Hbase to Flume Data Sink

-start hbase services

```
saurabh@saurabh-VirtualBox:~$ start-hbase.sh
starting master, logging to /usr/local/hbase/bin/../logs/hbase-saurabh-master-saurabh-VirtualBox.out
OpenJDK 64-Bit Server VM warning: ignoring option PermSize=128m; support was removed in 8.0
OpenJDK 64-Bit Server VM warning: ignoring option MaxPermSize=128m; support was removed in 8.0
```

-check the service

```
saurabh@saurabh-VirtualBox:~$ jps
22150 Application
23370 Jps
5693 HMaster
```

-create hbase.conf file

```
# Name the components on this agent
Agent1.sources = netcat-source
Agent1.channels = memory-channel
Agent1.sinks = logger-sink

# Describe/configure Source
Agent1.sources.netcat-source.type = netcat
Agent1.sources.netcat-source.bind = localhost
Agent1.sources.netcat-source.port = 44444

# Describe the sink
Agent1.sinks.logger-sink.type = logger
Agent1.sinks.logger-sink.type=hbase
Agent1.sinks.logger-sink.table=test
Agent1.sinks.logger-sink.columnFamily=CF1
Agent1.sinks.logger-sink.type=org.apache.flume.sink.hbase.AsyncHBaseSink
#Agent1.sinks.logger-sink.serializer=org.apache.flume.sink.hbase.SimpleHbaseEventSerializer

# Use a channel which buffers events in memory
Agent1.channels.memory-channel.type = memory
Agent1.channels.memory-channel.capacity = 1000
Agent1.channels.memory-channel.transactionCapacity = 100

# Bind the source and sink to the channel
Agent1.sources.netcat-source.channels = memory-channel
Agent1.sinks.logger-sink.channel = memory-channel
```

-open hbase shell

```
saurabh@saurabh-VirtualBox:~$ hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.3.3, ra75a458e0c8f3c60db57f30ec4b06a606c9027b4, Fri Dec 14 16:02:53 PST 2018
```

-create table in hbase

table and column-family name should be same as .conf file

```
hbase(main):028:0> create 'test','CF1'
0 row(s) in 1.2680 seconds

=> Hbase::Table - test
```

-list the table

```
hbase(main):001:0> list
TABLE
airflow
emp
pagalpanti
tab1
test
user
6 row(s) in 0.8780 seconds
```

-open another terminal using this cmd.
ctrl+alt+t
write the flume cmd

```
saurabh@saurabh-VirtualBox:~$ flume-ng agent --conf ./conf/ -f /home/saurabh/hbase.conf -Dflume.root.logger=DEBUG,console -n Agent1
Info: Including Hadoop libraries found via (/usr/local/hadoop/bin/hadoop) for HDFS access
Info: Excluding /usr/local/hadoop/share/hadoop/common/lib/slf4j-api-1.7.5.jar from classpath
Info: Excluding /usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar from classpath
Info: Including HBASE libraries found via (/usr/local/hbase/bin/hbase) for HBASE access
Info: Excluding /usr/local/hbase/bin/../lib/slf4j-api-1.7.7.jar from classpath
Info: Excluding /usr/local/hbase/bin/../lib/slf4j-log4j12-1.7.5.jar from classpath
Info: Excluding /usr/local/hadoop/share/hadoop/common/lib/slf4j-api-1.7.5.jar from classpath
Info: Excluding /usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar from classpath
Info: Including Hive libraries found via () for Hive access
```

