

relationships valuable to customer relationship management. Even at 140 characters per tweet, the high velocity (or frequency) of Twitter data ensures large volumes (over 8 TB per day). -generated /sensor data includes Call Detail Records (CDR), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), trading systems data.

(c) Variety

Traditional data formats tend to be relatively well defined by a data schema and change slowly. In contrast, non-traditional data formats exhibit a dizzying rate of change. As new services are added, new sensors deployed, or new marketing campaigns executed, new data types are needed to capture the resultant information. Big data typically refers to the following types of data: Traditional enterprise data includes customer information from CRM systems, transactional ERP data, web store transactions, and general ledger data. Machine-generated /sensor data includes Call Detail Records (CDR), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), and trading systems data. Social data includes customer feedback streams, micro-blogging sites like Twitter, and social media platforms like Facebook.

4. Proposed System

4.1 How system works?

Designing a web application MakeMyChoice for promoting the products based on user profile abstraction. The vendor products are advertised based on information given by user during registration. From the next time, this selection is done by applying map-reduce methods to the user clicks on various products :

- 1) Logging Module, vendor enters the details regarding all the products to be promoted.
- 2) User has to fill the registration form in which one has to enter details regarding age, gender, interests, etc. On the basis of this data, the advertisements are generated for the first login.
- 3) From the next login, the segregation of products to be displayed is done on the basis of tracked user clicks. This tracking of user clicks is the result of application of map-reduce methods.
 - These methods are a part of HADOOP technology.
 - The map method undertakes filtering and sorting of data.
 - The reduce method performs summary operation.
 - As a result, the promoted advertisements are displayed each time map-reduce methods are applied.
- 4) All the data to be filtered and summarized is stored in a specialized file system known as HDFS (HADOOP Distributed File System). It is a file system designed for storing very large files with streaming data access patterns, running on clusters on commodity hardware. In HDFS, data is laid out sequentially on your hard disk, reducing the number of seeks to read data.
- 5) The user and vendor can also update their profile.
- 6) The user can also search for new products. If available, it is displayed and if not, then particular message is shown.

4.2 Architecture of Proposed System

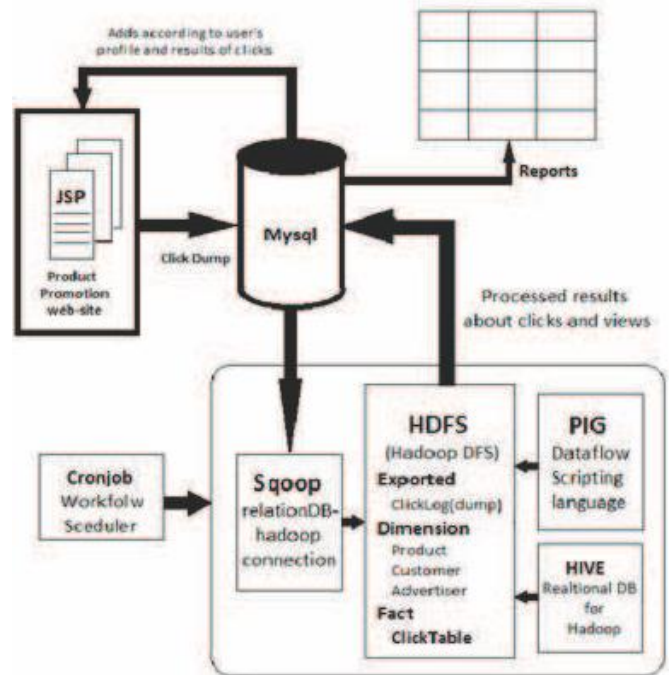


Figure 1: Architecture of Online Product Promotion System

The Online Product Promotion System is executed on Linux platform. The website is created for Actors to interact with system. This website provide to every user who logs in. The customer is shown with product of his interest (i.e. product which matches to his profile). The Advertiser is provided with reports with any of his products. The processing in background to achieve these targets is as follows:

1. The Website will be created to WebPages.
2. The data stored in Click-dump will be transferred to Hadoops file system (HDFS) using snoop. The imported table will be stored on the HDFS in date-wise folder. The table is stored in HDFS will help to calculate the targets.
3. This stored click-dump will be processed by Pig Script (written in Pig Latin). This pig script will return number of clicks to each product and send this information to hive by loading it into it.
4. The hive will provide SQL like command to execute queries. These queries will produce target results.
5. This process will be scheduled for whole day. The cronjob will be applied to execute whole process one time in day.

