Session 6

ADVANCE HBASE
Oozie and Flume

Assignment

Task1.1

Explain the below concepts with an example in brief.

- **Nosql Databases**

NoSQL database is the class of Database Management System which does not strictly follow principle of relational database system. These databases can accommodate not only structured data but also semi-structured and un-structured data. These database do not require table with fixed set of columns to store data and avoid join and support horizontal scaling. Below are the most popular NoSQL databases:

1. MongoDB - This is open source, highly scalable and agile NoSQL database
2. Redis - An open source, key value store of an advanced level.
3. Couch DB - This is powerful database for JSON based web applications. This database provides really powerful API to store JSON objects as documents in the database.
4. REVENDB - This DB is document oriented and schema free database. It provides extremely flexible and fast queries.
5. MemcacheDB - This is a distributed storage system of key value.
6. Neo4j - This is a NoSQL graph database which exhibits a high level of performance.
7. HBASE - HBase is a scalable, distributed and a big data store. This database can be used as real time and random access to data.
8. HyperGraphDB - This is an open source data storage system that is extensible, distributed, general purpose, portable and embeddable.
9. Cassandra - This is high available and scalable without compromising on performance, comes with fault tolerance and linear scalability along with best in class replication support.

- **Types of Nosql Databases**

There are 4 basic types of NoSQL databases:

1. Key-Value Store – The key value type basically, uses a hash table in which there exists a unique key and a pointer to a particular item of data. Example- Riak, Amazon S3 (Dynamo)

Key	Value
“BELLEVUE”	{“Office 2310, WA, Suite 2300, Bellevue - Bellevue Skyline Tower, United States”}

“VANCOUVER”	{“Office 2352, 1066, West Hasting, Street, 20th & 23rd Floors, Vancouver, BC V6E 3X2, Canada”}
“AUSTIN”	{“9020 N Capital of Texas Highway, Building # 01 Suite 220, Austin TX 78759, United States”}

2. Document-based Store- A document store quite similar to a key-value store, but the only difference is that the values stored (referred to as “documents”) provide some structure like XML, JSON. Example- CouchDB

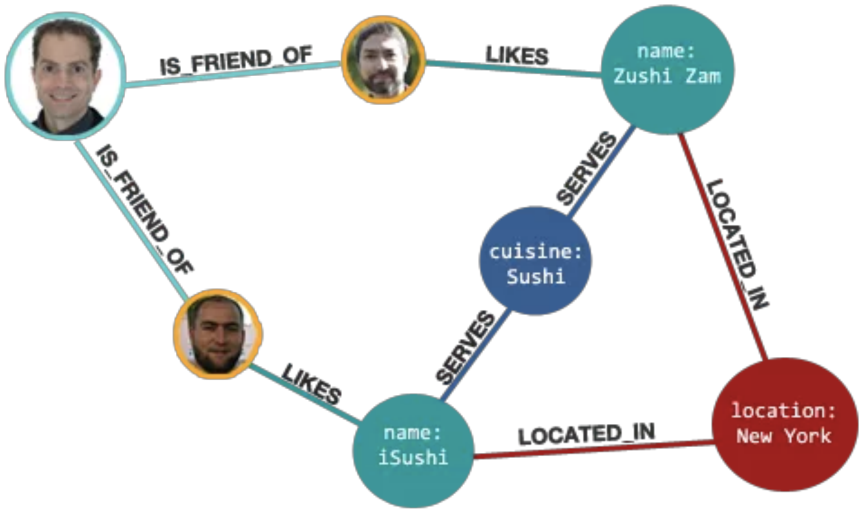
```
{
   data: [{
      manufacturer: 'Porsche',
      model: '911',
      price: 135000,
      wiki: 'http://en.wikipedia.org/wiki/Porsche_997',
      img: '2004_Porsche_911_Carrera_type_997.jpg'
   },{
      manufacturer: 'Nissan',
      model: 'GT-R',
      price: 80000,
      wiki:'http://en.wikipedia.org/wiki/Nissan_Gt-r',
      img: '250px-Nissan_GT-R.jpg'
   }]
}
```

3. Column-based Store- In column-oriented NoSQL database, data is stored in cells grouped in columns of data rather than as rows of data. Columns are logically grouped into column families. Column families can contain a virtually unlimited number of columns that can be created at runtime. Example- HBase, Cassandra

RowId	EmpId	Lastname	Firstname	Salary
1	10	Smith	Joe	40000
2	12	Jones	Mary	50000
3	11	Johnson	Cathy	44000
4	22	Jones	Bob	55000

10:001,12:002,11:003,22:004;
Smith:001,Jones:002,Johnson:003,Jones:004;
Joe:001,Mary:002,Cathy:003,Bob:004;
40000:001,50000:002,44000:003,55000:004;

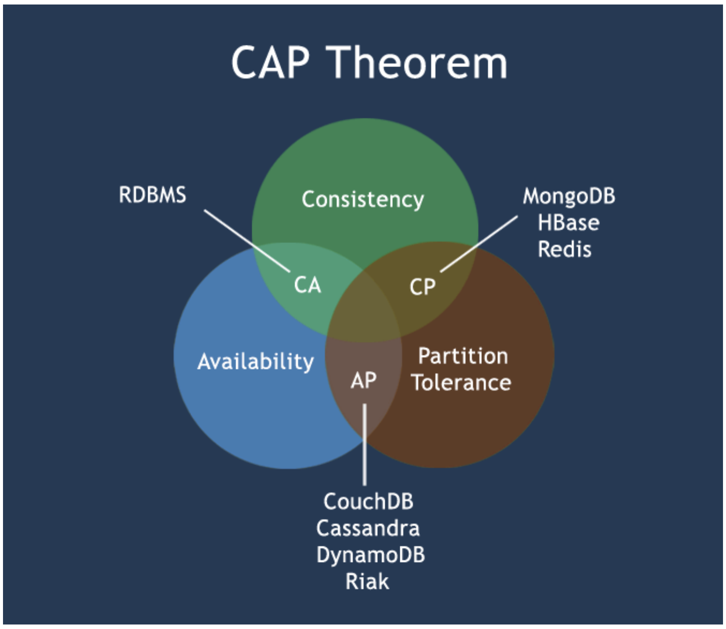
4. Graph-based - A flexible graphical representation is instead used which is perfect to address scalability concerns. Graph structures are used with edges, nodes and properties which provides index-free adjacency. Example- Neo4J



