**Create partitioned table (single bucket)**

hive (niit)> create table txnrecsByCat(txnno INT, txndate STRING, custno INT, amount DOUBLE,

> product STRING, city STRING, state STRING, spendby STRING)

> partitioned by (category STRING)

> row format delimited

> fields terminated by ','

> stored as textfile;

hive (niit)> set hive.exec.dynamic.partition.mode=nonstrict;

hive (niit)> set hive.exec.dynamic.partition=true;

hive (niit)> set hive.enforce.bucketing=true;

hive (niit)> from txnrecords txn INSERT OVERWRITE TABLE txnrecsByCat PARTITION(category) select txn.txnno, txn.txndate,txn.custno, txn.amount,txn.product,txn.city,txn.state, txn.spendby, txn.category DISTRIBUTE By category;

Query ID = hduser\_20171223200350\_dcedc51e-b0c6-4ded-a795-4dde27180b15

Total jobs = 1

Launching Job 1 out of 1

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 mapreduce.job.reduces=<number>

Starting Job = job\_1514000082488\_0028, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0028/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0028

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-12-23 20:03:53,905 Stage-1 map = 0%, reduce = 0%

2017-12-23 20:03:58,020 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.27 sec

2017-12-23 20:04:04,176 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.0 sec

MapReduce Total cumulative CPU time: 6 seconds 0 msec

Ended Job = job\_1514000082488\_0028

Loading data to table niit.txnrecsbycat partition (category=null)

Time taken for load dynamic partitions : 1237

Loading partition {category=Team Sports}

Loading partition {category=Jumping}

Loading partition {category=Exercise & Fitness}

Loading partition {category=Racquet Sports}

Loading partition {category=Dancing}

Loading partition {category=Outdoor Recreation}

Loading partition {category=Combat Sports}

Loading partition {category=Winter Sports}

Loading partition {category=Games}

Loading partition {category=Puzzles}

Loading partition {category=Gymnastics}

Loading partition {category=Indoor Games}

Loading partition {category=Outdoor Play Equipment}

Loading partition {category=Air Sports}

Loading partition {category=Water Sports}

Time taken for adding to write entity : 4

Partition niit.txnrecsbycat{category=Air Sports} stats: [numFiles=1, numRows=960, totalSize=65688, rawDataSize=64728]

Partition niit.txnrecsbycat{category=Combat Sports} stats: [numFiles=1, numRows=1630, totalSize=108080, rawDataSize=106450]

Partition niit.txnrecsbycat{category=Dancing} stats: [numFiles=1, numRows=414, totalSize=28499, rawDataSize=28085]

Partition niit.txnrecsbycat{category=Exercise & Fitness} stats: [numFiles=1, numRows=7394, totalSize=548520, rawDataSize=541126]

Partition niit.txnrecsbycat{category=Games} stats: [numFiles=1, numRows=3666, totalSize=258424, rawDataSize=254758]

Partition niit.txnrecsbycat{category=Gymnastics} stats: [numFiles=1, numRows=3196, totalSize=234784, rawDataSize=231588]

Partition niit.txnrecsbycat{category=Indoor Games} stats: [numFiles=1, numRows=2799, totalSize=188252, rawDataSize=185453]

Partition niit.txnrecsbycat{category=Jumping} stats: [numFiles=1, numRows=2015, totalSize=145347, rawDataSize=143332]

Partition niit.txnrecsbycat{category=Outdoor Play Equipment} stats: [numFiles=1, numRows=2910, totalSize=201470, rawDataSize=198560]

Partition niit.txnrecsbycat{category=Outdoor Recreation} stats: [numFiles=1, numRows=8383, totalSize=577634, rawDataSize=569251]

Partition niit.txnrecsbycat{category=Puzzles} stats: [numFiles=1, numRows=612, totalSize=44909, rawDataSize=44297]

Partition niit.txnrecsbycat{category=Racquet Sports} stats: [numFiles=1, numRows=1611, totalSize=106119, rawDataSize=104508]

Partition niit.txnrecsbycat{category=Team Sports} stats: [numFiles=1, numRows=6010, totalSize=405097, rawDataSize=399087]

Partition niit.txnrecsbycat{category=Water Sports} stats: [numFiles=1, numRows=5219, totalSize=365864, rawDataSize=360645]

Partition niit.txnrecsbycat{category=Winter Sports} stats: [numFiles=1, numRows=3181, totalSize=220705, rawDataSize=217524]

MapReduce Jobs Launched:

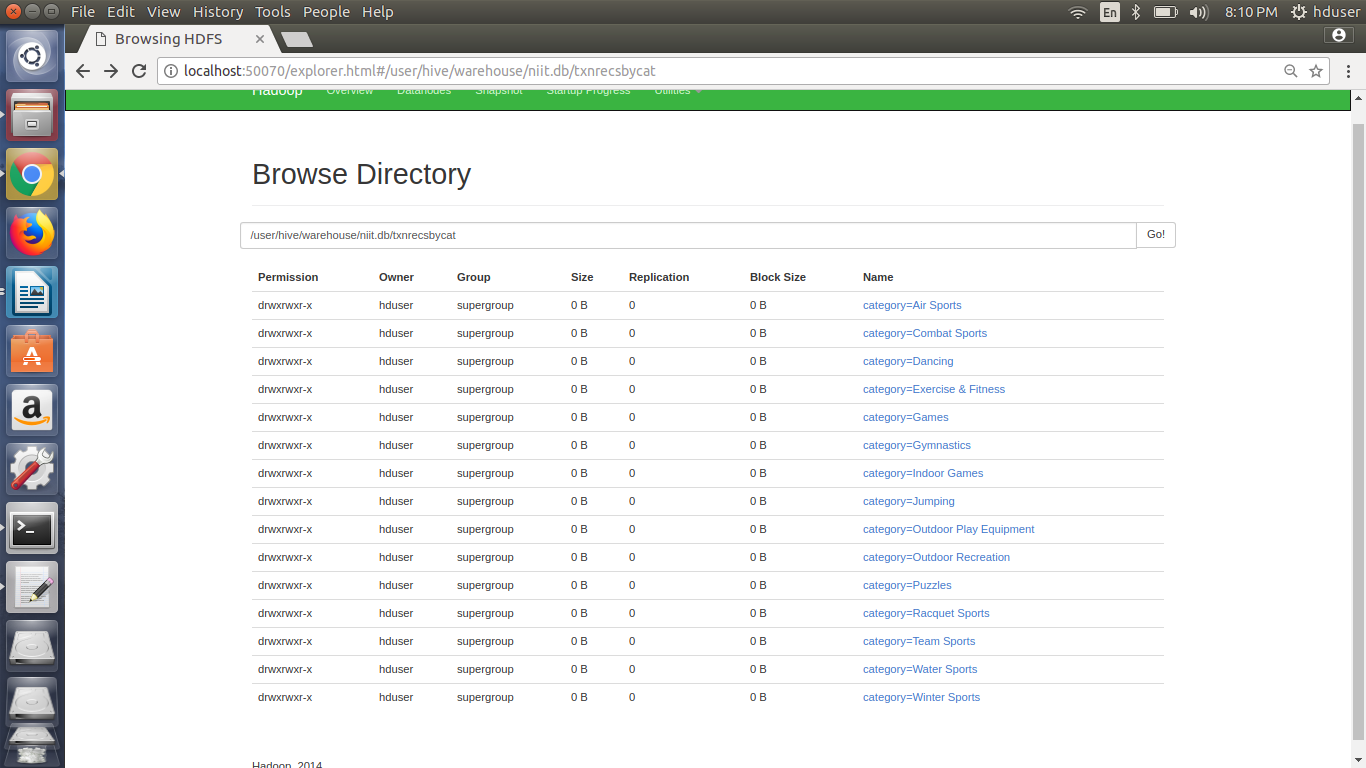
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.0 sec HDFS Read: 4427416 HDFS Write: 3500382 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 0 msec

OK

txn.txnno txn.txndate txn.custno txn.amount txn.product txn.city txn.state txn.spendby txn.category

Time taken: 17.882 seconds



**Without Partition:**

hive (niit)> select sum(amount) from txnrecords where category="Air Sports";

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.66 sec **HDFS Read: 4426656 HDFS Write: 18 SUCCESS**

Total MapReduce CPU Time Spent: 3 seconds 660 msec

OK

\_c0

99316.89999999994

Time taken: 14.011 seconds, Fetched: 1 row(s)

**With Partition:**

hive (niit)> select sum(amount) from txnrecsByCat where category="Air Sports";

