曙

珠江数码大数据分析平台

总体设计方案

项目编号：＿＿＿＿＿＿＿＿＿＿

项目名称：珠江数码大数据收视分析平台项目

填报单位：曙光信息产业（北京）有限公司

填报日期： 2015 年 10月 19日

验收单位：

验收日期：年月日（预）

年月日（终）

目录

[1 项目概述 1](#_Toc461195519)

[1.1 项目背景 1](#_Toc461195520)

[1.2 系统建设原则 1](#_Toc461195521)

[1.3 系统建设目标 2](#_Toc461195522)

[2 大数据平台总体设计 4](#_Toc461195523)

[2.1 平台整体架构 4](#_Toc461195524)

[2.2 系统关键技术指标 5](#_Toc461195525)

[2.2.1 数据共享平台 5](#_Toc461195526)

[2.2.2 收视行为分析 5](#_Toc461195527)

[2.2.3 用户画像及个性化服务 5](#_Toc461195528)

[2.2.4 智能运维 6](#_Toc461195529)

[3 大数据平台设计 6](#_Toc461195530)

[3.1 基础硬件 6](#_Toc461195531)

[3.2 数据存储 7](#_Toc461195532)

[3.2.1 分布式文件系统 7](#_Toc461195533)

[3.2.2 全文检索库 8](#_Toc461195534)

[3.3 数据处理 9](#_Toc461195535)

[3.3.1 内存计算 9](#_Toc461195536)

[3.3.2 索引引擎 11](#_Toc461195537)

[3.3.3 数据访问接口 12](#_Toc461195538)

[3.4 数据接入 13](#_Toc461195539)

[3.4.1 格式转换 13](#_Toc461195540)

[3.4.2 数据清洗 14](#_Toc461195541)

[3.4.3 数据关联 14](#_Toc461195542)

[3.4.4 数据校验 14](#_Toc461195543)

[4 大数据管理系统设计 14](#_Toc461195544)

[4.1 系统功能 14](#_Toc461195545)

[4.1.1 便捷的部署 14](#_Toc461195546)

[4.1.2 完善的管理 15](#_Toc461195547)

[4.1.3 全面的监控 15](#_Toc461195548)

[4.1.4 健全的告警 15](#_Toc461195549)

[4.1.5 多样的报表 16](#_Toc461195550)

[4.1.6 易用的调参 16](#_Toc461195551)

[4.1.7 快速的诊断 16](#_Toc461195552)

[4.1.8 良好的扩展 17](#_Toc461195553)

[4.2 系统设计 17](#_Toc461195554)

[4.2.1 架构设计 17](#_Toc461195555)

[5 收视行为分析系统 18](#_Toc461195556)

[5.1.1 常规收视指标分析 18](#_Toc461195557)

[5.1.2 基于用户的收视指标分析 20](#_Toc461195558)

[5.2 收视行为分析系统技术架构 22](#_Toc461195559)

[5.2.1 前端界面技术架构 22](#_Toc461195560)

[5.2.2 后台服务技术架构 23](#_Toc461195561)

[5.3 收视行为分析系统处理流程 28](#_Toc461195562)

[6 精准营销 30](#_Toc461195563)

[6.1 精准营销系统功能 30](#_Toc461195564)

[6.1.1 用户画像 30](#_Toc461195565)

[6.1.2 用户群体划分 30](#_Toc461195566)

[6.1.3 个性化推荐 30](#_Toc461195567)

[6.2 精准营销系统流程 31](#_Toc461195568)

[6.2.1 用户画像 31](#_Toc461195569)

[7 智能运维 32](#_Toc461195570)

[7.1 智能运维系统功能 32](#_Toc461195571)

[7.1.1 终端管理 32](#_Toc461195572)

[7.2 智能运维系统架构 32](#_Toc461195573)

[7.3 智能运维系统流程 33](#_Toc461195574)

[7.3.1 终端管理 33](#_Toc461195575)

1. 项目概述
   1. 项目背景

广州珠江数码集团有限公司（简称“珠江数码集团”）成立于1993年，是经广州市委、市政府授权，专门负责建设、维护和管理广州市行政区域内广播电视网络的主要运营机构。经过十多年的发展，珠江数码集团已成为华南地区最大规模的广播电视网络运营商之一，拥有有线电视用户超过260万户，宽带用户超过16万户，高清互动电视用户数正处于快速发展阶段。

珠江数码集团已建成完整覆盖广州各区（县级市）的有线传输与无线传输互为延伸、互为补充的广电宽带信息网络，实现了城区全程全网的双向覆盖，为广大市民提供有线数字电视、互联网接入服务（珠江宽频）、高清互动电视（Candytime甜果时光）、移动数字电视、CMMB手机电视、信息内容集成等多样化、跨平台的信息服务。

珠江数码集团正在从传统的有线电视运营商向“多媒体网络和信息服务运营商”作战略转型，集团下设综合服务、技术支持、市场、投资和分公司5大序列，共设15个部门、中心，7家分公司；旗下还拥有珠江在线多媒体信息有限公司和广州珠江数字电视及数字家庭技术应用研究院有限公司2家全资子公司，广州珠江数码番禺有限公司、花都有限公司、从化有限公司、增城有限公司4家控股子公司，珠江移动多媒体信息有限公司和南沙信息港有限公司2家参股子公司，已实现了集团化规模经营格局。

当前，珠江数码集团正紧紧把握“三网融合”所带来的发展机遇，加快高清交互数字电视的推广，将家庭电视机变成多媒体信息终端，为广大用户带来全新的数字媒体体验，并为广州信息化、数字化的发展提供新的途径，为政府部门架设通达市民的便捷通道。

由于众多新业务的推出和普及，需要建设一个大数据分析平台，通过这个平台可以整合集团各个大系统的所有和用户相关的数据，建立一套完整的数据分析及数据挖掘的系统平台，用于了解用户需求，促进各项业务的发展。

* 1. 系统建设原则

大数据平台利用普通软硬件设备，搭建存储和分析平台环境。

为了使大数据平台成为未来的数据知识运营平台，需要对数据资源进行整合以提供统一的数据访问入口，为应用提供服务。同时对于关键数据提供备份支持。对于大数据知识网络访问方式可以支持REST/SOAP接口，同时也需要提供基于REST/SOAP的管理接口。

平台软硬件设计遵循开放、兼容、高扩展性、高安全性的原则。从而使整个系统结构合理、技术先进、易于扩展，既能满足目前的信息服务需求，又能符合发展的要求。

架构松耦合，应用紧耦合：为了满足系统灵活扩展，架构上采取松耦合设计的方针，为了高效配合应用，应用关联紧密耦合。

高性能：为满足大数据系统的性能需求，针对不同子系统的技术特点，选择合适的技术，突破性能瓶颈、提升单台设备的处理性能。

高可扩展性：系统应具备良好的扩展能力，在适应当前需求的基础上，尽量为将来可预见和不可预见的功能、性能扩充留有余地；新功能、新业务的增加能够在不影响系统运行的情况下实现。在本方案设计中，针对不同子系统特点，采用分布式计算、SOA架构等技术提高系统的扩展性，将弹性计算理念贯穿于设计始终。

高可用性：在方案设计中，采用HA、负载均衡、副本式存储计算等冗余技术，从系统级、设备级、数据级和业务级等多个层面进行高可用设计，以提高整体可用性。保证主机、操作系统、网络、数据库、应用软件等系统7\*24小时平稳运行。对系统故障和运行环境变化，有较好的容错能力，能够在故障和灾难等情况下，最大程度地保持系统的基本功能。

高可管理性：充分考虑到系统管理对大系统上线运维的重要性，选择有过大规模设备（数万台）管理成功案例的成熟的管理系统作为系统管理工具，对物理设备进行状态监控、实时告警及审计管理；所有管理系统提供良好的用户操作界面，完备的帮助信息。系统参数的维护与管理通过操作界面实现。有全面、完善、便捷、统一的网管系统和网络性能监控系统，支持多种故障报警，一旦发生问题能够在最短的时间内处理解决。

