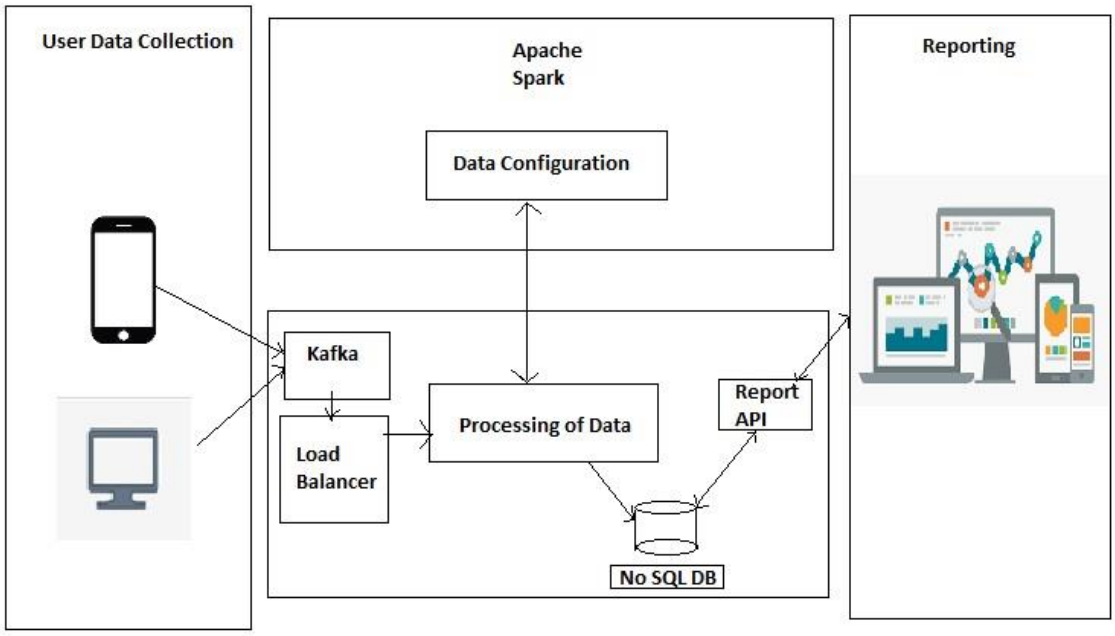
Design of Google Analytic like Backend System



This Google Analytic like Backend System has main four components:

1. **User Data Collection**

   Collects all user interaction data

   Google analytics use ga.js file which is load from the Google web server and then sets variable with user's account number. Similarly we need JavaScript file which provide user interaction data to our processing server. This file will be cache in user local memory so it will not affect performance during initial loading. If cookies are enabled in client side than we can also set important information like timestamp, device info, etc. into it. We also need some Publisher /Subscriber system which allows users to subscribe to it and publish data to any number of systems. We can use Apache Kafka for this purpose which provide a unified, high-throughput, low-latency platform for handling real-time data feeds.

2. **Processing of Data**

   Processes the user-interaction data with the configuration data

   We need to use Load balancer to improve the performance and reliability of a server environment by distributing the workload across multiple servers. Increased reliability is also achieved through this manner, in case some of backend servers become unavailable. We need database which can handle large amount of data still provide superior performance and highly

scalable. So will use NoSQL database such as MongoDB, CouchDB, Redis, or Apache Cassandra. This NoSQL Database will use to store this processed data and get when needed for report generating API.

3. **Data configuration**

   Manage how the data is processed

   We can write different micro service to define how the data is processed as per business requirement. We can use Apache Spark to provide an interface for programming entire clusters with implicit data parallelism and fault tolerance. It also used in reprocessing of data.

4. **Reporting**

   Show all the processed data in required format

   Create custom API to fetch the data from the No SQL database and show the report using third party report generating tools like JasperReport, BIRT, ReportServer etc.