



大数据环境下电信数据服务能力开放研究

刘 春¹, 邹海锋², 向 勇¹

(1. 中国电信股份有限公司广东研究院 广州 510630;

2. 深圳市远行科技有限公司 深圳 518000)

摘 要: 数据资源是电信运营商的核心资产,在数据激增的环境下,数据服务能力开放有利于运营商增加管道的价值和收入,但对其运营能力也提出了更高的要求。在研究大数据环境下数据服务能力开放业务需求的基础上,明确了数据服务内容,提出了数据服务能力开放平台的建设思路、体系架构和开放机制,对运营商加强能力开放平台建设具有积极的指导意义。

关键词: 大数据;数据服务;能力开放

doi: 10.3969/j.issn.1000-0801.2014.03.028

Research on Telecom Data Service Open Ability Under the Environment of Big Data

Liu Chun¹, Zou Haifeng², Xiang Yong¹

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

2. Shenzhen Vispractice Technology Corporation, Shenzhen 518000, China)

Abstract: Data resource is the core asset of telecom operators, under the environment of the data surging, opening data service ability is advantage to the operator to increase the value and revenue of pipeline, but it also need higher requirements of the operating capacity. On the basis of researching the requirements of data service ability open business under the environment of big data, the data service content was clearly introduced, the construction ideas, system architecture and open mechanism of the data service ability open platform were put forward.

Key words: big data, data service, open ability

1 引言

随着企业信息化建设的不断深化、社会化网络的兴起以及移动互联网等新一代信息技术的广泛应用,全球数据规模及其存储容量正在迅速增长,数据的类型也变得复杂多样。海量多样化的数据对信息的有效存储、快速读取、检索提出了挑战,且其中所蕴藏的巨大商业价值也引发了对数据处理、分析的巨大需求。因此,大数据的概念应运而生,引用维基百科的定义,“大数据是指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”。

当前,大数据已逐渐渗透到各个行业和业务职能领域,以利用数据价值为核心的商业服务正在不断涌现,如数据资源服务、数据分析服务、数据开放平台服务等,大数据将成为重要的生产因素,它的演进将与生产力的提升有直接关系。

对电信运营商而言,其业务特性使它具备了较强的大数据特征。然而,运营商业数据在快速积聚的同时,并没有从传送的大量上层应用内容中获得更多收益,却面临收入增速放缓的困境。要真正扭转这一局面,运营商必须转变过去简单粗放的网络运营方式,顺应移动互联网“开放、

共享、合作、高效”等特征,逐步向综合服务提供商转变,向集约化运营模式转变。在大数据时代,数据服务能力开放是这一转变的有力推手。

业务数据是运营商的核心资源,结合大数据技术,电信运营商开放自身的数据服务能力,为外部合作伙伴提供精确的用户行为收集与分析、提供精准广告投放推荐能力等数据分析服务,是强化流量经营、开展规模化信息服务运营的有效手段。这将有助于实现向综合服务提供商、集约化运营的有效转变,增加管道的价值和收入,进一步抓住未来广阔的信息化市场,摆脱被边缘化和底层化的危机,获取更大的商业价值。

2 业务需求

2.1 业务应用场景

在电信行业,利用数据的核心价值,实现数据服务能力开放,可以满足以下多种业务应用场景要求。

- 实时营销:实时捕捉用户通信行为,通过用户实时信令信息与营销目标用户场景,分别设置匹配信息向用户推送营销内容,达到在适当时间,以适当方式,向适当用户推广适当业务的目的。
- 手机定向广告推荐:通过与后向商家合作,向用户行为符合后向商家要求(如用户到达指定地点)的用户推送相关合作商家的广告信息,以赚取后向商家的广告费,丰富电信运营商的收入模式。
- 用户点击流分析:对用户的上网行为进行采集与存储,使用户上网行为信息变得可查询、可统计,统计分析的结果可以指导相关网页的页面布局、产品开发等。
- 信息推送:用户行为、兴趣特征、位置等综合分析,为合作伙伴提供有针对性的信息推送服务,合作伙伴可以基于这些信息,开发第三方应用或进一步的数据深加工服务(如店址选择)。
- 市场调研:通过用户行为了解市场动态,分析市场结构,为合作伙伴提供市场结构、用户多维特征、竞争情况等分析。

可见,如何在有效保护用户信息隐私的前提下,高效、及时、便捷地向自营单位、合作伙伴等外部对象提供所需的数据是以上业务应用的关键。大数据环境下的数据服务能力开放,即运营商通过定期或实时采集海量的运营数据,在异构多样的数据中进行处理和分析,以分析结果的

形式呈现,以集约化的能力开放平台向外部提供各种数据服务。

2.2 存在的问题

运营商在实现集约化的数据服务能力开放时,将会存在较多的挑战和困难,有如下几个方面。

- 技术要求高:移动互联网带来的海量数据对处理技术提出更高的要求。
- 整体分析难:业务平台割裂,数据散落地分布在各系统,难以支持整体分析,也不利于形成统一的客户行为视图。
- 能力分散:用户获取相关数据服务可能需要通过多个渠道或平台申请相关的数据服务能力,没有统一的服务能力开放平台,造成服务能力分散。
- 资源浪费:分散的服务能力,自然需要分散的计算、存储、网络资源,重复的服务能力开放必然造成资源的多倍使用。

电信运营商需要分析内外部数据集约化运营的共享需求和业务应用,以便于提供高度共享和集约化的数据服务能力。

2.3 总体需求

开放能力、创新模式并建立良好的合作机制,是电信运营商形成具有产业影响力的互联网生态系统的必经之路。数据服务能力开放平台的定位为提升运营商管道价值和提供综合信息服务的主体设施,是深化互联网流量经营的主体工具,是实现数据服务能力开放的主体平台。

- 建立基于横向业务网络体系架构的信息汇聚平台,通过对网络、应用客户端、业务平台和IT系统数据的采集,获取业务行为信息,采用大数据技术构建统一的用户行为信息数据库;采用建模技术实现数据挖掘和行为分析,为运营商的自营单位(如基地业务平台)或合作伙伴提供个性化数据分析能力,为精准营销和精准广告投放服务提供数据支持。
- 实现数据分析应用不同层次和权限的开放,如数据仓库、数据集市、分析模型、分析应用等;提供面向自营业务、第三方应用和合作伙伴等不同权限的能力开放;此外,考虑用户数据在传输、存储等方面的安全保障,并避免用户隐私保护在法律、技术上出现重大问题。



- 提供简单易用的开放接口,封装成标准统一的能力开放 API 供第三方调用,提供简单易用的文档说明,降低第三方开发者开发门槛;满足“一站式”开发、测试、运营的服务需求。
- 集约化的平台运营,按照一体化运营模式进行平台建设和流程优化,着重加强通信服务、电子渠道、用户行为分析等领域的全网集约化运营能力。

3 数据服务内容

数据服务能力以 API 形式提供给外部,首先要明确数据服务的形态,即通过数据服务开放平台可以提供的数据类型、形式等。本节通过分析业界数据开放的案例,结合电信行业特征,明确数据服务的内容。

3.1 业界数据服务开放

数据服务开放已经开始覆盖互联网企业、政府行业,其中较为主流的两个案例为美国的政府服务数据开放以及国内淘宝网在电商行业中的数据开放。

(1) 美国政府“一站式”数据下载

美国政府为了解决公众对政府服务信息查找困难、不便的问题,于 2009 年 5 月宣布实施“开放政府计划”,提出利用开放的网络平台公开政府信息、工作程序和决策过程,开通了“一站式”政府数据下载网站——Data.gov^[1],要点如下。

- 以各种标准数据接口提供易于发现、访问和理解的数据,方便用户下载,改善联邦数据利用率。
- 整合现有 1 100 多个数据中心,削减到 100~200 个,实现数据集中和节能降耗,并制定完备的法律体系保障信息的公开与获取。
- 采集与发布的政府数据主要有两类:原始数据集,涵盖美国的人口特征统计数据、GIS 信息、环保、教育、能源、地域、法令等相关主题的政府信息;综合数据集,通过原始数据与地理信息的叠加,生成地理空间信息服务和综合地理数据。
- 提供标准的 API,方便第三方机构开发特色应用,如福布斯杂志利用人口流动数据开发人口迁移的应用,方便查看地点间的人口迁出和迁入情况,为企业提供决策参考。

(2) 淘宝数据开放

淘宝网“大淘宝”战略中,生态商业系统的核心竞争能力是利用互联网信息技术累计的数据提高自身开展电子

商务业务的能力,即解读互联网数据能力和对快速市场的应急能力,通过数据开放提高其在电子商务业务中的核心竞争能力,从以下 3 个层面进行开放:

- 涉及电子商务行业的宏观数据以及让消费者了解最新消费风向标的数据,淘宝实行免费开放策略;
- 涉及各个行业市场情况、消费者行为研究等的商业数据,淘宝通过商业方式开放;
- 涉及消费者个人隐私、企业商业隐私的数据,淘宝绝对保护,防止任何泄漏。

淘宝数据的开放,将推进企业提高开展电子商务活动的核心竞争能力,有助于促进整个电子商务生态链中各个物种的成长。

3.2 电信运营商数据服务内容

3.2.1 数据服务开放考虑因素

总结以上两个案例,电信运营商实现数据服务能力开放,需要考虑表 1 中的 4 种因素。

表 1 数据服务开发的考虑因素

因素	说明
内容	实现数据查询和下载服务、数据分析服务、数据决策服务
原则	满足可靠性、便捷性、公开性等原则
提供方式	主要以 API 提供为主
管理措施	发布数据使用范围、数据安全权限管理、隐私管理等规范、措施

3.2.2 数据来源

在明确数据服务内容之前,首先总结一下运营商的大数据来源,主要分为以下 3 个方面。

(1) 用户提供的信息

包括用户的许可和偏好所得到的数据、用户与运营商的合同数据、用户社会网络数据、合同或用户注册中的直接内容、对用户的调研数据和用户使用的电信产品。

(2) 直接获取的数据

通过用户行为分析(使用的各类介质或传感器)可直接获得的数据,包括所使用的物理终端连接状态、终端类型、位置信息、用户网络流量数据、用户点击广告和推荐的数据、用户购买/下载的习惯、用户网站浏览行为、电视观看习惯以及 CDR(call detail record,由电话交换机产生的计算机记录,包含经过的每个电话呼叫的详细信息)。

(3) 推理分析挖掘数据

除了上述主动获取的数据之外,还需通过推理分析挖掘对运营商有用的数据,包括对用户离网的风险进行分

析、对用户社交圈数据进行分析 and 挖掘等。

3.2.3 数据服务内容

通过对业界数据服务的分析,结合电信运营商的业务需求和数据来源,总结出以下数据服务内容。

(1) 基础数据查询能力服务

提供对电信业务数据的高性能查询服务,实现对客户基础信息、位置信息、订购信息、终端信息、用户行为等数据的提供。

(2) 数据分析服务能力服务

提供对外数据分析能力服务,为外部提供数据分析结果,如用户的应用偏好(微博、游戏、阅读等)、消费习惯(历史消费情况、位置出现规律、付费习惯等)、行为偏好(如军事、旅游等)以及个性化信息推送(利用电信业务的综合行为分析结果,提供个性化推荐结果)。

(3) 数据运算能力服务

随着业务需求的不断多样化以及技术架构的进步,探索对外提供大数据运算服务,分为以下两方面。

- 数据分析挖掘能力:通过数据隔离、沙盒管控等措施,为合作伙伴提供快速的分析挖掘能力。
- 实时数据探索服务开放:利用电信的大数据综合分析结果,为合作伙伴提供实时数据探索分析服务。

4 能力开放平台架构研究

实现面向大数据的数据服务能力开放,需要建立基于横向业务网络体系架构的信息汇聚平台,以下对平台架构进行研究。

4.1 平台建设思路

平台建设思路介绍如下。

- 综合数据平台的提供者,拓展电信运营商的业务范围:综合企业内外部数据资源,对外开放数据服务,帮助合作伙伴提升精确营销能力;为行业客户提供大数据的解决方案和数据分析服务。
- 支撑商业模式创新,通过对外的数据服务支撑,更好地聚集价值链中的优秀合作伙伴;面向后向客户收费,进行收入分成或内容收费。
- 标准化操作,构建“表达标准化、实现模块化”的能力封装与开放体系。

4.2 平台架构

能力开放是对底层复杂的业务、技术、流程实现进行抽象和封装,对外提供开发运营环境^[2]。结合电信运营商的业务需求以及IT建设现状,提出数据能力开放平台架构,如图1所示。

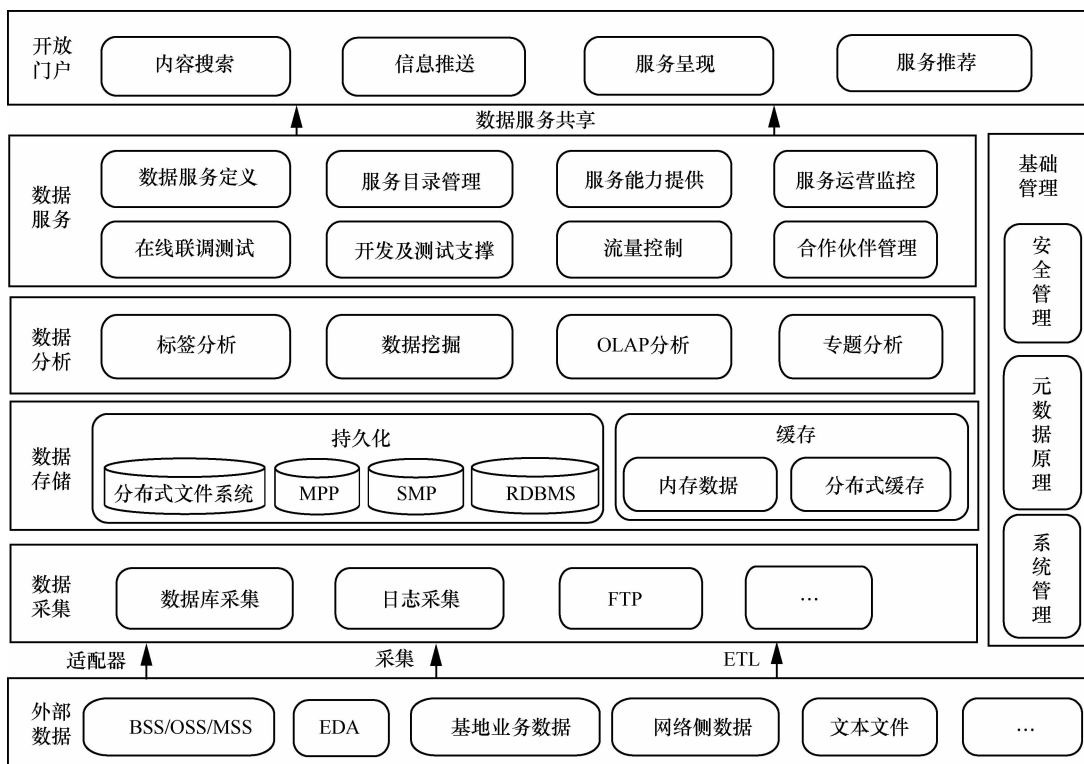


图1 数据服务能力开放平台架构



(1)数据源

数据源主要包括 IT 应用系统的业务数据(如 CRM 系统、计费系统、网管系统、数据仓库等)、电信移动终端数据(如连接状态、终端类型、位置信息等)、网络通信协议和数据(如 DPI、NetFlow、网络及拓扑结构、移动网 CDR 等)、业务平台数据(如业务订购信息、浏览点击行为等)。

(2)数据采集

数据采集层需要支持关系数据库、日志数据、文件数据、XML 文档数据等数据源的采集,满足联机/脱机、实时/定时/周期性等采集方式要求;采用任务驱动的采集机制,根据采集任务的触发条件和优先级等,加载、激活和协调采集任务的运行,要求所采用的方式对现有系统的影响最小。

(3)数据存储

从所采集的数据特征可知,其数据结构复杂多样且体量巨大,需要改变目前以结构化为主的单一存储方案;要求满足不同的存储要求,支持 SMP、MPP、RDBMS 等多种架构的数据库以及内存数据库、分布式缓存等技术能力。采用如下策略实现:对非结构化数据,采用分布式文件系统进行存储;对结构松散、无模式的半结构化数据,采用面向文档的分布式 key/value 存储引擎进行存储;对海量的结构化数据,采用 shared-nothing 分布式并行数据库系统进行存储;支持行、列混合存储;支持存储的在线、线性扩展;构建各存储引擎之间的连接器,使得非结构化数据在处理成结构化信息后,能方便地和分布式数据库中的关系型数据快速融通,保证大数据分析的敏捷性。

(4)数据处理

为了满足海量数据处理以及响应时效高的要求,需要具备分布式批处理和实时流处理的计算能力。

在分布式批处理层面,需要支持海量数据的并行化处理;自动处理节点/任务的故障检测和恢复;处理节点可以灵活加入和退出,自动感知节点状态并进行处理;计算任务灵活调度,从而提升执行性能。

在实时流处理方面,支持数据的连续注入、连续分析,实时对流数据进行分布式并行分析计算(达到毫秒/秒级);以极高性能分析处理结构化和非结构化动态数据流(如关系、文本、图片、视频等);支持事件驱动,捕获实时事件,触发相应的处理流程;具备高速数据传输、时延极低、速率极高等特性。

(5)数据分析

数据分析层主要对经过存储和处理的数据进行 OLAP

分析、搜索引擎、专题分析、数据挖掘、标签分析等操作。海量数据使得计算速度难以保证,数据结构变化导致计算模式变化。在实现以上数据分析时,需要融合 MapReduce、流计算等先进技术思路,要求支持非结构化数据分析(如文本分析和挖掘技术、视频分析技术、网络爬虫关键字分析技术、客户标签化处理等);支持非结构化数据向结构化数据转换;实现基于结构化和非结构化数据混合环境下的数据分析处理;具备分布式并发的大规模计算能力;提供数据分析模型管理和分析应用创建能力。

(6)数据服务

数据服务层将屏蔽底层针对各类数据服务需求的数据处理过程,将加工后的数据、应用等通过集中的数据/服务提供功能,为外部合作伙伴提供数据服务能力,简化数据共享逻辑,集约化数据分析能力。

数据服务能力的对外提供需要满足敏捷、安全可靠、能力复用的总体目标。功能层面上,要求满足共享需求的数据资源以服务的方式暴露到平台上,通过数据服务定义、数据服务目录、服务路由、传输加密、在线联调测试、流量控制等功能^[9],为接入的外部平台提供数据服务,同时也实现对服务从需求提出→开发→发布→部署上线→维护更新→下线全过程的服务生命周期的管理;技术层面上,满足异步、消息实时响应、发布订阅、并行处理、松耦合等核心技术需求。

(7)数据门户

数据门户是开放平台的数据应用统一入口,面向不同角色的使用者,可以实现数据服务内容的搜索和推荐,能够通过短信、邮件等方式把关键业务指标、系统分析应用推送给目标用户。

(8)基础管理

基础管理主要包含安全管理、元数据管理和系统管理等功能。元数据管理贯穿于大数据分析子平台的构建、运行和维护整个生命周期;安全管理是数据管理工作的重要组成部分,包括传输安全、存储安全和访问控制;系统管理提供对认证、权限、用户等基础信息的管理。

4.3 能力开放机制

从数据服务能力开放的业务特性看,数据服务主要提供一对多的服务共享方式,要满足高性能、高实时性的服务并行处理要求,且要与外部平台实现充分的解耦。业界能力开放平台在提供服务时,主要采用传统的 SOA 思路进行构建,此种方式存在以下几个方面的问题。

