

# Log Filtering

Project Presentation

# About

- Safesquid – a proxy server, generates logs for each and every access to internet.
- To interpret and analysis the N/W overview at particular instance is the aim of this project.
- We can summarize information such as frequent user, total data used, data used by user, who accessed what and when etc and can trace particular user to website and vice versa.
- And yes everyone wants results in a flip second, this project understand the needs and time of N/W administrator !

# Technology and Tools Used

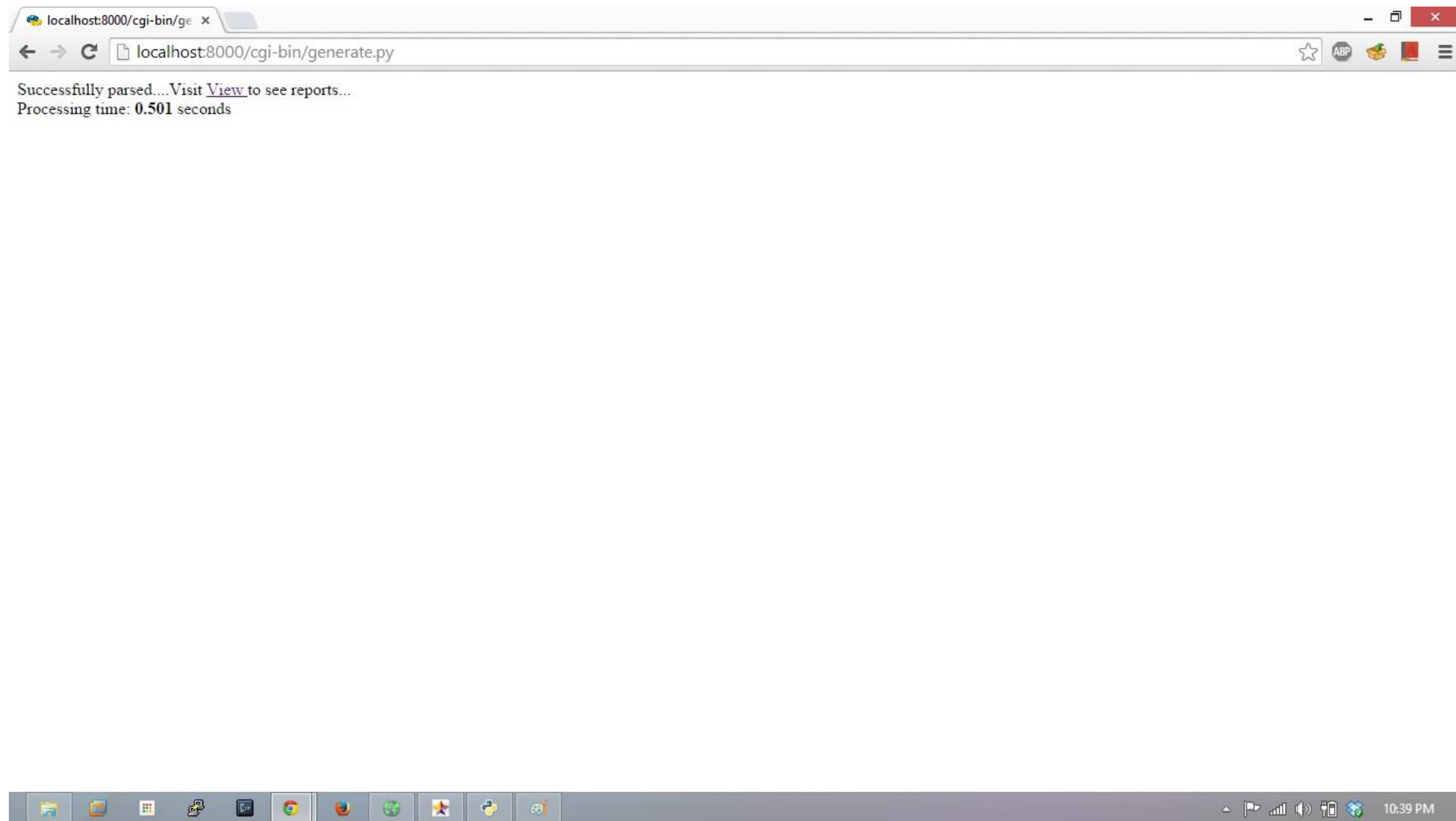
- Python 2.7
- MongoDB (NoSql Database)
- Ajax
- JQuery
- CGI Server

