NOTES:

1. I have taken screenshots from Cloudera VM as it is capable enough and just tried to save some cost associated with use of AWS EC2 instance.
2. Password for root user of MySQL database is “cloudera” in Cloudera VM so in ReportBolt.java, password variable is set to “cloudera”. Please feel free to change if you are going to run on AWS EC2 instance.
3. In order to avoid frequent updates in database, I have set limit to 100000. After 100000 Tuples are processed, results will then be stored in MySQL database. This limit is controlled via cnt_tuples_update variable in ReportBolt.java which can be changed if needed.
4. It took me quite some time to take all different screenshots due to which numbers under Emitted or Acked header etc. in various screenshots may not match. For example, number of Acked messages appearing for SentenceSpout in screenshot taken for topology summary may differ from number of Acked messages appearing in screenshot taken for SentenceSpout. Taking screenshot was quite time taking so there is a time difference between different screenshots and so numbers will mismatch on different screenshots.

Solution to Load Imbalance Problem

In order to solve load imbalance problem and to ensure uniform distribution of Tuples among all instances of WordCountBolt, I am using shuffleGrouping rather than fieldsGrouping.

Shuffle grouping: Tuples are randomly distributed across the bolt's tasks in a way such that each bolt is guaranteed to get an equal number of tuples.

In WordCountTopology.java, when WordCountBolt is set, shuffleGrouping is used due to which Tuples emitted by SplitSentenceBolt will be uniformly distributed among all 4 instances of WordCountBolt. This will be clear from WordCountBolt table’s screenshot at a later stage in this document.

builder.setBolt(COUNT_BOLT_ID, countBolt, 4).shuffleGrouping(SPLIT_BOLT_ID);

In WordCountBolt, I am just passing the tuples to ReportBolt without any processing. In ReportBolt, aggregation is being done. ReportBolt stores {word, count} as {key, value} pair in a map. If word is already present in map, its count is incremented by 1 and if not then word is added in map with count as 1. ReportBolt takes care of final word count and stores result in MySQL database.

I have created a database **‘upgrad’** in MySQL and a table **‘wordcounts’** in upgrad database. Commands to do the same, has been added in later section of this document. In ReportBolt, I have setup a kind of merge statement in which {word, count} will be updated if word already exists in table and will be inserted if word does not exists already in table. Table wordcounts has 2 columns namely word and count where word column is declared as primary key.

As per problem statement of this assignment, following components are maintained in WordCountTopology:

1. One instance of SentenceSpout
2. Three instances of SplitSentenceBolt
3. Four instances of WordCountBolt
4. One instance of ReportBolt

Entire Stream of Tuples coming from SplitSentenceBolt are uniformly distributed among all instances of WordCountBolt.

Topology stores results (count of words) in MySQL database.

Reliability API is also implemented to ensure at least once processing of input messages which is explained next.

Implementation of Reliability API

I have implemented Reliability API along with replay mechanism for failed tuples. In order to implement reliability API, following changes are done:

1. **SentenceSpout:**

   (a) When the Tuple is emitted from the Spout a unique message ID is attached to the Tuple.
   (b) This messageID is used by Storm to identify the Tuple later. I have used it as type of java.util.UUID.
   (c) The emit method of SpoutOutputCollector takes msgID as its second parameter.

2. **SplitSentenceBolt:**

   (a) Tuple is anchored in the bolt. When Tuple is emitted, original Tuple is passed as first argument to emit method of OutputCollector.
   (b) Once Tuple is processed, it is acknowledged.

3. **WordCountBolt:**

   (a) Tuple is anchored in the bolt. When Tuple is emitted, original Tuple is passed as first argument to emit method of OutputCollector.
   (b) Once Tuple is processed, it is acknowledged.

4. **ReportBolt:**

   (a) Once Tuple is processed, it is acknowledged.
   (b) It does not emit any tuple. It stores the word and count in MySQL database.

In order to implement **replay mechanism**, following changes are done:

1. **SentenceSpout:**

   (a) 2 Maps are created namely toSend and messages in open method of SentenceSpout extending BaseRichSpout.
   (b) Map messages stores all emitted Tuples where msgID is the key and Tuple is the value.
   (c) Map toSend stores all failed Tuples so can be replayed.
   (d) In nextTuple method, first it is checked if toSend has any Tuples. If yes, then all pairs are fetched from collection and all Tuples are emitted again (replayed) and then toSend is cleared. Further, new object of UUID is created namely msgID and using randomUUID method, random unique UUID is generated and it is attached with newly emitted Tuple which is also added in messages Map with msgID as key and Tuple as value. Sleep of 1 ms is added at the end to avoid overloading storm topology.
   (e) In ack() method, Acknowledged msgID is printed and it is removed from the messages Map.
   (f) In fail() method, Failed msgID is printed and it is then stored in toSend Map with msgID as key and Tuple as value which is obtained from messages Map using get method with msgID as argument.

2. **ReportBolt:**

   (a) I have setup a boolean namely “acking” to ‘true’ by default so every processed tuple is acknowledged. If we want to test replay of failed Tuples then set this boolean to ‘false’ and then every Tuple will fail and will then be replayed by SentenceSpout. Logs can be checked to see failed message IDs being replayed.

