

Στοιχεία Μελών Ομάδας:

Παπαδόπουλος Λάμπρος	AM:1054433	Έτος Σπουδών: 4 ^ο
Κωστορρίζος Δημήτριος	AM:1054419	Έτος Σπουδών: 4 ^ο

Εντολές για την σύνδεση στην HBASE και την δημιουργία των ζητούμενων Table

Αρχικά, συνδεθήκαμε, με την χρήση του OpenVPN GUI, στο σύστημα Okeanos χρησιμοποιώντας το config αρχείο που μας έστειλε, στην προσωπική συνομιλία στο Slack, ο κύριος Γεώργιος Σεργιάννης. Έπειτα, συνδεθήκαμε στο Apache Hadoop Cluster, στο οποίο υπήρχαν η μη σχεσιακή βάση δεδομένων HBASE και το εργαλείο σύνταξης SQL εντολών, sqlline.py. Με την βοήθεια του Apache Phoenix, οι σχεσιακές SQL εντολές μετατρέπονται σε μη σχεσιακές εντολές, που μπορεί να αναγνωρίσει η HBASE.

Καλώντας την εντολή: hbase shell μεταβήκαμε από το Apache Hadoop shell, στο hbase shell. Δημιουργήσαμε τα δύο άδεια table 'USER12.YELPBUSINESS', 'USER12.YELPCHECKIN' με τα αντίστοιχα column families, με τις εντολές παρακάτω και επιβεβαιώσαμε την δημιουργία τους με την εντολή: list.

Για το Table 'USER12.YELPBUSINESS': create 'USER12.YELPBUSINESS', 'BASE', 'ATTRIBUTES', 'HOURS'

Για το Table 'USER12.YELPCHECKIN': create 'USER12.YELPCHECKIN', 'PERHOUR'

Εντολές για την εισαγωγή δεδομένων στην HBASE

Αφού δημιουργήσαμε επιτυχώς τα Table, χρησιμοποιήσαμε το εργαλείο ImportTsv από το Apache Hadoop command line, με σκοπό να εισάγουμε τα περιεχόμενα των αρχείων .csv, που περιέχονται στον φάκελο dataset, στο κατάλληλο Table και Column Family. Παρακάτω παρουσιάζονται οι εντολές αυτές. Ως HBASE_ROW_KEY έχει ορισθεί η στήλη business_id του αρχείου και στα δύο Table. Για τα ονόματα των στηλών των Table έχει χρησιμοποιηθεί <<το όνομα του Column Family.(τελεία)>> το όνομα της στήλης στο αρχείο>> αφού έχουν αφαιρεθεί τυχόν underscore και έχουν μετατραπεί όλα τα γράμματα σε κεφαλαία. Ως column separator έχει ορισθεί ο separator του csv αρχείου, δηλαδή ,(κόμμα).

Εντολή για την εισαγωγή των δεδομένων, του yelp_business.csv, στο Column Family BASE του Table 'USER12.YELPBUSINESS':

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=,' -  
Dimporttsv.columns=HBASE_ROW_KEY,BASE:NAME,BASE:NEIGHBORHOOD,BASE:ADDRESS,BASE:CITY,B  
ASE:STATE,BASE:POSTALCODE,BASE:LATITUDE,BASE:LONGTITUDE,BASE:STARS,BASE:REVIEWCOUNT,BAS  
E:ISOPEN,BASE:CATEGORIES USER12.YELPBUSINESS ../hbase/dataset/yelp_business.csv
```

**Εντολή για την εισαγωγή των δεδομένων, του yelp_business_attributes.csv, στο Column Family
ATTRIBUTES του Table USER12.YELPBUSINESS:**

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=,' -  
Dimporttsv.columns=HBASE_ROW_KEY,ATTRIBUTES:ACCEPTSINSURANCE,ATTRIBUTES:BYAPPOINTMENT  
ONLY,ATTRIBUTES:BUSINESSACCEPTSCREDITCARDS,ATTRIBUTES:BUSINESSPARKINGGARAGE,ATTRIBUTE  
S:BUSINESSPARKINGSTREET,ATTRIBUTES:BUSINESSPARKINGVALIDATED,ATTRIBUTES:BUSINESSPARKING  
LOT,ATTRIBUTES:BUSINESSPARKINGVALET,ATTRIBUTES:HAIRSPECIALIZESINCOLORING,ATTRIBUTES:HAIR  
SPECIALIZESINAFRICANAMERICAN,ATTRIBUTES:HAIRSPECIALIZESINCURLY,ATTRIBUTES:HAIRSPECIALIZESI  
NPERMS,ATTRIBUTES:HAIRSPECIALIZESINKIDS,ATTRIBUTES:HAIRSPECIALIZESINEXTENSIONS,ATTRIBUTES:  
HAIRSPECIALIZESINASIAN,ATTRIBUTES:HAIRSPECIALIZESINSTRAIGHTPERMS,ATTRIBUTES:RESTAURANTSP  
RICERANGETWO,ATTRIBUTES:GOODFORKIDS,ATTRIBUTES:WHEELCHAIRACCESSIBLE,ATTRIBUTES:BIKEPA  
RKING,ATTRIBUTES:ALCOHOL,ATTRIBUTES:HASTV,ATTRIBUTES:NOISELEVEL,ATTRIBUTES:RESTAURANTS  
ATTIRE,ATTRIBUTES:MUSICDJ,ATTRIBUTES:MUSICBACKGROUNDMUSIC,ATTRIBUTES:MUSICNOMUSIC,ATT  
RIBUTES:MUSICKARAOKE,ATTRIBUTES:MUSICLIVE,ATTRIBUTES:MUSICVIDEO,ATTRIBUTES:MUSICJUKEBO  
X,ATTRIBUTES:AMBIENCEROMANTIC,ATTRIBUTES:AMBIENCEINTIMATE,ATTRIBUTES:AMBIENCECLASSY,A  
TTRIBUTES:AMBIENCEHIPSTER,ATTRIBUTES:AMBIENCEDIVEY,ATTRIBUTES:AMBIENCETOURISTY,ATTRIBU  
TES:AMBIENCETRENDY,ATTRIBUTES:AMBIENCEUPSCALE,ATTRIBUTES:AMBIENCECASUAL,ATTRIBUTES:RE  
STAURANTSGOODFORGROUPS,ATTRIBUTES:CATERS,ATTRIBUTES:WIFI,ATTRIBUTES:RESTAURANTSRESER  
VATIONS,ATTRIBUTES:RESTAURANTSTAKEOUT,ATTRIBUTES:HAPPYHOUR,ATTRIBUTES:GOODFORDANCI  
NG,ATTRIBUTES:RESTAURANTSTABLESERVICE,ATTRIBUTES:OUTDOORSEATING,ATTRIBUTES:RESTAURAN  
TSDelivery,ATTRIBUTES:BESTNIGHTSMONDAY,ATTRIBUTES:BESTNIGHTSTUESDAY,ATTRIBUTES:BESTNIG  
HTSFRIDAY,ATTRIBUTES:BESTNIGHTSWEDNESDAY,ATTRIBUTES:BESTNIGHTSTHURSDAY,ATTRIBUTES:BE  
SNIGHTSSUNDAY,ATTRIBUTES:BESTNIGHTSSATURDAY,ATTRIBUTES:GOODFORMEALDESSERT,ATTRIBUTE  
S:GOODFORMEALLATENIGHT,ATTRIBUTES:GOODFORMEALLUNCH,ATTRIBUTES:GOODFORMEALDINNER,  
ATTRIBUTES:GOODFORMEALBREAKFAST,ATTRIBUTES:GOODFORMEALBRUNCH,ATTRIBUTES:COATCHECK  
,ATTRIBUTES:SMOKING,ATTRIBUTES:DRIVETHRU,ATTRIBUTES:DOGSALLOWED,ATTRIBUTES:BUSINESSAC  
CEPTSBITCOIN,ATTRIBUTES:OPEN24HOURS,ATTRIBUTES:BYOBCORKAGE,ATTRIBUTES:BYOB,ATTRIBUTES:  
CORKAGE,ATTRIBUTES:DIETARYRESTRICTIONSDAIRYFREE,ATTRIBUTES:DIETARYRESTRICTIONSGLUTENFR  
EE,ATTRIBUTES:DIETARYRESTRICTIONSVEGAN,ATTRIBUTES:DIETARYRESTRICTIONSKOSHER,ATTRIBUTES:  
DIETARYRESTRICTIONSHALAL,ATTRIBUTES:DIETARYRESTRICTIONSISOYFREE,ATTRIBUTES:DIETARYRESTRIC  
TIONSVEGETARIAN,ATTRIBUTES:AGESALLOWED,ATTRIBUTES:RESTAURANTSCOUNTERSERVICE  
USER12.YELPBUSINESS ../hbase/dataset/yelp_business_attributes.csv
```

**Εντολή για την εισαγωγή των δεδομένων, του yelp_business_hours.csv, στο Column Family HOURS
του Table USER12.YELPBUSINESS:**

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=' -  
Dimporttsv.columns=HBASE_ROW_KEY,HOURS:MONDAY,HOURS:TUESDAY,HOURS:WEDNESDAY,HOURS  
:THURSDAY,HOURS:FRIDAY,HOURS:SATURDAY,HOURS:SUNDAY USER12.YELPBUSINESS  
../hbase/dataset/yelp_business_hours.csv
```

Εντολή για την εισαγωγή των δεδομένων, του yelp_checkin.csv, στο Column Family PERHOUR του Table USER12.YELPCHECKIN:

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=' -  
Dimporttsv.columns=PERHOUR:ID,HBASE_ROW_KEY,PERHOUR:WEEKDAY,PERHOUR:HOURL,PERHOUR:C  
HECKIN USER12.YELPCHECKIN ../hbase/dataset/yelp_checkin.csv
```

Τα αρχεία yelp_business.csv, yelp_business_attributes.csv, yelp_checkin.csv και yelp_checkin.csv περιέχουν τα ονόματα των στηλών ως header. Με την χρήση της παρακάτω εντολής διαγράψαμε τις εγγραφές που περιέχουν τα ονόματα των στηλών αυτά.

Διαγραφή της csv column header γραμμής – εγγραφής από τα Table:

3

```
deleteall 'USER12.YELPBUSINESS' , 'business_id'
```

```
deleteall 'USER12.YELPCHECKIN' , 'business_id'
```

Με την χρήση των εντολών:

```
scan 'USER12.YELPBUSINESS' , {FILTER=>"(RowFilter(=,'regexstring:business_id'))"}
```

```
scan 'USER12.YELPCHECKIN' , {FILTER=>"(RowFilter(=,'regexstring:business_id'))"}
```

Επιβεβαιώσαμε ότι οι συγκεκριμένες εγγραφές είχαν διαγραφεί από τα Table, δηλαδή οι εγγραφές που είχαν ως κλειδί το αλφαριθμητικό 'business_id'.

Αφού ολοκληρώσαμε την εισαγωγή των δεδομένων στην HBASE, χρησιμοποιήσαμε την παρακάτω εντολή ώστε να ενεργοποιήσουμε το command line του εργαλείου sqlline.py, προκειμένου να τρέξουμε τις εντολές για την διασύνδεσης της HBASE με το Phoenix, την δημιουργία Index στις στήλες των Table και την αναζήτηση εγγραφών.

Εντολή εκκίνησης του εργαλείου sqlline.py:

/usr/hdp/2.6.3.0-235/phoenix/bin/sqlline.py

Εντολές στο εργαλείο sqlline.py:

Για να διασυνδέσουμε την HBASE με το Phoenix δημιουργήσαμε ένα View για κάθε Table. Ωστόσο, λόγω της δυσκολίας κατανόησης του τρόπου αποθήκευσης(hex byte σε ακέραιους,timestamp,date,time), που εφαρμόζει η HBASE σε δεδομένα που δεν είναι αλφαριθμητικά, αποφασίσαμε ο στήλες των View, να είναι τύπος Varchar. Στις εντολές που χρειαζόνταν την αριθμητική αναπαράσταση των δεδομένων, εφαρμόσαμε safe typecast. Τα ονόματα των στηλών του View για το κάθε Table, είναι τα ίδια με τα ονόματα των στηλών στο HBASE Table. Παρακάτω παρουσιάζονται οι εντολές για την δημιουργία των View. Με την χρήση της εντολής: !tables επιβεβαιώσαμε την επιτυχή δημιουργία των επιθυμητών View.

Create View για το table USER12.YELPBUSINESS:

```
CREATE VIEW USER12.YELPBUSINESS ( ROWKEY VARCHAR PRIMARY KEY,BASE.NAME  
VARCHAR,BASE.NEIGHBORHOOD VARCHAR,BASE.ADDRESS VARCHAR,BASE.CITY VARCHAR,BASE.STATE  
VARCHAR,BASE.POSTALCODE VARCHAR,BASE.LATITUDE VARCHAR,BASE.LONGITUDE  
VARCHAR,BASE.STARS VARCHAR,BASE.REVIEWCOUNT VARCHAR,BASE.ISOPEN  
VARCHAR,BASE.CATEGORIES VARCHAR, ATTRIBUTES.ACCEPTSINSURANCE  
VARCHAR,ATTRIBUTES.BYAPPOINTMENTONLY VARCHAR,ATTRIBUTES.BUSINESSACCEPTSCREDITCARDS  
VARCHAR,ATTRIBUTES.BUSINESSPARKINGGARAGE VARCHAR,ATTRIBUTES.BUSINESSPARKINGSTREET  
VARCHAR,ATTRIBUTES.BUSINESSPARKINGVALIDATED VARCHAR,ATTRIBUTES.BUSINESSPARKINGLOT  
VARCHAR,ATTRIBUTES.BUSINESSPARKINGVALET VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINCOLORING  
VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINAFRICANAMERICAN  
VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINCURLY VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINPERMS  
VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINKIDS VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINEXTENSIONS  
VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINASIAN  
VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINSTRAIGHTPERMS  
VARCHAR,ATTRIBUTES.RESTAURANTSPRICERANGETWO VARCHAR,ATTRIBUTES.GOODFORKIDS  
VARCHAR,ATTRIBUTES.WHEELCHAIRACCESSIBLE VARCHAR,ATTRIBUTES.BIKEPARKING  
VARCHAR,ATTRIBUTES.ALCOHOL VARCHAR,ATTRIBUTES.HASTV VARCHAR,ATTRIBUTES.NOISELEVEL  
VARCHAR,ATTRIBUTES.RESTAURANTSATTIRE VARCHAR,ATTRIBUTES.MUSICDJ  
VARCHAR,ATTRIBUTES.MUSICBACKGROUNDMUSIC VARCHAR,ATTRIBUTES.MUSICNOMUSIC  
VARCHAR,ATTRIBUTES.MUSICKARAOKE VARCHAR,ATTRIBUTES.MUSICLIVE  
VARCHAR,ATTRIBUTES.MUSICVIDEO VARCHAR,ATTRIBUTES.MUSICJUKEBOX  
VARCHAR,ATTRIBUTES.AMBIENCEROMANTIC VARCHAR,ATTRIBUTES.AMBIENCINTIMATE  
VARCHAR,ATTRIBUTES.AMBIENCCLASSY VARCHAR,ATTRIBUTES.AMBIENCCHIPSTER  
VARCHAR,ATTRIBUTES.AMBIENCEDIVEY VARCHAR,ATTRIBUTES.AMBIENCETOURISTY  
VARCHAR,ATTRIBUTES.AMBIENCETRENDY VARCHAR,ATTRIBUTES.AMBIENCUPSCALE  
VARCHAR,ATTRIBUTES.AMBIENCecasual VARCHAR,ATTRIBUTES.RESTAURANTSGOODFORGROUPS  
VARCHAR,ATTRIBUTES.CATERS VARCHAR,ATTRIBUTES.WIFI  
VARCHAR,ATTRIBUTES.RESTAURANTSRESERVATIONS VARCHAR,ATTRIBUTES.RESTAURANTSTAKEOUT
```

```
VARCHAR,ATTRIBUTES.HAPPYHOUR VARCHAR,ATTRIBUTES.GOODFORDANCING
VARCHAR,ATTRIBUTES.RESTAURANTSTABLESERVICE VARCHAR,ATTRIBUTES.OUTDOORSEATING
VARCHAR,ATTRIBUTES.RESTAURANTSDELIVERY VARCHAR,ATTRIBUTES.BESTNIGHTSMONDAY
VARCHAR,ATTRIBUTES.BESTNIGHTSTUESDAY VARCHAR,ATTRIBUTES.BESTNIGHTSFRIDAY
VARCHAR,ATTRIBUTES.BESTNIGHTSWEDNESDAY VARCHAR,ATTRIBUTES.BESTNIGHTSTHURSDAY
VARCHAR,ATTRIBUTES.BESTNIGHTSSUNDAY VARCHAR,ATTRIBUTES.BESTNIGHTSSATURDAY
VARCHAR,ATTRIBUTES.GOODFORMEALDESSERT VARCHAR,ATTRIBUTES.GOODFORMEALLATENIGHT
VARCHAR,ATTRIBUTES.GOODFORMEALLUNCH VARCHAR,ATTRIBUTES.GOODFORMEALDINNER
VARCHAR,ATTRIBUTES.GOODFORMEALBREAKFAST VARCHAR,ATTRIBUTES.GOODFORMEALBRUNCH
VARCHAR,ATTRIBUTES.COATCHECK VARCHAR,ATTRIBUTES.SMOKING VARCHAR,ATTRIBUTES.DRIVETHRU
VARCHAR,ATTRIBUTES.DOGSALLOWED VARCHAR,ATTRIBUTES.BUSINESSACCEPTSBITCOIN
VARCHAR,ATTRIBUTES.OPEN24HOURS VARCHAR,ATTRIBUTES.BYOBCORKAGE
VARCHAR,ATTRIBUTES.BYOB VARCHAR,ATTRIBUTES.CORKAGE
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSDAIRYFREE
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSGLUTENFREE
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSVEGAN
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONS Kosher
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONS HALAL
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONS SOYFREE
VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONS VEGETARIAN VARCHAR,ATTRIBUTES.AGESALLOWED
VARCHAR,ATTRIBUTES.RESTAURANTS COUNTERSERVICE VARCHAR,HOURS.MONDAY
VARCHAR,HOURS.TUESDAY VARCHAR,HOURS.WEDNESDAY VARCHAR,HOURS.THURSDAY
VARCHAR,HOURS.FRIDAY VARCHAR,HOURS.SATURDAY VARCHAR,HOURS.SUNDAY VARCHAR );
```

5

Create View για το table USER12.YELPCHECKIN:

```
CREATE VIEW USER12.YELPCHECKIN (PERHOUR.ID VARCHAR, ROWKEY VARCHAR PRIMARY
KEY,PERHOUR.WEEKDAY VARCHAR,PERHOUR.HOUR VARCHAR,PERHOUR.CHECKIN VARCHAR );
```

Αφού δημιουργήσαμε τα View για τα δύο Table, συντάξαμε τα SQL Select που μας επιστρέφουν τα επιθυμητά αποτελέσματα. Με βάση τα select αυτά, δημιουργήσαμε τα παρακάτω Indexes ώστε να βελτιώσουμε τον χρόνο εξυπηρέτησης των ερωτημάτων. Το βασικό μας κριτήριο κατά την δημιουργία των Index αποτελεί η μοναδικότητα των τιμών της στήλης που στοχεύει το Indexing.

Εντολές για την βελτιστοποίηση στο PHOENIX ώστε να εξυπηρετούνται αποδοτικά τα ερωτήματα:

Index για το primary - row key του Table USER12.YELPBUSINESS: CREATE INDEX ROWKEYBUSINESSINDEX ON USER12.YELPBUSINESS(ROWKEY);

Index για το primary - row key του Table USER12. YELPCHECKIN: CREATE INDEX ROWKEYCHECKININDEX ON USER12.YELPCHECKIN(ROWKEY);

Index για το BASE.NAME του Table USER12.YELPCHECKIN: CREATE INDEX BASENAMEINDEX ON USER12.YELPBUSINESS(BASE.NAME);

Με την χρήση της εντολής: !outputformat csv ρυθμίσαμε το format του αρχείου των αποτελεσμάτων. Έπειτα, καλέσαμε την εκάστοτε εντολή !record <filename> με το όνομα του αρχείου που θα έχουν τα αποτελέσματα, ώστε τα αποτελέσματα της εντολής Select να οδηγηθούν στο αρχείο .csv. Εκτελέσαμε την αντίστοιχη εντολή Select και σταματήσαμε το recording με την εντολή !record. Η χρήση των Nested εντολών COALESCE(TO_NUMBER(REGEXP_SUBSTR(value, '^d+(\.d+)?'), 0), γίνεται προκειμένου να γίνει safe typecast σε αριθμητικά δεδομένα. Σε περίπτωση σφάλματος, η τιμή που προσμετράται στο αποτέλεσμα είναι το 0.

Εντολές για την λήψη των αποτελεσμάτων των 7 ερωτημάτων:

!outputformat csv

Εντολές για το ερώτημα Q1:

!record USER12Q1.csv

```
SELECT BASE.NAME AS Name,BASE.STATE AS State,BASE.STARS AS Stars FROM USER12.YELPBUSINESS
WHERE BASE.ISOPEN = '1' LIMIT 1000;
```

!record

Εντολές για το ερώτημα Q2:

!record USER12Q2.csv

```
SELECT BASE.NAME AS Name,BASE.ADDRESS AS Address,BASE.CITY AS City,BASE.REVIEWCOUNT AS
ReviewsCount FROM USER12.YELPBUSINESS WHERE BASE.CATEGORIES LIKE 'Drugstores' ORDER BY
BASE.REVIEWCOUNT DESC;
```

!record

Εντολές για το ερώτημα Q3:

!record USER12Q3.csv

```
SELECT SUM(REVIEWCOUNT) AS ReviewCountSum,CATEGORIES AS Categories FROM (SELECT  
COUNT(BASE.REVIEWCOUNT) AS REVIEWCOUNT,BASE.CATEGORIES AS CATEGORIES FROM  
USER12.YELPBUSINESS WHERE BASE.ISOPEN = '1' AND HOURS.MONDAY = '0:0-0:0' AND  
HOURS.TUESDAY = '0:0-0:0' AND HOURS.WEDNESDAY = '0:0-0:0' AND HOURS.THURSDAY = '0:0-0:0' AND  
HOURS.FRIDAY = '0:0-0:0' AND HOURS.SATURDAY = '0:0-0:0' AND HOURS.SUNDAY = '0:0-0:0' GROUP BY  
BASE.CATEGORIES) GROUP BY CATEGORIES;
```

!record

Εντολές για το ερώτημα Q4:

!record USER12Q4.csv

```
SELECT COUNT(BASE.NAME) AS NameCount,BASE.STATE AS State FROM USER12.YELPBUSINESS WHERE  
ATTRIBUTES.SMOKING = 'False' AND HOURS.SUNDAY != 'None' GROUP BY BASE.STATE;
```

!record

Εντολές για το ερώτημα Q5:

!record USER12Q5.csv

```
SELECT SUM(COALESCE(TO_NUMBER(REGEXP_SUBSTR(PERHOUR.CHECKIN, '^\\d+(\\.\\d+)?'), 0)) AS  
CheckInSum, PERHOUR.WEEKDAY AS Weekday, PERHOUR.HOUR AS Hour FROM USER12.YELPCHECKIN  
GROUP BY PERHOUR.WEEKDAY,PERHOUR.HOUR;
```

!record

Εντολές για το ερώτημα Q6:

!record USER12Q6.csv

```
SELECT SUM(COALESCE(TO_NUMBER(REGEXP_SUBSTR(CHECKIN, '^\\d+(\\.\\d+)?'), 0)) AS CheckInSum,  
BASE.CATEGORIES AS Categories FROM USER12.YELPBUSINESS FULL JOIN USER12.YELPCHECKIN ON  
USER12.YELPBUSINESS.ROWKEY = USER12.YELPCHECKIN.ROWKEY WHERE BASE.ISOPEN ='1' AND  
PERHOUR.WEEKDAY != 'Sat' AND PERHOUR.WEEKDAY != 'Sun' AND (PERHOUR.HOUR = '14:00' OR  
PERHOUR.HOUR = '15:00' OR PERHOUR.HOUR = '16:00') GROUP BY BASE.CATEGORIES;
```

!record

Εντολές για το ερώτημα Q7:

!record USER12Q7.csv

```
SELECT BASE.NAME AS Name,BASE.NEIGHBORHOOD AS Neighborhood,BASE.ADDRESS AS  
Address,BASE.CITY AS City,BASE.STATE AS State,BASE.POSTALCODE AS PostalCode,BASE.LATITUDE AS  
Latitude,BASE.LONGTITUDE AS Longitude,BASE.STARS AS Stars,BASE.REVIEWCOUNT AS  
ReviewCount,BASE.ISOPEN AS IsOpen,BASE.CATEGORIES AS Categories FROM USER12.YELPCHECKIN  
FULL JOIN USER12.YELPBUSINESS ON USER12.YELPBUSINESS.ROWKEY = USER12.YELPCHECKIN.ROWKEY  
WHERE PERHOUR.WEEKDAY = 'Sat' ORDER BY PERHOUR.CHECKIN DESC LIMIT 100;
```

!record

Για να κάνουμε λήψη των αρχείων με τα αποτελέσματα των προσωπικό μας υπολογιστή, χρησιμοποιήσαμε το λογισμικό FileZilla για να συνδεθούμε στο Apache Hadoop cluster και να τα μεταφέρουμε από τον προσωπικό μας φάκελο στον προσωπικό μας υπολογιστή.