create database assignment;

## Data Definition Language

create external table shop (shop\_id INT, name STRING)

row format delimited fields terminated by '\t' location '/user/training/shop'

Create external table logfile (shop\_id INT, time int, ip\_address string, session\_id string, visited\_page STRING, referer string)

row format delimited fields terminated by ';' location '/user/training/logfile'

create external table category (category\_id int, russian\_name string, english\_name string)

row format delimited fields terminated by '\t' location '/user/training/category'

create external table list (list\_id int, russian\_name string, english\_name string)

row format delimited fields terminated by '\t' location '/user/training/list'

create external table brand (brand\_id int, name string)

row format delimited fields terminated by '\t' location '/user/training/brand'

create external table topic (topic\_id int, russian\_name string, english\_name string)

row format delimited fields terminated by '\t' location '/user/training/topic'

create external table lat\_long (lati int, lat\_dir string, longi int, longi\_dir string, country\_code string)

row format delimited fields terminated by ',' location '/user/training/lat\_long'

create external table ip\_address (ip\_start string, ip\_end string, country\_id string)

row format delimited fields terminated by ',' location '/user/training/ip\_address'

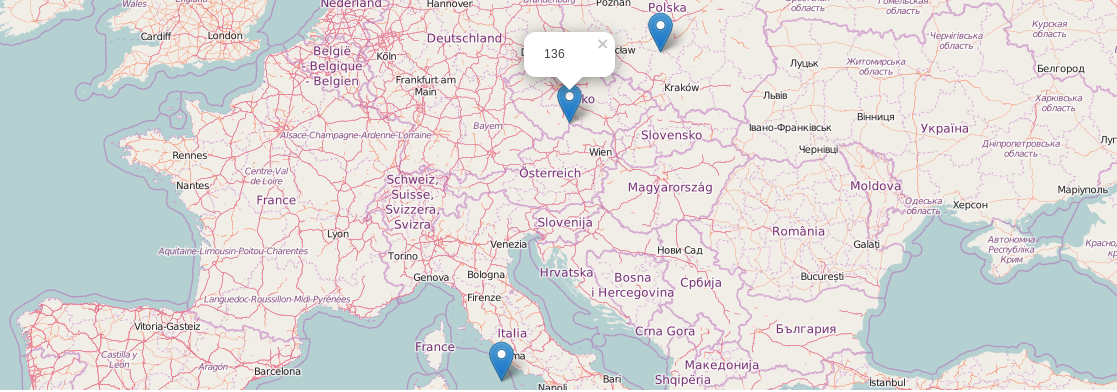
## Data Manipulation Language

**### Count of users from across the world – Demographic analysis**

select count(distinct(l.ip\_address)) as count, ll.lati, ll.longi from logfile l join ip\_address i join lat\_long ll

where l.ip\_address BETWEEN i.ip\_start AND i.ip\_end and i.country\_id = ll.country\_code

group by ll.lati, ll.longi



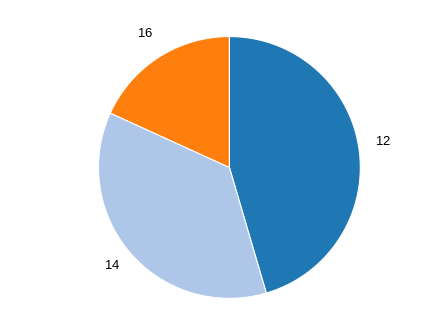
**### Proportion of buyers among the shops – Internal marketing strategy analysis**

select shop\_id,count(visited\_page)

from logfile

where visited\_page LIKE '%udaje.php%'

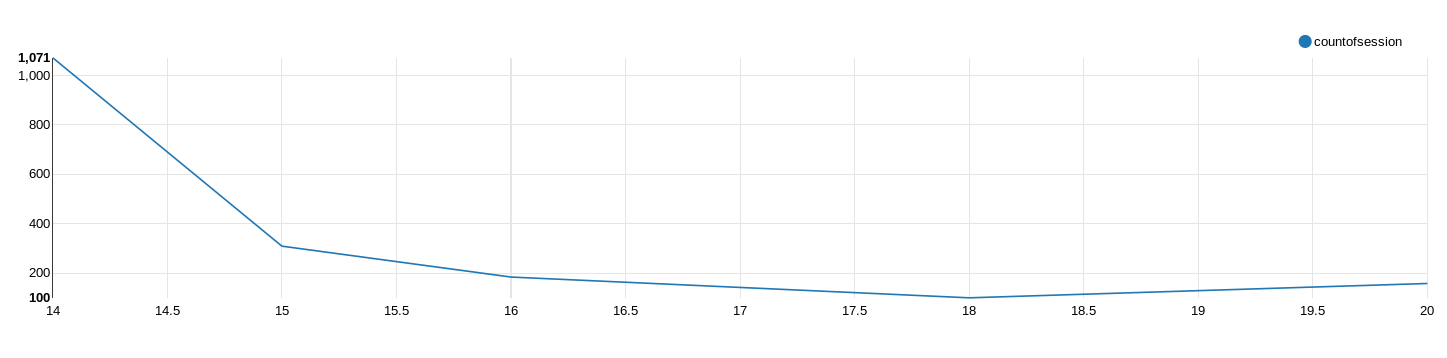
group by shop\_id



**### Identifying the peak traffic hours for buying**

SELECT hour(from\_unixtime(time)) as hours, count(distinct(session\_id)) as countOfSession FROM logfile

