



基于 PaaS 技术的大数据云化平台实践

徐海勇, 黄岩

(中国移动通信集团有限公司, 北京 100032)

摘要: 大数据平台作为数据存储、处理和服务的基础平台, 需支撑内部和外部多种大数据应用的开发及运行。提出一种把 PaaS (platform as a service, 平台即服务) 技术应用到大数据平台的方案, 详细描述了大数据 PaaS 云化平台的架构体系以及分层调度、多租户管理、Hadoop 任务全局调度等关键技术方案。实践证明, 该平台可以显著提高大数据应用的开发效率, 实现应用快速部署上线; 同时, 其良好的弹性伸缩能力可以有效地提高大数据应用系统资源的使用效率。

关键词: 大数据; PaaS; Hadoop; YARN 调度器; 云服务

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018043

Applications of a big data platform based on PaaS

XU Haiyong, HUANG Yan

China Mobile Communications Groups Co., Ltd., Beijing 100032, China

Abstract: As a platform of data storage, processing, and service, big data platform needs to support internal and external various big data applications. A solution applied PaaS technology to the big data platform was proposed. Details of the key technical solutions such as the architecture of PaaS cloud computing platform, hierarchical scheduling scheme, multi-tenant management and global scheduling of Hadoop tasks were described. It was proved that the platform could remarkably improve big data application development efficiency and help application deployment fast. In the meantime, good elastic scaling capability can effectively improve big data application system resource utilization.

Key words: big data, PaaS, Hadoop, YARN scheduler, cloud service

1 引言

中国移动作为全球规模最大的运营商, 拥有 9 亿户移动用户资源, 其用户使用通信服务的过程中产生的海量数据, 如用户 profile、通信话单、网络信令、上网行为等, 形成了宝贵的大数据资产。利用大数据技术挖掘海量数据背后隐藏的规

律, 可以有效地改善用户业务质量, 优化运营效率, 促进产品创新, 发挥巨大价值。随着数据规模的急速增长, 大数据平台规模的逐渐扩大, 应用需求日趋强烈, 如何协同多家应用开发者在集中、安全、可靠的平台上开发自己的应用, 成为关注的焦点。目前, 面临以下挑战。

首先, 中国移动传统的大数据平台为总部—

省级二级架构,数据呈现分散状态,而全网类应用如金融客户需要信用评估服务、政府部门需要全国人口统计服务等,都需要全网数据的支撑;其次,大数据应用上线时间长且重复开发,对于特定场景的大数据应用,如位置应用、精准营销应用、信用评估应用等,缺乏统一业务开发平台,导致存在各省份公司重复开发、重复运维的情况,浪费了投资,且技术力量分散,难以聚焦形成精品大数据应用;最后,现网中多个烟囱式的数据分析平台存在数据和资源相互隔离、无法共享的问题,当突发数据处理任务时,即使其他平台有大量闲置资源,也无法调配资源,造成业务阻塞。

随着企业IT架构向云化转型,IT资源逐步由共享资源池取代,同时应用的开发迭代周期从年变成周,需要给开发人员和管理人员提供支撑服务,这些需求引发了近年来PaaS(platform as a service)^[1,2]技术的蓬勃发展。谷歌(Google)基于自己的Borg^[3]平台推出开源Kubernetes^[4],为容器化Docker^[5]的应用提供资源调度、部署运行、服务发现、扩容缩容等整套功能,实现大规模CloudNative^[6]应用的集中管理。随着大数据技术应用广泛和日趋成熟,开源软件在资源和任务调度、多租户管理、自动运维等方面已经取得众多成果,如Hadoop^[7]2.0 重构MapReduce框架,以YARN^[8]作为通用的资源调度层,支持MapReduce、Spark、Storm、MPI^[9-11]等更多编程框架,实现了不同类型的数据处理任务共享底层集群资源。

本文提出了一种大数据平台云化解决方案:基于PaaS的理念和技术实现大数据平台云化服务,将大数据平台的数据采集、分析、处理及展现能力作为PaaS能力对外开放。同时,面向大数据业务开发者,提供开发环境和作业执行环境,使开发者可以使用PaaS中各种资源、开发工具、软件中间件等能力,PaaS的弹性伸缩及大数据平

