

# 从点状应用到大数据统一平台

钱岭,孙少陵,石在辉

中移(苏州)软件技术有限公司,江苏 苏州 215163

## 摘要

企业在大数据引入过程中,由于平台技术不成熟、缺乏统一技术架构、组织架构不适应等问题,形成了应用竖井和复杂的系统混搭结构,造成了资源利用率不均、数据重复存储、运维成本高、难以形成企业级数据共享等问题。目前大数据平台技术已经逐步成熟,大数据统一平台逐步成为企业分析系统的核心。深入分析了企业级统一大数据平台的整合思路,从管理、业务和技术3个层面入手,分别从组织架构体系建设、大数据平台运营流程重构与实施、大数据平台技术规划与实现等角度提出对策建议。最后,给出中国移动在统一大数据平台整合方面的实践经验。

## 关键词

大数据统一平台;BI+;Hadoop;集中化经营分析系统

中图分类号:TP311.1

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2017032

## *Towards unified big data platform*

**QIAN Ling, SUN Shaoling, SHI Zaihui**

China Mobile (Suzhou) Software Technology Co., Ltd., Suzhou 215163, China

## *Abstract*

Due to problems like immaturity of platform technology, lack of unified technical framework and inadaptation of organization structure, silo applications and complicated mashup structures were formed when enterprises were introducing big data technology. With maturity of big data platform technology, unified big data platform was gradually becoming the core of enterprise analysis system. The thoughts of enterprise-level big data platform integration from three aspects, management, operation and technology were analyzed. Suggestions about organization structure development, reconstruction and implementation of big data platform operation process, technology planning and realization of platform were provided. At last, practical experiences of unified big data platform integration of China Mobile were presented.

## *Key words*

unified big data platform, BI+, Hadoop, centralized BI system

## 1 引言

“数据是新的石油”已是全世界的共识,我国也陆续出台了《促进大数据发展行动纲要》等政策法规,并在“十三五”规划中将“国家大数据战略”列入十四大重点战略。国内数据生产、交易、消费环节初步形成,各行各业都开始把大数据作为产业转型升级的有力武器。受益于多方环境利好,大数据市场增长潜力巨大。同时,因为诸多原因,企业在实际开展大数据建设时也面临很多问题,比如数据条块分割、重复存储和计算、运维成本居高不下、扩容压力大、大量点状应用系统等典型问题<sup>[1]</sup>。

## 2 大数据平台整合问题

大数据最先是由于企业商业智能(business intelligence, BI)系统升级换代而出现的<sup>[2]</sup>。一些数据分析压力大的先发企业在BI建设过程中,率先在生产环境引入大数据技术,如Hadoop、MapReduce、HBase等<sup>[3-5]</sup>,形成BI+系统。但受技术成熟度、市场环境变化、企业战略调整、部分业务应用的紧迫性等因素影响,几乎都会出现成长的烦恼,比如:系统为了满足业务需求,在缺乏规划的情况下建设大量点状应用竖井式建设;随着技术的发展,形成了数据仓库、MPP(massively parallel processor)、Hadoop等纷繁复杂的技术体系;很难按照预定的IT规划路线图发展,往往需要在实践中优化和重构。

在大中型企业中,多个数据竖井支撑点状应用现象比较普遍。以中国移动通信集团公司(以下简称中国移动)为例,同时

存在商用数据仓库、Hadoop、MPP或者多种数据架构混搭方式,支撑不同的业务应用,如图1所示。

随着大数据应用的不断深化,点状应用的建设模式已经无法满足发展要求,具体表现为以下几点。

- 数据交换难:数据分散在竖井中,难以形成多维度关联,数据价值降低。
- 建设成本高:数据重复存储和处理,资源无法共享和统一调度。
- 管理开销大:每个集群均需要单独的维护人员,如果厂商/架构不同,复杂度更高。
- 安全风险高:数据分散管理,增加了数据泄露的风险。

企业在引入大数据的同时均会遇到上述问题。如何建设统一的企业级大数据平台,对资源、服务、数据和用户进行统一管理,提供统一的应用开发和运行环境,是所有大数据应用企业需要解决的共性问题。