# What changed from v1 to v2 ?

- Removed File I/O and introduced NoSql storage
- Improvised Regex and pattern matching
- Static Graphs to Dynamic Graphs on the go...
- Dynamic summary in tables
- User friendly UI
- Dynamic Graphs and Data summary
- Storage efficient and High performance

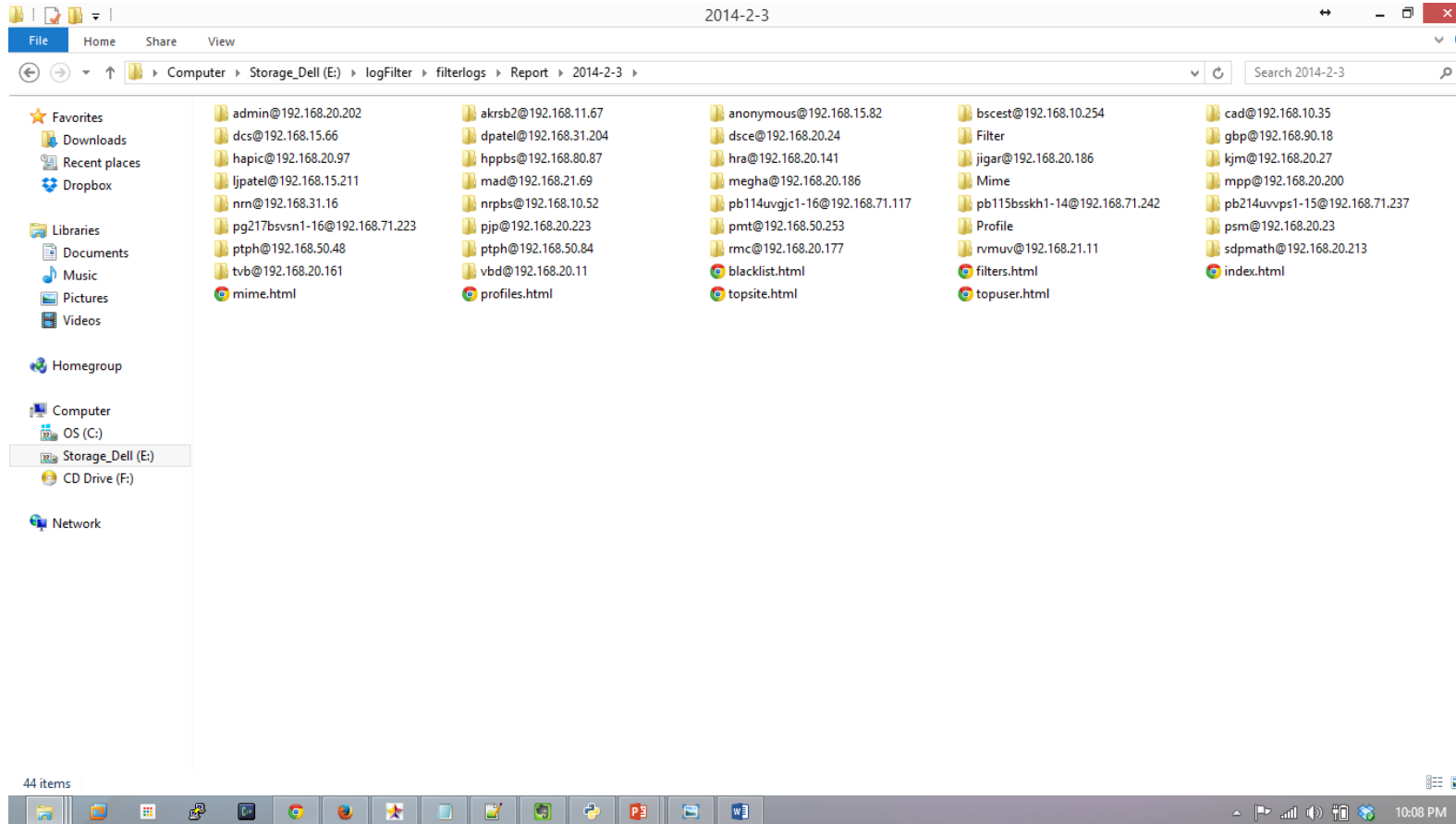
# Performance & Features

Version 1	Version 2
~ 1 minute to process 1 MB log file	~ 0.50 second to process 1 MB log file
Generates only static report	~ <1 second to query 600,000 logs
File Storage	Handles by MongoDB's underlying storage feature
No interactive UI	Dynamic UI
No search feature	Search by users, websites and data used
Can't provide summarized report of more than one logs files	Provide summarized report of more than one log files
Only Top 10 fields were displayed	Uptill Top 100 can be displayed