Submission of Storm Topology

(A)

After copying provided StormAssignmentWordCount-0.0.1-SNAPSHOT.jar in home directory (/home/ec2-user) of ec2-user on EC2 instance or in home directory of cloudera user (/home/cloudera) on Cloudera VM, just cross verify:

[cloudera@quickstart ~]$ ls -lrt StormAssignmentWordCount-0.0.1-SNAPSHOT.jar
-rwxr-xr-x 1 cloudera cloudera 1004815 Jan 11 03:06 StormAssignmentWordCount-0.0.1-SNAPSHOT.jar

```
[cloudera@quickstart ~]$ ls -lrt StormAssignmentWordCount-0.0.1-SNAPSHOT.jar
-rwxr-xr-x 1 cloudera cloudera 1004815 Jan 11 03:06 StormAssignmentWordCount-0.0.1-SNAPSHOT.jar
```

Or

ls -lrt /home/ec2-user/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar

(B)

Go to directory where Apache Storm is installed.

[cloudera@quickstart ~]$ cd apache-storm-1.2.1

```
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ cd apache-storm-1.2.1
[cloudera@quickstart apache-storm-1.2.1]$
```

Or

cd /home/ec2-user/apache-storm-1.2.1

(C)

In order to submit storm topology in local mode, follow below command:

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm jar /home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar com.upgrad.wordcount.WordCountTopology

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm jar /home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar com.upgrad.wordcount.WordCountTopology
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm jar /home/ec2-user/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar com.upgrad.wordcount.WordCountTopology

(D)

Before we submit topology in production mode, we need to start nimbus, supervisor, ui and logviewer using below commands:

(a)

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm nimbus

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm nimbus
Running: /usr/java/jdk1.8.0_161/bin/java -server -Ddaemon.name=nimbus -Dstorm.options= -Dstorm.home=/home/cloudera/apache-storm-1.2.1 -Dstorm.log.dir=/home/
loudera/apache-storm-1.2.1/logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.conf.file= -cp /home/cloudera/apache-storm-1.2.1/*:/home/
loudera/apache-storm-1.2.1/lib/*:/home/cloudera/apache-storm-1.2.1/extlib/*:/home/cloudera/apache-storm-1.2.1/extlib-daemon/*:/home/cloudera/apache-storm-1.
.1/conf -Xmx1024m -Dlogfile.name=nimbus.log -DLog4jContextSelector=org.apache.logging.log4j.core.async.AsyncLoggerContextSelector -Dlog4j.configurationFile=
home/cloudera/apache-storm-1.2.1/log4j2/cluster.xml org.apache.storm.daemon.nimbus
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm nimbus

(b)

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm supervisor

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm supervisor
Running: /usr/java/jdk1.8.0_161/bin/java -server -Ddaemon.name=supervisor -Dstorm.options= -Dstorm.home=/home/cloudera/apache-storm-1.2.1 -Dstorm.log.dir=/h
me/cloudera/apache-storm-1.2.1/logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.conf.file= -cp /home/cloudera/apache-storm-1.2.1/*:/h
me/cloudera/apache-storm-1.2.1/lib/*:/home/cloudera/apache-storm-1.2.1/extlib/*:/home/cloudera/apache-storm-1.2.1/extlib-daemon/*:/home/cloudera/apache-stor
-1.2.1/conf -Xmx256m -Dlogfile.name=supervisor.log -Dlog4j.configurationFile=/home/cloudera/apache-storm-1.2.1/log4j2/cluster.xml org.apache.storm.daemon.su
ervisor.Supervisor
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm supervisor

(c)

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm ui

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm ui
Running: /usr/java/jdk1.8.0_161/bin/java -server -Ddaemon.name=ui -Dstorm.options= -Dstorm.home=/home/cloudera/apache-storm-1.2.1 -Dstorm.log.dir=/home/clou
era/apache-storm-1.2.1/logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.conf.file= -cp /home/cloudera/apache-storm-1.2.1/*:/home/clou
era/apache-storm-1.2.1/lib/*:/home/cloudera/apache-storm-1.2.1/extlib/*:/home/cloudera/apache-storm-1.2.1/extlib-daemon/*:/home/cloudera/apache-storm-1.2.1:
home/cloudera/apache-storm-1.2.1/conf -Xmx768m -Dlogfile.name=ui.log -DLog4jContextSelector=org.apache.logging.log4j.core.async.AsyncLoggerContextSelector -
log4j.configurationFile=/home/cloudera/apache-storm-1.2.1/log4j2/cluster.xml org.apache.storm.ui.core
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm ui

(d)

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm logviewer

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm logviewer
Running: /usr/java/jdk1.8.0_161/bin/java -server -Ddaemon.name=logviewer -Dstorm.options= -Dstorm.home=/home/cloudera/apache-storm-1.2.1 -Dstorm.log.dir=/ho
e/cloudera/apache-storm-1.2.1/logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.conf.file= -cp /home/cloudera/apache-storm-1.2.1/*:/ho
e/cloudera/apache-storm-1.2.1/lib/*:/home/cloudera/apache-storm-1.2.1/extlib/*:/home/cloudera/apache-storm-1.2.1/extlib-daemon/*:/home/cloudera/apache-storm
1.2.1:/home/cloudera/apache-storm-1.2.1/conf -Xmx128m -Dlogfile.name=logviewer.log -DLog4jContextSelector=org.apache.logging.log4j.core.async.AsyncLoggerCon
extSelector -Dlog4j.configurationFile=/home/cloudera/apache-storm-1.2.1/log4j2/cluster.xml org.apache.storm.daemon.logviewer
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm logviewer