## 3 统一大数据平台整合思路

对于大中型企业来说,大数据平台整合是一项系统工程,需要统筹考虑管理、业务和技术等诸多方面的要求,有针对性地进行组织管理调整、业务流程设计和平

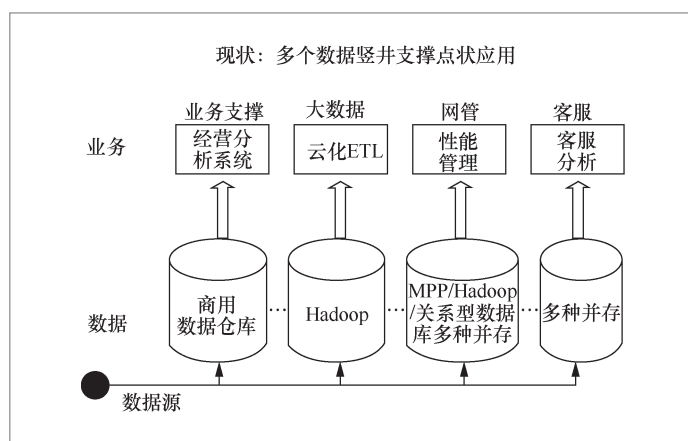


图1 多个数据竖井支撑点状应用示例

台技术实现<sup>[6]</sup>。

### 3.1 大数据平台整合考虑要素

**管理层面：**建立管理体系，确定管理模式，包括对原体系的变更；管理数据获取，建立利于大数据获取的组织架构和流程；管理控制数据质量，建立对各个域数据质量的统一管理体系；管理数据变现，依据法律法规，制定统一的数据变现规则；管理数据安全，针对各种条例开展审计工作，保障业务稳定运营。

**业务层面：**建立大数据运营体系；定义大数据运营的商务模式，明确服务对象、盈利模式等；定义对内、对外的产品形态，并根据业务形态定义产品；定义运营优化体系和办法，通过持续优化，提高运营效益和利润率。

**技术层面：**建立大数据技术体系，明确研发模式、技术来源、技术架构等内容；明确定义大数据平台的技术体系；明确符合各种数据处理、业务需求的技术架构和产品；针对技术架构，为各种技术选择研发模式，并开展研发；定义应用迁移方式，确定现有业务大数据改造途径。

### 3.2 大数据平台整合思路

(1) 建设与大数据平台整合运营相适应的组织架构和体系

构建企业级大数据能力需要成立大数据管理控制（以下简称管控）、大数据运营和大数据技术机构。其中，管控机构负责大数据战略和整体规划；运营机构负责大数据业务规划、数据管理、建设和业务运营等工作；技术机构负责大数据技术研发和运营支撑工作。相关机构协同工作，共同建立面向大数据运营的体系。

大数据管控机构是公司级别的管理单

位，需明确整个公司层面的大数据战略，建立大数据组织架构调整，推进大数据跨部门采集和应用、实施统筹、服务保障和考核等体系，为大数据运营保驾护航。

大数据运营机构是公司级别的生产单位，需承担大数据业务规划，建设企业级大数据业务体系，包括建设大数据系统、从各个领域采集大数据、统一治理企业级数据模型、在大数据系统上开展各种业务等，为公司内外部提供服务。

大数据技术机构是公司级别的研发和支撑单位，需承担大数据系统技术方案和系统集成工作，包括支撑运营机构完成技术和实施方案、技术研发和采购、系统集成工作，在系统上线后，还需支撑数据分析、系统技术服务和运维支撑工作。

**整合实施要点：**在整合前，各部门遵从自己的内部流程；整合后，需要统一遵循公司的统一流程和体系（包括数据获取、应用构建、应用运营和维护等），可以基于统一大数据平台快速开展，运营过程需要经由管控和运营机构审核与保障。

(2) 大数据平台运营流程重构与实施  
构建企业大数据中心之后，需求单位可以按照云计算方式申请大数据中心的计算能力和数据资源，在正常的申请范围内，可以快速实现业务开通和上线。在大数据中心管辖范围之外的数据，还需要大数据管控单位的审核开通。

