

DOI: 10.16516/j.gedi.issn2095-8676.2016.S1.001

基于大数据的 DSS 融合架构研究

冯国平, 解文艳, 王海吉, 黄翔

(中国能源建设集团广东省电力设计研究院有限公司, 广州 510663)

摘要: 传统决策支持系统(DSS)技术架构多为传统的基于 SMP 的数据库集群架构或 MPP 架构, 存储和计算能力构建成本高, 扩展性受限、并发性不高, 大规模离线计算和流式计算场景下存在性能瓶颈。通过分析大数据时代 DSS 面临的挑战, 提出基于大数据的 DSS 融合架构。这种架构支持结构化数据、半结构化数据和非结构化数据分析处理, 全面满足关系运算、大规模离线分布式计算和流计算要求。

关键词: 决策支持系统; MPP; Hadoop; 融合架构

中图分类号: TP311.13

文献标志码: A

文章编号: 2095-8676(2016)S1-0001-04

Research on Fusion Architecture of DSS Based on Big Data

FENG Guoping, XIE Wenyan, WANG Haiji, HUANG Xiang

(China Energy Engineering Group Guangdong Electric Power Design Institute Co., Ltd., Guangzhou 510663, China)

Abstract: Traditional architectures of decision support system are most based on database cluster of SMP or MPP with a high cost of storage and computing. It is difficult to expand, and the performance bottleneck of large scale offline computing and streaming computing. By analyzing the challenges of DSS, this paper puts forward a fusion architecture of DSS based on big data, which satisfies the requirements of relational operations, large-scale offline distributed computing and streaming computing. It supports the analysis and processing of structured data, semi-structured data and unstructured data.

Key words: decision support system; MPP; hadoop; fusion architecture

企业决策支持系统(DSS)是以管理学、运筹学、控制论和行为科学为基础, 以人工智能技术、智能仿真技术和信息网络技术为手段, 针对企业经营管理行为中的决策问题, 辅助企业经营管理开展决策活动的信息化系统。基于丰富的数据资源, DSS 通过建立决策分析模型, 识别待决策的问题, 明确决策目标, 为决策人员提供各种决策方案, 并对各种决策方案进行量化评价, 并能进行交互式的分析和比较, 支撑企业经

营管理正确的决策, 促进企业决策科学化、智能化^[1]。

1 决策支持系统发展

DSS 为企业各层级提供辅助决策支持, 其服务的对象不仅包括企业的战略决策层, 还包括企业的经营管理层和业务操作层。DSS 一般具有如下特点: (1)构建、管理并应用各种决策模型; (2)收集和管理与决策问题有关的企业内外部信息资源; (3)提供友好的人机交互界面; (4)支持决策主体之间的横向协同。

1.1 DSS 体系架构

一般 DSS 的体系架构包括数据源、数据仓库、BI 平台、决策分析应用以及标准体系、运营管理体系和安全体系等部分, 如图 1 所示。

数据源包括企业的内部数据和外部数据, 内部

收稿日期: 2016-02-01

基金项目: 中国能建广东院科技项目“南方电网公司电源资产管理信息系统信息化解决方案研究及充电设施典型设计”(EX03881W)

作者简介: 冯国平(1980), 男, 湖北云梦人, 高级工程师, 硕士, 主要从事电网企业信息化规划、大数据相关研究工作(e-mail) feng-guoping@gedi.com.cn。

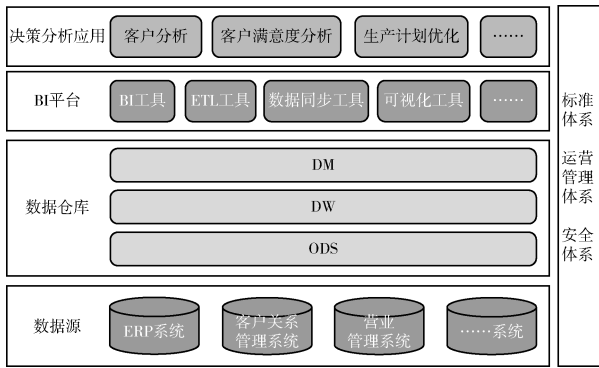


图1 决策支持系统体系架构

Fig. 1 System architecture of DSS

数据包括企业的经营管理数据和生产运行数据，外部数据包括情报数据、上下游合作伙伴数据等。这些数据从类型上包括结构化数据、半结构化数据和非结构化数据。数据仓库包括 ODS、DW 和 DM，是数据转换、存取、管理和加工的场所，通常包括一些业务分析模型^[2]。BI 平台为分析应用提供工具支撑，包括 OLAP、ETL 和报表展现等工具。决策分析应用基于数据仓库和 BI 平台，通过建立决策分析的预测、监控模型，为企业提供辅助决策手段^[3]。