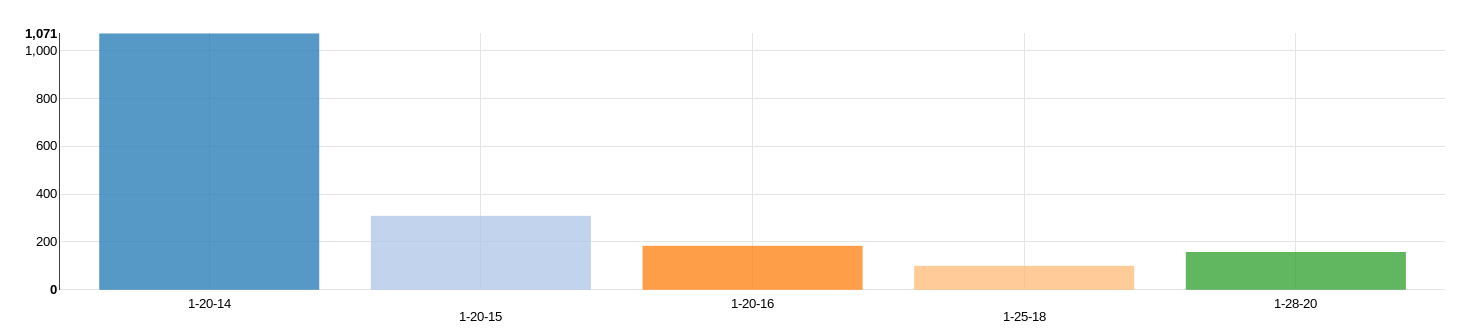
group by hour(from\_unixtime(time))



**### Identifying the peak traffic date (month-day-hour) for buying**

SELECT concat(month(from\_unixtime(time)),'-',day(from\_unixtime(time)),'-',hour(from\_unixtime(time))) as date,count(distinct(session\_id)) as countOfSession FROM logfile

group by concat(month(from\_unixtime(time)),'-',day(from\_unixtime(time)),'-',hour(from\_unixtime(time)))



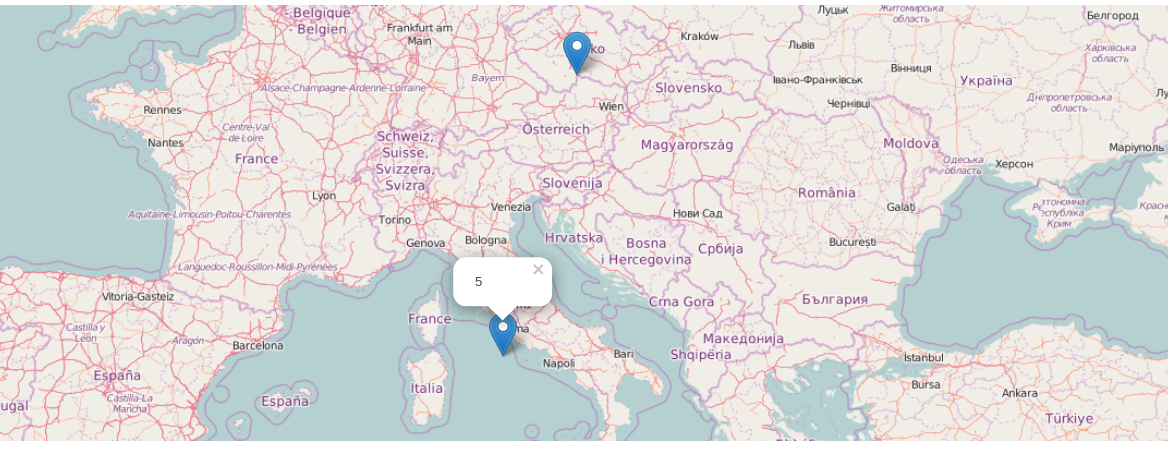
select count(visited\_page LIKE '%udaje.php%') as noOfPurchases, ll.lati, ll.longi

from logfile l join ip\_address i join lat\_long ll

where visited\_page LIKE '%udaje.php%'

and l.ip\_address BETWEEN i.ip\_start AND i.ip\_end and i.country\_id = ll.country\_code

group by ll.lati, ll.longi



select country\_id, count(\*) from ip\_address

group by country\_id

select shop\_id, count(distinct(session\_id)) as count from logfile

group by shop\_id

SELECT from\_unixtime(time) as new\_time, shop\_id, ip\_address, session\_id, visited\_page, referer

FROM logfile where shop\_id = 14

select l.ip\_address, i.country\_id from logfile l join ip\_address i

where l.ip\_address BETWEEN i.ip\_start AND i.ip\_end limit 100

select day(from\_unixtime(time)) as days, count(distinct(session\_id)) as countOfSession FROM logfile

group by day(from\_unixtime(time))

SELECT hour(from\_unixtime(time)) as hours, count(distinct(session\_id)) as countOfSession FROM logfile

group by hour(from\_unixtime(time))

select countOfSession, days, hours from

(SELECT day(from\_unixtime(time)) as days,

hour(from\_unixtime(time)) as hours,

count(distinct(session\_id)) as countOfSession

FROM logfile)

group by days, hours

select o.shop\_id, o.name, l.session\_id, l.visited\_page from logfile l

join obchod o

on l.shop\_id = o.shop\_id

limit 50;

select shop\_id, session\_id, count(session\_id) from logfile

group by shop\_id, session\_id having count(session\_id)>50

# Additional SQL Commands

#### Overwriting from another table

insert overwrite table c\_part partition (country) select \* from c\_ip;