高效性：在系统设计中，充分考虑了计算设备、存储设备等各种资源整合，实现跨业务应用的设备复用，提高系统中各类设备的使用效率，构建高效系统，保护用户投资。

先进性：在设计中，充分考虑实用和技术发展的趋势，采用先进的技术架构实现系统建设。所选用的服务器、网络设备、海量数据存储与分析系统、基础设施设备、大规模管理系统云操作系统等硬件设备以及支撑平台，都有选择地采用行业中主流并领先的产品和技术。

安全可控性：尽可能采用国产 、自主、可控的产品和技术，以保证系统安全可控。

* 1. 系统建设目标

基于廉价的服务器集群搭建信息资讯平台，改善现有非结构化数据离线服务水平；突破在传统技术手段下无法解决的海量数据存储弹性扩展能力差、备份恢复难度大的问题；利用结构化与非结构化历史数据资源集中优势，为业务提供更集成、更长时间跨度、更多维度的数据查询服务，提升相关业务的查询访问效率，改善客户体验；使离线数据真正上线并得以充分利用，提高大量半结构化非结构化数据利用率，为业务发展提供基础。

系统层面：基于当前的状况和业界的实践与趋势来看，Hadoop作为半结构化、非结构化咨询信息平台具有绝佳的契合度。针对上述的系统特点，我们需要构建基于开源Apache Hadoop的框架的存储，管理和分析Internet级别数据量的半结构化和非结构化数据的强大易用方案，具备企业级别的管理能力，工作流管理，安全管理，可视化挖掘与展现等能力，能与现有基础设施和大数据流计算技术无缝的集成，形成具有领先的数据分析，机器学习以及文本数据分析挖掘能力的大数据处理平台，来更好的应对复杂的，海量数据带来的挑战。

1. **数据共享平台**

整合集团各个大系统的所有和用户相关的数据，包括用户收视行为数据（来自新媒体），用户数据（来自运支中心），用户服务情况（投诉报障工单数据，来自运支中心），用户终端的设备数据（电平、流量等，来自网维中心）等，将上述的数据进行统一存储和检索，为前端应用系统提供丰富的ETL类服务，并支持跨系统数据的分析应用。

1. **收视行为分析和精准营销**

收拾行为分析主要是传统收视行为分析和互联网行业用户画像的结合，依据当前用户的收视习惯数据和用户评论、用户属性、广告点击等相关信息，完成基于排行榜和用户浏览历史的节目和广告推荐，含实时榜单、实时推荐等。

精准营销主要是建立用户档案，为业务部门提供精准化营销、用户挽留、产品推介等市场服务。开发针对每个用户的数据化视图，建立用户的各项数据档案，将其标签化、积分化。这个视图面向一线的营销人员、服务人员使用。他们打开一个用户的视图，可以看到这个用户是否是“高消费能力”用户，是否是“对服务不满”的用户，是否是个“电视活跃用户”等。

该类型应用需要持续改善，主要体现在个性化推荐系统的优化和智能广告插入。个性化推荐的改善主要是新的分析算法的引入，用户画像、用户社交关系、用户群体画像的权重再分配等，从而形成更加精准的推荐覆盖。而智能广告的插入，不仅仅与用户喜好和当前播放节目有关，也与广告定价相关，实现广告资源的精细化管理。

1. **大数据智能运维**

主要是对市场管理、经营分析和运维管理提供支撑。具体来讲，

* 主动服务，实现对客户挽留等业务的支持，提高挽留的成功率，增强社区经理等一线人员的服务效果；
* 经营分析，整合品牌订阅、工单信息等信息，更好地分析、预测用户的消费行为，提高整体效益；
* 运维管理，主要是与Modem质量检测系统、呼叫中心、IT业务系统、中间件等信息源数据的整合和再利用。

1. **数据分析即服务Big Data as a Service**

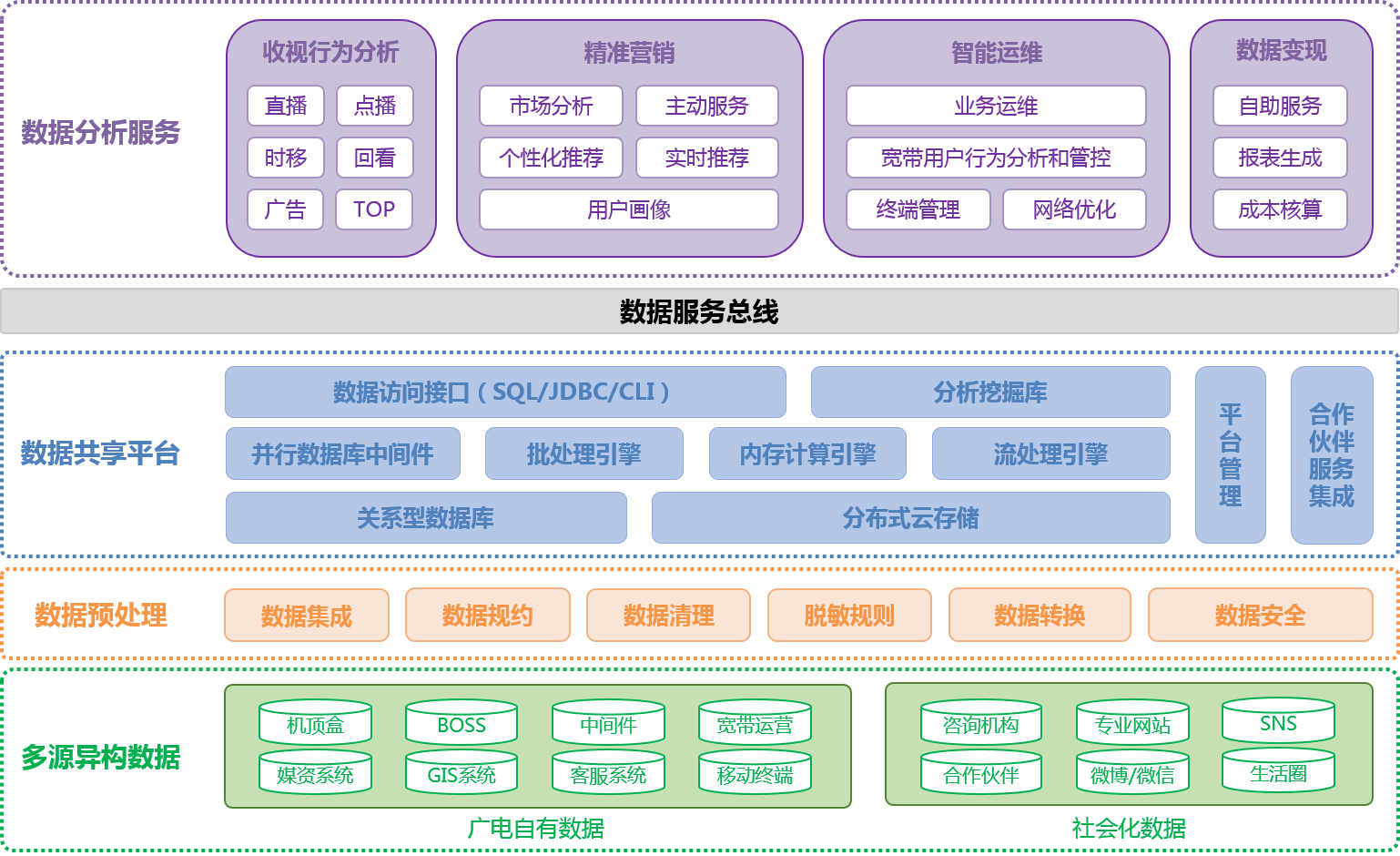
将大数据分析作为一项基础服务提供给各个业务应用。需要完成统一服务框架的建设、标准服务接口的制定、各类业务建模工具的兼容性支持、标准化和高可定制模板设计器等。

大数据分析平台作为珠江数码云计算平台（统一运维平台）的有机组成部分，将与其他公有服务（如转码、打包、技审等）一起，渗透到全网IP化业务流程的各个环节。

本次项目建设，乙方需要完成大数据分析平台的整体部署、调试及基本的POC测试，对各组件选型和角色规划提供相关技术咨询支持；项目合作过程中针对项目功能需求产生的源代码、文档、关键设计或技术的知识产权为甲方独有。

