

评语：该方案解决了某银行数据增长之后的性能问题。技术选型非常稳健，对于解决此类问题具有典型的参考意义。

大数据平台在金融行业的典型应用

罗圣美, 戚晨, 王敏, 叶郁文

中兴通讯股份有限公司, 江苏 南京 210012

摘要

金融行业需要大数据技术支持数字化转型,但大数据技术体系复杂、组件快速变化、产品有效运维等问题越来越成为金融行业数字化转型的瓶颈。阐述了金融行业数据处理中面临的问题,结合银行业务发展需求,分析了基于开源技术的商用大数据平台GoldenData HD的系统架构、技术难点和主要特性。通过具体应用案例,展示了GoldenData HD的应用效果,该平台有效提升了数据存储容量、查询处理速度和应用业务创新,为金融企业的数字化转型提供了有力支撑。

关键词

大数据平台;数字化转型;金融行业

中图分类号:TP391

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2018023

Typical application in financial industry of big data platform

LUO Shengmei, QI Chen, WANG Min, YE Yuwen

ZTE Corporation, Nanjing 210012, China

Abstract

Financial industry increasingly needs big data technology to support digital transformation, but the complex architecture, rapid change, weak management of big datasytem are becoming the bottleneck. The financial industry problems of data processing, and analysis system architecture, technical difficulties and main characteristics of the commercial big data platform Golden Data HD which based on the open source technology, with the development of banking business needs, were expounded. Using the specific application cases, the application effect of the Golden Data HD was shown. The platform can effectively enhance the data storage, query processing and application of business innovation, and provide strong support for the digital transformation of the financial industry in the era of big data.

Key words

big data platform, digital transformation, financial industry

1 引言

移动互联网时代,传统金融企业面临多方面的压力。利率市场化,同业竞争加剧,导致利润空间下降、IT成本压缩;传统银行IT架构采购和维护成本过高,具有更低成本优势的IT进入了金融企业的视野;同时,IOE(IBM, Oracle, EMC)对金融业形成事实垄断,产品闭源不可控,IT系统依赖性过大,失去议价能力,安全风险越来越大。种种问题迫使金融企业更加关注新技术的发展,需要尽快引入新技术,完善移动渠道的访问能力,增加服务渗透力,加快新业务部署上线速度。

2 金融行业面临的问题

相比中小型金融企业,银行具有庞大的客户数量,其数字化转型难度更为巨大。挑战主要包括数据种类越来越多样,数据量日趋庞大,无论是数据存储还是数据查询,在软硬件等方面都遇到了瓶颈。用户的应用和分析结果呈整合趋势,对实时性和指令响应时间的要求越来越高;同时,数据处理的模型也越来越复杂,算法的复杂性相应提高,这些都需要从数据采集、数据管理、数据处理(包括数据传输)等方面改进优化^[1]。下面以中国某银行为例,分析其在数据处理和应用中存在的问题,具体如下。

(1)传统工具数据存储空间成为瓶颈

随着业务的日益发展,金融企业出现了大量的冷数据、低价值数据和历史数据。这些数据已经超出了传统数据存储工具软件的管理上限,同时还消耗了价格昂

贵的服务器、数据库的有效存储空间。随着数据增长速度的加快,此问题日益成为金融业务改造和拓展的巨大障碍。

(2)传统工具的数据加工效率日益低下

在信用风险管理、客户关系管理、财务分析、合规管理、运行监控、数据仓库等领域,每天都需要进行TB、10 TB甚至100 TB级数据的加工,传统存储过程加工周期越来越长,已经无法满足应用的要求。

(3)应用系统的客户体验越来越差

数据量的剧增、加工效率的下降导致金融应用系统的客户体验越来越差。此类情况在很多金融应用中存在,部分金融应用不得不将实时查询方式改造为离线查询方式,进而导致客户体验进一步下降。