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.31 sec **HDFS Read: 73710 HDFS Write: 18 SUCCESS**

Total MapReduce CPU Time Spent: 2 seconds 310 msec

OK

\_c0

99316.90000000005

Time taken: 12.911 seconds, Fetched: 1 row(s)

**Create partitioned table (with multiple buckets)**

hive (niit)> create table txnrecsByCat2(txnno INT, txndate STRING, custno INT, amount DOUBLE,

> product STRING, city STRING, state STRING, spendby STRING)

> partitioned by (category STRING)

> clustered by (state) into 10 buckets

> row format delimited

> fields terminated by ','

> stored as textfile;

OK

Time taken: 0.55 seconds

**hive (niit)> from txnrecords txn INSERT OVERWRITE TABLE txnrecsByCat2 PARTITION(category) select txn.txnno, txn.txndate,txn.custno, txn.amount,txn.product,txn.city,txn.state, txn.spendby, txn.category DISTRIBUTE By category;**

Query ID = hduser\_20171223212731\_75ff62c6-5b09-4479-ab1b-c62ced59cc3d

Total 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 mapreduce.job.reduces=<number>

Starting Job = job\_1514000082488\_0031, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0031/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0031

**Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1**

2017-12-23 21:27:36,074 Stage-1 map = 0%, reduce = 0%

2017-12-23 21:27:40,250 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.17 sec

2017-12-23 21:27:44,353 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.41 sec

MapReduce Total cumulative CPU time: 5 seconds 410 msec

Ended Job = job\_1514000082488\_0031

Launching Job 2 out of 2

Number of reduce tasks determined at compile time: 10

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 mapreduce.job.reduces=<number>

Starting Job = job\_1514000082488\_0032, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0032/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0032

**Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 10**

2017-12-23 21:27:53,826 Stage-2 map = 0%, reduce = 0%

2017-12-23 21:27:57,939 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.97 sec

2017-12-23 21:28:06,255 Stage-2 map = 100%, reduce = 10%, Cumulative CPU 6.66 sec

2017-12-23 21:28:07,294 Stage-2 map = 100%, reduce = 20%, Cumulative CPU 10.71 sec

2017-12-23 21:28:10,422 Stage-2 map = 100%, reduce = 40%, Cumulative CPU 19.46 sec

2017-12-23 21:28:12,536 Stage-2 map = 100%, reduce = 50%, Cumulative CPU 23.94 sec

2017-12-23 21:28:14,633 Stage-2 map = 100%, reduce = 60%, Cumulative CPU 28.26 sec

2017-12-23 21:28:15,665 Stage-2 map = 100%, reduce = 80%, Cumulative CPU 37.57 sec

2017-12-23 21:28:16,686 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 47.16 sec

MapReduce Total cumulative CPU time: 47 seconds 160 msec

Ended Job = job\_1514000082488\_0032

**Loading data to table niit.txnrecsbycat2 partition (category=null)**

**Time taken for load dynamic partitions : 3424**

**Loading partition {category=Indoor Games}**

**Loading partition {category=Outdoor Play Equipment}**

**Loading partition {category=Outdoor Recreation}**

**Loading partition {category=Exercise & Fitness}**

**Loading partition {category=Air Sports}**

**Loading partition {category=Combat Sports}**

**Loading partition {category=Dancing}**

**Loading partition {category=Puzzles}**

**Loading partition {category=Winter Sports}**

**Loading partition {category=Water Sports}**

**Loading partition {category=Gymnastics}**

**Loading partition {category=Racquet Sports}**

**Loading partition {category=Team Sports}**

**Loading partition {category=Jumping}**

**Loading partition {category=Games}**

Time taken for adding to write entity : 3

Partition niit.txnrecsbycat2{category=Air Sports} stats: [numFiles=10, numRows=960, totalSize=65688, rawDataSize=64728]

Partition niit.txnrecsbycat2{category=Combat Sports} stats: [numFiles=10, numRows=1630, totalSize=108080, rawDataSize=106450]

Partition niit.txnrecsbycat2{category=Dancing} stats: [numFiles=10, numRows=414, totalSize=28499, rawDataSize=28085]

Partition niit.txnrecsbycat2{category=Exercise & Fitness} stats: [numFiles=10, numRows=7394, totalSize=548520, rawDataSize=541126]

Partition niit.txnrecsbycat2{category=Games} stats: [numFiles=10, numRows=3666, totalSize=258424, rawDataSize=254758]

Partition niit.txnrecsbycat2{category=Gymnastics} stats: [numFiles=10, numRows=3196, totalSize=234784, rawDataSize=231588]

Partition niit.txnrecsbycat2{category=Indoor Games} stats: [numFiles=10, numRows=2799, totalSize=188252, rawDataSize=185453]

Partition niit.txnrecsbycat2{category=Jumping} stats: [numFiles=10, numRows=2015, totalSize=145347, rawDataSize=143332]

Partition niit.txnrecsbycat2{category=Outdoor Play Equipment} stats: [numFiles=10, numRows=2910, totalSize=201470, rawDataSize=198560]

Partition niit.txnrecsbycat2{category=Outdoor Recreation} stats: [numFiles=10, numRows=8383, totalSize=577634, rawDataSize=569251]

Partition niit.txnrecsbycat2{category=Puzzles} stats: [numFiles=10, numRows=612, totalSize=44909, rawDataSize=44297]

Partition niit.txnrecsbycat2{category=Racquet Sports} stats: [numFiles=10, numRows=1611, totalSize=106119, rawDataSize=104508]

Partition niit.txnrecsbycat2{category=Team Sports} stats: [numFiles=10, numRows=6010, totalSize=405097, rawDataSize=399087]

Partition niit.txnrecsbycat2{category=Water Sports} stats: [numFiles=10, numRows=5219, totalSize=365864, rawDataSize=360645]

Partition niit.txnrecsbycat2{category=Winter Sports} stats: [numFiles=10, numRows=3181, totalSize=220705, rawDataSize=217524]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.41 sec HDFS Read: 4426099 HDFS Write: 4915372 SUCCESS

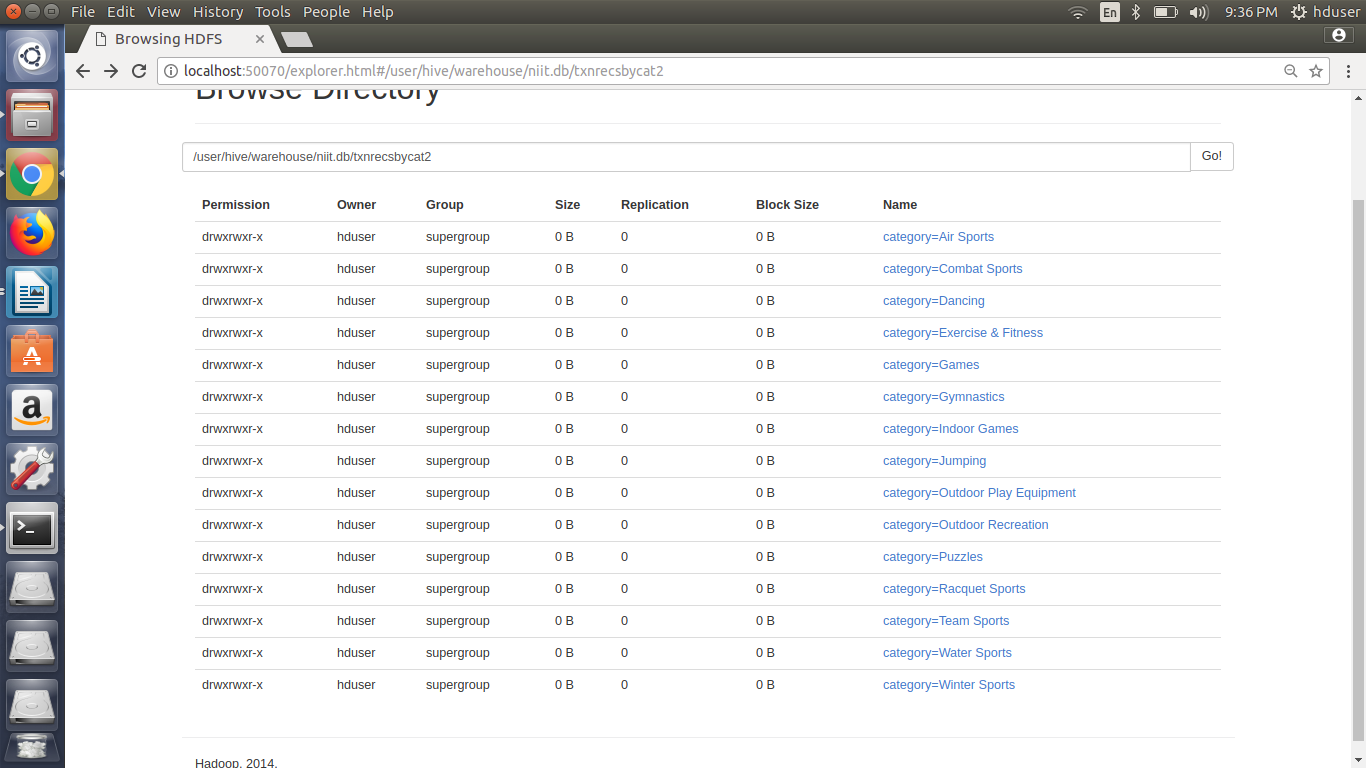
Stage-Stage-2: Map: 1 Reduce: 10 Cumulative CPU: 47.16 sec HDFS Read: 4959955 HDFS Write: 3509114 SUCCESS

