



专栏：网络新技术

电信运营商大数据能力开放平台建设

张海峰, 董昭, 李娟, 齐磊, 顾慧琼

(中国移动通信集团设计院有限公司, 北京 100080)

摘要: 大数据正加速在各行各业的应用, 数据的开放有助于运营商内部的聚合和共享, 有助于服务千变万化的应用场景, 对运营商开拓新业务具有重要意义。分析了大数据和云计算进行结合的背景, 提出面向云计算的大数据开放平台整体架构和核心能力, 对大数据开放平台的建设思路进行阐述, 并对运营商在具体工程中进行大数据 PaaS 平台的规划和待解决问题进行阐述, 为相关研究人员提供参考。

关键词: 大数据; 能力开放平台; 云计算; PaaS

中图分类号: TP319

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017106

Construction of big data open platform for telecom operators

ZHANG Haifeng, DONG Zhao, LI Juan, QI Lei, GU Huiqiong

China Mobile Group Design Institute Co., Ltd., Beijing 100080, China

Abstract: Big data is accelerating the application across various industries, the opening of the data will help operators in the internal aggregation and sharing, and help to serve the ever-changing application scenarios, this is important for operators to develop new business. The background of combining big data with cloud computing was analyzed, overall architecture of big data platform for cloud computing was proposed, then the big data open platform implementation scheme was focused on, the platform planning and the problems to be solved in the project were carried out, and the reference for the related researchers were provided.

Key words: big data, open platform, cloud computing, PaaS

1 引言

随着近些年全球互联网的快速发展, 大数据也逐渐从技术驱动转变为企业的商业驱动, 如数据资产服务、数据分析服务、数据开放平台服务。未来几年, 企业经营优化和数据资产变现将成为主要诉求, 基于大数据的商业模式创新将会爆发式增长。

在大数据成为趋势、成为国家战略的今天, 如何最大限度地发挥大数据的价值成为人们思考的问题。无论是互联网企业、电信运营商还是数量众多的初创企业, 大数据的开放与变现显得尤为重要。在探索大数据商业模式的同时, 大数据正加速在各行各业的应用, 大数据不仅为人们的购物、出行、交友提供了帮助, 甚至还在各类重要的事件中发挥作用^[1]。

收稿日期: 2017-03-15; 修回日期: 2017-04-05



2 大数据开放的诉求及其内容

2.1 大数据应用前景

电信运营商掌握了丰富的用户身份数据、语音数据、视频数据、流量数据和位置数据，数据的海量性、多元性和实时性使运营商具有经营大数据的先天优势。随着智能手机和高速网络的普及，运营商能够获得的用户行为数据还将更为丰富。运营商的大数据应用主要分为面向内部的应用和面向外部的应用两部分。

(1) 面向内部的应用

- 促进市场营销：通过用户偏好分析，及时、准确进行业务推荐，包括流量经营、存量经营、集团客户经营、终端营销。
- 改善用户体验：通过大数据分析用户行为，运营商可以改进产品设计，并通过用户偏好分析，及时、准确进行业务推荐，强化客户关怀，这样就可以不断改善用户体验，增加用户的信息消费以及对运营商的黏性。
- 优化网络质量：在网络层面，可以通过大数据分析网络的流量、流向变化趋势，及时调整资源配置，同时还可以分析网络日志，进行全网络优化，不断提升网络质量和网络利用率^[2]。

(2) 面向外部的应用

电信运营商也从传统BI系统逐渐向大数据平台进行转型，未来的商业模式也不仅局限内部应用，更多的是面向政府、旅游、银行、交通等各行各业的开放合作。

- 实时营销：通过对用户基本信息、业务信息、消费信息、位置信息、行为信息的综合分析，形成全方位的用户画像。根据外部公司的需求，为互联网行业提供数据产品服务，指导用户精准洞察、广告精准投放，实现数据增值。

- 位置应用：通过离线挖掘与实时分析相结合的方式，对海量移动用户运动轨迹进行精准计算，分析特定场景（如地铁、公交、商场、美容院等）下人群的聚类特征，应用于城市规划、商圈选址、智慧旅游、交通规划、人流预警等领域。
- 研究报告：通过与行业的深入结合，实现数据跨行业的关联与分析，生成特定需求的行业分析报告。

