

智能电网大数据信息平台研究

黄翔, 陈志刚

(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: 智能电网已成为世界各国发展的重点。智能化运行, 依赖于广域、高频的智能监控设备感知和控制电网运行。然而, 海量的监控数据, 远远超出了传统数据库处理能力, 对数据处理提出了前所未有的挑战。大数据技术打破了传统串行化、集中化的数据处理方式, 采用分布式并行化的计算模型, 解决了数据中心横向扩展难的问题, 在大规模数据存储、处理和分析领域取得了巨大的成功。重点分析了智能电网建设对信息系统数据处理方面带来的挑战, 讨论了利用大数据技术解决智能电网海量实时数据的现实性与可行性, 建立了智能电网信息化平台建设的理论和技术基础。

关键词: 智能电网; 大数据; 信息平台

中图分类号: TM769

文献标志码: A

文章编号: 2095-8676(2015)01-0017-05

Research on Big Data Information Platform for Smart Grid

HUANG Xiang, CHEN Zhigang

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

Abstract: Smart grid has become one of the main development trends of power grid all over the world. However, smart control depends on waste area and high frequency monitor. The huge size of the monitor data has exceeded traditional database's ability because it can only handle limited data size with the centralized and serialized data process method. Big data technologies using distributed parallelized approaches overcome the above shortcuts, and make a big success in many practical areas. In this paper, we focus on the challenges in smart grid data process, and discuss the methods how to adapt the technologies for smart grid. Our target is to establish an information platform for smart grid construction in near future.

Key words: smart grid; big data; information platform

随着节能减排、绿色能源和可持续发展成为关注的焦点, 世界各国对下一代电网均给予了极大的关注。近几年来, 我国对智能电网建设也给予了高度重视, 将“建设国家智能电网”写入了发展规划。智能电网^[1](如图 1 所示), 通过数字化信息网络将发电、输电、变电、配电、供电、售电以及终端用户的各种电气设备连接在一起, 通过智能化控制实现精确供能、对应供能、互助供能和互补供能, 以提高能源利用效率和能源供应安全水平, 同时降低污染与温室气体排放。

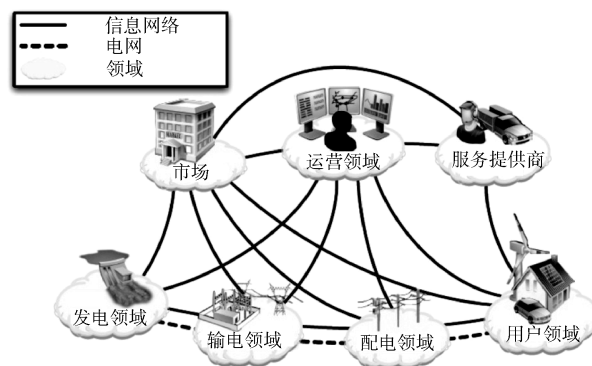


图 1 智能电网概念模型

Fig. 1 Smart Grid Conceptual Model

收稿日期: 2014-10-20

基金项目: 中国博士后基金资助项目(2014M552268); 广东省自然科学基金资助项目(S2013040011941)

作者简介: 黄翔(1982), 男, 湖南常德人, 博士, 主要从事智能电网, 大数据与云计算研究(e-mail) huangxiang@gedi.com.cn。

智能精准的控制依赖于大量高速监控装置的引入。新型的监测装置不仅涵盖一次系统设备, 还囊括了二次系统设备。监测数据不仅包括实时在线状态数据, 还包括设备基本信息、试验数据、运行数

据、缺陷数据、巡检记录、带电测试数据等离线信息^[2]，数据量极大、可靠性和实时性要求高，远远超出传统电网状态监测的范畴。面对这些海量的、分布式的、异构的、复杂的状态数据，常规的数据存储与管理方法会遇到极大的困难。随着传统数据处理技术日渐显现出处理能力的不足，大数据技术迅速得到广泛的关注和重视。目前，已经在互联网、通信、政府、医疗、零售等众多行业取得巨大的成功。大数据技术以其超大规模的处理能力、低廉的建造价格和优良的可靠性与扩展性，已成为了海量数据处理的关键技术。

