**ЗАДАНИЕ:**

Посмотреть что не так в конфигурации NetCat Flume agent которого я сделал. Описать и аргументировать.

# Naming the components on the current agent

NetcatAgent.sources = Netcat r1

NetcatAgent.channels = MemChannel c1

NetcatAgent.sinks = LoggerSink k1

# insert timestamp

NetcatAgent.sources.Netcat.interceptors = ts\_interceptor

NetcatAgent.sources.Netcat.interceptors.ts\_interceptor.type = timestamp

# Describing/Configuring the source r1

NetcatAgent.sources.Netcat.type = netcat

NetcatAgent.sources.Netcat.bind = 89.208.220.216

NetcatAgent.sources.Netcat.port = 7401

# Describing/Configuring the sink k1

NetcatAgent.sinks.LoggerSink.type = hdfs

NetcatAgent.sinks.LoggerSink.hdfs.path= /NetCat/snap\_day=%Y%m%d/

# NetcatAgent.sinks.LoggerSink.hdfs.path= /NetCat/

NetcatAgent.sinks.LoggerSink.hdfs.filePrefix=flume\_data

NetcatAgent.sinks.LoggerSink.hdfs.fileType=SequenceFile

NetcatAgent.sinks.LoggerSink.hdfs.codeC=gzip

# Describing/Configuring the channel c1

NetcatAgent.channels.MemChannel.type = memory

NetcatAgent.channels.MemChannel.capacity = 10000

NetcatAgent.channels.MemChannel.transactionCapacity = 10

# Bind the source and sink to the channel

NetcatAgent.sources.Netcat.channels = MemChannel c1

NetcatAgent.sinks.LoggerSink.channel = MemChannel c1

ОТВЕТ:

Похоже на то, что в конфигурации мы попытались настроить sink вида LoggerSink

NetcatAgent.sinks = LoggerSink

и при этом задаем ему type

NetcatAgent.sinks.LoggerSink.type = hdfs

Согласно документации в разделе о **LoggerSink** сказано:

|  |  |  |
| --- | --- | --- |
| **type** | – | The component type name, needs to be logger |

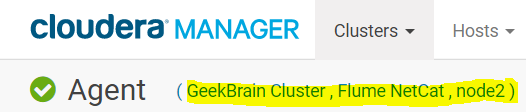
a1.sinks **=** k1

a1.sinks.k1.type **=** logger

следовало бы задать ему type = HDFS

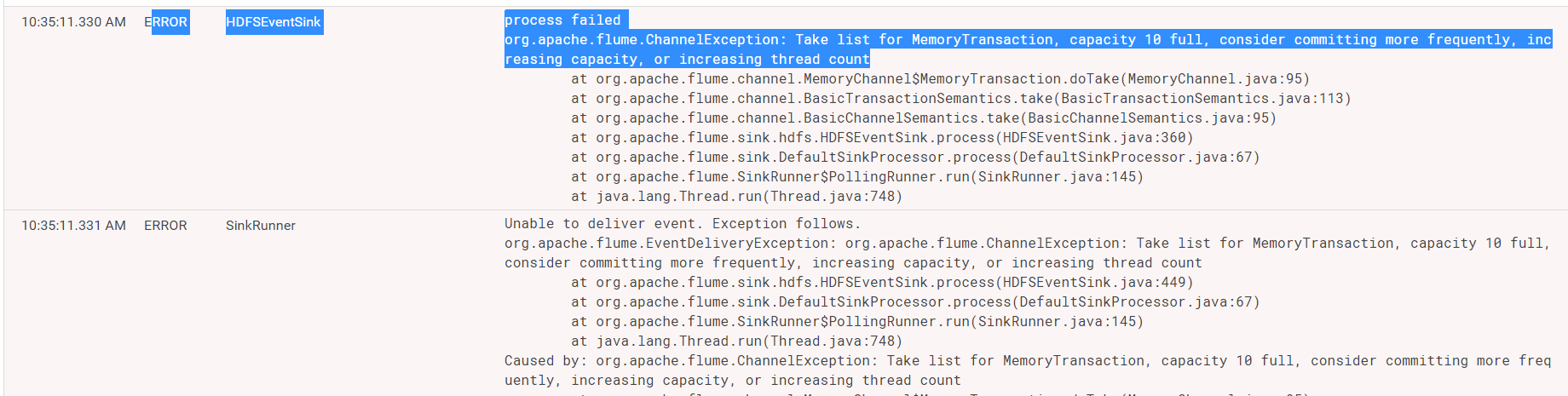
NetcatAgent.sinks = HDFS

В дополнение к вышесказанному посмотрим логи у Flume



Внимание привлекли следующие места:

Agent configuration for 'NetcatAgent' has no configfilters.



Возможно узким местом является строка:

NetcatAgent.channels.MemChannel.transactionCapacity = 10

**ЗАДАНИЕ**

2. Создать любой Flume поток используя Flume сервис соотвествующего номера.

