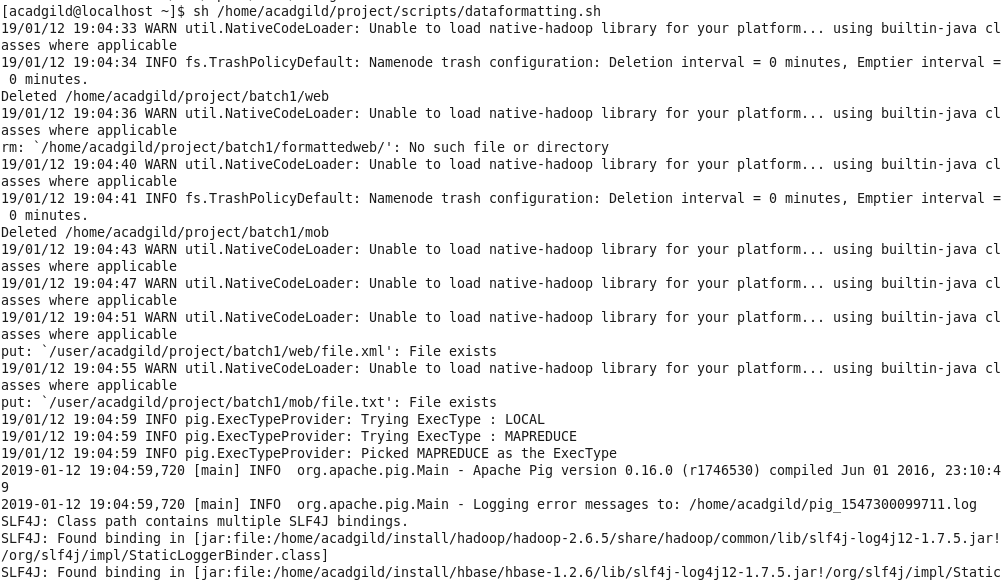
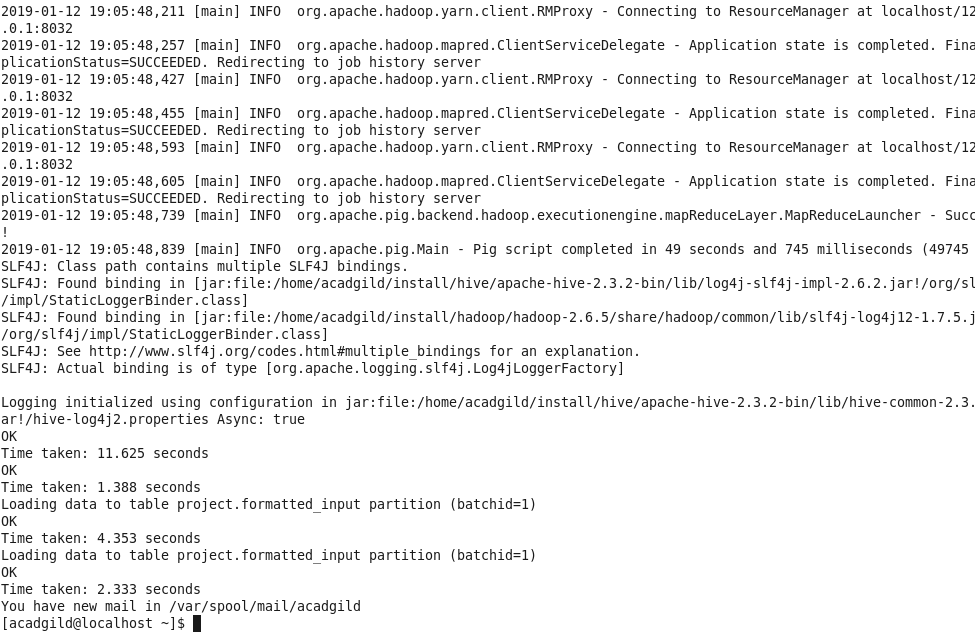
**Formatting the data:**

Note : The scripts are placed in separate folder named as Scripts – Kinldy refer to it.