- **CAP Theorem**

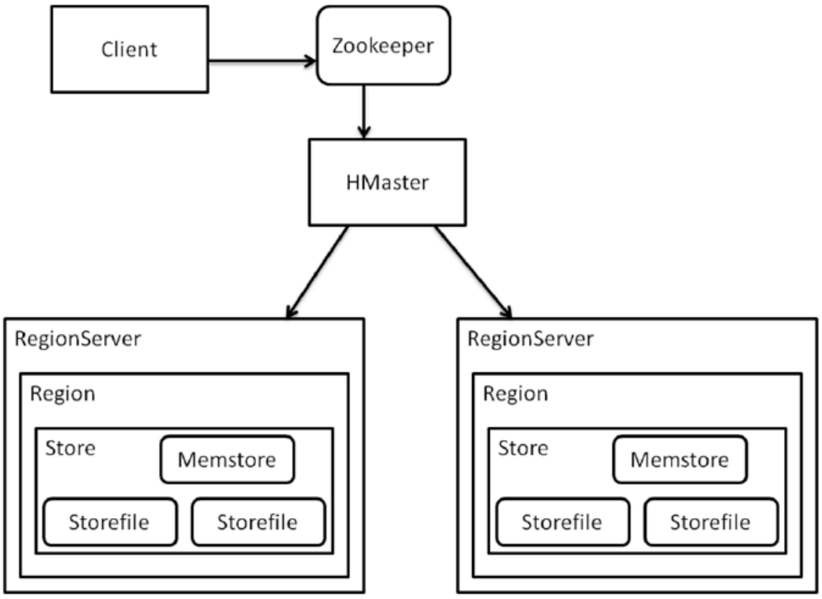
The CAP theorem, also known as Brewer's theorem, states that it is impossible for a distributed computer system to simultaneously provide all three of the following guarantees:

1. Consistency (all nodes see the same data at the same time)
2. Availability (a guarantee that every request receives a response about whether it was successful or failed)
3. Partition tolerance (the system continues to operate despite arbitrary message loss or failure of part of the system)



- **HBase Architecture**

The HBase cluster has one Master node, called HMaster and multiple Region Servers called HRegionServer. Each Region Server contains multiple Regions called HRegions. Also HBase uses ZooKeeper as a distributed coordination service to maintain server state in the cluster.
Data in HBase is stored in Tables and these Tables are stored in Regions. When a Table becomes too big, the Table is partitioned into multiple Regions. These Regions are assigned to Region Servers across the cluster. Each Region Server hosts roughly the same number of Regions.



- **HBase vs RDBMS**

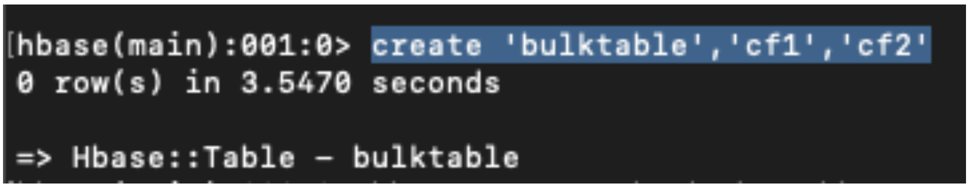
HBASE	RDBMS
Column-Oriented	Row-Oriented
Flexible schema, columns can be added run time	Fixed schema
Good with sparse table	Not optimised for sparse table
No query language	SQL
Not transactional	Transactional
Wide table	Narrow table
De-Normalized data	Normalized data
Join using Map-Reduce	Optimized join

Task1.2

ImportTSV Data from HDFS into HBase

Step-1

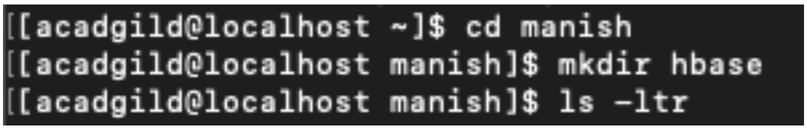
```
create ‘bulktable','cf1','cf2'
```



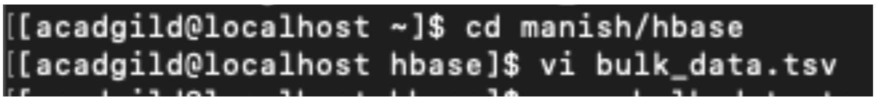
```
[hbase(main):001:0> create 'bulktable','cf1','cf2'
 0 row(s) in 3.5470 seconds

 => Hbase::Table - bulktable
```

