Performance Analysis of a Website using Key Performance Indicators

Navin Kumar Tyagi
Department of Computer Science,
MIT Bulandshahr, India

Anil Kumar Solanki
Department of Computer Science,
BIET Jhansi, India

Manoj Kumar Sharma
Department of Computer Science,
MIT Bulandshahr, India

ABSTRACT
Web analytics is the measurement of visitor behaviour on a website and it can be used to collect basic visitor information such as number of visitors and visit duration etc. This basic information can then be combined to create meaningful key performance indicators which are used to analyze the performance of the website. In this paper we analyzed the log files of a website using a log analyzer tool. We used different key performance indicators like visit depth, frequency of visit and visitor ratio for our analysis.

Keywords
Web analytics, Key performance indicators, Visit depth, Deep log analyzer.

1. INTRODUCTION
Modern websites have evolved significantly in the past few years. Most of the websites consists multiple elements like images, video, audio, external JavaScript etc. Websites are categorized as E-commerce; lead generation, content/media and support/self service websites on the basis of their objectives [1]. As a part of marketing strategy it is crucial to track the performance of a website to monitor the success of that website. In order to fulfil this we should know about the behaviour of visitors on the website. Web analytics plays an important role in the measurement and analysis of web usage data. Different web analytics metrics can be collected which can be used to develop key performance indicators (KPIs).It can be treated as a versatile analytic model which measures several metrics against each other to define visitor trends. Key performance indicators use the combination of metrics to identify visitor behaviour on a website. This information can be used by an organization to align its website goal with business goals.

We analyzed the log files of the Trinity Touch website (www.trinitytouch.com).We used Deep Log analyzer tool for analyzing the log files of the web server. The objective of analysis was to assess the visitor usage behaviour on the website.

In Section2 we presented the related work done in this area. In section3 used methodology is given. In section4 we described different key performance indicators for a website. Results and observation are given in section 5 and finally we conclude the paper in section 6.

2. RELATED WORK
Arlitt and Jin [2] first characterized the workload of a large commercial website (World cup 1998).In [3] study about dynamics of server content and client accesses of a large commercial website was presented and it was found that file modifications were more frequent than file creations, file popularity followed a Zipf distribution, and popularity of documents tended to decrease as they became older. The server workloads of a customizable university web portal analysed in [4] and it has been deduced that client side latencies can be reduced by actively prefetching a few non shared personalized channels. A study of workload characterization of a conference website was presented in [5]. They focused on usage behaviour, client errors, client network properties, bandwidth, and robot activity of the site. Rowbottom et.al.[6] introduced the Web server log file and assessed its potential as a research instrument to measure and interpret the use of corporate reporting system. Rockman et.al [7] analyzed the log files of UCMP teacher website to assess visitor characteristics, navigation and search patterns, relative popularity of pages, sections, and other features of the site.

3. METHODOLOGY

3.1 Log file analysis
In web analytics, log files and page tagging are two well known methods for data collection. In log files method web server log files are analyzed. Each request of a client which he/she submit to server is recorded and saves to a log file. To understand the usage patterns of a website these log files need to be analyzed. This technique is known as server side data collection. Page tagging method uses the visitor's web browser to collect web usage data. A snippet of JavaScript code is placed in each page of website. When a page is requested, the embedded JavaScript runs and collects data which are sent to a remote server. Cookies are used to track user activities like visit duration and return visits. This technique is known as client side data collection.

Web usage data can be exhibited in different manners and following are the most common formats used widely.

NCSA Common Log
NCSA Combined Log
NCSA Separate Log
W3C Extended Log

In NCSA common log format, information includes the client IP address, client identifier, visitor user name, date and time, HTTP request, Status code, and the number of bytes transferred during the request. Same information as the common log format is contained in combined log format. In addition to this it also consists three additional fields: the referring URL, web browser and operating system information, and the cookie. The separate log format or 3-Log format contains the same information as the combined log, but this information is divided into three separate files namely the access log, the referral log, and the agent log. W3C format provides better control and manipulation of data and it is different from other formats. It produces a log file readable by most web analytics tools. The extended format consists user defined fields and identifiers and on the other hand default values are represented by a dash "-"[8]. Table1 and Table2 show four different types of log formats [9], [10].