In this paper, these three Actors play an important role.

- (a) **Customer:** Customer is expected to see advertisements of his own interest (i.e. products which matches to his profile), when customer logs in to website. The customer also given facility to update his profile, modify his interest.
- (b) **Advertiser:** Advertiser is expected to get reports of products. These reports have to include position of product in market, number of hits to products and so on. Advertiser should be able to add/delete/update the product information from his account.
- (c) **Administrator:** Administrator is main controller of the website. He is given full access to website. He can view

customer/advertiser details. He can also send reports to admin as per advertiser's request.

5. Results

The proposed system gives the following functionality :

1. Advertiser Sign In

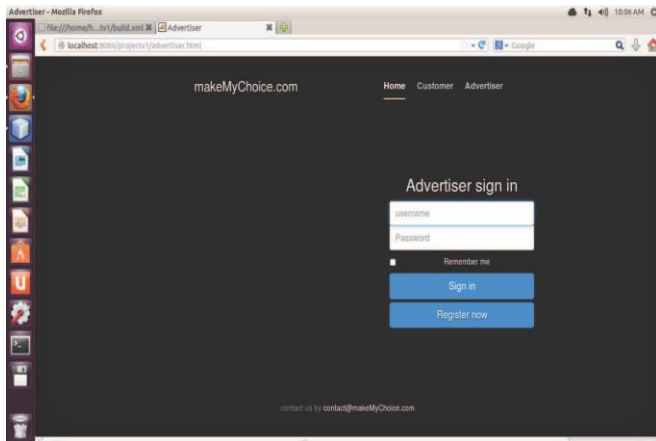


Figure 2: Advertiser Sign In

2. Advertiser dashboard

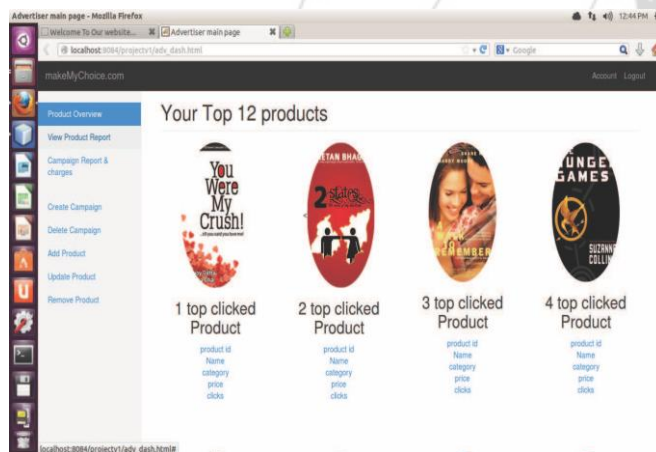


Figure 3: Advertiser dashboard

3. Create Campaign

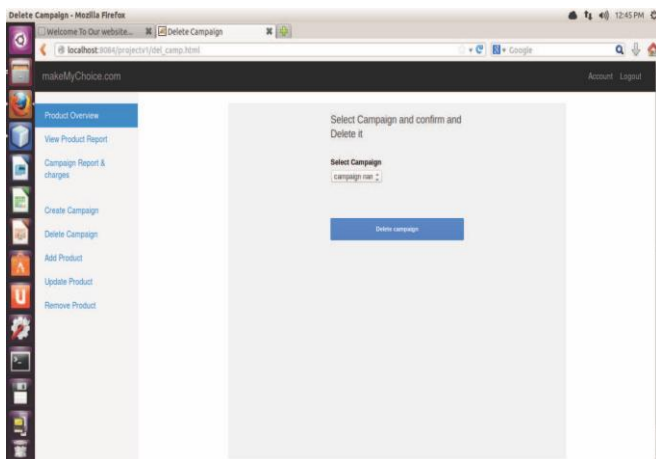


Figure 4: Create Campaign

4. Add Product

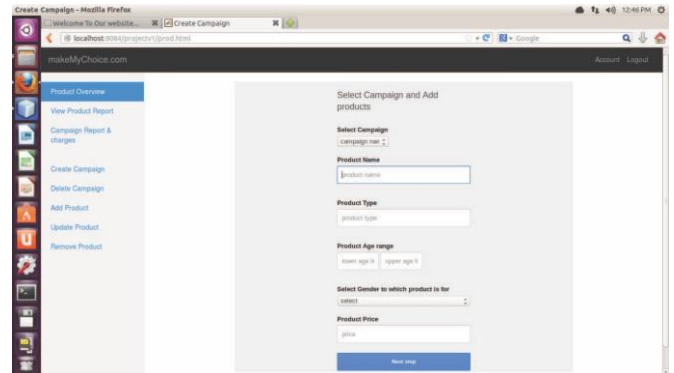


Figure 5: Add Product

5. Delete Product

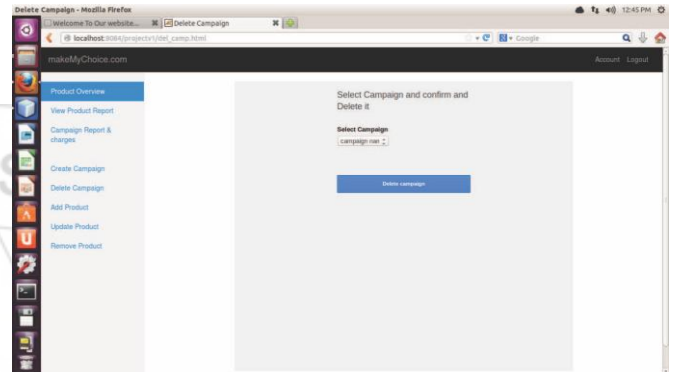


Figure 6: Delete Product

6. Report of Products



Figure 7: Report of Products

6. Conclusion

As the world is turning towards use of internet for every day-to-day activity, need for viewing and selecting products of one's choice is of prime importance. The list of irrelevant advertisements frustrates the user, which proves to be the main reason for failures of most sites. But our website makes it smooth for the users to select products by filtering available products based on individual customer's interests.