Version 2 – process time

# How data is stored ?



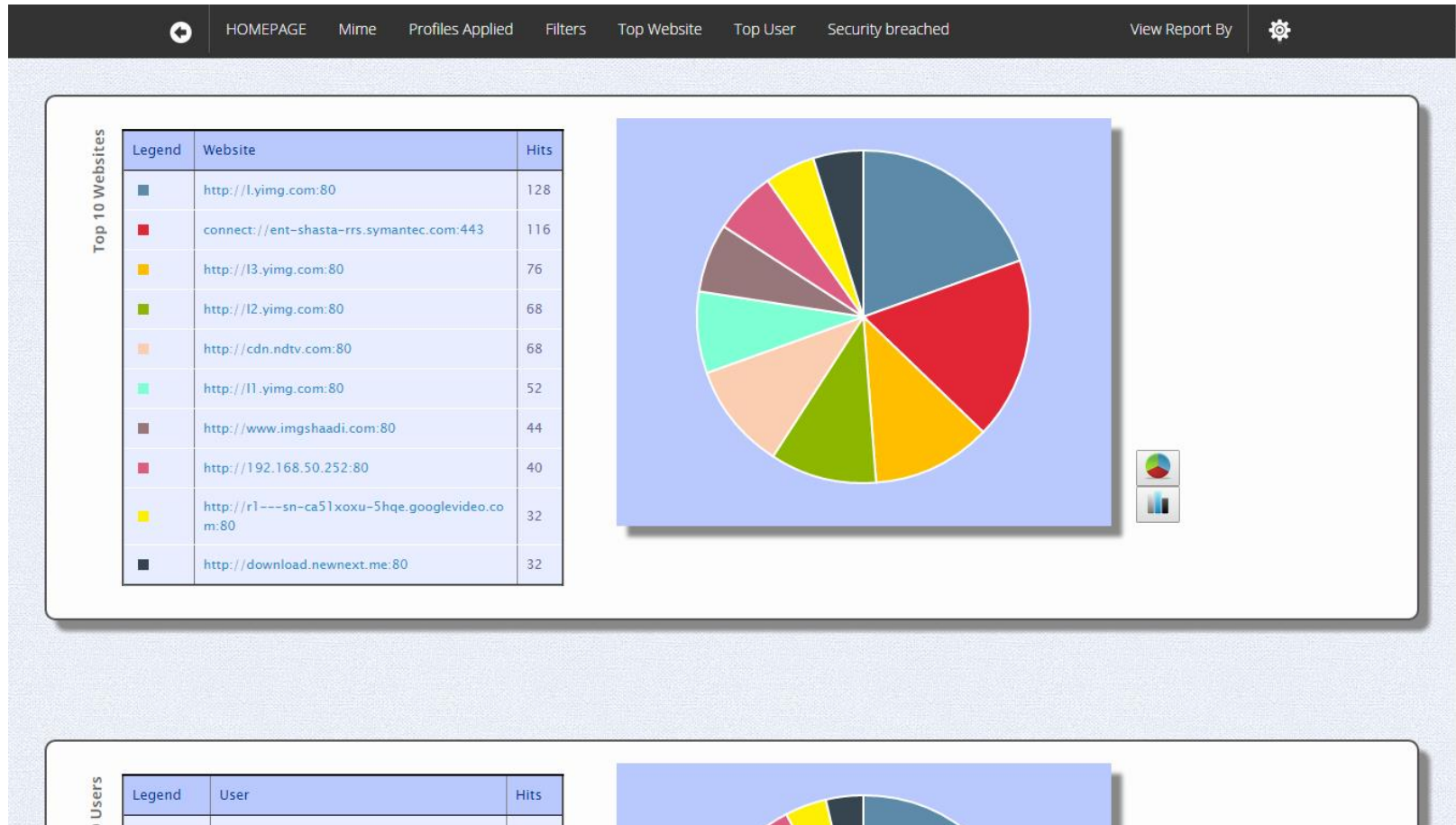
Version 1 - File System Storage

```
C:\Windows\system32\cmd.exe
> use logfilter
switched to db logfilter
> db.logs.findOne()
{
  "_id" : ObjectId("545bab4136306b2a24abdc3d"),
  "username" : "pm",
  "client_ip" : "192.168.20.23",
  "method_url" : "GET http://www.rakshashaktiuniversity.edu.in:80/images/closelabel.gif",
  "httpstatus" : "200",
  "interface" : "10.0.0.5:8080",
  "mime" : "image/gif",
  "filter_name" : "-_-",
  "date" : "03/Feb/2014",
  "id1" : "1391426555.892",
  "elapsed_time" : "4248",
  "byte_transferred" : 979,
  "user_agent" : "Mozilla/5.0 (Windows NT 5.1; rv:14.0) Gecko/20100101 Firefox/14.0.1",
  "filter_profile" : "staff-uvp,http,cachable,block_finance,Mozilla_Firefox,allowed-https,Block_Porn_Keyword,Block_Proxy_Keywords,staff-limit,DYNAMIC_LENGTH,imgfi",
  "time" : "16:52:40"
}
```

