Fall 2015

Telecom Data Analysis

Sai Roopak Sarva
Governors State University

Anudeep Masetty
Governors State University

Vinay Reddy Kondam
Governors State University

Table of contents

1 Project Description

1.1 Project Abstract

The telecommunications industry regularly uses data analytics in fields such as customer analysis and network optimization. For financial analysis such as identifying risks, which could negatively impact an entity's financial performance, communications service providers have traditionally used statistical sampling techniques that cover only short time periods and a limited subset of data.

Given the massive number of transactions processed by telecommunications companies; and the costs and complexity involved in their operations, data analytics offers a valuable opportunity for enhancing the frameworks and procedures they adopt to drive profitability and minimize unnecessary downside risk.

1.2 Competitive Information

All other Service providers are having terabytes of data and scattered across the organization. In order to exploit the full potential of this stored data, service providers must have solutions that can help them correlate, process and decipher the actionable information. This is not possible without big data and advanced analytics.

1.3 Relationship to Other Applications/Projects

This project is relevant to the existing techniques of analyzing data using different databases like MYSQL but not same in the platform and architecture used in this project. MySQL is playing a key role in many big data platforms. Based on estimates from a leading Hadoop vendor, MySQL is a core component of the big data pipeline in over 80% of deployments

1.4 Assumptions and Dependencies

Assumptions are made based on rapid data growth in the future of a telecom Industry. The telecom sector's use of data analytics tools is expected to grow at a compound annual growth rate of 28.28 percent over the next four years.

1.5 Future Enhancements

The one possible evolution which made is moving this platform to cloud based engines might help burden of managing cluster environment. The mobile market is poised for a huge explosion of new business. More than half of the world's population of 7 billion is using mobile connections at present, and estimates suggest another billion will join in over the next two to three years. With many subscribers using more than one device, the total number of mobile connections, also around 7 billion at present, is increasing sharply.

1.6 Definitions and Acronyms

HDFS – Hadoop Distributed File System
MR – MapReduce

2 *Technical Description*

This application scope is to use Hadoop infrastructure to analyze the data which is coming to Hadoop cluster and stored in HDFS as files and to find out the solution for business problems

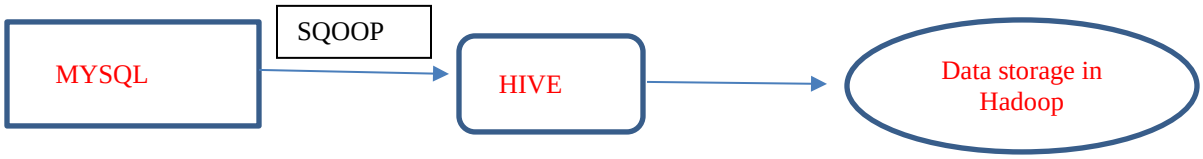
Telecom Data Analysis will help telecom companies to do market analysis and predictive analysis based on existing users and their orders information. The current systems cannot handle the large volume in a reasonable time. The new system will use Hadoop framework to handle processing of semi-structured and large volumes of data by parallelism techniques

These are the components used in this project:-

Initial Data Storage : MySQL
Data Ingestion : Sqoop (SQL+Hadoop),MapReduce
Data Storage from Hadoop : HDFS and HIVE tables
Data Analysis : HIVE and MapReduce

2.1 *Project/Application Architecture*

Architecture : Starts by importing data from RDBMS systems to Hadoop – Creating tables on top of imported data – Apply analytical queries to get business insights



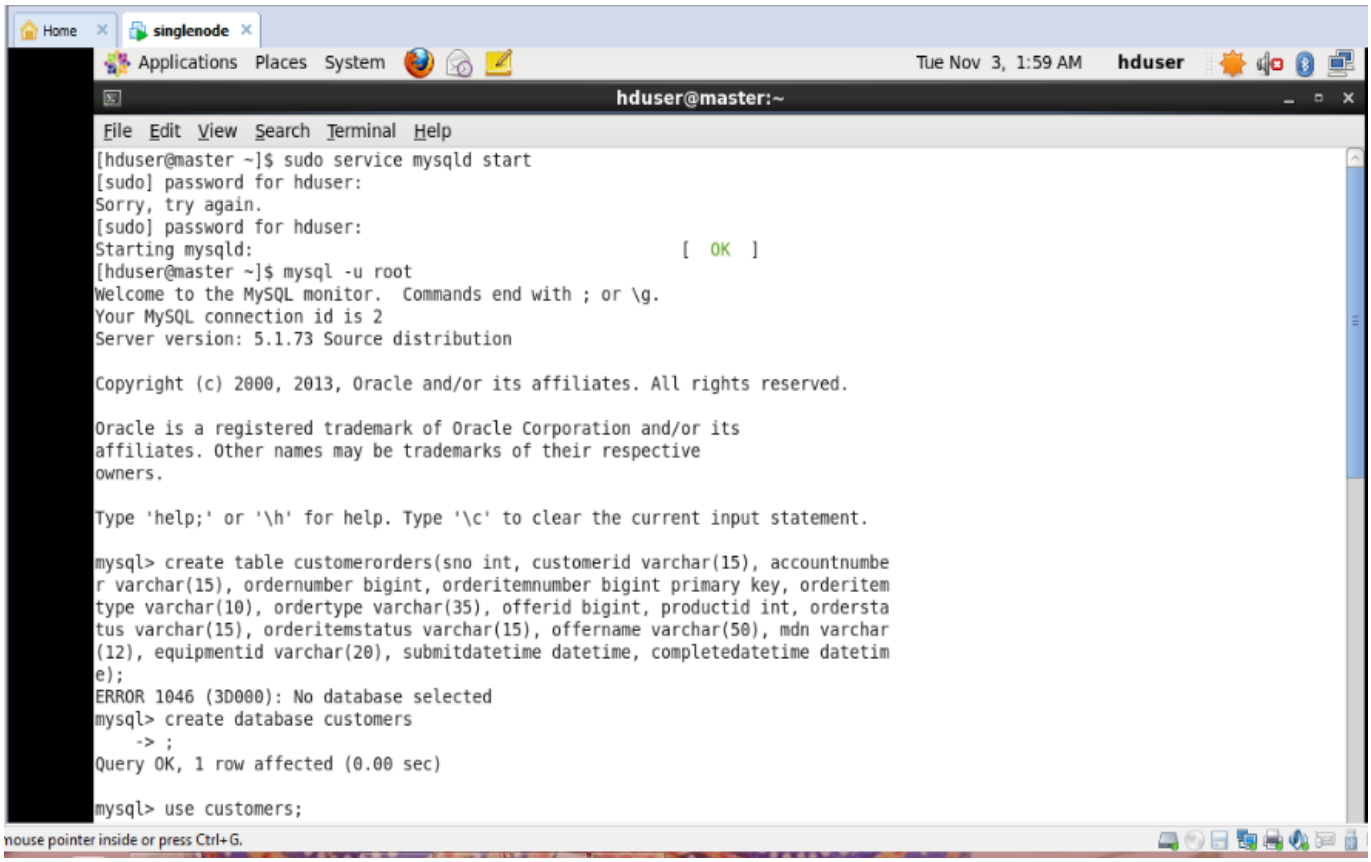
2.2 *Project/Application Information flows*

