



基于大数据的电信运营商业务精确运营平台的构建

刘 洁¹, 王 哲²

(1. 中国电信股份有限公司广东研究院 广州 510630; 2. 中国电信股份有限公司广东分公司 广州 510081)

摘 要: 大数据作为云计算的一种实际应用, 在电信运营商的运营支撑中有着举足轻重的作用。基于移动互联网的发展趋势, 首先分析电信运营商面临的挑战和建设难题, 基于大数据分析的精确运营平台的重要性和必要性, 提出引入大数据精确运营平台的业务网络架构和构建大数据精确运营平台的思路和方法。

关键词: 大数据; 用户行为分析; 云数据库; 分布式计算; 并行计算; 云计算; 业务网络

doi: 10.3969/j.issn.1000-0801.2013.03.005

Construction of Carrier Business Precise Operating Platform Based on Big Data

Liu Jie¹, Wang Zhe²

(1. Guangdong Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China;

2. Guangdong Branch of China Telecom Co., Ltd., Guangzhou 510081, China)

Abstract: As a practical application of cloud computing, big data plays a more and more important role in the carrier's operations. With the high-speed development of mobile internet, the challenges faced by the carriers were analyzed, and then the importance and necessity of constructing a business precise operating platform based on big data was promoted. Also, the new business network architecture with the introduction of such a big data platform was suggested. Some ideas and methods of how to build such a precise operating platform based on big data were suggested as well.

Key words: big data, user behavior analysis, cloud database, distributed computing, parallel computing, cloud computing, business network

1 概述

1.1 移动互联网大环境

移动互联网的发展呈现明显的去电信化趋势, 如OTT、电子商务、基于LBS的商业服务和社交应用。越来越多的用户群体倾向用微信、QQ等社交应用手段满足沟通和信息传送需求, 电信运营商的传统业务越来越受到威胁。在此大环境中, 电信运营商只有围绕自身的资源优势, 构建用户/业务可识别、管理和控制的综合平台, 实现通信管道的智能化, 才能提升核心资源和优势资源的价值, 更

好地适应移动互联网的发展趋势。

对比电信网络和互联网络的特点, 不难看出, 电信运营商的优势在于网络资源、用户信息资源和管理支撑体系。电信运营商掌握更为详细的用户信息, 把控更为深入的用户使用基础电信业务和通信管道的行为特征。在“去电信化”的需求下, 电信运营商需要结合这些信息和第三方应用使用行为, 主动挖掘用户需求并推荐切合用户实际的应用。有更好的资源支撑, 是运营商相对互联网服务商的最大优势, 也是运营商构建前后向运营环境的最好切入点。要做到这一点, 一个基本的要求就是要有一个支撑精

确运营的数据关联分析的大数据平台。

1.2 大数据定义

什么是大数据? 维基百科中将大数据定义为“无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合”。业界用4个V概括大数据的基本特征,即Volume(量大)、Variety(多样化)、Value(价值密度低)和Velocity(处理要求快)。量大是指数据规模至少在TB级;多样化,一方面指数据有不同的来源,另一方面指数据类型有结构化、半结构化和非结构化等多种形式;价值密度低指这些数据的单一来源或单一维度并不能带来有效的价值,只有形成一定规模并关联起来,才能体现大的价值;处理要求快主要指数据分析的实时性要求很高。

1.3 电信应用需求

运营商建设的业务数据关联分析平台是一个大数据平台,这点不难理解,因为它不仅需要对网络信令、流量日志、客户数据、用户行为数据、用户地理数据和应用喜好等数据进行分析,更需要对这些动静态数据进行关联分析和数据挖掘,数据量是海量的;但只要及时将数据之间的内在关联挖掘出来,数据的价值将大大提升,带来很大的商机。海量数据的分析将生成动态控制策略,一方面为智能管道提供网络控制的数据支撑,另一方面为应用的智能推送和用户主动服务提供支撑。

2 业务网络新架构

2.1 现有业务网络架构

电信运营商经过多年的网络改造和建设,基本上改变了传统的纵向业务网络,建成了管理统一、能力明晰、开放的业务网络架构。业务网络主要包括业务及能力管理域、能力接入和开放域、基础能力域以及应用平台等几部分。

业务及能力管理域集合各类增值业务平台的公共管理功能,包括认证、鉴权、话单生成和用户、CP/SP、业务的管理以及业务订购等功能;能力接入和开放域提供业务的统一接入功能,开放的业务生成、测试、执行环境,实现能力共享,降低业务生成门槛;基础能力域提供各种基本业务能力和适配功能,包括语音、短信、彩信、WAP、下载、定位、彩铃、流媒体、支付等;应用平台实现各种具体应用、业务逻辑的执行环境和个性化的管理功能。可见,现有的业务网络架构已经从根本上改变了以往单一业务独立管理的纵向架构,为业务融合和业务合作提供了良好环境,在一定程度上适应了互联网时代的通信服务要求;而管理域

作为核心的运营支撑部分,实现了业务的统一管理,包括集中的业务订购、统一的业务认证/鉴权、灵活统一的计费结算和统一的产品和用户管理等。

在目前的横向架构中,虽然实现了统一管理,但这种管理是相对静态的管理,缺乏对业务之间关联性和用户使用业务的行为特征的分析,是一种被动服务的方式,不能很好地适应移动互联网大环境中去电信化和有效开放的运营要求。

2.2 引入大数据平台的业务网络架构

基于上文的分析,提出在业务网络架构中引入基于大数据的业务精确运营平台(以下简称大数据平台)。引入大数据平台的业务网络架构如图1所示(灰色阴影部分表示新增内容)。

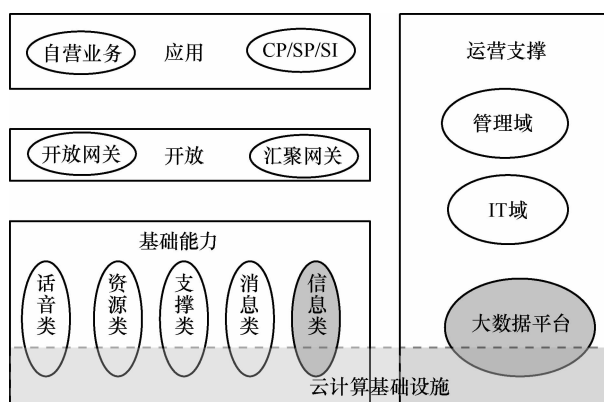


图1 业务网络新架构

新架构中的大数据平台定位于刻画用户的全息视图,对用户属性、承载网络、用户行为特征、用户业务喜好和业务订购特点及用户、业务等进行统计分析和关联分析,同时结合用户终端和接入业务的特点以及智能管道的承载信息,为智能管道的动态联动、电信业务和第三方应用的智能推介、用户接入业务的智能升级推荐等提供有力的数据支撑,从而实现前后向收费的移动互联网新模式。经过处理的数据加上包装,可以作为一种能力,提供调用和对第三方开放的功能。

组网方面,大数据平台与信令网元、终端自注册平台、支撑系统、业务管理平台、承载层信息感知网元和控制网元等均有连接,如与CRM(客户关系管理)、ISMP(统一业务管理)、AAA(认证授权计费)、CCG/DPI(内容计费网关/深度分组检测)等网元有连接关系,并从这些平台中收集静态和动态信息。特别要指出的是,这些信息包括大量的动态信息,如漫游信息、开关机信息、位置信息等,这些动



态信息提供了用户的行为日志,能够从中挖掘出更深层次的信息:如感知用户使用业务的喜好和惯性、用户使用业务的流量曲线等,从而提高业务精确运营的可能性和业务之间动态关联的可行性。另外,大数据平台与推送网关(WAP/push 网关、短信中心等)、OCS(在线计费系统)、PCG(策略控制网关)等网元也有连接,以实现业务的智能推送、根据动态业务情况实现实时计费或者分发网络控制指令等。

运营商大数据平台的主要应用场景如下。

- 基于用户上网日志,把握热点内容和应用,制定流量策略,支撑差异化服务和带宽控制,以客户体验为核心发展流量经营。
- 分配客户标签,刻画用户全息视图,实现根据用户喜好、位置、行为等的业务交叉推荐,支撑精确营销。
- 通过对 OTT 替代性业务的分析,掌握业务发展趋势,为制定电信业务的发展策略提供数据支撑和依据。
- 通过对前、后向流量的关联分析,实现与合作伙伴的合作定向营销。
- 根据业务和网络流量对数据进行分析,为 IT 系统的优化提供依据,使用户的优惠套餐定制化,改善用户体验,减少用户投诉。

可见,引入大数据平台的业务网络架构突破了传统业务网络专注于单一业务逻辑的实现和静态业务提供的局限,从运营的角度出发,重新定位了业务网络并提高了对业务网络的整体功能要求。不难理解,随着更多不同来源数据的引入(如第三方应用平台的用户个性化数据),业务网络可以利用大数据平台提供更多切合用户需求的个性化服务。

3 大数据平台的构建

3.1 大数据平台网络架构

大数据平台的组网架构如图 2 所示。

从图 2 可以看出,大数据平台作为业务网络的一个网元,与周边系统的众多网元有连接关系,其组网架构已经突破了业务域的范围,并从运营的角度切入,对业务域、网络域和 IT 域实现了有机的关联。

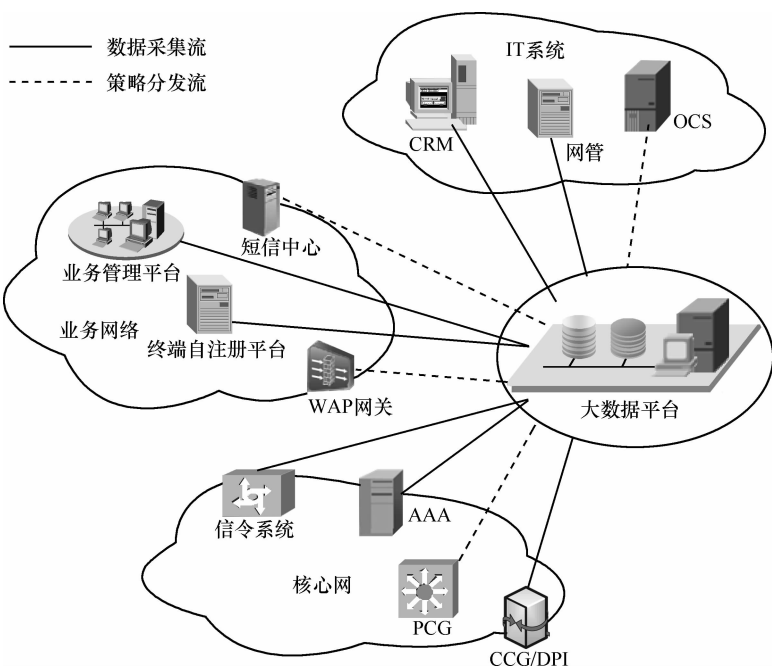


图 2 大数据平台的组网架构

根据上面的需求分析,大数据平台的功能架构如图 3 所示。

功能架构包括数据采集层、数据存储层、数据处理层和数据运营服务层。其中,数据采集层提供到各个原始信息网元的接口,并与各网元实时同步,获取各类信息。数据存储层将根据这些海量的基础信息进行数据识别、分类处理,如对用户的分类(如企业客户、家庭客户、个人客户等)、对业务的分类(如视频业务、话音业务、社交类业务、支付类业务等)、对接手段的分类(如 ADSL 接入、光纤接入、3G 接入、WLAN 接入等),并采用云数据库的方式进行有效存放。数据处理层对静态数据和动态信息进行分析

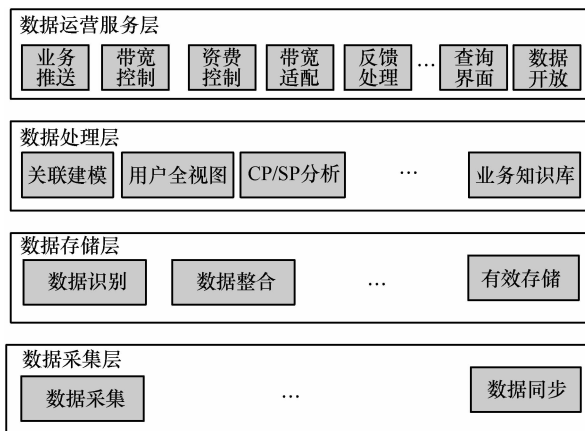


图 3 大数据平台功能架构

挖掘,根据关联性构建关联模型,生成知识库和用户、CP/SP/SI的全视图。最后,数据运营服务层结合运营层面的要求,一方面实现信息能力的开放和共享,另一方面根据运营策略执行策略指令,将信息提供给网络控制设备或者业务推送网元。数据运营服务层具备反馈机制和流程,支撑市场营销策略的修订和控制策略的灵活动态调整。

3.2 大数据平台的技术

基于集中、大容量、多样化和实时性等特点,大数据平台目前只能采用云计算的方式实现。具体来说,通过云数据库构建海量数据的存储空间,基于分布式计算实现海量的数据分析处理。从技术上,该平台的软硬件必须满足以下条件:

- 具备海量的协议识别能力;
- 支持多种数据采集工具;
- 支持常见和新型的数据挖掘模型;
- 兼容结构化和非结构化数据;
- 采用分布式海量处理技术;
- 支持存储虚拟化,支持数据的冗错、负载均衡和灵活扩展;
- 支持数据存取和查询,支持以认证数据等为关键字整合用户数据。

相应地,大数据平台的设计理念和技术如下。

(1) 基于 SOA 理念

基于 SOA(面向服务架构)理念,遵循松耦合、模块化的原则,解除数据和应用之间的耦合性。上层的应用逻辑基于实际的运营策略,调用底层的数据。数据和应用逻辑分离的方式提高了灵活性,同时适应个性化的需求,可以随时根据运营策略产生新的应用逻辑。

(2) 支持多协议识别

考虑不同的应用,通信网络从低层往上,支持的协议有上千种。理论上,大数据平台应该支持所有的通信协议,如 SNMP、WAP、DNS、Telnet、P2P、SIP、RADIUS 等。具体操作上可以将数据采集层和数据存储层的数据采集模块和数据识别模块合设,以实现数据的统一和规范化。

(3) 支持数据的有效采集

数据采集是指通过网络爬虫或标准接口等方式从数据源上获取数据信息。对于网站等非结构化数据,一般通过网络爬虫或者开放的 API,该方法可以实现将网页中的非结构化数据以结构化的方式存储到本地数据文件,并支持附件与正文的自动关联。数据采集工具有

很多种,目前在电信网络中采用较多的是 DPI 或标准接口的方式。

(4) 支持数据关系的挖掘

由于数据之间的关联关系并不遵循一定的模式和规则,大数据平台要支持各种类型的数据挖掘技术,包括假设检验、方差分析、回归分析、逻辑回归、聚类分析、因子分析等统计方法和决策树、向量机、神经网络、朴素贝叶斯等分析模型和方法。

(5) 使用流数据处理技术

使用流数据处理技术,将数据视为流动的,在动态过程中进行分析和计算。流数据处理技术不仅支持结构化数据的处理,同时支持非结构数据的连续流处理。流数据处理的由事件或数据触发,始终在线,结合分布式计算,非常适合大数据平台的实时性和动态处理数据的需求。

(6) 使用云数据库

传统电信 IT 系统的数据为结构化数据,而基于 Internet 应用的数据多为非结构化数据,如 Web、E-mail、SNS、文档共享等应用产生的数据。正是因为数据量很大,大数据的类型很复杂,如存在结构化、半结构化、非结构化数据,而且数据有不同的来源和标准,数据量大小与实时性要求不同,存放和处理的困难很大。云数据库提供了基础条件。云数据库采用云计算的相关技术,屏蔽了底层存储的差异,是存储虚拟化的一种表现形式、云计算的一种 SaaS 模式;云数据库支持数据节点动态伸缩与热插拔,提供数据的容错机制和安全管理,特别适合大数据的要求;同时兼容非结构化和结构化数据。对于互联网等需要大容量存储需求的业务,可使用新型非关系型数据库技术,进行分布式存储和处理。云数据库结合分布式存储的方式,使数据模式更加灵活,而且提高了可用性和容错性。

(7) 采用并行处理和分布式计算技术

云数据库存储了海量的数据,涉及大量的计算,因此需要一种不同于传统方式的编程方式和数据处理方法来支撑。通常采用分布式计算或并行处理技术。分布式计算在应用上通常结合新型的非关系型数据库,适合海量数据的存取管理,在支持在线分析处理(OLAP)方面具有较大的优势。并行处理技术可以支持同时的在线分析处理和在线事务处理(OLTP)。在大数据的场景下,并行处理只有部署在云基础设施的基础上,才能具备良好的可伸缩性;不同



于传统意义上的并行处理,更像分布式计算。业界典型的数据处理解决方案架构 Hadoop 是一种基于新型的非关系型数据库的大数据处理平台,采用了分布式计算技术。

以上几种构建技术是相辅相成的。总之,综合利用云计算模式、分布式技术和云数据库,不需要复杂的模型和算法就可以处理大数据。目前大数据平台的多种技术如何更有效地融汇贯通,业界还在继续探索,大数据在模型构建、体系架构、安全和性能优化方面还有很多需要深入探讨和完善的地方。