传统IT运营流程模式向大数据中心运营流程模式演进过程如图2所示。

**整合实施要点：**整合前，各部门遵从工程项目建设流程，需要经历预算审批、招投标、集成、数据开通等复杂的流程，时间周期长，不确定性高；整合后，可以按照技术服务方式构建大数据业务系统，周期短、效率高、更安全。

(3) 大数据平台技术规划与实现  
企业级大数据平台包括管理、业务、

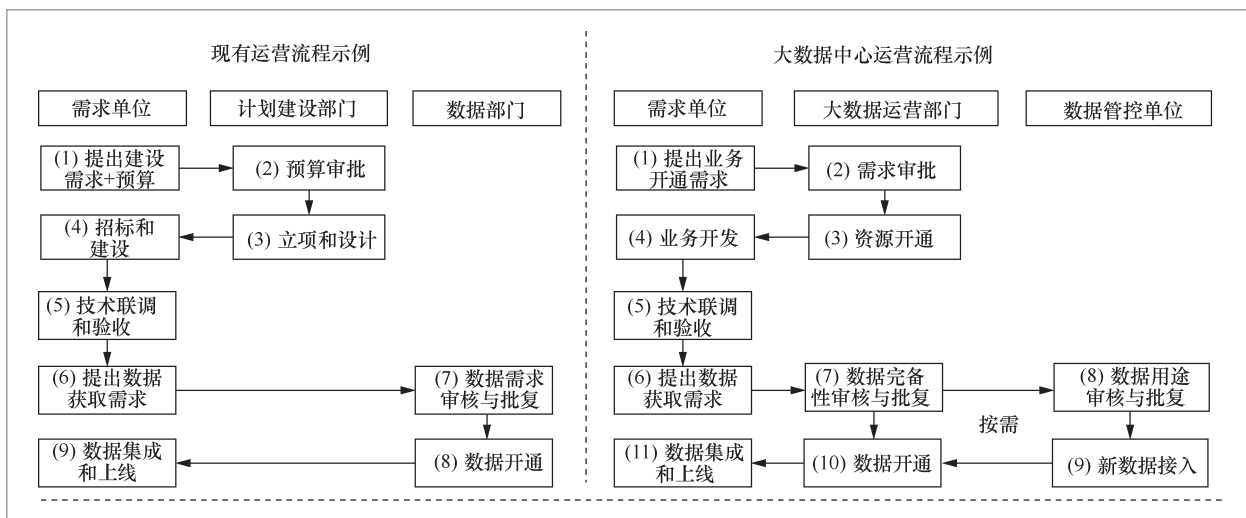


图2 大数据平台运营流程示例

数据、开发等所有相关功能，需要具备全局的管理思维、详尽的功能视图设计以及整合复杂的技术体系。大数据平台技术规划与实现可参照开放组体系结构框架（the open group architecture framework, TOGAF）等业界通用的IT架构设计方法论。如图3所示，以中国移动为例，大数据功能视图包括多源数据、大数据平台和应用、不同类型的使用者。对应到技术架构，包括数据资产融合、大数据技术合理选型、大数据平台构建、大数据服务模式等技术主题。

整合实施要点：整合后，需要开展现有应用迁移，一般涉及3种模式，即将应用的计算部分迁移到新系统；将应用的计算和原有数据都迁移到新系统；新建与平台相符合的应用。企业应根据实际情况，有序有目标地进行迁移。

## 4 大数据平台整合实践

中国移动集团公司包括31家省公司和多家专业公司，IT系统已各自独立建设运营十几年，数据体量巨大，种类繁多，位置

分散，应用多样。各单位共有超过1 000套数据仓库、MPP和Hadoop等大数据系统，由超过百家集成商承建，导致资源分散、版本众多、数据重复存储、平台资源重复投资、运营不能统一等问题，专业IT人员无法复用。为应对当前公司面临的大数据挑战，支撑公司万物互联战略和数字化服务商转型的发展，中国移动统筹制定公司大数据发展战略和规划，逐步整合现有应用，加快大数据基础能力和开放平台建设，实现全网数据资源的统一管理和运营，支持对内、对外多种应用。

### 4.1 大数据平台整体思路

(1) 先立后破，加快统一平台建设

中国移动集团总部推进集中化经分大数据平台建设，汇聚全网数据。初期支撑总部、各专业公司和部分省公司的大数据内外部应用需求，长期目标是构建“逻辑集中、物理分散”的1+N超大规模集群大数据平台。