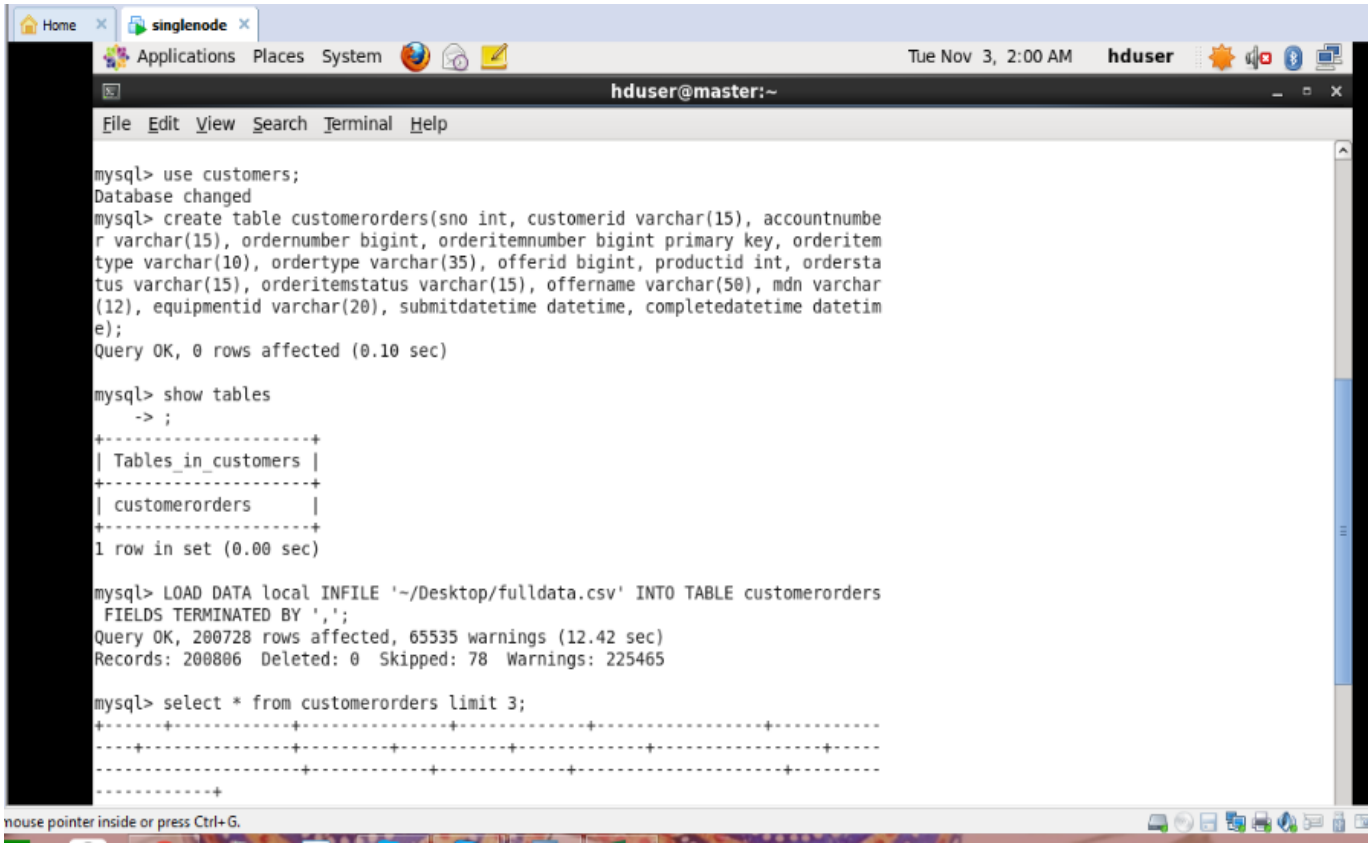
Step: I Initial Data Storage

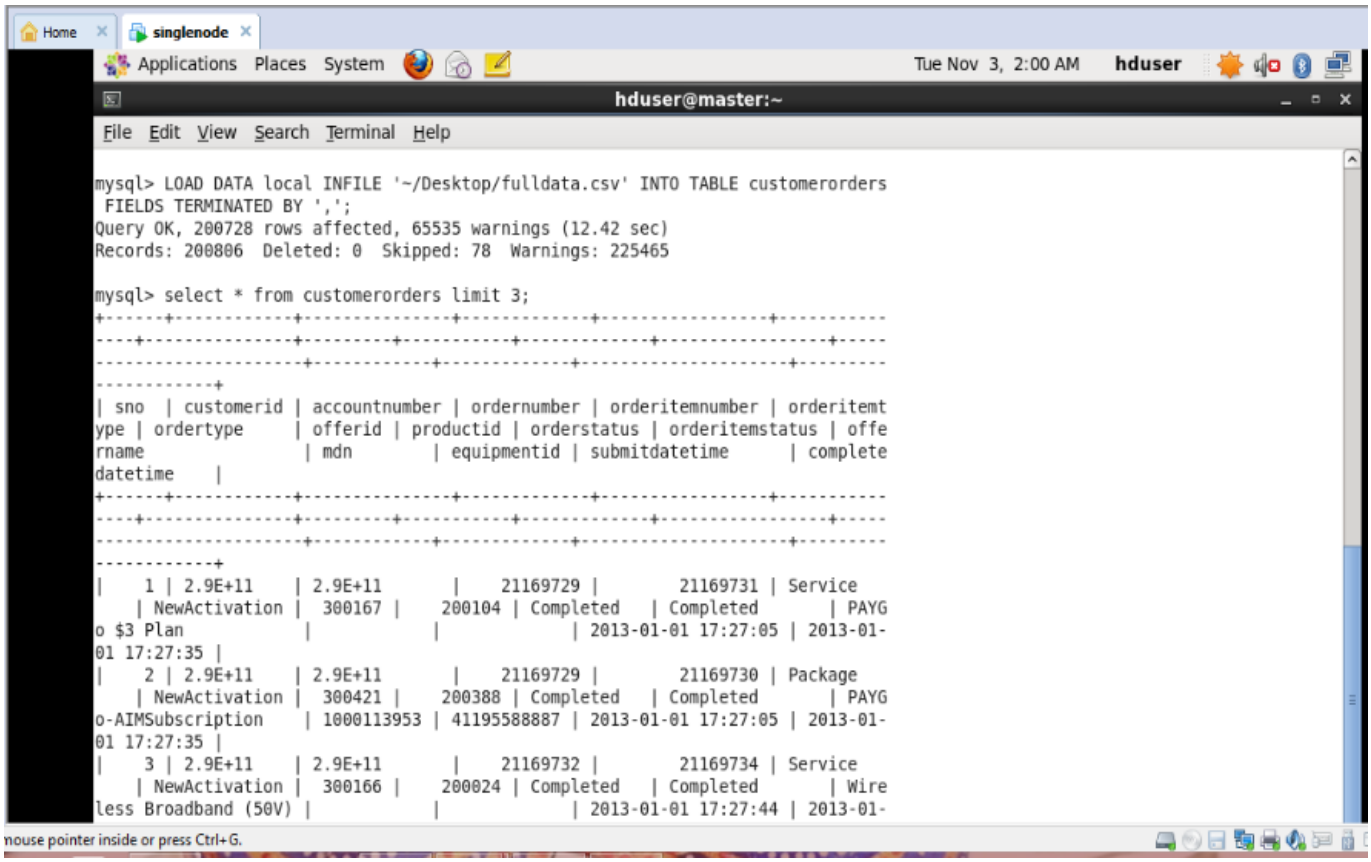
- Generally the telecom company consists of the customers' orders data in the form of a excel file where it has 200078 records in it.
- In this step we need to store the data in MySQL database by creating a database called customers and a table called customer orders
- So that data from excel file can be loaded into MySQL through logging into MySQL