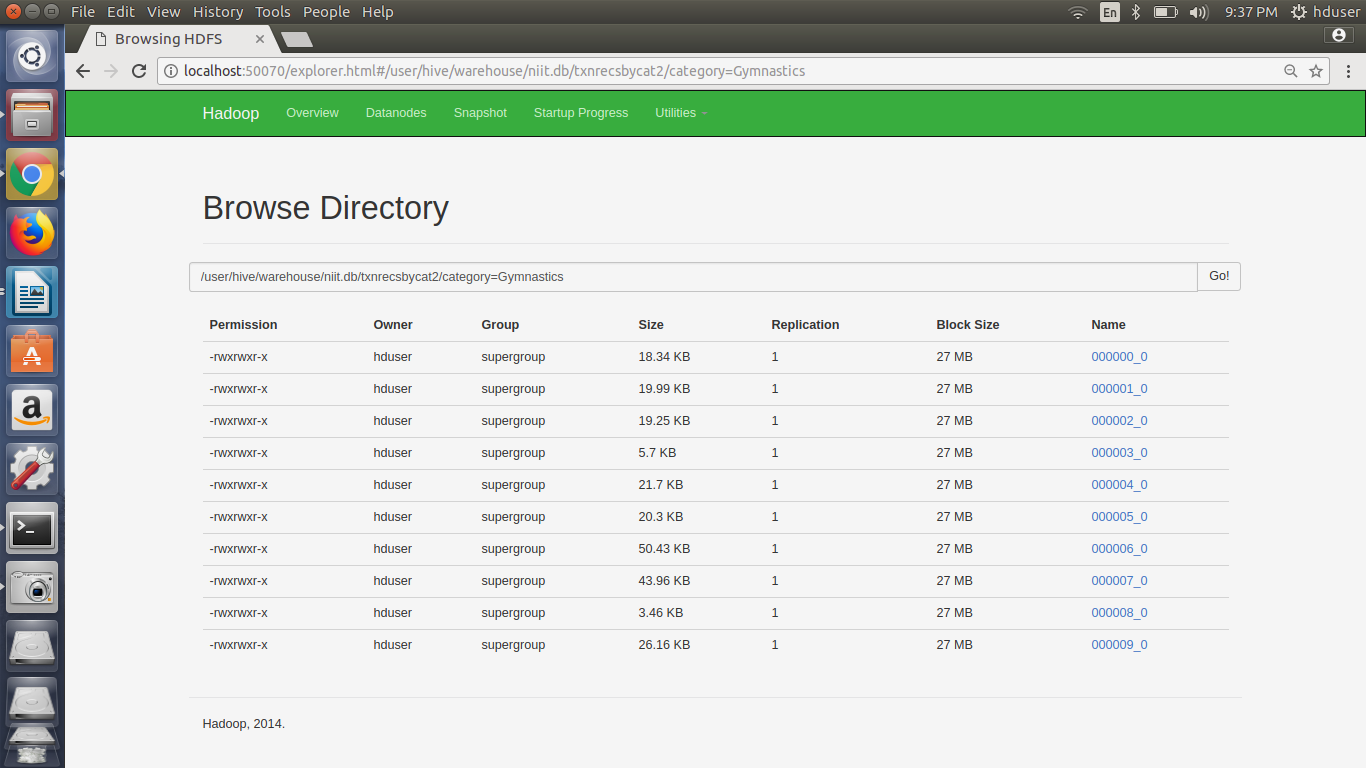
Total MapReduce CPU Time Spent: 52 seconds 570 msec

OK

txn.txnno txn.txndate txn.custno txn.amount txn.product txn.city txn.state txn.spendby txn.category

Time taken: 51.189 seconds





Create partitioned table (single bucket) on a derived column

**hive (niit)> create table txnrecsByCat3(txnno INT, txndate STRING, custno INT, amount DOUBLE,**

**> category string, product STRING, city STRING, state STRING, spendby STRING)**

**> partitioned by (month string)**

**> row format delimited**

**> fields terminated by ','**

**> stored as textfile;**

OK

Time taken: 0.032 seconds

**hive (niit)> from txnrecords txn INSERT OVERWRITE TABLE txnrecsByCat3 PARTITION(month) select txn.txnno, txn.txndate,txn.custno, txn.amount,txn.product,txn.city,txn.state, txn.spendby, txn.category, substring(txn.txndate,1,2) as month DISTRIBUTE By substring(txndate,1,2);**

Query ID = hduser\_20171223213932\_b6481642-4186-4340-b9c5-821e5a79e511

Total jobs = 1

Launching Job 1 out of 1

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 mapreduce.job.reduces=<number>

Starting Job = job\_1514000082488\_0033, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0033/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0033

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-12-23 21:39:36,113 Stage-1 map = 0%, reduce = 0%

2017-12-23 21:39:41,255 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.11 sec

2017-12-23 21:39:46,405 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.08 sec

MapReduce Total cumulative CPU time: 8 seconds 80 msec

Ended Job = job\_1514000082488\_0033

Loading data to table niit.txnrecsbycat3 partition (month=null)

Time taken for load dynamic partitions : 629

**Loading partition {month=05}**

**Loading partition {month=08}**

**Loading partition {month=10}**

**Loading partition {month=07}**

**Loading partition {month=12}**

**Loading partition {month=01}**

**Loading partition {month=03}**

**Loading partition {month=09}**

**Loading partition {month=11}**

**Loading partition {month=06}**

**Loading partition {month=04}**

**Loading partition {month=02}**

Time taken for adding to write entity : 5

Partition niit.txnrecsbycat3{month=01} stats: [numFiles=1, numRows=4268, totalSize=361072, rawDataSize=356804]

Partition niit.txnrecsbycat3{month=02} stats: [numFiles=1, numRows=3834, totalSize=324576, rawDataSize=320742]

Partition niit.txnrecsbycat3{month=03} stats: [numFiles=1, numRows=4360, totalSize=369112, rawDataSize=364752]

Partition niit.txnrecsbycat3{month=04} stats: [numFiles=1, numRows=4175, totalSize=352386, rawDataSize=348211]

Partition niit.txnrecsbycat3{month=05} stats: [numFiles=1, numRows=4205, totalSize=355646, rawDataSize=351441]

Partition niit.txnrecsbycat3{month=06} stats: [numFiles=1, numRows=4171, totalSize=352168, rawDataSize=347997]

Partition niit.txnrecsbycat3{month=07} stats: [numFiles=1, numRows=4252, totalSize=359250, rawDataSize=354998]

Partition niit.txnrecsbycat3{month=08} stats: [numFiles=1, numRows=4250, totalSize=357930, rawDataSize=353680]

Partition niit.txnrecsbycat3{month=09} stats: [numFiles=1, numRows=4160, totalSize=352024, rawDataSize=347864]

Partition niit.txnrecsbycat3{month=10} stats: [numFiles=1, numRows=4181, totalSize=352615, rawDataSize=348434]

Partition niit.txnrecsbycat3{month=11} stats: [numFiles=1, numRows=3994, totalSize=338528, rawDataSize=334534]

Partition niit.txnrecsbycat3{month=12} stats: [numFiles=1, numRows=4150, totalSize=350691, rawDataSize=346541]

MapReduce Jobs Launched:

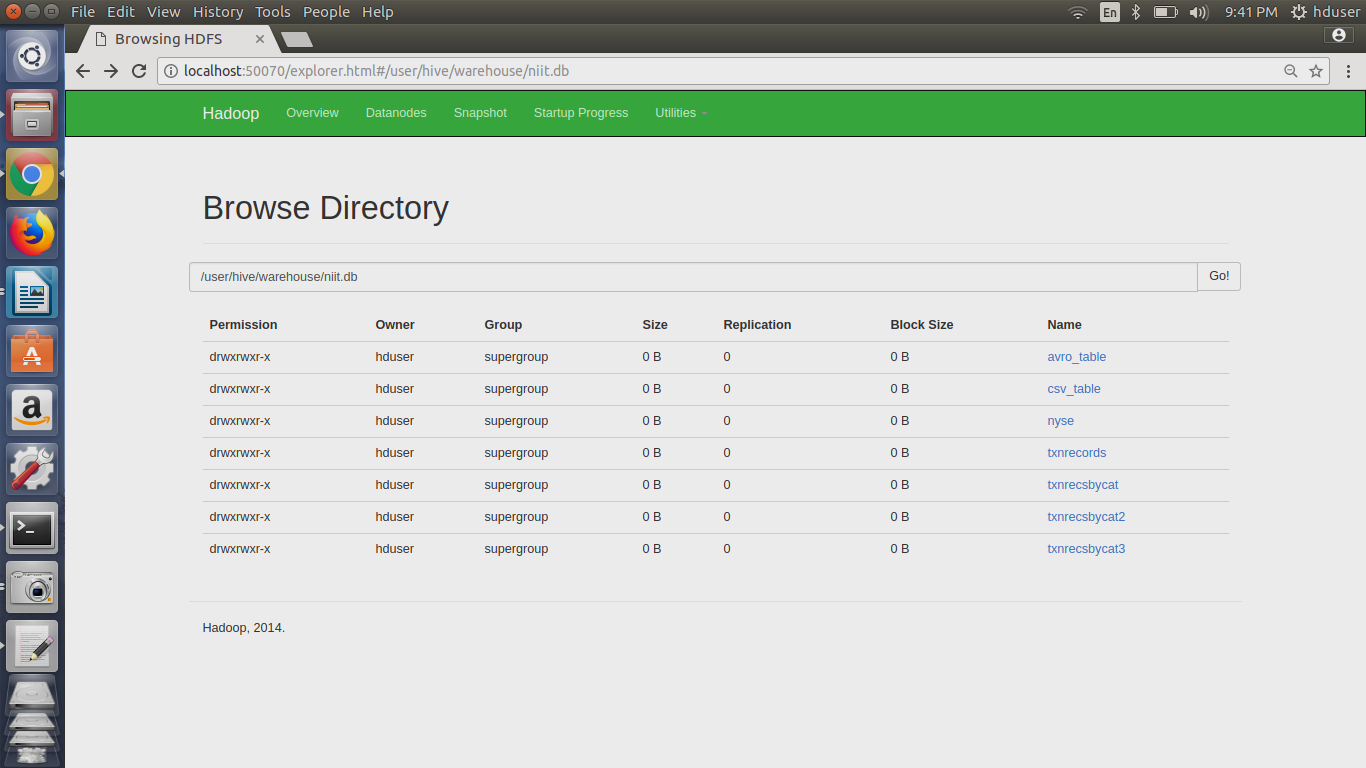
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.08 sec HDFS Read: 4428283 HDFS Write: 4226659 SUCCESS

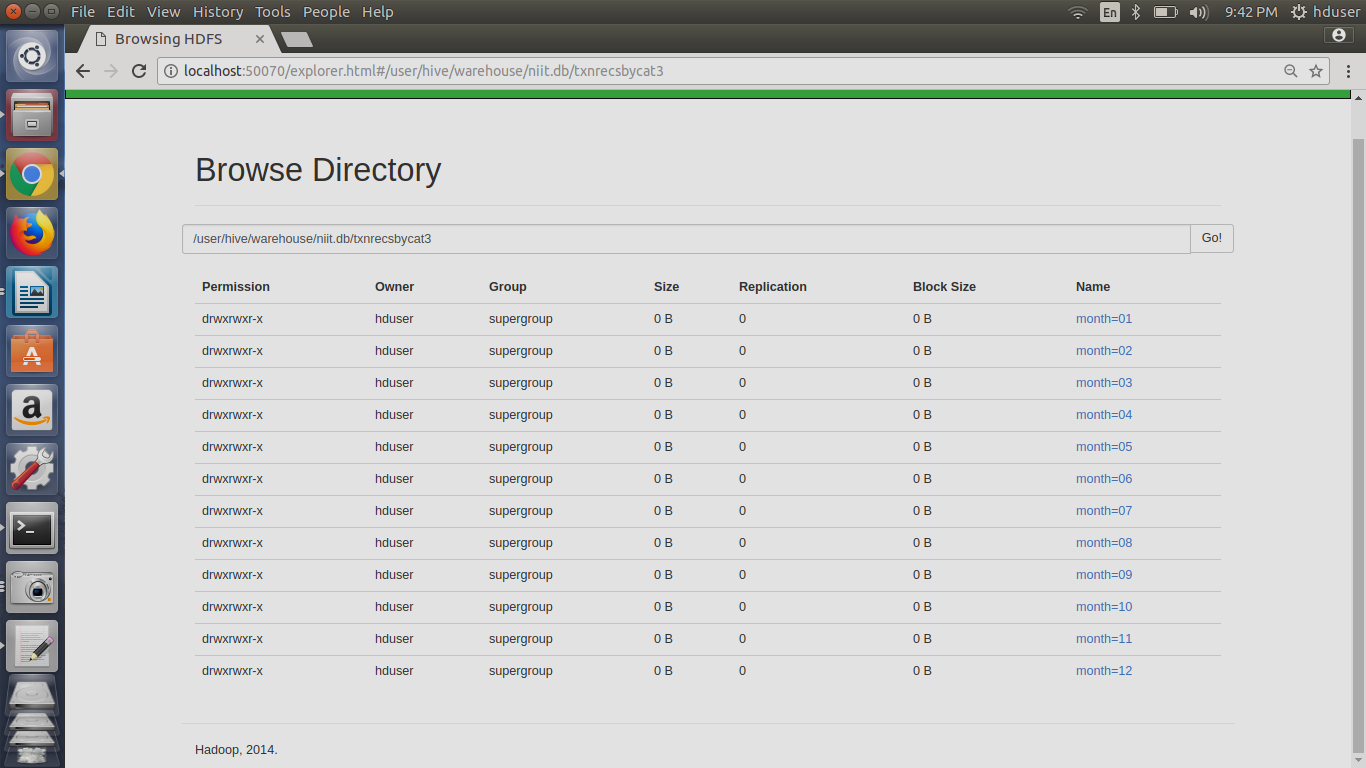
Total MapReduce CPU Time Spent: 8 seconds 80 msec