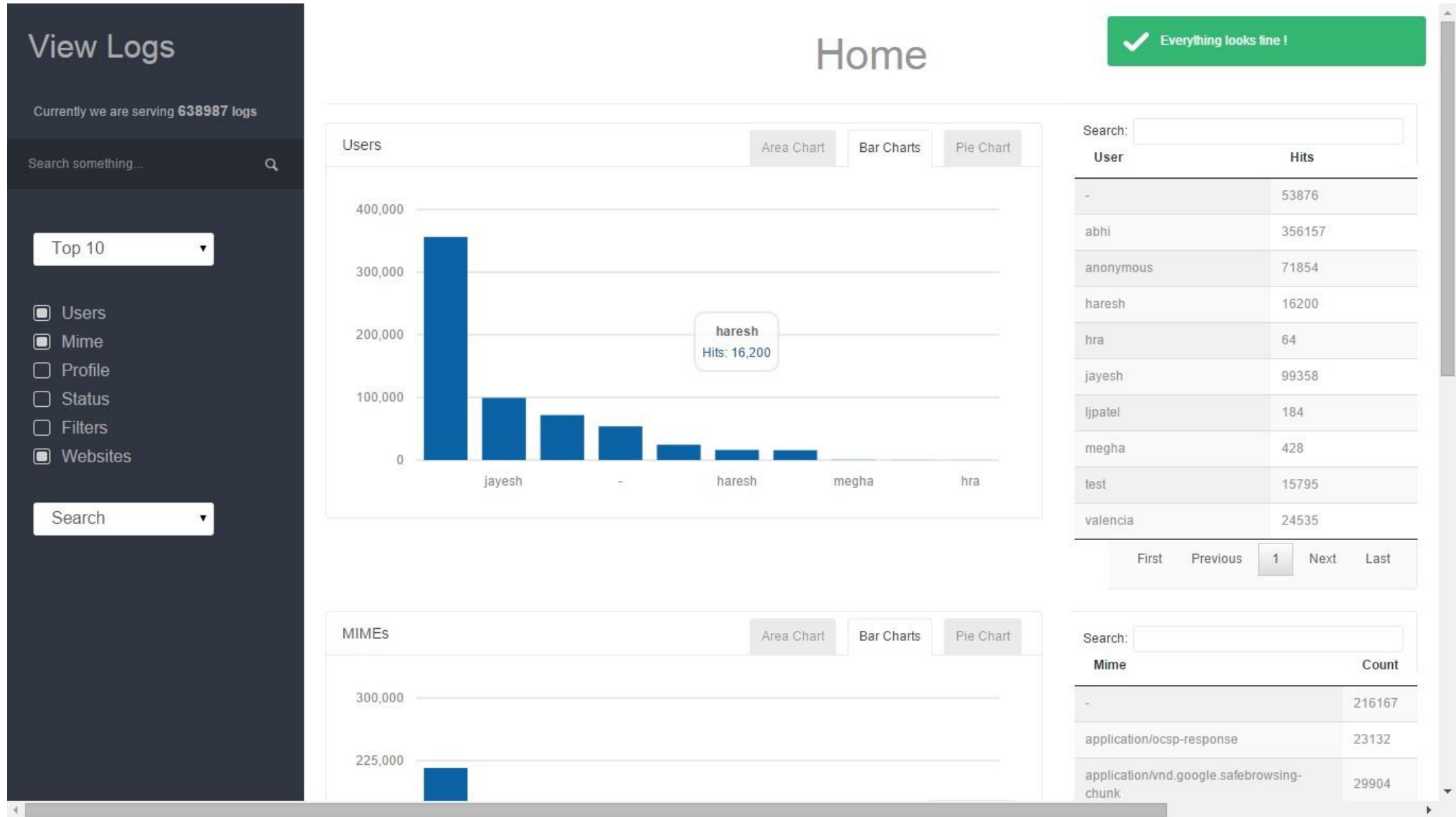
Version 2 – NoSql Document structure



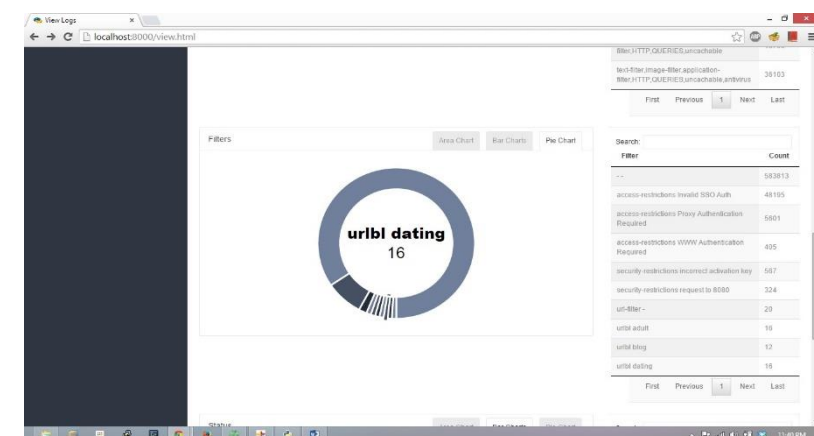