A1 | 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	1	2.90E+11	2.90E+11	21169729	21169731	Service	NewActiv	300167	200104	Complete	Complete	PAYGo $3 Plan			########	########	
2	2	2.90E+11	2.90E+11	21169729	21169730	Package	NewActiv	300421	200388	Complete	Complete	PAYGo-All	1E+09	4.12E+10	########	########	
3	3	2.90E+11	2.90E+11	21169732	21169734	Service	NewActiv	300166	200024	Complete	Complete	Wireless Broadband (50V)			########	########	
4	4	2.90E+11	2.90E+11	21169732	21169733	Package	NewActiv	300401	200288	Complete	Complete	BBMM	1E+09	1E+11	########	########	
5	5	2.90E+11	2.90E+11	21169735	21169736	Package	NewActiv	300421	200388	Complete	Complete	PAYGo-All	1E+09	7.53E+10	########	########	
6	6	2.90E+11	2.90E+11	21169735	21169737	Service	NewActiv	300167	200280	Complete	Complete	All-In-Monthly PAYGo (PM2)			########	########	
7	7	2.90E+11	2.90E+11	21169738	21169739	Package	NewActiv	300401	200288	Complete	Complete	BBMM	1E+09	9.86E+10	########	########	
8	8	2.90E+11	2.90E+11	21169738	21169740	Service	NewActiv	300166	200024	Complete	Complete	Wireless Broadband (50V)			########	########	
9	9	2.90E+11	2.90E+11	21169741	21169743	Service	NewActiv	300166	200024	Complete	Complete	Wireless Broadband (50V)			########	########	
10	10	2.90E+11	2.90E+11	21169741	21169742	Package	NewActiv	300401	200288	Complete	Complete	BBMM	1E+09	1E+10	########	########	
11	11	2.90E+11	2.90E+11	21169744	21169746	Service	NewActiv	300166	201236	Complete	Complete	BBMM 20 (BBMM20A)			########	########	
12	12	2.90E+11	2.90E+11	21169744	21169745	Package	NewActiv	300401	200288	Complete	Complete	BBMM	1E+09	1E+10	########	########	
13	13	2.90E+11	2.90E+11	21169747	21169748	Package	Registrati	300401	200288	Complete	Complete	BBMM	1E+09	1E+10	########	########	
14	14	2.90E+11	2.90E+11	21169747	21169749	Service		300166	201235	Complete	Complete	BBMM 45 (BBMM45A)			########	########	
15	15	2.90E+11	2.90E+11	21169750	21169755	Service	NewActiv	300168	201357	Complete	Complete	Wireless $50 Smart (BBY5)			########	########	
16	16	2.90E+11	2.90E+11	21169750	21169753	Service		300173	200100	Complete	Complete	Multi-Product Plan			########	########	
17	17	2.90E+11	2.90E+11	21169750	21169751	Package	NewActiv	300165	200011	Complete	Complete	Broadban	1E+09	9E+10	########	########	
18	18	2.90E+11	2.90E+11	21169750	21169754	Service	NewActiv	300168	201129	Complete	Complete	ADRSmartphone All_In (ADR1)			########	########	
19	19	2.90E+11	2.90E+11	21169750	21169756	Service	NewActiv	300173	200515	Complete	Complete	Multiline-Broadband Bundle C			########	########	
20	20	2.90E+11	2.90E+11	21169750	21169752	Package	NewActiv	300163	200013	Complete	Complete	Voice Sub	1E+09	9E+10	########	########	
21	21	2.90E+11	2.90E+11	21169757	21169758	Package	NewActiv	300401	200288	Complete	Complete	BBMM	1E+09	1E+10	########	########	
22	22	2.90E+11	2.90E+11	21169757	21169759	Service	NewActiv	300166	201235	Complete	Complete	BBMM 45 (BBMM45A)			########	########	
23	23	2.90E+11	2.90E+11	21169760	21169762	Service		300166	201236	Complete	Complete	BBMM 20 (BBMM20A)			########	########	




Home
singlenode
Applications Places System
Tue Nov 3, 2:00 AM hduser
hduser@master:~
File Edit View Search Terminal Help
mysql> use customers;
Database changed
mysql> create table customerorders(sno int, customerid varchar(15), accountnumbe
r varchar(15), ordernumber bigint, orderitemnumber bigint primary key, orderitem
type varchar(10), ordertype varchar(35), offerid bigint, productid int, ordersta
tus varchar(15), orderitemstatus varchar(15), offername varchar(50), mdn varchar
(12), equipmentid varchar(20), submitdatetime datetime, completedatetime datetim
e);
Query OK, 0 rows affected (0.10 sec)
mysql> show tables
-> ;
| Tables_in_customers |
| customerorders |
1 row in set (0.00 sec)
mysql> LOAD DATA local INFILE '~/Desktop/fulldata.csv' INTO TABLE customerorders
FIELDS TERMINATED BY ',';
Query OK, 200728 rows affected, 65535 warnings (12.42 sec)
Records: 200806 Deleted: 0 Skipped: 78 Warnings: 225465
mysql> select * from customerorders limit 3;
nouse pointer inside or press Ctrl+G.


Home
singlenode
Applications Places System
Tue Nov 3, 2:00 AM hduser
hduser@master:~
File Edit View Search Terminal Help
mysql> LOAD DATA local INFILE '~/Desktop/fulldata.csv' INTO TABLE customerorders
FIELDS TERMINATED BY ',';
Query OK, 200728 rows affected, 65535 warnings (12.42 sec)
Records: 200806 Deleted: 0 Skipped: 78 Warnings: 225465
mysql> select * from customerorders limit 3;
| sno | customerid | accountnumber | ordernumber | orderitemnumber | orderitemt
ype | ordertype | offerid | productid | orderstatus | orderitemstatus | offe
rname | mdn | equipmentid | submitdatetime | complete
datetime |
| 1 | 2.9E+11 | 2.9E+11 | 21169729 | 21169731 | Service
| NewActivation | 300167 | 200104 | Completed | Completed | PAYG
o $3 Plan | | | 2013-01-01 17:27:05 | 2013-01-
01 17:27:35 |
| 2 | 2.9E+11 | 2.9E+11 | 21169729 | 21169730 | Package
| NewActivation | 300421 | 200388 | Completed | Completed | PAYG
o-AIMSubscription | 1000113953 | 41195588887 | 2013-01-01 17:27:05 | 2013-01-
01 17:27:35 |
| 3 | 2.9E+11 | 2.9E+11 | 21169732 | 21169734 | Service
| NewActivation | 300166 | 200024 | Completed | Completed | Wire
less Broadband (50V) | | | 2013-01-01 17:27:44 | 2013-01-
nouse pointer inside or press Ctrl+G.