标准体系、运营管理体系和安全体系也是决策支持系统体系架构的重要组成部分，明确数据的标准规范和安全要求，规范决策支持系统的运营管理组织和制度等。

1.2 DSS 技术架构演进

随着人们对信息处理规律认识的不断提高，面对不断变化的环境需求，管理者迫切需要在生产管理中积极发挥作用，要求更加智能化的信息系统来支持决策。DSS 于 20 世纪 70 年代提出，80 年代得到发展。从技术架构的发展，将 DSS 分为三个发展阶段。

第一个阶段为单机系统阶段。受限于计算机软硬件水平发展水平，这种技术架构采用基于关系数据库的单机系统实现关系运算，主要支持小规模静态统计分析，满足简单的报表统计和展现需求。

第二个阶段为基于 SMP 的数据库集群架构阶段。根据计算机体系结构的发展，采用基于 SMP 结构实现基于关系数据库的集群。这种技术架构可支持复杂的报表统计、数据可视化和 OLTP 业务。

第三个阶段是 MPP 架构阶段。这种架构由许

多松耦合的处理单元组成，其最大的特点在于不共享资源。MPP 架构侧重实现大规模的统计分析，支持大规模数据分析，但是节点扩展有限，构建成本较高。

2 传统 DSS 技术架构

本文将上述三个阶段的技术架构归为传统 DSS 技术架构。

2.1 基于 SMP 的数据库集群架构

传统 DSS 技术架构基本采用关系数据库 RDBMS + 小型机 + 高端阵列模式，服务器为 SMP 架构，如图 2 所示。在基础设施层面，基本采用运行在 Unix 操作系统上计算能力强、价格昂贵的小型机，与 X86 服务器采用的 CISC 指令集不同，而是采用 RISC 指令集。在物理存储方面，主要是通过存储阵列实现大规模存储，存储的成本较高。

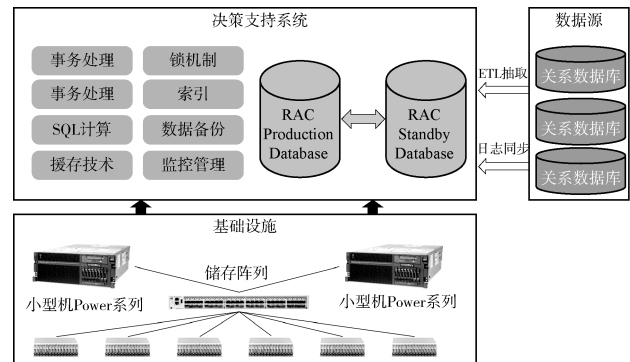


图2 基于 SMP 的数据库集群技术架构

Fig. 2 Database cluster based on SMPP

在数据采集方面，采用 ETL 工具实现数据的抽取、转换和加载，或者通过专业的数据库日志同步工具，如 GoldenGate 实现数据的准实时增量同步。在数据存储方面，以成熟的关系数据库为主。一般通过数据库双机集群方式实现多节点负载均衡，保障高可用，并通过并行执行技术提高事务响应时间。

在计算方面，主要是通过关系数据库以 SQL 方式实现关系运算，数据计算能力受限于服务器的性能，对大规模离线计算和流式计算会存在性能瓶颈。

基于数据仓库的决策支持系统一般是 T + 1 模式（分析人员今天只能看到前一天或更早的数据），侧重满足静态分析需求。由于传统架构扩展性较

差、性能存在瓶颈等, 支持的数据规模难以超过 PB 级, 即席查询、实时分析等对性能和实时性要求较高的操作难以实现, 不能满足大量一线人员参与数据分析需求。

2.2 基于 MPP 技术的 DSS 架构

MPP (Massively Parallel Processing, 即大规模并行处理系统) 由若干个松耦合处理单元组成, 每个单元内的 CPU 都有自己私有的计算资源, 包括主机、操作系统、内存、存储等都是自我控制, 在每个单元内都有操作系统和管理数据库的实例副本。节点之间的信息交互是通过节点通信网络实现。通过将数据分布到多个节点上实现大规模数据的存储, 通过并行查询处理来提高查询性能。MPP 技术架构如图 3 所示。