(E)

Once above all processes are started, we can submit storm topology in production mode. In order to submit storm topology in production mode, follow below command:

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm jar /home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar com.upgrad.wordcount.WordCountTopology "WordCountTopology"

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm jar /home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar com.upgrad.wordcount.WordCountTopology "W
rdCountTopology"
Running: /usr/java/jdk1.8.0_161/bin/java -client -Ddaemon.name= -Dstorm.options= -Dstorm.home=/home/cloudera/apache-storm-1.2.1 -Dstorm.log.dir=/home/cloude
a/apache-storm-1.2.1/logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.conf.file= -cp /home/cloudera/apache-storm-1.2.1/*:/home/cloude
a/apache-storm-1.2.1/lib/*:/home/cloudera/apache-storm-1.2.1/extlib/*:/home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar:/home/cloudera/apache-storm
1.2.1/conf:/home/cloudera/apache-storm-1.2.1/bin -Dstorm.jar=/home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar -Dstorm.dependency.jars= -Dstorm.dep
ndency.artifacts={} com.upgrad.wordcount.WordCountTopology WordCountTopology
806  [main] WARN  o.a.s.u.Utils - STORM-VERSION new 1.2.1 old null
863  [main] INFO  o.a.s.StormSubmitter - Generated ZooKeeper secret payload for MD5-digest: -5452446045004683895:-7109163967801423948
944  [main] WARN  o.a.s.u.NimbusClient - Using deprecated config nimbus.host for backward compatibility. Please update your storm.yaml so it only has config
nimbus.seeds
981  [main] INFO  o.a.s.u.NimbusClient - Found leader nimbus : quickstart.cloudera:6627
998  [main] INFO  o.a.s.s.a.AuthUtils - Got AutoCreds []
998  [main] WARN  o.a.s.u.NimbusClient - Using deprecated config nimbus.host for backward compatibility. Please update your storm.yaml so it only has config
nimbus.seeds
1001 [main] INFO  o.a.s.u.NimbusClient - Found leader nimbus : quickstart.cloudera:6627
1045 [main] INFO  o.a.s.StormSubmitter - Uploading dependencies - jars...
1045 [main] INFO  o.a.s.StormSubmitter - Uploading dependencies - artifacts...
1052 [main] INFO  o.a.s.StormSubmitter - Dependency Blob keys - jars : [] / artifacts : []
1064 [main] INFO  o.a.s.StormSubmitter - Uploading topology jar /home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar to assigned location: /home/cloud
ra/apache-storm-1.2.1/Data/nimbus/inbox/stormjar-0406bd7a-34ac-4798-8088-0b5111416cca.jar
Start uploading file '/home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar' to '/home/cloudera/apache-storm-1.2.1/Data/nimbus/inbox/stormjar-0406bd7a-
4ac-4798-8088-0b5111416cca.jar' (1004815 bytes)
[==================================================] 1004815 / 1004815
File '/home/cloudera/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar' uploaded to '/home/cloudera/apache-storm-1.2.1/Data/nimbus/inbox/stormjar-0406bd7a-34ac-47
8-8088-0b5111416cca.jar' (1004815 bytes)
1093 [main] INFO  o.a.s.StormSubmitter - Successfully uploaded topology jar to assigned location: /home/cloudera/apache-storm-1.2.1/Data/nimbus/inbox/stormj
r-0406bd7a-34ac-4798-8088-0b5111416cca.jar
1093 [main] INFO  o.a.s.StormSubmitter - Submitting topology WordCountTopology in distributed mode with conf {"storm.zookeeper.topology.auth.scheme":"digest
,"storm.zookeeper.topology.auth.payload":"-5452446045004683895:-7109163967801423948","topology.workers":3,"topology.eventlogger.executors":3}
1094 [main] WARN  o.a.s.u.Utils - STORM-VERSION new 1.2.1 old 1.2.1
1160 [main] INFO  o.a.s.StormSubmitter - Finished submitting topology: WordCountTopology
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm jar /home/ec2-user/StormAssignmentWordCount-0.0.1-SNAPSHOT.jar com.upgrad.wordcount.WordCountTopology “WordCountTopology”

(F)

After running for some time, storm topology can be killed using below command:

[cloudera@quickstart apache-storm-1.2.1]$ bin/storm kill WordCountTopology

```
[cloudera@quickstart apache-storm-1.2.1]$ bin/storm kill WordCountTopology
Running: /usr/java/jdk1.8.0_161/bin/java -client -Ddaemon.name= -Dstorm.options= -Dstorm.home=/home/cloudera/apache-storm-1.2.1 -Dstorm.log.dir=/home/cloude
ra/apache-storm-1.2.1/logs -Djava.library.path=/usr/local/lib:/opt/local/lib:/usr/lib -Dstorm.conf.file= -cp /home/cloudera/apache-storm-1.2.1/*:/home/cloud
era/apache-storm-1.2.1/lib/*:/home/cloudera/apache-storm-1.2.1/extlib/*:/home/cloudera/apache-storm-1.2.1/extlib-daemon/*:/home/cloudera/apache-storm-1.2.1/
conf:/home/cloudera/apache-storm-1.2.1/bin org.apache.storm.command.kill_topology WordCountTopology
4827 [main] WARN  o.a.s.u.NimbusClient - Using deprecated config nimbus.host for backward compatibility. Please update your storm.yaml so it only has config
 nimbus.seeds
4915 [main] INFO  o.a.s.u.NimbusClient - Found leader nimbus : quickstart.cloudera:6627
4965 [main] INFO  o.a.s.c.kill-topology - Killed topology: WordCountTopology
```

Or

/home/ec2-user/apache-storm-1.2.1/bin/storm kill WordCountTopology

(G)

Check results in MySQL

In order to store and view results in MySQL, lets follow below steps:

(i) Log in to MySQL database, using credentials ‘root/cloudera’ if working on Cloudera VM or ‘root/123’ if working on EC2 instance.

(ii) Create new database namely upgrad.
create database upgrad;

(iii) Change current database to upgrad.
use upgrad;

(iv) Create a new table namely wordcounts.
create table wordcounts (word VARCHAR(250) PRIMARY KEY, count BIGINT);

(v) Your database is setup to store results.

Above steps need to be performed before storm topology is submitted either in local mode or in production mode. Above steps are pre-requisites for the storm topology execution.

Below are the screenshots taken for MySQL taken during execution of storm topology in production mode from time to time:

```
Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use upgrad;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> describe wordcounts;
+-------+--------------+------+-----+---------+-------+
| Field | Type         | Null | Key | Default | Extra |
+-------+--------------+------+-----+---------+-------+
| word  | varchar(250) | NO   | PRI | NULL    |       |
| count | bigint(20)   | YES  |     | NULL    |       |
+-------+--------------+------+-----+---------+-------+
2 rows in set (0.00 sec)

mysql> select * from wordcounts;
+-----------+--------+
| word      | count  |
+-----------+--------+
| a         | 150000 |
| ate       | 150000 |
| beverages | 150000 |
| cold      | 150000 |
| cow       | 150000 |
| dog       | 300001 |
| dont      | 300000 |
| fleas     | 300000 |
| has       | 150000 |
| have      | 150000 |
| homework  | 150000 |
| i         | 449999 |
| like      | 299999 |
| man       | 150000 |
| my        | 300001 |
| the       | 150000 |
| think     | 150000 |
+-----------+--------+
17 rows in set (0.00 sec)
```

```
mysql> select * from wordcounts;
+-----------+--------+
| word      | count  |
+-----------+--------+
| a         | 150000 |
| ate       | 150000 |
| beverages | 150000 |
| cold      | 150000 |
| cow       | 150000 |
| dog       | 300001 |
| dont      | 300000 |
| fleas     | 300000 |
| has       | 150000 |
| have      | 150000 |
| homework  | 150000 |
| i         | 449999 |
| like      | 299999 |
| man       | 150000 |
| my        | 300001 |
| the       | 150000 |
| think     | 150000 |
+-----------+--------+
17 rows in set (0.00 sec)

mysql> select * from wordcounts;
+-----------+--------+
| word      | count  |
+-----------+--------+
| a         | 158333 |
| ate       | 158333 |
| beverages | 158334 |
| cold      | 158334 |
| cow       | 158333 |
| dog       | 316667 |
| dont      | 316666 |
| fleas     | 316667 |
| has       | 158334 |
| have      | 158333 |
| homework  | 158333 |
| i         | 475000 |
| like      | 316667 |
| man       | 158333 |
| my        | 316667 |
| the       | 158333 |
| think     | 158333 |
+-----------+--------+
17 rows in set (0.00 sec)
```

After some more time:

```
mysql> select * from wordcounts;
+-----------+--------+
| word      | count  |
+-----------+--------+
| a         | 216666 |
| ate       | 216667 |
| beverages | 216667 |
| cold      | 216667 |
| cow       | 216667 |
| dog       | 433334 |
| dont      | 433333 |
| fleas     | 433333 |
| has       | 216667 |
| have      | 216666 |
| homework  | 216666 |
| i         | 650001 |
| like      | 433333 |
| man       | 216667 |
| my        | 433333 |
| the       | 216667 |
| think     | 216666 |
+-----------+--------+
17 rows in set (0.00 sec)

mysql> select * from wordcounts;
+-----------+--------+
| word      | count  |
+-----------+--------+
| a         | 225000 |
| ate       | 225000 |
| beverages | 225000 |
| cold      | 225000 |
| cow       | 225000 |
| dog       | 450000 |
| dont      | 450000 |
| fleas     | 450001 |
| has       | 225000 |
| have      | 225000 |
| homework  | 225000 |
| i         | 674999 |
| like      | 450000 |
| man       | 225000 |
| my        | 450000 |
| the       | 225000 |
| think     | 225000 |
+-----------+--------+
17 rows in set (0.00 sec)
```