- 实时性:主要基于消息的同步模式机制,其流程处理方式为顺序、线性,即只有前面步骤处理完成才能发起后续步骤和服务调用,这影响到服务调用的性能和时效。
- 服务效率:针对一对多的服务共享场景,会开发多个服务且增加路由服务,服务提供方必须要知道所有消费方,并对所有消费方进行服务调用和数据分发,这无疑给平台增加了性能压力。
- 耦合度:企业服务总线能起到松耦合的作用,但其建设过程是渐进的。当前的消息机制多采用同步模式,当服务消费/提供者两方中的任何一方存在系统异常时,都将直接导致另一方的功能系统和服务调用失败,业务系统间没有彻底实现解耦。

SOA 能把线性、可预测内容的服务连接起来,但缺乏对动态实时业务做出应答的能力。数据服务提供需要满足异步、消息实时响应、发布订阅、并行处理、松耦合等核心技术需求,而通过事件驱动架构(event driven architecture, EDA)模式建立的系统可对动态实时业务做出有效的处理,允许复式、不可预测、异步的事件并行地发生和在单一的活动中被触发,并且由于服务可以像事件一样被触发,因此 EDA 实际上是 SOA 的一种有效补充^[4]。

经过以上分析,本文所述平台的数据服务能力采用 EDA 服务共享机制,具备以下特性。

- 异步:支持异步活动,服务发布后不必再关心消费方是否能收到响应,也不必在源和目的端之间维持一条活动的链路。
- 发布/订阅:支持一对多或多对多的服务交互,在平台中发布一个关于事件的信息到网络,许多其他已经订阅和授权的系统可以收到消息并做出相应的动作。
- 解耦:实现与外部系统的充分解耦,服务在两个系统间交互时不需要知道对方的详细信息。

在实现以上服务机制时,主要满足以下功能要求。

- 异步消息支撑:平台架构需要保证当外部服务请求事件发生时,能传送异步消息。采用 JMS 实现消息的创建、发送、接收和读取,通过 JMS 接口,消息生产者和消费者能够以点对点或发布/订阅模型发送和接收消息。
- 事件管理:将服务提供方/消费方所发生的工作定义为事件,平台的体系结构必须保证有相应的功能

可以识别、定义和聚合事件,以便于统一管理,这包括在事件发生时和发生后都可以访问,并且可以在特定的业务场景中识别和访问一系列事件。

5 结束语

本文分析了在大数据环境下,电信运营商数据服务能力开放的业务需求、数据服务内容,重点研究了数据服务能力开放平台的建设思路、体系架构,针对数据服务提供的业务特点,研究了集约化运营环境下运营商数据服务共享的机制。本文的研究成果对电信运营商能力开放平台的建设具有较好的指导意义,为运营商向集约化运营、综合服务提供的战略方向转型提供了技术参考。而在本文的研究领域中,对于数据安全性问题,特别是涉及个人隐私的内容,后续将做更进一步的研究。

参考文献

- 1 刘增明,贾一苇. 美国政府 Data.gov 和 Apps.gov 的经验与启示. 电子政务, 2011(4)
- 2 吴钢. 移动互联网时代电信运营商的商业模式——能力开放. 信息通信技术, 2011(1)
- 3 刘斌,王桂荣,华竹轩. 移动互联网环境下电信 IT 能力开放研究. 广东通信技术, 2011(12)
- 4 季云峰. 基于 SOA 的 EDA 的研究和实现. 软件, 2012, 33(7)
- 5 刘秋生,李红贵. 基于事件驱动 SOA 架构的企业应用集成模式研究. 中国管理信息化, 2009, 12(4)
- 6 赵钧. 构建电信物联网开放数据服务体系思考. 电信科学, 2012, 28(2)

[作者简介]



刘春,男,中国电信股份有限公司广东研究院工程师,一直从事数据仓库、数据挖掘及大数据技术等方面的技术研究工作。

邹海锋,男,现就职于深圳市远行科技有限公司,一直从事数据分析、挖掘以及 IT 规划等方面的研究工作。

向勇,男,中国电信股份有限公司广东研究院高级工程师,主要研究方向为电信 IT 系统研发及数据应用等,目前专注于大数据相关技术的研究及应用。

(收稿日期:2013-09-24)