Step-2

```
mkdir -p manish/hbase
```



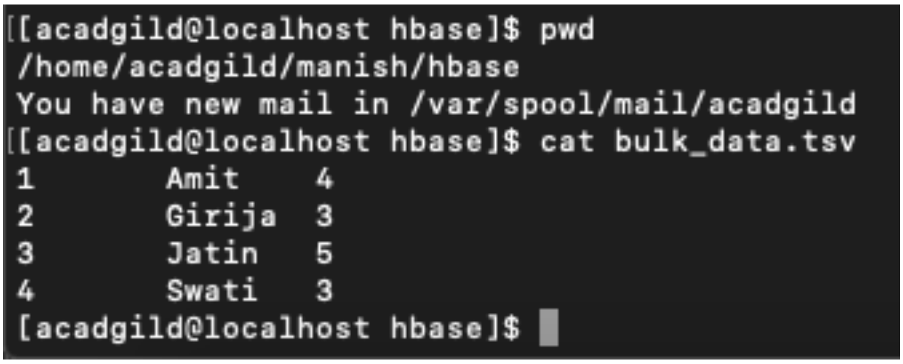
```
[[acadgild@localhost ~]$ cd manish
[[acadgild@localhost manish]$ mkdir hbase
[[acadgild@localhost manish]$ ls -ltr
```



```
[[acadgild@localhost ~]$ cd manish/hbase
[[acadgild@localhost hbase]$ vi bulk_data.tsv
```

Step-3

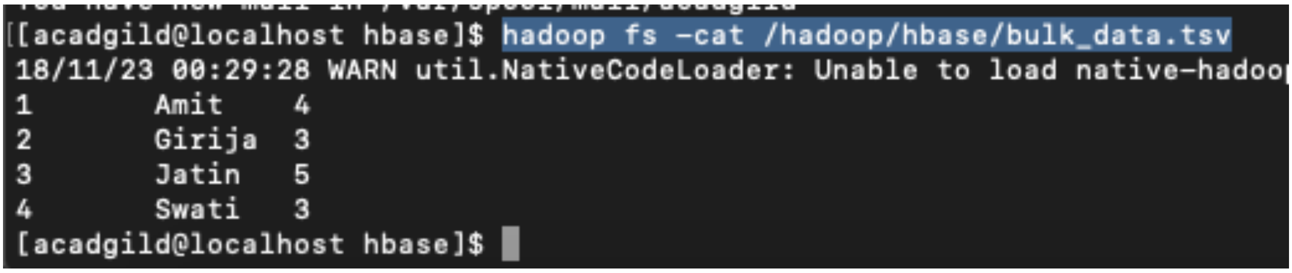
```
cat /home/acadgild/manish/hbase/bulk_data.tsv
```



```
[[acadgild@localhost hbase]$ pwd
 /home/acadgild/manish/hbase
 You have new mail in /var/spool/mail/acadgild
[[acadgild@localhost hbase]$ cat bulk_data.tsv
 1       Amit    4
 2       Girija  3
 3       Jatin   5
 4       Swati   3
 [acadgild@localhost hbase]$ █
```

Step-4

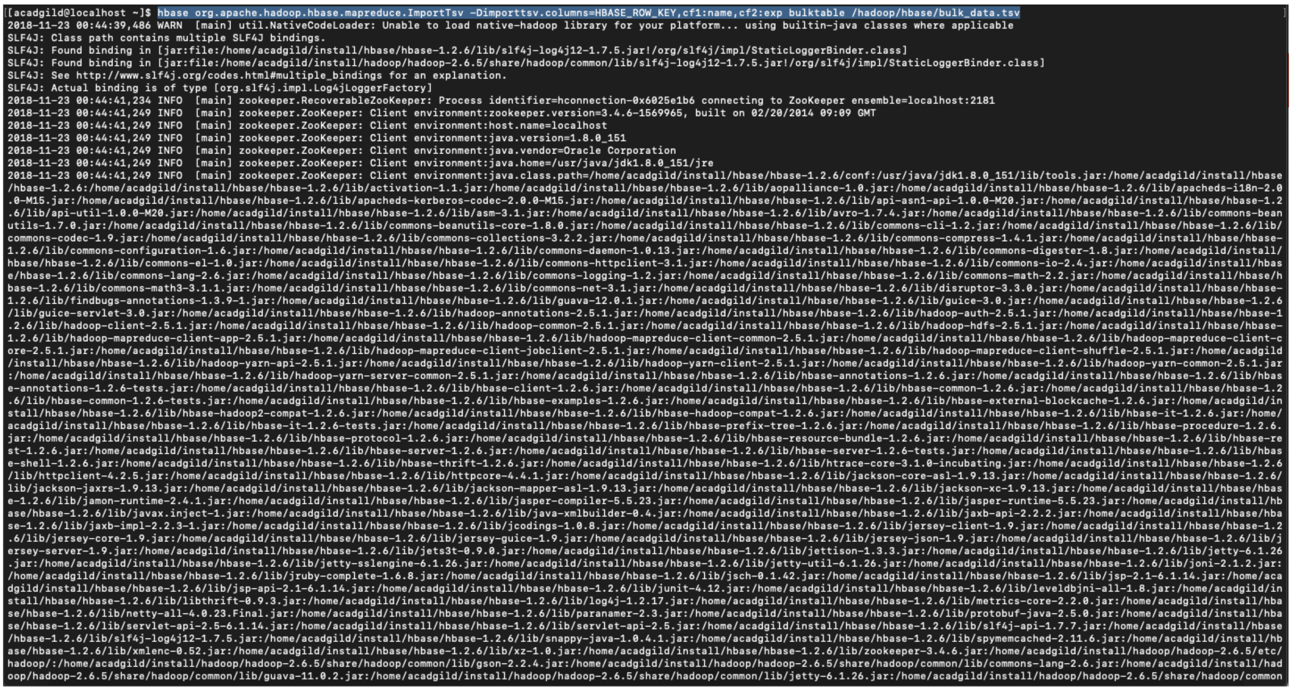
```
hadoop fs -mkdir -p /hadoop/hbase
hadoop fs -put bulk_data.tsv /hadoop/hbase
hadoop fs -cat /hadoop/hbase/bulk_data.tsv
```