Before killing storm topology, another screenshot for MySQL:

```
mysql> select * from wordcounts;
+-----------+--------+
| word      | count  |
+-----------+--------+
| a         | 225000 |
| ate       | 225000 |
| beverages | 225000 |
| cold      | 225000 |
| cow       | 225000 |
| dog       | 450000 |
| dont      | 450000 |
| fleas     | 450001 |
| has       | 225000 |
| have      | 225000 |
| homework  | 225000 |
| i         | 674999 |
| like      | 450000 |
| man       | 225000 |
| my        | 450000 |
| the       | 225000 |
| think     | 225000 |
+-----------+--------+
17 rows in set (0.00 sec)

mysql> select * from wordcounts;
+-----------+---------+
| word      | count   |
+-----------+---------+
| a         |  566666 |
| ate       |  566667 |
| beverages |  566667 |
| cold      |  566667 |
| cow       |  566666 |
| dog       | 1133334 |
| dont      | 1133333 |
| fleas     | 1133333 |
| has       |  566667 |
| have      |  566667 |
| homework  |  566667 |
| i         | 1699999 |
| like      | 1133333 |
| man       |  566667 |
| my        | 1133334 |
| the       |  566667 |
| think     |  566666 |
+-----------+---------+
17 rows in set (0.00 sec)
```

Storm UI – First Page - Cluster Summary

Storm UI

Cluster Summary

Version	Supervisors	Used slots	Free slots	Total slots	Executors	Tasks
1.2.1	1	3	1	4	15	15

Nimbus Summary

Search:

Host	Port	Status	Version	UpTime
localhost	6627	Offline	Not applicable	Not applicable
quickstart.cloudera	6627	Leader	1.2.1	47m 49s

Showing 1 to 2 of 2 entries

Topology Summary

Search:

Name	Owner	Status	Uptime	Num workers	Num executors	Num tasks	Replication count	Assigned Mem (MB)	Scheduler Info
WordCountTopology	cloudera	ACTIVE	30m 58s	3	15	15	1	2496	

Showing 1 to 1 of 1 entries

Supervisor Summary

Search:

Host	Id	Uptime	Slots	Used slots	Avail slots	Used Mem (MB)	Version
quickstart.cloudera (log)	9f39df22-0359-4fcf-816e-bd6221bc65c7	47m 36s	4	3	1	2496	1.2.1

Showing 1 to 1 of 1 entries

Nimbus Configuration

Show 20 entries

Search:

Key	Value
backpressure.disruptor.high.watermark	0.9
backpressure.disruptor.low.watermark	0.4
backpressure.znode.timeout.secs	30
backpressure.znode.update.freq.secs	15
client.blobstore.class	"org.apache.storm.blobstore.NimbusBlobStore"
dev.zookeeper.path	"/tmp/dev-storm-zookeeper"
drpc.authorizer.acl.filename	"drpc-auth-acl.yaml"
drpc.authorizer.acl.strict	false
drpc.childopts	"-Xmx768m"
drpc.http.creds.plugin	"org.apache.storm.security.auth.DefaultHttpCredentialsPlugin"
drpc.http.port	3774
drpc.https.keystore.password	""

drpc.https.keystore.type	"JKS"
drpc.https.port	-1
drpc.invocations.port	3773
drpc.invocations.threads	64
drpc.max_buffer_size	1048576
drpc.port	3772
drpc.queue.size	128
drpc.request.timeout.secs	600

Showing 1 to 20 of 207 entries

Previous 1 2 3 4 5 … 11 Next

Storm Topology Summary

Storm UI

Search WordCountTopology-2-1547204841: Search Search Archived Logs:

Topology summary

Name	Id	Owner	Status	Uptime	Num workers	Num executors	Num tasks	Replication count	Assigned Mem (MB)	Scheduler Info
WordCountTopology	WordCountTopology-2-1547204841	cloudera	ACTIVE	38m 11s	3	15	15	1	2496	

Topology actions

Activate Deactivate Rebalance Kill Debug Stop Debug Change Log Level

Topology stats

Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
10m 0s	4567274	4567568	9.908	431319	
3h 0m 0s	14696220	14696220	12.315	1386100	
1d 0h 0m 0s	14696220	14696220	12.315	1386100	
All time	14696220	14696220	12.315	1386100	

Spouts (All time)

Search:

Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error	Error Time
SentenceSpout	1	1	1390020	1390020	12.315	1386100	0				

Showing 1 to 1 of 1 entries

Bolts (All time)

Search:

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error	Error Time
ReportBolt	1	1	0	0	0.013	0.004	6653240	0.003	6653240	0				
SplitSentenceBolt	3	3	6654340	6654340	0.006	0.023	1386360	0.024	1386360	0				
WordCountBolt	4	4	6651860	6651860	0.009	0.009	6655140	0.009	6655120	0				

Showing 1 to 3 of 3 entries

Worker Resources

Search: Toggle Components

Host	Supervisor Id	Port	Uptime	Num executors	Assigned Mem (MB)	Components
quickstart.cloudera	9f39df22-0359-4fcf-816e-bd6221bc65c7	6701	37m 47s	5	832	3 components
quickstart.cloudera	9f39df22-0359-4fcf-816e-bd6221bc65c7	6702	37m 47s	5	832	2 components
quickstart.cloudera	9f39df22-0359-4fcf-816e-bd6221bc65c7	6700	37m 47s	5	832	3 components