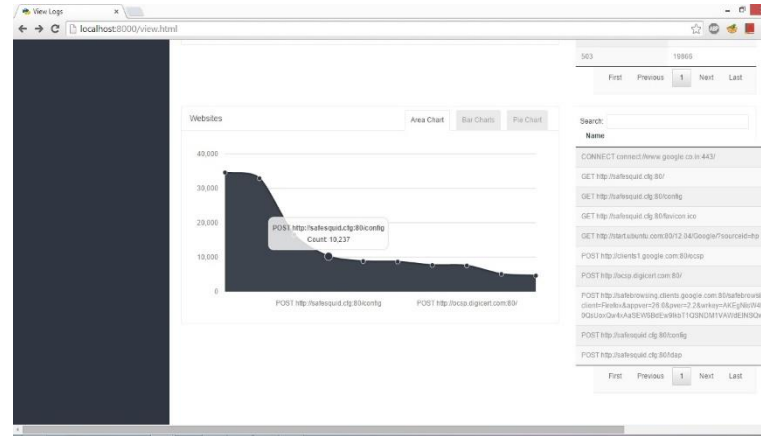
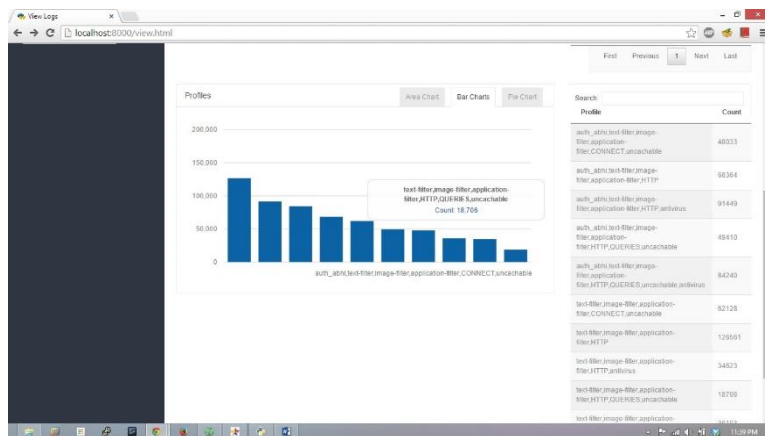
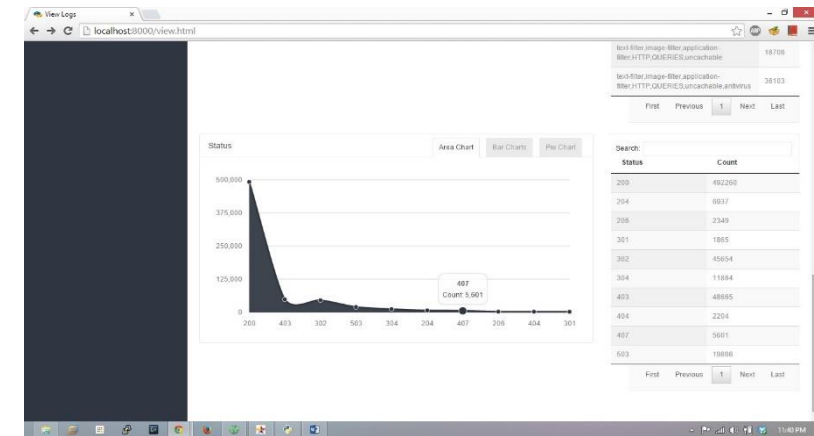