各省公司同期开展大数据平台省级系统建设，加快原有烟囱系统的迁移。为中国移动集团统一大数据平台奠定资源基础，实现统一资源管理、数据管理和开放平台

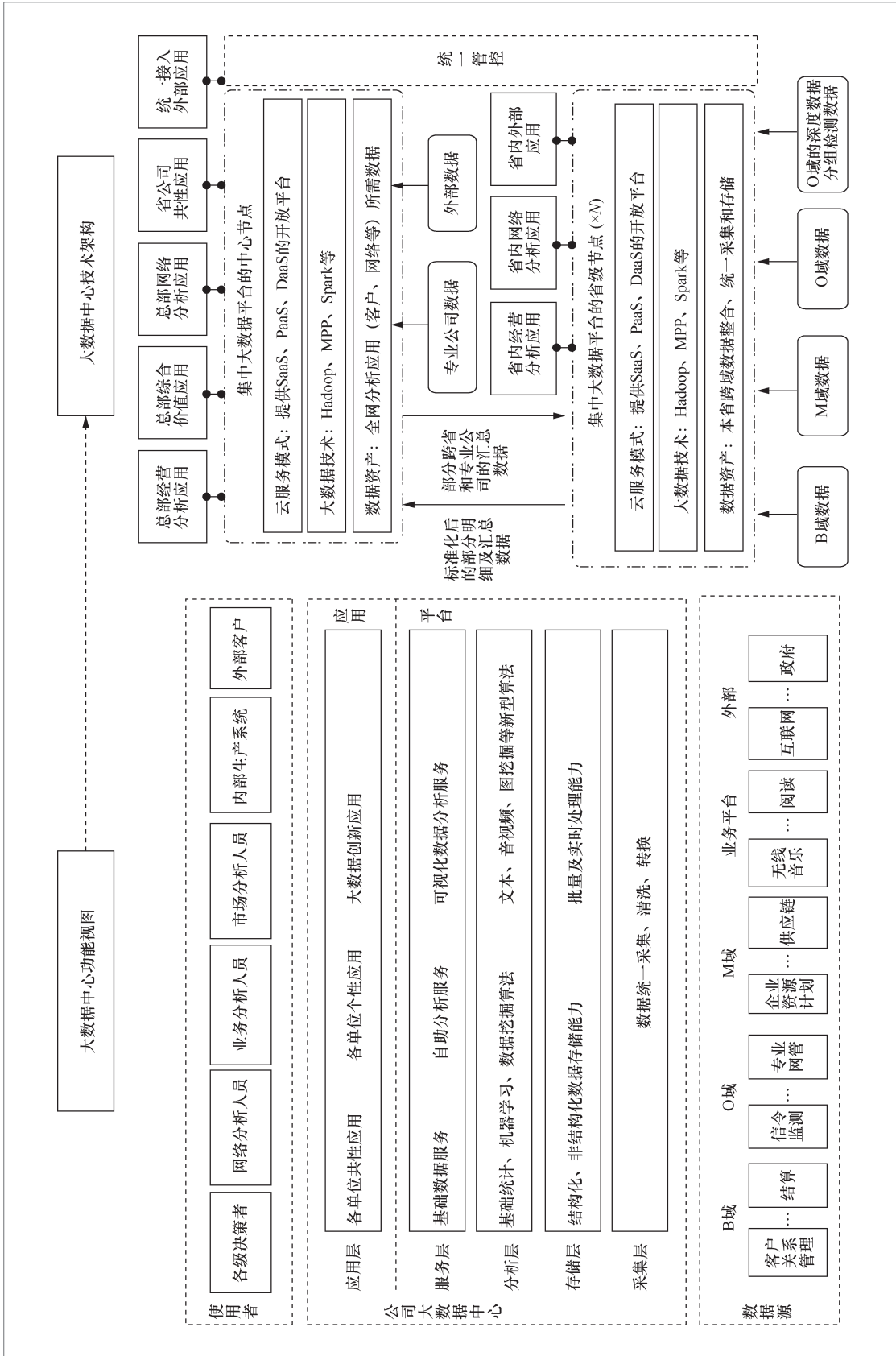


图 3 中国移动大数据中心功能视图和技术架构

能力,支撑多部门大数据应用建设。

### (2) 加快数据融合

中国移动集团统一制定数据治理框架和规则,各省公司加快内部数据聚合,完成B域、O域和M域数据的统一采集和存储,推进跨域数据融合和跨域数据治理<sup>[7]</sup>。

### (3) 探索多种服务模式

积极探索软件即服务 (software as a service, SaaS)、平台即服务 (platform as a service, PaaS)、数据即服务 (data as a service, DaaS) 等对内对外服务模式。持续深耕内部大数据价值挖掘,提高大数据对经营决策、数字化服务创新、渠道集中化运营、网络运维和优化、市场营销和服

务、精细化管理等各方面工作的智能化支撑水平。同时,鼓励各省联合创新,推动优秀应用跨省快速迁移。积极开展对外服务,以中国移动为核心,与合作伙伴一起打造生态圈,服务政府、金融、医疗、工业等多行业多领域。

### (4) 构建安全保障体系

为应对大数据应用服务过程中数据滥用和个人隐私安全风险,中国移动建立了完善的大数据安全保障体系,目标是保护大数据权属性、保密性、完整性、可用性、可追溯性,实现大数据“可管、可控、可信”,保护公司各领域大数据资产及用户隐私。大数据安全保障体系框架如图4所示。