台处理能力的结合,实现资源效率和开发效率最大化的目标。

2 大数据PaaS平台定位与架构

2.1 功能定位

基于PaaS技术的大数据云化平台功能涵盖集中数据平台、应用开发平台、数据服务平台三大功能,具体如下。

- 集中数据平台,汇聚中国移动网络域(O域)、业务支撑域(B域)、管理域(M域)、专业公司及外部引入数据,定义统一的数据模型,形成数据的汇聚及全网服务。
- 应用开发平台,提供数据分析常用的开发工具、软件中间件、任务部署工具等开发环境,业务开发者在平台上在线选择所需的开发环境资源,开发自己的大数据作业和应用。支持多个开发团队并行开发,并支持开发成员之间开发协同。
- 数据服务平台,为业务单位和合作伙伴的数据分析作业、大数据应用提供运行执行环境。从资源池中分配计算、存储和网络等资源,作业执行完毕时,资源重新回收资源池,形成资源共享,租户在大数据云化平台上可以直接对外提供数据应用及服务。

2.2 整体架构

基于PaaS的大数据云化平台包括四层一域,整体架构如图1所示。

- IaaS(infrastructure as a service,基础设施即服务)层,采用OpenStack架构,统一接入VMware、Fusionsphere等主流厂商的Hypervisor虚拟化软件,统一管理CPU、内存、存储、网络等IT基础设施资源。以OpenStack标准化接口,向PaaS层提供IaaS基础资源。
- PaaS层,负责应用和服务生命周期管理、资源调度、多租户管理、软件仓库等PaaS

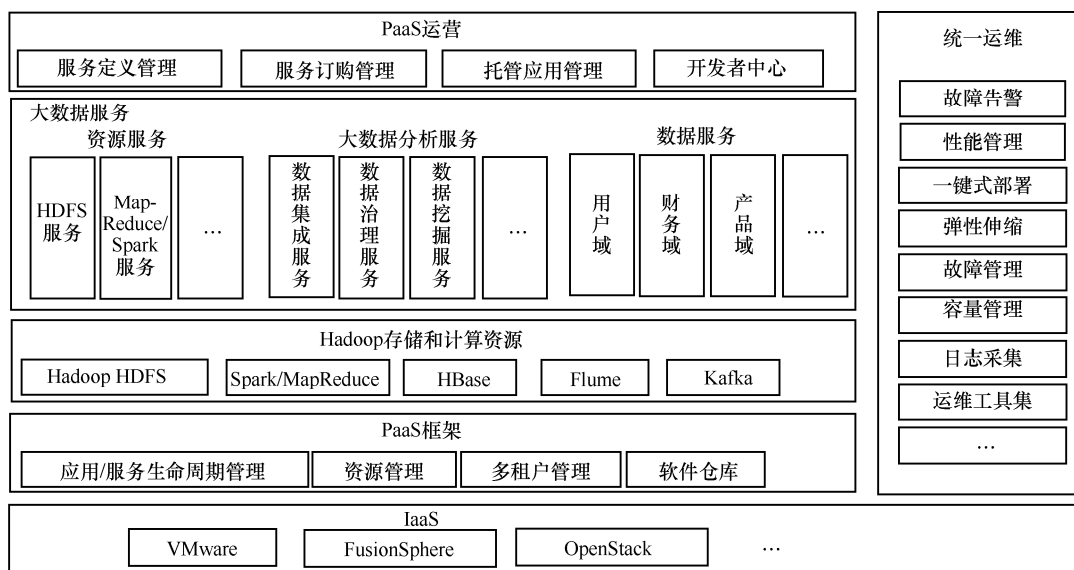


图1 基于 PaaS 的大数据云化平台基础架构

公共能力，为平台开发者调用访问。

- Hadoop 存储和计算资源池层，数据分布式存储和分布式计算的核心执行引擎，以 Hadoop 为核心，包括 Hadoop 协议族中 HDFS、MapReduce、Spark、HBase、Kafka、Flume 等组件。
- 大数据服务层，将基础资源、大数据开发工具、数据资产等封装成服务，供平台开发者使用。同时，对 PaaS 平台管理服务上架、发布和状态监控，对平台开发者提供使用平台服务能力的统一入口。
- 统一运维域，负责整个平台运维，包括应用和服务的部署、弹性伸缩、故障告警监测等，为平台运维人员提供运维工具和管理工具，同时将运维能力开放给平台租户，使租户可以管理自己的应用、服务和资源。