1. 大数据平台总体设计
   1. 平台整体架构

在广泛调研的基础上，建议采用如下的平台整体架构：



数据由多维数据源经过统一的数据抽取和转换平台对数据进行抽取、格式转换、脱敏等操作。通过数据抽取、转换和加载工具，将清洗好的数据（含结构化数据和非结构化数据）加载到数据共享平台中进行存储。由于汇聚了多源异构的数据，数据量非常大，数据共享平台需要具有良好的性能和可扩展性。

数据共享平台中对其内部数据进行最终的数据分析和数据挖掘，要求支持根据不同的业务分析和挖掘需求建立不同的模型，支持自助设计面向具体业务的查询任务流，支持查询分析服务的二次开发，支持增强的数据安全性。

数据分析服务通过建模对数据共享平台中经过初步加工的数据进行进一步的分析挖掘，通过大数据可视化工具对分析挖掘结果进行展现，展现形式要求丰富多样，可以是图表（含柱状图、折线图、雷达图、地图等展现方式），也可以是分析报告等易于理解的形式，方便数据转化为信息，进而为业务决策提供支撑。

* 1. 系统关键技术指标
     1. 数据共享平台
* 实现关系型数据库、分布式文件系统、列式数据库、内存数据库以及全文检索库等多种数据存储方式。
* 实现并行SQL计算框架、流计算框架、内存计算框架、离线计算框架、图计算框架及数据挖掘算法库，并通过资源调度模块对集群的计算资源和处理作业进行统一分配和调度，提供对实时、近实时及离线等多种处理场景的支持。
* 实现数据接入、格式转换、数据清洗、数据关联合并、数据校验、数据传输以及日志记录等功能。
* 实现多种数据访问接口，包括客户端和开发工具包（SDK）等。
  + 1. 收视行为分析
* 终端收视数据整合
* 实现分钟/小时/天/周/月/年多时间粒度统计分析
* 实现节目观看统计分析，广告投放效果分析
* 实现用户收视个性化行为分析，包括开机、观看、搜索等行为分析。
  + 1. 用户画像及个性化服务
* 整合数据资源，基于用户个体进行分析，提取用户特征，针对用户的基本资料、收视行为、消费行为、网络、设备以及喜好等进行分析。
* 。
  + 1. 智能运维
* 终端在线监测与智能判断通过终端信息采集系统、营销系统的数据共享与交互，利用大数据分析技术进行终端故障诊断/预警、设备整体工况分析，实现终端异常分析、故障分析、各类事件分析等功能。

1. 大数据平台设计
   1. 基础硬件

|  |  |  |
| --- | --- | --- |
| **大数据平台数据节点服务器** | 品牌 | 国产服务器品牌 |
| 高度 | 机架式服务器，高度≤4U |
| CPU类型 | 配置≥2颗Intel Xeon E5-2600v3系列CPU，每CPU≥12核，主频≥2.5GHz； |
| 内存槽 | 支持扩展24根内存插槽，最大1.5TB容量 |
| 内存掉电保护 | 支持NV DIMM,硬件设计支持 |
| 内存容量 | 配置≥16\*16GB DDR4 内存 |
| 硬盘槽位 | 提供≥24个硬盘槽位 |
| 硬盘容量 | 配置2块600GB 10K 2.5" 6Gbps SAS热插拔硬盘，配置18块4TB 3.5" 6Gbps SAS热插拔硬盘 |
| RAID卡 | 配置SAS RAID卡，支持Raid0、1、5、6等多种RAID模式 |
| PCI I/O插槽 | 提供≥4个PCI-E 3.0 扩展插槽 |
| 千兆网卡 | 提供≥2个RJ45千兆网口 |
| 万兆网卡 | 提供≥2个万兆以太网口 |
| 管理功能 | 支持IPMI2.0和KVM Over IP功能，独立管理网口 |
| SD卡 | 主板支持SD卡插槽，可实现BMC日志扩充存储 |
| 电源 | 1+1冗余电源 |
| 光驱 | DVD-RW光驱 |
| 硬件管理接口 | 支持标准IPMI智能平台管理接口 |
| 售后服务 | 提供不少于3年原厂商技术支持，提供正式承诺函 |
| **大数据平台计算点服务器** | 品牌 | 国产服务器品牌 |
| 高度 | 机架式服务器，高度≤2U |
| CPU类型 | 配置≥2颗Intel Xeon E5-2600v3系列CPU，每CPU≥12核，主频≥2.5GHz； |
| 内存槽 | 支持扩展24根内存插槽，最大1.5TB容量 |
| 内存掉电保护 | 支持NV DIMM,硬件设计支持 |
| 内存容量 | 配置≥16\*16GB DDR4 内存 |
| 硬盘槽位 | 提供≥12个硬盘槽位 |
| 硬盘容量 | 配置2块600GB 10K 2.5" 6Gbps SAS热插拔硬盘，配置6块240GB SSD硬盘 |
| 硬盘控制器 | 配置SATA硬盘控制器，支持Raid0、1、5、6等多种RAID模式 |
| PCI I/O插槽 | 提供≥4个PCI-E 3.0 扩展插槽 |
| 千兆网卡 | 提供≥2个RJ45千兆网口 |
| 管理功能 | 支持IPMI2.0和KVM Over IP功能，独立管理网口 |
| SD卡 | 主板支持SD卡插槽，可实现BMC日志扩充存储 |
| 电源 | 1+1冗余电源 |
| 光驱 | DVD-RW光驱 |
| 硬件管理接口 | 支持标准IPMI智能平台管理接口 |
| 售后服务 | 提供不少于3年原厂商技术支持，提供正式承诺函 |

* 1. 数据存储

大数据平台需要满足对海量多样性数据的存储需求，包括结构化、半结构化、非结构化数据，并且要满足各类型业务场景的处理需求。

* + 1. 分布式文件系统

分布式文件系统能够支持GB级、TB级的超大文件的存储，同时还需要满足如下特性：

* 系统的存储容量能够动态不停机扩容，扩容时现有系统可以不间断正常运行，不受扩容影响。扩容时无需迁移数据，无需要求用户对数据进行重新分区,避免硬盘和数据损坏。
* 具有存储分级特性，至少支持SSD和SATA两级存储，同时能够支持第三方磁带库备份。
* 支持安全认证及文件访问权限控制，对存放在分布式文件系统中的数据根据认证用户进行读/写访问控制。
* 支持文件夹级别的加密，对应用级透明访问。
* 能够提供容错机制，数据自动保存多个副本，能够支持设置数据副本，当副本丢失后，能够自动恢复副本。
* 分布式文件系统中不允许存在单点故障，任何一个角色故障后都应有备份机器承担原失效节点工作。
* 能够支持GB级、TB级文件超大文件的存储。
* 提供数据负载均衡工具，尽量均匀的将数据分布在集群中各个节点上。
* 支持文件删除恢复机制，当用户或应用删除某个文件时，并没有立刻从分布式文件系统中删除，而是将此文件重命名，转移到回收站目录下，只要文件还存在于回收站目录下即可被迅速的恢复。文件在回收站中的保存时间是可配置的，当超过这个时间是，将会彻底的删除此文件。
  + 1. 全文检索库

全文索引库就是一组索引文件的集合。分为两部分：索引创建（Indexing）和搜索过程（Search）。

* 全文分词检索

全文检索的方法主要分为按字检索和按词检索两种。

1）面对英文文档，选用按字检索方式。对文章中每个字创建索引。

2）对于中文没有空格作为断词标志，所以需要按词检索方法。

* 索引创建

全文检索库对一系列被索引文件（此处所指即数据库数据或者文档数据）构建索引结构。

1. 被索引文件经过语法分析和语言处理形成一系列词(Term)。
2. 经过索引创建形成词典和反向索引表。
3. 通过索引存储将索引写入硬盘。

* 搜索过程

1. 用户输入查询语句。
2. 对查询语句经过语法分析和语言分析得到一系列词(Term)。
3. 通过语法分析得到一个查询树。
4. 通过索引存储将索引读入到内存。
5. 利用查询树搜索索引，从而得到每个词(Term)的文档链表，对文档链表进行交差，并得到结果文档。
6. 将搜索到的结果文档对查询的相关性进行排序。
7. 返回查询结果给用户。