3 大数据平台在金融行业应用中的瓶颈

最近几年,大数据的商业化应用逐步落地^[2]。通过合理运用大数据技术,对掌控的数据资产进行加工、处理、整合、利用和反馈等,不断挖掘数据资产的价值,为创新金融的内外监管、提高服务效率、支撑产品优化、新产品开发、提供高质量服务,提供了全新的思维和技术支撑^[3]。通过批量处理、实时数据流分析和各类即时查询等大数据技术,对丰富的数据资源进行开发和利用,是金融创新的主流选择^[4]。但在目前常见的Cloudera公司的大数据平台发行版(CDH)、Intel公司的大数据平台发行版(IDH)等平台产品使用过程中存在较多的客户化服务问题,具体如下。

- 新技术、组件的研究、集成、开发、运维成本非常高,周期长。在没有有效技

术支持的情况下，只能一点点地摸索新组件的使用，不但增加了研发周期，同时还需要较长时间的跟踪、积累，才能将其投入关键领域。部分大数据组件成熟度较低，导致应用风险更大、周期更长。

- 应用二次改造、升级、缺陷弥补难度大。大数据技术发展速度非常快，组件的版本不断更新。每个新版本的引入都需要全面的测试、验证，才能投入研发。而应用的二次改造、升级难度就更大。

- 开源大数据组件运维、监控复杂，故障恢复、数据灾备等功能不完善，无法进行集中管理。

- 开源大数据平台缺少有效的培训服务和技术支持。技术人员缺少大数据开发技能学习环境，掌握大数据组件开发技术周期长，产品运维过程中解决问题效率低下，严重影响应用研发周期的有效控制。

综上所述，缺少跟随开源路线的高质量大数据平台化产品和技术支持，在规划架构、技术支持、故障恢复、生产运维等方面会导致应用产品开发成本难以控制，投产应用产品稳定性欠佳，产品运维严重脱节。

4 大数据平台GoldenData HD

中兴大数据平台GoldenData HD是基于开源技术的商用大数据平台，是笔者所在单位近10年大数据项目实践中的技术沉淀。该平台涵盖了数据的采集、存储计算、分析挖掘、应用建模、可视化展现以及运维管理等能力，其架构如图1所示。