Showing 1 to 3 of 3 entries

Topology Visualization

Show Visualization

Topology Configuration

Show 20 entries

Search:

Key	Value
backpressure.disruptor.high.watermark	0.9
backpressure.disruptor.low.watermark	0.4
backpressure.znode.timeout.secs	30
backpressure.znode.update.freq.secs	15
client.blobstore.class	"org.apache.storm.blobstore.NimbusBlobStore"
dev.zookeeper.path	"/tmp/dev-storm-zookeeper"

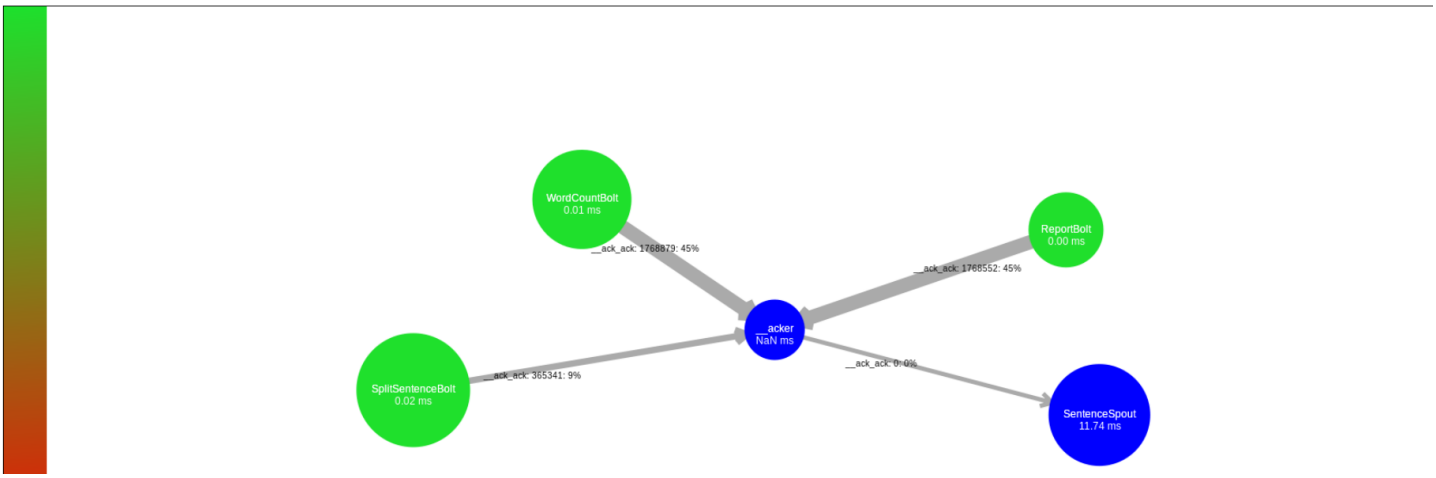
drpc.authorizer.acl.filename	"drpc-auth-acl.yaml"
drpc.authorizer.acl.strict	false
drpc.childopts	"-Xmx768m"
drpc.http.creds.plugin	"org.apache.storm.security.auth.DefaultHttpCredentialsPlugin"
drpc.http.port	3774
drpc.https.keystore.password	""
drpc.https.keystore.type	"JKS"
drpc.https.port	-1
drpc.invocations.port	3773
drpc.invocations.threads	64
drpc.max_buffer_size	1048576
drpc.port	3772
drpc.queue.size	128
drpc.request.timeout.secs	600