* 索引文件的维护

支持对索引文件的插入、删除、更新等操作，由于更新操作代价较高，一般不进行更新操作，而是使用“先删除，后创建”的方式代替更新操作。

* 索引文件的存放

支持将索引文件存放在本地磁盘目录和分布式文件系统目录中。

* 1. 数据处理
     1. 内存计算

内存计算实质上就是CPU直接从内存而非硬盘上读取数据，并对数据进行计算、分析。是对传统数据处理方式的一种加速，是实现商务智能中海量数据分析和实施数据分析的关键应用技术。

1. 计算框架

内存计算服务采用集群模式，包括计算节点和管理节点，管理节点支持主备高可用模式，集群节点支持线性扩展，扩展节点线性提升集群性能。作业运行过程中，管理节点的主备节点可以进行切换，计算节点宕机，网络异常不能导致作业的运行失败。

采用基于内存的计算引擎，作业中尽量减少持久化的过程以充分利用内存计算的优势。提供丰富的计算原子，在此基础上通过对计算原子的合理组合，构建出一张有向无环图（DAG），经过逻辑计划和物理计划的优化后最终形成作业。

计算节点子任务计算失败后，根据作业子任务间的依赖关系能够重新计算失败的子任务并保证数据的完整性，正确性。对于运算时间很长，运算量大等，有一定概率出错的的作业，通过提供checkpoint检查点，将内存中的中间结果持久化到硬盘上，当子任务失败需要重新计算时，就可以从整个任务的checkpoint处计算，避免耗时作业重新计算，产生不必要的的时间消耗。对于经常访问的数据（迭代算法），可以通过显示调用方法的方式将数据缓存到内存中。

1. 存储级别

数据处理时提供多种存储级别，MEMORY\_ONLY、MEMORY\_AND\_DISK、DISK\_ONLY，以满足不同机器配置的应用。

1. 计算原子

提供对内存，cpu等计算资源的参数配置。并提供一个容错的、并行的数据结构，将数据，中间结果根据存储级别存储到磁盘和内存中，并能控制数据的分区。

基于此抽象数据结构，提供丰富的数据处理原子，包括像分组，聚合，关联，过滤等常用操作，同时支持用户自定义函数实现的类似map，reduce函数，通过合理的组合以支持常见的数据运算。

1. 与Hadoop系统集成

能够与Hadoop系统混合部署；支持Hadoop所使用的各种文件格式( 如 Text\line Sequence\RCFile\Avro\Json\Parquet）；支持Snappy, Gzip,Lzo等多种压缩格式；可以直接访问HDFS和HBase中的数据,可利用Hive的metestore中的元数据。

1. 功能支持

提供多种语言的支持:R,python,java

支持通用SQL查询接口：

同时支持分析型和事务型任务；支持多用户高并发查询；支持耗时从毫秒到小时的查询任务（实时查询以及批处理）；支持SQL 92标准；支持ODBC/JDBC；

1. 数据类型

支持多种数据类型

|  |  |
| --- | --- |
| 类别 | 数据类型 |
| 时间类型 | TimestampType， DateType |
| 数字类型 | ByteType，ShortType，IntegerType，LongType，FloatType，DoubleType，DecimalType |
| 字符类型 | StringType |
| 二进制类型 | BooleanType |
| 布尔类型 | BLOB, CLOB, NCLOB, TEXT |
| 复杂类型 | ArrayType ，MapType ，StructType，dataType |

* 数字类型

ByteType：代表一个字节的整数。范围是-128到127

ShortType：代表两个字节的整数。范围是-32768到32767

IntegerType：代表4个字节的整数。范围是-2147483648到2147483647

LongType：代表8个字节的整数。范围是-9223372036854775808到9223372036854775807

FloatType：代表4字节的单精度浮点数

DoubleType：代表8字节的双精度浮点数

DecimalType：代表任意精度的10进制数据。通过内部的java.math.BigDecimal支持。BigDecimal由一个任意精度的整型非标度值和一个32位整数组成

* 字符类型

StringType：代表一个字符串值

* 二进制类型

BinaryType：代表一个byte序列值

* 布尔类型

BooleanType：代表boolean值

* Datetime类型

TimestampType：代表包含字段年，月，日，时，分，秒的值

DateType：代表包含字段年，月，日的值

* 复杂类型

ArrayType(elementType, containsNull)：代表由elementType类型元素组成的序列值。 containsNull 用来指明 ArrayType 中的值是否有null值。

MapType(keyType, valueType, valueContainsNull)：表示包括一组键 - 值对的值。通过keyType表示key数据的类型，通过valueType表示value数据的类型。 valueContainsNull 用来指明 MapType 中的值是否有null值。

StructType(fields):表示一个拥有 StructFields (fields) 序列结构的值StructField(name, dataType, nullable):代表 StructType 中的一个字段，字段的名字通过 name 指定， dataType 指定field的数据类型， nullable 表示字段的值是否有null值。

* + 1. 索引引擎

平台使用的索引引擎为ElasticSearch。

ElasticSearch是一种开放源码的、基于 Lucene Java 的搜索服务器，易于加入到 Web 应用程序中。ElasticSearch提供了层面搜索(就是统计)、命中醒目显示并且支持多种输出格式（包括XML/XSLT 和JSON等格式）。它易于安装和配置，而且附带了一个基于HTTP 的管理界面。可以使用 Solr 的表现优异的基本搜索功能，也可以对它进行扩展从而满足企业的需要。ElasticSearch的特性包括：

1. 高级的全文搜索功能
2. 专为高通量的网络流量进行的优化
3. 基于开放接口（XML、HTTP、JSON）的标准
4. 综合的HTML管理界面
5. 可伸缩性－能够有效地复制到另外一个Solr搜索服务器
6. 通过JVM监控服务器统计信息
7. 近实时性收索
8. 使用XML配置达到灵活性和适配性
9. 可扩展的插件体系

SolrCloud是基于Solr和Zookeeper的分布式搜索方案。是Solr的基于Zookeeper一种部署方式。

* 数据导入

1. 支持使用DIH(DataImportHandler)从数据库导入数据。
2. 支持CSV文件导入，因此Excel数据也能轻松导入。
3. 支持JSON格式文档。
4. 支持XML格式文档。
5. 还能以编程的方式来自定义导入。

* 更新数据

同一份文档重复导入，自动替换字段id相同的数据。

* 删除数据

1. 支持通过字段id来删除指定文档(document)。
2. 支持通过相关查询删除匹配的文档(document)。

* 查询数据

1. 支持通过HTTP的GET请求获取。
2. 支持对查询结果排序。
3. 支持在网页搜索中高亮搜索结果。

* 文本分析

支持对文本中语句分词以及运用各种转换后（如：小写转换、复数移除、词干提取）进行索引。

* 索引字段类型

支持索引字段的数据类型有int、string、long、float、double、text\_en、text\_general、date、boolean。

* + 1. 数据访问接口
       1. 客户端

平台提供命令行客户端ShellCli进行管理。

1. 集群管理

客户端可以查看集群拓扑信息、服务状态信息，进行集群配置和服务管理。

1. 数据查询

通过客户端可以进行数据查询操作，包括结构化数据和非结构化数据。

* + - 1. 开发工具包标准
* RestAPI标准

提供符合Restful标准的服务接口，可以对数据对象进行管理和查询，进行作业提交和调度，读取结果集等操作。

* 1. 数据接入

为了提供企业级数据接入功能，要求数据接入平台稳定可靠、性能良好、具有良好的可扩展性、易于管理和维护，能够实现数据可靠、安全、高效的接入Hadoop大数据平台。

数据接入的源根据不同的业务需求有不同的接入方式。

1. 对于大数据的实时数据接入，需要使用消息队列作为缓冲，防止数据的丢失，然后在流处理框架中拉取数据，进行数据的基本处理，以完成数据的接入。这时对数据接入平台的要求是：数据不丢失、安全可靠、实时处理。
2. 接入普通的本地文本数据。对本地文件的处理要求可以应对本地文件的格式的多样性。对于跨节点、跨网络的数据文件的接入则需要更多网络环境的支持，同时选择合适的接入工具也很重要。
3. 接入传统数据库中的数据。对传统数据库种类的支持、提取速度等都是衡量传统数据库数据接入好坏的标准
4. 接入FTP服务器中的数据。
   * 1. 格式转换