大数据平台GoldenData HD具有如下主要特性。

- 成熟稳定：基于多年的技术和经验

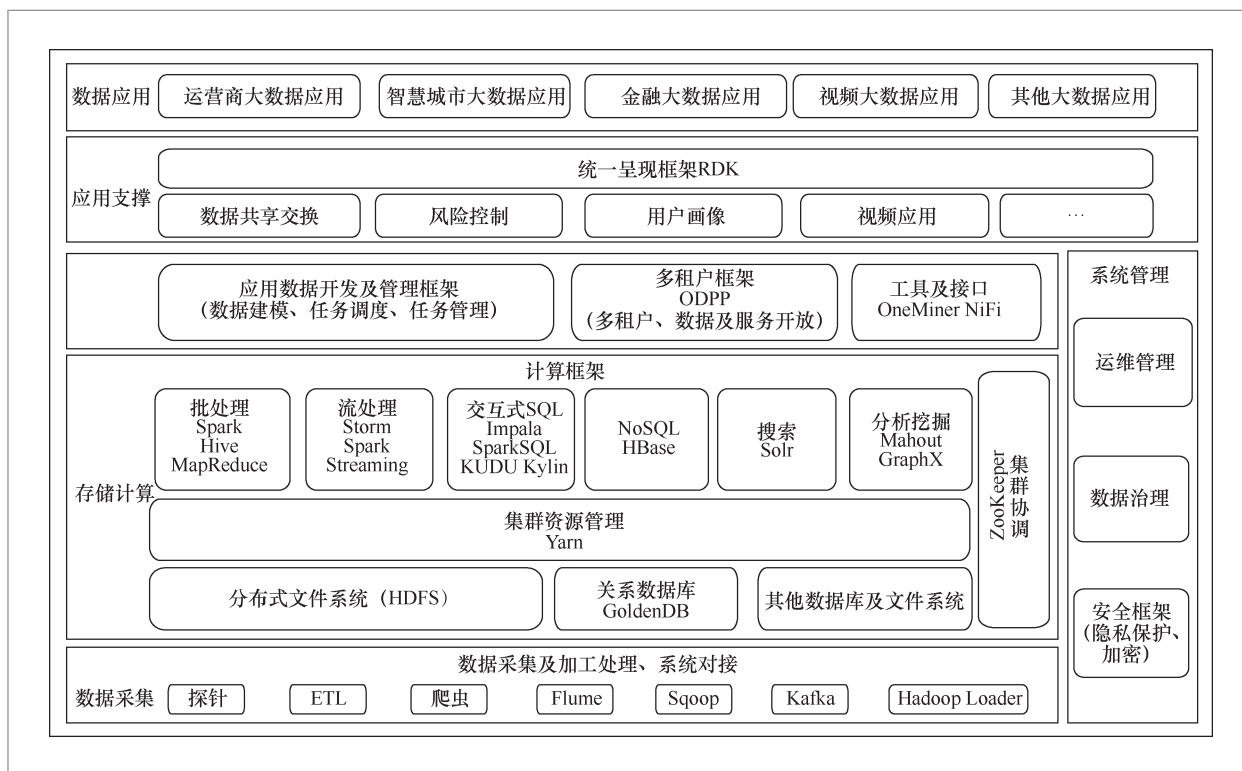


图1 GoldenData HD 架构

积累,平台的功能、性能、可靠性和稳定性都经过了极为苛刻的市场验证,能够满足金融行业高可靠性的要求。

- **开放包容:** 集成超过20个组件,其中约80%为独立的开源软件。在融合的同时,进行了大量的创新和优化,使之更具商用价值,GoldenData HD组件功能优化结果如图2所示。

- **统一运维:** 统一大数据管理系统,一站式满足系统快速部署、集中管理及可靠运维需要。

- **安全可靠:** 精心设计的安全架构,构建高枕无忧的应用访问环境。

- **访问便捷:** 自主研发的开放数据处理平台(open data processing platform, ODP),在大数据基础平台之上架设统一访问层。

- **开发速度快:** 平台集成了大数据挖掘分析相关的各种基础算法库、开发工

具、可视化工具,可以快速实现应用开发,降低开发难度。

5 应用案例及效果

目前GoldenData HD已经在某银行的多个研发部门应用,逐步替换CDH、IDH等已有产品。通过全面测试,平台功能的完备性、性能的可靠性得到了某银行的全面认可。目前构建在GoldenData HD大数据平台上的文件管理、信用风险、身份鉴别、柜面业务改造、对公客户关系管理、合规审计、个人客户关系管理等业务都已经投产上线,管理会计、对公客户关系、风险控制等多个业务领域也正在研发中。截至2017年年底,在生产、开发、测试等环境中,部署GoldenData HD集群节点近400个,不但解决了某银行在大数据技

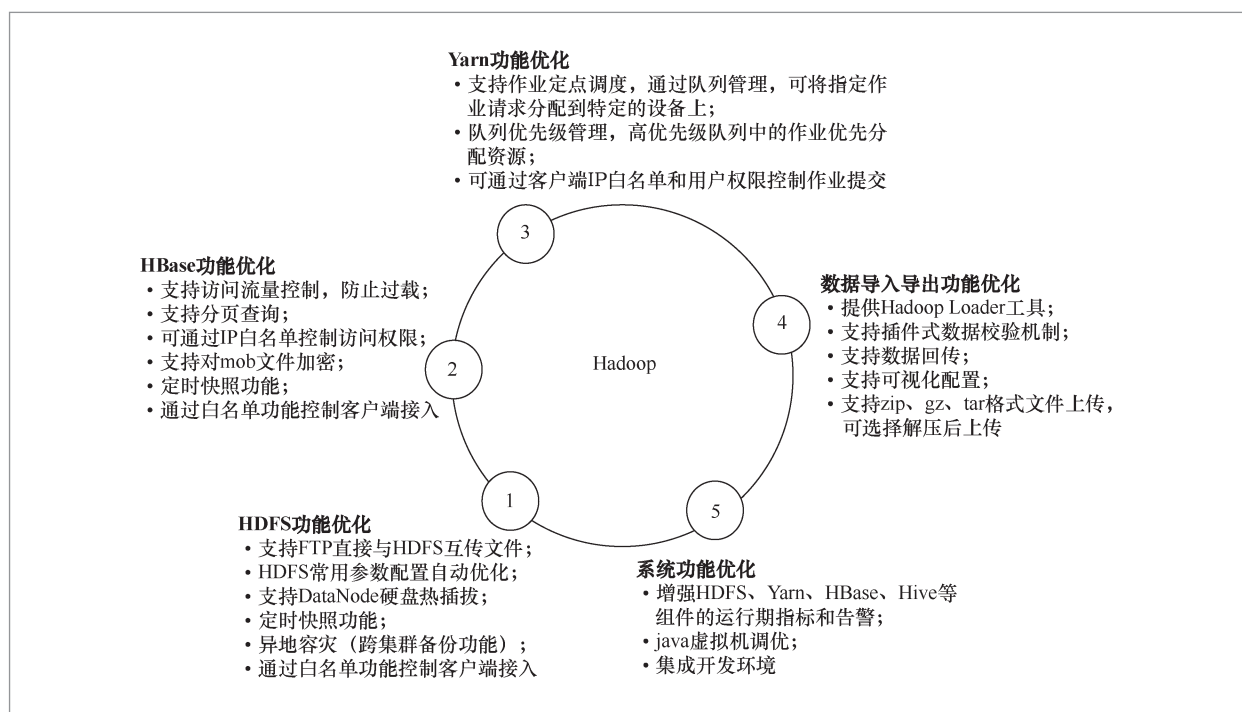


图2 GoldenData HD 组件功能优化结果

术应用过程中的难题,也有效提升了业务开发的速度和质量,满足了金融业务的多样性运行和测试需求。

下面简单介绍基于GoldenData HD的3个典型应用案例及效果。

(1) 文件管理系统——业务支柱产品

文件管理系统是最先移植到大数据平台上的应用业务。此系统在某银行信息化架构中,用来管理、存储各应用系统产生的各种类型文件,如图片、报表、照片等。系统部署在某银行上海数据中心80个节点的集群中,使用了Hadoop分布式文件系统(Hadoop distributed file system, HDFS)、Yarn、ZooKeeper、HBase、Hive、Solr等组件,为全银行提供文档高效存储查询服务。基于GoldenData HD的文件管理系统架构如图3所示。

其设计思路如下:

- 通过服务接口接收各应用系统发送的小文件;
- 接收的文件在GoldenData HD中的Solr组件中建立索引;
- 若是小文件,存储在GoldenData HD中的HBase组件,并根据业务特征建立RowKey;
- 若是大文件,则直接存储在GoldenData HD中的HDFS上;
- 应用系统发送查询请求后,通过Solr快速定位文件存储信息;
- 根据文件存储信息分别从HBase、HDFS中提取文件提交至应用系统。

系统投产至今,GoldenData HD除支撑系统架构和投产外,还支持了两次重大改造:一是2017年5月份,实现了在线Solr扩容,将索引数据承载能力提升了一倍;二是2017年年底,实现了系统从千兆网集群到万兆网集群的平滑升级迁移,将文件数据处理效率提升了3倍。在两次改造过程中,GoldenData HD提供的迁移、Solr扩

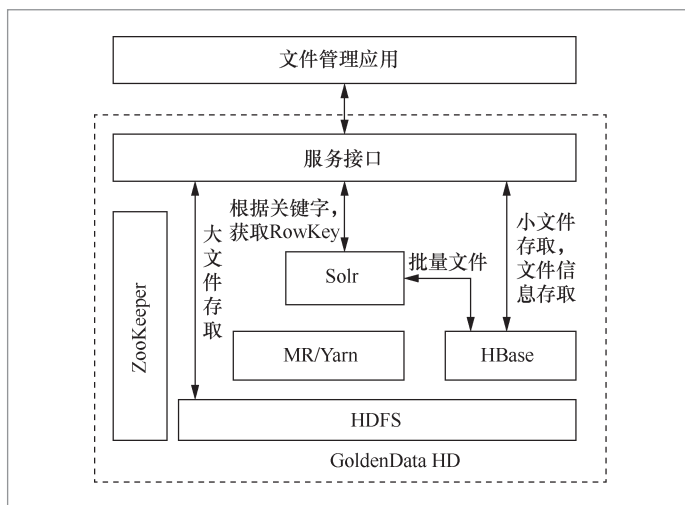


图3 基于 GoldenData HD 的文件管理系统

容、数据恢复等支撑工具,有效缩短了扩容、迁移过程周期,并且保障了改造过程的顺利完成。

通过扩容和迁移,有效提升了系统的负载能力。通过2年多的推广应用,文件管理平台从辅助性产品提升为业务支柱性产品,目前已经支撑5个省份的网点柜面身份鉴别业务改造、2个省份的网点柜面业务无纸化改造。截至2017年年底,使用的存储容量已经超过90 TB,预计2018年年底使用容量将超过2 PB。

(2) 信用风险管理系统——交易图谱构建

信用风险管理系统是银行业大数据应用的核心内容,原有系统受限于传统架构,难以处理多达30亿条的法人客户交易数据,仅月度数据查询一次就超过20 min,无法满足实际业务需要。

2016年年初,在系统规划阶段,笔者团队技术专家和某银行技术专家一起,首先分析了业务需求、数据特征,之后依据查询并发能力、查询效率等指标的对比,选定了HBase作为查询服务的载体组件。但HBase对多条件组合查询的支持能力存

在先天的严重不足,导致项目无法推进。

笔者团队技术专家在进行深入的研究后,提出了组合实现方案,包括业务查询方式的微调、合理的HBase的数据行键值(RowKey)设计方法、基于业务数据特征的HBase数据分区(region)划分规则及定制开发的HBase协处理器功能,将复杂的组合查询条件降维后和HBase的RowKey进行匹配,在集群中各数据分区服务(region server)功能上进行分布式并行过滤与查询计算,实现了查询过程的降维处理要求,将每次查询的数据范围控制在50 GB以内,有效提升了数据查询效率。

系统采用了HDFS、Yarn、HBase、WAS(Web应用系统)、外部数据库组成的混合架构,实现了查询反馈效率的提升,使升级后的系统达到了预定性能目标。基于GoldenData HD的信用风险系统如图4所示。

系统投产后,对比原系统,基于月度数据的单次查询效率由超过20 min下降到不

足1 s,受到了银行内信用风险部门业务人员的称赞。

(3) 基于大数据的多维分析平台

某银行现有的多维分析应用是基于Cognos实现的。由于Cognos元数据服务承载能力低,难以承担全银行查询并发压力,故而不得不采取了总行/分行部署架构,此种架构造成了运维难度巨大的问题。同时,随着数据量的快速增长,数据方体(cube)的加工周期快速增长,部分cube的加工周期达到以天为单位。无法有效承载某银行新业务模式赋予的重任。2016年年初,某银行立项了新一代多维分析平台项目,以彻底取代Cognos。

在研究过程中,笔者团队技术专家与某银行专家一起,选择了GoldenData HD的Kylin为核心组件,并进行了深入的功能和性能研究。通过实际业务数据测试,针对大数据量的cube, Kylin的构建速度比Cognos提升近10倍,并且证明了Kylin的高并发负载能力、标准SQL支持能力、

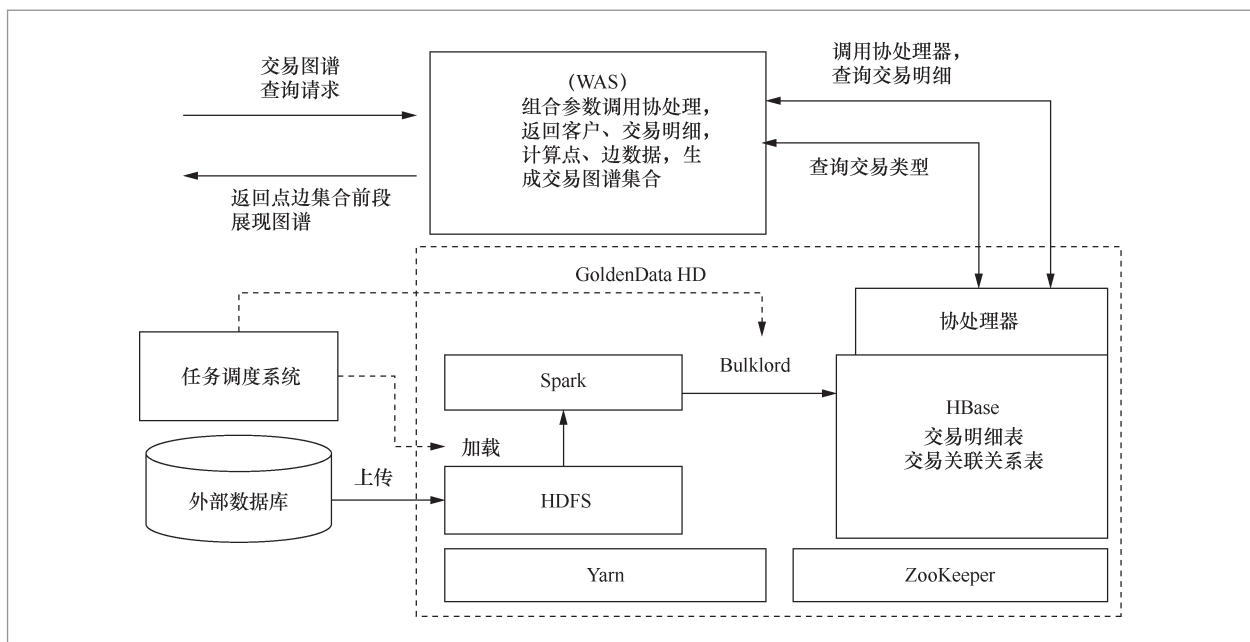


图4 基于 GoldenData HD 的信用风险系统

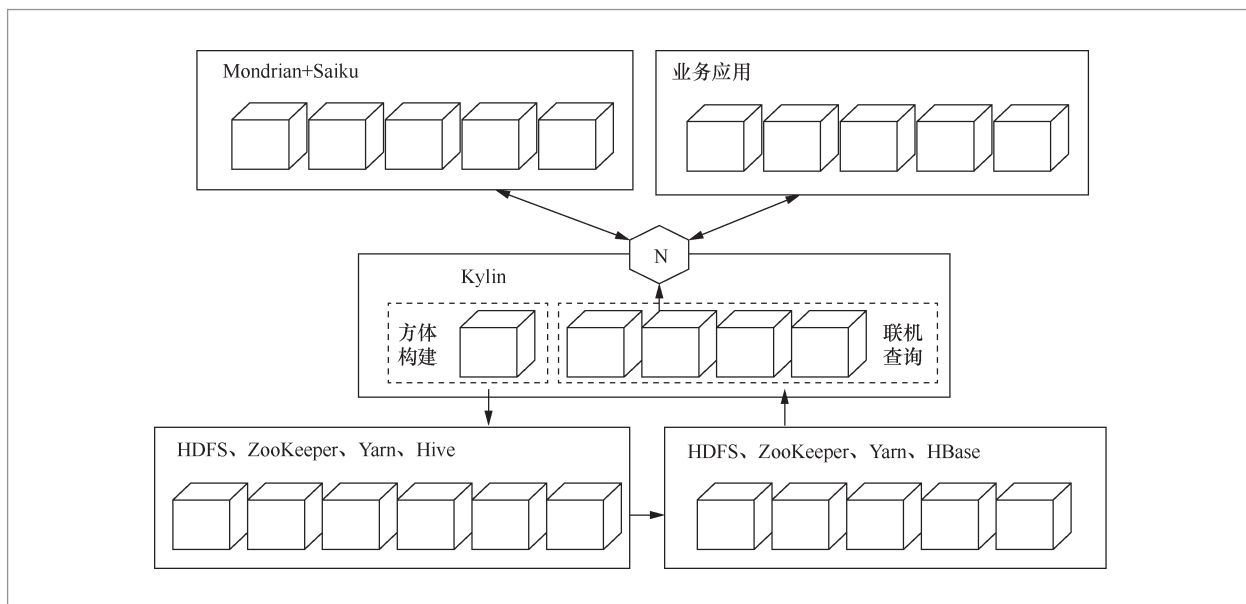


图 5 基于 GoldenData HD 的多维分析系统

GoldenData HD 集群的快速横向扩展能力，充分满足了某银行当前以及将来的多维分析业务需求。

2017年7月，以GoldenData HD为基础，具备支撑百亿级数据能力的多维分析平台顺利投产，以云服务模式满足各领域商业智能分析需求。基于GoldenData HD的多维分析系统如图5所示。

投产后不久，某银行管理会计系统的cube就在此平台上设计并构建，为管理会计系统提供了新一代多维数据服务。后续在各业务系统升级改造过程中，新一代多维分析平台将逐步取代原有的Cognos，实现全银行集中的大数据多维分析服务平台。

6 结束语

基于对金融行业数字化转型中面临的实际问题的研究，某银行联合笔者所在单位进行大数据技术应用分析，采用GoldenData HD平台进行生产系统改造，构建了多个开

放的数据存储、查询和分析系统，有效提升了海量数据存储和业务应用查询的速度，为某银行在大数据时代的数字化转型和跨越式发展提供了有力的支撑。

参考文献:

- [1] 刘伟娜, 谷晓飞, 监文慧. 大数据应用研究[J]. 合作经济与科技, 2017(9): 26-27.
LIU W N, GU X F, JIAN W H. Research on large data application[J]. Co-operative Economy & Science, 2017(9): 26-27.
- [2] SCHOENBERG V M, CUKIER K. 大数据时代: 生活、工作、思维的大变革[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013.
SCHOENBERG V M, CUKIER K. Big data: a revolution that will transform how we live, work, and think[M]. Translated by SHENG Y Y, ZHOU T. Hangzhou: Zhejiang People's Publishing House, 2013.
- [3] 侯敬文, 程功勋. 大数据时代我国金融数据的服务创新[J]. 财经科学, 2015(10): 26-35.
HOU J W, CHENG G X. Service innovation of China's financial data in big data

time[J]. Finance & Economics, 2015(10): 26-35.

- [4] 王文, 李明, 戚晨. 浅析金融大数据平台的架构与建设[J]. 金融电子化, 2015(12): 86-87.

WANG W, LI M, QI C. Analysis of the structure and construction of large financial data platform[J]. Financial Computerizing, 2015(12): 86-87.

作者简介



罗圣美 (1971-), 男, 博士, 中兴通讯股份有限公司首席架构师, 中国计算机学会大数据专家委员会委员, 主要研究方向为云计算和大数据。



戚晨 (1973-), 男, 中兴通讯股份有限公司云计算与大数据产品总经理, 主要负责产品经营和市场规划。



王敏 (1975-), 男, 中兴通讯股份有限公司云计算与政企业务产品高级架构师, 主要研究方向为大数据在金融行业的应用。



叶郁文 (1974-), 男, 中兴通讯股份有限公司云计算及政企产品规划部部长, 主要研究方向为云计算和大数据在政企领域的行业应用。

收稿日期: 2018-02-09