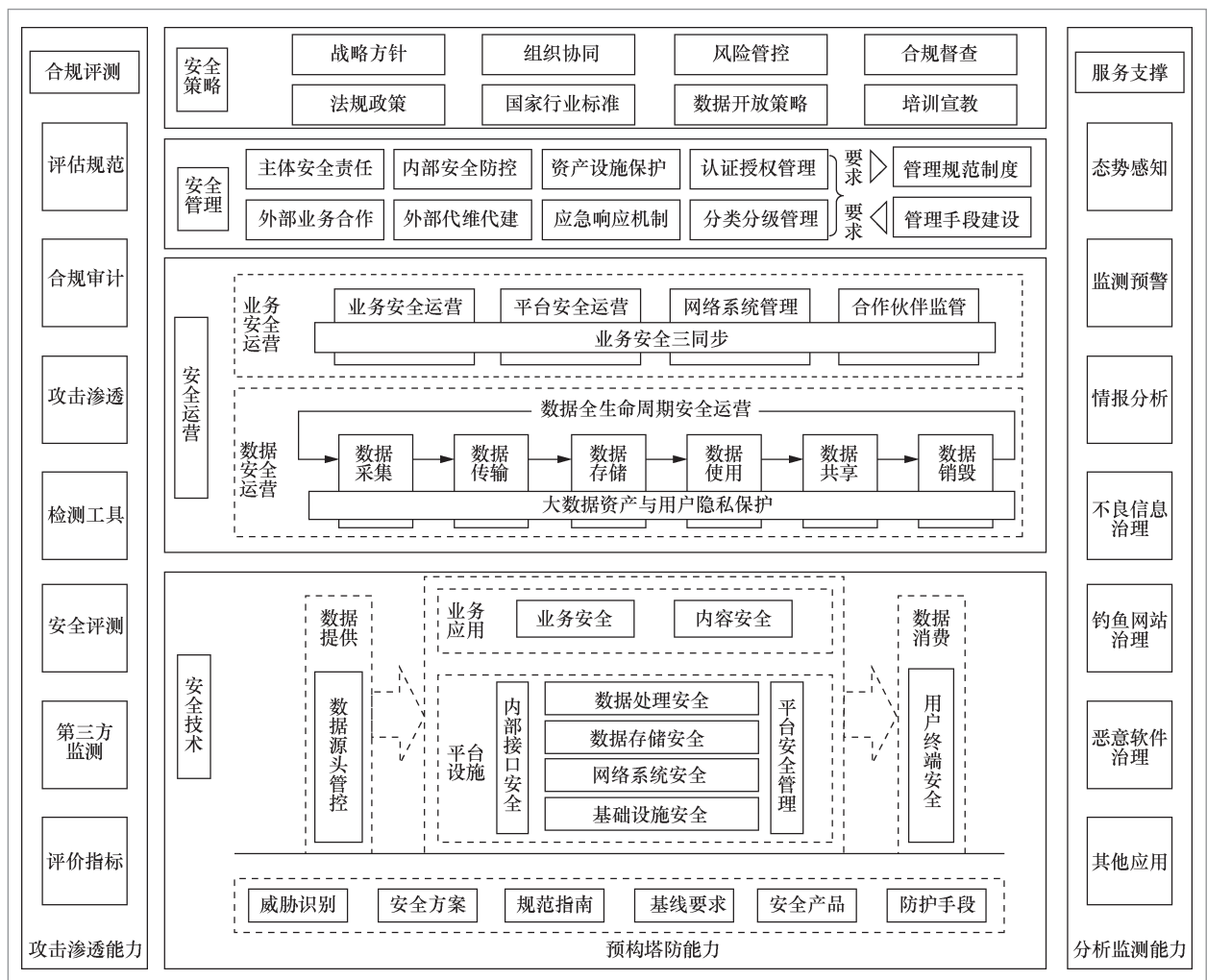


图4 中国移动大数据安全保障体系框架

中国移动大数据安全保障体系涉及安全策略、安全管理、安全运营、安全技术、合规评测、服务支撑六大体系。同时,对用户个人信息的各个处理环节施行严格规定与落实,具体措施举例如下:

- 对客户信息包含的内容进行界定、分类及分级;
- 明确信息安全管理责任部门及职责,对各部门的职责进行严格要求和细致规定,明确相关岗位角色及权限;
- 对客户敏感信息操作进行严格管理,对于涉及用户敏感信息的关键操作,严格遵守金库模式保护要求,采取“关键操作、多人完成、分权制衡”的原则,实现操作与授权分离;
- 设立客户信息安全检查制度;
- 不断提高客户信息系统技术管控水平;
- 严控第三方信息安全风险。

#### 4.2 大数据平台架构设计

如图5所示,中国移动大数据平台采用“3域4层”的实施架构,构建复杂的多租户系统。各个模块之间充分解耦,共同支撑上层百花齐放的应用。

3域包括管理域(提供统一管理能力,包括资源、租户、维护、数据、安全、

应用等)、服务域(提供统一调度能力,包括计算、存储、I/O、开放数据能力等)、开发测试域(提供统一开发能力,包括工具、需求、流程、测试和上线等)。4层包括采集层、计算和存储层、能力和接口层、应用层。