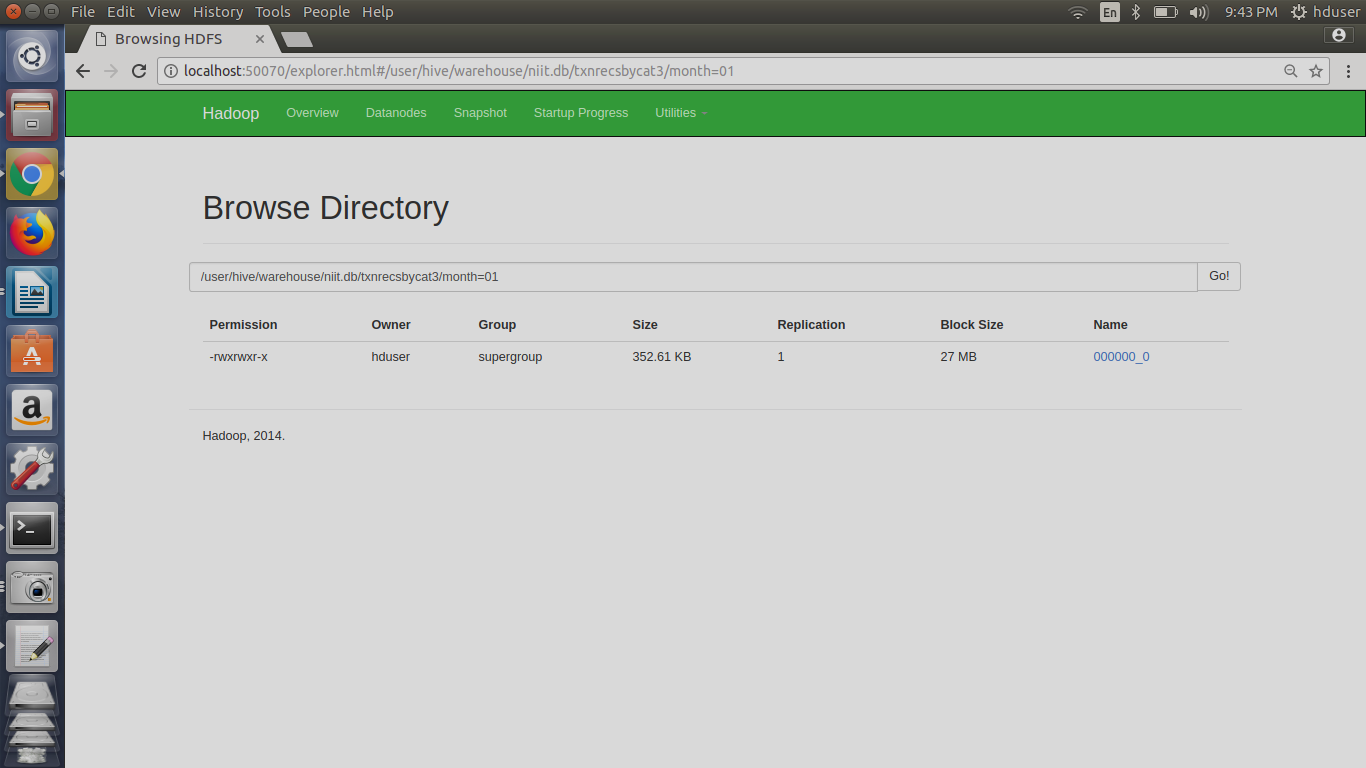
OK

txn.txnno txn.txndate txn.custno txn.amount txn.product txn.city txn.state txn.spendby txn.category month

Time taken: 15.863 seconds







H4. Create partitioned table (single bucket) on multiple columns

hive (niit)> create table txnrecsByCat4(txnno INT, txndate STRING, custno INT, amount DOUBLE,

> product STRING, city STRING, state STRING)

> partitioned by (category STRING, spendby string)

> row format delimited

> fields terminated by ','

> stored as textfile;

OK

Time taken: 0.264 seconds

hive (niit)> from txnrecords txn INSERT OVERWRITE TABLE txnrecsByCat4 PARTITION(category,spendby) select txn.txnno, txn.txndate,txn.custno, txn.amount,txn.product,txn.city,txn.state, txn.category, txn.spendby DISTRIBUTE By category, spendby;

Query ID = hduser\_20171223214629\_c948f73b-3b53-454d-ba78-7fd24cdc9f7b

Total jobs = 1

Launching Job 1 out of 1

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 mapreduce.job.reduces=<number>

Starting Job = job\_1514000082488\_0034, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0034/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0034

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-12-23 21:46:34,225 Stage-1 map = 0%, reduce = 0%

2017-12-23 21:46:38,436 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.34 sec

2017-12-23 21:46:43,570 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.25 sec

MapReduce Total cumulative CPU time: 6 seconds 250 msec

Ended Job = job\_1514000082488\_0034

Loading data to table niit.txnrecsbycat4 partition (category=null, spendby=null)

Time taken for load dynamic partitions : 2159

Loading partition {category=Gymnastics, spendby=credit}

Loading partition {category=Puzzles, spendby=cash}

Loading partition {category=Outdoor Play Equipment, spendby=credit}

Loading partition {category=Team Sports, spendby=cash}

Loading partition {category=Games, spendby=credit}

Loading partition {category=Outdoor Recreation, spendby=credit}

Loading partition {category=Air Sports, spendby=cash}

Loading partition {category=Gymnastics, spendby=cash}

Loading partition {category=Outdoor Play Equipment, spendby=cash}

Loading partition {category=Indoor Games, spendby=cash}

Loading partition {category=Water Sports, spendby=credit}

Loading partition {category=Games, spendby=cash}

Loading partition {category=Dancing, spendby=cash}

Loading partition {category=Dancing, spendby=credit}

Loading partition {category=Racquet Sports, spendby=cash}

Loading partition {category=Team Sports, spendby=credit}

Loading partition {category=Indoor Games, spendby=credit}

Loading partition {category=Combat Sports, spendby=credit}

Loading partition {category=Air Sports, spendby=credit}

Loading partition {category=Jumping, spendby=credit}

Loading partition {category=Exercise & Fitness, spendby=credit}

Loading partition {category=Puzzles, spendby=credit}

Loading partition {category=Exercise & Fitness, spendby=cash}

Loading partition {category=Racquet Sports, spendby=credit}

Loading partition {category=Combat Sports, spendby=cash}

Loading partition {category=Outdoor Recreation, spendby=cash}

Loading partition {category=Winter Sports, spendby=cash}

Loading partition {category=Water Sports, spendby=cash}

Loading partition {category=Winter Sports, spendby=credit}

Loading partition {category=Jumping, spendby=cash}

Time taken for adding to write entity : 5

Partition niit.txnrecsbycat4{category=Air Sports, spendby=cash} stats: [numFiles=1, numRows=123, totalSize=7493, rawDataSize=7370]

Partition niit.txnrecsbycat4{category=Air Sports, spendby=credit} stats: [numFiles=1, numRows=837, totalSize=51721, rawDataSize=50884]

Partition niit.txnrecsbycat4{category=Combat Sports, spendby=cash} stats: [numFiles=1, numRows=223, totalSize=13237, rawDataSize=13014]

Partition niit.txnrecsbycat4{category=Combat Sports, spendby=credit} stats: [numFiles=1, numRows=1407, totalSize=83879, rawDataSize=82472]

Partition niit.txnrecsbycat4{category=Dancing, spendby=cash} stats: [numFiles=1, numRows=59, totalSize=3648, rawDataSize=3589]

Partition niit.txnrecsbycat4{category=Dancing, spendby=credit} stats: [numFiles=1, numRows=355, totalSize=22071, rawDataSize=21716]

Partition niit.txnrecsbycat4{category=Exercise & Fitness, spendby=cash} stats: [numFiles=1, numRows=898, totalSize=59987, rawDataSize=59089]

Partition niit.txnrecsbycat4{category=Exercise & Fitness, spendby=credit} stats: [numFiles=1, numRows=6496, totalSize=438571, rawDataSize=432075]

Partition niit.txnrecsbycat4{category=Games, spendby=cash} stats: [numFiles=1, numRows=510, totalSize=32222, rawDataSize=31712]

Partition niit.txnrecsbycat4{category=Games, spendby=credit} stats: [numFiles=1, numRows=3156, totalSize=201560, rawDataSize=198404]

Partition niit.txnrecsbycat4{category=Gymnastics, spendby=cash} stats: [numFiles=1, numRows=453, totalSize=29912, rawDataSize=29459]

Partition niit.txnrecsbycat4{category=Gymnastics, spendby=credit} stats: [numFiles=1, numRows=2743, totalSize=183406, rawDataSize=180663]

Partition niit.txnrecsbycat4{category=Indoor Games, spendby=cash} stats: [numFiles=1, numRows=376, totalSize=22415, rawDataSize=22039]

Partition niit.txnrecsbycat4{category=Indoor Games, spendby=credit} stats: [numFiles=1, numRows=2423, totalSize=146996, rawDataSize=144573]

Partition niit.txnrecsbycat4{category=Jumping, spendby=cash} stats: [numFiles=1, numRows=264, totalSize=17122, rawDataSize=16858]

Partition niit.txnrecsbycat4{category=Jumping, spendby=credit} stats: [numFiles=1, numRows=1751, totalSize=114648, rawDataSize=112897]

Partition niit.txnrecsbycat4{category=Outdoor Play Equipment, spendby=cash} stats: [numFiles=1, numRows=410, totalSize=25509, rawDataSize=25099]

Partition niit.txnrecsbycat4{category=Outdoor Play Equipment, spendby=credit} stats: [numFiles=1, numRows=2500, totalSize=156411, rawDataSize=153911]

Partition niit.txnrecsbycat4{category=Outdoor Recreation, spendby=cash} stats: [numFiles=1, numRows=1247, totalSize=76683, rawDataSize=75436]

Partition niit.txnrecsbycat4{category=Outdoor Recreation, spendby=credit} stats: [numFiles=1, numRows=7136, totalSize=444764, rawDataSize=437628]

Partition niit.txnrecsbycat4{category=Puzzles, spendby=cash} stats: [numFiles=1, numRows=91, totalSize=6072, rawDataSize=5981]

Partition niit.txnrecsbycat4{category=Puzzles, spendby=credit} stats: [numFiles=1, numRows=521, totalSize=34735, rawDataSize=34214]

Partition niit.txnrecsbycat4{category=Racquet Sports, spendby=cash} stats: [numFiles=1, numRows=218, totalSize=12674, rawDataSize=12456]

Partition niit.txnrecsbycat4{category=Racquet Sports, spendby=credit} stats: [numFiles=1, numRows=1393, totalSize=82604, rawDataSize=81211]

Partition niit.txnrecsbycat4{category=Team Sports, spendby=cash} stats: [numFiles=1, numRows=806, totalSize=48337, rawDataSize=47531]

Partition niit.txnrecsbycat4{category=Team Sports, spendby=credit} stats: [numFiles=1, numRows=5204, totalSize=316302, rawDataSize=311098]

Partition niit.txnrecsbycat4{category=Water Sports, spendby=cash} stats: [numFiles=1, numRows=742, totalSize=46683, rawDataSize=45941]

Partition niit.txnrecsbycat4{category=Water Sports, spendby=credit} stats: [numFiles=1, numRows=4477, totalSize=284132, rawDataSize=279655]

Partition niit.txnrecsbycat4{category=Winter Sports, spendby=cash} stats: [numFiles=1, numRows=429, totalSize=26538, rawDataSize=26109]

Partition niit.txnrecsbycat4{category=Winter Sports, spendby=credit} stats: [numFiles=1, numRows=2752, totalSize=172758, rawDataSize=170006]

MapReduce Jobs Launched:

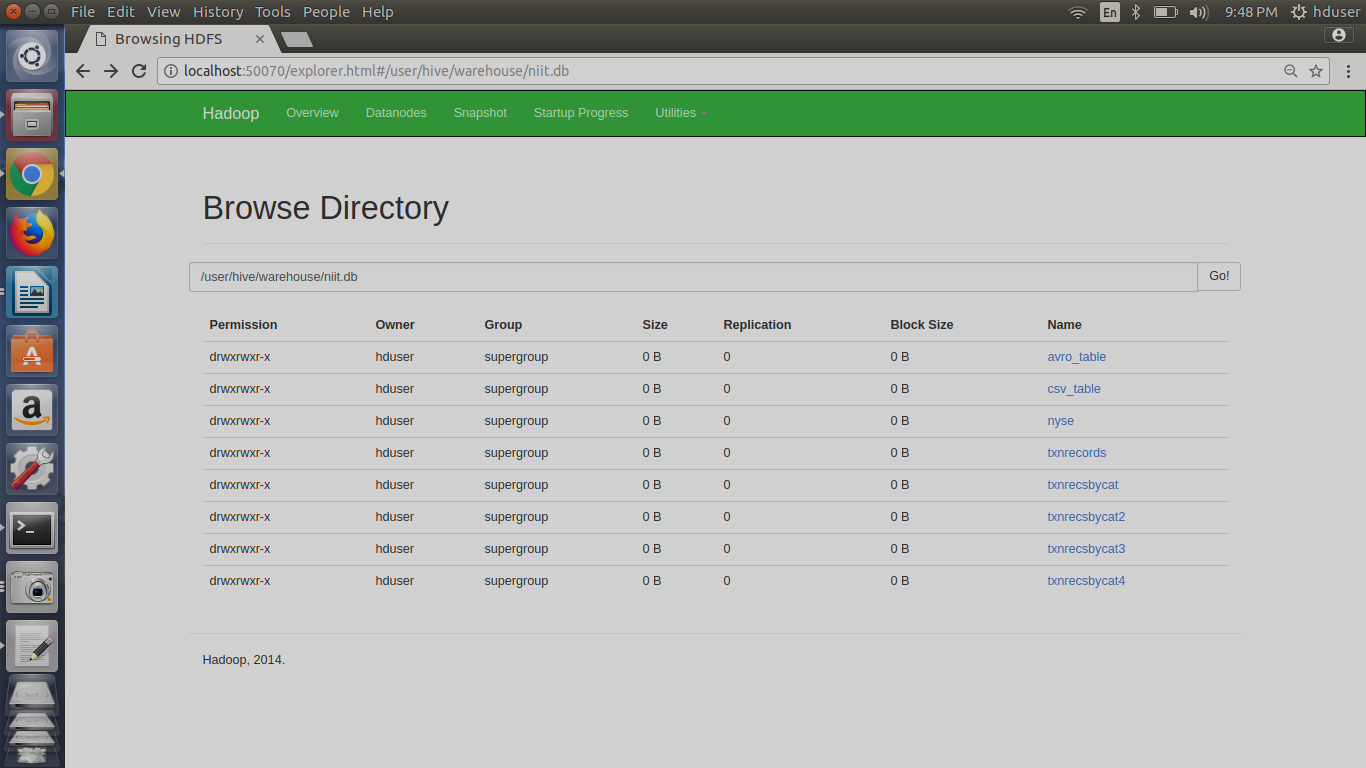
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.25 sec HDFS Read: 4427505 HDFS Write: 3165457 SUCCESS

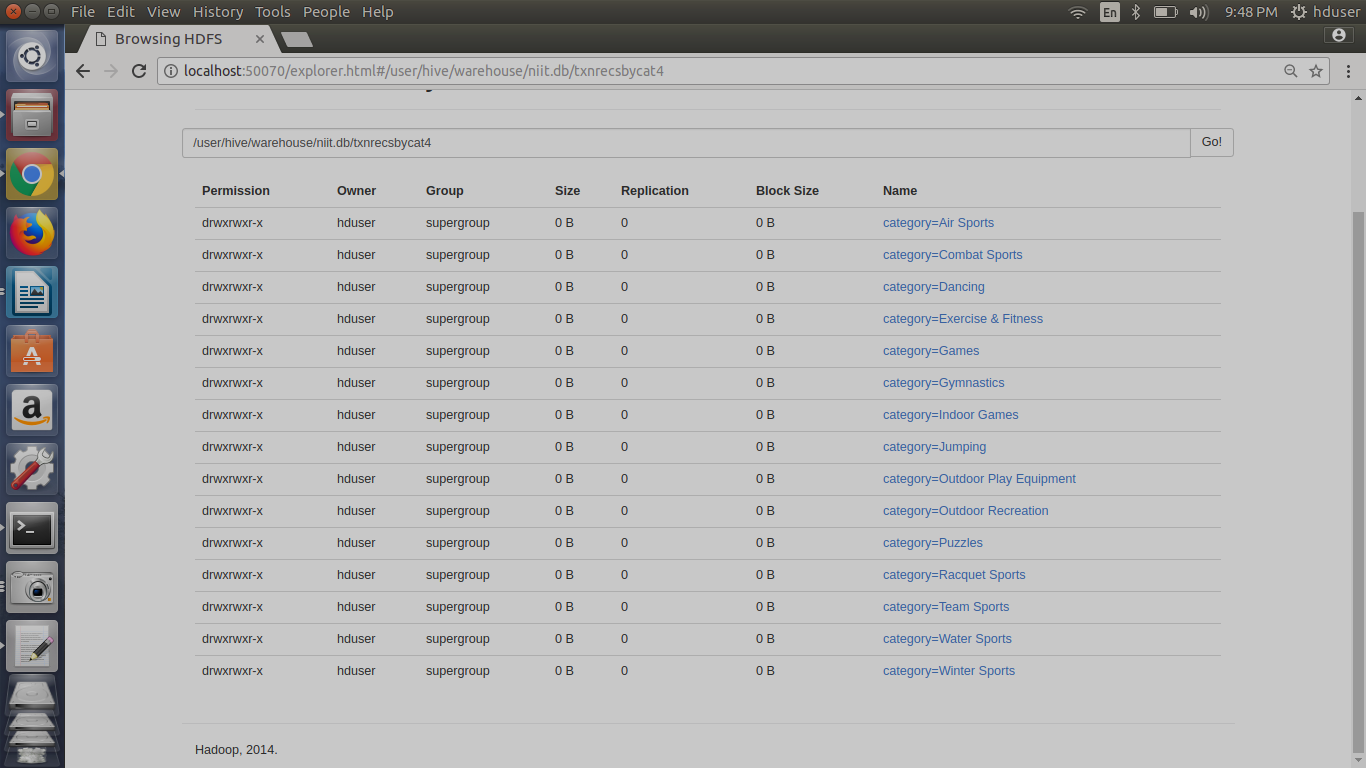
Total MapReduce CPU Time Spent: 6 seconds 250 msec

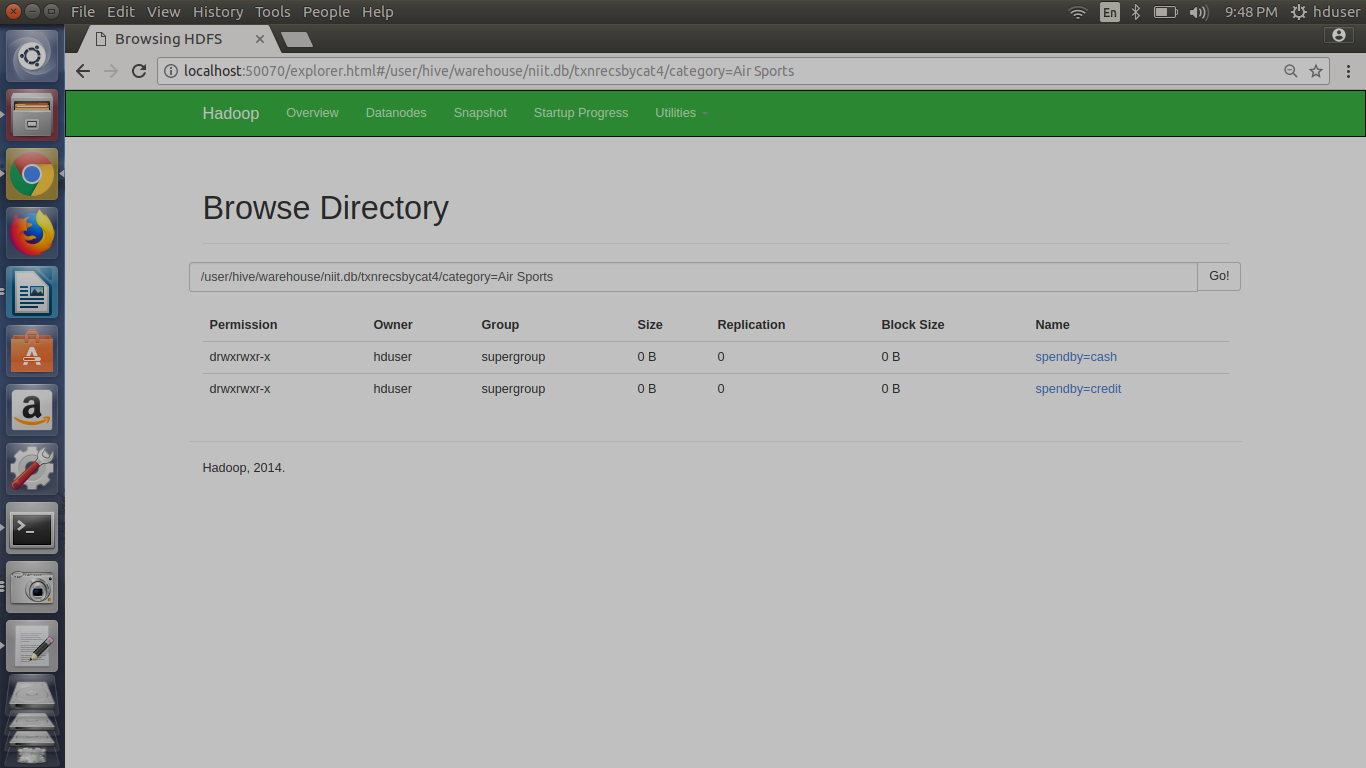
OK

txn.txnno txn.txndate txn.custno txn.amount txn.product txn.city txn.state txn.category txn.spendby

Time taken: 18.574 seconds







hive (niit)> create table txnrecsByCat5(txnno INT, txndate STRING, custno INT, amount DOUBLE,

> product STRING, city STRING, state STRING, spendby STRING)

> partitioned by (category STRING)

> row format delimited

> fields terminated by ','

> stored as textfile;

OK

Time taken: 0.046 seconds

static partitioning

hive (niit)> from txnrecords txn INSERT OVERWRITE TABLE txnrecsByCat5 PARTITION(category='Gymnastics') select txn.txnno, txn.txndate,txn.custno, txn.amount,txn.product,txn.city,txn.state, txn.spendby where txn.category='Gymnastics';

Query ID = hduser\_20171223214955\_3ba1f0a0-0e66-4717-b8da-fa8a9627dc06

Total jobs = 3

Launching Job 1 out of 3

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_1514000082488\_0035, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0035/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0035

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2017-12-23 21:49:59,235 Stage-1 map = 0%, reduce = 0%

2017-12-23 21:50:04,374 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.84 sec

MapReduce Total cumulative CPU time: 2 seconds 840 msec

Ended Job = job\_1514000082488\_0035

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to: hdfs://localhost:54310/user/hive/warehouse/niit.db/txnrecsbycat5/category=Gymnastics/.hive-staging\_hive\_2017-12-23\_21-49-55\_765\_646719793743637648-1/-ext-10000

Loading data to table niit.txnrecsbycat5 partition (category=Gymnastics)

Partition niit.txnrecsbycat5{category=Gymnastics} stats: [numFiles=1, numRows=3196, totalSize=234784, rawDataSize=231588]

MapReduce Jobs Launched:

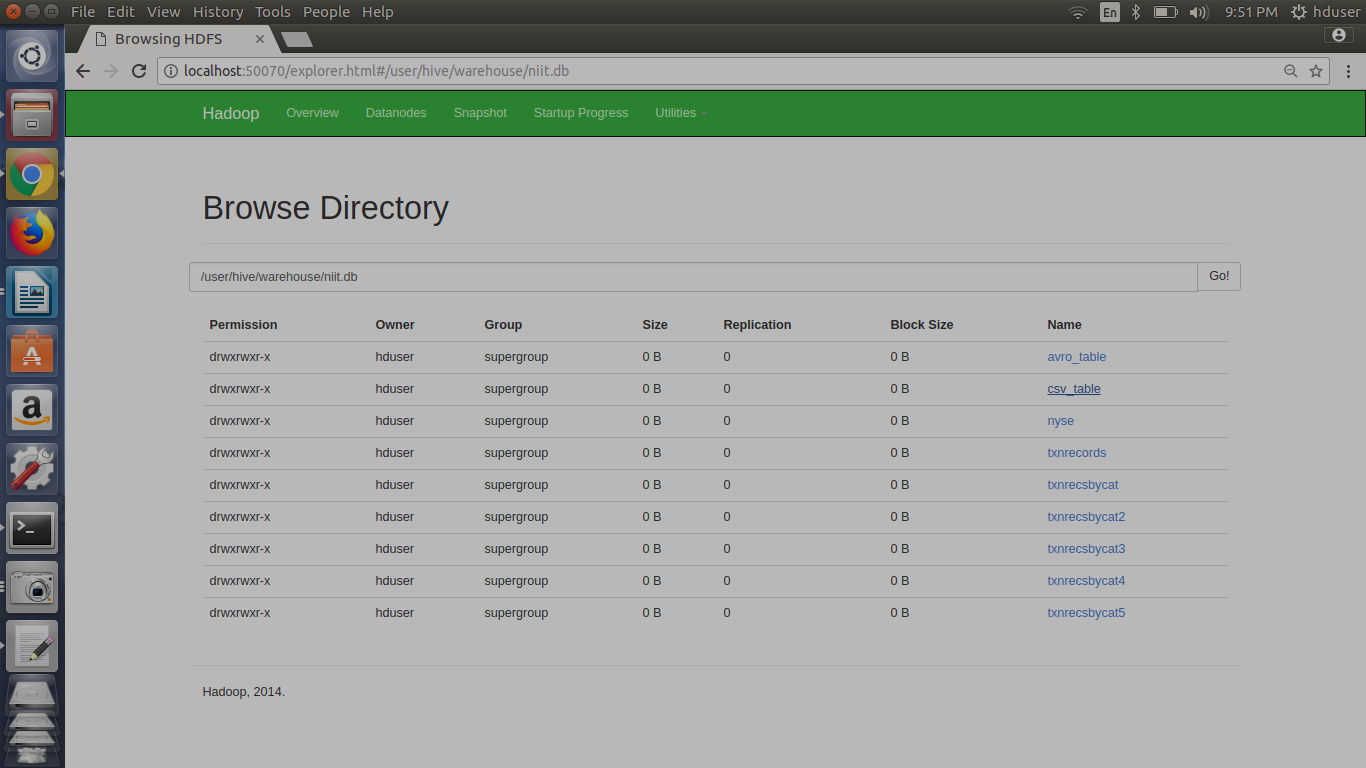
Stage-Stage-1: Map: 1 Cumulative CPU: 2.84 sec HDFS Read: 4423285 HDFS Write: 234884 SUCCESS

