|  |  |  |
| --- | --- | --- |
| S. No. | Task | Description |
| 1 | Problem statement | Analyze Apache Web Logs to convert it into structured format and extract KPI information. |
| 2 | Solution Architecture |  |
| 3 | Software/Tools specification | - WinSCP for Copying data over to CloudX  - HDFS  - Spark Application |
| 4 | Solution Description: Steps for processing and analysing log data | Files copied in HDFS:    Read Files into Hadoop    Code for Parsing the Data and transforming it into Structured format. Then using SQLContext for querying the data for KPIs    Output processed data into HDFS    KPI: Category wise Count    KPI: Sub-Category Wise Count    KPI: Request Status Count    KPI: Unique IP Count    KPI: Load the data into MySQL   * Create empty table in MySQL      * Sqoop command to export the HDFS data dir      * Data Loaded into MySQL |