Analyzing Web Access Logs using Spark with Hadoop

International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

Volume 180
-
Number 1

Year of Publication: 2017

Authors:
Vandita Jain, Tripti Saxena, Vineet Richhariya

10.5120/ijca2017915904

Abstract

Web usage mining is a process for finding a user navigation patterns in web server access logs. These navigation patterns are further analyzed by various data mining techniques. The discovered navigation patterns can be further used for several things like identifying the frequent patterns of the user, predicting the future request of user, etc. and in the recent years there are huge growth in electronic commerce websites like flipkart, amazon, etc. with an huge amount of online shopping websites, it is necessary to notice that how many users are actually reaching to the websites. When user’s access any online website, web access logs are generated on the server. Web access logs data helps us to analyze user behavior that contain information like ip address, user name, url, timestamp, bytes transferred. It is very meaningful to analyze the web access logs which helps us in knowing the emergency trends on electronic commerce. These ecommerce websites generates petabytes of log data every day which is not possible by traditional tools and techniques to store and analyze such log data. In these paper we proposed an hadoop framework which is very reliable for storing such huge amount of data in to HDFS and than we can analyze the unstructured logs data using apache spark framework.
Analyzing Web Access Logs using Spark with Hadoop

to find user behaviour. And in these paper we can also analyze the log data using mapreduce framework and finally we can compare the performance on spark and mapreduce framework on analyzing the log data.

References

1. Dr.S.Suguna, M.Vithya, J.I.Christy Eunaicy, “Big Data Analysis in E-commerce System Using HadoopMapReduce” in 2016 IEEE.

Index Terms

Computer Science Wireless
Keywords

Hadoop, HDFS, Mapreduce, Log analysis, spark, user behaviour.