

<PayPay Analytics >

# High-Level Technical Design

Version 1.0

Author: Juan Lu

09/16/2019

Document Number: <a000001>

Contract Number: <b000001 >

Table of Contents

[1. Introduction 1](#_Toc441656787)

[2. The System 2](#_Toc441656788)

[2.1 Functional Description 2](#_Toc441656789)

[2.2 User Community Description 2](#_Toc441656790)

[2.3 Technical Architecture 2](#_Toc441656791)

[3. Goals, Objectives, and Rationale for New or Significantly Modified System 3](#_Toc441656792)

[3.1 Project Purpose 3](#_Toc441656793)

List of Figures

[Figure 1 - High-Level Architecture](#_Toc441656833) 3

List of Tables

[Table 1 – Module and Technologies 6](#_Toc441656834)

## Introduction

Design Question: Design A Google Analytic like Backend System. We need to provide Google Analytic like services to our customers. Pls provide a high level solution design for the backend system. Feel free to choose any open source tools as you want.

The system needs to:

1. handle large write volume: Billions write events per day.
2. handle large read/query volume: Millions merchants want to get insight about their business. Read/Query patterns are time-series related metrics.
3. provide metrics to customers with at most one hour delay.
4. run with minimum downtime.
5. have the ability to reprocess historical data in case of bugs in the processing logic.

## The System

Instructions: If applicable, this section describes the current system that is being replaced, enhanced, or upgraded.

### Functional Description

Paypay Analytics is implemented with "page tags", in this case, called the Google Analytics Tracking Code, which is a snippet of JavaScript code that the website owner adds to every page of the website. The tracking code runs in the client browser when the client browses the page (if JavaScript is enabled in the browser) and collects visitor data and sends it to a Paypay data collection center server as part of a request for a web beacon.

The file does not usually have to be loaded, however, due to browser caching. Assuming caching is enabled in the browser, it downloads ga.js only once at the start of the visit. Furthermore, as all websites that implement Google Analytics with the ga.js code use the same master file from Google, a browser that has previously visited any other website running Google Analytics will already have the file cached on their machine.

In addition to transmitting information to a Google server, the tracking code sets a first party cookie (If cookies are enabled in the browser) on each visitor's computer. This cookie stores anonymous information called the ClientId. Before the launch of Universal Analytics, there were several cookies storing information such as whether the visitor had been to the site before (new or returning visitor), the timestamp of the current visit, and the referrer site or campaign that directed the visitor to the page (e.g., search engine, keywords, banner, or email)..

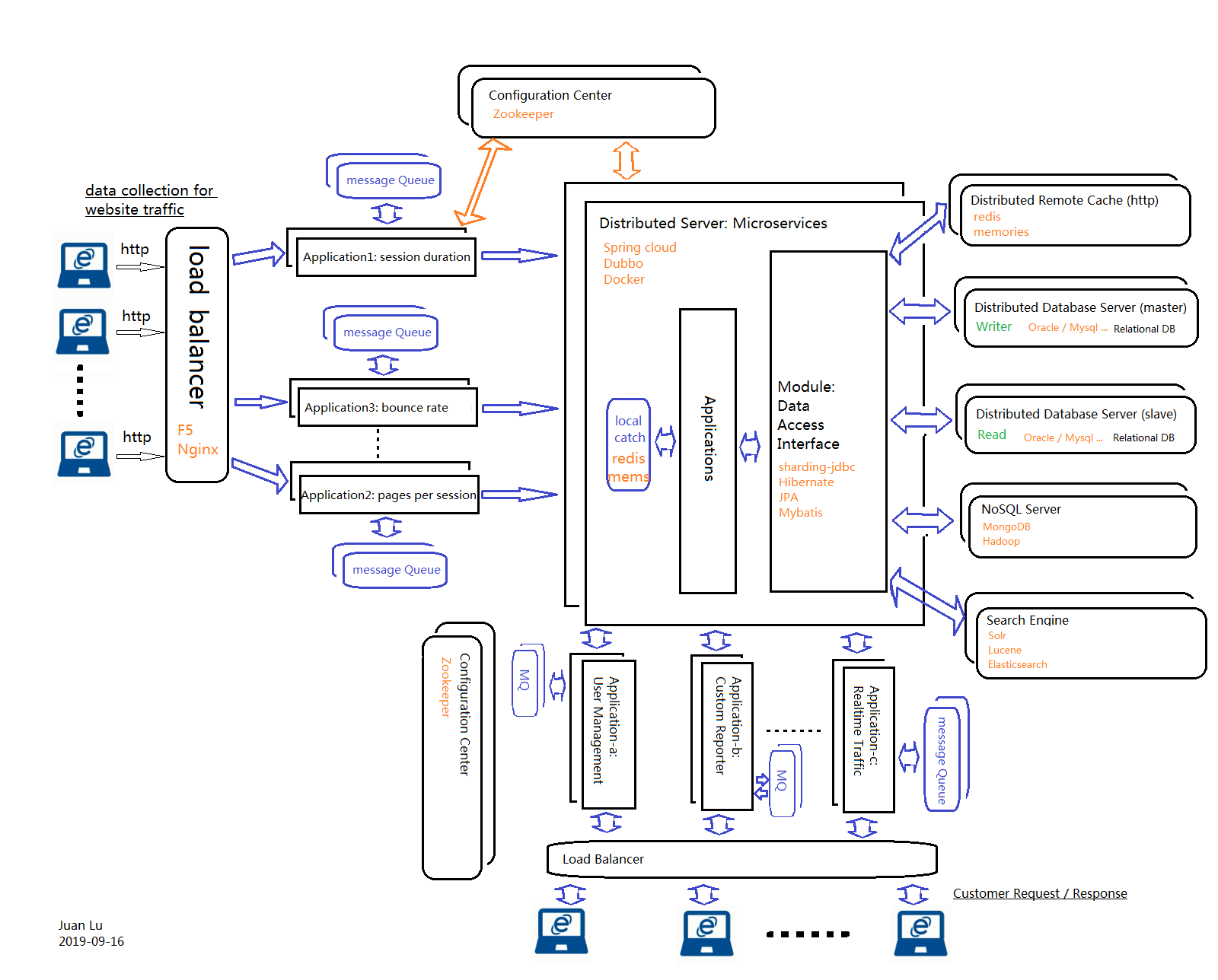
*The customers of Paypay Analitics could use the panel to monitor the traffic and obtain the report of the analytics data of the website based on those collected data from the website users.*

### User Community Description

The one who need track and analyze the traffic of a website could be our customer.

### Technical Architecture

The technical architecture of the system see the below chart. Arrow shows the flow of data.



## Goals, Objectives, and Rationale for New or Significantly Modified System

Providing customer straightforward analyzed traffic data reports through a visualized panel with charts and tables.

### Project Purpose

Analyzed the traffic conditions of website, for example, the frequency of website usage, the usage peak along time, etc. to assist the customer make the market goal and directions, etc.

Table – Module and Technologies

| User Group | Module | Technologies |
| --- | --- | --- |
| <Web User> | Load balancer | F5 , nginx |
| <Web User> | cookie | javascript |
| <Web Owner> | panel | Html, css, javascript |
| / | microserice | Spring cloud, Dubbo, Docker |
| / | Configuration center | zookeeper |
| / | cache | Redis, memories |
| / | Search engine | Lucene, solr, elastic search |
| / | No SQL | MongoDB, Hadoop |
| / | Big data analysis | python |
| / |  |  |