The e-commerce field is emerging rapidly. Advertisers need a way to promote their products in market. This way is provided by personalized websites like this one. The reports provided by our website makes it easier for them to know the status of their products and hence take necessary measures in order to come up for the faced losses.

To summarize, our website is working as an adapter between the advertiser and customer

References

- [1] Extraction of user profile based on the hadoop framework. IEEE, 2009. ZHAI Yan-dong YANG Bin HUANG Lan*, WANG Xiao-wei.
- [2] Multicluster Hadoop Distributed File System
- [3] Tomaši, J. Ugovšek, A. Rashkovska and R. Trobec Jožef Stefan Institute/Department of Communication Systems, Ljubljana, Slovenia Ivan.Tomasic@ijs.si, janez@ugovsek.info, Aleksandra.Rashkovska@ijs.si, Roman.Trobec@ijs.si J2EE Framework Perspective for Security Augmentation Pawan Kumar Verma, Student Member, IEEE Neetu Singh, Student Member, IEEE Department of Computer Science & Engineering Jaypee University of Information Technology, Wagnaghat, India. E-mail: abes.pawan@gmail.com, neetu.tanwar29@gmail.com.
- [4] Automated Detection of Client-State Manipulation Vulnerabilities Anders Møller Department of Computer Science Aarhus University amoeller@cs.au.dk Mathias Schwarz, Department of Computer Science Aarhus University schwarz@cs.au.dk
- [5] S.N. Srirama, P. Jakovits, and E. Vainikko, "Adapting scientific computing problems to clouds using MapReduce," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 184-192, 2012.
- [6] "Improving the storage efficiency of small files in cloud storage", S. Yu, X. Gui, R. Huang, and W. Zhuang *Hsi-An Chiao Tung Ta Hsueh/Journal of Xi'an Jiaotong University*, vol. 45, no. 6, 2011.
- [7] ZHONG Zi-chang. "Constructing distributed storage and distributed computing cluster with Hadoop", <http://www.zhongzichang.com/archives/83>.