2.3 业务流程

开发者使用大数据 PaaS 平台，主要包括资源申请及应用开发两个主要步骤，具体如下。

(1) 资源申请流程

开发者首先注册成为平台租户，然后创建开发项目；再为项目申请 Hadoop 资源、各种大数据

分析工具服务。所需的资源申请批准后，业务单位即可进入大数据开发。申请流程如图 2 所示。

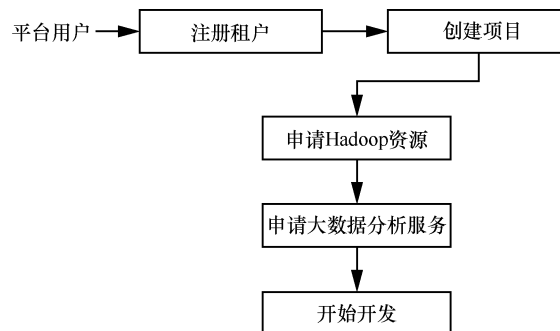


图2 资源申请流程

开发者在平台的服务市场中选择所需的资源和服务界面，如图 3 所示。

(2) 大数据应用开发流程

业务单位开发人员获得所需资源后，进入应用开发阶段。开发者需要经过公共数据资源申请、数据模型定义、其他私有数据上传、数据处理逻辑集成开发及分析、最终形成数据应用。开放数据访问 API 或者形成 SaaS 类应用展示结果等几大基本过程，具体如图 4 所示。