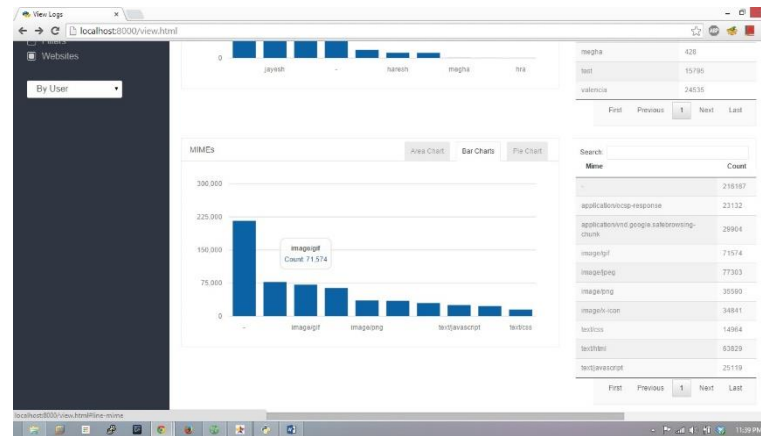
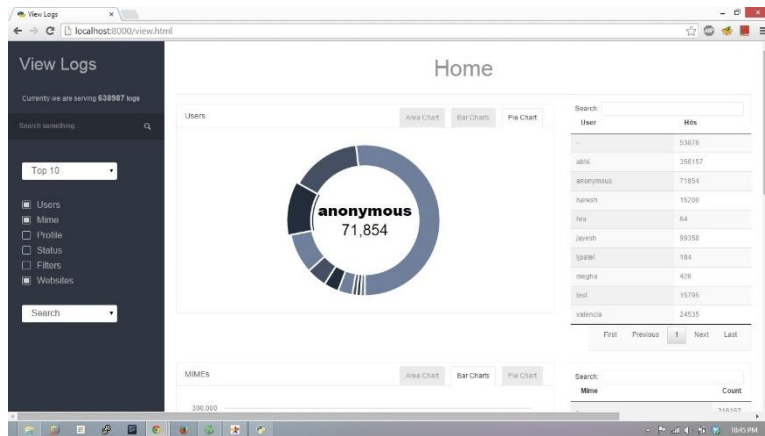
# User Interface



