

READ ME FILE

Log analysis project #3

Udacity Full Stack Nanodegree

OVERVIEW

The purpose of this project is to connect to the News database and to obtain results of 3 specific questions, listed below:

What are the most popular articles of all time?

Who are the most popular authors of all time?

Which days did more than 1% of request lead to errors?

INSTRUCTIONS

In order to run this project you need to make sure you have downloaded

- Python 3
- Vagrant
- Virtual box

Launching the virtual machine:

- Download the database file
https://d17h27t6h515a5.cloudfront.net/topher/2016/August/57b5f748_newsdata/newsdata.zip
- Unzip the file
- Vagrant up
- Vagrant ssh
- Cd /vagrant
- Psql -d news -f newsdata.sql
- \dt (view of all the tables and data)

Name of the file: logs.py in order to run the queries

CREATE VIEW COMANDS

QUERY 1

```
SELECT articles.title, COUNT(*) as views FROM articles JOIN log on articles.slug = substring(log.path,10) WHERE path != '/' GROUP by substring(log.path, 10),articles.title ORDER BY views desc LIMIT 3;
```

QUERY 2

```
SELECT authors.name, count(log.path) as views FROM authors left join articles on authors.id = articles.author left JOIN log on log.path like concat('%', articles.slug) GROUP BY authors.name ORDER BY views desc;
```

QUERY 3

```
create view
a1 as SELECT to_char(time,'FMMonth FMDD, YYYY'), COUNT(status) as total FROM log WHERE
status = '404 NOT FOUND' GROUP BY to_char(time,'FMMonth FMDD, YYYY;

create view
r2 as SELECT to_char(time,'FMMonth FMDD, YYYY'), COUNT(status) as total FROM log GROUP
BY to_char(time,'FMMonth FMDD, YYYY;

create view
rate_val as SELECT r2.to_char, cast(a1.total*100 as float)/ cast(r2.total as float) as
val FROM a1,r2 WHERE a1.to_char = r2.to_char;

create view
SELECT * FROM rate_val WHERE val > 1;
```

EXPECTED VIEW RESULTS

#Query1

What are the most popular articles of all time?

```
news=> select articles.title, count(*) as views from articles join log on articles.slug = substring(log.path,10) where path != '/' group by substring(log.path, 10),
articles.title order by views desc limit 3;
      title                | views
-----+-----
Candidate is jerk, alleges rival | 338647
Bears love berries, alleges bear | 253801
Bad things gone, say good people | 170098
(3 rows)
```

#Query2

Who are the most popular authors of all time?

```
news=> select authors.name, count(log.path) as views from authors left join articles on
authors.id = articles.author left join log on log.path like concat('%', articles.slug)
group by authors.name order by views desc;
      name                | views
-----+-----
Ursula La Multa          | 507594
Rudolf von Treppenwitz   | 423457
Anonymous Contributor    | 170098
Markoff Chaney           | 84557
(4 rows)
```

#Query3

Which days did more than 1% of request lead to errors?

```
news=> select * from rate_val where val > 1;
      to_char                | val
-----+-----
July 17, 2016 | 2.26268624680273
(1 row)
```

```
news=> select round(2.2626862468027, 2);
      round
-----
2.26
```

DBNAME= "news"

REFERENCES

I used this websites as additional resources to complete this project:

https://www.youtube.com/watch?v=8LnWXxYYB_4

<https://docs.python.org/2/library/functions.html#int>

<https://www.vagrantup.com/docs/cli/>

<https://stackoverflow.com/questions/1108742/sql-round-function>

<https://stackoverflow.com/questions/5420789/how-to-install-psycpg2-with-pip-on-python>

<https://stackoverflow.com/questions/28677670/why-isnt-pycharms-autocomplete-working-for-libraries-i-install>

<https://discussions.udacity.com>