2.2 大数据开放的意义

大数据平台的使用对象往往需要大数据平台提供一定的二次开发能力，包括数据、开发工具、分析工具、计算存储资源等一系列的环境，以更好地迭代服务千变万化的应用场景，因此大数据平台必须以开放化的方式进行建设。将大数据平台开放可以获取如下收益：

- 有助于企业内数据的聚合和共享，从而减少数据重复采集、重复处理、重复存储，降低大数据整体开发运维成本；
- 有助于打造开放的、共享的、百家争鸣的大数据应用生态环境；
- 有效支撑不同业务部门的个性化应用需求；
- 有益于大数据对外变现和产业合作能力，让数据产生更大价值。

2.3 开放平台的目标

虽然电信运营商的数据保有量大，但在推动数据对外合作方面也遇到很多困难，比如现有数据散落在多个管理域中，数据割裂，重复存储；平台能力方面没有进行统一规划，各部门及各省公司重复进行能力建设；系统模式单一，开发周期长，运维模式僵化。

因此，面向未来的大数据平台核心诉求如下。

(1) 数据资产全网统一

完成对全网B/O/M三域数据以及互联网DPI数据、各业务平台数据的统一采集、统一存储和

统一治理,形成企业级统一的数据资产,有效支撑数据的对外开放。

(2) 系统架构转型

打破现有烟囱式和省分独立建设模式,以“纵向拆分,横向整合”为原则,实现应用与平台的解耦,实现跨系统的能力共享。

(3) 能力对外开放

构建标准、开放的企业级大数据平台,满足各专业公司、省公司、外部开发者和使用者对全网数据日益迫切的个性化分析需求。

(4) 开发运维提升

转换传统的开发模式,引入合作伙伴,实现应用的快速开发;建设自动化的运维能力,有效支撑应用的快速部署。

2.4 开放的内容

大数据开放平台重点关注大数据平台的能力如何向应用开发者开放,应用开发者以何种形式使用这些能力以及为达成这些能力的开放平台必须具备和关注的管理能力。大数据开放平台的开放内容包括两部分。

(1) 数据开放能力

在可管、可控的前提下,实现业务支撑域、管理信息域、网络支撑域、DPI 数据及业务平台数据的开放。

(2) 数据处理能力

通过多租户的方式,提供大数据平台的计算资源、存储资源、网络资源、中间件、分析工具等能力,确保租户在共享能力的同时实现一定的隔离。

3 大数据开放平台架构及核心能力

3.1 大数据与云计算的结合

云计算按照其定义,具备五大基本特征:资源的共享、按需的便捷服务、广泛的网络访问、快速的弹性伸缩和可度量的服务。可以看到,云计算正是解决大数据对外开放的利器。

云计算提供基于多租户共享资源的消费模式以及弹性扩展能力,使得数据资产和平台能力实现了共享,同时为开发者提供了很好的便捷性,这使它成为大数据分析的理想选择。有了云计算技术,大数据的价值才能得到更好的转化。因此大数据和云计算的结合是必然的。

云计算分为 IaaS、PaaS 和 SaaS 3 种服务模式。电信运营商在云计算的发展过程中也是从 IaaS 入手,已经初步建设了集中的 IaaS 平台,实现了云存储、云主机、云数据中心等基础资源的统一管理。

随着云计算技术的日趋成熟,本文发现 IaaS 提供了硬件资源层的能力与服务,本身并不面向应用。IaaS 需要用户自行构建开发测试环境,自行管理和运维平台软件,天然与应用之间产生鸿沟。而当 PaaS 以自动化部署、一整套中间件服务、全流程的 DevOps 出现时,本文发现 PaaS 才是真正贴近于应用的平台。因为“更快、更便捷、更灵活的用户体验”才是云计算一直以来的核心。