4 大数据平台应用实例

这里举个简单的例子:某用户特别喜欢观看节目,经常在 IPTV 上看大片,也常关注新的影视产品。根据互联网上的最新消息,影片《泰囧》马上就要在影院首映,估计高清版 2 个月后可以 IPTV 上线。通过分析,大数据平台发现用户 A 是一个影迷,经常反复收看喜欢看的电影。大数据平台还发现, A 经常用“院线通”订购首映影票,并且喜欢看完电影之后订购附近餐饮或者下载电影彩铃。据此,大数据平台针对与用户 A 有共同喜好特性的用户生成一种运营支撑应用,向该类用户推送影片《泰囧》的放映消息和订购信息;用户通过“院线通”订购电影票之后,在适当的时间和地点为用户推送大众点评优惠订餐信息和彩铃下载链接或彩铃套餐优惠信息。随着高清版《泰囧》在 IPTV 的上线,大数据平台根据用户 A 目前使用 ADSL 2 Mbit/s 接入的业务信息,推送免费试看 8 Mbit/s 接入的《泰囧》高清片段的链接,并提示用户 A 升级网络接入带宽到 8 Mbit/s,同时提供接入业务的升级手段。

该场景的相关业务流程如图 4 所示。

5 结束语

总之,基于大数据的运营商精确运营平台通过采用大数据技术,精确了解用户行为及管道流量组成,提供业务分析、用户建模等基础能力,为运营商提供业务个性化推荐、用户行为分析以及精准营销(动态套餐优化、定向营销

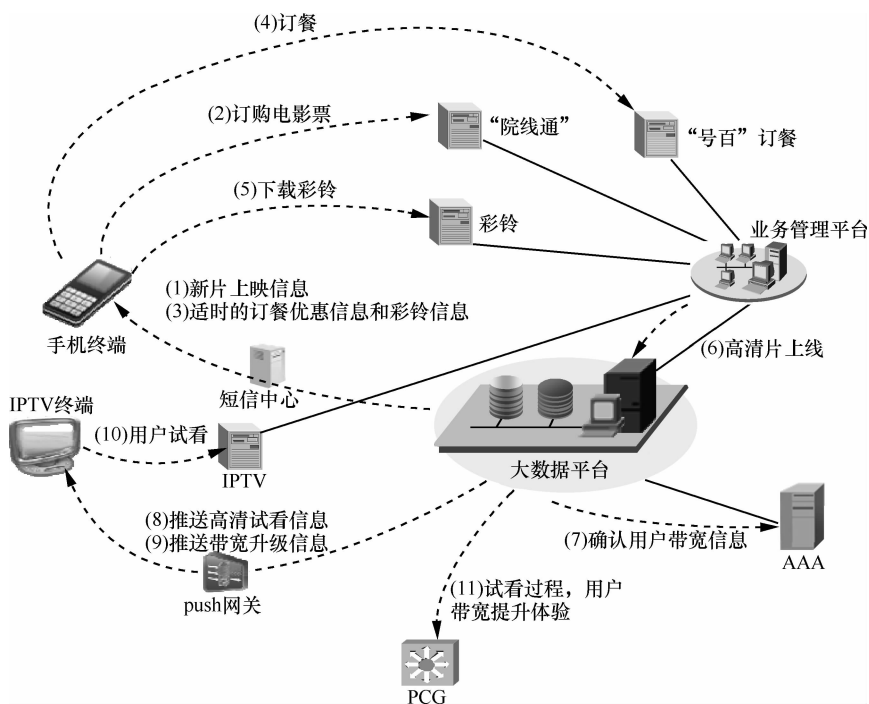


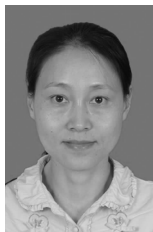
图 4 大数据平台应用举例

等)运营能力,将助力提升运营商的运营能力,提高移动互联网环境下运营商的竞争力。

参考文献

- 林子雨, 赖永炫, 林琛等. 云数据库研究. 软件学报, 2012, 23(5)
- 余长慧, 潘和平. 商业智能及其核心技术. 计算机应用研究, 2002(9)

[作者简介]



刘洁,女,硕士,中国电信股份有限公司广东研究院高级工程师、网络规划一级工程师,主要从事业务网络的研究、规划工作,发表学术论文 30 余篇,近两年主要负责云计算体系架构和基于云的业务网络架构研究工作。



王哲,女,博士,中国电信股份有限公司广东分公司高级工程师、网络发展部副总经理,主要从事电信宽带、业务及核心网络技术发展和运营规划工作,主持中国电信股份有限公司广东分公司云计算的应用和试点相关工作。

(收稿日期:2013-03-05)