Table 1. *NCSA* Log Comparison

NCSA Common Log	125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html HTTP/1.0" 200 1043
NCSA Combined Log	125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html HTTP/1.0" 200 1043 "http://www.ibm.com/" "Mozilla/4.05 [en] (WinNT; I)" "USERID=CustomerA;IMPID=01234"
NCSA Separate Log	Common Log: 125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html HTTP/1.0" 200 1043 Referral Log: [10/Oct/1999:21:15:05 +0500] "http://www.ibm.com/index.html" Agent Log: [10/Oct/1999:21:15:05 +0500] "Microsoft Internet Explorer - 5.0"

Table 2. W3C Extended Log File

W3C Extended Log	#Software: Microsoft Internet Information Services 6.0 #Version: 1.0 #Date: 2002-05-24 20:18:01 #Fields: date time c-ip cs-username s-ip s-port cs-method cs-uri-stem cs-uri-query sc-status sc-bytes cs-bytes time-taken cs(User-Agent) cs(Referrer) 2002-05-24 20:18:01 172.224.24.114 - 206.73.118.24 80 GET /Default.htm - 200 7930 248 31 Mozilla/4.0+(compatible;+MSIE+5.01;+Windows+2000+Server) http://64.224.24.114/

Log file method does not require any changes to the website or any extra software installation to create the log files. Web servers automatically create these logs and store them in the server which can give any company freedom to alter their web analytics tool and strategies as well. This method does not require any extra bandwidth when loading a page and it is also possible to log both page request successes and failures. The apache HTTP server provides a variety of different mechanism for logging everything which happens on your server from the initial request, through the URL mapping process, to the final resolution of the connection including any errors that may have occurred in the process. In addition to this, third party modules provides logging capabilities or inject entries into the existing log files and applications such as CGI programs, PHP scripts or other handlers may send message to the server error log.

3.2 Trace data overview

We have collected the data of Trinity Touch website (www.trinitytouch.com) for the period of three months (92 days), between 1 August 2010 and 31 October 2010. We have got 1 GB of data during this period. Tables 3 summarize the dataset used in our study. The server logs contained 1344440 hits and 15.2 GB of volume was transferred during this period. On average the website received 14613 hits per day and 170 MB of data was transferred daily. Table 3 also shows the visit, visitor activity, and error hits.

Table 3.Summary of Dataset

Number of hits	1344440
Number of visits	36774
Number of unique visitors	6206
Visitors who visited once	4006
Repeat visitors	2200
Number of page views	611645
Average visits per day	400
Average visit duration	15.38 Min
Average visits per visitor	6.49
Outgoing traffic	15.23 GB
Average traffic per day	169.51
Average visitors per day	67
Number of pages per visit	16.6
Error hits	5222

4. KEY PERFORMANCE INDICATORS

Metrics are the numerical representations of data collected from a website. On the other hand key performance indicators are measured by ratio of two metrics and tied to a business strategy. A business can save both time and money by selecting key performance indicators on the basis of website type [11]. First we should understand the business goals of the company and then we have to determine which KPIs will provide the most insight. For a website to be beneficial information gathered from its visitors does not merely show what has happened in the past but it is also able to improve the site for future visitors. The four Ms of operational management [12] can be used to facilitate effective selection of key performance indicators.

Motivate

To make sure that goals are relevant to every stakeholder.

Manage

Encouragement, collaboration and involvement required for achieving the desired goals.

Monitor

Tracking of key performance indicators and quick response to the problems occurred.

Measure

Root causes of the problems are identified and assumptions associated with the strategy are tested.

Followings are the some key performance indicators for a content/media website which we have taken for analysis purposes.

Visit Depth: It is also referred to as depth of visit or path length. It is the measurement of the ratio between page views and unique visitors or how many pages a visitor accesses in each visit. Generally visitors with higher visit depth are interacting more with the website.

Returning visitor ratio*:* It is the ratio of unique visitors to total visits. A factor in customer loyalty, this KPI measures the effectiveness of a website to bring visitor back. A lower ratio for the KPI is best as a lower number indicates more repeat visitors who are interested and trust the content of the site [13],[14].

New visitor ratio*:* It is the measurement of new visitors to unique visitors and is used to determine if a web site is attracting new people.

When measuring this KPI, the age of the site plays an important role as newer site will attract new people. Another factor to consider is customer retention. However the new visitor ratio should decrease over the time as the returning visitor ratio increases.

Page depth: This is the ratio of page views for a specific page and the number of unique visitors to that page. Its measurement focuses more on page popularity. Average page depth can be used to measure interest in specific areas of a web site over the time. If one particular page on a web site has a high page depth, it is an indication that page is of particular interests to visitors.

5. RESULT ANALYSIS

In this work we analyzed the server log data of Trinitytouch website by taking different key performance indicators. In figure 1 visit depth on daily basis has been shown. Some major spikes have been observed in the month of august. One major peak was observed on 31/8/2010 with 113 page views. Pattern showing significant number of page views per visit for most of the days which indicates that visitors are interacted and engaged with the site.