按照统一的数据标准和既定的格式转换规则，对结构化数据进行如下格式转换处理。

代码转换：将不同来源数据中的用代码表示的数据，统一转换成平台中使用的代码格式。

数据截断：超出字段规定长度的数据，从后向前超出部分截断（需处理半字符情况），需截断字段的全部内容（并加字段描述）另存文件，同一条记录的多个需截断字段合并为一个文件，合并时需制定一个格式用以区分。数据截断时保留内容中要有“特殊字符”能够表明字段被截断过。

格式统一：用于比对的字段需要做格式统一，特定字段全角转半角。格式统一后的字段单独存储，原字段予以保留。

对不完整、无效数据予以丢弃并记录日志。

为规范数据元素查询、布控的统一格式，在保存原始数据的基础上，对数据产生过程是由网络用户人工输入的数据（如用户填写的网站注册资料）等情况，按照元数据规则进行转换处理。

* + 1. 数据清洗

基于预定策略，对结构化和非结构化进行去重，数据的类型检查及过滤。

* 垃圾过滤

对接入的各个数据源的各类垃圾数据进行垃圾过滤。

基于样本过滤。用户提供垃圾样本，系统根据样本对数据进行过滤。

基于账号列表过滤。设置经常发送的账号列表白名单，系统根据账号列表对数据进行过滤。

* 数据去重

数据重复主要有以下两种情况：结构化数据中数据重复，非结构化数据中的数据重复，如大量群发邮件中的相同的附件。

结构化数据和非结构化数据采用不同的去重方法。

* + 1. 数据关联

大数据平台数据类型多样，在进行数据交换时，可设置关联规则，将导入的数据与平台内存储的数据进行关联。例如：如针对上网数据，对ADSL账号关联和IP地址进行关联，对手机号码、IMSI号和硬件序列号进行关联等。

* + 1. 数据校验

在进行数据交换时，数据交换层会对计算发送数据的校验值，待接收方获取数据后，通过此校验值验证数据的正确性和完整性。

1. 大数据管理系统设计
   1. 系统功能

大数据平台中将运行着规模众多的硬件设备，包括服务器设备、网络设备、存储设备等，同时平台中将同时运行着多种大数据相关软件，包括hadoop各组件、新闻数据处理软件、用户行为分析软件等各种应用软件。大数据管理系统将提供高效、稳定、统一的管理功能，及时有效的对系统中的各种软硬件资源进行监控管理。针对大数据平台的特点，曙光大数据管理系统的设计将实现如下功能：

* + 1. 便捷的部署

对于大数据平台中的各种软件系统，在大规模硬件环境下，管理系统会提供便捷的部署安装功能，实现基于Web的一键安装功能，帮助管理员轻松完成软件的安装、升级、卸载、扩容等部署工作。部署的方式结合曙光多年来运维软件的经验，将提供向导式和一键式部署两种方式。一键式部署方式中管理员只需将相关配置通过xls上传就可一键完成系统部署工作；向导式部署方式将提供基于Web的步进式向导页面，引导管理员设置各软件的相关配置，进而完成一体化的部署工作。管理系统中将定义统一的标准部署接口，从而适配不同软件的部署开发工作。管理系统将统一处理异构软件、异构环境的异同，对管理员屏蔽差异，进而实现便捷的部署功能。

* + 1. 完善的管理

对于大数据平台中数量众多的软件，大数据管理系统将针对不同的软件或组件实现完善的管理功能，对所有功能提供基于Web的控制台操作页面，并支持良好的UI操作，使得管理员能够轻松驾驭大数据平台。对常用的管理功能，如软件的启停、软件中某类服务的启停、节点角色配置、高可用设置、负载均衡等，管理系统也将提供一键式的操作，降低管理员的工作复杂度，减少管理工作量。在基于Web的控制台中，管理系统将提供对分布式文件系统的管理功能，包括文件系统的启停、文件浏览、文件上传等功能；对于任务调度系统，如MapReduce，将提供作业的创建、提交、监控、结果查看等全生命周期的管理功能。对于应用软件，也将提供对应软件的各种管理功能。

* + 1. 全面的监控

面对纷繁复杂的大数据平台，管理系统将提供全面的监控功能，对系统中各个层次的软硬件状态、性能等进行全方位的监控，并以直观的方式加以展现。设备监控将提供各服务器节点、存储设备、网络设备的状态和运行性能监控，以及设备中各部件的基本信息监控，如CPU、网卡、硬盘等部件的信息；组件监控将提供大数据各组件HDFS、MapReduce、HBase、Hive等服务的状态、性能、运行情况的监控；同时管理系统将提供集群层次的监控，对集群的整体性能、状态进行聚合的监控展现。管理系统将利用数据可视化工具来展现监控的数据，辅助管理员确定是否有负载均衡、失控进程或硬件故障的问题，以及将会出现故障的趋势，及时发现和制止，对风险加以控制。

* + 1. 健全的告警

大数据管理系统在对平台中各软硬件资源全面监控的基础上，将提供健全的告警机制，在故障发生的第一时间将告警通知给管理员。管理系统提供多种方式的告警探测方式，可以对各种状态异常直接进行告警，也可以基于某项监控指标设定阈值进行告警，如磁盘空间利用率到90%需告警。在告警收集方式上，管理系统支持各种API的调用，SNMP接口的调用等多种手段，从而应对不同类型软硬件设备监控的需求。在告警通知上，管理系统提供Web页面、短信、邮件等多种方式结合的通知手段，并允许管理员对告警发送的规则进行设置，进而可以针对不同的告警发送给相应的管理人员。同时对于大规模环境下的告警，管理系统将实现完善的告警过滤、去重等功能，从而应对在大规模故障引起告警风暴时对系统内告警的及时响应。

* + 1. 多样的报表

管理系统将对监控的数据进行保存，并对历史数据提供各种维度的报表分析功能，从而辅助分析系统的运行趋势、故障风险等。报表模块通过时序图、柱状图、饼图、堆栈图、多轴对比图、对指标时序对比图等各种方式可以反应系统各层级设备、应用、软件的当前可用性、历史可用性、资源使用情况等，从而评测出系统整体的可用行；通过报表的单指标历史趋势和多指标融合对比，可以反应系统各个层面状态，预测系统可能的问题，从而辅助制定处理预案；通过报表的基于应用特性的分析，可用优化配置应用和资源，实现对历史监控数据的挖掘功能。多样的报表功能，将更完善的基于历史监控数据把握系统的整体运行态势，辅助分析系统的性能状态。

* + 1. 易用的调参

对于大数据平台，其运行着若干软件，管理系统支持对每种软件在运行期根据实际情况进行灵活的设置，对存在相互关联的参数会同时修改；对部分参数修改后会重启相应的软件使其生效。通过集中式的参数管理功能，管理员可以轻松的实现大数据平台内各种软件的参数修改，管理系统自动完成相关参数在各软件、各服务节点上的设置并让其生效。对于在调优阶段，管理员通过易用的调参功能，能够及时有效的调整系统参数，从而使得系统能够运行在最佳状态。

* + 1. 快速的诊断

对于大数据平台，其设备众多，运行软件众多，软件之间尤其还存在着相互依赖的关系。当系统出现故障时，管理系统将提供快速的诊断工具，帮助用户快速准确、及时有效的找到故障的根本原因。管理系统将提供全面的日志查询、搜索功能，能够将各软件的日志信息进行汇总分析，并直观的提供Web浏览界面，方便用户查询；管理系统借助监控、告警、报表等功能，全面的监控系统的状态，及时的将告警信息通知给用户；另外管理系统将深入到各软件内部，在软件内部进入内部探测，第一时间将各软件运行中的潜在风险、故障及时的呈现出来。通过快速诊断的机制，在系统出现问题时，管理员可以较快的找出症结所在，及时修正及时解决。