大数据集成开发环境支持图形化开发，如图 5 所示。

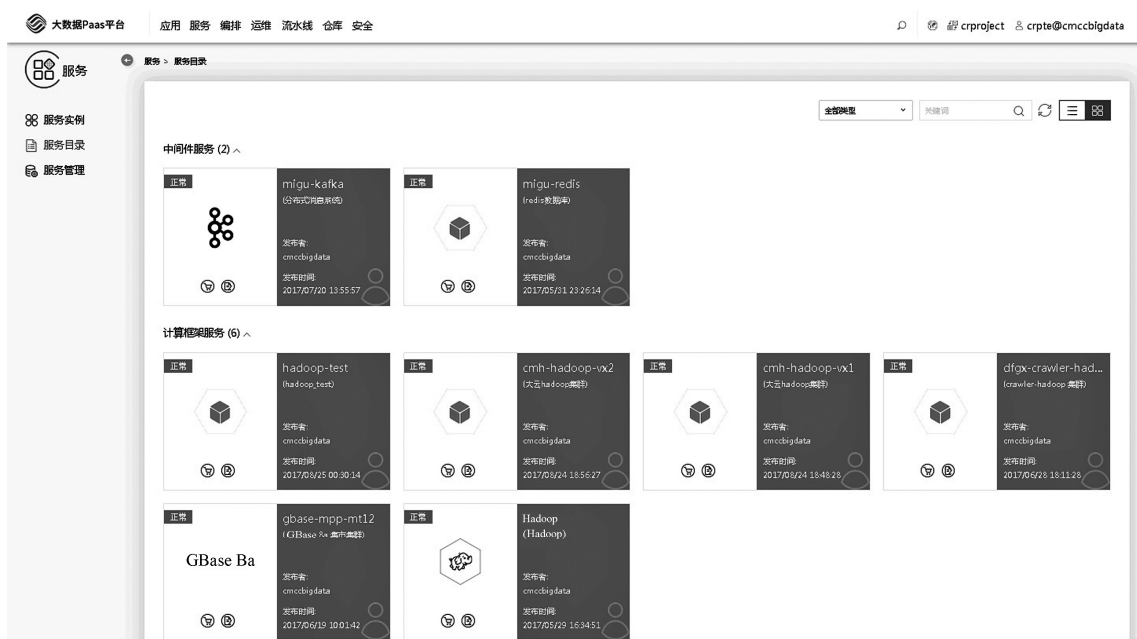


图3 开发者资源申请界面

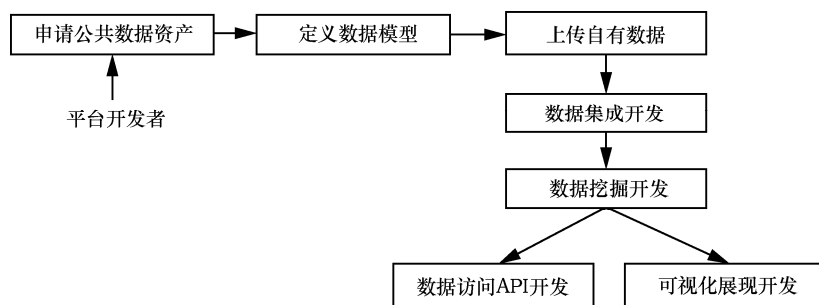


图4 开发者应用开发流程

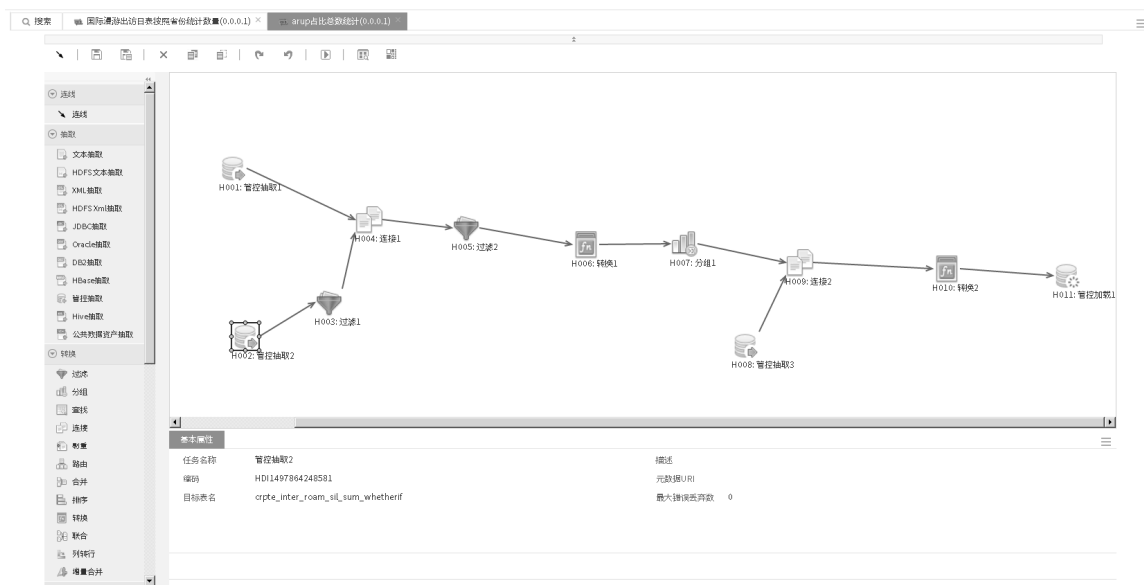


图5 开发者应用开发界面



3 关键技术

基于 PaaS 平台的大数据云化平台,提供了统一的数据存储、开发环境、应用部署对外服务等基本能力,其关键技术包括多应用类型的分层调度、基于 IAM (identity and access management, 统一身份认证服务) 的多租户管理以及全局优化的 YARN 调度技术。

3.1 多应用类型的分层调度

一个完整的大数据业务系统包含多种应用形态:既有框架类的应用,如 MapReduce 作业;也有普通的交易类应用,如营销门户等。如何同时管理和调度这两类的应用,成为研究焦点,本文实现了两种方案。

一种是独立调度,即对不同类型的应用提供不同的资源调度器,资源调度器之间相互独立。

Hadoop 集群使用 YARN 管理调度自己的资源,支持框架类应用;PaaS 另外部署以 Kubernetes 为核心的调度器,管理虚拟机、容器等资源支持普通应用,独立调度技术方案如图 6 所示。

另一种为两层调度方式,即将两种类型应用的资源调度器叠加起来,从而实现资源共享的目的,同时也实现应用的统一管理。需要进一步提升租户间的资源共享程度,就需要采取多层资源调度的方式。比如,以 Kubernetes 作为统一的 1 层资源调度器,以 YARN/Mesos 作为针对大数据应用的两层资源调度器,如图 7 所示。