中国移动各省公司在大数据引入早期构建了大量的大数据点状应用,如ETL(数据仓库技术,用来描述将数据从来源端经过抽取、转换、加载至目的端的过程)、数据挖掘、账详单查询、营销类应用,随着业务增长,均造成数据膨胀、性能下降、利用率不均等问题,需要扩容;同时点状应用数量增加,大幅度增加了管理成本,需要整合。为解决点状应用整合管理的问题,中国移动设计了大数据运营管理中心(big data operations management center, BDOC)平台产品,整合大数据点状系统和统一运营运维管理,如图6所示。

大数据平台总体架构设计的考虑因素如下:

- 为满足数据量爆炸式增长和数据安全备份需要,大数据中心应是一中心、多集群的架构;
- 需建设统一的大数据管理中心,实现统一管理、统一调度、统一开发、多租户管理;

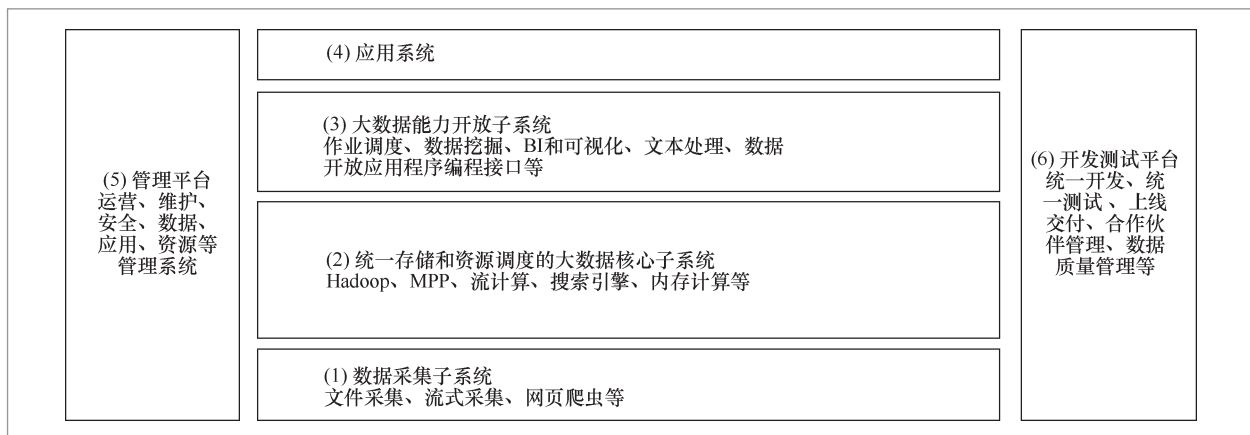


图5 “3域4层”实施架构

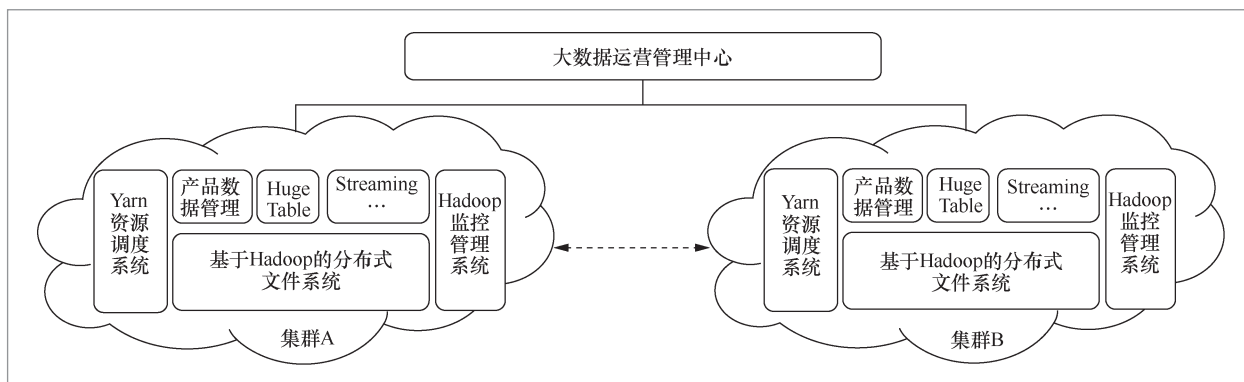


图6 统一大数据平台运营管理

- 对多个大数据集群（包括异地异构）的各种资源进行管理及统一调度；

- 为大数据管理人员、资源运营人员、数据使用者、应用开发者、运维人员等提供统一门户。

### 4.3 自主大数据核心能力构建

2014年3月，中国移动成立中移（苏州）软件技术有限公司（又称中国移动苏州研发中心，以下简称“苏研”），负责中国移动统一大数据平台的建设工作，对内外部客户提供大数据领域的咨询、规划、方案、建设、集成、研发、运维等全方位服务。