* + 1. 良好的扩展

管理系统将设计成插件方式的软件体系架构，借助ESB、SOA定义相关功能的标准规范和接口，从而成为一个一体化的软件集成平台。管理系统可以将符合规范标准的的开源的模块和厂商的模块无缝的集成到系统中，可以定制开发相应的插件，将异构的软件通过适配开发集成到系统中。管理系统采用基于微内核的软件体系结构，实现模块、插件的生命周期动态管理，可具有良好的可扩展、可维护性，支持模块级升级功能，支持动态在线升级功能。一体化的软件集成平台，将使得整个大数据平台可以长期发展，得到良好的扩展，以适应不断变化的需求。

* 1. 系统设计
     1. 架构设计

曙光大数据管理系统将融合曙光在大规模集群监控领域多年的技术积累，采用先进的模块化软件体系架构、层次化设计路线、插件式基础架构，实现一体化的软件集成管理平台。管理系统将以实现大数据管理的PAAS平台为基线，从运维、运行、运用三个层面全方位实现大数据管理的各种功能。系统整体架构如下图所示：



整个系统横向以PAAS平台的设计思想，将构建一体化集成平台、统一调度平台和应用开发平台。一体化集成平台将针对大数据平台中需要监控管理的软件，以插件式模块化的架构路线，构建一体化的集成平台，可以集成对各种软件的管理功能；统一调度平台将构建大数据平台的任务调用平台，实现各种类型作业的定义、执行、管理等功能，实现资源的统一调度；应用开发平台，将针对新华社的需求，融合各种应用软件，使得各种软件可以在大数据平台中很好的得以运行。

管理系统纵向将在各种硬件设备的基础上分为运维、运行、运用三个层次构建各种管理功能。运维层将实现网管类的功能需求，包括设备监控、服务监控、运行监控、自维护、软件部署、服务管理、节点管理、配置管理、调优诊断、告警管理、统计报表等功能，通过运维层的功能，完成大数据系统日常运维所需的管理功能；运行层将实现大数据系统中数据处理的相关功能，包括文件浏览、文件上传、文件下载、作业编写、作业提交、作业调度、作业提交、数据查询、数据导入、数据转换等ETL功能，通过运行层的功能，用户能够使用大规模数据处理的相关功能；运用层将对各种数据处理后的结果进行全面的分析展现，包括各种Web Portal、移动终端、数据分析、数据可视化工具，通过运用层的功能，能为用户提供最终的数据分析结果，帮助用户挖掘数据背后的价值。

1. 收视行为分析系统
   * 1. 常规收视指标分析
        1. 常规收视指标分析内容

常规收视指标主要针对频道（直播，点播），节目（直播，VOD），广告进行相关统计指标的计算，这其中主要涉及到的指标有以下：

1. **收视人数**

到达率是指在特定时间段内，收看过某频道（节目）的总人次数

1. **收视率**

收视率是指在某一特定时段收看某频道（节目）的平均人数或占总体电视推及人口的百分比。百分之一的收视率又被称为一个收视点。

1. **到达率**

到达率是指在特定时间段内，收看过某频道（节目）的不重复的人数，或其占总体电视推及人口的百分比。一般情况下，到达条件被设定为“至少收看某频道（节目）1分钟以上”，或者“至少收看某广告1次以上”。依据不同的需求，设置这一条件将会得到不同的到达率和接触度。

1. **收视时长**

收视时长通常指平均每天（日平均收视时长）或平均每周（周平均收视时长）实际收视观众的收视时长总和与总体电视推及人口的比值，计时单位一般是分钟或小时。

1. **收视份额**

收视份额是指特定时段内收看某频道（节目）的人数占同一时段所有收看电视人数的百分比。也即特定时段内某频道的收视率占所有频道总收视率的百分比，常被称为相对收视率，因为各时间段的电视总收视率是变化的。

1. **观众集中度**

观众集中度是将某频道（节目）的全部观众中的目标观众比例，与总体推及人口中同一目标人群的比例相比较所得的比值，如果这一比值大于100，说明目标观众的收视高于平均水平，反之，则低于平均水平。观众集中度表示的是目标观众相对于总体观众的收视集中程度，以此来反映目标观众对特定频道（节目）的收视倾向。

1. **观众忠实度**

观众忠实度是指同一频道（节目）的收视率与到达率的比值，经常用于考察观众在该频道（节目）上平均停留的时间长短。该指标值的变化幅度在0-100％之间，数值越大，表明观众对该频道（节目）的忠实程度越高

1. **开机率**

开机率是指一段时间内有过开机行为的用户的比率，开机率旨在表示所有用户整体的活跃度

* + - 1. 常规收视指标配置参数

**a．趋势报表**

以上各指标均支持选择以下条件，用以生成对应的报表

1. 指标（包括指标选项）
2. 查询类型
   1. 直播节目（可根据名称检索定位到直播节目，可复选多个）
   2. 直播频道（可根据名称检索定位到直播频道，可复选多个）
   3. 点播节目（可根据名称检索定位到点播节目，点播节目包，

支持节目包按树状层次结构展开，可复选多个）

* 1. 广告（可根据广告位，广告名称检索，广告位支持按树状层次

结构展开，可复选多个）

1. 时间范围
   1. 可以按照日期控件预制模板选择
   2. 支持自定义日期选择
2. 时间粒度
   1. 可以按照分钟/小时/天/周/月/年进行统计
   2. 可以自定义时间段模板，比如按十五分钟时间段进行统计

**b．TOP报表**

以上指标均支持展示TOP报表，支持选择以下条件：

1. 指标（包括以上指标选项）
2. 查询类型
3. 直播节目
4. 直播频道
5. VOD节目
6. VOD节目包
7. 广告名称
8. 广告位

3. 时间范围

a) 可选择日期范围和时间范围，粒度可以到分钟级别

4. Top榜数量

* + 1. 基于用户的收视指标分析
       1. 基于用户的收视指标分析内容

基于用户的收视指标分析，有助于针对每个数字电视用户进行更为精确的行为判断，并可以此为基础，实现对用户的收视行为标签化，有助于更好的开展针对用户使用习惯的个性化营销服务。

针对用户的收视指标，主要从直播、点播、回看，广告等用户行为进行相关分析：

1. 直播
   1. 用户开机后，观看的直播频道名称、节目、类型、时长
   2. 通过对比用户开机和观看第一个板块的时间差，计算用户对直播板块的关注时长
   3. 当付费节目预览系统开通时，记录用户使用预览系统的时长、次数、频道名称、观看节目内容
2. 点播（含点播，热播，专题）
   1. 了解用户使用点播版块下的类目明细，观看时间、观看时长、观看节目名称、持续时长
   2. 将点播节目与不同供应商匹配起来，了解不同供应商提供的节目受欢迎程度，包括供应商提供的节目类型、数量、语言、用户观看的数量、停留时间、付费的比例
   3. 将点播节目与BOSS数据匹配起来，包括片区、楼盘、入网方式、套餐、入网时间，了解不同因素下的点播倾向
3. 回看
   1. 了解用户使用回看的频道类型、名称、观看节目名称、观看时间、观看时长、持续时长
   2. 将回看节目与不同的频道、节目类型匹配起来，了解不同类型的节目对用户使用回看的吸引力
   3. 将回看节目与BOSS数据匹配起来，包括片区、楼盘、入网方式、套餐、入网时间，了解不同因素下的倾向
      * 1. 基于用户的收视指标配置参数

以上各指标均支持选择以下条件，用以生成对应的报表

1. 指标类别（包括以上指标选项）
2. 查询类型
   1. 直播
   2. 点播
   3. 回看
   4. 广告
3. 时间范围
   1. 可以按照日期控件预制模板选择
   2. 支持自定义日期选择
   3. 收视行为分析系统技术架构

