

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', 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:~

```
[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', 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 ~]$
```

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
```

Flume4_10 - Clc X gb_hadoop/lesson (348) Tiesto X gb_hadoop/lesson R Stream

89.208.221.132:7180 Flume4_10 - C

cloudera MANAGER Clusters Hosts Diagnostics Audits Charts Admin

Flume4_10 (GeekBrains Cluster) Actions

Status Instances Configuration Commands Metric Details Charts Library Audits Q

Search

Filters

SCOPE

Flume4_10 (Service-Wide)	18
Agent	48

CATEGORY

Advanced	10
Flume-NG Solr Sink	3
Logs	4
Main	10
Monitoring	23
Performance	1
Ports and Addresses	1
Resource Management	5
Security	3
Stacks Collection	5

HDFS Service Flume4_10 (Service-Wide)

☒ HDFS

☐ none

HBase Service Flume4_10 (Service-Wide)

☒ HBase

☐ none

Kafka Service Flume4_10 (Service-Wide)

☒ Kafka

☐ none

Solr Service Flume4_10 (Service-Wide)

☒ none

1 Edited

Windows Taskbar: Start, Task View, Store, Edge, Yandex, Mail, File Explorer, Yandex, RemoteApp, Lock, File Explorer, Word, Teams

Копируем и вставляем код

Configuration File

Agent Default Group

```
# 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/tail
#Flume4_10.sources.ExecSource.command = /bin/tail
Flume4_10.sources.ExecSource.interceptors = Times
Flume4_10.sources.ExecSource.interceptors.Timestamp

#SINKS

# Describing/Configuring the HDFS sink
Flume4_10.sinks.HdfsSink.type = hdfs
#Flume4_10.sinks.HdfsSink.hdfs.path= /flume/Flume
Flume4_10.sinks.HdfsSink.hdfs.path= /home/student
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 = 1000

# 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
```



Сохраним и запустим

Flume4_10 - Clc X gb_hadoop/lesson (349) Dream X gb_hadoop/lesson Stream

← 89.208.221.132:7180 Flume4_10 - C

cloudera MANAGER Clusters Hosts Diagnostics Audits Charts Administration

Flume4_10 (GeekBrains Cluster) Actions

Status Restart Command

Status **Finished** Context [Flume4_10](#) Jun 27, 7:31:57 PM 24.83s

Successfully restarted service.

Completed 2 of 2 step(s).

☒ Show All Steps ☐ Show Only Failed Steps ☐ Show Running Steps

- Execute command Stop on service Flume4_10
- Execute command Start on service Flume4_10

Ports and Addresses 1
Resource Management 5
Security 3
Stacks Collection 5

STATUS

- Error 0
- Warning 0
- Edited 0
- Non-default 8

Solr Service Flume4_10 (Service-Wide)
☒ none

System User Flume4_10 (Service-Wide)
flume

Windows taskbar icons: Start, Task View, Microsoft Store, Edge, Yandex, Mail, File Explorer, Yandex, RemoteApp, Lock, File Explorer, Word, PowerPoint.

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/st
[student4_10@manager ~]$ hdfs dfs -ls -r /home/student4_10/hbase/20-
Found 13 items
-rw-r--r--  3 flume supergroup      209 2020-06-27 22:32 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:32 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:32 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:32 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
-rw-r--r--  3 flume supergroup    1061 2020-06-27 22:31 /home/st
[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_7:Message", "Student4_10", "Users", "UsersTable", "UsersTable:video4", "log_table", "mytest:videos", "shlyapka_hbase", "student2_5", "student3_9_db", "syslog"]
```

hbase(main):002:0> scan student4_10

NameError: undefined local variable or method `student4_10' for #<Object>

hbase(main):003:0> scan 'Student4_10'

ROW	COLUMN+CELL
default105c7444-94f7-492d-8817-ef6ea6ec6bcd	column=Message:pCol, timestamp=1514601: finished 0anacron
default1bb9cff5-0e19-4027-8e60-e21bbbfa5ebe	column=Message:pCol, timestamp=1522981: starting 0anacron
default1f52efc8-f37f-41d5-9d6b-44cf3f48400b	column=Message:pCol, timestamp=1522990: finished 0anacron
default2deaadf9-ee99-42a7-89a7-9d45a37b70ad	column=Message:pCol, timestamp=1522990: finished 0anacron (run-parts /etc/cron.hourly)
default5c64f5bf-9ee7-4d70-a868-553b6408452e	column=Message:pCol, timestamp=1522990: finished 0anacron (run-parts /etc/cron.hourly)
default922fdb46-c3c2-45aa-9aaa-413eed20f384	column=Message:pCol, timestamp=1522990: finished 0anacron (run-parts /etc/cron.hourly)
defaultbd6a476d-e424-4ec4-98c1-1bba61918472	column=Message:pCol, timestamp=1522925: starting 0anacron
defaultbfc8286e-db17-4c3a-9d07-e247ef959c07	column=Message:pCol, timestamp=1522934: finished 0anacron
defaultca2128e5-b855-4431-9a27-0a23d60eb0fd	column=Message:pCol, timestamp=1514563: starting 0anacron
incRow	column=Message:iCol, timestamp=1514563: starting 0anacron

10 row(s) in 0.1710 seconds

hbase(main):004:0> █

Данные загружены в HBASE