• Тип источника источник – exeс

• Тип канала – file

• Тип слива – hdfs

**ВЫПОЛНЕНИЕ**

Всё будет выполняться под пользователем Flume

рекомендовано писать в корневую папку **/flume** (на нее есть разрешения в HDFS у пользователя Flume – см выше)

Для задачи воспользуемся потоком **Flume-5**

Источник: exec

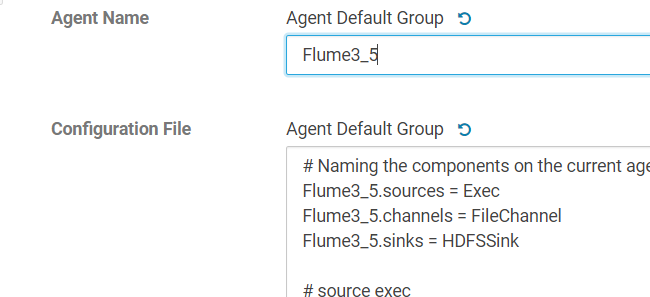
Это означает, что мы посылаем некую команду в консоль и Flume должен читать ее результат

Канал: file – это означает, что результаты будут сохраняться на локальной файловой системе

В качестве команды попробуем прочитать файл

**tailf /var/log/cron**

Зайдем в свойства потока и зададим ему конфигурацию:



# Naming the components on the current agent

Flume3\_5.sources = Exec

Flume3\_5.channels = FileChannel

Flume3\_5.sinks = HDFSSink

# source exec

Flume3\_5.sources.Exec.type = exec

Flume3\_5.sources.Exec.channels = FileChannel

Flume3\_5.sources.Exec.command = tailf /var/log/cron

# insert timestamp

Flume3\_5.sources.Exec.interceptors = tsi

Flume3\_5.sources.Exec.interceptors.tsi.type = timestamp

# channel

Flume3\_5.channels.FileChannel.type = file

#Flume3\_5.channels.FileChannel.checkpointDir = /tmp/flume/flume3\_5/checkpoint

#Flume3\_5.channels.FileChannel.dataDirs = /tmp/flume/flume3\_5/data

Flume3\_5.channels.FileChannel.checkpointDir = /tmp/student3\_5/checkpoint

Flume3\_5.channels.FileChannel.dataDirs = /tmp/student3\_5/data

Flume3\_5.channels.FileChannel.capacity = 1000

Flume3\_5.channels.FileChannel.transactionCapacity = 100

# sinks

Flume3\_5.sinks.HDFSSink.type = hdfs

Flume3\_5.sinks.HDFSSink.channel = FileChannel

Flume3\_5.sinks.HDFSSink.hdfs.useLocalTimeStamp = true

Flume3\_5.sinks.HDFSSink.hdfs.path = /flume/student3\_5/log/%y-%m-%d

Flume3\_5.sinks.HDFSSink.hdfs.filePrefix = hdfs-st35-

# File size to trigger roll, in bytes (256Mb)

Flume3\_5.sinks.HDFSSink.hdfs.rollSize = 1000

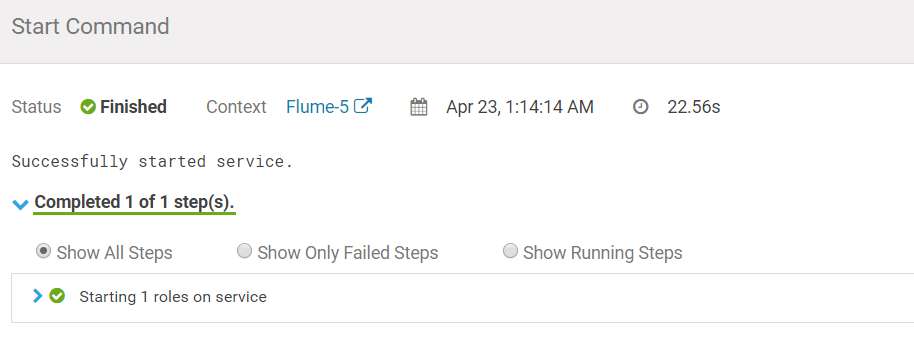
Flume3\_5.sinks.HDFSSink.hdfs.rollInterval = 0

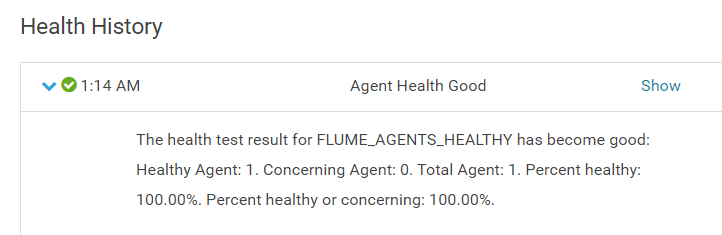
Flume3\_5.sinks.HDFSSink.hdfs.rollCount = 0

Flume3\_5.sinks.HDFSSink.hdfs.fileType = SequenceFile

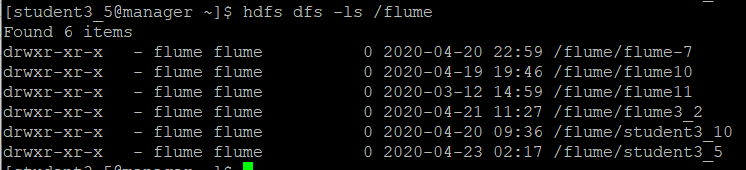
Flume3\_5.sinks.HDFSSink.hdfs.codeC = gzip

Запускаем поток





ПРОВЕРИМ содержимое целевой директории



Посмотрим содержимое файла:

