INFO7275 Final Project
Student: Dongyue Li
NUID: 001632075

1. Dataset Description

In this project, I used Yelp Challenge Dataset from this website:
https://www.yelp.com/dataset_challenge

There are 5 JSON files, and each file is composed of a single object type, one json-object per-line. The total size of data I used is a single 952M tar file. Here is the detail information about the tar file.

Name	Date Modified	Size	Kind
Dataset_Challenge_Ac...ataset_Agreement.pdf	Jan 13, 2016, 13:09	397 KB	Adobe...cument
Yelp_Dataset_Challenge_Terms_round_8.pdf	Aug 24, 2016, 12:42	102 KB	Adobe...cument
yelp_academic_dataset_business.json	Aug 3, 2016, 12:53	77.3 MB	JSON
yelp_academic_dataset_checkin.json	Aug 3, 2016, 12:53	28 MB	JSON
yelp_academic_dataset_review.json	Aug 3, 2016, 12:54	2.29 GB	JSON
yelp_academic_dataset_tip.json	Aug 3, 2016, 12:54	130.7 MB	JSON
yelp_academic_dataset_user.json	Aug 3, 2016, 12:54	287 MB	JSON

1) Business (77.3M)
2) Check-in(28M)
3) Review(2.29GB)
4) Tip(130.7M)
5) User(287M)

Here is the Data summary:

- 2.7M reviews and 649K tips by 687K users for 86K businesses
- 566K business attributes, e.g., hours, parking availability, ambience.
- Social network of 687K users for a total of 4.2M social edges.
- Aggregated check-ins over time for each of the 86K businesses

I calculated the number of business, users and reviews.

85,539 business
686,556 users
2,685,066 reviews

2. Problem statement and summary

In this project, I implemented 6 design patterns.
1) Summarization patterns (Min, Max, Count, Average)
2) Filtering patterns (Top ten, Distinct)
3) Join patterns (Reduce-side join)
4) Data organization patterns (Partitioning, Binning)
5) Metapatterns(Job chaining)
6) Inverted Index

The highlights of my project are:
1) In addition to run project locally and in pseudo distributed mode, I also put and run project on AWS which Hadoop, Hbase and Hue have been configured
2) Store MapReduce output data in HBase
3) Use Hue to visualize hbase data.
4) Create Graph database using Neo4j

My project including following parts:
Mapper,
Reducer,
TableReducer,
Combiner,
Partitioner,
Custom Writable,
Custom WritableComparator,
Custom WritableComparable

Here are the analysis I performed in my project:
1) Given Review dataset, determine the first and last time a user reviewed and the total number of reviews from that user.
2) Given Business dataset, output the information of the top 20 restaurants in Arizona based on stars.
3) Redo part2 using other techniques and compare the running times.
4) Given a set of Business information and a set of tips information, enrich each business with their tips information.
5) Given Business dataset, bin the business information into four bins based on their Categories such as restaurants, shopping, food, nightlife. Also, create a separate bin for business information which state is Arizona.
6) For each Business Category, to order the Category descendingly by their average stars and their average review count. (Secondary Sort)
7) Find the distinct user in User dataset.

8) Create Inverted Index for Business Categories.
9) Use Neo4j to create graph database about Category, Stars and Count.

3. Detailed Description of each analysis

1) Given Review dataset, determine the first and last time a user reviewed and the total number of reviews from that user.

---teJGnwK07UO6_oJfbRw

cf: first time reviewed	cf: last time reviewed	cf: total number of reviews
2016-02-13	2016-02-13	1

--0HEXd4W6bJI8k7E0RxTA

cf: first time reviewed	cf: last time reviewed	cf: total number of reviews
2013-07-13	2013-07-13	2

--0KsjlAThNWua2Pr4HStQ

cf: first time reviewed	cf: last time reviewed	cf: total number of reviews
2014-07-19	2014-12-06	6

2) Given Business dataset, output the information of the top 20 restaurants in Arizona based on stars.

-zzS9n4EaKMBlHUG6784qw

cf: Business Name	cf: Full Address	cf: Stars
My Slice of the Pie Pizzeria	3724 E Indian SchoolPhoenix, AZ 85018	5.0

0D5T1XBQBlWnsr57GBTU7A

cf: Business Name	cf: Full Address	cf: Stars
P.A.K Crepes & Coffee	7540 W Indian School Rd.Ste A4Phoenix, AZ 85033	5.0

2aq8b0Rpp60NUr4F6r_Mrw

cf: Business Name	cf: Full Address	cf: Stars
Culver's	1540 N Verrado WayBuckeye, AZ 85396	5.0

AOWv902YtBNzcWMhixVf8Q

cf: Business Name	cf: Full Address	cf: Stars
The Original Marketplace	7272 E Indian School RdSte 105Scottsdale, AZ 85251	5.0

/usr/local/hadoop-2.7.3 — -bash java -Dproc_thrift -XX:....th

Top20Restaurants

```
                FILE: Number of write operations=0
                HDFS: Number of bytes read=4740578088
                HDFS: Number of bytes written=1683342
                HDFS: Number of read operations=68
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=12
        Map-Reduce Framework
                Map input records=9427
                Map output records=20
                Map output bytes=1784
                Map output materialized bytes=1830
                Input split bytes=130
                Combine input records=0
                Combine output records=0
                Reduce input groups=1
                Reduce shuffle bytes=1830
                Reduce input records=20
                Reduce output records=20
                Spilled Records=40
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=53
                Total committed heap usage (bytes)=1055916032
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=841671
        File Output Format Counters
                Bytes Written=0
Top20Restaurants Running time: 8018
```

3) Redo part2 using other techniques and compare the running times.

02rrDia_FTc8jN7LGqzIbQ

cf: Business Name	cf: Full Address	cf: Stars
Corral West Arena	14401 W Arena DrGoodyear, AZ 85338	5.0

6DnA9e4LHVafOyneQ2SjrA

cf: Business Name	cf: Full Address	cf: Stars
El Pacifico Botanas	1615 W Bethany Home RdPhoenix, AZ 85015	5.0

6ZqAwNXbJcC8uV1cJn438A

cf: Business Name	cf: Full Address	cf: Stars
The Salsa Lady	5618 E Skyline DrSan Tan Valley, AZ 85140	5.0

81d4adPvMloqVoiPopuhqA

cf: Business Name	cf: Full Address	cf: Stars
Mingo's Louisiana Kitchen	1245 W. BaselinePhoenix, AZ 85041	5.0

```
16/12/13 10:52:27 INFO mapreduce.Job: Counters: 35
        File System Counters
                FILE: Number of bytes read=1788068
                FILE: Number of bytes written=3270075
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=154502644
                HDFS: Number of bytes written=1702
                HDFS: Number of read operations=13
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
        Map-Reduce Framework
                Map input records=85901
                Map output records=9427
                Map output bytes=860540
                Map output materialized bytes=879417
                Input split bytes=149
                Combine input records=0
                Combine output records=0
                Reduce input groups=9
                Reduce shuffle bytes=879417
                Reduce input records=9427
                Reduce output records=20
                Spilled Records=18854
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=45
                Total committed heap usage (bytes)=758120448
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=77251322
        File Output Format Counters
                Bytes Written=1702
```

part-r-00000 (9)

```
JUr34w08NZz2I3GYenPgiQ  Betty's Rooste  10503 W Thunderbird BlvdSte 102Sun City, AZ 85351   5.0
iPY5hB0TSd8rIiC2xHhnFQ  Blimpie Subs & Salads   7650 S McClintock DrSte 112Tempe, AZ 85284  5.0
B44BU83xIGbjT-25hXvPWw  Frenchys Caribbean Dogs E McDowell RdPhoenix, AZ 85004  5.0
y5L1AEfncjbh1BX2hKRBzA  T&J Gyro and Burger 18635 N 35th AveSte 107Phoenix, AZ 85027    5.0
3ue6ZqXYoJKi7Utt_aYj7g  Bacus Bros Hotdogs and Beer 414 S Mill AveSTE #114Tempe, AZ 85281   5.0
GLjXNuPJL6oP9M36ExKoMQ  Simon's Hot Dogs    4280 Drinkwater BlvdSte 200Scottsdale, AZ 85251 5.0
JzKG_Upd49TVOneDmesELw  Monroe's Downtown   3 W Monroe StPhoenix, AZ 85003  5.0
CJwnoQnL1Q8GmV4HQBQZzg  Leoni's Focaccia    7116 E Mercer LnSte 1Scottsdale, AZ 85254   5.0
OfBBUUWZtO1HgWc4BWlCyg  City Square Deli & Grill    22601 N 19th AveSte 121Phoenix, AZ 85027    5.0
81d4adPvMloqVoiPopuhqA  Mingo's Louisiana Kitchen   1245 W. BaselinePhoenix, AZ 85041   5.0
5CWKlH0uyOUF5EZIyewMXg  Chi Chi's Tacos Phoenix, AZ 85013   5.0
kS3J2SwkKF4yX1OS5Wg01w  Papa John's Pizza   25101 N Lake Pleasant PkwyPeoria, AZ 85383  5.0
TXePwn52cUezrgY95AtJPw  Taqueria Castillo   4217 W Camelback RdPhoenix, AZ 85019    5.0
N6sqymeJNMa2TaZ7dzrypw  Skyline Liquor and Deli 9950 E Broadway RdMesa, AZ 85208    5.0
IGKGTqSR3IUvgZLl62R2ZA  Oasis Crossroads Bar & Grill    5030 W McDowell RdSte 39Phoenix, AZ 85035   5.0
e37EkihO4Zayy4e8U0eT7w  Combo Express Mexican Food  4568 E Cactus RdPhoenix, AZ 85032   5.0
o6w5mssU2sGOceaqPWOczQ  Samurai Sam's Teriyaki Grill    920 E University DrSte D103Tempe, AZ 85281  5.0
tcQrOolSjtS5gcVcEbmwoQ  Green Chile Grill   Phoenix, AZ 85027   5.0
E5CDCVDHpNzTPh5F3Pbppg  Fàn Boy at ASU  1290 S Normal AveTempe, AZ 85281    5.0
j5zlUy0uST-V5ae2kF9ttg  Chili Rush  1440 S Country Club Dr.Ste 10Mesa, AZ 85210 5.0
```

/usr/local/hadoop-2.7.3 — -bash java -Dproc_thrift -XX:....th

Top20Restaurants

```
16/12/11 23:06:36 INFO mapreduce.Job: Counters: 35
        File System Counters
                FILE: Number of bytes read=1866781452
                FILE: Number of bytes written=1881405241
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=4896764074
                HDFS: Number of bytes written=3366684
                HDFS: Number of read operations=98
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=24
        Map-Reduce Framework
                Map input records=9427
                Map output records=9427
                Map output bytes=917122
                Map output materialized bytes=936019
                Input split bytes=134
                Combine input records=0
                Combine output records=0
                Reduce input groups=9
                Reduce shuffle bytes=936019
                Reduce input records=9427
                Reduce output records=20
                Spilled Records=18854
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=1061158912
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=841671
        File Output Format Counters
                Bytes Written=0
Top20RestaurantsRedo Running time: 6487
```

Top20Restaurants Running Time	Top20RestaurantsRedo Running Time
8018	6487

In the first Top20Restaurants algorithm I used TreeMap and combiner to get the top 20 restaurants.

In the second Top20RestaurantsRedo algorithm. In the mapper, ouput<value, key>, and in the reducer output<key, value>. So, MapReduce will sort by value and in my reducer, I just used a counter to output the first 20 records.

There is one thing we must pay attention is that: In the default behavior Tree sort will ignore the duplicate values which will not give the correct results. If you use TreeMap to get Top N, and your values for key are duplicated, you need to write your own comparator with TreeMap to include the duplicate values and sort them.

In my opinion, the reason why the first algorithm is a little bit slower is that the use of TreeMap.

4) Given a set of Business information and a set of tips information, enrich each business with their tips information.

--0ZoBTQWQra1FxD4rBWmg

Business: city	Business: name	Business: review_count	Business: stars	Business: state	Tip: date	Tip: likes	Tip: text
Phoenix	Bourquin Residential Group	4	2.0	AZ	2015-10-29	0	Don't ignore the reviews!!! This management company is awful. They rented me a home that was not up to code and I've been trying to get a workin...

--1emggGHgoG6ipd_RMb-g

Business: city	Business: name	Business: review_count	Business: stars	Business: state	Tip: date	Tip: likes	Tip: text
Las Vegas	Sinclair	4	4.0	NV	2014-07-17	0	Beer is cheaper then gas

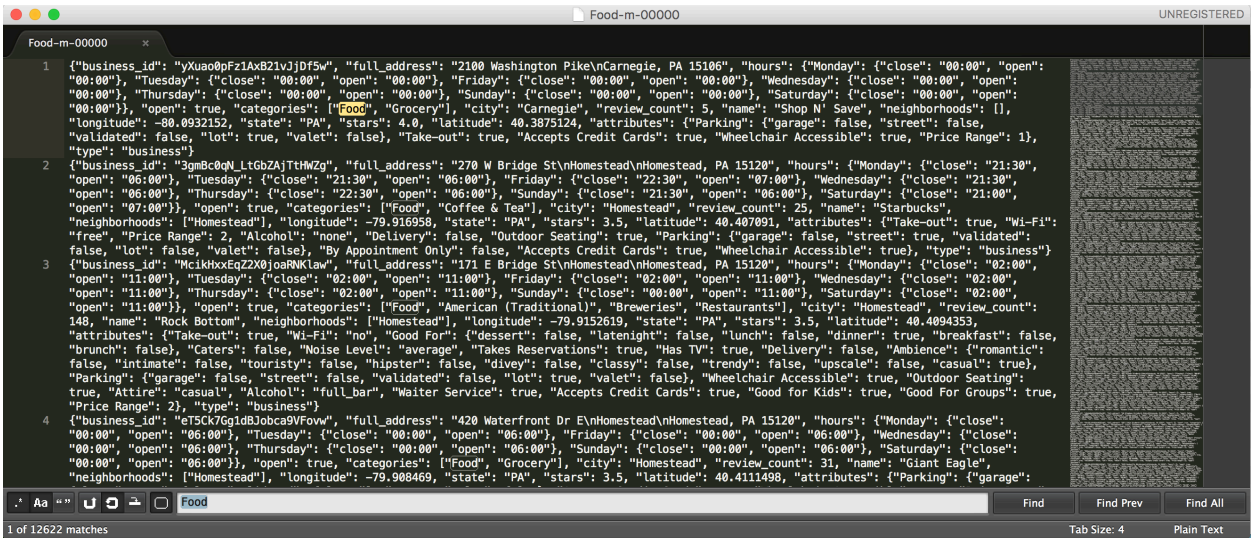
--4Pe8BZ6gj57VFL5mUE8g

Business: city	Business: name	Business: review_count	Business: stars	Business: state	Tip: date	Tip: likes	Tip: text
Phoenix	Office Max	9	3.0	AZ	2016-06-23	0	Items not in stock. :(why advertise on shelf? And on sale!?!??!

5) Given Business dataset, bin the business information into four bins based on their Categories such as restaurants, shopping, food, nightlife. Also, create a separate bin for business information which state is Arizona.

/INFO7275FinalProject/binningOutput Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	ubuntu	supergroup	8.87 MB	12/12/2016, 12:02:45 AM	1	128 MB	Food-m-00000
-rw-r--r--	ubuntu	supergroup	6.88 MB	12/12/2016, 12:02:45 AM	1	128 MB	Nightlife-m-00000
-rw-r--r--	ubuntu	supergroup	32.05 MB	12/12/2016, 12:02:45 AM	1	128 MB	Restaurants-m-00000
-rw-r--r--	ubuntu	supergroup	9.41 MB	12/12/2016, 12:02:45 AM	1	128 MB	Shopping-m-00000
-rw-r--r--	ubuntu	supergroup	2.86 MB	12/12/2016, 12:02:45 AM	1	128 MB	State_AZ-m-00000
-rw-r--r--	ubuntu	supergroup	0 B	12/12/2016, 12:02:45 AM	1	128 MB	_SUCCESS
-rw-r--r--	ubuntu	supergroup	0 B	12/12/2016, 12:02:45 AM	1	128 MB	part-m-00000



6) For each Business Category, to order the Category descendingly by their average stars and their average review count. Finally, partition the results based on average stars' range.

/INFO7275FinalProject/secondarySortCategoryCountAvgStarsOutput Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	dongyueli	supergroup	0 B	12/11/2016, 11:07:23 PM	1	128 MB	_SUCCESS
-rw-r--r--	dongyueli	supergroup	20 B	12/11/2016, 11:07:22 PM	1	128 MB	part-r-00000
-rw-r--r--	dongyueli	supergroup	29 B	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00001
-rw-r--r--	dongyueli	supergroup	296 B	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00002
-rw-r--r--	dongyueli	supergroup	620 B	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00003
-rw-r--r--	dongyueli	supergroup	2.31 KB	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00004
-rw-r--r--	dongyueli	supergroup	6.63 KB	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00005
-rw-r--r--	dongyueli	supergroup	7 KB	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00006
-rw-r--r--	dongyueli	supergroup	3.52 KB	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00007
-rw-r--r--	dongyueli	supergroup	1010 B	12/11/2016, 11:07:23 PM	1	128 MB	part-r-00008

part-r-00008

```
5.0 42  Photo Booth Rentals
5.0 5   IV Hydration
5.0 4   Real Estate Photography
5.0 4   Party Bike Rentals
5.0 3   Bookkeepers
5.0 3   Placenta Encapsulations
5.0 3   House Sitters
5.0 3   Party Characters
5.0 3   Rolfing
5.0 2   Sports Psychologists
5.0 2   Contract Law
5.0 2   Process Servers
5.0 2   Game Truck Rental
5.0 2   Auto Security
5.0 2   Ayurveda
5.0 2   Childproofing
5.0 2   Feng Shui
5.0 2   Home Energy Auditors
5.0 2   Public Relations
5.0 1   3D Printing
5.0 1   Town Car Service
5.0 1   Cannabis Tours
5.0 1   Community Gardens
5.0 1   Awnings
5.0 1   DIY Auto Shop
5.0 1   Day Camps
5.0 1   Solar Panel Cleaning
5.0 1   Editorial Services
5.0 1   Ethic Grocery
5.0 1   Art Tours
5.0 1   Art Restoration
5.0 1   Flyboarding
5.0 1   Rodeo
```

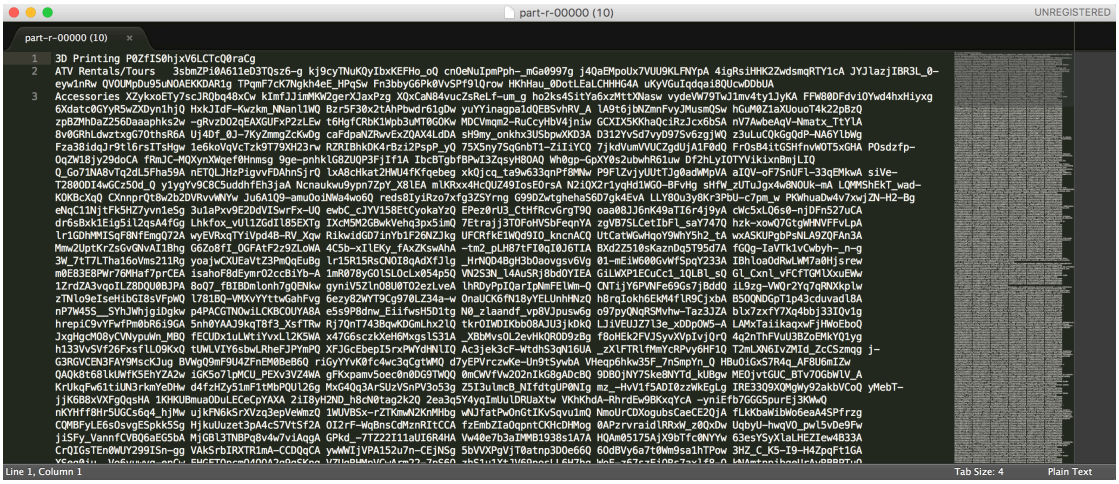
Line 1, Column 1

7) Find the distinct user in User dataset.

part-r-00000 (2)

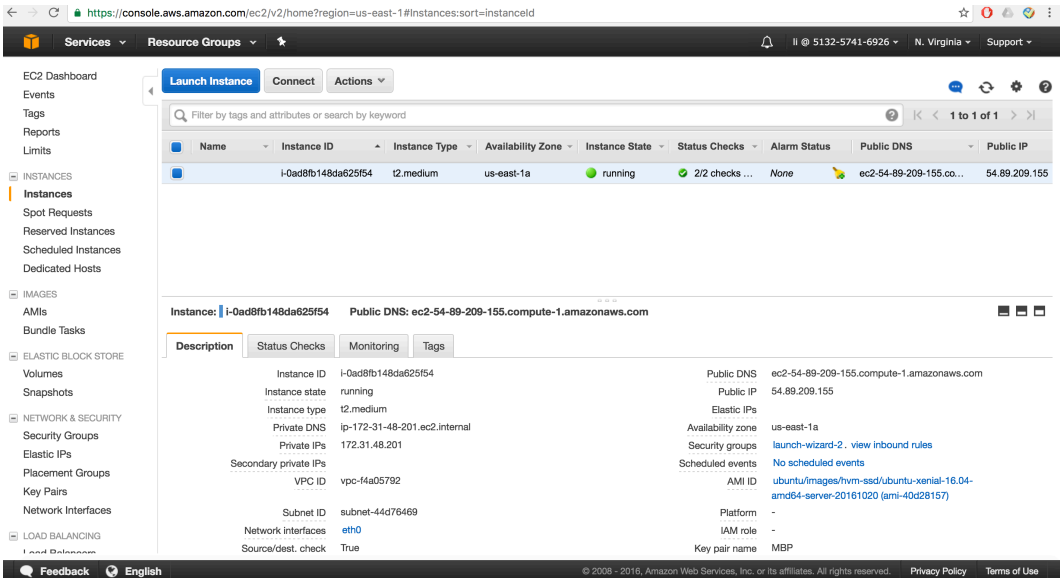
```
686556
```

8) Create Inverted Index for Business Categories.



4. Highlights of my project

AWS



HDFS

ec2-54-89-209-155.compute-1.amazonaws.com:50070/explorer.html#/

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/ Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:02:52 AM	0	0 B	INFO7275FinalProject
drwxr-xr-x	ubuntu	supergroup	0 B	12/11/2016, 11:47:46 PM	0	0 B	hbase
drwxr-xr-x	ubuntu	supergroup	0 B	12/11/2016, 4:00:00 PM	0	0 B	user

Hadoop, 2016.

ec2-54-89-209-155.compute-1.amazonaws.com:50070/explorer.html#/INFO7275FinalProject

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/INFO7275FinalProject Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:02:45 AM	0	0 B	binningOutput
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:02:49 AM	0	0 B	categoryCountAvgStarsOutput
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:03:02 AM	0	0 B	distinctUserOutput
drwxr-xr-x	ubuntu	supergroup	0 B	12/11/2016, 11:56:35 PM	0	0 B	input
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:02:52 AM	0	0 B	secondarySortCategoryCountAvgStarsOutput
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:01:43 AM	0	0 B	topOutput
drwxr-xr-x	ubuntu	supergroup	0 B	12/12/2016, 12:01:49 AM	0	0 B	topRedoOutput

Hbase

ec2-54-89-209-155.compute-1.amazonaws.com:16010/master-status

APACHE HBASE Home Table Details Local Logs Log Level Debug Dump Metrics Dump HBase Configuration

Tables

User Tables System Tables Snapshots

4 table(s) in set. [Details]

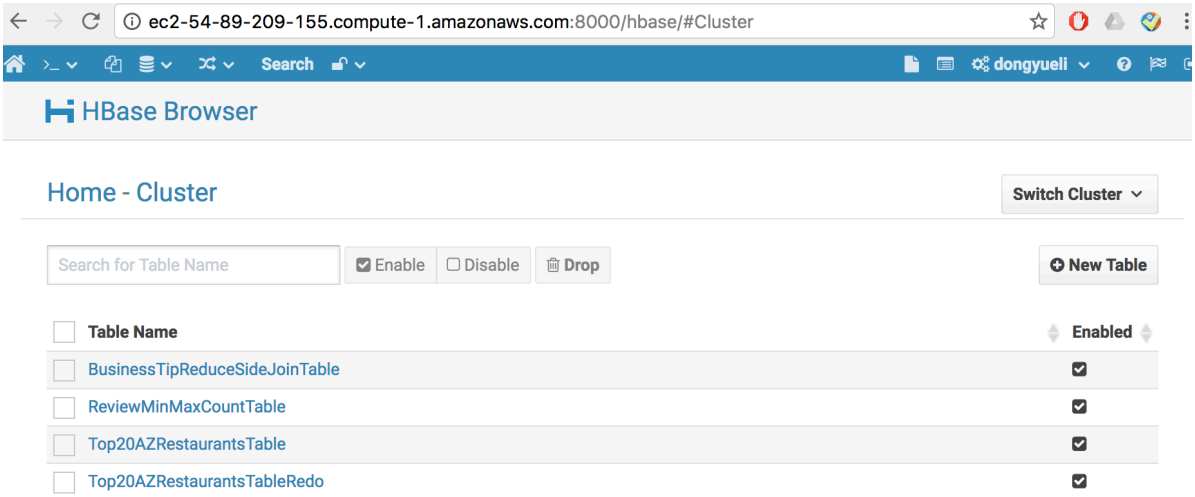
Namespace	Table Name	Online Regions	Offline Regions	Failed Regions	Split Regions	Other Regions	Description
default	BusinessTipReduceSideJoinTable	1	0	0	0	0	'BusinessTipReduceSideJoinTable', {NAME => 'Business'}, {NAME => 'Tip'}
default	ReviewMinMaxCountTable	1	0	0	0	0	'ReviewMinMaxCountTable', {NAME => 'cf'}
default	Top20AZRestaurantsTable	1	0	0	0	0	'Top20AZRestaurantsTable', {NAME => 'cf'}
default	Top20AZRestaurantsTableRedo	1	0	0	0	0	'Top20AZRestaurantsTableRedo', {NAME => 'cf'}

```
MacBook-Pro:hbase-1.2.4 dongyueli$ bin/hbase shell
2016-12-12 21:36:46,108 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for yo
le
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.4, r67592f3d062743907f8c5ae00dbbe1ae4f69e5af, Tue Oct 25 18:10:20 CDT 2016

hbase(main):001:0> list
TABLE
BusinessTipReduceSideJoinTable
ReviewMinMaxCountTable
Top20AZRestaurantsTable
Top20AZRestaurantsTableRedo
4 row(s) in 0.1670 seconds
```

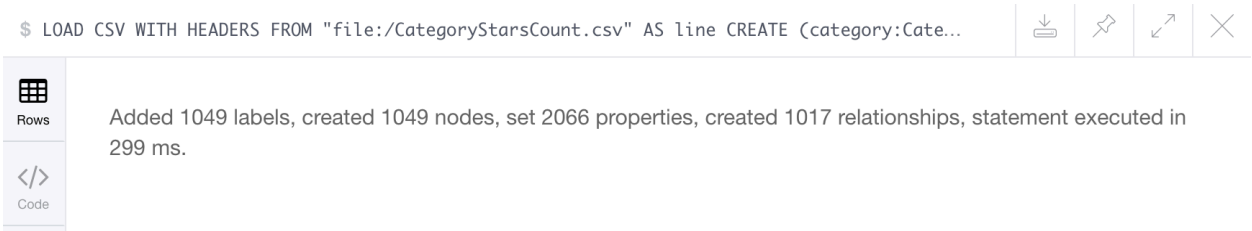
```
=> ["BusinessTipReduceSideJoinTable", "ReviewMinMaxCountTable", "Top20AZRestaurantsTable", "Top20AZRestaurantsTableRedo"]
hbase(main):002:0> scan 'BusinessTipReduceSideJoinTable
```

```
hbase-1.2.4 — java -Dproc_shell -XX:OnOutOfMemoryError=kill -9 %p -XX:+UseConcMarkSweepGC -Dhbase.ruby.sources=/usr/local/hbase-1.2.4/bin/../lib/ruby -Dhbase.log.dir=/usr/lo...
/usr/local/hadoop-2.7.3 — -bash | java -Dproc_shell -XX:OnOutOf...cal/hbase-1.2.4/bin/../bin/hirb.rb | /usr/local/hbase-1.2.4 — -bash
zzlpdQ2GPcXj7mo0lut3Vg                  column=Tip:likes, timestamp=1481515634451, value=
zzlpdQ2GPcXj7mo0lut3Vg                  column=Tip:text, timestamp=1481515634451, value=
zzokIU5MGrt-l_U-ZzZiug                  column=Business:city, timestamp=1481515634451, value=Montr\xC3\xA9al
zzokIU5MGrt-l_U-ZzZiug                  column=Business:name, timestamp=1481515634451, value=LaSalle College
zzokIU5MGrt-l_U-ZzZiug                  column=Business:review_count, timestamp=1481515634451, value=5
zzokIU5MGrt-l_U-ZzZiug                  column=Business:stars, timestamp=1481515634451, value=1.0
zzokIU5MGrt-l_U-ZzZiug                  column=Business:state, timestamp=1481515634451, value=QC
zzokIU5MGrt-l_U-ZzZiug                  column=Tip:date, timestamp=1481515634451, value=2016-01-26
zzokIU5MGrt-l_U-ZzZiug                  column=Tip:likes, timestamp=1481515634451, value=0
zzokIU5MGrt-l_U-ZzZiug                  column=Tip:text, timestamp=1481515634451, value=Shittty school
zzrm5HEoYKEsfdi8XxSXuQ                  column=Business:city, timestamp=1481515634451, value=Las Vegas
zzrm5HEoYKEsfdi8XxSXuQ                  column=Business:name, timestamp=1481515634451, value=Kaplan Robert P, DO
zzrm5HEoYKEsfdi8XxSXuQ                  column=Business:review_count, timestamp=1481515634451, value=10
zzrm5HEoYKEsfdi8XxSXuQ                  column=Business:stars, timestamp=1481515634451, value=2.0
zzrm5HEoYKEsfdi8XxSXuQ                  column=Business:state, timestamp=1481515634451, value=NV
zzrm5HEoYKEsfdi8XxSXuQ                  column=Tip:date, timestamp=1481515634451, value=2015-02-27
zzrm5HEoYKEsfdi8XxSXuQ                  column=Tip:likes, timestamp=1481515634451, value=0
zzrm5HEoYKEsfdi8XxSXuQ                  column=Tip:text, timestamp=1481515634451, value=Not going anywhere for a while... grab a Snickers. You wait almost as long
                                        here as you do going to a Concentra on workers comp issues but the doctor is frendly and knowledgeable and I  would say wor
                                        th the wait.
zztd3r04hvMTZh3ol7VC4g                  column=Business:city, timestamp=1481515634451, value=Montr\xC3\xA9al
zztd3r04hvMTZh3ol7VC4g                  column=Business:name, timestamp=1481515634451, value=Br\xC3\xBBlerie aux quatre vents
zztd3r04hvMTZh3ol7VC4g                  column=Business:review_count, timestamp=1481515634451, value=8
zztd3r04hvMTZh3ol7VC4g                  column=Business:stars, timestamp=1481515634451, value=4.0
zztd3r04hvMTZh3ol7VC4g                  column=Business:state, timestamp=1481515634451, value=QC
zztd3r04hvMTZh3ol7VC4g                  column=Tip:date, timestamp=1481515634451, value=2015-02-23
zztd3r04hvMTZh3ol7VC4g                  column=Tip:likes, timestamp=1481515634451, value=0
zztd3r04hvMTZh3ol7VC4g                  column=Tip:text, timestamp=1481515634451, value=One of my favorite coffee stops
zzucnDH33KlmbkI4rLyQTA                  column=Business:city, timestamp=1481515634451, value=Phoenix
zzucnDH33KlmbkI4rLyQTA                  column=Business:name, timestamp=1481515634451, value=Ski Pro
zzucnDH33KlmbkI4rLyQTA                  column=Business:review_count, timestamp=1481515634451, value=30
zzucnDH33KlmbkI4rLyQTA                  column=Business:stars, timestamp=1481515634451, value=4.0
zzucnDH33KlmbkI4rLyQTA                  column=Business:state, timestamp=1481515634451, value=AZ
zzucnDH33KlmbkI4rLyQTA                  column=Tip:date, timestamp=1481515634451, value=2011-01-07
zzucnDH33KlmbkI4rLyQTA                  column=Tip:likes, timestamp=1481515634451, value=0
zzucnDH33KlmbkI4rLyQTA                  column=Tip:text, timestamp=1481515634451, value=Extra 10% off all winter sale items through Sunday!
zzxb0YNkRvH1UgifuR3BvQ                  column=Business:city, timestamp=1481515634451, value=Phoenix
zzxb0YNkRvH1UgifuR3BvQ                  column=Business:name, timestamp=1481515634451, value=Sonora Quest Laboratories
zzxb0YNkRvH1UgifuR3BvQ                  column=Business:review_count, timestamp=1481515634451, value=3
zzxb0YNkRvH1UgifuR3BvQ                  column=Business:stars, timestamp=1481515634451, value=5.0
zzxb0YNkRvH1UgifuR3BvQ                  column=Business:state, timestamp=1481515634451, value=AZ
zzxb0YNkRvH1UgifuR3BvQ                  column=Tip:date, timestamp=1481515634451, value=
zzxb0YNkRvH1UgifuR3BvQ                  column=Tip:likes, timestamp=1481515634451, value=
zzxb0YNkRvH1UgifuR3BvQ                  column=Tip:text, timestamp=1481515634451, value=
85901 row(s) in 44.5490 seconds

hbase(main):003:0>
```

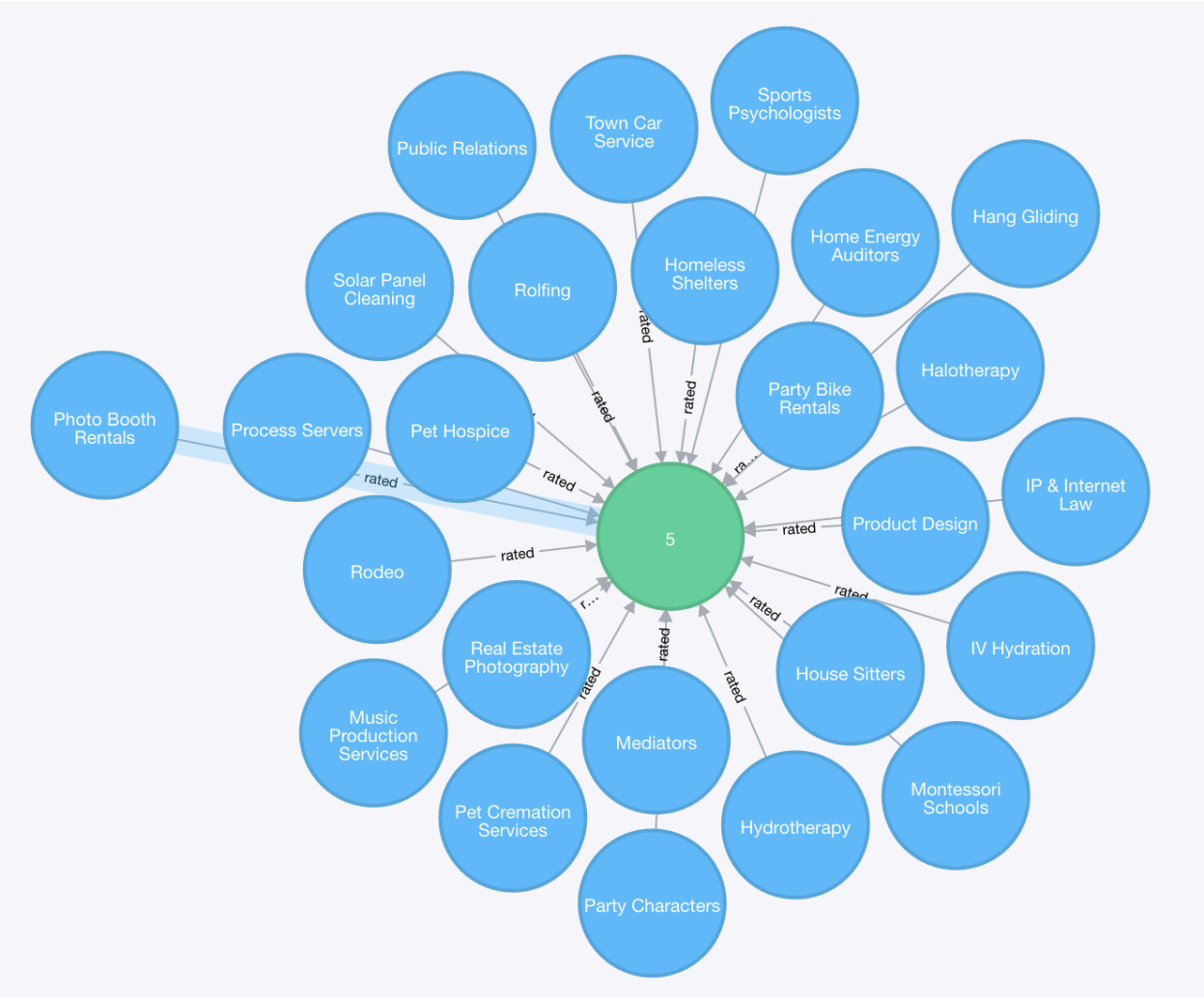
Hue


ec2-54-89-209-155.compute-1.amazonaws.com:8000/hbase/#Cluster
Search
dongyueli
HBase Browser
Home - Cluster
Switch Cluster
Search for Table Name
Enable
Disable
Drop
New Table
Table Name
Enabled
BusinessTipReduceSideJoinTable
ReviewMinMaxCountTable
Top20AZRestaurantsTable
Top20AZRestaurantsTableRedo

Neo4j


$ LOAD CSV WITH HEADERS FROM "file:/CategoryStarsCount.csv" AS line CREATE (category:Cate...
Rows
Code
Added 1049 labels, created 1049 nodes, set 2066 properties, created 1017 relationships, statement executed in 299 ms.

$ MATCH (n:Category) RETURN n LIMIT 25
*(25)
Category(25)
Graph
Rows
Text
Code
Engraving
Home Network Installation
Caricatures
Trophy Shops
Talent Agencies
Face Painting
Private Investigation
Olive Oil
Wine Tours
Color:
Size:
Caption:
<id>

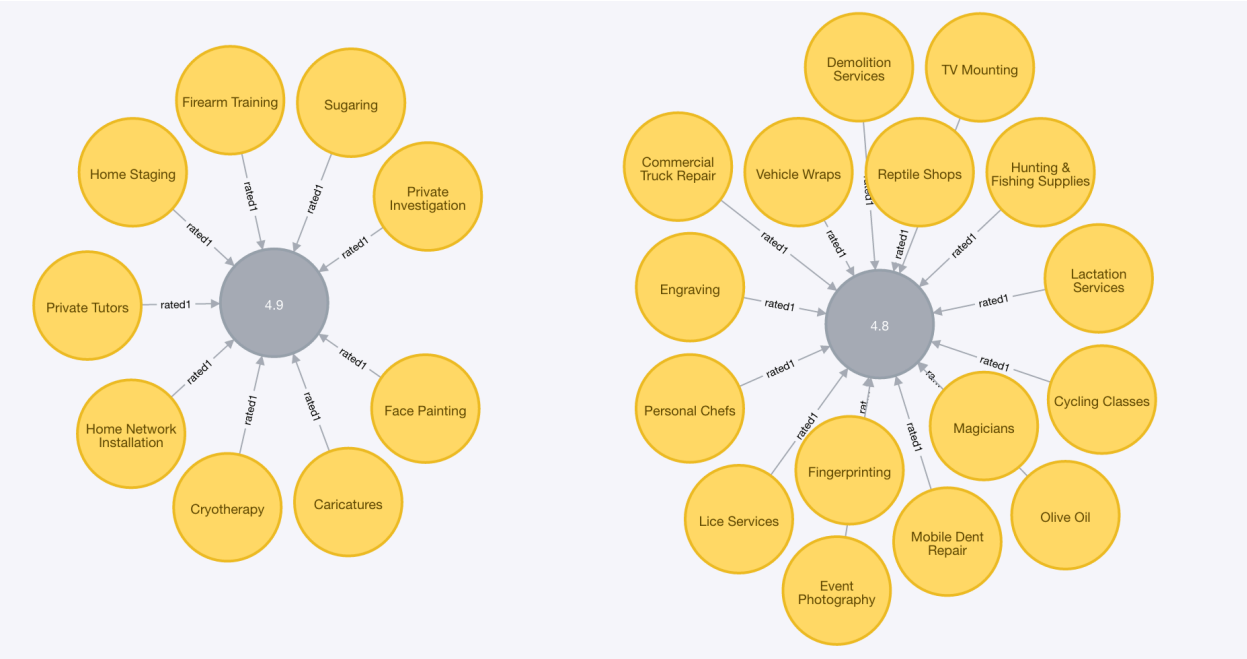



```
$ LOAD CSV WITH HEADERS FROM "file:/CategoryStarsCountPart.csv" AS line CREATE (category:...
```

Rows

Code

Added 168 labels, created 168 nodes, set 331 properties, created 163 relationships, statement executed in 156 ms.


Firearm Training
Sugaring
Home Staging
Private Investigation
Private Tutors
4.9
Home Network Installation
Face Painting
Cryotherapy
Caricatures
rated1
Demolition Services
TV Mounting
Commercial Truck Repair
Vehicle Wraps
Reptile Shops
Hunting & Fishing Supplies
Engraving
4.8
Lactation Services
Personal Chefs
Cycling Classes
Magicians
Fingerprinting
Lice Services
Olive Oil
Mobile Dent Repair
Event Photography