**Description:** This Project aims to collect, aggregate, process, and report the web server logs. Web server logs contain the information about user clicks, user demographics and products information. We used this data to derive some insights like which products are doing well, which pages are popular, which users are active, which users are inactive, page clicks count, time spent by each user (per month, per week, per day, per year) etc. This POC objective is to use the Open Distributed Computing Platform like Apache Hadoop and Hadoop Ecosystem to ingest data, to store data on commodity hardware, to process data in parallel and finally to report the processed data.

**The Process flow:**

* We collected some few Giga bytes log data.
* We stored this data in Hadoop Distributed File System.
* We wrote MapReduce programs, Pig scripts for processing.
* We reported this processed data to other Systems.

**Responsibilities:**

* Requirements gathering
* Hadoop Cluster Setup
* Data Movement from Web Servers to Hadoop Cluster.
* Written a few MapReduce Programs and Pig scripts