图 1 说明了 IaaS 和 PaaS 的关系。

当更多的应用开发者开始转向微服务的时候,也就说明他们选择了 PaaS。因此,未来的大数据要做到开放、灵活,一定是要拥抱 PaaS 的。

3.2 平台目标架构

由第 3.1 节可知,未来企业级大数据开放平台需要依托在以 PaaS 为核心的基础上,整体平台架构如图 2 所示。

企业大数据 PaaS 平台主要包括:集成平台 I-PaaS、面向应用的能力组件平台 A-PaaS、API 开放平台以及 PaaS 管理平台。

(1) I-PaaS

提供集成和构建复合应用的能力,因为企业级 PaaS 平台不仅仅是应用部署平台,而且是复杂多租户环境和复杂应用环境下的共享基础设施平台。I-PaaS 解决的是多个应用间如何集成和交互的问题,满足应用从开发态到集成态的一系列调

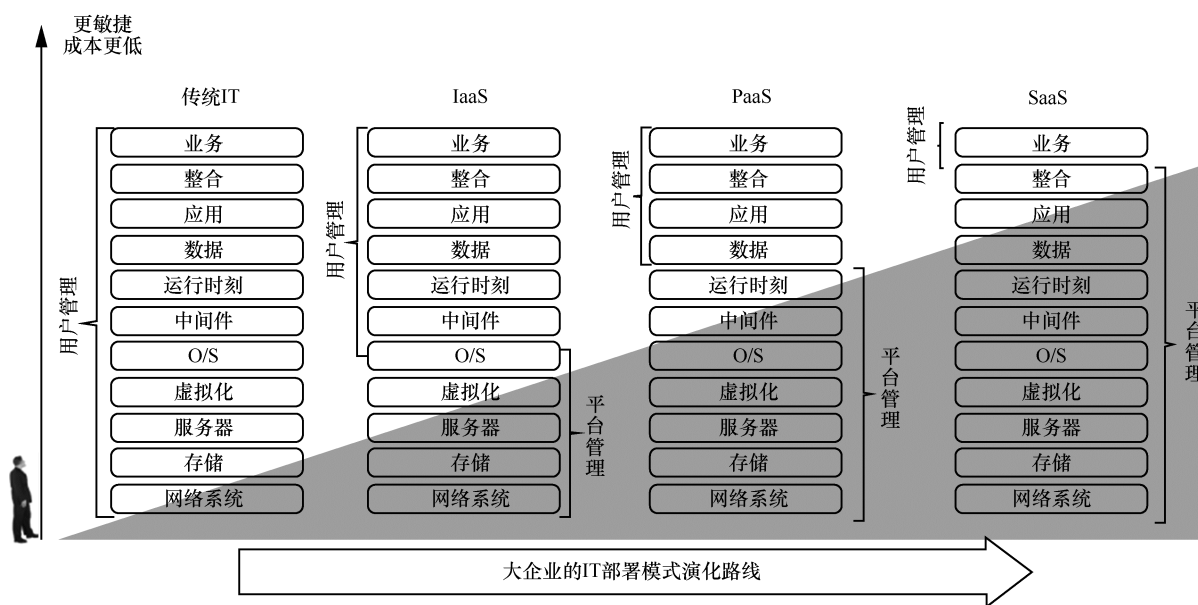


图1 IaaS 和 PaaS 的关系

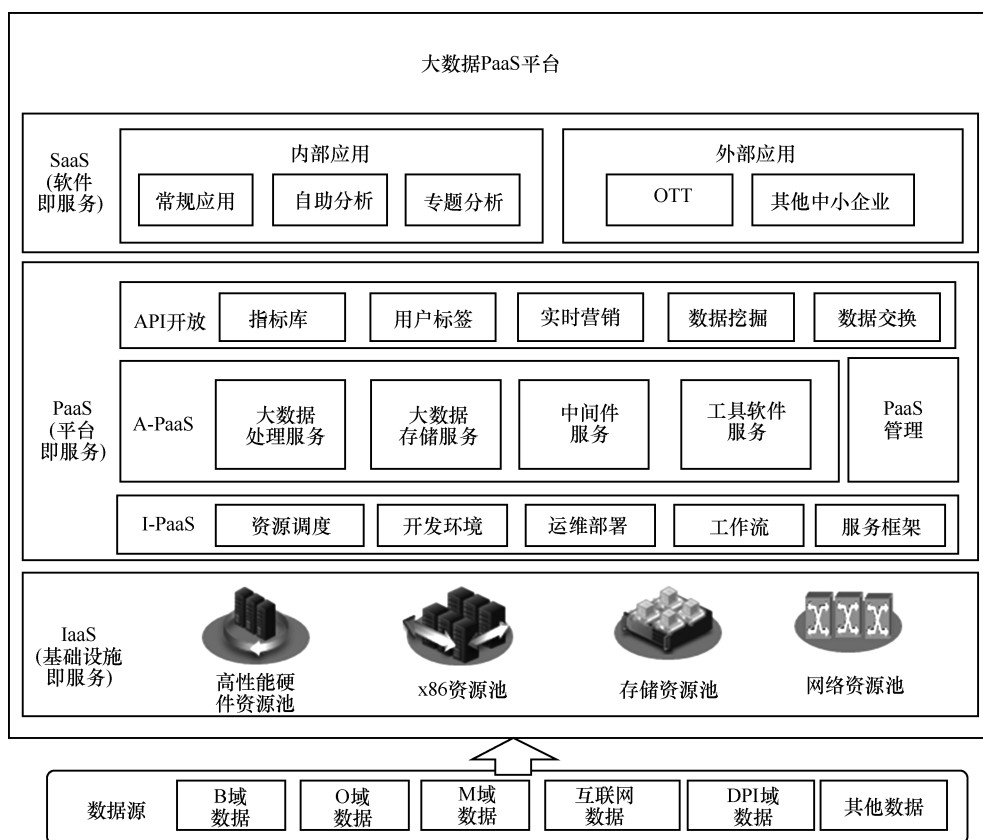


图2 企业大数据 PaaS 平台整体架构

度、集成、服务发现及工作流的问题。

(2) A-PaaS

用来承载和管理不同的应用，提供应用相关

的能力组件。A-PaaS 提供诸如 MPP、RDB、HDFS 等大数据分析能力，报表统计、自助分析、中间件等工具软件的能力以及多租户访问的支持

能力。

(3) API 开放平台

用来实现将底层的能力组件通过 API 的方式提供对外访问。

(4) PaaS 管理平台