本文着重讨论智能电网大数据特点，结合电力系统对数据处理的要求，分析现有大数据技术应用在电力行业的可行性与可能性，给出推荐的方案，为下一代智能电网信息平台的建设奠定基础。

1 智能电网大数据

目前，智能电网建设还处于理论研究和小规模尝试阶段，世界各国对智能电网的定义和认识也不尽相同。但是总的来说，智能电网是通过传感器把各种设备连接到一起，形成一个双向互动的信息化网络，进而对信息进行整合分析，实现智能化控制，到达降低成本，提高效率、优化管理的目的。大数据产生于电力系统的各个环节^[3]，包括：

1)发电侧。随着数字化电厂建设的推广，海量的生产数据被保存下来，用于分析生产运行，优化控制策略和故障诊断分析等。此外，随着大量新能源的建设，对环境气象数据的要求也越来越高，这类数据的存储规模也在急剧增长。

2)输变电侧。随着同步相量监测系统(Phasor Measurement Unit, PMU)的大规模使用，输变电网能更全面的了解电网运行暂态的变化，以便更实时的控制电网运行。PMU采样频率非常高。一个PMU设备一天可收集6千万个数据点，数据量约0.6 GB，实际应用中，大面积的部署将产生TB级的数据。

3)用电侧。与传统电网相比，智能电网下用户与供电公司双向互动^[4]，并参与到电力系统的运行和管理中。利用智能电表，电力公司可以实时了解用户用电情况，并实时通知用户用电成本、实时电价、电网状态、计划停电等信息。目前，智能电表

可每隔5~10 min向电网发送实时用电信息。例如法国电力公司所部属的3500万个智能电表，每月产生的数据超过300 TB。

面对TB甚至PB级的智能电网数据，传统的数据库技术暴露的问题日益显著。其中，最显著的瓶颈在于串行的数据处理方式，难以应对爆炸增长的数据规模。根据实际经验，在当前的主流硬件配置下，传统数据库处理能力大致在10 TB以内。

2 智能电网大数据平台

考虑到电网数据海量增长的特点，本文提出了一种分级混合存储的智能电网信息化大数据基础平台(以下简称“大数据平台”)。

2.1 分级存储

智能电网存储多种类型的数据，比如生产运行数据、营销计量数据、资产数据、管理规划数据等。它们不论从数据格式和使用频率上都存在很大的差异。

本文从数据格式、数据规模和使用频度等方面出发，设计了分级混合的智能电网大数据信息平台。分级混合存储系统按照数据使用频度、数据格式和数据规模的不同，将其分布在不同级别的存储介质之上，使其在经济性、可靠性和性能等方面达到平衡。大数据平台主要功能划分如图2所示，最上层是传统的应用服务层，中间是分级存储服务层，统一管理其下的存储资源。

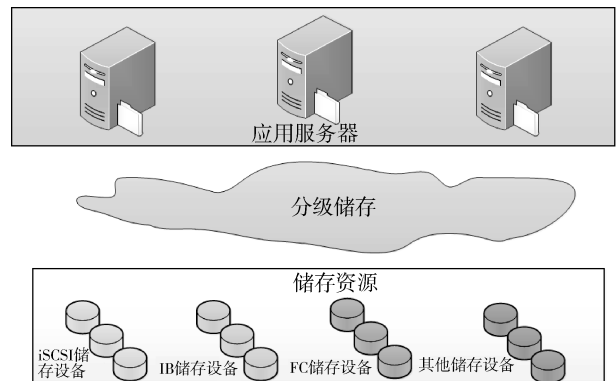


图2 智能电网海量数据分级存储

Fig. 2 Big Data Hierarchical Storage for Smart Grid

2.2 分级混合大数据平台技术架构

分级混合大数据平台实现方案如图3所示。该平台考虑了智能电网海量的半结构化和非结构化数

据, 同时根据不同数据的性能需求优化数据读取的速度。

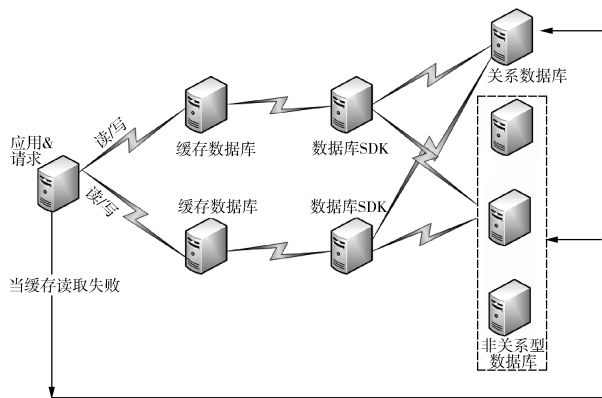


图 3 数据库系统实现方案

Fig. 3 Implementation of Databases System

该实现方案主要包括三个部分: 缓存数据库、关系数据库、非关系型数据库。

1) 缓存数据库。主要是通过数据缓存的方式提高数据访问速度。缓存的数据都加载在内存当中, 只存储热点数据。对于缓存中不存在的数据, 访问申请将被重定向到持久化存储数据库(关系型、非关系型数据库)中去读取数据。缓存数据库采用分布式部署, 使用哈希技术管理数据信息。

2) 关系型数据库。主要用于兼容电网当中现有的业务系统, 保证电网现有业务系统中存储的数据在未来仍然可用。这类数据库主要存储一些结构化、对索引要求比较高的数据。

3) 非关系型数据库。非关系型数据库(HBase)主要用于解决智能电网中海量半结构化和非结构化数据的持久化存储需求^[5]。通过采取比关系型数据库弱的一致性模型, 数据模型采用键值对或者多维表, 采用集群部署方式, 并提供日志重做来实现失效恢复。

2.3 大数据存储方案

智能电网海量的信息、数据需要可靠存储, 同时还需要支持大量的快速的访问申请。如果使用传统的单机存储, 可能会成为整个系统的瓶颈和障碍。本文采用分布式文件系统(HDFS), 将数据的存储和访问分布在大量服务器之中, 用作非关系型数据库的存储基础, 如图 4 所示。

该架构设计可以提供以下特点:

1) 高吞吐量访问。系统以数据块为基本单位, 每个数据块都有若干备份数据块。数据块分布在不

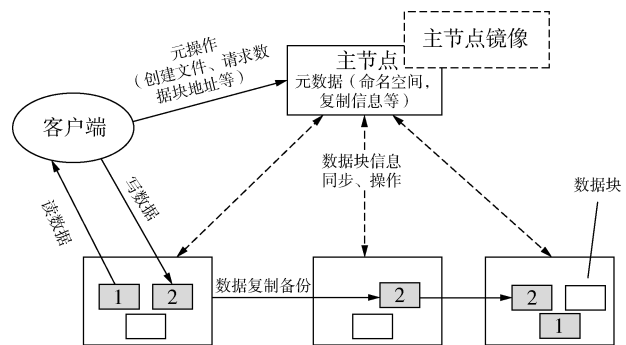


图 4 基础架构

Fig. 4 Infrastructure

同机架的服务器上, 数据访问时, 系统将使用网络最近和访问量最小的服务器。另外, 备份数据块也能提供访问, 从而极大地提高了数据吞吐量。

2) 无缝容量扩充。系统将数据块的索引存放在主节点, 内容存放在数据节点。当需要扩充时, 只要增加数据节点的数量, 系统即会自动将新的服务器添加进主节点, 然后利用数据均衡算法会将数据块迁移到新的数据节点。整个过程无需系统宕机和人工干预。