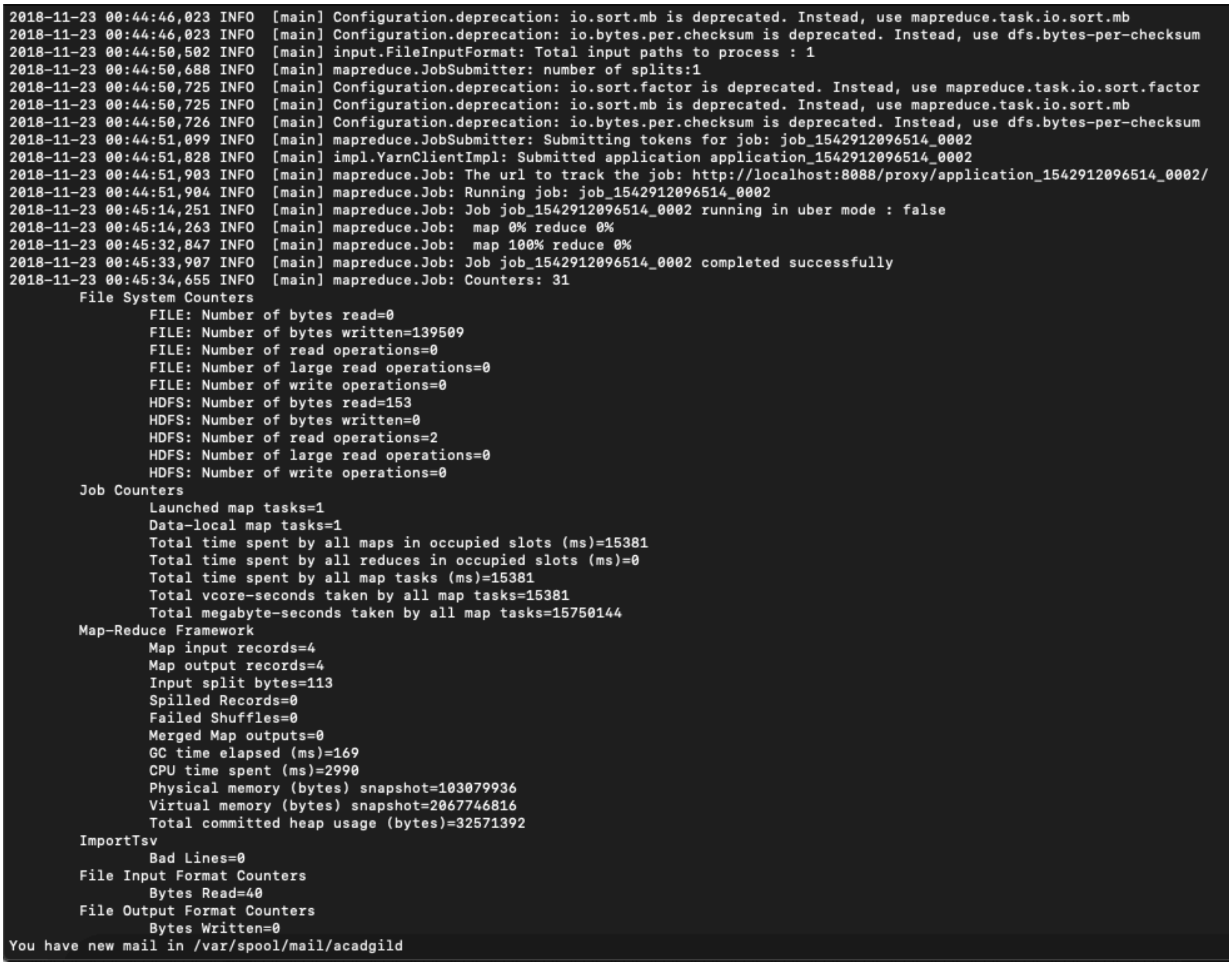
```
[[acadgild@localhost hbase]$ hadoop fs -cat /hadoop/hbase/bulk_data.tsv
 18/11/23 00:29:28 WARN util.NativeCodeLoader: Unable to load native-hadoo
 1       Amit    4
 2       Girija  3
 3       Jatin   5
 4       Swati   3
 [acadgild@localhost hbase]$ █
```

Step-5

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -
Dimporttsv.columns=HBASE_ROW_KEY,cf1:name,cf2:exp bulktable /hadoop/
hbase/bulk_data.tsv
```

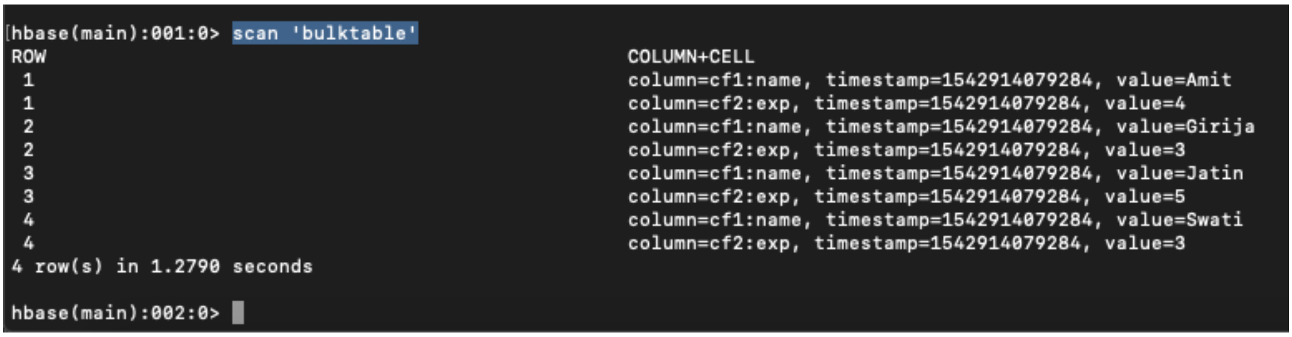