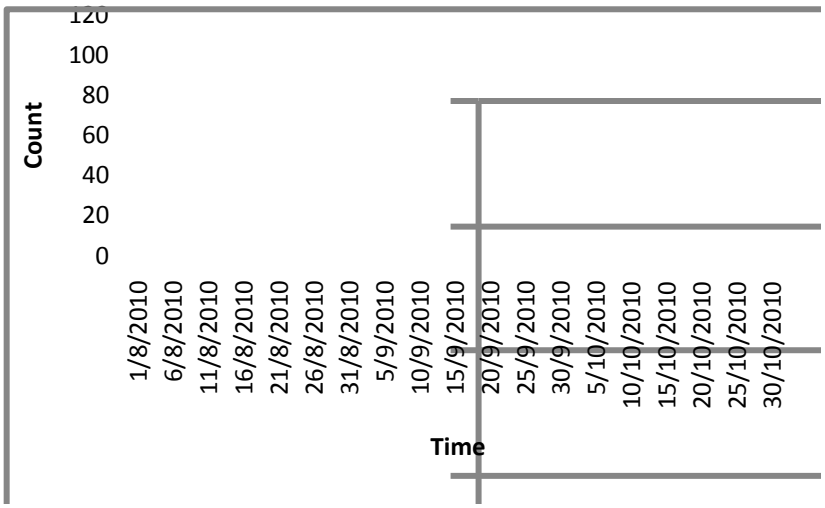

Fig 1: Visit depth

Figure 2 shows frequency of visits per day. On average two visits were observed most of the days. Two peaks with approximately 50 visits were observed in the month of august (11 and 26 august 2010). Higher number of visits on these two dates may be caused either by high bounce rate or click fraud which means visitors used to generate visits to the site without having genuine interest. This KPI shows the effectiveness of web site to bring the visitors back (more repeat visitors) or in other words it shows the customer loyalty for the site.

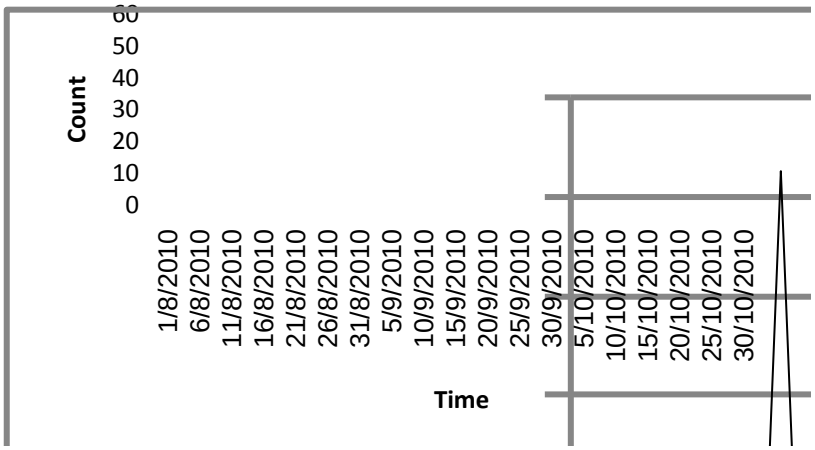

Fig 2: Frequency of visit

Figure 3 illustrates the New and Repeat visitors' ratio and patterns suggested that site is attracting the new people. Some spikes in new visitors' ratio and corresponding dips in repeat visitors' ratio have been observed throughout the traced period. Overall pattern is fluctuating.

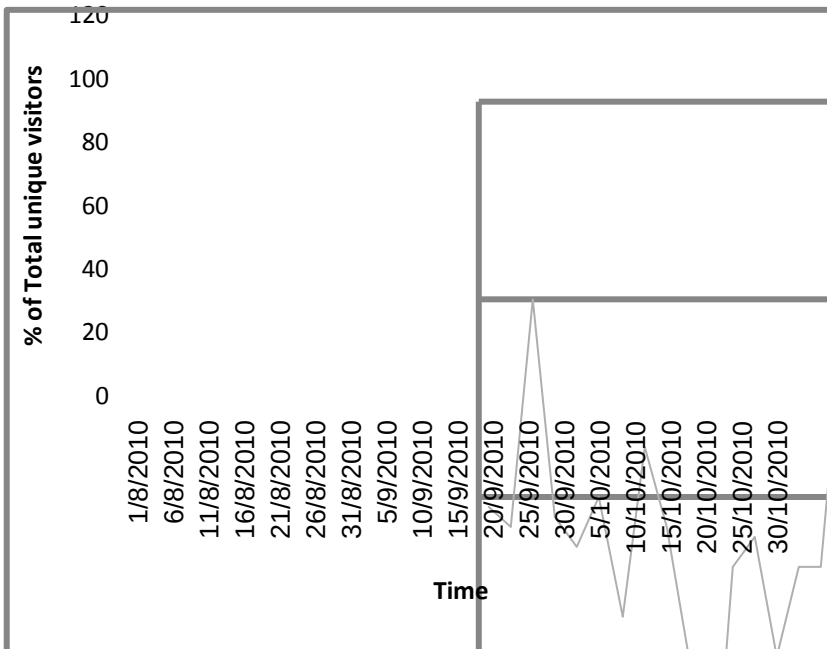

Fig 3: New & repeat visitor ratio

Figure 4 shows the average visit duration per day for three month period. It can be observed from the figure that most of the days visit length varies between 0 to 50 minutes. The huge spike was observed on 26/10/2010 which is perhaps due to the person not closing the browser. Another huge peak was observed on 19/8/2010 due to the same reason.