作为中国移动大数据核心技术的研发机构，苏研在大数据平台技术方面基于Hadoop、Spark等开源软件构建，采用开放技术架构，坚持核心模块自主研发，兼容业界主流大数据商用产品。在应用开发方面，已经与产业界广泛合作，无缝对接中国移动内部应用需求，快速交付现网可用产品，并实现既有应用迁移，在市场营销、企业体验、网络优化和IT系统优化等多个领域，苏研都有成熟的产品和解决方案支撑。

苏研目前已有24项大数据产品<sup>[8-12]</sup>，目前已在27个单位落地，累计部署近2 000个节点。其中10项产品和3项服务（集成实施、数据开发、算法和分析）进入集团大

数据核心能力清单。苏研大数据产品体系架构如图7所示。2016年苏研进入中国移动集团的大数据核心能力清单包括：大云大数据运营管理中心BC-BDOC以及资源整合服务（含集成服务和数据开发服务）；大云大数据平台BC-Hadoop（含Spark等开源软件）；大云数据库/数据仓库系列，HugeTable多引擎大数据仓库版本、MPP版本和RDB关系型数据库版本（含联机分析处理（on-line analytical processing, OLAP）、联机事务处理（on-line transaction processing, OLTP）数据库）；大云流计算产品BC-Streaming；企业级搜索引擎BC-SE；大云大数据挖掘产品BC-PDM/BC-ETL以及算法建模服务（含算法和分析服务）；舆情监控产品。

## 5 结束语

中国移动作为大型传统行业的企业代表，对大数据有旺盛的需求。传统的全盘外包的IT建设模式难以适应大数据的发展需求，中国移动采用了统一规划、集中运营、核心能力自主研发、强化安全管控的发展思路。在公司级平台建成前，允许省公司和专业机构各自发挥能动性，按照统一规范建设省级平台，支撑生产，这