提供诸如多租户管理、安全管理、运维监控等能力。

3.3 核心能力

大数据 PaaS 平台需具备的核心能力包括如下几点。

3.3.1 多租户

多租户的概念最早起源于软件领域,指一个软件实例服务于多个用户的架构,每个用户称为一个租户。多租户带来资源高度共享模式,提高资源利用率,降低单位资源成本。

建设基于多租户的能力开放框架需要具备如下功能。

(1) 租户管理

企业级省大数据平台通过租户开放的模式实现计算资源、存储资源以及数据的隔离管控,以保障资源的安全使用。租户的管理包括对资源的物理隔离和安全控制、对租户作业的调度。

(2) 租户权限管理

租户权限管理,是通过租户资源级权限设置、数据级权限设置两类权限管理,实现对数据开放、资源开放、工具开放 3 种基于多租户的开放能力的权限管控,确保租户内的资源、数据等访问安全。

(3) 租户资源调度管理

资源管理包含:资源分配(空间、权限、优先级)、数据隔离和安全控制;作业调度包含:租户作业优先级控制,业务优先级控制。在大数据平台多租户、多应用并存的情况下,作业的统一调度可以提高系统运行效率和资源的利用率。统一调度应能管控集群内的所有资源,包括计算资源和存储资源。统一调度包括对数据采集、数

据交互、数据处理等各类作业的调度,并支持跨系统的作业依赖和调度,对作业优先级和并发进行控制,同时支持可视化监控和管理功能。

基于 Hadoop 平台的大数据多租户实现方式:采用租户的方式,调用 Hadoop 的组件 Yarn、组件 HDFS、组件 Thrift 实现多租户资源的自动分配管理方式管理。包括以下几部分。