```
[acadgild@localhost ~]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.columns=HBASE_ROW_KEY,cf1:name,cf2:exp bulktable /hadoop/hbase/bulk_data.tsv
2018-11-23 00:44:39,486 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/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]
2018-11-23 00:44:41,234 INFO  [main] zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x6025e1b6 connecting to ZooKeeper ensemble=localhost:2181
2018-11-23 00:44:41,249 INFO  [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-1569965, built on 02/20/2014 09:09 GMT
2018-11-23 00:44:41,249 INFO  [main] zookeeper.ZooKeeper: Client environment:host.name=localhost
2018-11-23 00:44:41,249 INFO  [main] zookeeper.ZooKeeper: Client environment:java.version=1.8.0_151
2018-11-23 00:44:41,249 INFO  [main] zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
2018-11-23 00:44:41,249 INFO  [main] zookeeper.ZooKeeper: Client environment:java.home=/usr/java/jdk1.8.0_151/jre
2018-11-23 00:44:41,249 INFO  [main] zookeeper.ZooKeeper: Client environment:java.class.path=/home/acadgild/install/hbase/hbase-1.2.6/conf:/usr/java/jdk1.8.0_151/lib/tools.jar:/home/acadgild/install/hbase
/hbase-1.2.6:/home/acadgild/install/hbase/hbase-1.2.6/lib/activation-1.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/aopalliance-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/apacheds-i18n-2.0
.0-M15.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/api-asn1-api-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2
.6/lib/api-util-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/asm-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/avro-1.7.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-bean
utils-1.7.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-core-1.8.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-cli-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/
commons-codec-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-collections-3.2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-compress-1.4.1.jar:/home/acadgild/install/hbase/hbase-
1.2.6/lib/commons-configuration-1.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-daemon-1.0.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-digester-1.8.jar:/home/acadgild/install/
hbase/hbase-1.2.6/lib/commons-el-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-httpclient-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-io-2.4.jar:/home/acadgild/install/hbas
e/hbase-1.2.6/lib/commons-lang-2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-logging-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math-2.2.jar:/home/acadgild/install/hbase/h
base-1.2.6/lib/commons-math3-3.1.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-net-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/disruptor-3.3.0.jar:/home/acadgild/install/hbase/hbase-
1.2.6/lib/findbugs-annotations-1.3.9-1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guava-12.0.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6
/lib/guice-servlet-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-annotations-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-auth-2.5.1.jar:/home/acadgild/install/hbase/hbase-1
.2.6/lib/hadoop-client-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-hdfs-2.5.1.jar:/home/acadgild/install/hbase/hbase-
1.2.6/lib/hadoop-mapreduce-client-app-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-c
ore-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-jobclient-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-shuffle-2.5.1.jar:/home/acadgild
/install/hbase/hbase-1.2.6/lib/hadoop-yarn-api-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-client-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-common-2.5.1.jar
:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-server-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-annotations-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbas
e-annotations-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-client-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2
.6/lib/hbase-common-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-examples-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-external-blockcache-1.2.6.jar:/home/acadgild/in
stall/hbase/hbase-1.2.6/lib/hbase-hadoop2-compat-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-hadoop-compat-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-it-1.2.6.jar:/home/
acadgild/install/hbase/hbase-1.2.6/lib/hbase-it-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-prefix-tree-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-procedure-1.2.6.
jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-protocol-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-resource-bundle-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-re
st-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-server-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-server-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbas
e-shell-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-thrift-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/htrace-core-3.1.0-incubating.jar:/home/acadgild/install/hbase/hbase-1.2.6
/lib/httpclient-4.2.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/httpcore-4.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-core-asl-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/
lib/jackson-jaxrs-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-mapper-asl-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-xc-1.9.13.jar:/home/acadgild/install/hbase/hbas
e-1.2.6/lib/jamon-runtime-2.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jasper-compiler-5.5.23.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jasper-runtime-5.5.23.jar:/home/acadgild/install/hb
ase/hbase-1.2.6/lib/javax.inject-1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/java-xmlbuilder-0.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jaxb-api-2.2.2.jar:/home/acadgild/install/hbase/hba
se-1.2.6/lib/jaxb-impl-2.2.3-1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jcodings-1.0.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-client-1.9.jar:/home/acadgild/install/hbase/hbase-1.2
.6/lib/jersey-core-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-guice-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-json-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/j
ersey-server-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jets3t-0.9.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jettison-1.3.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-6.1.26
.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-sslengine-6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-util-6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/joni-2.1.2.jar:
/home/acadgild/install/hbase/hbase-1.2.6/lib/jruby-complete-1.6.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsch-0.1.42.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsp-2.1-6.1.14.jar:/home/aca
dgild/install/hbase/hbase-1.2.6/lib/jsp-api-2.1-6.1.14.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/junit-4.12.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/leveldbjni-all-1.8.jar:/home/acadgild/in
stall/hbase/hbase-1.2.6/lib/libthrift-0.9.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/log4j-1.2.17.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/metrics-core-2.2.0.jar:/home/acadgild/install/hba
se/hbase-1.2.6/lib/netty-all-4.0.23.Final.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/paranamer-2.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/protobuf-java-2.5.0.jar:/home/acadgild/install/hba
se/hbase-1.2.6/lib/servlet-api-2.5-6.1.14.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/servlet-api-2.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-api-1.7.7.jar:/home/acadgild/install/hbase
/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/snappy-java-1.0.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/spymemcached-2.11.6.jar:/home/acadgild/install/hb
ase/hbase-1.2.6/lib/xmlenc-0.52.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/xz-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/zookeeper-3.4.6.jar:/home/acadgild/install/hadoop/hadoop-2.6.5/etc/
hadoop/:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/gson-2.2.4.jar:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/commons-lang-2.6.jar:/home/acadgild/install/had
oop/hadoop-2.6.5/share/hadoop/common/lib/guava-11.0.2.jar:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/jetty-6.1.26.jar:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common
```



