**Log Analysis**

**1.0**

# Contents

[1. Contents 1](#_Toc383539260)

[2. Change History 2](#_Toc383539261)

[3. Object 2](#_Toc383539262)

[4. Purpose 2](#_Toc383539263)

[5. Non-Functional Requirements 2](#_Toc383539264)

[6. Versions 3](#_Toc383539265)

[7. System Configuration 3](#_Toc383539266)

[8. Workflow diagram 4](#_Toc383539267)

# Change History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Change Purpose** | **Changed By** | **Version** | **Contact** |
| 24-Mar-14 | Initial Version | Murthy | 1.0 | Ksn.moorthy@gmail.com |
|  |  |  |  |  |

# Object

This document outlines the problem statement of the Log Analysis project.

# Purpose

The following data will be pulled out from the logs.

* Total hits for a particular day.
* Total hits from a particular IP address
* IP address wise total hits.
* Which IP address has made more number of hits?
* Which day has made more number of hits?
* Max number of hits IP address of
* Browser wise total hits.
* Date wise total products sold.

# Non-Functional Requirements

* Flume - Data Ingestion
* Oozie - Work flow
* Map Reduce - Pull the required data and convert into desired format
* Hive - Load the converted data into Hive tables and perform queries
* HBase - Maintain the data that will be available to dashboards
* A Web technology - Display the analyzed data

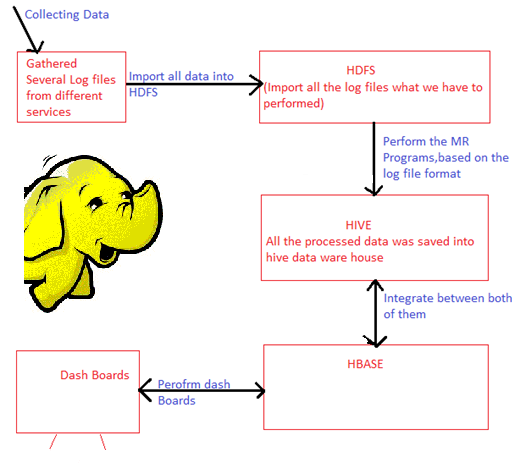
# Versions

* Hadoop - 1.0.4
* Flume -
* Oozie - 3.3.1
* Hive - 0.9
* CentOS - 5

# System Configuration

* 4 Node Cluster
* 8 GB RAM
* 1 TB
* Quad Core Processor

# Workflow diagram



More details are to be updated please correct above details if any, and add details. Don’t forget update change history.