Showing 1 to 20 of 215 entries

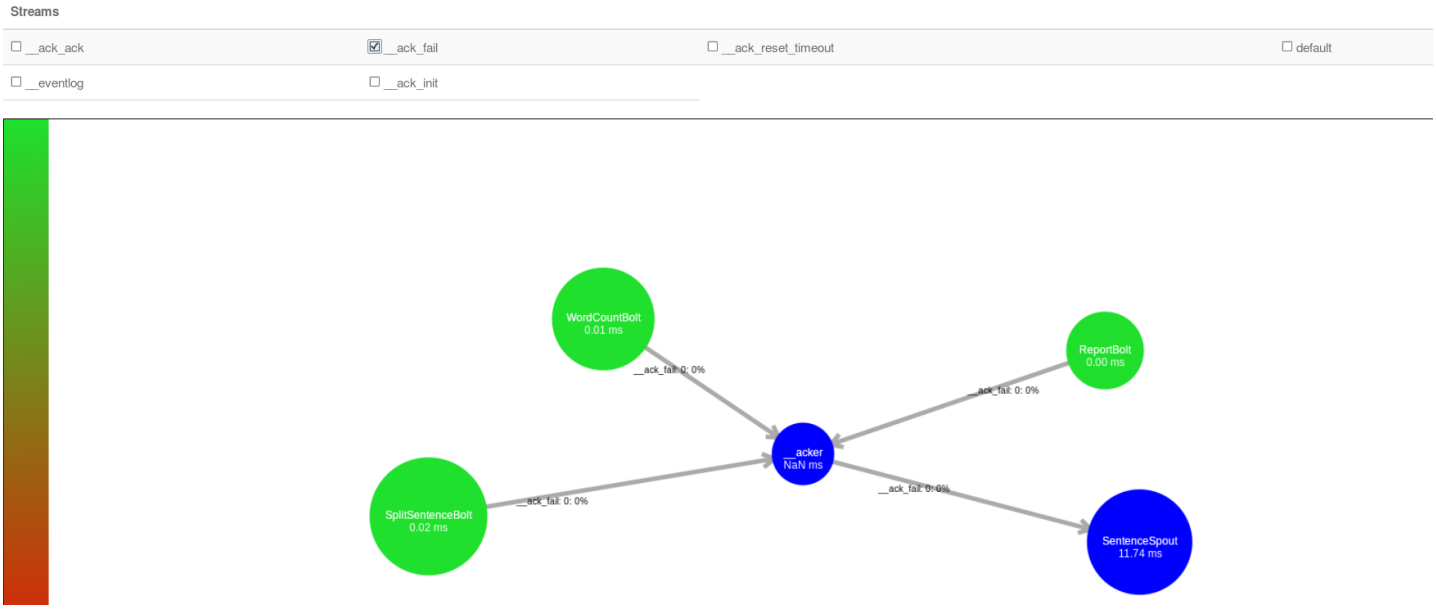
Show System Stats

Storm Topology Visualization

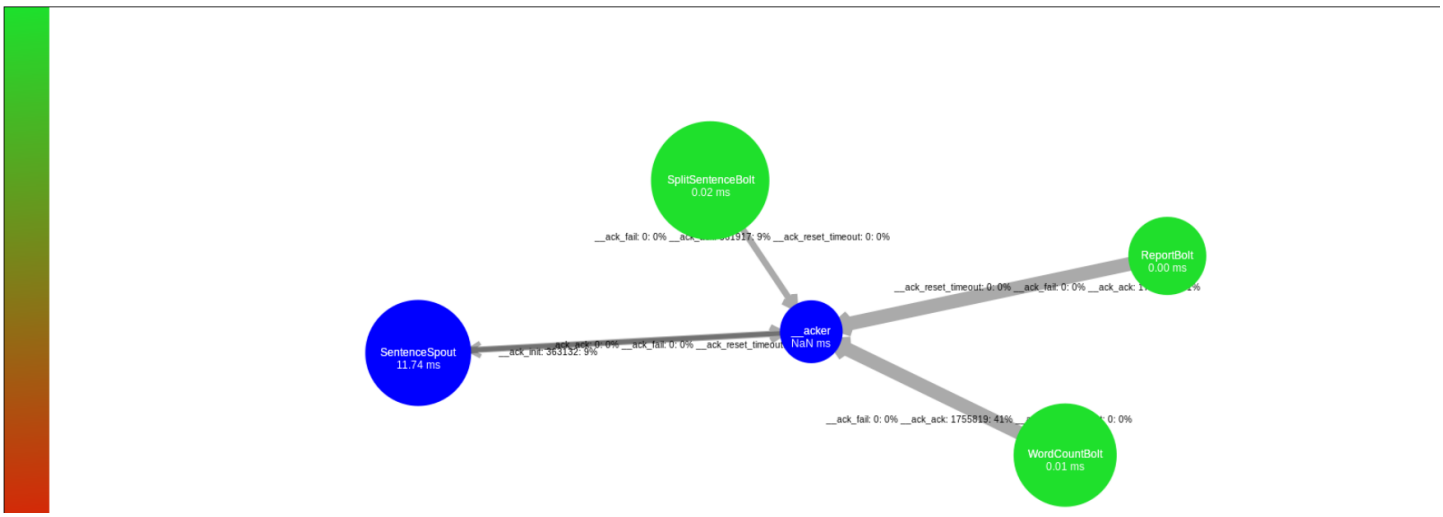
⇨ **default**



⇨ **__ack_ack**



⇨ **__ack_fail**



⇨ **__ack_ack + __ack_fail + __ack_init + __act_reset_timeout**

SentenceSpout

Storm UI

Component summary

Id	Topology	Executors	Tasks	Debug
SentenceSpout	WordCountTopology	1	1	events

Component actions

Debug | Stop Debug

Spout stats

Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
10m 0s	390121	390121	11.268	393539	0
3h 0m 0s	2239620	2239620	11.892	2237760	0
1d 0h 0m 0s	2239620	2239620	11.892	2237760	0
All time	2239620	2239620	11.892	2237760	0

Output stats (All time)

Search:

Stream	Emitted	Transferred	Complete latency (ms)	Acked	Failed
default	2239620	2239620	11.892	2237760	0

Showing 1 to 1 of 1 entries

Profiling and Debugging

Use the following controls to profile and debug the components on this page.