3.2 基于 IAM 的多租户管理

当大数据云化平台把各种能力以云服务方式对外提供时,存在两种典型的云服务形态——逻辑多租和物理多租。

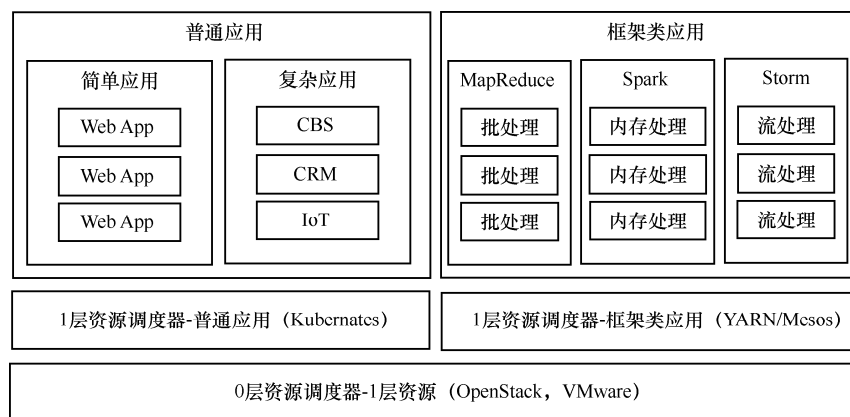


图 6 独立调度技术方案

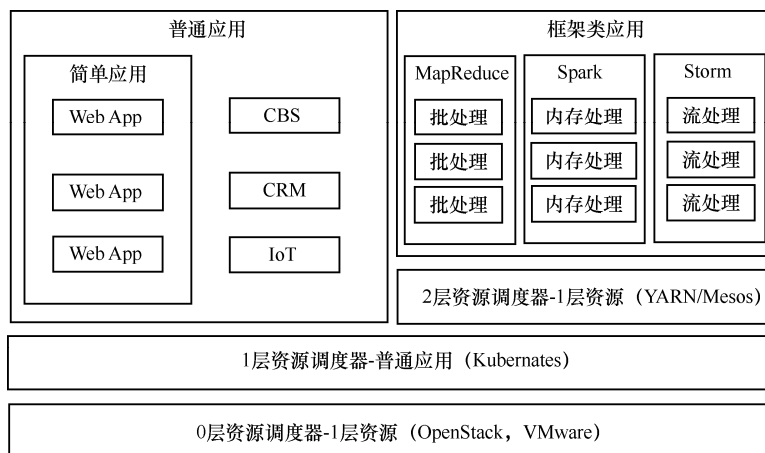


图 7 两层调度技术方案

- 逻辑多租服务：所有租户应用共享同一个服务实例，租户共享服务实例以及实例中的数据、计算能力和存储。
- 物理多租服务：每个租户申请一套独立的服务实例，租户间不共享服务实例。

在逻辑多租的场景下，存在对租户的多层级多域管理的需求。为了能够实现租户可以创建自己独立管理的应用（物理多租场景），又能够在集团租户的空间内创建应用（逻辑多租场景），

参考 OpenStack Keystone^[11]架构实现了 IAM。

基于 IAM 的能力，同一个租户可以在不同的租户空间中创建应用（项目），并对自己创建的应用进行管理，并接受租户空间管理员的统一管理，如图 8 所示。

以上多租户管理方案在对外服务时，解决了租户对公共数据的访问问题，租户可在管理者租户创建的域中创建自己的项目来访问管理者租户中的数据资源，同时租户也可以选择创建自己的

