

浅谈通信运营信息的大数据采集及存储系统和方法

张 维

中国电信股份有限公司东海分公司 江苏 连云港 222300

【摘要】在新时代发展背景下,科技与经济快速发展,逐渐出现了较多新技术与新产品。在信息时代下,必须深入研究和应用大数据技术,以此促进各行业领域的发展。此次研究主要是探讨分析通信运营信息的大数据采集及存储系统和方法,希望能够对相关人士起到参考性价值。

【关键词】通信运营信息;大数据采集;存储系统;方法

在大数据发展过程中,人们的生活发生了翻天覆地的变化,例如购物方式的变化和支付方式的变化等。上述改变内容都为社会生产生活提供便利性。然而需要注意的是,在大数据时代下,也为社会发展带来了一定的负面影响。所以必须深入探究大数据采集与存储,这样能够更好地应用通信运营信息大数据。

1 通信运营下的大数据平台

在通信运营信息下,采集与存储数据有助于掌握时代的发展方向,制定出更好地发展规划。大数据技术在收集信息方面,主要包括网络信息、用户信息和管理系统及用户关系信息等。

对于数据源来说,我国现有通信运营商都具备相应的数据量,且数据宽度和广度都非常大。国内三大运营商公布的2019年8月份运营数据显示,中国移动的移动用户总数达到9.41279亿户、4G用户总数达到7.45869亿户、宽带用户总数达到1.80815亿户,中国电信的移动用户总数达到3.2806亿户、4G用户总数达到2.7210亿户、宽带用户总数达到1.5174亿户,中国联通的移动用户总数达到3.24646亿户、4G用户总数达到2.47794亿户、宽带用户总数达到0.83975亿户。移动、电信、联通三大运营商移动用户总数达到15.93985亿户,远远超出中国大陆总人口数,相当于每个中国大陆公民都参与在运营商大数据中,涵盖了各行各业的从业人员,更有企业、组织所需要的潜在客户资源。如此庞大的用户群体,为通信运营大数据网络带来了海量数据。

通信运营大数据平台多由数据采集层、源层、存储层、分析层以及平台管理系统等构成。对于数据采集层来说,其主要应用业务系统数据和第三方数据,联合业务数据库终端接口、文件传输协议以及分辨率信息等。通过处理数据和提取转换等方式,能够将其传输到数据平台层。数据收集多应用不同格式数据源,对结构化信息与非结构化信息进行覆盖。通过此种方式可以实现数据采集的周期性,还能够确保数据具备稳定来源。对于数据存储层来说,数据存储多应用分布式存储方式,将数据存储到服务平台上,数据存储能够科学分析数据,之后实现数据的合理分配与存储。对于数据分析处理层来说,在处理大数据时,为了促进信息处理效率的提升,需要采用离线方式和分布式处理方式,实时计算和处理信息流。大数据编程模型由于具备较高的灵活性,因此被应用到离线处理工作中。实时大数据处理技术主要是应用分布

式处理技术实现,对于数据处理分析的要求比较高。因此在分析和处理数据时,必须全面提升数据的有效利用率。对于数据汇聚层来说,其主要是利用分布式开源数据库,存储和查询数据,以此提升数据查询检索的便利性,还能够简化数据查询的繁琐性。对于数据服务层来说,其能够为第三方提供数据查询服务、分析服务以及可视化服务,确保第三方能够从数据信息服务平台内获取相关信息,这样能够促进第三方服务的发展。对于用户应用层来说,能够为用户提供短信推送服务、项目退浆服务以及日志推送服务等,确保用户能够及时了解和掌握产品信息。对于平台管理系统来说,其能够直接管理通信运营大数据,包含数据采集系统、维护系统和安全系统。

2 通信运营大数据收集与存储方法

2.1 大数据信息收集方法

在通信运营大数据平台中,主要是通过网络数据采集、系统日志采集以及其他数据采集方法,最常应用的日志数据采集方法为Hadhoop、Scribe等方式。不同通信运营商都具备各自的系统日志数据采集方法。网络数据采集方法主要是利用网络方式完成数据采集,需要将信息采集网站数据纳入到数据采集中,这样能够收集数据并传输到数据库中。对于其他数据采集方法来说,主要是与相关研究机构进行合作,应用特殊数据采集方式采集数据。

第一,采集文件统一命名与格式:应用大数据所涉及的网元与专业比较多,且持续时间长,为了防止重复性数据收集与处理问题,需要对文件名称与格式进行规定。

第二,统一命名采集文件。对于数据文件来说,包含被采集系统、区域、数据类型、专业以及采集网元设备类型等。数据上报打包的文件名称为tar.gz。数据文件名不同字段的定义。

第三,采集文件统一格式。打包文件格式多应用tar.gz。文件格式,tar.gz代表采用tar命令打包,通过gzip命令进行压缩,将数据文件打包为一个整体。数据文件格式所应用的常规格式为CSV,文件数据中应用“|”或者“,”隔开,针对明确要求的厂家私有文件格式数据,可以不需要应用CSV文件格式。

2.2 大数据信息存储方法

在存储通信运营大数据时,主要是应用网格存储技术、输入输出技术和虚拟存储技术,以上三组技术能够有效存储大

数据, 以此维护数据存储的安全性, 还能够便于搜索和查找数据。虚拟存储技术多是由于虚拟磁盘技术, 以此确保数据应用的合理性, 还能够明显降低数据存储成本, 对存储方式进行优化。网格式存储方式能够高效存储数据, 能够确保数据存储的条理性。

2.3 监测采集状态

通过监测数据采集状态, 能够有效维护数据及时性和完整性, 可以通过采集平台对不同采集任务的数据质量进行监控。

第一, 针对数据完整性监控来说, 对于配置数据和性能数据, 可以判断采集适配器采集数据的完整性、历史数据的完整性, 还能够监控文件大小和数目。对于告警数据来说, 可以按照解析完整性, 对数据解析成功率进行判断。

第二, 针对数据及时性监控来说, 针对性能与告警等实时数据, 可以按照数据产生时间、接收时间, 对数据整个环节的及时性进行判断, 且了解数据不同环节延迟是否满足及时性要求。此外, 通过设定及时性阈值, 能够对数据采集与呈现过程的及时性进行判断, 并且做好监控措施。

第三, 监测数据采集接口状态。对制定时间段的接口调用成功率以及应用频次进行统计。针对非实时数据的上传来说, 可以显示出采集文件的名称、数据类型以及文件大小, 还能够明确数据采集状态、接收时间以及接口连接状态等。接

口检查点也包含在其中, 但是并非局限于上述数据。

第四, 监测采集设备状态。对数据采集机的运行状态进行监控, 还能够监控采集机的资源使用情况。采集器适配器能够通过心跳通信保持连接。当适配器发生中断时, 可以利用心跳检测进行自动化连接。当多次连接仍未成功时, 则通过告警方式转移到呈现界面中, 处理相关故障。此外, 还能够对数据采集机的中央处理器及磁盘使用情况进行实时监控, 明确采集机的负荷, 防止由于负荷过大所致采集数据质量低下问题。当数据大于设定阈值时, 会自动化产生告警, 并且转移到集中显示界面中, 处理相关情况。

3 结束语

综上所述, 在大数据时代发展下, 极大影响了社会生产与生活, 所以必须注重大数据收集与存储技术, 以此确保数据存储的可靠性和安全性, 从根本上提升数据有效利用率。

参考文献:

- [1]张牧,黄伟,邹茜.无线网络应用于大数据采集与监控的可靠性探讨[J].信息化建设. 2015(12).
- [2]王鹏,傅子明,洪训山,韦景康.基于安卓终端的多模信号频域分析及装置研发[J].机电信息.2018(06).
- [3]管宇旻,汤亚杰,徐剑桥.计量检测数据采集与智能化管理[J].上海计量测试. 2017(04).