Step : II

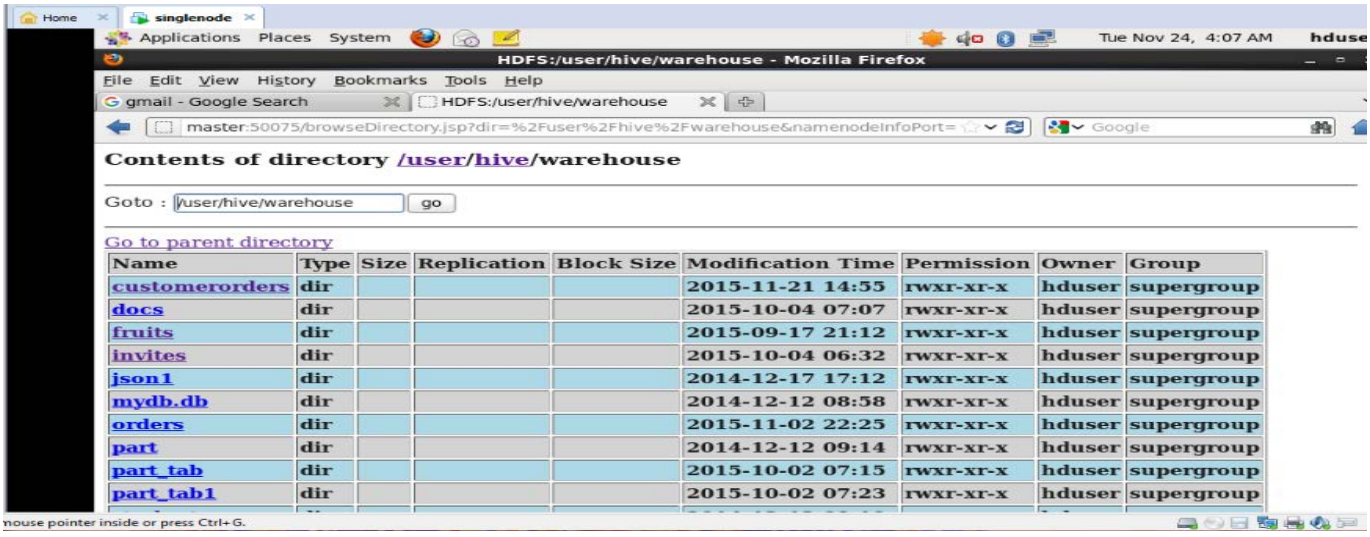
Data Ingestion

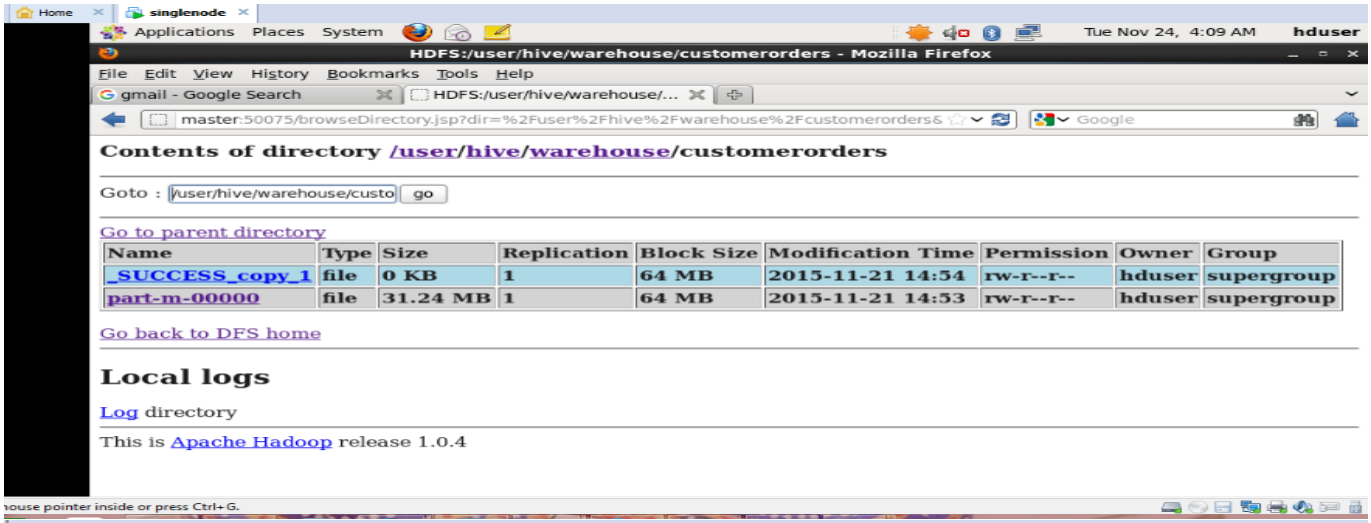
- Data ingestion is the process of obtaining, importing, and processing data for later use or storage in a database.
- In this step the data is imported from MySQL database to hive ecosystem by using a tool called SQOOP
- Where SQOOP is a used in HADOOP to import the data from other sources

