Abstract

There is an invariable progress and extension of the World Wide Web which has resulted into the generation of log files having enormous magnitude of data. Log files incorporate traits of user behavior, therefore it is essential to analyze log data and acquire knowledge from it. Web mining techniques primarily focuses on deciphering and scrutinizing the navigational behavior of user form various aspects and ascertaining the hidden knowledge from these web logs. As log files over the web are outsized, storage becomes a constraint wherein effective techniques such as virtual database prove to be ineffectual for the same. Conversely, Hadoop offers a large scale distributed batch processing infrastructure that provides adequate data storage,
Trend Analysis based on Access Pattern over Web Logs using Hadoop

distributive and analogous processing, isolation of process and fault tolerant on occurrences of data loss. This paper characterizes on the dominant approach for managing the large chunk of web log data using Hadoop MapReduce which reduces the response time for throughput generation, loads the log data effectively and ensures reliability. The primary focus of the paper is to construct log analysis system which depicts trends based on the users browsing mode using Hadoop MapReduce which facilitates handling of heterogeneous query execution on log file.

References

- Praveen Kumar, Dr Vijay Singh Rathore, "Efficient Capabilities of Processing of Big Data using Hadoop Map Reduce", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014
- Anuja Pandit, Amruta Deshpande, Prajakta Karmarkar, "Log Mining Based on Hadoop’s Map and Reduce Technique", International Journal on Computer Science and Engineering (IJCSE) Vol. 5 No. 04 Apr 2013

Index Terms

Computer Science
Web Services

Keywords

Cloudera Hadoop MapReduce Log Files Web Mining MySql Database
Hadoop Distributed File System

Trend Analysis.