**基于大数据平台的电商系统搭建**

**2020年5月**

# 目录

[1 项目概述 3](#_Toc39272580)

[2 建设目标及原则 3](#_Toc39272581)

[2.1 建设目标 3](#_Toc39272582)

[2.2 建设原则 3](#_Toc39272583)

[3 需求分析 4](#_Toc39272584)

[3.1 概述 4](#_Toc39272585)

[3.2 功能需求 4](#_Toc39272586)

[3.2.1 GMV成交总额 4](#_Toc39272587)

[4 概要设计 4](#_Toc39272588)

[4.1 概述 4](#_Toc39272589)

[4.2 系统框架设计 4](#_Toc39272590)

[4.2.1 技术架构设计 4](#_Toc39272591)

[4.2.2 软件功能结构设计 5](#_Toc39272592)

[4.3 应用系统设计 5](#_Toc39272593)

[4.3.1 设计原则 5](#_Toc39272594)

[4.3.2 功能模块设计 5](#_Toc39272595)

[4.4 接口设计 5](#_Toc39272596)

[4.4.1 外部接口 5](#_Toc39272597)

[4.4.2 内部接口 5](#_Toc39272598)

[5 详细设计 5](#_Toc39272599)

[5.1 系统概述 5](#_Toc39272600)

[5.2 系统数据流程 6](#_Toc39272601)

[5.3 功能设计 6](#_Toc39272602)

[5.3.1 GMV成交总额 6](#_Toc39272603)

[6 系统关键技术 6](#_Toc39272604)

[6.1 Hadoop平台 6](#_Toc39272605)

1. **项目概述**

现如今，随着人们对互联网行业的深入研究，互联网技术在近几年得到了极大的普及，电商网站的产生更是让互联网逐步成为人们日常生活中必不可少的一部分，它便利了人们的生活，提高了人们的生活品质。

随着电商网站日益增长的用户体验，数据量增长势头也是只增不减，对这些海量数据进行加工处理使其成为有用的数据成为了一个巨大的挑战，但同时也带来了巨大的技术创新和商业机遇。正确的数据分析结果可以帮助网站能对自身得到更全新的认识，然后对网站的不足之处做出正确的判断，及时纠正，为网站创造更多的收益价值。

1. **建设目标及原则**
   1. **建设目标**

本系统是基于Hadoop框架，围绕Nginx服务器、MySQL数据库、Hive数据仓库、HighCharts图库表、JavaScript脚本语言、CSS样式表等众多前后端技术，实现电商网站用户日志数据的解析处理，最终生成一系列的数据指标，并对此进行可视化展示，为电商网站的改进提供一定的参考。

总之，该系统建设的总体目标是：

* 对电商日志数据进行处理；
* 分析和挖掘电商日志数据信息；
* 为电商网站带来巨大的商业价值；
* 提高电商网站的核心竞争力。
  1. **建设原则**

**开放性原则：**

运用开放的标准规范整合各类系统，为实现系统信息交换和共享提供标准接口。

**集约性原则：**

充分利用已有建设成果，辅以必要的补充完善，逐步提升系统智慧化水平。

**复用性原则：**

采用模块化和平台化研制方法构建应用服务体系，提高软件复用和重用能力，方便建设成果推广。

1. **需求分析**
   1. **概述**
   2. **功能需求**
      1. **GMV成交总额**

GMV（Gross Merchandise Volume），是一定时间段内的成交总额（比如一天、一个月、一年），包括付款和未付款的部分。

1. **概要设计**
   1. **概述**
   2. **系统框架设计**

* + 1. **技术架构设计**

系统主要由三部分组成，分别为数据采集传输模块、数据存储计算模块和数据展示模块。

数据采集传输模块：主要由nginx、Flume、Sqoop等工具组成，负责将用户操作的各项数据进行收集传输。

数据存储计算模块：主要由Hadoop框架组成，负责过滤、解析上一步收集到的数据信息，并保存计算后的结果。

数据展示模块：主要由JavaScript、HTML、CSS组成，负责对平台展示界面的编写以及调取数据api接口，将数据分析结果进行展示。

* + 1. **软件功能结构设计**

系统按照功能可以分为

* 1. **应用系统设计**
     1. **设计原则**
     2. **功能模块设计**
  2. **接口设计**
     1. **外部接口**
     2. **内部接口**

1. **详细设计**
   1. **系统概述**

面向服务的体系结构（SOA）是一个组件模型，具备敏捷性、重用性和低耦合性等特点，使得它能很好地应用于智慧营区“多维立体智能管控平台”，它将应用程序的不同功能单，通过这些服务之间定义良好的接口和契约联系起来。基于SOA架构技术的多维立体智能管控平台划分为感知层、传输层、服务层和应用层构成，通过整合各个服务层，对多维立体智能管控平台的核心业务的变化做出快速反应，呈现出可以支持有机业务框架的能力

在本系统中，基础接入子系统包括卡口子系统、共享单车子系统、摆渡车子系统、绿化灌溉子系统、停车引导子系统、微信公众号和手机APP、公共信息显示子系统、公共区域无线接入子系统、背景音乐子系统。

智慧服务系统通过人、车、物、环境有机的结合在一起，彼此之间数据互通，构建一套从惠民需求到惠民服务的一整套服务体系。

* 1. **系统数据流程**

系统

* 1. **功能设计**
     1. **GMV成交总额**
        1. 系统设计

1. **系统关键技术**
   1. **Hadoop平台**

Hadoop是一个由Apache基金会所开发的分布式系统基础架构。用户可以在不了解分布式底层细节的情况下，开发分布式程序。充分利用集群的威力进行高速运算和存储。Hadoop实现了一个分布式文件系统（Hadoop Distributed File System），简称HDFS。HDFS有高容错性的特点，并且设计用来部署在低廉的（low-cost）硬件上；而且它提供高吞吐量（high throughput）来访问应用程序的数据，适合那些有着超大数据集（large data set）的应用程序。HDFS放宽了（relax）POSIX的要求，可以以流的形式访问（streaming access）文件系统中的数据。Hadoop的框架最核心的设计就是：HDFS和MapReduce。HDFS为海量的数据提供了存储，而MapReduce则为海量的数据提供了计算