```
2018-11-23 00:44:46,023 INFO  [main] Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
2018-11-23 00:44:46,023 INFO  [main] Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2018-11-23 00:44:50,502 INFO  [main] input.FileInputFormat: Total input paths to process : 1
2018-11-23 00:44:50,688 INFO  [main] mapreduce.JobSubmitter: number of splits:1
2018-11-23 00:44:50,725 INFO  [main] Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
2018-11-23 00:44:50,725 INFO  [main] Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
2018-11-23 00:44:50,726 INFO  [main] Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2018-11-23 00:44:51,099 INFO  [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1542912096514_0002
2018-11-23 00:44:51,828 INFO  [main] impl.YarnClientImpl: Submitted application application_1542912096514_0002
2018-11-23 00:44:51,903 INFO  [main] mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1542912096514_0002/
2018-11-23 00:44:51,904 INFO  [main] mapreduce.Job: Running job: job_1542912096514_0002
2018-11-23 00:45:14,251 INFO  [main] mapreduce.Job: Job job_1542912096514_0002 running in uber mode : false
2018-11-23 00:45:14,263 INFO  [main] mapreduce.Job:  map 0% reduce 0%
2018-11-23 00:45:32,847 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2018-11-23 00:45:33,907 INFO  [main] mapreduce.Job: Job job_1542912096514_0002 completed successfully
2018-11-23 00:45:34,655 INFO  [main] mapreduce.Job: Counters: 31
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=139509
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=153
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=2
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=15381
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=15381
                Total vcore-seconds taken by all map tasks=15381
                Total megabyte-seconds taken by all map tasks=15750144
        Map-Reduce Framework
                Map input records=4
                Map output records=4
                Input split bytes=113
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=169
                CPU time spent (ms)=2990
                Physical memory (bytes) snapshot=103079936
                Virtual memory (bytes) snapshot=2067746816
                Total committed heap usage (bytes)=32571392
        ImportTsv
                Bad Lines=0
        File Input Format Counters
                Bytes Read=40
        File Output Format Counters
                Bytes Written=0
You have new mail in /var/spool/mail/acadgild
```

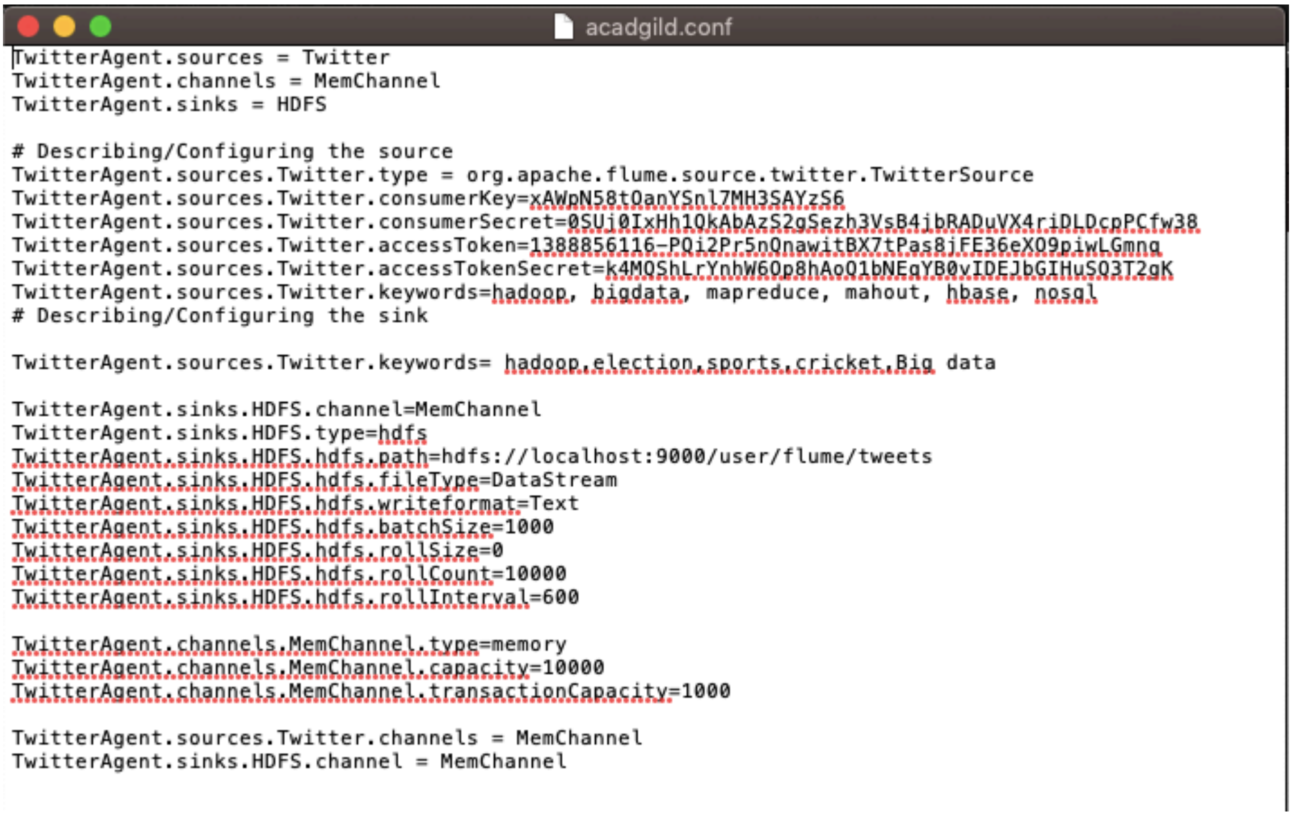
```
scan ‘bulktable'
```



