service has a configuration value of 'Flume4_10 (Service-Wide)' and a radio button selection for 'HDFS', 'HBase', 'Kafka', or 'none'. The 'HDFS' option is selected for HDFS Service, 'HBase' for HBase Service, 'Kafka' for Kafka Service, and 'none' for Solr Service. A 'Filters' sidebar on the left shows a tree view of the configuration hierarchy. At the bottom, there is a 'Save Changes' button and a 'Reason for change...' field.

Service	Configuration Value	Selected Option	Show All Descriptions
HDFS Service	Flume4_10 (Service-Wide)	<input checked="" type="radio"/> HDFS	?
HBase Service	Flume4_10 (Service-Wide)	<input checked="" type="radio"/> HBase	?
Kafka Service	Flume4_10 (Service-Wide)	<input checked="" type="radio"/> Kafka	?
Solr Service	Flume4_10 (Service-Wide)	<input checked="" type="radio"/> none	?

Копируем и вставляем код



Сохраним и запустим



Загрузим данные

Данные загруженные в HDFS

```
student4_10@manager:~  
[student4_10@manager ~]$ hdfs dfs -ls -r /home/student4_10/hbase/  
Found 1 items  
drwxr-xr-x - flume supergroup 0 2020-06-27 22:32 /home/student4_10/hbase/20-06-27  
[student4_10@manager ~]$ hdfs dfs -ls -r /home/student4_10/hbase/20-06-27  
Found 13 items  
-rw-r--r-- 3 flume supergroup 209 2020-06-27 22:32 /home/student4_10/hbase/20-06-27/events.1593297065905  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:32 /home/student4_10/hbase/20-06-27/events.1593297065904  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:32 /home/student4_10/hbase/20-06-27/events.1593297065903  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:32 /home/student4_10/hbase/20-06-27/events.1593297065902  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065901  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065900  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065899  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065898  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065897  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065896  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065895  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065894  
-rw-r--r-- 3 flume supergroup 1061 2020-06-27 22:31 /home/student4_10/hbase/20-06-27/events.1593297065893  
[student4_10@manager ~]$
```

```
student4_10@manager:~  
student2_6_logs  
student3_5  
student3_8  
student3_9  
student3_9_db  
syslog  
27 row(s) in 0.2710 seconds  
  
=> ["Project_S2_3", "Stud3_2", "Student3_10", "Student3_12", "Student3_12dfg", "Student3_13", "Student3_3", "Student3_6_Table", "Stude  
nt3_7:Message", "Student4_10", "Users", "UsersTable", "UsersTable:videos", "carbon hbase", "exec_date", "flume_cronlog", "foo_table3.1  
4", "log_table", "mytest:videos", "shlyapka_hbase", "student2_5", "student2_6_logs", "student3_5", "student3_8", "student3_9", "studen  
t3_9_db", "syslog"]  
hbase(main):002:0> scan student4_10  
NameError: undefined local variable or method `student4_10' for #<Object:0x46c10083>  
  
hbase(main):003:0> scan 'Student4_10'  
ROW COLUMN+CELL  
default105c7444-94f7-492d-8817-ef column=Message:pCol, timestamp=1593297133807, value=Jun 27 22:01:01 node3 run-parts(/etc/cron.hourl  
6ea6ec6bcd y)[14601]: finished 0anacron  
default1bb9cff5-0e19-4027-8e60-e2 column=Message:pCol, timestamp=1593297133807, value=Jun 27 20:01:01 node3 run-parts(/etc/cron.hourl  
1bbbfa5ebe y)[22981]: starting 0anacron  
default1f52efc8-f37f-41d5-9d6b-44 column=Message:pCol, timestamp=1593297133807, value=Jun 27 20:01:01 node3 run-parts(/etc/cron.hourl  
cf3f48400b y)[22990]: finished 0anacron  
default2deaad9-ee99-42a7-89a7-9d column=Message:pCol, timestamp=1593297133807, value=Jun 27 22:01:01 node3 CROND[14563]: (root) CMD  
45a37b70ad (run-parts /etc/cron.hourly)  
default5c64f5bf-9ee7-4d70-a868-55 column=Message:pCol, timestamp=1593297133807, value=Jun 27 20:01:01 node3 CROND[22981]: (root) CMD  
3b6408452e (run-parts /etc/cron.hourly)  
default922fdb46-c3c2-45aa-9aaa-41 column=Message:pCol, timestamp=1593297133807, value=Jun 27 21:01:01 node3 CROND[2925]: (root) CMD (  
3eed20f384 run-parts /etc/cron.hourly)  
defaultbd6a476d-e424-4ec4-98c1-1b column=Message:pCol, timestamp=1593297133807, value=Jun 27 21:01:01 node3 run-parts(/etc/cron.hourl  
ba61918472 y)[2925]: starting 0anacron  
defaultbfc8286e-db17-4c3a-9d07-e2 column=Message:pCol, timestamp=1593297133807, value=Jun 27 21:01:01 node3 run-parts(/etc/cron.hourl  
47ef959c07 y)[2934]: finished 0anacron  
defaultca2128e5-b855-4431-9a27-0a column=Message:pCol, timestamp=1593297133807, value=Jun 27 22:01:01 node3 run-parts(/etc/cron.hourl  
23d60eb0fd y)[14563]: starting 0anacron  
incRow column=Message:iCol, timestamp=1593297134157, value=\x00\x00\x00\x00\x00\x00\x00\x00  
10 row(s) in 0.1710 seconds  
  
hbase(main):004:0>
```

Данные загружены в HBASE