Version 1 - Static Interface




# Dynamic Charts and Tables




# Search

View Logs

Currently we are serving 638987 logs

Search something... 

By User 

Enter User name

List of websites visited by haresh

Search:

Website	Clicks
CONNECT connect://accounts.google.com:443/	81
CONNECT connect://analytics.twitter.com:443/	81
CONNECT connect://api.digitalocean.com:443/	81
CONNECT connect://apis.google.com:443/	81
CONNECT connect://clients5.google.com:443/	81
CONNECT connect://d.adroll.com:443/	405
CONNECT connect://dnn506yrbagrg.cloudfront.net:443/	81
CONNECT connect://googleads.g.doubleclick.net:443/	162
CONNECT connect://help.ubuntu.com:443/	486
CONNECT connect://ib.adnxs.com:443/	81

Showing 1 to 10 of 112 entries

First Previous 1 2 3 4 5 ... 12 Next Last

Websites blocked

Search:

Blocked Website	Clicks
No data available in table	

Showing 0 to 0 of 0 entries

First Previous Next Last

Search by user

# View Logs

Currently we are serving 638987 logs

Search something ...

By Website

Enter website

solidtechreview

About 70 results (0.815 seconds)

Search:

Website	Clicks
GET http://solidtechreview.com:80/	81
GET http://www.google-analytics.com:80/r/collect?v=1&_v=30&a=596522373&t=pageview&_s=1&dl=http%3A%2F%2Fwww.solidtechreview.com%2F&ul=en-us&de=UTF-8&dt=Solid%20Tech%20Review&sd=24-bit&sr=1366x768&vp=1288x678&je=0&fl=11.2%20r202&_u=MEAAAAQAI~&jid=1968626072&cid=694391100.1414684702&tid=UA-42129205-1&_r=1&z=696396875	81
GET http://www.solidtechreview.com:80/	162
GET http://www.solidtechreview.com:80/cdn-cgi/inexp/dok2v=1613a3a185/cloudflare/rocket.js	81
GET http://www.solidtechreview.com:80/cdn-cgi/pe/bag2?r[]=http%3A%2F%2Fwww.google-analytics.com%2Fanalytics.js	81
GET http://www.solidtechreview.com:80/cdn-cgi/pe/bag2?r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fjquery.mousewheel.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Fjs_composer%2Fassets%2Fisotope%2Fjquery.isotope.min.js%3Fver%3D3.4.12&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fjquery.nicescroll.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fjquery.colorbox.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fsite.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Ffoundation.min.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fscrolling.js	81
GET http://www.solidtechreview.com:80/cdn-cgi/pe/bag2?r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fmodernizr.foundation.