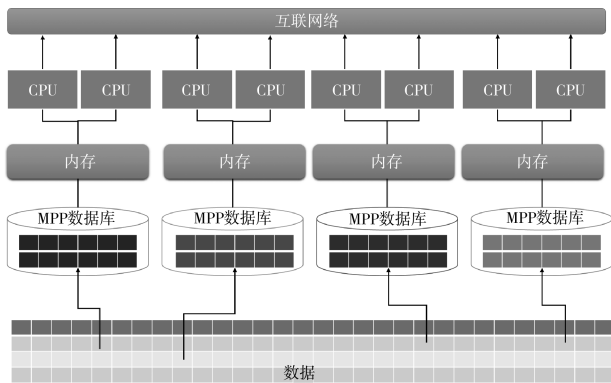


图 3 MPP 技术架构
Fig. 3 MPP architecture

这种采用 MPP 架构的新型数据库集群的主要设计目标是横向扩展, 数据库支持严格的关系模型, 支持事务处理, 保持数据的强一致性。MPP 架构的主要特征包括: (1) 任务并行执行; (2) 数据分布式存储(本地化); (3) 分布式计算; (4) 私有资源; (5) 横向扩展; (6) Shared Nothing 架构。

一般采用 MPP 架构的产品大都可以支撑 PB 级别的数据分析, 这是传统数据仓库技术难以实现的。从性能来讲, MPP 数据库在多维复杂查询的性能确实要优于 HIVE/HBASE/IMPALA 等, 但是在扩展性方面, 实际支持的节点数有限(很少能超过 100 个节点), 与 Hadoop 可以支持的集群节点规模相比存在较大差距。MPP 架构严格遵循 CAP 原则: 一致性(Consistency): 数据一致更新, 所有数据更新都保持同步; 可用性(Availability), 保持好的响应性能; 分区容错性(Partition tolerance)。对于一

般分布式系统而言, 在架构上极难兼顾上述 CAP 原则。MPP 数据库还是基于原 RDMBS 扩展而来, RDBMS 追求一致性必然造成分区容错性较差。当集群规模很大业务数据太多时, MPP 数据库的元数据管理相当困难, 系统一旦出错很难恢复。

在并发性方面, MPP 数据库的核心原理是将大的查询通过分解为多个子查询, 分布到底层执行, 最后再合并结果。对于单个查询来说该方法调用了整个系统的能力, 因而查询比较快, 但是整个系统支持的并发不高。因此 MPP 架构的应用场景适合小集群、低并发, 难以应对大数据量的实时分析场景。

3 基于大数据的 DSS 融合架构

3.1 大数据时代 DSS 面临的挑战

传统的 DSS 基于关系数据库、数据仓库、OLAP、BI、ETL 等工具和技术, 实现数据资源的采集、存储、加工、分析和应用。随着大数据时代的来临, 数据呈现体量大、类型多等特点, DSS 在数据存储计算能力、数据处理与转换、系统柔性扩展等方面存在瓶颈。传统的 DSS 技术架构已经无法满足当前决策分析的要求。

随着大数据时代的来临, 数据呈爆炸式增长, 传统基于 SMP 的数据库集群或 MPP 架构的系统面临新的挑战及需求^[4]:

- 1) 随着企业业务的不断发展, 移动互联网、物联网等新技术应用, DSS 面临海量数据处理需求。
- 2) 大数据应用往往是结构化与非结构化数据的融合。非结构化数据占企业数据总量的 80%, DSS 面临着海量非结构化数据存储、处理、分析需求。
- 3) 大数据应用不仅需要企业的经营管理数据, 还包括了流数据资源, 如状态信息、气象信息等, DSS 需要满足对流数据的存储、处理、分析需求。

3.2 基于大数据的融合架构

在大数据时代, 企业一般将 DSS 定位为企业的大数据平台。传统 DSS 架构已经不能满足大数据时代决策分析需求。本文结合企业实际需求, 从系统构建的成本、数据存储计算的能力要求出发, 提出了基于大数据的 DSS 融合架构, 如图 4 所示。

上述融合架构保留了传统基于关系数据库的数据仓库, 又拓展了基于分布式文件系统构建海量批量数据的分布式计算和流计算功能, 即具备高价值

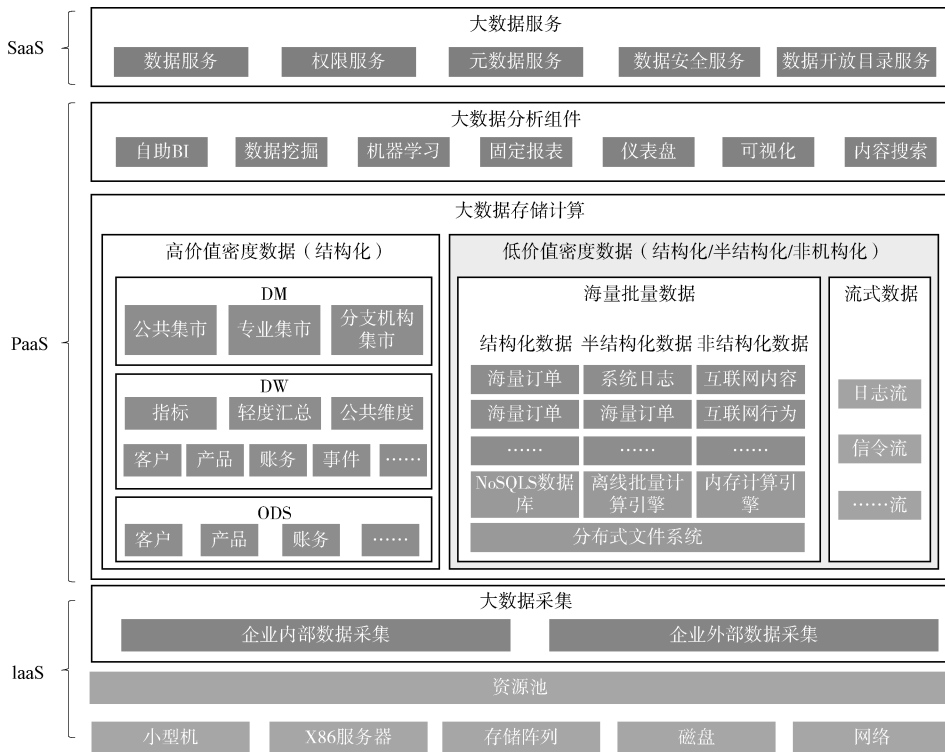


图 4 DSS 融合技术架构
Fig. 4 Fusion architecture of DSS

密度的结构化数据处理能力，也具备结构化、半结构化和非结构化以及流式数据处理能力。

3.3 融合架构的特点

基于大数据的 DSS 融合架构具有以下特点：

1) 提升了海量数据的存储和计算能力。采用廉价的 X86 服务器构建 Hadoop 分布式系统，在保证数据安全的前提下，可实现 PB 级的大规模数据存储和大规模离线分布式计算，同时还扩展了流计算能力。

2) 兼顾了传统关系运算的应用需求。由于 SQL 计算效率高，基于 SQL 的关系运算应用场景多，市场上支持的 BI 工具也很广泛，因此有必要兼顾传统关系运算的应用需求。

3) 构建成本低。传统架构为了提高计算性能，大型企业多采购高端的商用小型机，系统建设成本昂贵。上述融合架构既利旧原有计算存储设备，又采用成本低廉的 X86 服务器进行扩展。

4) 可扩展性强。数据库集群架构扩展能力较差，MPP 架构具有一定扩展能力，但是节点数达到一定级别后就难以扩展。而基于大数据的融合架构

具有很强的扩展性，基本不受限制。

4 结论

基于大数据的 DSS 融合架构即保留了满足 ACID 事务处理要求的关系运算能力，又引进以 MapReduce/Hadoop 为代表的大规模批量离线计算技术，能够以插件形式配置多种数据计算引擎，提供高并发的、大数据处理能力，这种架构是当前众多企业最适当的选择。随着大数据相关技术的不断进步，分布式架构对关系运算以及 ACID 事务支持能力不断增强，未来 DSS 架构也许会逐渐从融合架构向完全分布式架构转变。

参考文献：

- [1] 刘博元, 范文慧, 肖田元. 决策支持系统研究现状分析 [J]. 系统仿真学报, 2011(7): 241-244.
LIU B Y, FAN W H, XIAO T Y. Development of decision support system [J]. Journal of System Simulation, 2011(7): 241-244.
- [2] 陈曦, 王执铨. 决策支持系统理论与方法研究综述 [J]. 控制与决策, 2006, 21(9): 961-968.