- 提供最小分配资源定义、共享资源定义和资源单位定义以及对应的 API; PaaS 管控平台根据 Hadoop 提供的 API 完成资源信息查询和分配功能。
- 数据共享与安全控制:实现不同 Hadoop 应用忙闲时资源复用,数据只保存一份;不同租户间的数据做到有效隔离,不能有非法的跨系统间互访。
- 资源优先级控制:根据租户优先级,来分配不同的资源容量。
- 业务优先级保护:统计分析型离线业务,优先级低于清单查询型实时业务。
- PaaS 管控平台基于 Hadoop 实现多租户能力,通过集成 Hadoop 平台的租户管理界面,实现租户的新增、修改、删除等功能,包括:支持在 Hadoop 平台的租户管理界面上新增、删除、修改租户;支持通过 Hadoop 平台提供的接口查询租户信息,包括租户对应的资源信息。

3.3.2 业务敏捷性

随着应用的微服务化与持续集成的兴起,传统的开发运维模式受到很大冲击,基于手工方式实现开发、测试、部署的时间周期长,而且软件版本和平台环境也无法做到完全统一。借助 Dorker 的成功和资源调度技术的实现, PaaS 平台可以轻松实现从开发测试到生产运维的自动化管理,实现应用快速上线。

(1) DevOps

DevOps 是一组过程、方法与系统的统称,用



于促进开发（应用程序/软件工程）、技术运营和质量保障部门之间的沟通、协作与整合。它的出现是由于软件行业日益清晰地认识到为了按时交付软件产品和服务，开发和运营工作必须紧密合作而提出的。

DevOps 不仅仅是自动化，更是一种新的文化，用于促进业务、开发、运营的沟通和协作。DevOps 的目标是通过建立持续交付的流水线，从代码开发到持续测试、持续集成、持续部署的能力。DevOps 加速软件开发，减少软件开发到上线的时间，并使开发和运维协作得更紧密。

DevOps 的价值具备以下几点。

- 代码的提交直接触发：消除等待时间，快速反馈。
- 全开发流程高效自动化：稳定，快速，交付结果可预测。
- 持续进行自动化回归测试：提升交付质量。
- 设施共享并按需提供：资源利用最大化。
- 每个变化对应一个交付管道：使问题定位和调试变得简单。

(2) Docker

Docker 是容器技术的一种，它运行于 Linux 宿主机之上，每个运行的容器都是相互隔离的，也被称为轻量级虚拟技术或容器型虚拟技术^[3]，是基于容器技术的轻量级虚拟化解决方案。

Docker 容器技术的部署架构如图 3 所示。

Docker 的优势在于秒级启动，秒级停止，空间资源占用极少。可实现集成级别的隔离，可在普通服务器上建立上百个 Docker 实例，并加快开发测试部署的速度，简化版本管理工作。它可解决资源利用效率低、单物理机多应用无法有效隔离、运维部署不便、版本管理复杂、迁移成本高等问题。

Docker 的价值包括以下几点。

- 更快的交付部署：以“集装箱”为单位来管理应用，使开发、测试、运维直接使用同一镜像来部署。
- 高效的虚拟化：直接使用操作系统的内核隔离功能，是轻量的虚拟化技术，实现应用隔离和多租户环境。
- 轻松迁移扩展：一次创建，随处运行，可以运行在任意平台。
- 便捷的管理：减少人工维护和手工更新的工作，修改以增量的方式进行更新。

3.3.3 运维自动化

传统基于 IaaS 的应用需要单独部署中间件、数据库、应用程序等，且大规模分布式环境下的应用部署不具备自动化能力。PaaS 平台支持一键部署、故障恢复、弹性伸缩的能力，全面支持运维自动化。