[acadgild@localhost ~]$ sh /home/acadgild/project/scripts/dataformatting.sh  
19/01/12 19:04:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
19/01/12 19:04:34 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.  
Deleted /home/acadgild/project/batch1/web  
19/01/12 19:04:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
rm: `/home/acadgild/project/batch1/formattedweb/': No such file or directory  
19/01/12 19:04:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
19/01/12 19:04:41 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.  
Deleted /home/acadgild/project/batch1/mob  
19/01/12 19:04:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
19/01/12 19:04:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
19/01/12 19:04:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
put: `/user/acadgild/project/batch1/web/file.xml': File exists  
19/01/12 19:04:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
put: `/user/acadgild/project/batch1/mob/file.txt': File exists  
19/01/12 19:04:59 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL  
19/01/12 19:04:59 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE  
19/01/12 19:04:59 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType  
2019-01-12 19:04:59,720 [main] INFO  org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49  
2019-01-12 19:04:59,720 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/acadgild/pig\_1547300099711.log  
SLF4J: Class path contains multiple SLF4J bindings.  
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: 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: See [http://www.slf4j.org/codes.html#multiple\_bindings](http://www.slf4j.org/codes.html" \l "multiple_bindings" \t "_blank) for an explanation.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
2019-01-12 19:05:00,607 [main] WARN  org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
2019-01-12 19:05:01,137 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/acadgild/.pigbootup not found  
2019-01-12 19:05:01,519 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2019-01-12 19:05:01,519 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:01,519 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020  
2019-01-12 19:05:02,686 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session: PIG-dataformatting.pig-a7fab280-dc28-4bfd-9aad-cd571eba1f59  
2019-01-12 19:05:02,686 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false  
2019-01-12 19:05:02,892 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:03,997 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:04,438 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:04,620 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator  
2019-01-12 19:05:04,794 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN  
2019-01-12 19:05:04,915 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:04,937 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2019-01-12 19:05:05,102 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}  
2019-01-12 19:05:05,273 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752  
2019-01-12 19:05:05,435 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false  
2019-01-12 19:05:05,539 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1  
2019-01-12 19:05:05,539 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1  
2019-01-12 19:05:05,700 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:06,006 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:06,801 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job  
2019-01-12 19:05:06,825 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent  
2019-01-12 19:05:06,826 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2019-01-12 19:05:06,826 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress  
2019-01-12 19:05:06,844 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process  
2019-01-12 19:05:06,975 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication  
2019-01-12 19:05:07,598 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/project/lib/piggybank.jar to DistributedCache through /tmp/temp-2082373811/tmp1732776960/piggybank.jar  
2019-01-12 19:05:07,774 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp-2082373811/tmp-616480840/pig-0.16.0-core-h2.jar  
2019-01-12 19:05:07,874 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-2082373811/tmp-1295264571/automaton-1.11-8.jar  
2019-01-12 19:05:07,922 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-2082373811/tmp-732119197/antlr-runtime-3.4.jar  
2019-01-12 19:05:08,061 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp-2082373811/tmp1027425772/joda-time-2.9.3.jar  
2019-01-12 19:05:08,113 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job  
2019-01-12 19:05:08,164 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.  
2019-01-12 19:05:08,168 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche  
2019-01-12 19:05:08,169 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []  
2019-01-12 19:05:08,414 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.  
2019-01-12 19:05:08,419 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address  
2019-01-12 19:05:08,439 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:08,525 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - [mapred.task.id](http://mapred.task.id/" \t "_blank) is deprecated. Instead, use [mapreduce.task.attempt.id](http://mapreduce.task.attempt.id/" \t "_blank)  
2019-01-12 19:05:08,531 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - [fs.default.name](http://fs.default.name/" \t "_blank) is deprecated. Instead, use fs.defaultFS  
2019-01-12 19:05:09,427 [JobControl] WARN  org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set.  User classes may not be found. See Job or Job#setJar(String).  
2019-01-12 19:05:09,644 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2  
2019-01-12 19:05:09,645 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 2  
2019-01-12 19:05:09,712 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2019-01-12 19:05:09,822 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2019-01-12 19:05:10,202 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1547297823566\_0006  
2019-01-12 19:05:10,672 [JobControl] INFO  org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2019-01-12 19:05:10,969 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1547297823566\_0006  
2019-01-12 19:05:11,456 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: [http://localhost:8088/proxy/application\_1547297823566\_0006/](http://localhost:8088/proxy/application_1547297823566_0006/" \t "_blank)  
2019-01-12 19:05:11,456 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1547297823566\_0006  
2019-01-12 19:05:11,456 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B  
2019-01-12 19:05:11,456 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[5,5],A[-1,-1],B[7,4] C:  R:   
2019-01-12 19:05:11,505 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2019-01-12 19:05:11,505 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1547297823566\_0006]  
2019-01-12 19:05:43,936 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete  
2019-01-12 19:05:43,937 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1547297823566\_0006]  
2019-01-12 19:05:46,959 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:47,125 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2019-01-12 19:05:47,801 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:47,818 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2019-01-12 19:05:47,969 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces  
2019-01-12 19:05:47,970 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:47,989 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2019-01-12 19:05:48,201 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete  
2019-01-12 19:05:48,209 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:   
  
HadoopVersion    PigVersion    UserId    StartedAt    FinishedAt    Features  
2.6.5    0.16.0    acadgild    2019-01-12 19:05:06    2019-01-12 19:05:48    UNKNOWN  
  
Success!  
  
Job Stats (time in seconds):  
JobId    Maps    Reduces    MaxMapTime    MinMapTime    AvgMapTime    MedianMapTime    MaxReduceTime    MinReduceTime    AvgReduceTime    MedianReducetime    Alias    Feature    Outputs  
job\_1547297823566\_0006    1    0    11    11    11    11    0    0    0    0    A,B    MAP\_ONLY    /home/acadgild/project/batch1/formattedweb,  
  
Input(s):  
Successfully read 40 records (13915 bytes) from: "/user/acadgild/project/batch1/web"  
  
Output(s):  
Successfully stored 40 records (2480 bytes) in: "/home/acadgild/project/batch1/formattedweb"  
  
Counters:  
Total records written : 40  
Total bytes written : 2480  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job\_1547297823566\_0006  
  
  
2019-01-12 19:05:48,211 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:48,257 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2019-01-12 19:05:48,427 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:48,455 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2019-01-12 19:05:48,593 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/[127.0.0.1:8032](http://127.0.0.1:8032/" \t "_blank)  
2019-01-12 19:05:48,605 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2019-01-12 19:05:48,739 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2019-01-12 19:05:48,839 [main] INFO  org.apache.pig.Main - Pig script completed in 49 seconds and 745 milliseconds (49745 ms)  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.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](http://www.slf4j.org/codes.html" \l "multiple_bindings" \t "_blank) for an explanation.  
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]  
  
Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true  
OK  
Time taken: 11.625 seconds  
OK  
Time taken: 1.388 seconds  
Loading data to table project.formatted\_input partition (batchid=1)  
OK  
Time taken: 4.353 seconds  
Loading data to table project.formatted\_input partition (batchid=1)  
OK  
Time taken: 2.333 seconds  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ 