

基于大数据云分析平台的企业经营分析系统建设^①

汶拴劳

(上海宝信软件股份有限公司, 上海 201203)

通讯作者: 汶拴劳, E-mail: 609076377@qq.com

摘要: 针对现有经营分析系统的海量数据存储、海量数据处理及企业业务发展需求, 我们设计并构建了基于大数据云分析平台的企业经营分析系统. 本文重点介绍了该系统的建设技术路线、总体框架图、系统实现及运行效果. 结果证明, 该系统能够较好的提高数据综合利用效率, 为企业运营提供数据支撑.

关键词: 经营分析系统; 大数据云分析平台; 海量数据处理; 企业运营

引用格式: 汶拴劳. 基于大数据云分析平台的企业经营分析系统建设. 计算机系统应用, 2018, 27(10): 54-60. <http://www.c-s-a.org.cn/1003-3254/6526.html>

Enterprise Business Analysis System Construction Based on Big Data Cloud Analysis Platform

WEN Shuan-Lao

(Shanghai Baosight Software Co. Ltd., Shanghai 201203, China)

Abstract: For the existing business analysis system of the massive data storage, massive data processing, and business development needs, we designed and built the enterprise business analysis system based on big data cloud analysis platform. This study focuses on the construction of the system technology route, the overall framework of the system to achieve and operational results. The results show that the system can improve the efficiency of data utilization and can provide better data support for the operation of the enterprise.

Key words: business analysis system; big data cloud analysis platform; massive data processing; business operations

1 前言

随着大数据、云计算、移动应用、可视化、人工智能等新技术的出现, 企业经营分析系统呈现出数据来源多样性、数据类型丰富化、数据处理实时化、数据存储海量化等特点. 这些日积月累的数据形成了一个企业的巨大“宝藏”, 如何对其进行有针对性的开发, 挖掘出有价值的信息, 形成企业知识, 指导企业的经营决策, 以适应大数据新时代业务发展需求, 全面支持企业各个层面的经营决策, 是钢铁企业迫切需要解决的问题^[1].

以财务分析为中心理念建设的企业经营分析系统为企业正常运营发挥了关键作用, 但在使用了近十年之后, 经营分析系统已不能满足企业业务发展需求: 对

项目精细化管理缺乏支撑, 信息化存在信息“孤岛”, 影响跨部门数据贯通. 经营分析系统已无法适应大数据新时代的企业全方位业务分析的需求, 系统应用功能也主要以报表方式呈现, 无法满足新形势下的用户自助分析及可视化分析需求, 系统架构采用小型机+传统盘阵+关系型数据库的 IOE 架构^[2], 主要不足方面如下:

1) 数据信息不完整: 数据主要来源于企业内部 ERP 信息系统, 缺少外部数据信息, 缺少对项目精细化管理支撑, 还无法构建面向多层次、多维度、跨领域的信息架构.

2) 架构不够灵活: 现有小型机+传统磁盘+关系型数据库的 IOE 架构, 该架构需要通过一个狭窄的数据

^① 收稿时间: 2017-12-08; 修改时间: 2018-01-04; 采用时间: 2018-02-26; csa 在线出版时间: 2018-09-28

管道将所有的 I/O 信息过滤到共享磁盘子系统, 在进行大量数据的存取时, 无法高效管理磁盘的并发, 处理缓存的同步, 产生数据处理的性能瓶颈^[3].

3) 应用缺乏实时性: 分析数据多为事后处理, 无法及时反应业务数据状态, 无法满足财务日常管理实时性分析需求.

4) 维护成本高: 面对海量数据, 传统经营分析系统负荷急剧增加, 引发 I/O 性能下降, 系统数据处理时间变长、前端应用访问性能变慢等问题, 企业面临着系统存储容量和计算能力的长期扩容压力, 运维成本及投资成本上升.

基于以上分析可以看出, 现有经营分析系统已经不能适应大数据时代的企业业务发展需要, 构建高性能、低成本的基于大数据云分析平台的企业经营分析系统势在必行, 也是企业未来发展的趋势.

2 系统建设技术路线及总体框架图

2.1 设计思路

本系统的设计思路如下: 系统硬件基础平台基于宝之云数据中心先进的虚拟化资源分配技术, 可以通过在线横向扩展 x86 计算节点连接成集群, 利用高可用资源调度技术, 使之在集群上运行^[4], 支持高并发访问和海量数据处理. 海量数据处理工具采用基于 x86 平台的 MPP 数据库集群和基于 x86 平台的 Hadoop 集群, 借助于云计算的支撑, 通过分布式框架和技术, 支撑海量数据处理, 把计算任务分配到要处理的数据所在的节点上运行, 让 CPU 尽可能处理内存的数据或本地硬盘的数据, 避免网络通讯开销, 显著提高分布式系统的性能, 实现对海量数据的高速并行处理. 应用开发工具采用宝信信息系统平台 iPlat4J 及宝信移动应用平台 iPlat4M, 他们都是基于 J2EE 三层架构的技术路线, 是拥有完全自主知识产权, 为支持企业实现业务流程信息化、应用程序移动化而构建的完整平台体系架构, 它通过一系列全面的服务, 帮助企业快速开发应用程序及将应用、数据和业务流程移动化到主流移动设备上.

2.2 技术路线

大数据云分析平台软件技术架构如图 1.

1) 数据抽取层:

数据抽取层部署 IBM DataStage 软件, 完成将源数据从业务系统抽取至数据存储层的功能.

2) 数据存储层:

数据存储层部署 IBM DB2 Data Warehouse 和 Hadoop 软件, 组成 MPP 数据库集群和 Hadoop 集群, 对外提供数据存储和大数据处理服务.

3) 分析服务层:

分析服务层部署 IBM Cognos BI 数据分析软件, 通过软件集群功能组成数据分析集群, 对外提供数据分析功能.

4) 元数据管理:

贯穿整个云分析平台各层之间的元数据管理功能, 安装 IBM DB2 数据库, 提供元数据管理服务.

5) 信息服务层:

信息服务层采用独立的虚拟机进行部署, 安装 WebSphere 中间件组成集群, 分别在其上部署相应的基于宝信信息系统平台 iPlat4J 和宝信移动应用平台 iPlat4M 开发的信息服务应用程序, 通过服务调用方式访问分析服务层的服务, 通过 JDBC 接口访问数据存储层的数据.

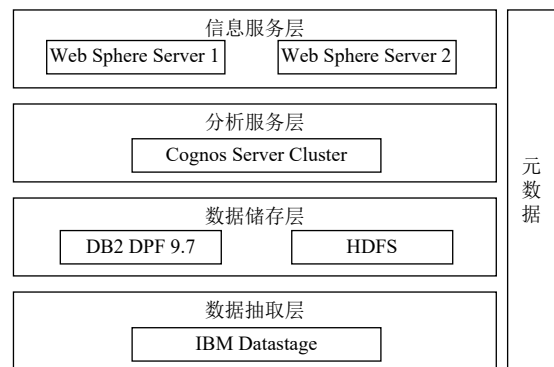


图 1 大数据云分析平台软件技术架构

2.3 系统总体框架图

为满足大数据时代企业业务发展需要, 企业经营分析系统按照新一代“云平台、服务化”架构模式进行架构设计, 采用分布式、大数据、云计算等新技术, 构建高性能、高可用、可扩展的分布式应用系统, 基于大数据云分析平台的企业经营分析系统总体框架如图 2.

按照系统功能职责划分, 基于大数据云分析平台的企业经营分析系统总体框架由数据采集云、数据存储云、数据分析云三个平台层构成, 主要从功能分担的角度描述构建该系统所涉及到的技术, 数据的流向及组织.

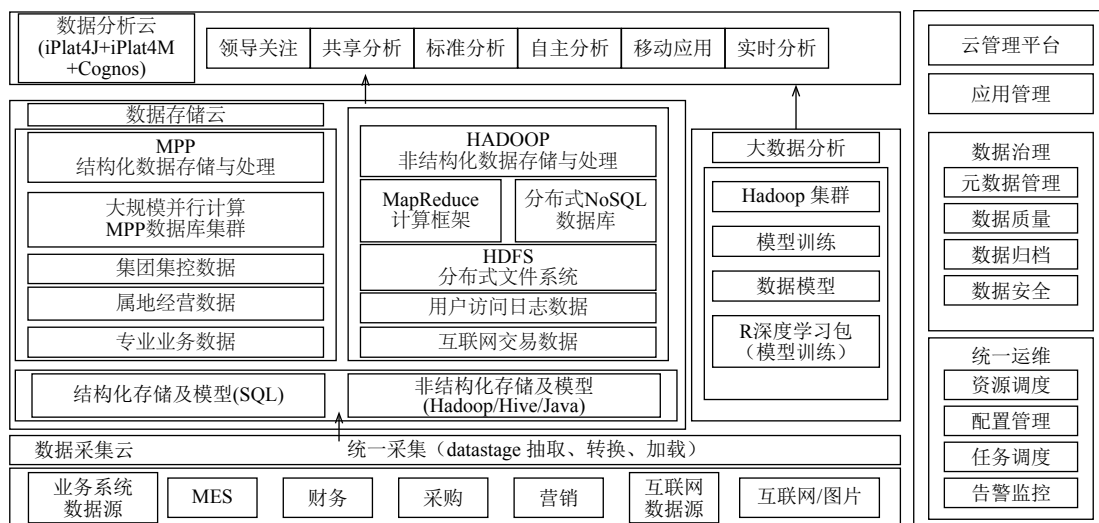


图2 企业经营分析系统总体框架图

2.4 系统各层功能

1) 数据采集云通过云平台 ETL 工具将业务系统原始数据抽取到数据存储云中, 执行源数据的清洗、格式转换和合并计算等功能^[5]. 从业务系统抽取原始数据, 需要考虑以下因素: 元数据注册、增量抽取与加载和调度方式.

2) 数据存储云主要存储业务系统原始数据、主题明细层数据、主题汇总层数据、自助分析数据. 数据整合通过 x86 平台的 MPP 数据库集群和基于 x86 平台的 Hadoop 集群对业务操作数据进行清洗和一系列的转换, 然后装载到数据存储云. 数据存储云按照共享、标准原则, 集中存储企业集团共享数据和标准化数据, 为经营分析系统应用提供基础数据.

3) 数据分析云按照集团共享分析云、标准化分析云、个性化分析云相结合的方式, 实现领导关注、业务分析报表体系、自助分析、数据挖掘等应用功能, 用户借助企业 WEB 信息门户、移动信息门户, 以可视

化方式快速获取数据信息.

3 企业经营分析系统的实现

为应对海量数据存储、海量数据处理及满足企业的工程、服务、研发三大板块业务的发展需求. 基于大数据云分析平台的经营分析系统跨系统采集业务数据, 采用云 ETL 替代原 ETL, 应用云 MPP 数据库并行架构进行数据存储与数据处理, 构建共享应用统一平台, 移动应用统一平台以实现面向多层次、多维度、跨领域的信息架构, 符合企业应用安全体系的认证授权管理, 具体实现如下.

3.1 采用云 ETL 替代原 ETL^[2]

ETL 即数据的抽取 (Extract)、转换 (Transform)、装载 (Load), 基于大数据云分析平台构建企业数据统一采集平台, 实现定时抽取业务系统合同、项目、库存、科研、固定资产、财务原始数据和前端应用在线实时调用数据抽取服务, 实现数据准实时抽取采集, 如图 3.

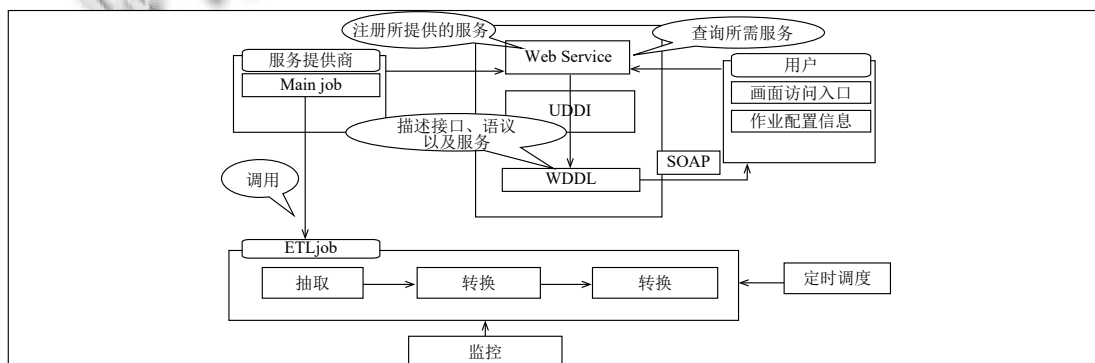


图3 企业数据统一采集平台

1) 作业定时调度

按照业务系统分类, 每个系统创建 Job Sequence 作业, 在 Datastage 中配置调度时间进行调度. 如图 4 所示, 不同数据表之间的抽取依赖, 可以用数据库记

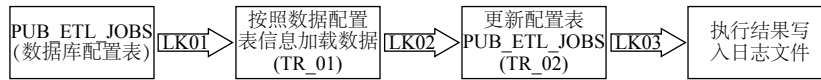


图 4 基于数据库记录方式抽取数据表

2) 实时数据抽取

创建 Datastage Server job 以 Service 方式发布在 Server 服务器上并对外提供实时数据抽取服务, 用户通过使用前端应用功能画面调用 Server job 服务, 实现对 JOB 进行控制, 执行数据抽取作业、执行数据整合作业并将执行状态写入系统日志.

采用云 ETL 替代原 ETL, 使的 Datastage 在从业务系统抽取数据至 DB2 数据库时, 数据加载的高效性得以充分体现, 数据加载速度可达到 1.2G 每分钟, 具体性能参数可参考表 1. 数据抽取性能的提升, 降低了系统抽取异常事件的发生, 有效的降低了系统运维成本. 同时, 实时数据抽取功能, 解决了应用层缺乏实时性的缺点, 满足了财务人员日常财务活动分析需求.

表 1 数据抽取性能参数

应用场景	列数	行数	大小 (MB)	消耗时间 (秒)	速度 (条/秒)	速度 (GB/分)
使用云 ETL	55	182845	4864	237	77 150	1.2
使用传统 ETL	55	182845	4864	698	26 195	0.41

3.2 大规模并行数据处理

构建基于大数据云分析平台的企业经营分析系统“混搭结构”, 选择基于 x86 平台的 MPP 数据库集群与基于 x86 平台的 Hadoop 集群相结合的方式处理海量业务数据, Hadoop 成为了典型的非结构化大数据批量处理架构, 由 HDFS 负责静态数据的存储, 并通过 MapReduce 将计算逻辑分配到各数据节点进行数据计算和价值发现^[6]. 使企业经营分析系统具备完备的数据处理能力, 实现对结构化数据和非结构化数据的有效挖掘和分析^[3].

海量结构化业务数据处理平台部署 IBM DB2 Data Warehouse 软件, 组成 MPP 数据库集群, 将业务系统合同、项目、库存、科研、固定资产、财务数据

录的方式, 每条数据库记录的数据表抽取执行完毕后, 更新数据库配置表的抽取时间及执行状态. 该作业最终用 Sequence Job 进行调用, 实现数据抽取自动化.

抽取到数据存储云, 创建数据表结构时选择数据分布均匀的分区主键, 将数据均匀分区到不同的物理节点上, 按照规划的数据分析主题模型对数据分布式存储和计算, 然后把各节点计算的结果汇总^[7], 以满足企业对海量结构化业务数据处理需求, 基于大数据云分析平台的分布式数据处理, 充分发挥了 MPP 集群海量数据处理优势, 通过主题作业性能测试, 主题作业执行效率比传统平台快 6 倍以上, 批次作业执行性能参考数据如表 3 所示, 数据处理效率的提升减少了作业执行时间, 前端应用响应时间控制在 3 秒以内, 有效的保证了系统的可用性, 提高了用户的满意度.

表 2 测试环境及配置

MPP(海量数据处理平台)	SAS(传统平台)
DELL PowerEdge R720	DELL PowerEdge R720
2 颗 CPU E5-2650/台	16 core CPU/台
16 core/台	16 core/台
4 节点	2 节点
64 G 内存/台	64 G 内存/台

表 3 数据处理性能参考数据

批次作业	MPP(海量数据处理平台)	SAS(传统平台)
pre_comp_product	0:27:46	1:32:02
favd0008	0:01:17	1:38:02
pe_sales10	0:45:16	1:32:02
pe_settle10	0:03:27	0:55:59
pe_inventory10	0:01:12	0:33:39
pe_fi_inventory10	0:00:50	0:58:10
pe_fi_delivery10	0:05:37	0:43:28
pe_contract10	0:01:47	0:49:19
pre_comp_product	0:27:46	1:32:02
favd0008	0:01:17	1:38:02

海量非结构化业务数据处理平台部署 Hadoop 软件, 组成 Hadoop 集群, 它可利用集群资源, 以高并行度处理用户上传日志、网页爬虫、网页分析等文件数据, 并将处理后的结果按照结构化格式回写到 MPP 数据库, 满足企业对海量非结构化业务数据处理需求.

3.3 共享应用统一平台

为了支撑企业的工程、服务、研发三大板块业务的发展,基于大数据云分析平台的企业经营分析系统以合同、项目为主线的数据分析体系,实现营销合同、项目过程管理、采购合同、项目核算跟踪分析,并实现业务与财务的数据贯通、信息共享需求,功能架构如图5.

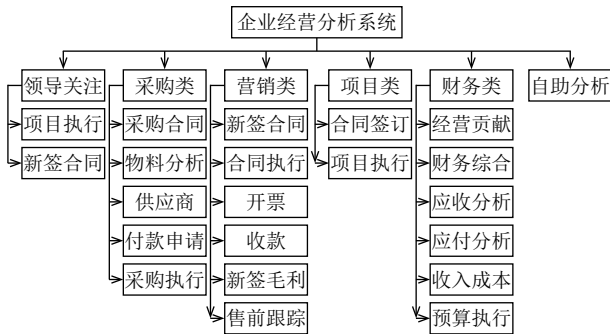


图5 系统功能架构

1) 领导关注

基于大数据云分析平台的企业经营分析系统领导关注模块,将企业领导关注的主营收入、利润、可控费用、新签合同、新签毛利、应收账款、尚未确认收

入、项目状态、资金占用等关键指标,纳入企业核心监控指标体系.采用可视化技术将指标以更直观的方式呈现给领导,为企业经营管理决策服务,如图6.

2) 报表体系

基于大数据云分析平台的企业经营分析系统以财务为中心,以业务为驱动,以营销、工程、服务、采购管理一体化为核心,以合同、项目为主线,构建企业级报表分析体系.通过业务和财务的数据贯通,实现业务与财务活动的双向跟踪分析,提升业务管理水平的同时,也有力的支撑了财务管理的核心内容.同时根据不同类型项目特点,实现项目过程精细化管理.

企业经营分析系统报表数据模型开发工具采用 Cognos FrameWork Manager,它可以连接数据存储云数据库,对业务数据分析模型进行建模,为 Cognos 报表、多维分析等应用提供统一的数据视图,并将生成的元数据包发布到 Cognos 服务器上.报表开发工具采用 Cognos Report Studio,在已发布的元数据包的基础上,开发固定报表、交互报表、交叉报表、多维分析等,Cognos 报表样例如图7所示.

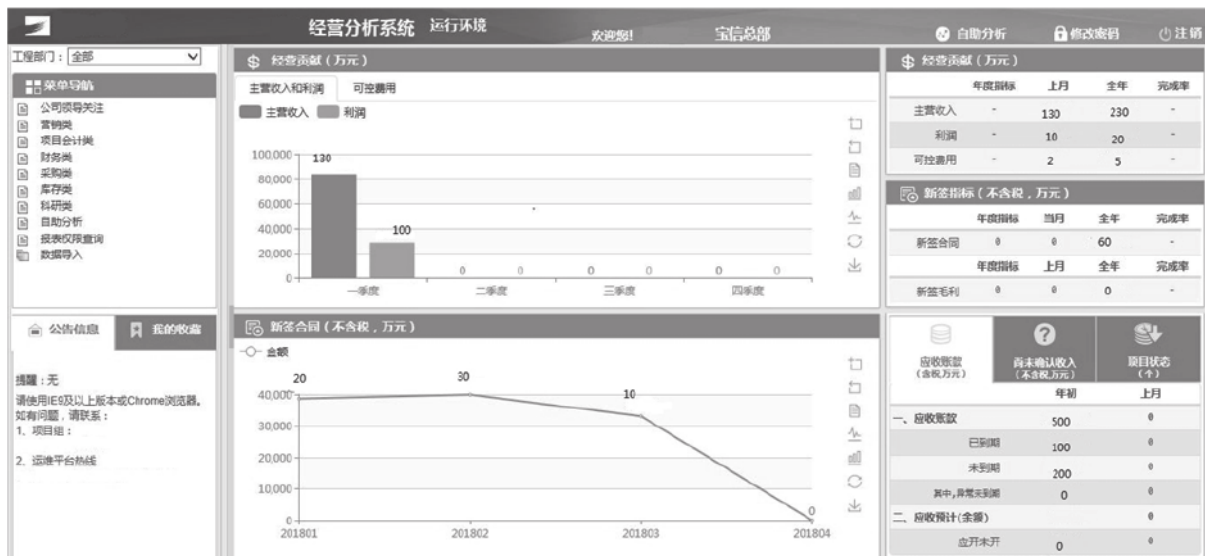


图6 领导关注

3) 自助分析

部分用户由于自身业务的需要必须对数据进行更深入、全面的研究和分析,因此,需要在现有固定报表的基础上进行个性化的分析,自助分析就是为了满足用户个性化分析的需求.基于大数据云分析平台的企业经营分析系统自助分析工具采用基于宝信信息系统

平台 iPlat4J 开发的自助分析应用,完成多维数据计算,自行开发多维数据分析功能^[8].自助分析应用可以将存储云作为数据源,用户通过各种拖拽式操作,设置统计维度、指标及筛选条件,并按照用户设置的报表样式(如:表格、饼图、柱形图、折线图)生成自定义报表,以满足用户个性化需求.同时,系统开发人员由全程定

制报表, 转向专注于后台数据准备工作, 按照业务主题组织好业务数据并将数据表权限开放给用户, 由用户主导按需定制报表, 如图 8 所示.

3.4 移动应用统一平台

基于大数据云分析平台的移动应用统一平台对经营分析系统部分报表功能进行了移动化扩展, 目前功能包括新签合同完成情况 (按营销部门)、新签合同完成情况 (按工程部门)、项目收款跟踪分析等功能. 移动应用功能上线后, 进一步提升了企业协同办公能力.

移动应用统一平台基于宝信移动应用平台

iPlat4M 开发移动报表, iPlat4M 是位于移动终端系统平台与企业应用系统之间的中间层, 它提供后端业务系统快速集成、多种前端平台覆盖和网络通讯适配等服务, 支持 iPad、iPhone、Android 智能手机, 在实现简洁高效、互联互通、资源共享的同时, 满足企业在数据分析、授权管控等方面移动化扩展的要求. 移动应用统一平台采用 HTML5 开发框架, 支持使用 HTML、CSS 和 JavaScript 构建前端移动报表应用. 用户通过安装在移动终端上的基座程序统一访问移动分析应用, 移动应用门户如图 9 所示.

合同规模	2018年合同(01-01月)				2017年合同(01-01月)				2016年合同(01-01月)			
	金额	比例	数	比例	金额	比例	数	比例	金额	比例	数	比例
大于1000万	1001	20%	8	4.49%	100	70.34%	3	2.78%	1000	37.66%	2	1.31%
大于300万小于等于1000万	450	20%	9	5.06%	300	15.07%	8	7.41%	500	27.31%	8	5.30%
大于100万小于等于300万	200	20%	20	11.24%	200	7.55%	16	14.81%	200	20.21%	21	13.91%
小于等于100万	50	20%	141	79.21%	50	6.24%	81	75.00%	100	14.82%	120	79.47%
总计	1500	100%	178	100%	650	100%	108	100%	1800	100%	151	100%

图 7 Cognos 报表样例

图 8 自助分析工具

3.5 统一身份认证管理系统

为了满足企业对使用业务数据安全管理的的要求, 在企业级层面建立了一套统一认证管理系统, 负责对企业内部所有业务系统提供实时在线用户身份认证服务^[9], 实现了用户采用一套用户名和密码即可访问不同应用

系统的效果, 统一身份认证管理系统如图 10 所示.

宝信信息系统平台 iPlat4J 平台组件 ePass 企业认证和授权管理产品, 实现了与 IBM Cognos BI 数据分析软件的数据管理产品的认证、授权的集成, 基于大数据云分析平台的企业经营分析系统基于统一的用户

授权管理界面,对系统功能、数据资源进行授权和管理,用户登录进系统后,获取登录用户拥有的角色,按照角色检核用户资源操作权限并进行权限控制。

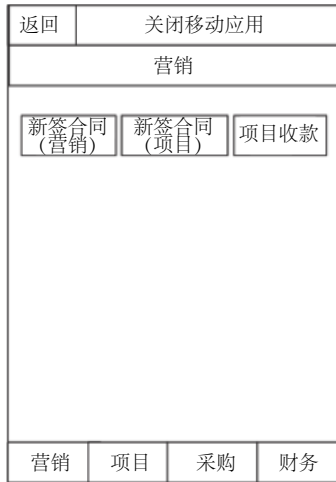


图9 移动应用统一平台

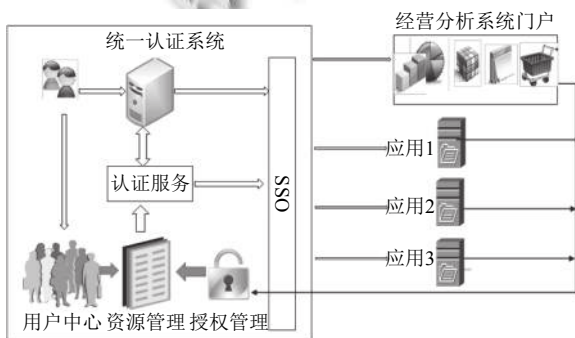


图10 统一身份认证管理系统

4 企业经营分析系统运行效果

基于大数据云分析平台的企业经营分析系统,遵循“整体规划、分步实施”的策略,于2013年投入使用,2016年全面建成,系统的广泛应用,取得了以下效果:

1)跨系统采集、管理、分析合同、项目、采购、资产、财务业务数据,提供面向业务主题的多层次、多方面、多维度信息架构,全面支撑企业各个层面的经营决策。

2)灵活的混合平台架构既支持传统数据分析也支持大数据分析。具备高效快速的海量数据存储和大数据处理能力,解决了系统性能瓶颈问题,提升了作业执行效率,降低了企业运维成本。

3)实时数据抽取功能,变离线分析为在线分析,解决了应用层缺乏实时性的缺点,满足了财务人员日常

财务活动分析需求。

4)企业在报表功能、授权管控等方面实现了移动化扩展的要求,用户可以通过主流移动化设备访问报表数据并跨地域办公,进一步提升了企业在线协同办公能力。

5)数据安全方面实现了统一认证管理系统,用户采用一套用户名和密码即可访问不同应用系统的效果,实现了企业多系统的集成和单点登录效果。

5 小结

为应对海量数据存储、海量数据处理及支撑企业工程、服务、研发三大板块业务发展需求的背景下,设计并构建了基于大数据云分析平台的企业经营分析系统。本文重点介绍了该系统的建设技术路线、总体框架图、系统实现及应用效果。基于大数据云分析平台的企业经营分析系统的应用,满足了用户日常业务数据分析、信息共享和业务协同需求,支持企业绩效评估、经营分析、经营决策、战略设计以及各部门的业务运作优化等。在后期的系统建设方面我们还要结合用户的使用效果,继续加强应用深度和应用广度两个方面的拓展。从未来的发展看,大数据资产对企业业务发展越来越重要,与第三方企业合作,构建大数据生态体系和能力开放平台的数据租赁、数据分析、数据分享服务等将是下阶段的研究方向。

参考文献

- 1 于戈,鲍玉斌,嵇晓. 数据仓库工程方法论. 沈阳: 东北大学出版社, 2003
- 2 孙淳晔. 基于大数据平台的企业级经营分析系统建设探讨. 电信工程技术与标准化, 2015, (1): 78-81. [doi: 10.3969/j.issn.1008-5599.2015.01.019]
- 3 王昕, 张金刚. MPP 数据库在经营分析系统中的应用探讨. 互联网天地, 2006, (3): 56-60.
- 4 覃雄派, 王会举, 杜小勇, 等. 大数据分析—RDBMS 与 MapReduce 的竞争与共生. 软件学报, 2012, 23(1): 55-56.
- 5 余长慧, 潘和平. 商业智能及其核心技术. 计算机应用研究, 2002, (9): 14-16, 26. [doi: 10.3969/j.issn.1001-3695.2002.10.005]
- 6 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述. 软件学报, 2014, 25(9): 1889-1908.
- 7 杨安萍, 唐小三. 基于云计算的业务海量数据处理方法. 广播电视信息, 2015, (7): 53-55. [doi: 10.3969/j.issn.1007-1997.2015.07.018]
- 8 张旭, 董有田. OLAP 多维数据分析与应用研究. 黑龙江科技学院学报, 2002, 12(3): 17-19. [doi: 10.3969/j.issn.1671-0118.2002.03.006]
- 9 杨鸣. 宝钢集团统一认证系统实现. 中小企业管理与科技, 2013, (3): 239-241. [doi: 10.3969/j.issn.1673-1069.2013.03.164]