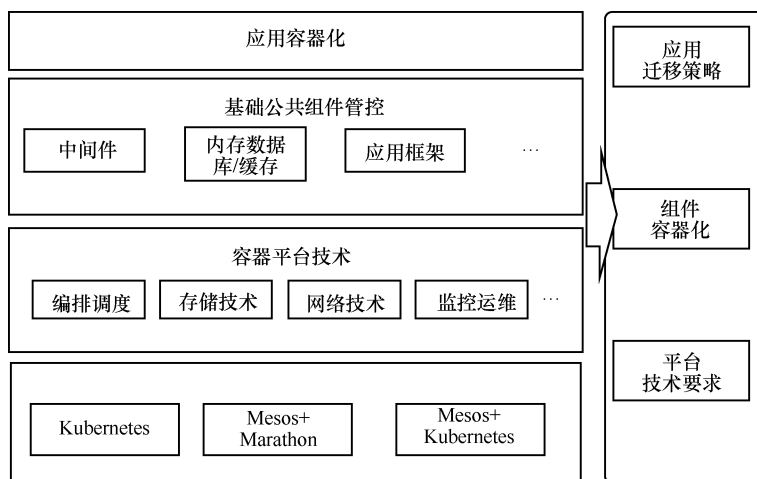


图 3 Docker 容器技术的部署架构

3.3.4 丰富的能力组件

PaaS 平台提供诸如 MPP、RDB、HDFS 等大数据分析能力, 报表统计、自助分析、ETL 等工具软件的能力, Redis、Kafka 等中间件的能力。让开发者聚焦业务代码的开发, 通用的能力用 PaaS 平台来统一提供。

3.3.5 数据资产服务

大数据平台中最有价值的部分就是数据本身, 一方面要实现数据资产的统一视图, 另一方面也要做到数据的方便获取和安全隔离, 确保数据的有效共享。

数据开放包括如下几部分。

(1) 现有数据资产盘点, 包括资料类数据、行为类数据、终端类数据、收入类数据、交往圈数据、位置类数据、上网行为类数据、客户标签类数据及集团客户类数据。

(2) 构建开放平台引入合作伙伴开发者。

(3) 实现多样化数据服务模式。

- PaaS (平台即服务) 适用于第三方带私有的数据到数据工厂跟企业数据进行融合加工提取信息。建设数据工厂, 包括数据融合、应用开发、数据目录及开发者门户。
- DaaS (数据即服务) 适用于需求方查询调用企业内部的数据。
- SaaS (软件即服务) 利用企业数据能力构建应用供第三方使用, 建立应用商店, 如应用开发、应用注册及应用下载。

4 大数据开放平台演进思路

近年来, 电信运营商已经初步构建了“传统数据仓库+MPP+Hadoop”的混搭架构。面对未来大数据开放平台的建设需求, 需要重点推动全网数据整合和统一大数据平台基础能力, 构建多样化的大数据服务能力, 对内支撑精细化运营管理与精准营销, 对外联动数字化生态, 支撑整体大数据开放战略。

电信运营商大数据开放平台的演进主要包括如下几个阶段。

(1) 采取“先立后破”, 推动集中平台能力

针对企业内部多套分析系统, 需要打破目前各个系统独立建设、重复存储、数据不共享的模式, 建设跨域统一的数据基础资源层, 完成对跨域数据的统一采集、统一存储和统一计算, 实现企业内部数据统一视图。

(2) 初步构建服务化能力

①构建平台能力

按照先易后难, 初步构建具备核心能力的 PaaS 平台。实现 I-PaaS 中的资源调度、运维部署、多租户隔离等框架功能, 实现 A-PaaS 中的一部分组件能力, 如 Spark、HDFS、中间件、自助分析等能力, 实现一部分 API 对外访问能力。

②推动关键产品的优先部署

形成部分共性应用在总部节点支撑, 对外提供全网能力级的 DaaS、SaaS 产品, 如征信、位置能力, 构建大数据产品统一发布门户。探索总部与各省、各专业公司在统一服务、众包化全网应用开发等方面的协同机制。

(3) 规模推进与优化

①大数据 PaaS 平台规模化

逐步实现统一的开发环境、统一的调度工具以及一些复杂组件服务的接入管理, 按照“成熟一个, 接入一个”的原则, 不断地丰富 PaaS 平台的组件, 不断地完善平台功能, 实现平台整体资源的弹性伸缩能力。除 Hadoop 平台外, 增加更多组件能力, 如 MPP、可视化、数据挖掘等。