收视行为分析系统分为前端报表展示界面和后端服务两部分

* + 1. 前端界面技术架构

前端界面基于可扩展的前端框架，支持报表控件的自由选择，模块化拖曳，

Dashboard定制等功能，支持曲线趋势图，折线图，饼图，柱状图，云图，雷达图等不同视图展示方式。

前段系统将采用动态模块化插件架构：利用插件机制，实现对不同服务的扩展，可以为每一种服务、设备写对应的插件，新增服务时，只需为其量身定制插件即可实现对其兼容管理；采用OSGi模块化技术，基于构件和服务的思想，实现软件模块的“热插拔”，从而实现系统的在线扩展、升级机制，构建坚实的软件平台。

1. **插件架构体系**

插件架构体系的核心理念是基于松散的模块积累方式，通过新增插件以及扩展原有插件的方法来完成系统的实现。一个插件模块的更新不需要对整个系统进行重新编译，不会影响其它插件模块。

插件架构体系的优点非常明显，象硬件一样的即插即用。在开发的时候只需划分好模块，各个插件开发人员只需遵循接口协议，就能开发出互不影响的插件模块，非常方便开发和调试；由于其灵活性，可以实现系统的灵活定制，当新增功能或者修改功能时，只需对相应插件进行修改即可实现，为软件的发展带来了极大的灵活度；并且无论对于公司还是业界而言都是技术良好的积累方式，为公司或业界提供统一而规范的开发方式以及稳定的内核架构等等，这些优点无论对于公司还是业界来说都是非常重要的。

OSGI

发布服务

监听器

注册插件

插件调度器

插件管理器

DB

服务扩展

插件A

图 插件架构体系示意图

系统将结合服务器监控管理的具体需求，实现灵活的插件架构体系；构建统一的插件管理器、插件调度器，开发人员只需按照相应的框架要求，编写单个插件，即可将插件集成运行与系统中。本着关注点分离的理念，插件框架实现者聚焦与插件架构机制的实现，而插件编写者则关注与具体服务器设备监控管理方式的实现，分工协作，将极大提高系统的效率。

* + 1. 后台服务技术架构

后端服务的逻辑架构如下图所示



图 收视行为分析后台服务逻辑架构图

其中，几个重要的组件介绍如下

* **ElasticSearch**：

ElasticSearch是一个基于Lucene的搜索服务器。它提供了一个分布式多用户能力的全文搜索引擎，基于RESTful web接口。Elasticsearch是用Java开发的，并作为Apache许可条款下的开放源码发布，是第二流行的企业搜索引擎。设计用云计算中，能够达到实时搜索，稳定，可靠，快速，安装使用方便。

我们建立一个网站或应用程序，并要添加搜索功能，令我们受打击的是：搜索工作是很难的。我们希望我们的搜索解决方案要快，我们希望有一个零配置和一个完全免费的搜索模式，我们希望能够简单地使用JSON通过HTTP的索引数据，我们希望我们的搜索服务器始终可用，我们希望能够一台开始并扩展到数百，我们要实时搜索，我们要简单的多租户，我们希望建立一个云的解决方案。Elasticsearch旨在解决所有这些问题和更多的问题。其中有以下重要组件：

**cluster**

代表一个[集群](http://baike.baidu.com/view/302477.htm" \t "_blank)，集群中有多个节点，其中有一个为主节点，这个主节点是可以通过选举产生的，主从节点是对于集群内部来说的。es的一个概念就是去中心化，字面上理解就是无中心节点，这是对于集群外部来说的，因为从外部来看es集群，在逻辑上是个整体，你与任何一个节点的通信和与整个es集群通信是等价的。

**shards**

代表索引分片，es可以把一个完整的索引分成多个分片，这样的好处是可以把一个大的索引拆分成多个，分布到不同的节点上。构成分布式搜索。分片的数量只能在索引创建前指定，并且索引创建后不能更改。

**replicas**

代表索引副本，es可以设置多个索引的副本，副本的作用一是提高系统的[容错性](http://baike.baidu.com/view/2700299.htm" \t "_blank)，当某个节点某个分片损坏或丢失时可以从副本中恢复。二是提高es的查询效率，es会自动对搜索请求进行负载均衡。

**recovery**

代表数据恢复或叫数据重新分布，es在有节点加入或退出时会根据机器的负载对索引分片进行重新分配，挂掉的节点重新启动时也会进行数据恢复。

**river**

代表es的一个数据源，也是其它存储方式（如：数据库）同步数据到es的一个方法。它是以插件方式存在的一个es服务，通过读取river中的数据并把它索引到es中，官方的river有couchDB的，RabbitMQ的，Twitter的，Wikipedia的。

**gateway**

代表es索引快照的存储方式，es默认是先把索引存放到内存中，当内存满了时再持久化到本地硬盘。gateway对索引快照进行存储，当这个es集群关闭再重新启动时就会从gateway中读取索引备份数据。es支持多种类型的gateway，有本地文件系统（默认），[分布式文件系统](http://baike.baidu.com/view/771589.htm" \t "_blank)，Hadoop的HDFS和amazon的s3[云存储](http://baike.baidu.com/view/2044736.htm)服务。

**discovery.zen**

代表es的自动发现节点机制，es是一个基于p2p的系统，它先通过广播寻找存在的节点，再通过[多播](http://baike.baidu.com/view/378050.htm" \t "_blank)协议来进行节点之间的通信，同时也支持[点对点](http://baike.baidu.com/view/1145124.htm)的交互。

**Transport**

代表es内部节点或集群与客户端的交互方式，默认内部是使用tcp协议进行交互，同时它支持http协议（json格式）、[thrift](http://baike.baidu.com/view/1698865.htm" \t "_blank)、servlet、memcached、zeroMQ等的[传输协议](http://baike.baidu.com/view/441895.htm" \t "_blank)（通过[插件](http://baike.baidu.com/view/18979.htm" \t "_blank)方式集成）。

* **HDFS：**

Hadoop分布式文件系统(HDFS)被设计成适合运行在通用硬件(commodity hardware)上的分布式文件系统。它和现有的分布式文件系统有很多共同点。但同时，它和其他的分布式文件系统的区别也是很明显的。HDFS是一个高度容错性的系统，适合部署在廉价的机器上。HDFS能提供高吞吐量的数据访问，非常适合大规模数据集上的应用。HDFS放宽了一部分POSIX约束，来实现流式读取文件系统数据的目的。HDFS在最开始是作为Apache Nutch搜索引擎项目的基础架构而开发的。HDFS是Apache Hadoop Core项目的一部分。

有如下设计特点：

**硬件故障**

硬件故障是常态，而不是异常。整个HDFS系统将由数百或数千个存储着文件数据片断的服务器组成。实际上它里面有非常巨大的组成部分，每一个组成部分都很可能出现故障，这就意味着HDFS里的总是有一些部件是失效的，因此，故障的检测和自动快速恢复是HDFS一个很核心的设计目标。

**数据访问**

运行在HDFS之上的应用程序必须流式地访问它们的数据集，它不是运行在普通文件系统之上的普通程序。HDFS被设计成适合批量处理的，而不是用户交互式的。重点是在数据吞吐量，而不是数据访问的反应时间，POSIX的很多硬性需求对于HDFS应用都是非必须的，去掉POSIX一小部分关键语义可以获得更好的[数据吞吐率](http://baike.baidu.com/view/4854370.htm" \t "_blank)。

**大数据集**

运行在HDFS之上的程序有很大量的数据集。典型的HDFS文件大小是GB到TB的级别。所以，HDFS被调整成支持大文件。它应该提供很高的聚合数据带宽，一个[集群](http://baike.baidu.com/view/302477.htm" \t "_blank)中支持数百个[节点](http://baike.baidu.com/view/47398.htm)，一个集群中还应该支持千万级别的文件。

**简单一致性模型**

大部分的HDFS程序对文件操作需要的是一次写多次读取的操作模式。一个文件一旦创建、写入、关闭之后就不需要修改了。这个假定简单化了数据一致的问题和并使高吞吐量的数据访问变得可能。一个Map-Reduce程序或者[网络爬虫程序](http://baike.baidu.com/view/1431312.htm" \t "_blank)都可以完美地适合这个模型。