 flume-ng agent --conf ./conf/ -f /home/saurabh/hbase.conf -Dflume.root.logger=DEBUG,console
-n Agent1

```
sources:/usr/lib/jvm/java-8-openjdk-amd64/lib/tools.jar:/contrib/capacity-scheduler/*.jar:/conf:/lib/*' -Djava.library.path=:/usr/local/hadoop/lib/native:/usr/local/hadoop/lib/native org.apache.flume.node
.Application -f /home/saurabh/hbase.conf -n Agent1
19/05/06 13:47:41 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
19/05/06 13:47:41 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:/home/saurabh/hbase.conf
19/05/06 13:47:41 INFO conf.FlumeConfiguration: Added sinks: logger-sink Agent: Agent1
19/05/06 13:47:41 INFO conf.FlumeConfiguration: Processing:logger-sink
19/05/06 13:47:41 INFO conf.FlumeConfiguration: Processing:logger-sink
19/05/06 13:47:41 INFO conf.FlumeConfiguration: Processing:logger-sink
19/05/06 13:47:41 INFO conf.FlumeConfiguration: Processing:logger-sink
19/05/06 13:47:41 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [Agent1]
19/05/06 13:47:41 INFO node.AbstractConfigurationProvider: Creating channels
19/05/06 13:47:41 INFO channel.DefaultChannelFactory: Creating instance of channel memory-channel type memory
19/05/06 13:47:41 INFO node.AbstractConfigurationProvider: Created channel memory-channel
19/05/06 13:47:41 INFO source.DefaultSourceFactory: Creating instance of source netcat-source, type netcat
19/05/06 13:47:41 INFO sink.DefaultSinkFactory: Creating instance of sink: logger-sink, type: org.apache.flume.sink.hbase.AsyncHBaseSink
19/05/06 13:47:41 INFO hbase.AsyncHBaseSink: No serializer defined, Will use default
19/05/06 13:47:41 INFO hbase.AsyncHBaseSink: The write to WAL option is set to: true
19/05/06 13:47:42 INFO node.AbstractConfigurationProvider: Channel memory-channel connected to [netcat-source, logger-sink]
19/05/06 13:47:42 INFO node.Application: Starting new configuration:{ sourceRunners:{netcat-source=EventDrivenSourceRunner: { source:org.apache.flume.source.NetcatSource{name:netcat-source,state:IDLE} }}
sinkRunners:{logger-sink=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@46120ec7 counterGroup:{ name:null counters:{} } }} channels:{memory-channel=org.apache.flume.channel.MemoryChannel{
name: memory-channel}} }
19/05/06 13:47:42 INFO node.Application: Starting Channel memory-channel
19/05/06 13:47:42 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: memory-channel: Successfully registered new MBean.
19/05/06 13:47:42 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: memory-channel started
19/05/06 13:47:42 INFO node.Application: Starting Sink logger-sink
19/05/06 13:47:42 INFO node.Application: Starting Source netcat-source
19/05/06 13:47:42 INFO source.NetcatSource: Source starting
19/05/06 13:47:42 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: logger-sink: Successfully registered new MBean.
19/05/06 13:47:42 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: logger-sink started
19/05/06 13:47:42 INFO hbase.AsyncHBaseSink: Callback pool created
19/05/06 13:47:42 INFO source.NetcatSource: Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/127.0.0.1:44444]
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-1569965, built on 02/20/2014 09:09 GMT
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:host.name=saurabh-VirtualBox
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.version=1.8.0_191
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.home=/usr/lib/jvm/java-8-openjdk-amd64/jre
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.class.path=./conf/:/usr/local/flume/lib/jopt-simple-3.2.jar:/usr/local/flume/lib/commons-collections-3.2.1.jar:/usr/local/flume/lib/jack
son-databind-2.3.1.jar:/usr/local/flume/lib/parquet-column-1.4.1.jar:/usr/local/flume/lib/hive-serdes-1.0-SNAPSHOT.jar:/usr/local/flume/lib/flume-ng-auth-1.6.0.jar:/usr/local/flume/lib/paranamer-2.3.jar:/
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.library.path=:/usr/local/hadoop/lib/native:/usr/local/hadoop/lib/native
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.io.tmpdir=/tmp
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:java.compiler=<NA>
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:os.name=Linux
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:os.arch=amd64
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:os.version=4.15.0-47-generic
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:user.name=saurabh
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:user.home=/home/saurabh
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Client environment:user.dir=/home/saurabh
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Initiating client connection, connectString=localhost:2181 sessionTimeout=5000 watcher=org.hbase.async.HBaseClient$ZKClient@374e8391
19/05/06 13:47:43 INFO async.HBaseClient: Opening socket connection to server localhost/127.0.0.1:2181. Will not attempt to authenticate using SASL (unknown error)
19/05/06 13:47:43 INFO async.HBaseClient: Need to find the -ROOT- region
19/05/06 13:47:43 INFO hbase.AsyncHBaseSink: waiting on callback
19/05/06 13:47:43 INFO zookeeper.ClientCnxn: Socket connection established to localhost/127.0.0.1:2181, initiating session
19/05/06 13:47:43 INFO zookeeper.ClientCnxn: Session establishment complete on server localhost/127.0.0.1:2181, sessionid = 0x16a8bed3aea0018, negotiated timeout = 5000
19/05/06 13:47:43 INFO async.HBaseClient: Connecting to .META. region @ 127.0.1.1:30859
19/05/06 13:47:43 INFO zookeeper.ZooKeeper: Session: 0x16a8bed3aea0018 closed
19/05/06 13:47:43 INFO zookeeper.ClientCnxn: EventThread shut down
19/05/06 13:47:43 INFO async.HBaseClient: Added client for region RegionInfo(table="test", region_name="test,,1557129624284.6ed37b43e0dac18ece48d3ec660b58fa.", stop_key=""), which was added to the regions
  cache.  Now we know that RegionClient@156889438(chan=[id: 0xc1d4ff00, /127.0.0.1:18690 => /127.0.1.1:30859], #pending_rpcs=0, #batched=0, #rpcs_inflight=0) is hosting 1 region.
19/05/06 13:47:43 INFO hbase.AsyncHBaseSink: callback received
19/05/06 13:47:43 INFO hbase.AsyncHBaseSink: table found
```

open another terminal using this cmd
ctrl+alt+t
and type

```
saurabh@saurabh-VirtualBox:~$ nc localhost 44444
azmat
OK
```

-switch the terminal where the hbase terminal open and enter

```
hbase(main):032:0> scan 'test'
ROW                                              COLUMN+CELL
 default4f40bba8-2ea5-40af-96f8-6a294fb0beb8     column=CF1:pCol, timestamp=1557129647978, value=azmat
 default55f40b6d-3b20-4956-8ff6-cc60247e582e     column=CF1:pCol, timestamp=1557129629708, value=
 incRow                                          column=CF1:iCol, timestamp=1557129647979, value=\x00\x00\x00\x00\x00\x00\x00\x02
3 row(s) in 0.0430 seconds
```

insert the data

we can insert continously data

```
saurabh@saurabh-VirtualBox:~$ nc localhost 44444
azmat
OK
he
OK
```

Done !