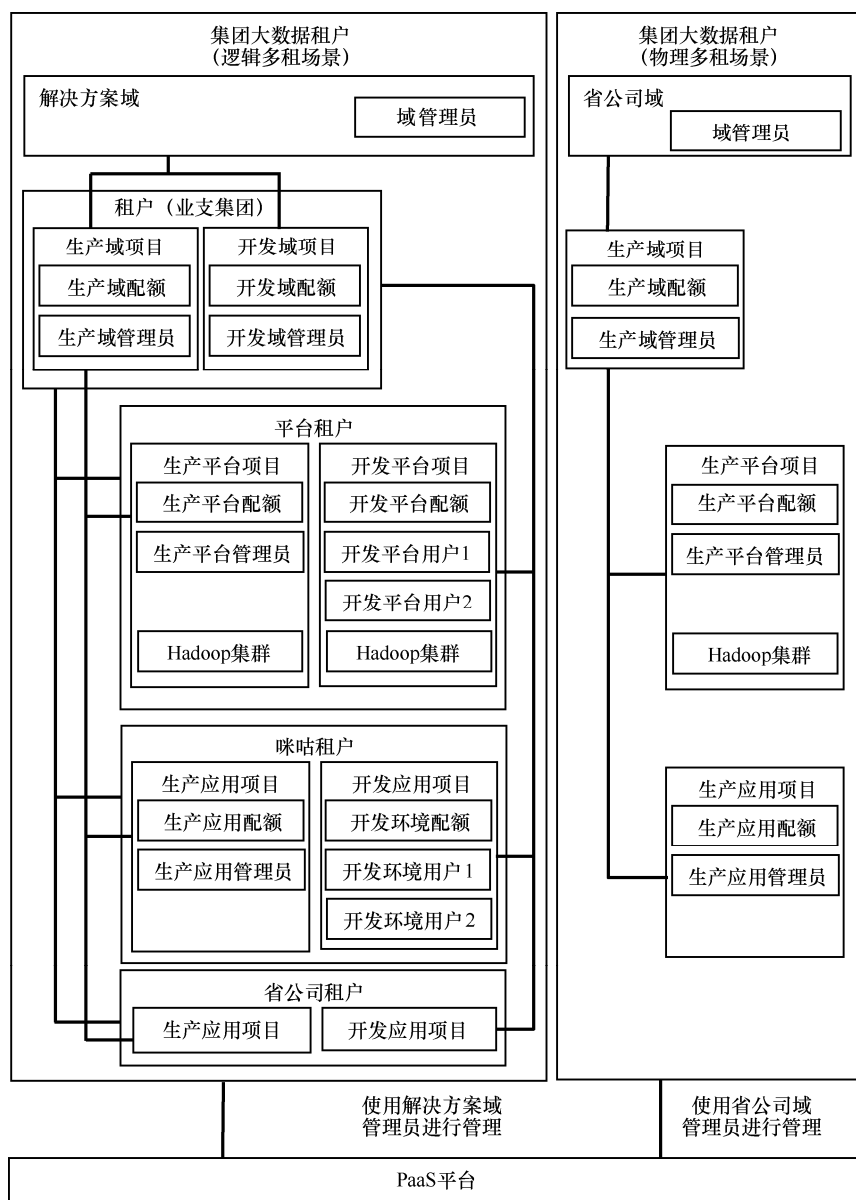


图 8 基于 IAM 的多租户管理模型



域，建立自己的大数据平台。

3.3 全局优化的 YARN 调度技术

开源的 Hadoop YARN 系统中目前包括两大类调度器：公平调度器（fair scheduler）和容量调度器（capacity scheduler），开源的调度器都是基于计算节点心跳驱动的资源反向匹配作业的调度机制，每个计算节点定期发送心跳来通知 YARN 的资源管理器（resource manager）自身的节点状态并同时启动调度器，为这个节点分配作业。这种调度机制把调度的周期同心跳结合在一起，当集群规模增大时，会遇到系统扩展性以及调度性能瓶颈。另外，因为采用了资源反向匹配作业的调度机制，开源调度器在调度精度上也有局限性，例如计算正向数据分配准确性低，系统也无法支持基于负载的调度策略等，其主要原因是调度器在选择作业时，缺乏全局的资源视图，很难做到

最优选择。

本文提出了一种全局的 YARN 调度算法以优化任务调度、提高调度性能、系统资源利用率和大规模集群扩展程度。全局优化的调度器内部采用了与开源不同的调度机制，其调度流程采用了从作业到资源的正向匹配方法，通过转有调度线程，查看所有待处理作业，通过对全局作业和资源的了解匹配合适的任务，这样每个调度的作业都有全局的资源视图，可以很大程度地提高调度的精度，对比流程如图 9 所示。

同时，全局调度器引入了专门的调度线程，把调度同心跳剥离开，避免了在超大规模集群时容易出现系统心跳风暴问题。相比开源调度器，优化的调度器在系统吞吐量、利用率、数据亲和性等方面都有很大提升。