3) 高度容错。数据在写入时被复制成多份, 存储在不同的物理服务器上。在读写时, 会自动进行数据校验, 发现数据校验错误将重新进行复制。另外, 针对于主节点单点故障, 对主节点增加了镜像功能。主节点服务器出现故障时, 可瞬间迁移到镜像服务器。

另外, 考虑到一些电网历史数据使用的频率会随着时间的推移越来越低, 根据数据对失真的容忍程度, 本文采用有损和无损相结合的压缩方案。使得在尽可能小地影响数据的可用性的前提下, 最大化数据存储的效率。比如传感器产生的日志文件, 可采用有损压缩处理, 而对于另一些音像文件, 可以进行有损压缩。

2.4 大数据平台管理系统

大数据平台管理系统的架构如图 5 所示, 该分级存储系统逻辑上主要包括客户端、管理器和存储资源代理三个部分, 客户端、管理器和存储资源代理通过以太网相连。

客户端即应用服务器。客户端是访问智能电网海量数据分级存储系统的入口, 它向电网管理者提供一个标准的接口, 通过该接口可以透明的访问系统中的数据, 并能进行相应的数据管理操作如数据

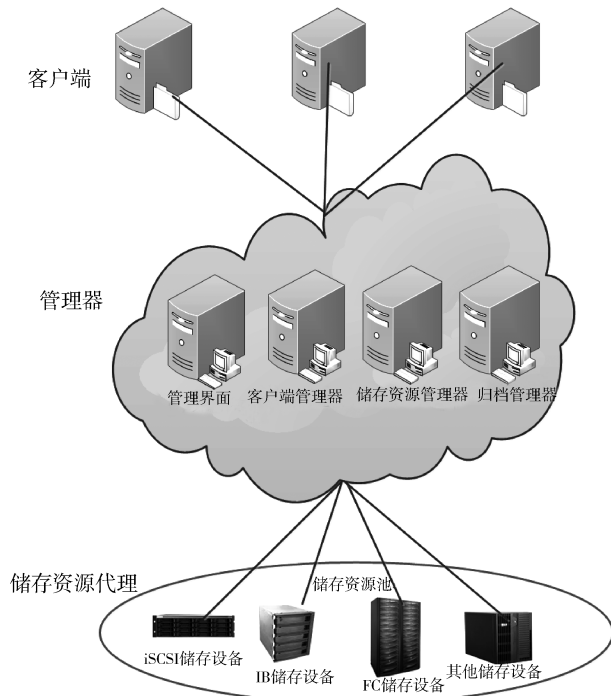


图5 智能电网海量数据管理平台

Fig. 5 Big Data Management System for Smart Grid

扩容、迁移、删除和恢复等。

存储资源代理是存储资源的一个代理，用于管理存储资源的基本信息（如存储容量和存储性能等信息），所有的存储资源逻辑上构成一个存储资源池，存储资源池包括 iSCSI 存储设备、IB 存储设备、FC 存储设备、普通 PC 服务器和磁带库等，存储资源池中的这些设备可以动态地加入系统或从系统中移除，因此能方便地实现动态扩容。

客户端管理器、存储资源管理器、归档管理器和界面共同构成管理器，在物理上可运行在一台或者多台机器上，管理器是分级存储系统的核心，管理器的主要功能是屏蔽下层存储资源的异构性，为上层应用提供统一的接口，并实现基于上层访问频度和下层存储资源性能的分级存储管理。客户端管理器周期性地收集来自客户端的应用需求信息，如逻辑卷性能、逻辑卷访问热度、CPU 使用率、可用空间、总空间、可用空间利用率等，从而实现了基于访问频度的分级存储管理。

存储资源管理器周期性地收集来自存储资源代理的存储资源特征信息如存储容量和存储性能等，并虚拟化这些存储资源，形成一个逻辑上的虚拟存储单元供客户端访问，从而实现了基于存储资源性