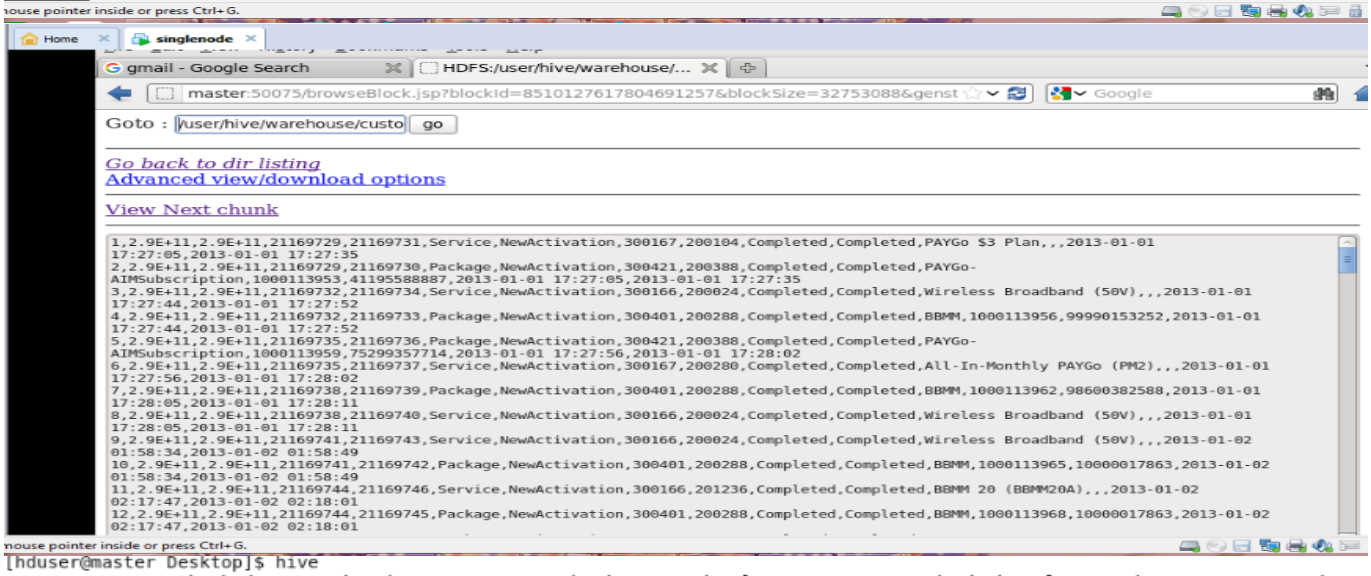
Step : III

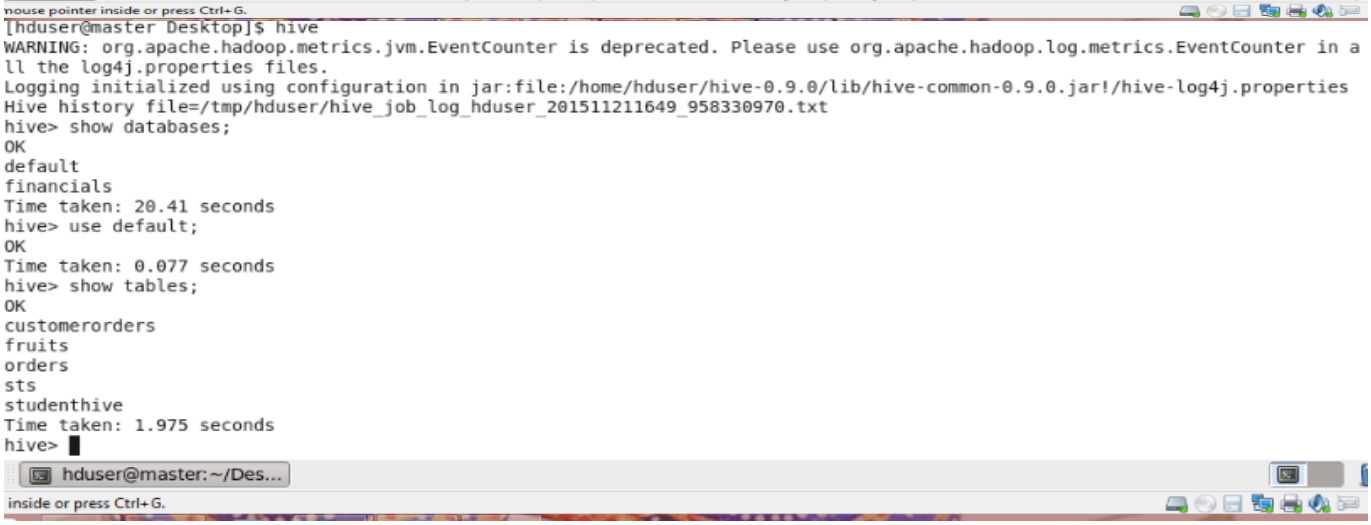
Data Storage in Hadoop

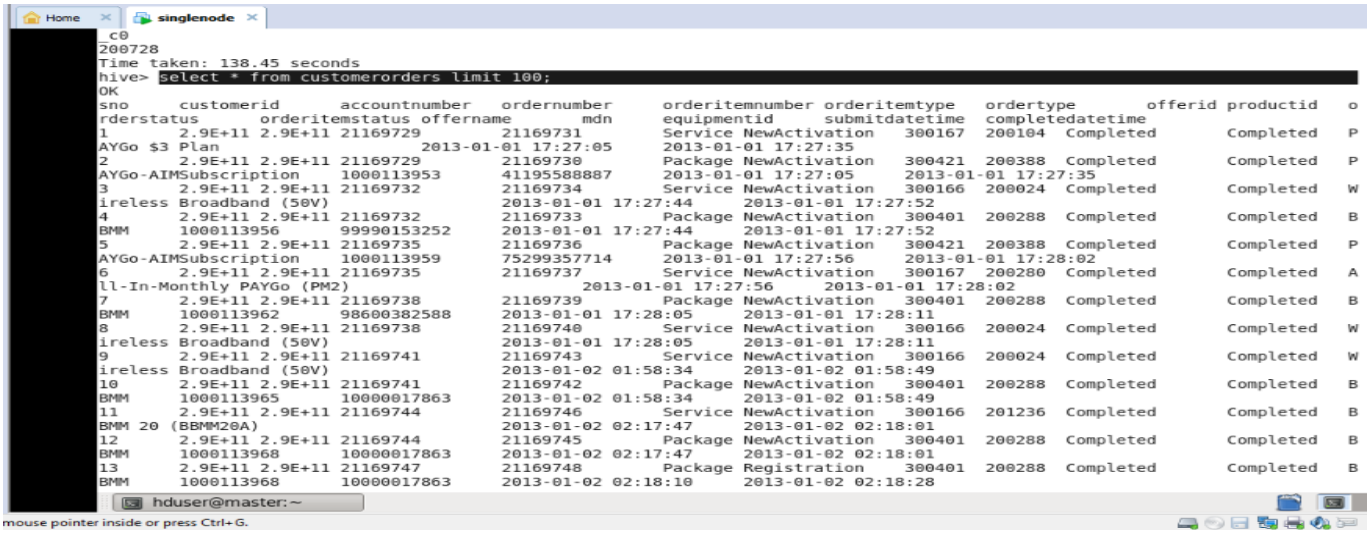
- In this step we used HDFS and hive components to make sure the data got imported into hive successfully or not.
- By using the web interface we can access the customerorders file which got imported into hive
- Once we get into hive by using several queries we checked the data got imported or not.




Home
singlenode
Applications Places System
Tue Nov 24, 4:09 AM hduser
HDFS:/user/hive/warehouse/customerorders - Mozilla Firefox
File Edit View History Bookmarks Tools Help
gmail - Google Search
HDFS:/user/hive/warehouse/...
master:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse%2Fcustomerorders&
Google
Contents of directory /user/hive/warehouse/customerorders
Goto : /user/hive/warehouse/custo go
Go to parent directory
Name Type Size Replication Block Size Modification Time Permission Owner Group
_SUCCESS_copy_1 file 0 KB 1 64 MB 2015-11-21 14:54 rw-r--r-- hduser supergroup
part-m-00000 file 31.24 MB 1 64 MB 2015-11-21 14:53 rw-r--r-- hduser supergroup
Go back to DFS home
Local logs
Log directory
This is Apache Hadoop release 1.0.4


Home
singlenode
gmail - Google Search
HDFS:/user/hive/warehouse/...
master:50075/browseBlock.jsp?blockId=8510127617804691257&blockSize=32753088&genst
Google
Goto : /user/hive/warehouse/custo go
Go back to dir listing
Advanced view/download options
View Next chunk
1,2.9E+11,2.9E+11,21169729,21169731,Service,NewActivation,300167,200104,Completed,Completed,PAYGo $3 Plan,,,2013-01-01 17:27:05,2013-01-01 17:27:35
2,2.9E+11,2.9E+11,21169729,21169730,Package,NewActivation,300421,200388,Completed,Completed,PAYGo-AIMSubscription,1000113953,41195588887,2013-01-01 17:27:05,2013-01-01 17:27:35
3,2.9E+11,2.9E+11,21169732,21169734,Service,NewActivation,300166,200024,Completed,Completed,Wireless Broadband (50V),,,2013-01-01 17:27:44,2013-01-01 17:27:52
4,2.9E+11,2.9E+11,21169732,21169733,Package,NewActivation,300401,200288,Completed,Completed,BBMM,1000113956,99990153252,2013-01-01 17:27:44,2013-01-01 17:27:52
5,2.9E+11,2.9E+11,21169735,21169736,Package,NewActivation,300421,200388,Completed,Completed,PAYGo-AIMSubscription,1000113959,75299357714,2013-01-01 17:27:56,2013-01-01 17:28:02
6,2.9E+11,2.9E+11,21169735,21169737,Service,NewActivation,300167,200280,Completed,Completed,All-In-Monthly PAYGo (PM2),,,2013-01-01 17:27:56,2013-01-01 17:28:02
7,2.9E+11,2.9E+11,21169738,21169739,Package,NewActivation,300401,200288,Completed,Completed,BBMM,1000113962,98600382588,2013-01-01 17:28:05,2013-01-01 17:28:11
8,2.9E+11,2.9E+11,21169738,21169740,Service,NewActivation,300166,200024,Completed,Completed,Wireless Broadband (50V),,,2013-01-01 17:28:05,2013-01-01 17:28:11
9,2.9E+11,2.9E+11,21169741,21169743,Service,NewActivation,300166,200024,Completed,Completed,Wireless Broadband (50V),,,2013-01-02 01:58:34,2013-01-02 01:58:49
10,2.9E+11,2.9E+11,21169741,21169742,Package,NewActivation,300401,200288,Completed,Completed,BBMM,1000113965,10000017863,2013-01-02 01:58:34,2013-01-02 01:58:49
11,2.9E+11,2.9E+11,21169744,21169746,Service,NewActivation,300166,201236,Completed,Completed,BBMM 20 (BBMM20A),,,2013-01-02 02:17:47,2013-01-02 02:18:01
12,2.9E+11,2.9E+11,21169744,21169745,Package,NewActivation,300401,200288,Completed,Completed,BBMM,1000113968,10000017863,2013-01-02 02:17:47,2013-01-02 02:18:01

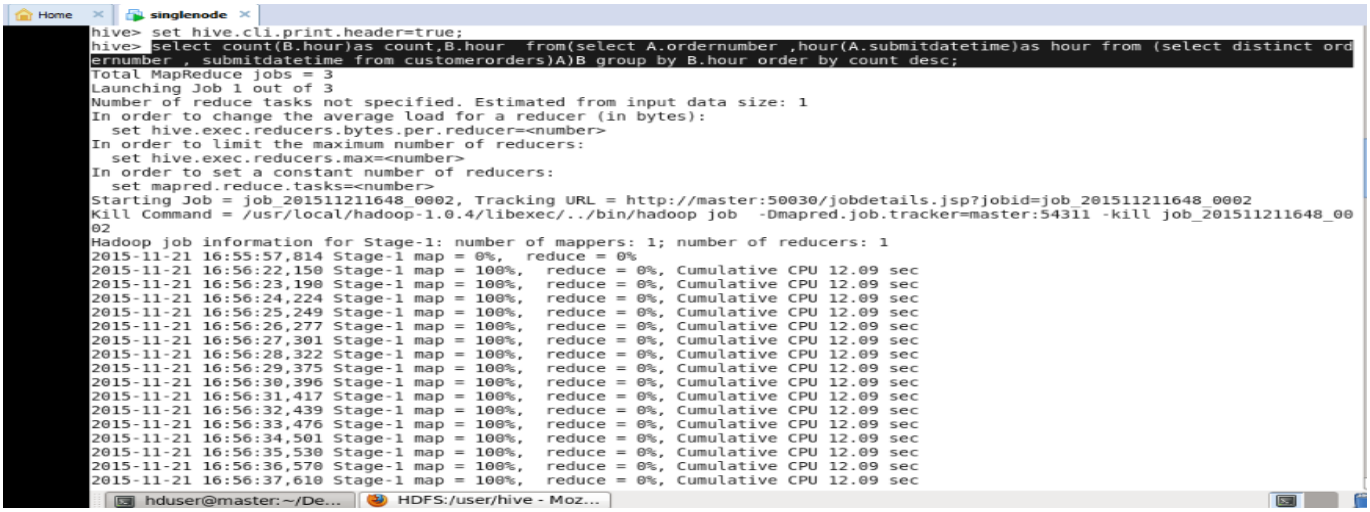

[hduser@master Desktop]$ hive
WARNING: org.apache.hadoop.metrics.jvm.EventCounter is deprecated. Please use org.apache.hadoop.log.metrics.EventCounter in all the log4j.properties files.
Logging initialized using configuration in jar:file:/home/hduser/hive-0.9.0/lib/hive-common-0.9.0.jar!/hive-log4j.properties
Hive history file=/tmp/hduser/hive_job_log_hduser_201511211649_958330970.txt
hive> show databases;
OK
default
financials
Time taken: 20.41 seconds
hive> use default;
OK
Time taken: 0.077 seconds
hive> show tables;
OK
customerorders
fruits
orders
sts
studenthive
Time taken: 1.975 seconds
hive>
hduser@master:~/Des...