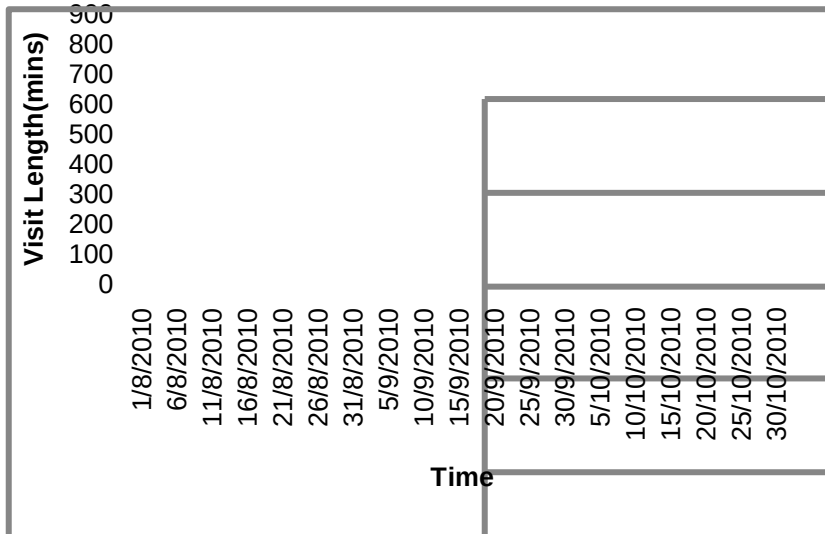

Fig 4: Visit Length

6. CONCLUSIONS

In this paper we analyzed the server log data of Trinity Touch web site using Deep log analyzer tool. We analyzed the usage behaviour of the visitors using key performance indicators of a content/media web site. Usage patterns shows that visitors are engaged with the site and also liking the contents of the site. They are also showing loyalty to the site as we found good percentage of repeat visitors. New visitor ratio KPI shows that web site is also attracting the new people.

7. REFERENCES

[1] Danielle Booth and Bernard J.Jansen. A review of methodologies for analyzing websites.

[2] M. Arlitt and T.Jin. A workload characterization study of the 1998 world cup web site. IEEE Network, 14 (3), 2000.

[3] V. Padmanabhan and L. Qiu. The content and access dynamics of a busy web site: Findings and implications. In proc. ACM SIGCOMM, 2000.

[4] W. Shi, Y. Wright, E. Collins, and V. Karamcheti. Workload Characterization of a personalized web site and its implications for dynamic content caching. In Proc. WCW, 2002.

[5] Aniket Mahanti, Leanne Wu and Carey Williamson. Workload Characterization of the WWW2007 Conference Web site.

[6] N. Rowbottom, A. Allam, and A. Lymer. An Exploration of the potential for studying the usage of investor relations information through the analysis of web server log.

[7] Rockman et.al. Evaluation of UCMP Evolution web site usage statistics. September 2005.

[8] Hallam-Baker, P.M. and Behlendorf, B. Extended log file format. February 1999. http://www.w3.org/TR/WD logfile.html

[9] IBM. (2004, May 19). Log File Formats. Retrieved, from http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA_info45/en_US/HTML/guide/c-logs.html.

[10] Microsoft. (2005, August 22). W3C Extended Log File Examples. Retrieved from http://technet2.microsoft.com/WindowsServer/en/library/b5b8a519-8f9b-456b-9040-018358f2c0c01033.mspx?mfr=true

[11] McFaddan, C. Optimizing the online business channel with web analytics. July 2005. http://www.webanalyticsassociation.org/en/art/?9

[12] Becher, J.D. Why metrics centric performance management solutions fall short. DM Review magazine.

[13] Mason, N. (2007, February 6). Customer Loyalty Improves Retention. Retrieved from http://www.clickz.com/showPage.html?page=3624868(11)

[14] Nicholas, D., Huntington, P. & Williams, P. Evaluating Metrics for Comparing the Use of Web sites: A Case Study of 2 Consumer Health Web sites, Journal of Information Science 2002; 28/1:63--75.