②统一架构体系、统一数据治理储备

对现有架构体系进行优化, 逐步建设集中控制、分散部署的大数据平台架构。同时, 对子节点大数据平台的元数据、数据质量等提供统一的数据治理。

(4) 纵向统一

随着技术的不断成熟, 大数据 PaaS 平台逐渐

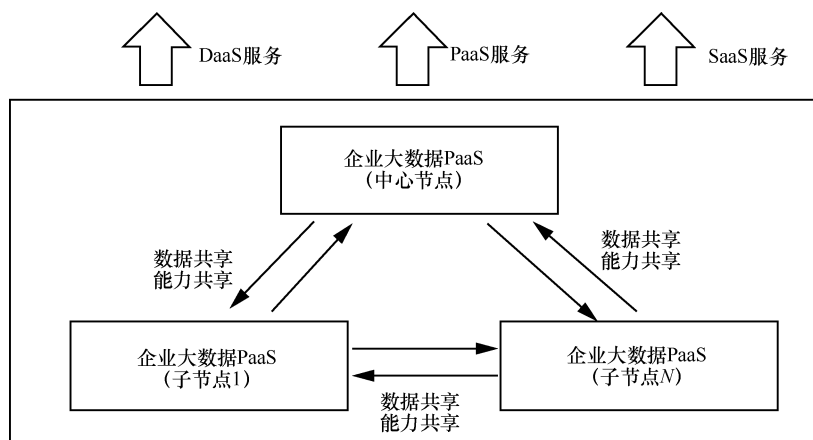


图4 纵向统一的大数据 PaaS 平台

由一点向多点进行演进，实现“逻辑集中、物理分布”的统一大数据 PaaS 平台，实现省级平台到总部平台的纵向汇聚与对接，实现数据的分散处理与调度的统一管理。以一点接入的模式实现对外部合作伙伴提供统一的 DaaS、PaaS、SaaS 服务，如图4所示。

5 结束语

建设大数据开放平台是一个十分艰难的过程，受限于开源技术、企业组织架构、安全策略等因素。随着云计算、大数据技术的不断完善，未来一定会吸引更多的合作伙伴，形成优势互补的大数据生态圈，引领大数据应用向各行业深度发展，百花齐放。

参考文献：

- [1] 刘春, 邹海锋, 向勇. 大数据环境下电信数据服务能力开放研究[J]. 电信科学, 2014, 30(3): 156-161.
LIU C, ZOU H F, XIANG Y, et al. Research on telecom data service open ability under the environment of big data[J]. Telecommunications Science, 2014, 30(3): 156-161.
- [2] 韩晶, 张智江, 王健全, 等. 面向统一运营的电信运营商大数据战略[J]. 电信科学, 2014, 30(11): 154-158.
HAN J, ZHANG Z J, WANG J Q, et al. The unified-operation-oriented big data strategy for telecom operators[J]. Telecommunications Science, 2014, 30(11): 154-158.
- [3] 谷红勋, 杨珂. 基于大数据的移动用户行为分析系统与应用案例[J]. 电信科学, 2016, 32(3): 139-146.
GU H X, YANG K. Mobile user behavior analysis system and

applications based on big data[J]. Telecommunications Science, 2016, 32(3): 139-146.

[作者简介]



张海峰（1979-），男，中国移动通信集团设计院有限公司高级工程师，主要研究方向为 IT 支撑系统、大数据技术及组网。



董昭（1977-），男，中国移动通信集团设计院有限公司网络所副所长、高级工程师，主要研究方向为 IT 支撑系统技术及组网。



李娟（1978-），女，中国移动通信集团设计院有限公司高级工程师，主要研究方向为 IT 支撑系统、大数据技术及组网。

齐磊（1982-），男，中国移动通信集团设计院有限公司高级工程师，主要研究方向为 IT 支撑系统技术及组网。

顾慧琼（1982-），女，中国移动通信集团设计院有限公司高级工程师，主要研究方向为 IT 支撑系统技术及组网。