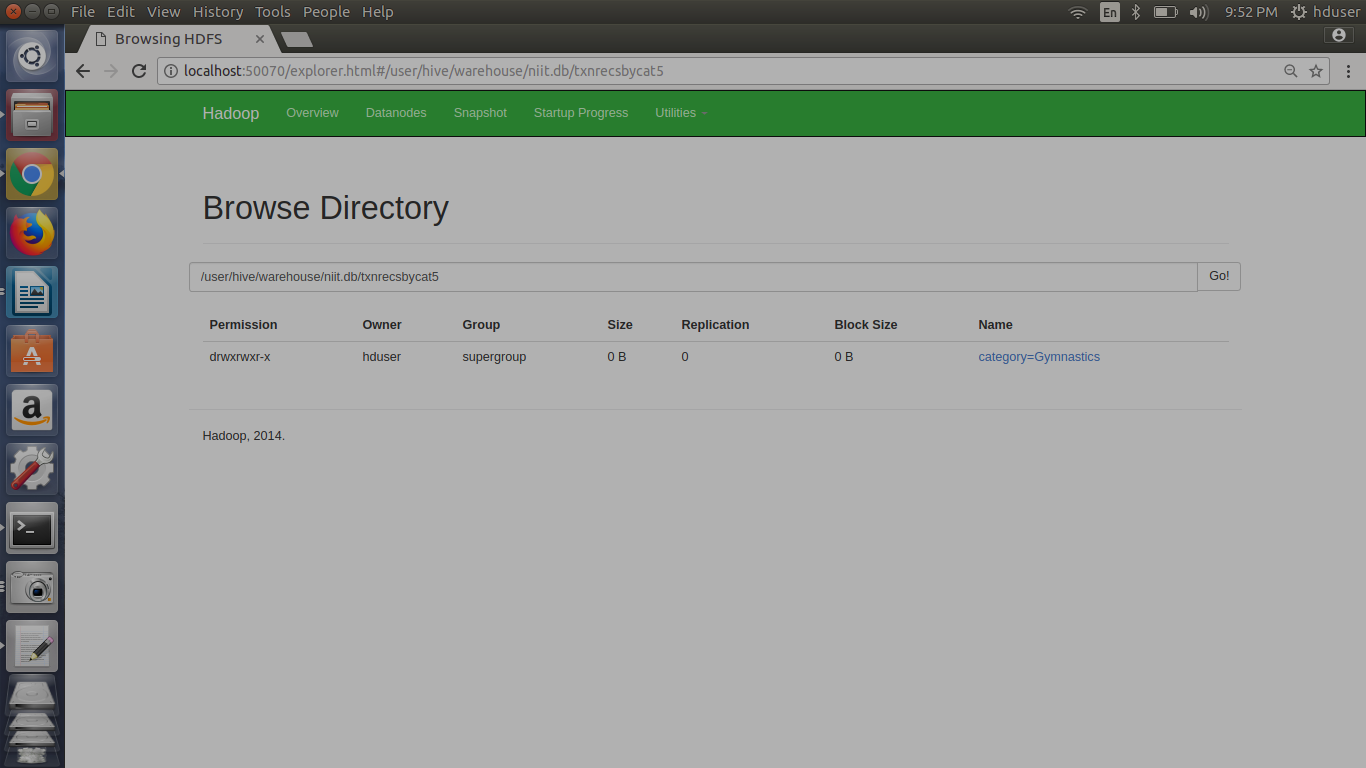
Total MapReduce CPU Time Spent: 2 seconds 840 msec

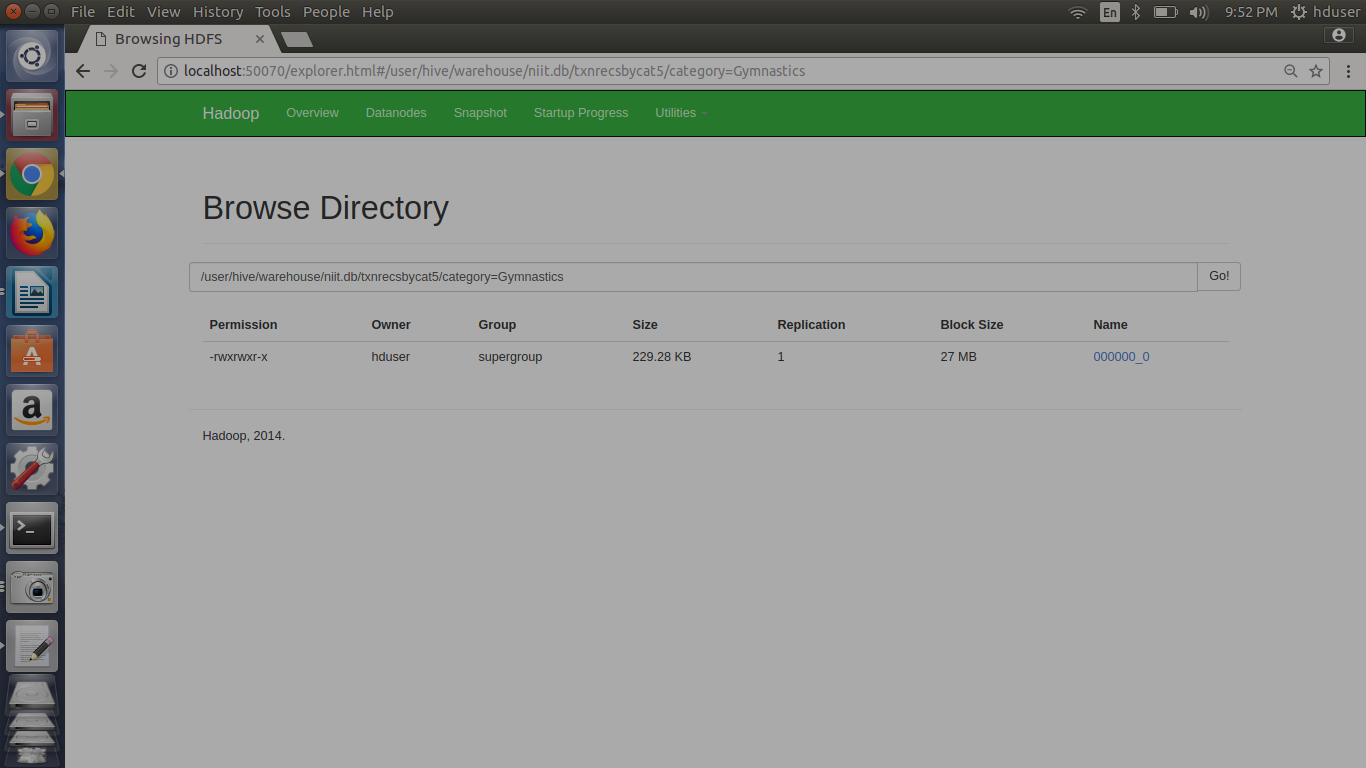
OK

txn.txnno txn.txndate txn.custno txn.amount txn.product txn.city txn.state txn.spendby

Time taken: 9.794 seconds







hive (niit)> from txnrecords txn INSERT OVERWRITE TABLE txnrecsByCat5 PARTITION(category='Team Sports') select txn.txnno, txn.txndate,txn.custno, txn.amount,txn.product,txn.city,txn.state, txn.spendby where txn.category='Team Sports';

Query ID = hduser\_20171223215304\_1db6b02a-576a-4c5c-b732-3533beaeac0c

Total jobs = 3

Launching Job 1 out of 3

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_1514000082488\_0036, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0036/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0036

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2017-12-23 21:53:08,580 Stage-1 map = 0%, reduce = 0%

2017-12-23 21:53:12,720 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.97 sec

MapReduce Total cumulative CPU time: 2 seconds 970 msec

Ended Job = job\_1514000082488\_0036

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to: hdfs://localhost:54310/user/hive/warehouse/niit.db/txnrecsbycat5/category=Team Sports/.hive-staging\_hive\_2017-12-23\_21-53-04\_680\_2059400008502098422-1/-ext-10000

Loading data to table niit.txnrecsbycat5 partition (category=Team Sports)

Partition niit.txnrecsbycat5{category=Team Sports} stats: [numFiles=1, numRows=6010, totalSize=405097, rawDataSize=399087]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 2.97 sec HDFS Read: 4423293 HDFS Write: 405198 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 970 msec

OK

txn.txnno txn.txndate txn.custno txn.amount txn.product txn.city txn.state txn.spendby