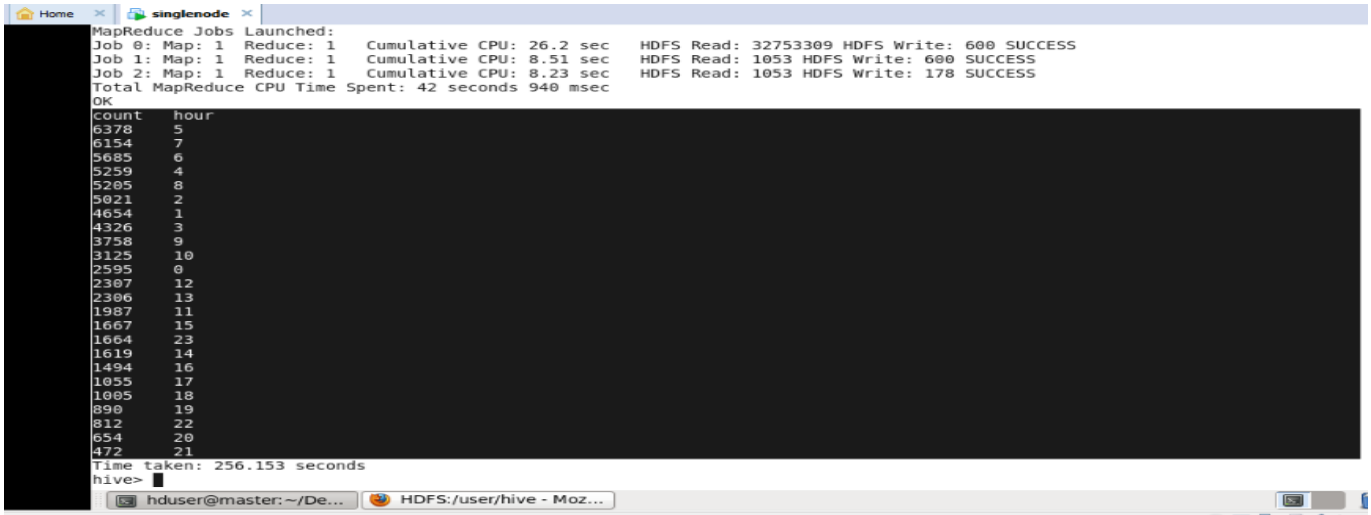


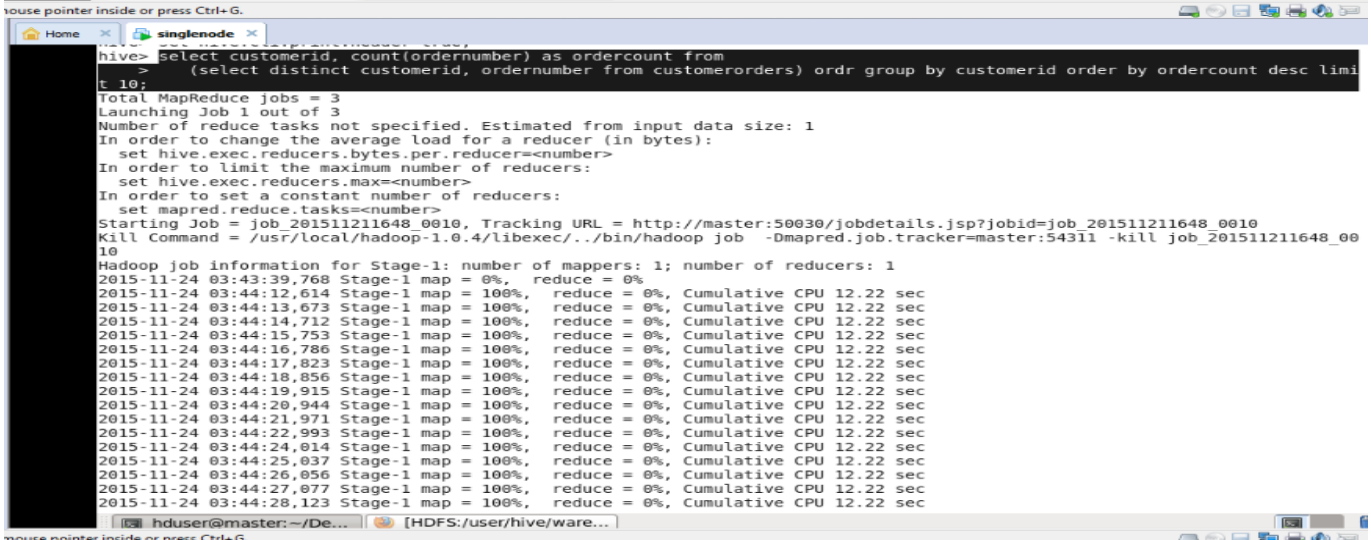
Step : IV

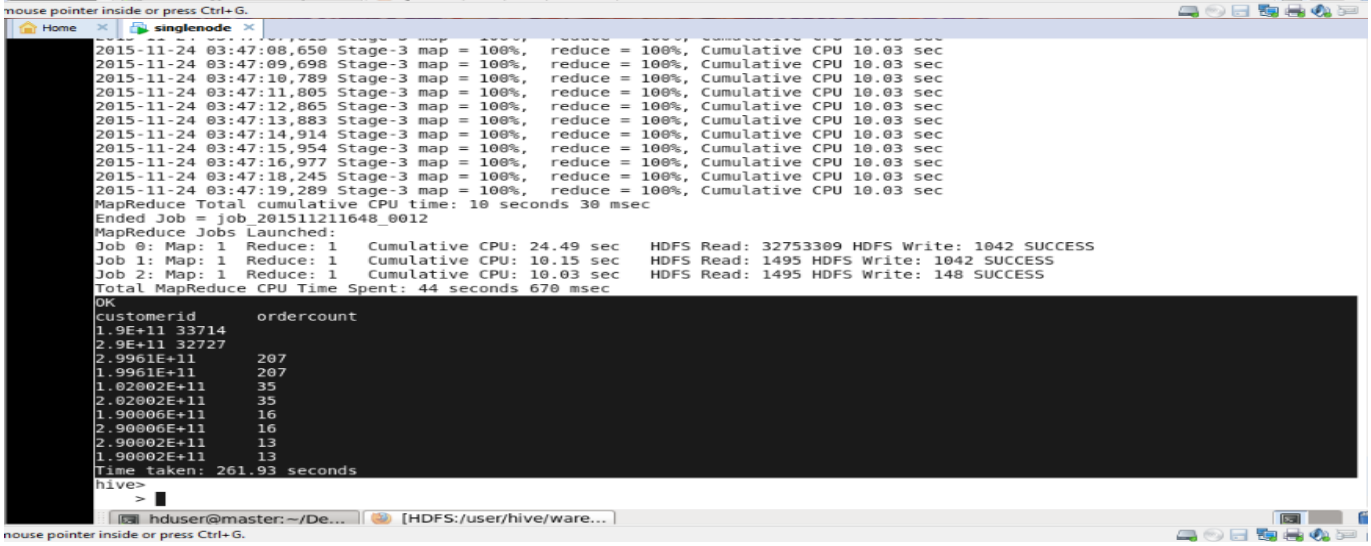
Data Analysis

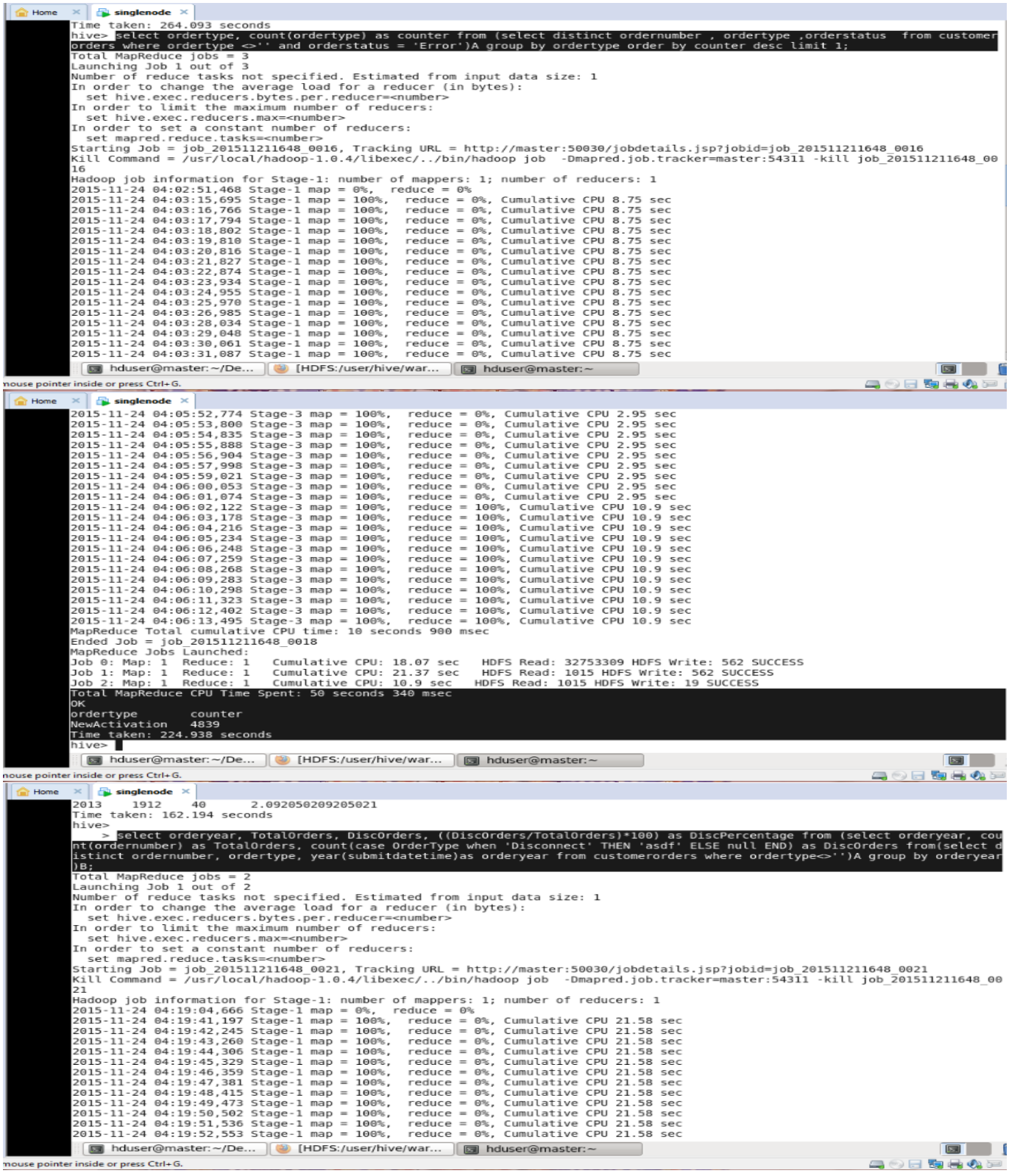
- In this step we used hive and map reduce components to analyze the data and to get the required reports
- Here Hive converts SQL queries into MapReduce jobs by using its own feature called HiveQL
- By executing selected four different hive queries on Hive. We analyzed the customer data

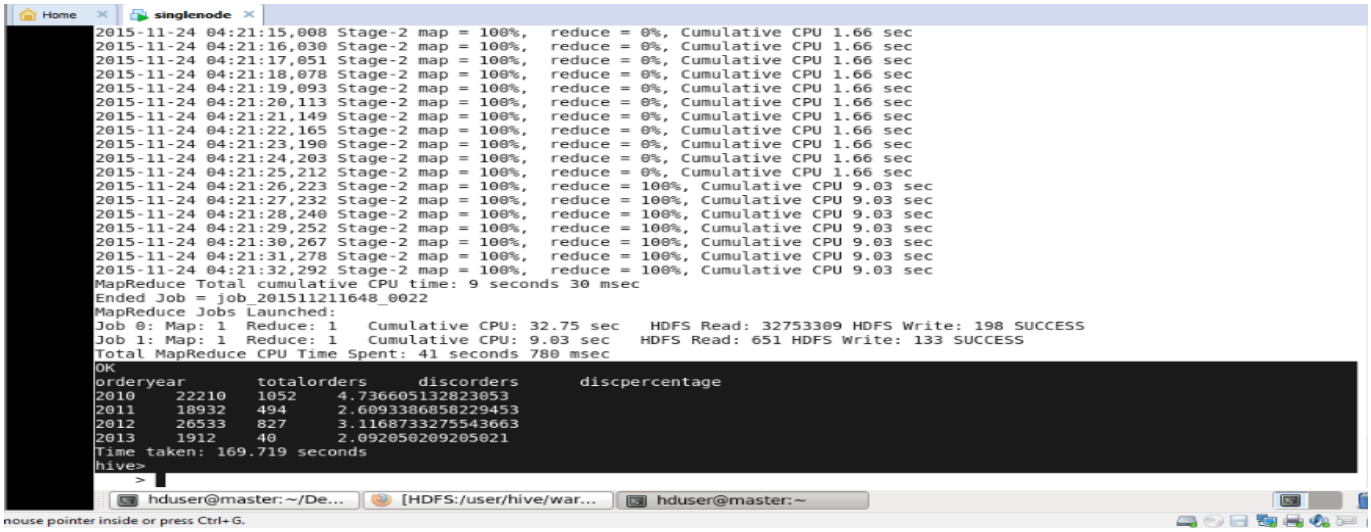



Home
singlenode
MapReduce Jobs Launched:
Job 0: Map: 1 Reduce: 1 Cumulative CPU: 26.2 sec HDFS Read: 32753309 HDFS Write: 600 SUCCESS
Job 1: Map: 1 Reduce: 1 Cumulative CPU: 8.51 sec HDFS Read: 1053 HDFS Write: 600 SUCCESS
Job 2: Map: 1 Reduce: 1 Cumulative CPU: 8.23 sec HDFS Read: 1053 HDFS Write: 178 SUCCESS
Total MapReduce CPU Time Spent: 42 seconds 940 msec
OK
count hour
6378 5
6154 7
5685 6
5259 4
5205 8
5021 2
4654 1
4326 3
3758 9
3125 10
2595 0
2307 12
2306 13
1987 11
1667 15
1664 23
1619 14
1494 16
1055 17
1005 18
890 19
812 22
654 20
472 21
Time taken: 256.153 seconds
hive>
hduser@master:~/De...
HDFS:/user/hive - Moz...
mouse pointer inside or press Ctrl+G.


Home
singlenode
hive> select customerid, count(ordernumber) as ordercount from
> (select distinct customerid, ordernumber from customerorders) ordr group by customerid order by ordercount desc limi
t 10;
Total MapReduce jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapred.reduce.tasks=<number>
Starting Job = job_201511211648_0010, Tracking URL = http://master:50030/jobdetails.jsp?jobid=job_201511211648_0010
Kill Command = /usr/local/hadoop-1.0.4/libexec/../bin/hadoop job -Dmapred.job.tracker=master:54311 -kill job_201511211648_00
10
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2015-11-24 03:43:39,768 Stage-1 map = 0%, reduce = 0%
2015-11-24 03:44:12,614 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:13,673 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:14,712 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:15,753 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:16,786 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:17,823 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:18,856 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:19,915 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:20,944 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:21,971 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:22,993 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:24,014 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:25,037 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:26,056 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:27,077 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
2015-11-24 03:44:28,123 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.22 sec
hduser@master:~/De...
[HDFS:/user/hive/ware...
mouse pointer inside or press Ctrl+G.


