

Log Analysis Challenge

Introduction

Please find attached an example production log which represents an approximation of one day of heyzap.com hits.

The aim of this task is to do some analytics on this log file.

You should:

- Only spend 3 hours on doing this.
- Your script should be written in a way that it can run once per day

Should be easily configurable to add new data points that you want to analyze

- The output should be readable text output, or html
- The example log data is 67 megabytes, but your program should accommodate a log file of several gigabytes in size.

Data to analyse: Counters

Unique and total page views counters should be built for:

Hits on front page (<http://www.heyzap.com>)

Hits on payments page (<http://www.heyzap.com/payments>)

Hits on payments item look up (http://www.heyzap.com/payments/get_item)

Game plays with the Weebly embed key.

Weebly game plays can be determined by the embed key '12affbbace', where the controller is 'heyzap', the action is 'index' and the permalink to a game is not empty.

Data to analyse: Funnels

When a unique user goes through a certain set of pages that makes a funnel. It is interesting to analyse funnels in order to determine how to optimize pages and where users tend to drop off.

The following two funnels would be useful to have data for:

Publishers front page funnel

Publishers take the following path:

1. <http://www.heyzap.com/>
2. http://www.heyzap.com/publishers/new_site

3. http://www.heyzap.com/publishers/get_embed...

How many uniques go through that path. At each step how many fall off.

Developer front page funnel

Developers take the following path:

1. <http://www.heyzap.com/>
2. <http://www.heyzap.com/developers>
3. http://www.heyzap.com/developers/new_game
4. http://www.heyzap.com/developers/import_games or http://www.heyzap.com/developers/new_inventory_item or
http://www.heyzap.com/developers/upload_game_simple

How many uniques go through that path. At each step how many fall off.

For step 4, developers can take three different options. It would be good to know which option they take and which developers just leave at this point.

Input

It would be good to be able to change what counters and funnels are being analysed. Ideally this can be done in a separate YAML file. Or doing it at the top of the script is also sufficient.

Output

Text output is fine.