Status / Timeout (Minutes)	Actions
10	JStack Restart Worker Heap

Executors (All time)

Search:

Id	Uptime	Host	Port	Actions	Emitted	Transferred	Complete latency (ms)	Acked	Failed
[2-2]	59m 9s	quickstart.cloudera	6701	files	2239620	2239620	11.892	2237760	0

Showing 1 to 1 of 1 entries

Errors

Search:

Time	Error Host	Error Port	Error
No data available in table			

Showing 0 to 0 of 0 entries

Show System Stats

SplitSentenceBolt

Storm UI

Component summary

Id	Topology	Executors	Tasks	Debug
SplitSentenceBolt	WordCountTopology	3	3	events

Component actions

Debug | Stop Debug

Bolt stats

Search:

Window	Emitted	Transferred	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
10m 0s	1876483	1875848	0.022	391223	0.022	391026	0
3h 0m 0s	12116020	12116020	0.023	2524160	0.023	2524140	0
1d 0h 0m 0s	12116020	12116020	0.023	2524160	0.023	2524140	0
All time	12116020	12116020	0.023	2524160	0.023	2524140	0

Showing 1 to 4 of 4 entries

Input stats (All time)

Search:

Component	Stream	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
SentenceSpout	default	0.023	2524160	0.023	2524140	0

Showing 1 to 1 of 1 entries

Output stats (All time)

Search:

Stream	Emitted	Transferred
default	12116020	12116020

Showing 1 to 1 of 1 entries

Profiling and Debugging

Use the following controls to profile and debug the components on this page.

Status / Timeout (Minutes)	Actions
10	JStack Restart Worker Heap

Executors (All time)

Search:

Id	Uptime	Host	Port	Debug	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
[3-3]	1h 6m 31s	quickstart.cloudera	6702	☐ files	4041960	4041960	0.005	0.025	840820	0.025	840820	0
[4-4]	1h 6m 31s	quickstart.cloudera	6700	☐ files	4032880	4032880	0.004	0.022	840640	0.023	840640	0
[5-5]	1h 6m 31s	quickstart.cloudera	6701	☐ files	4041180	4041180	0.005	0.023	842700	0.022	842680	0

Showing 1 to 3 of 3 entries

Errors

Search:

Time	Error Host	Error Port	Error
No data available in table			

Showing 0 to 0 of 0 entries

Show System Stats

WordCountBolt

Storm UI

Component summary

Id	Topology	Executors	Tasks	Debug
WordCountBolt	WordCountTopology	4	4	events

Component actions

Debug | Stop Debug

Bolt stats

Search:

Window	Emitted	Transferred	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
10m 0s	1868719	1868719	0.008	1866520	0.008	1866818	0
3h 0m 0s	12680280	12680280	0.009	12687220	0.008	12687220	0
1d 0h 0m 0s	12680280	12680280	0.009	12687220	0.008	12687220	0
All time	12680280	12680280	0.009	12687220	0.008	12687220	0

Showing 1 to 4 of 4 entries

Input stats (All time)

Search:

Component	Stream	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
SplitSentenceBolt	default	0.009	12687220	0.008	12687220	0

Showing 1 to 1 of 1 entries

Output stats (All time)

Search:

Stream	Emitted	Transferred
default	12680280	12680280

Showing 1 to 1 of 1 entries

Profiling and Debugging

Use the following controls to profile and debug the components on this page.

Status / Timeout (Minutes)	Actions
10	JStack Restart Worker Heap

Executors (All time)

Search:

Id	Uptime	Host	Port	Debug	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
[6-6]	1h 9m 34s	quickstart.cloudera	6702	☐ files	3172620	3172620	0.006	0.009	3173220	0.008	3173220	0
[7-7]	1h 9m 34s	quickstart.cloudera	6700	☐ files	3162100	3162100	0.007	0.009	3167660	0.009	3167660	0
[8-8]	1h 9m 34s	quickstart.cloudera	6701	☐ files	3183560	3183560	0.007	0.010	3174980	0.010	3174980	0
[9-9]	1h 9m 34s	quickstart.cloudera	6702	☐ files	3162000	3162000	0.004	0.007	3171360	0.006	3171360	0

Showing 1 to 4 of 4 entries

Errors

Search:

Time	Error Host	Error Port	Error
No data available in table			

Showing 0 to 0 of 0 entries

Show System Stats

ReportBolt

Storm UI

Component summary

Id	Topology	Executors	Tasks	Debug
ReportBolt	WordCountTopology	1	1	events

Component actions

Debug | Stop Debug

Bolt stats

Search:

Window	Emitted	Transferred	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
10m 0s	0	0	0.004	1867562	0.003	1867555	0
3h 0m 0s	0	0	0.004	13062120	0.003	13062120	0
1d 0h 0m 0s	0	0	0.004	13062120	0.003	13062120	0
All time	0	0	0.004	13062120	0.003	13062120	0

Showing 1 to 4 of 4 entries

Input stats (All time)

Search:

Component	Stream	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
WordCountBolt	default	0.004	13062120	0.003	13062120	0

Showing 1 to 1 of 1 entries

Output stats (All time)

Search:

Stream	Emitted	Transferred
No data available in table		

Showing 0 to 0 of 0 entries

Profiling and Debugging

Use the following controls to profile and debug the components on this page.

Status / Timeout (Minutes)	Actions
10	JStack Restart Worker Heap

Executors (All time)

Search:

Id	Uptime	Host	Port	Debug	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
[1-1]	1h 11m 34s	quickstart.cloudera	6700	☐ files	0	0	0.011	0.004	13062120	0.003	13062120	0

Showing 1 to 1 of 1 entries

Errors

Search:

Time	Error Host	Error Port	Error
No data available in table			

Showing 0 to 0 of 0 entries

Show System Stats

END OF DOCUMENT