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-includes%2Fjs%2Fjquery%2Fjquery.js%3Fver%3D1.11.0&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-includes%2Fjs%2Fjquery-migrate.min.js%3Fver%3D1.2.1&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Ffancy-box%2Fjquery.fancybox.js%3Fver%3D1.2.6&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Ffancy-box%2Fjquery.easing.js%3Fver%3D1.3&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Frevslider%2Frs-plugin%2Fjs%2Fjquery.themepunch.plugins.min.js%3Fver%3D3.9.2&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Fjs%2Fjquery.themepunch.revolution.min.js%3Fver%3D3.9.2&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Fassets%2Fjs%2Fapp.js&r[]=http%3A%2F%2Fcode.jquery.com%2Fui%2F1.9.1%2Fjquery-ui.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Fno-right-click-images-plugin%2Fno-right-click-images.js&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fthemes%2FOneTouch%2Finc%2Fhomepage_builder%2Fassets%2Fjs%2Faqpb-view.js%3Fver%3D1414684682&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-includes%2Fjs%2Fjquery%2Fui%2Fjquery.ui.core.min.js%3Fver%3D1.10.4&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-includes%2Fjs%2Fjquery%2Fui%2Fjquery.ui.widget.min.js%3Fver%3D1.10.4&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-includes%2Fjs%2Fjquery%2Fui%2Fjquery.ui.tabs.min.js%3Fver%3D1.10.4&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Fjs_composer%2Fassets%2Fjquery-ui-tabs-rotate%2Fjquery-ui-tabs-rotate.js%3Fver%3D3.4.12&r[]=http%3A%2F%2Fwww.solidtechreview.com%2Fwp-content%2Fplugins%2Fjs_composer%2Fassets%2Fjs_composer_front.js%3Fver%3D3.4.12	81
GET http://www.solidtechreview.com:80/wp-content/plugins/fancy-box/jquery.fancybox.css?ver=1.2.6	81

Search by website

Search By Website

localhost:8000/websiteDetails.html?url=http://solidtechreview.com:80/

Search

Star

ABP

Icons

Menu

View Logs

Currently we are serving 638987 logs

Search something...

Top 10

☐ Users

☐ Mime

☐ Profile

☐ Status

☒ Website

☐ Filter

Search

http://solidtechreview.com:80/

About 1 results (0.770 seconds)

Search:

User	Date	Click
valencia	30/Oct/2014	81

Showing 1 to 1 of 1 entries

First

Previous

1

Next

Last

Taskbar

12:01 AM

Website Details

Search By Website

localhost:8000/searchByData.html

View Logs

Currently we are serving 638987 logs

Search something...

By Data Used

From 10 To 100

show

About 21005 MB data has been used till now

Search:

Username	Data Used
cad	17 MB
hapi	22 MB

Showing 1 to 2 of 2 entries

First Previous 1 Next Last

12:03 AM

Search by Data

# MongoDB Query

```
db.aggregate([
  {'$match':{'username':{'$regex':search_str}}},
  {'$group':{'_id':'$method_url','count':{'$sum':1}}},
  {'$sort':{'count':-1}}
])
```



# Conclusion

- Extracting useful info from large dataset in no-time ensures admin that internet is used for right reasons.
- With various options and search features, admin can narrow down search fields.
- Fast processing and dynamic data, helps admin to get information without going through logs.
- NoSql database provides fast response with unstructured and complex data than any other relational DBs.