

Урок 6. NoSQL в Hadoop. Hbase/Cassandra

Выполнил Колеганов Н.Д.

Задание

1. Модифицировать любой Flume поток используя Flume сервис соответствующего номера.
Тип источника источник -- ехес
Тип канала -- file
Тип слива -- hbase
2. Убедиться что данные поступают из источника и реально сохраняются

1. Настройки для слива Hbase

```
# HBase Sink
Flume4_2.sinks.HBaseSink.type = hbase
Flume4_2.sinks.HBaseSink.channel = FileChannel
Flume4_2.sinks.HBaseSink.table = flume_cronlog
Flume4_2.sinks.HBaseSink.columnFamily = log_cf
```

Вот такая команда

```
Flume4_2.sources.Exec.command = tailf /home/student4_2/pub/123.txt
```

```
[student4_2@node3 pub]$ cat /var/log/cron > /home/student4_2/pub/123.tx
[student4_2@node3 pub]$ cat /var/log/cron > /home/student4_2/pub/123.tx
[student4_2@node3 pub]$ cat /var/log/cron > /home/student4_2/pub/123.tx
[student4_2@node3 pub]$
```

Далее hbase shell → scan flume_cronlog

```
hbase(main):008:0> scan 'flume_cronlog'
ROW                                COLUMN+CELL
default0001a632-3d22-4988-8b11-12289042412 column=log_cf:pCol, timestamp=1585743786451, value=Mar 31 03:24:01 node1 run-parts(/etc/cron.daily)[1731]: finished logrotate
3
default000253e5-5344-4ffe-bf97-f8d4f14fcb7 column=log_cf:pCol, timestamp=1586352948432, value=Apr 7 20:01:01 node1 run-parts(/etc/cron.hourly)[25097]: starting 0anacro
e
default00254163-7639-4c55-b1c2-075357abf30 column=log_cf:pCol, timestamp=1586612261470, value=Apr 6 20:01:01 node1 run-parts(/etc/cron.hourly)[8134]: finished 0anacro
d

2
defaultffff98ec1-5cae-4a28-a0ae-6646c7ac620 column=log_cf:pCol, timestamp=1585742791760, value=Mar 31 20:01:01 node1 run-parts(/etc/cron.hourly)[5738]: finished 0anacro
9
defaultfffa5856-6fbd-4d41-8b6f-46c8455ac7e column=log_cf:pCol, timestamp=1585744291818, value=Mar 30 11:01:02 node1 run-parts(/etc/cron.hourly)[4519]: finished 0anacro
8
incRow                                column=log_cf:iCol, timestamp=1586612263186, value=\x00\x00\x00\x00\x00\x008\xA8
14505 row(s) in 6.1010 seconds
hbase(main):002:0>
```