Home
singlenode
2015-11-24 03:47:08,650 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:09,698 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:10,789 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:11,805 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:12,865 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:13,883 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:14,914 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:15,954 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:16,977 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:18,245 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
2015-11-24 03:47:19,289 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
MapReduce Total cumulative CPU time: 10 seconds 30 msec
Ended Job = job_201511211648_0012
MapReduce Jobs Launched:
Job 0: Map: 1 Reduce: 1 Cumulative CPU: 24.49 sec HDFS Read: 32753309 HDFS Write: 1042 SUCCESS
Job 1: Map: 1 Reduce: 1 Cumulative CPU: 10.15 sec HDFS Read: 1495 HDFS Write: 1042 SUCCESS
Job 2: Map: 1 Reduce: 1 Cumulative CPU: 10.03 sec HDFS Read: 1495 HDFS Write: 148 SUCCESS
Total MapReduce CPU Time Spent: 44 seconds 670 msec
OK
customerid ordercount
1.9E+11 33714
2.9E+11 32727
2.9961E+11 207
1.9961E+11 207
1.02002E+11 35
2.02002E+11 35
1.90006E+11 16
2.90006E+11 16
2.90002E+11 13
1.90002E+11 13
Time taken: 261.93 seconds
hive>
>
hduser@master:~/De...
[HDFS:/user/hive/ware...
mouse pointer inside or press Ctrl+G.


Time taken: 264.093 seconds
hive> select ordertype, count(ordertype) as counter from (select distinct ordernumber , ordertype ,orderstatus from customerorders where ordertype <>'' and orderstatus = 'Error')A group by ordertype order by counter desc limit 1;
Total MapReduce jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapred.reduce.tasks=<number>
Starting Job = job_201511211648_0016, Tracking URL = http://master:50030/jobdetails.jsp?jobid=job_201511211648_0016
Kill Command = /usr/local/hadoop-1.0.4/libexec/../bin/hadoop job -Dmapred.job.tracker=master:54311 -kill job_201511211648_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2015-11-24 04:02:51,468 Stage-1 map = 0%, reduce = 0%
2015-11-24 04:03:15,695 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:16,766 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:17,794 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:18,802 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:19,810 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:20,816 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:21,827 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:22,874 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:23,934 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:24,955 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:25,970 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:26,985 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:28,034 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:29,048 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:30,061 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
2015-11-24 04:03:31,087 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.75 sec
hduser@master:~/De...
[HDFS:/user/hive/war...
hduser@master:~
mouse pointer inside or press Ctrl+G.
2015-11-24 04:05:52,774 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:05:53,800 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:05:54,835 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:05:55,888 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:05:56,904 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:05:57,998 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:05:59,021 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:06:00,053 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:06:01,074 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2015-11-24 04:06:02,122 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:03,178 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:04,216 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:05,234 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:06,248 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:07,259 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:08,268 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:09,283 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:10,298 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:11,323 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:12,402 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
2015-11-24 04:06:13,495 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 10.9 sec
MapReduce Total cumulative CPU time: 10 seconds 900 msec
Ended Job = job_201511211648_0018
MapReduce Jobs Launched:
Job 0: Map: 1 Reduce: 1 Cumulative CPU: 18.07 sec HDFS Read: 32753309 HDFS Write: 562 SUCCESS
Job 1: Map: 1 Reduce: 1 Cumulative CPU: 21.37 sec HDFS Read: 1015 HDFS Write: 562 SUCCESS
Job 2: Map: 1 Reduce: 1 Cumulative CPU: 10.9 sec HDFS Read: 1015 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 50 seconds 340 msec
OK
ordertype counter
NewActivation 4839
Time taken: 224.938 seconds
hive>
hduser@master:~/De...
[HDFS:/user/hive/war...
hduser@master:~
mouse pointer inside or press Ctrl+G.
2013 1912 40 2.092050209205021
Time taken: 162.194 seconds
hive>
> select orderyear, TotalOrders, DiscOrders, ((DiscOrders/TotalOrders)*100) as DiscPercentage from (select orderyear, count(ordernumber) as TotalOrders, count(case OrderType when 'Disconnect' THEN 'asdf' ELSE null END) as DiscOrders from(select distinct ordernumber, ordertype, year(submitdatetime)as orderyear from customerorders where ordertype<>'')A group by orderyear)B;
Total MapReduce jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapred.reduce.tasks=<number>
Starting Job = job_201511211648_0021, Tracking URL = http://master:50030/jobdetails.jsp?jobid=job_201511211648_0021
Kill Command = /usr/local/hadoop-1.0.4/libexec/../bin/hadoop job -Dmapred.job.tracker=master:54311 -kill job_201511211648_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2015-11-24 04:19:04,666 Stage-1 map = 0%, reduce = 0%
2015-11-24 04:19:41,197 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:42,245 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:43,260 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:44,306 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:45,329 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:46,359 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:47,381 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:48,415 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:49,473 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:50,502 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:51,536 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
2015-11-24 04:19:52,553 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.58 sec
hduser@master:~/De...
[HDFS:/user/hive/war...
hduser@master:~
mouse pointer inside or press Ctrl+G.



2.3 Interactions with other Projects (if Any)

It can interacted with other reporting tools to visualize the results.

2.4 Interactions with other Applications

Interaction with other RDMS system like MySQL is needed here to ingest the data.

2.5 Capabilities

MySQL should provide capabilities to support retrieving the data from its database from Hadoop. HIVE should be able to handle very big data in less response time.

2.6 Risk Assessment and Management

Risk associated with this project is when data ingestion is happening, Developer may loss some data if any services of Hadoop are not running or any nodes are failed. To make sure to not to happen this keep poll cluster to ensure all services are up and running.

3 Project Requirements

3.1 Identification of Requirements

Need at least single node running Hadoop cluster with MySQL installed and Hadoop 1.0.4 installed.
And HIVE any version have installed and successfully running on cluster .

3.2 Operations, Administration, Maintenance and Provisioning (OAM&P)

There is no special skill requirement needed for maintenance and administration until unless user is familiar with basic HDFS commands.

3.3 Security and Fraud Prevention

There no special requirement to take for security here , Bu default Hadoop provides key based communication between nodes.

3.4 Release and Transition Plan

Project will be deployed to the client by ingesting HIVE results to other reporting tools to create user interface for all results in the form of tables.

4 Project Design Description

Given the massive number of transactions processed by telecommunications companies; and the costs and complexity involved in their operations, data analytics offers a valuable opportunity for enhancing the frameworks and procedures they adopt to drive profitability and minimize unnecessary downside risk.

For simpler and faster processing of only relevant data, service providers need an advanced analytics driven big data solution that will help them to achieve timely and accurate insights using data mining and predictive analytics, text mining, forecasting and optimization capability to continuously drive innovation and help service providers make the best possible decisions.

5 Project Internal/external Interface Impacts and Specification

As we discuss about the project external interface impact, it deals with the MySQL which plays a major impact in importing the data into Hive. In case if the importing got failed or import was not successful then the analyzing of data cannot be done and we need to stop the project at same time, which generate a heavy loss to company and client. Coming to project Internal Interface , it impacts very less when compared to the external interface

6 Project Design Units Impacts

There will be no impact on MySQL database. Once the ingestion done there will be impact on HIVE database and table will be updated.

6.1 Functional Area/Design Unit

6.1.1 Functional Overview

This project starts with importing the data from MySQL to HIVE, after that several queries are applied to analyze the data in a functional way.

6.1.2 Impacts

This project completely deals in between MySQL an HIVE, SO no other elements will get effected.

6.1.3 Requirements

- Hadoop distribution.(Hadoop 1.0.4)

- MapReduce, Sqoop, Hive.
- VM ware workstation 9.0/VM ware Player 5 or greater.
- SSH Server as well as Client.
- JDK 1.7

7 References

http://birtanalytics.actuate.com/telecom-analytics

http://spotfire.tibco.com/blog/?p=18318