全局调度器和开源调度器功能对比见表 1。

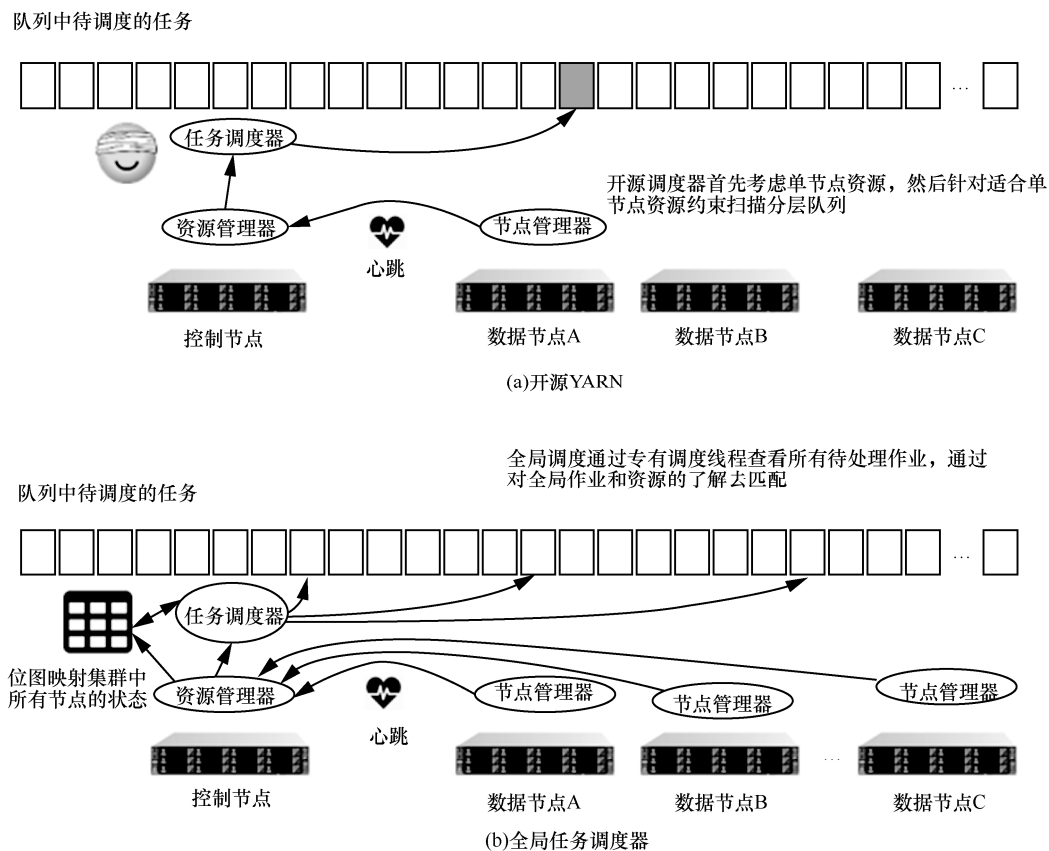


图 9 全局任务调度与开源任务调度流程

表1 开源调度器与全局调度器优势对比

领域	开源调度器	全局调度器
多租户调度	在同构集群上, 只能选择容量调度器或公平调度器两者之一, 且 FusionInsight 集群当前不支持公平调度器。容量调度器只支持百分比方式配置, 而公平调度器只支持绝对值方式	支持异构集群和多资源池; 在单一策略框架中同时支持百分比调度和绝对值方式配置资源需求; 支持预留, 以保证直接访问资源
数据位置感知调度	从节点到作业的调度策略导致降低数据本地命中率, 潜在影响应用的执行性能	从作业到节点的调度策略, 可具有更精确的数据位置感知, 数据本地化调度的作业命中率比较高
基于机器负载的均衡调度	不支持	在调度时考虑机器的负载和资源分配情况, 做到均衡调度
租户内多用户公平调度	不支持	租户内用户的公平调度, 支持关键字 default、others
作业等待原因	不支持	作业等待原因信息可显示为什么作业需等待

4 基于 PaaS 大数据云化平台的应用实践

4.1 平台建设及部署

基于上述的整体架构及关键技术, 中国移动进行了基于 PaaS 的大数据云化平台的建设实践。在数据层面实现了各省 B、O、M 三域数据整合, 构建了中国移动公共的数据资产模型。基础能力层面实现了 Hadoop 服务、公共数据资产服务、数据治理服务、数据集成服务、统一调度服务、数据探索服务、数据 API 开发服务和数据挖掘服务等, 以供省级公司和专业公司申请使用。平台共部署集群规模上百节点, 存储数据量数十 PB, 由租户开发大数据应用, 支撑了应用的快速迭代开发及上线。

4.2 平台应用及效果

经过为期半年的应用实践, 目前基于云化的大数据 PaaS 平台主要承载了信用分应用和畅游业务两大业务应用, 初步验证了大数据 PaaS 平台的能力。

信用分应用通过对移动用户的身份特征、消费能力、人脉关系、信用历史、行为偏好这 5 个维度客观地呈现个人的信用价值。通过信用分应用可以优化移动内部服务, 并扩展外部合作。信用分应用覆盖 11 亿用户和 200 多个脱敏标签数据, 共包含 12 个月的数据周期。该应用仅通过 4 周就在 PaaS 平台上完成了需求分析、数据导入、9 轮建模评估、6 轮联调测试和上线试商用, 上线后, 信用分应用

于 2017 年 8 月底开始试商用, 到同年 10 月底累计访问量已经超过 8 亿次。9-11 月信用分查询次数增长趋势如图 10 所示。

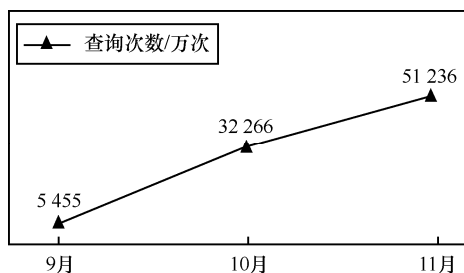


图 10 信用分查询次数增长趋势

在组织的 iPhone 购机活动、“双十一”、开发者大会等市场营销活动中, 基于 PaaS 的大数据平台利用资源的弹性扩展能力, 有效支撑了活动过程中用户访问量的激增。后续将支撑多渠道信用购机、国际漫游信用免押金等多种场景。

畅游业务主要通过分析国际漫游用户的通话和流量, 利用机器学习识别沉默用户, 开展国际漫游精准营销。畅游应用基面向北京、上海等 4 省市 300 万国际漫游移动用户, 基于用户信息、国际漫游用户访问信息、漫游通话交往圈信息等公共数据资产, 租户仅 3 周就完成了业务的需求分析、数据导入、模型开发、界面开发和上线发布。畅游业务实现了国际漫游沉默用户分析等应用, 充分利用营销资源进行沉默客户激活。畅游业务的展示界面如图 11 所示。



图 11 畅游业务系统展示

4.3 应用实践总结

基于 PaaS 的大数据云化平台的应用效果表明,该平台可以有效地为多租户提供统一的全网数据资源;在大数据开发过程中提升了开发效率,缩短了应用的上线周期;弹性扩展能力可以有效地利用资源支撑业务突发。表 2 从硬件资源准备时间、开发工具准备时间、数据准备时间、应用部署时间及资源利用率几大方面,分析了基于 PaaS 的大数据云化平台与传统大数据平台的差异。