```
hbase(main):001:0> scan 'bulktable'
ROW                                                   COLUMN+CELL
 1                                                    column=cf1:name, timestamp=1542914079284, value=Amit
 1                                                    column=cf2:exp, timestamp=1542914079284, value=4
 2                                                    column=cf1:name, timestamp=1542914079284, value=Girija
 2                                                    column=cf2:exp, timestamp=1542914079284, value=3
 3                                                    column=cf1:name, timestamp=1542914079284, value=Jatin
 3                                                    column=cf2:exp, timestamp=1542914079284, value=5
 4                                                    column=cf1:name, timestamp=1542914079284, value=Swati
 4                                                    column=cf2:exp, timestamp=1542914079284, value=3
4 row(s) in 1.2790 seconds

hbase(main):002:0>
```

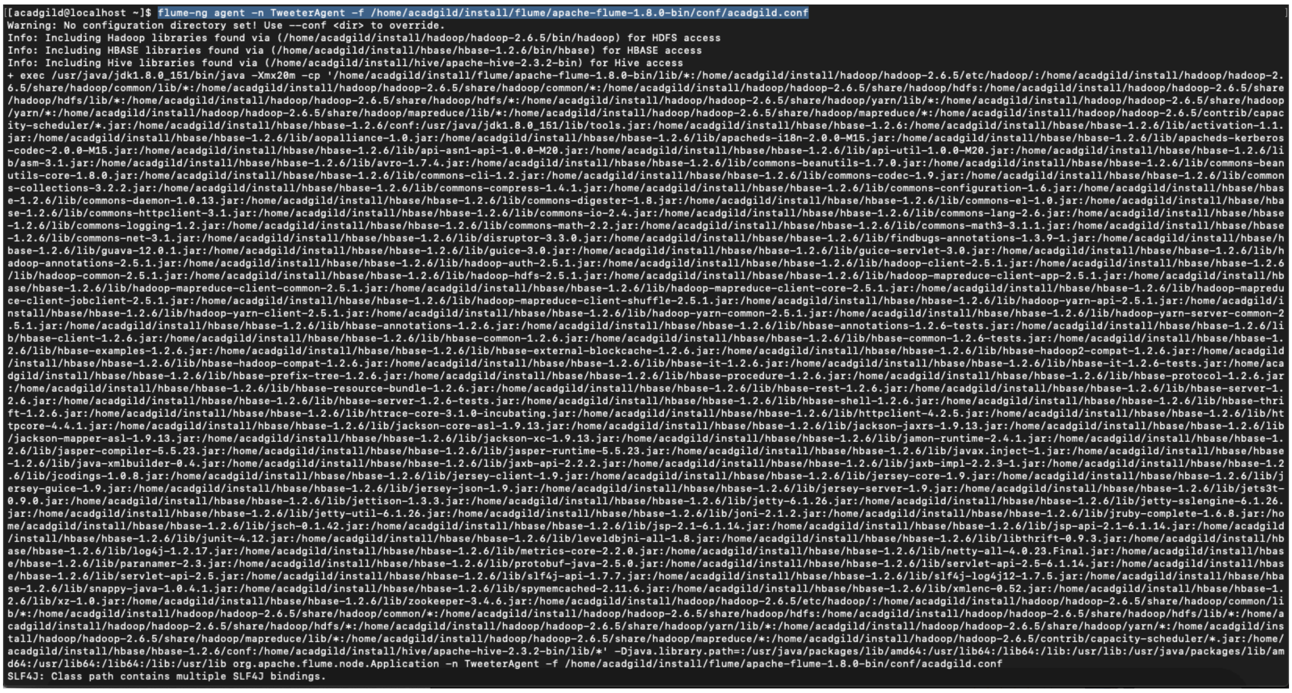
Task 2.1

Create a flume agent that streams data from Twitter and stores in the HDFS.



```
TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=xAWpN58tOanYSnl7MH3SAYzS6
TwitterAgent.sources.Twitter.consumerSecret=0SUj0IxHh1QkAbAzS2gSezh3VsB4jbRADuVX4riDLDcpPCfw38
TwitterAgent.sources.Twitter.accessToken=1388856116-PQi2Pr5nQnawitBX7tPas8jFE36eXO9piwLGmnq
TwitterAgent.sources.Twitter.accessTokenSecret=k4MQShLrYnhW6Op8hAoQ1bNEqYB0vIDEJbGIHuSQ3T2gK
TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql
# Describing/Configuring the sink

TwitterAgent.sources.Twitter.keywords= hadoop,election,sports,cricket,Big data

TwitterAgent.sinks.HDFS.channel=MemChannel
TwitterAgent.sinks.HDFS.type=hdfs
TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:9000/user/flume/tweets
TwitterAgent.sinks.HDFS.hdfs.fileType=DataStream
TwitterAgent.sinks.HDFS.hdfs.writeformat=Text
TwitterAgent.sinks.HDFS.hdfs.batchSize=1000
TwitterAgent.sinks.HDFS.hdfs.rollSize=0
TwitterAgent.sinks.HDFS.hdfs.rollCount=10000
TwitterAgent.sinks.HDFS.hdfs.rollInterval=600

TwitterAgent.channels.MemChannel.type=memory
TwitterAgent.channels.MemChannel.capacity=10000
TwitterAgent.channels.MemChannel.transactionCapacity=1000

TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sinks.HDFS.channel = MemChannel
```

```
flume-ng agent -n TweeterAgent -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/acadgild.conf
```