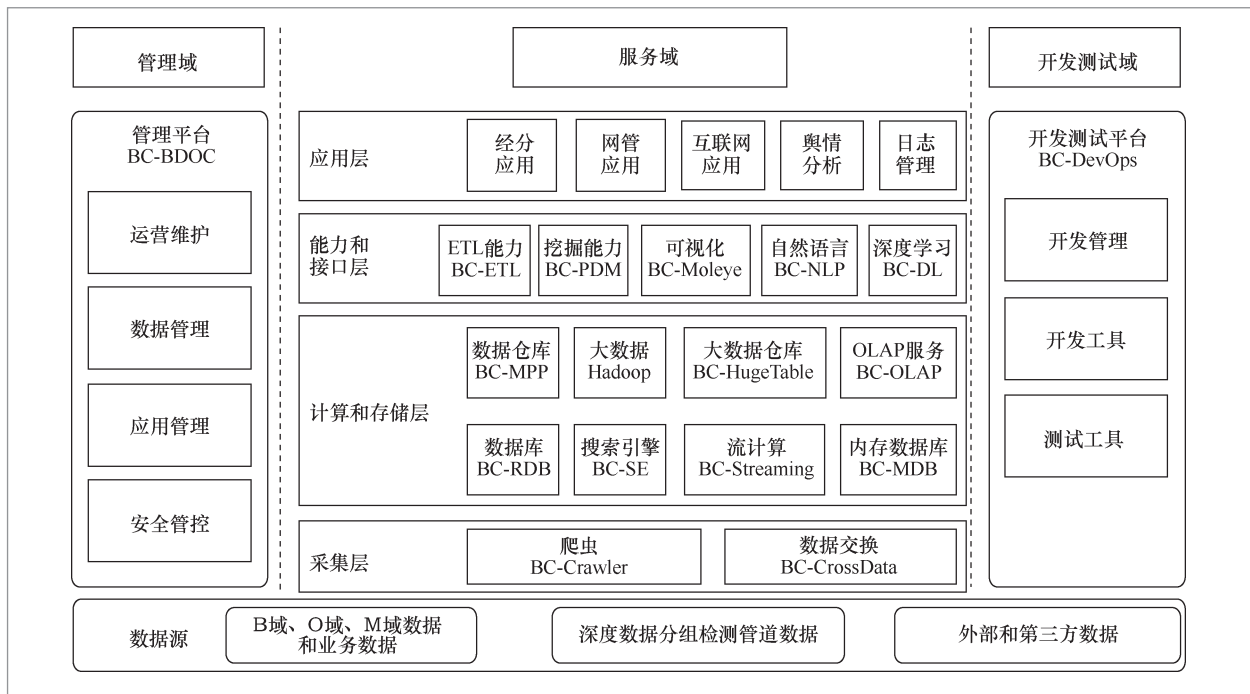


图7 苏研大数据产品体系架构

种发展途径目前已经取得了阶段性成果。传统企业可以借鉴中国移动的大数据发展模式，有序发展，真正发挥大数据价值。

## 参考文献:

- [1] 李国杰. 对大数据的再认识[J]. 大数据, 2015, 1(1): 2015001.  
LI G J. Further understanding of big data[J]. Big Data Research, 2015, 1(1): 2015001.
- [2] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908.  
CHENG X Q, JIN X L, WANG Y Z, et al. Survey on big data system and analytic technology[J]. Journal of Software, 2014, 25(9): 1889-1908.
- [3] GHEMAWAT S, GOBIOFF H, LEUNG S T. The Google file system[C]//The 19th ACM Symposium on Operating Systems Principles, October 19-22, 2003, New York, USA. New York: ACM Press, 2003: 1-15.
- [4] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[C]// The 6th Conference on Symposium on Operating Systems Design & Implementation, December 6-8, 2004, San Francisco, CA, USA. New York: ACM Press, 2004: 10.
- [5] CHANG F, DEAN J, GHEMAWAT S, et al. Bigtable: a distributed storage system for structured data[J]. ACM Transactions on Computer Systems, 2008, 26(2): 15.
- [6] 孟祥飞, 冯景华, 赵洋, 等. 应用驱动的大数据融合平台建设[J]. 大数据, 2017, 3(2): 67-77.  
MENG X F, FENG J H, ZHAO Y, et al. Application-oriented integration platform construction on big data[J]. Big Data Research, 2017, 3(2): 67-77.
- [7] 马朝辉, 聂瑞华, 谭昊翔, 等. 大数据治理的数据模式与安全[J]. 大数据, 2016, 2(3): 83-95.  
MA C H, NIE R H, TAN H X, et al. Research on data schema and security in data governance[J]. Big Data Research, 2016, 2(3): 83-95.
- [8] 孙少陵, 周大, 钱岭. 云数据仓库高性能查询

- 技术研究[J]. 邮电设计技术, 2011(10): 23-26.
- SUN S L, ZHOU D, QIAN L. High performance query technique of cloud data warehouse[J]. Designing Techniques of Posts and Telecommunications, 2011(10): 23-26.
- [9] 周大, 钱岭, 郭磊涛, 等. HugeTable: 一种面向电信行业的云数据仓库[C]// 2011中国计算机学会服务计算学术会议, 济南, 中国, 2011-08-18. 济南:[出版者不详], 2011: 1-4.
- ZHOU D, QIAN L, GUO L T, et al. HugeTable: telecom oriented data warehouse[C]// 2011CCF National Conference on Service Computing, August 18, 2011, Jinan, China. Jinan: [s.n.], 2011: 1-4.
- [10] BAO Y, WANG Z, BAI Q, et al. BC-BSP: a BSP-based system with disk cache for large-scale graph processing[J]. Open Cirrus Summit, 2012, 7215(1): 35-39.
- [11] GUO L T, SUN H W, LUO Z G. A data distribution aware task scheduling strategy for mapreduce system[C]// The 1st International Conference on Cloud Computing, December 1-4, 2009, Beijing, China. Berlin: Springer Press, 2009: 694-699.
- [12] YU L, DUAN S Q, SHEN C W, et al. BC-PDM: data mining, social network analysis and text mining system based on cloud computing[C]// The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 12-16, 2012, Beijing, China. New York: ACM Press, 2012: 1496-1499.

## 作者简介



**钱岭** (1972-), 男, 博士, 中移(苏州)软件技术有限公司大数据部高级工程师、总经理, 专注大数据产品规划、研发和应用全流程、软件过程改进等, 曾获得中国通信学会科学技术奖二等奖, 出版《敏捷开发知识体系》《云计算: 深刻改变未来》《功能点分析——成功软件项目的测量实践》和《自适应软件开发: 一种管理复杂系统的协作模式》4本著作, 发表多篇论文。



**孙少陵** (1972-), 男, 中移(苏州)软件技术有限公司高级工程师、副总经理, 主持研发“大云”云计算/大数据产品, 曾获得国家科技进步奖二等奖及多项省部级奖励。



**石在辉** (1983-), 男, 中移(苏州)软件技术有限公司大数据产品部方案架构师, 主要从事解决方案、战略规划、产品规划等工作。

收稿日期: 2017-03-24