参考文献:

- [1] 赵云飞, 陈金富. 层次分析法及其在电力系统中的应用 [J]. 电力自动化设备, 2004, 24(9): 85-89.
ZHAO Y F, CHEN J F. Analytic hierarchy process and its application in power system [J]. Electric Power Automation Equipment, 2004, 24(9): 85-89.
- [2] 袁昕. 基于 AHP 的北京城区电网可靠性评价体系研究与应用 [D]. 华北电力大学, 2012.
- [3] 李金芳. 基于 AHP/DEA 的 20 kV 中压配电方案研究及评估 [D]. 华北电力大学, 2009.
- [4] 柳顺, 杜树新. 基于数据包络分析的模糊综合评价方法 [J]. 模糊系统与数学, 2010, 24(2): 93-96.
LIU S, DU S X. Fuzzy comprehensive evaluation based on data envelopment analysis [J]. Fuzzy Systems and Mathematics, 2010, 24(2): 93-96.
- [5] 林济铿, 李童飞, 赵子明, 等. 基于熵权模糊综合评价模型的电力系统黑启动方案评估 [J]. 电网技术, 2012, 36(2): 51-54.
LIN J K, LI T F, ZHAO Z M, et al. Assessment on power system black-start schemes based on entropy-weighted fuzzy comprehensive evaluation model [J]. Power System Technology, 2012, 36(2): 51-54.
- [6] 李志铿, 杨海森, 陈超. 计及转供电概率的配电网可靠性评估 [J]. 南方能源建设, 2014, 1(1): 30-33.
LI Z K, YANG H S, CHEN C. Reliability evaluation for distribution system considering supply probability [J]. Southern Energy Construction, 2014, 1(1): 30-33.
- [7] 梁赞, 冯永青. 基于可信性理论的有源配电网可靠性分析 [J]. 南方能源建设, 2015, 2(2): 62-65.
LIANG Y, FENG Y Q. Reliability assessment of distribution networks with distributed generators based on credibility theory [J]. Southern Energy Construction, 2015, 2(2): 62-65.
- (责任编辑 高春萌)

(上接第 4 页 Continued from Page 4)

- CHEN X, WANG Z Q. Overview of theory and methods of decision support systems [J]. Control and Decision, 2006, 21(9): 961-968.
- [3] 蒋良孝, 蔡之华. 基于数据仓库的数据挖掘研究 [J]. 计算机技术与自动化, 2003(3): 102-105.
JIANG L X, CAI Z H. The Research of data mining based on data warehouse [J]. Computing Technology and Automation, 2003(3): 102-105.
- [4] 王珊, 王会举, 覃雄派, 等. 架构大数据: 挑战、现状与展望 [J]. 计算机学报, 2011, 34(10): 1741-1752.
WANG S, WANG H J, QIN X P, et al. Architecting big data: challenges, studies and forecasts [J]. Chinese Journal of Computers, 2011, 34(10): 1741-1752.
- [5] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战 [J]. 计算机研究与发展, 2013, 50(1): 146-169.
Meng X F, C X. Big data management: concepts, technique and challenges [J]. Journal of Computer Research and Development, 2013, 50(1): 146-169.
- (责任编辑 张春文)

(上接第 8 页 Continued from Page 8)

- [2] 张植华, 李健, 林毓, 等. 网格化城市配电网目标网架动态构建方法 [J]. 陕西电力, 2015, 43(2): 23-29.
- [3] 李海涛, 孙波, 王轩. 配电网网格化规划方法及其应用 [J]. 电力系统及其自动化学报, 2015(S1): 33-37.
- [4] 沈健, 盛佳蓉, 张麟, 等. 基于 AHP 的北京城区电网可靠性评价体系研究与应用 [D]. 北京: 华北电力大学, 2012.
- [5] 李金芳. 基于 AHP/DEA 的 20 kV 中压配电方案研究及评估 [D]. 北京: 华北电力大学, 2009.
- [6] 柳顺. 基于数据包络分析的模糊综合评价方法及其应用 [D]. 杭州: 浙江大学, 2010.
- [7] 林济铿, 李童飞, 赵子明, 等. 基于熵权模糊综合评价模型的电力系统黑启动方案评估 [J]. 电网技术, 2012, 36(2): 51-54.
- [8] 龚剑, 胡乃联, 崔翔, 等. 基于 AHP-TOPSIS 评判模型的岩爆倾向性预测 [J]. 岩石力学与工程学报, 2014, 33(7): 1442-1448.
- [9] 侯志东, 吴祈宗. 基于 Hausdauff 度量的模糊 TOPSIS 方法研究 [J]. 数学的实践与认识, 2005, 35(3): 233-237.
- (责任编辑 郑文棠)