**Mining of Web Server Logs Using Big Data Technologies**

**Abstract** : Big Data is an emerging growing dataset beyond the ability of a traditional database tool. Hadoop rides the big data where the massive quantity of information is processed using cluster of commodity hardware. Web server logs are semi-structured files generated by the computer in large volume usually of flat text files. It is utilized efficiently by Mapreduce as it process one line at a time. This paper performs the session identification in log files using Hadoop in a distributed cluster. Apache Hadoop Mapreduce a data processing platform is used in pseudo distributed mode and in fully distributed mode. The framework effectively identifies the session utilized by the web surfer to recognize the unique users and pages accessed by the users. The identified session is analyzed and statistical report is generated .

**Project Guide :**

Dr. Bipin Bihari Jayasingh

**Project Team :**

Ashish R Francis -11B81A1214

T Karthik -11B81A1226

B R Rishinder Kumar -11B81A1259