* **HBASE：**

HBase是一个分布式的、面向列的开源数据库，该技术来源于 Fay Chang 所撰写的Google论文“Bigtable：一个结构化数据的[分布式存储系统](http://baike.baidu.com/view/1911305.htm" \t "_blank)”。就像Bigtable利用了Google文件系统（File System）所提供的分布式数据存储一样，HBase在Hadoop之上提供了类似于Bigtable的能力。HBase是Apache的Hadoop项目的子项目。HBase不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。另一个不同的是HBase基于列的而不是基于行的模式。

* **HIVE：**

Hive是建立在 Hadoop 上的数据仓库基础构架。它提供了一系列的工具，可以用来进行数据提取转化加载（ETL），这是一种可以存储、查询和分析存储在 Hadoop 中的大规模数据的机制。Hive 定义了简单的类 SQL 查询语言，称为 HQL，它允许熟悉 SQL 的用户查询数据。同时，这个语言也允许熟悉 MapReduce 开发者的开发自定义的 mapper 和 reducer 来处理内建的 mapper 和 reducer 无法完成的复杂的分析工作。

Hive 没有专门的数据格式。 Hive 可以很好的工作在 Thrift 之上，控制分隔符，也允许用户指定数据格式。

* **SPARK：**

Spark是UC Berkeley AMP lab所开源的类Hadoop MapReduce的通用并行框架，Spark，拥有Hadoop MapReduce所具有的优点；但不同于MapReduce的是Job中间输出结果可以保存在内存中，从而不再需要读写HDFS，因此Spark能更好地适用于数据挖掘与机器学习等需要迭代的MapReduce的算法。

Spark 是一种与 [Hadoop](http://baike.baidu.com/view/908354.htm) 相似的开源集群计算环境，但是两者之间还存在一些不同之处，这些有用的不同之处使 Spark 在某些工作负载方面表现得更加优越，换句话说，Spark 启用了内存分布数据集，除了能够提供交互式查询外，它还可以优化迭代工作负载。

Spark 是在 [Scala](http://baike.baidu.com/view/1588150.htm) 语言中实现的，它将 Scala 用作其应用程序框架。与 Hadoop 不同，Spark 和 Scala 能够紧密集成，其中的 Scala 可以像操作本地集合对象一样轻松地操作分布式数据集。

尽管创建 Spark 是为了支持分布式数据集上的迭代作业，但是实际上它是对 Hadoop 的补充，可以在 Hadoop 文件系统中并行运行。通过名为 Mesos 的第三方集群框架可以支持此行为。Spark 由加州大学伯克利分校 AMP 实验室 (Algorithms, Machines, and People Lab) 开发，可用来构建大型的、低延迟的数据分析应用程序。

* **执行调度器：**

执行调度器，用来接收从web服务的请求，根据请求的具体的类型，对请求进行解析，然后再调用相应的服务中，目前有三种类型的请求，分别对应三种类型的服务。

1. 常规收视指标分析请求，对应ElasticSearch 服务
2. 基于用户的指标分析请求，对应与HBase服务
3. 数据加工作业，对应于MR/Hive/Spark作业请求

执行调度器得到服务请求的最终结果，返回到Web服务层

* **Web服务接口：**

Web服务层用来提供与Web端请求的交互逻辑的服务接口，处理提交的分析作业，返回最后的分析结果用于前端界面呈现。

Web服务层还支持以RESTful API的形式进行数据的交互，提供除Web界面之外另一种提交分析请求的途径。

* 1. 收视行为分析系统处理流程

整个收视行为分析系统的处理流程如下图所示：

图 收视行为分析整体流程图

其中

1. 日志等原始数据存储在HDFS上，以分割好的文本文件方式存放，方便Hive/MR/Spark进行处理。
2. 加工数据存放在HBase中，以用户ID为rowkey组织，根据某用户进行查询性能响应在毫秒级
3. 收视行为以倒排索引的方式存储在本地文件系统中，由ElasticSearch管理，具有自动分片，线性扩展的特性，基于索引的过滤，聚集，分组的统计响应时间均在秒级，可以达到对入库数据实时生成指标报表的性能。
4. 精准营销
   1. 精准营销系统功能
      1. 用户画像

基于用户属性信息（年龄、性别等）、终端信息（类型、状态）、位置信息（小区、生活圈等）、收视行为、上网行为轨迹、工单信息（安装、报修、投诉等）、消费信息等丰富的数据，为每个用户打上人口统计学特征、收视行为、上网行为和兴趣爱好标签

具体包含以下信息

1. 基本资料：年龄，地址，楼盘，联系方式
2. 业务属性：数字电视，模拟电视，宽带，互动电视
3. 收视行为特征：开机次数，平均开机时长，直播收看习惯，广告收看情况，点播收看习惯，喜爱的节目类型
4. 套餐：套餐订购，套餐使用
5. 消费：消费水平，消费等级，支付习惯
6. 工单情况：故障报修和工单投诉情况
7. 网络使用：网络使用带宽，网络延迟
8. 设备情况：设备的软硬件版本
   * 1. 用户群体划分

用户群体细分主要依据用户特征对用户进行群体聚类，将具有相似特性的用户划分为一个特征群体。

用户群体细分的依据为用户画像各维度特征，通过构成特征向量，根据余弦相似度对所有用户进行距离上的迭代计算得到最后细分的用户群体。

通过用户群体细分，数量巨大的用户被划分到若干个群体中，以群体来代替个体，在能有效保证推荐效果的同时，大大降低推荐中矩阵空间的维度，降低整个计算的复杂度，能带来明显效率上的提升。

* + 1. 个性化推荐

在用户画像和用户群体划分的基础上对用户特征进行深入理解，建立用户与品牌/产品/套餐、终端类型的精准匹配，并在推送渠道、推送时机、推送方式上满足用户的需求，实现精准营销。

1. 支持基于节目和基于用户的协同过滤推荐；
2. 支持基于节目内容的个性化推荐，比如通过分析客户消费信息、收视行为和品牌/产品订阅，结合当前节目的内容属性，在合适的时机（开机、音量条、频道切换等）推送合适的节目；
3. 提供RESTful API方式对外数据查询、展示接口，接口协议采用HTTP协议实现。
   1. 精准营销系统流程
      1. 用户画像

用户特征提取的主要流程如下图所示。



图 用户画像流程图

其中：

1. 用户消费工单行为数据以定期文件交换的方式，增量导入到系统HDFS中。
2. 分析过程采用MapReduce实现
3. 最终的结果存放在HBase用户画像特征库中。
4. 智能运维
   1. 智能运维系统功能
      1. 终端管理

主要是用户终端使用行为分析、终端在线监测与智能判断。

其中，终端使用行为分析根据终端开关机、时长、时段、日志信息等使用信息，研究客户的终端使用习惯，为精准营销打下基础；

终端在线监测与智能判断通过终端信息采集系统、营销系统的数据共享与交互，利用大数据分析技术进行终端故障诊断/预警、设备整体工况分析，实现终端异常分析、故障分析、各类事件分析等功能

* 1. 智能运维系统架构

终端管理的数据主要来源于终端设备产生的日志信息，对于这些日志信息，可以通过分析日志内容中的ERROR信息，已经日志记录的统计信息，与正常设备的统计值进行对比，得到相应的结果，对于出现严重问题的设备，需要告警提示。

业务运维的数据主要来源于用户的工单数据，由于工单数据既包含结构化字段，又包含纯文本的非结构化字段，需要对其进行文本提取，根据其语义信息，只能判断工单类型，并自动做出对应的操作。

对于，用户多次在工单中提到的关键词信息，需要记录下来，当某关键词累计到一个阈值时，发出告警提示，辅助线下人员有针对性的开展定点服务。

其总体的架构如下图所示



图 智能运维系统架构

* 1. 智能运维系统流程
     1. 终端管理

终端管理整体流程如下图所示：



图设备系统流程

其中：

1. 终端行为日志存储在HDFS文件系统上
2. 统计分析采用MR/Spark作业并行化，加速处理性能。