```
[acadgild@localhost ~]$ flume-ng agent -n TweeterAgent -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/acadgild.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop) for HDFS access
Info: Including HBASE libraries found via (/home/acadgild/install/hbase/hbase-1.2.6/bin/hbase) for HBASE access
Info: Including Hive libraries found via (/home/acadgild/install/hive/apache-hive-2.3.2-bin) for Hive access
+ exec /usr/java/jdk1.8.0_151/bin/java -Xmx20m -cp '/home/acadgild/install/flume/apache-flume-1.8.0-bin/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/etc/hadoop/:/home/acadgild/install/hadoop/hadoop-2.
6.5/share/hadoop/common/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs:/home/acadgild/install/hadoop/hadoop-2.6.5/share
/hadoop/hdfs/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop
/yarn/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/*:/home/acadgild/install/hadoop/hadoop-2.6.5/contrib/capac
ity-scheduler/*.jar:/home/acadgild/install/hbase/hbase-1.2.6/conf:/usr/java/jdk1.8.0_151/lib/tools.jar:/home/acadgild/install/hbase/hbase-1.2.6:/home/acadgild/install/hbase/hbase-1.2.6/lib/activation-1.1.
jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/aopalliance-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/apacheds-i18n-2.0.0-M15.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/apacheds-kerberos
-codec-2.0.0-M15.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/api-asn1-api-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/api-util-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/li
b/asm-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/avro-1.7.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-1.7.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-bean
utils-core-1.8.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-cli-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-codec-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/common
s-collections-3.2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-compress-1.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-configuration-1.6.jar:/home/acadgild/install/hbase/hbas
e-1.2.6/lib/commons-daemon-1.0.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-digester-1.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-el-1.0.jar:/home/acadgild/install/hbase/hba
se-1.2.6/lib/commons-httpclient-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-io-2.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-lang-2.6.jar:/home/acadgild/install/hbase/hbase
-1.2.6/lib/commons-logging-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math-2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math3-3.1.1.jar:/home/acadgild/install/hbase/hbase
-1.2.6/lib/commons-net-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/disruptor-3.3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/findbugs-annotations-1.3.9-1.jar:/home/acadgild/install/hbase/h
base-1.2.6/lib/guava-12.0.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-servlet-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/h
adoop-annotations-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-auth-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-client-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6
/lib/hadoop-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-hdfs-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-app-2.5.1.jar:/home/acadgild/install/hb
ase/hbase-1.2.6/lib/hadoop-mapreduce-client-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-core-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapredu
ce-client-jobclient-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-shuffle-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-api-2.5.1.jar:/home/acadgild/i
nstall/hbase/hbase-1.2.6/lib/hadoop-yarn-client-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-server-common-2
.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-annotations-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-annotations-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/li
b/hbase-client-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.
2.6/lib/hbase-examples-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-external-blockcache-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-hadoop2-compat-1.2.6.jar:/home/acadgild
/install/hbase/hbase-1.2.6/lib/hbase-hadoop-compat-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-it-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-it-1.2.6-tests.jar:/home/aca
dgild/install/hbase/hbase-1.2.6/lib/hbase-prefix-tree-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-procedure-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-protocol-1.2.6.jar
:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-resource-bundle-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-rest-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-server-1.
2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-server-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-shell-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-thri
ft-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/htrace-core-3.1.0-incubating.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/httpclient-4.2.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/ht
tpcore-4.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-core-asl-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-jaxrs-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib
/jackson-mapper-asl-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-xc-1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jamon-runtime-2.4.1.jar:/home/acadgild/install/hbase/hbase-1.
2.6/lib/jasper-compiler-5.5.23.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jasper-runtime-5.5.23.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/javax.inject-1.jar:/home/acadgild/install/hbase/hbase
-1.2.6/lib/java-xmlbuilder-0.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jaxb-api-2.2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jaxb-impl-2.2.3-1.jar:/home/acadgild/install/hbase/hbase-1.2
.6/lib/jcodings-1.0.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-client-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-core-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/j
ersey-guice-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-json-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-server-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jets3t-
0.9.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jettison-1.3.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-sslengine-6.1.26.
jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-util-6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/joni-2.1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jruby-complete-1.6.8.jar:/ho
me/acadgild/install/hbase/hbase-1.2.6/lib/jsch-0.1.42.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsp-2.1-6.1.14.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsp-api-2.1-6.1.14.jar:/home/acadgild
/install/hbase/hbase-1.2.6/lib/junit-4.12.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/leveldbjni-all-1.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/libthrift-0.9.3.jar:/home/acadgild/install/hb
ase/hbase-1.2.6/lib/log4j-1.2.17.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/metrics-core-2.2.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/netty-all-4.0.23.Final.jar:/home/acadgild/install/hbas
e/hbase-1.2.6/lib/paranamer-2.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/protobuf-java-2.5.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/servlet-api-2.5-6.1.14.jar:/home/acadgild/install/hbas
e/hbase-1.2.6/lib/servlet-api-2.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-api-1.7.7.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar:/home/acadgild/install/hbase/hba
se-1.2.6/lib/snappy-java-1.0.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/spymemcached-2.11.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/xmlenc-0.52.jar:/home/acadgild/install/hbase/hbase-1.
2.6/lib/xz-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/zookeeper-3.4.6.jar:/home/acadgild/install/hadoop/hadoop-2.6.5/etc/hadoop/:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/li
b/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs/lib/*:/home/a
cadgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/*:/home/acadgild/ins
tall/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/*:/home/acadgild/install/hadoop/hadoop-2.6.5/contrib/capacity-scheduler/*.jar:/home/
acadgild/install/hbase/hbase-1.2.6/conf:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/*' -Djava.library.path=:/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib:/usr/java/packages/lib/am
d64:/usr/lib64:/lib64:/lib:/usr/lib org.apache.flume.node.Application -n TweeterAgent -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/acadgild.conf
SLF4J: Class path contains multiple SLF4J bindings.
```