表 2 开发模式优势对比

对比项	传统开发模式	大数据云化开发模式
硬件资源准备时间	1~3 个月	小时级别
开发工具准备时间	1~3 周	小时级别
数据准备时间	1~3 个月	小时级别
应用部署时间	小时级别	分钟级别
资源利用率	资源独占, 利用率低	资源池化共享, 资源利用率高

5 结束语

本文提出了一种基于 PaaS 技术的大数据云化平台方案,分析了中国移动大数据应用中的问题和挑战,给出了大数据 PaaS 平台的系统框架。针对租户模型、服务接入、资源调度、弹性伸缩等

关键环节,给出了技术解决方案。基于该平台,实现了多个开发单位的应用开发和应用部署上线,同时业务单位在平台实际开展了开发和应用部署,证明该方案达到了预期效果。

在今后的工作中,在应用设计中结合微服务松耦合、上线敏捷、应用弹性的优势,在 PaaS 能力方面构建深度学习框架及机器学习能力等成为研究的重点。未来中国移动会着力于打造良好的大数据生态圈,吸引更多的合作伙伴基于该平台构建百花齐放的应用。

参考文献:

- [1] Gartner. Platform as a service(PaaS)[EB/OL]. (2014-04-28) [2017-11-28]. <https://www.gartner.com/it-glossary/platform-as-a-service-paas/>.
- [2] 齐磊,张海峰,张天骁,等. 基于容器技术的 PaaS 云平台方案[J]. 电信科学, 2017, 33(4): 177-182.
QI L, ZHANG H F, ZHANG T X, et al. PaaS cloud platform based on container technology[J]. Telecommunications Science, 2017, 33(4): 177-182.
- [3] VERMA A, PEDROSA L, KORUPOLU M, et al. Large-scale cluster management at Google with Borg[C]//10th European Conference on Computer Systems, April 21-24, 2015, Bordeaux, France. New York: ACM Press, 2015: 18.
- [4] 张基,李大中,张呈宇,等. DC/OS 关键技术与应用场景[J]. 电信科学, 2016, 32(12): 13-19.
ZHANG J, LI D Z, ZHANG C Y, et al. Key technologies and application scenario of DC/OS[J]. Telecommunications Science,

- 2016, 32(12): 13-19.
- [5] BERNSTEIN D. Containers and cloud: from LXC to Docker to Kubernetes[J]. IEEE Cloud Computing, 2015, 1(3): 81-84.
- [6] BALALAE A, HEYDARNOORI A, JAMSHIDI P. Microservices Architecture enables DevOps: migration to a cloud-native architecture[J]. IEEE Software, 2016, 33(3): 42-52.
- [7] 王峰, 雷葆华. Hadoop 分布式文件系统的模型分析[J]. 电信科学, 2010, 26(12): 95-99.
WANG F, LEI B H. Modeling and analysis of Hadoop distributed file system[J]. Telecommunications Science, 2010, 26(12): 95-99.
- [8] KONAR M. Apache Hadoop YARN: yet another resource negotiator[C]//ACM Symposium on Cloud Computing, Oct 1-3, 2013, Santa Clara, CA, USA. New York: ACM Press, 2013: 5.
- [9] LU X, LIANG F, WANG B, et al. DataMPI: Extending MPI to Hadoop-like big data computing[C]//International Parallel and Distributed Processing Symposium, May 19-23, 2014, Phoenix, Arizona, USA. Piscataway: IEEE Press, 2014: 829-838.
- [10] 郑文武, 李先绪, 黄执勤. 云计算中的并行计算技术分析[J]. 电信科学, 2012, 27(12): 31-38.
ZHENG W W, LI X X, HUANG Z Q. Technolical analysis of parallel computing in cloud computing[J]. Telecommunications Science, 2012, 27(12): 31-38.

- [11] 李小宁, 李磊, 金连文, 等. 基于 OpenStack 构建私有云计算平台[J]. 电信科学, 2012, 27(9): 1-8.

LI X N, LI L, JIN L W, et al. Constructing a private cloud computing platform based on OpenStack[J]. Telecommunications Science, 2012, 27(9): 1-8.

[作者简介]



徐海勇 (1970-), 男, 中国移动通信集团有限公司业务支撑系统部总经理、高级工程师, 主要研究方向为移动通信、互联网、大数据。



黄岩 (1976-), 男, 中国移动通信集团有限公司业务支撑系统部信息管理处经理、高级工程师, 主要研究方向为大数据、能力开放。