能的分级存储管理。

归档管理器主要负责备份数据的管理，主要记录客户端与备份信息，如备份时间点等之间的映射关系。管理界面是管理者和分级存储系统之间的交互接口，它能向管理者实时呈现业务负载信息和存储资源信息等，管理者通过这些信息可以为指定应用服务制定相应的策略规则，并能人工启动数据管理操作如数据迁移、数据备份、数据删除等，以便有效地对系统进行管理。

2.5 大数据管理关键技术

针对智能电网中海量数据的分级存储系统，涉及许多技术，如数据增量扫描技术、数据分级策略和分级存储中的缓存技术等。

2.5.1 数据增量扫描技术

超大规模文件系统的进行全名字空间扫描时非常耗时的任务，比如对于一个 10 亿文件的系统，每秒大约扫描 5 千个文件，那么全部扫描大约需要 27 小时。但是对于大规模系统，其文件访问率是非常低的，通常不足 1%。因此，只需要将经常访问的文件放在热点区域，即可极大的提高文件系统访问效率。本文通过增量扫描技术周期性增量扫描系统元数据来获得文件信息与文件访问情况，包括被访问文件的统计数据（包括访问次数和文件大小）、总访问热度等。通过增量扫描，读取最近访问的文件，避免扫描整个文件系统。

2.5.2 数据分级策略

数据分级策略决定数据存储的位置，级别高的数据存储在快速存储设备之上，级别低的存储在慢速设备。数据分级的指标项包括数据访问时间、频率、规模、生命周期等多个属性。数据分级最大的困难是初始化，合理的初始化分级将会极大的减少数据在不同级别存储设备间的迁移。

本文采用静态初始化和动态更新的方式对数据进行分级。首先是静态初始化，对数据类型进行分类，每类数据记录其分级情况，数据在不同级别存储设备上最佳迁移周期。之后，将系统在运行过程中分为短期优化和长期规划两个阶段。对于热点数据将会导入内存缓存，然后按频率退出内存，并记录下整个数据迁移过程。最后，对整个类型的数据进行统计分析，发现其数据整个生命周期中活跃的阶段，然后对各阶段进行分级。后续数据进入系统，即可按历史规律对数据进行更新。

2.5.3 分级存储中的缓存技术

为了快速响应请求并承受大量用户的并发请求, 缓存技术就变得至关重要, 直接影响系统整体性能。本文的高速缓存采用 Redis/Memcached^[6] 实现。Redis/Memcached 将热点数据缓存在内存中, 采用 key/value 的方式保存数据。运行过程中会周期性的把最少访问的数据写入下一层级的存储。基于 Redis/Memcached 设计的告诉缓存系统能够有效减少下层存储系统数据读写的操作, 从而提高数据访问的速度, 优化数据的访问效率, 是对海量数据存储的强大补充。

3 智能电网大数据应用前景

目前, 大数据已受到广泛的关注和重视, 在互联网、电信、零售、政府、医疗等领域取得巨大成功。电力行业大数据发展相对较晚, 搭建大数据平台的目的是提供电网海量数据存储和处理的能力, 利用好数据资源是摆在电力行业面前的巨大挑战。

对于电网的大数据应用, 本文认为可以从三大角度, 四个方向去考虑(如图 6 所示)。首先, 从公司自身出发, 针对服务对象和第三方机构这三个角度出发分析。比如, 公司内部, 可以考虑资产优化、管理优化、经营优化等优化配置, 利用大数据的数据整合和大规模系统分析能力, 将条块划分的各个业务模块整合起来, 形成有序优化的整体性电网运营模式; 对电力用户, 可以考虑节能、账单优化、个性化服务以及交互模式创新。利用掌握的用户群体性数据, 分析用户的用电习惯, 给出优化用

电方案, 提高用户满意度水平, 加速电费资金回笼; 对发电企业, 可以考虑发电计划优化, 新能源接入优化, 直购电过网优化和交互模式创新。利用大数据知识发现、数据挖掘与智能控制技术优化电厂发电计划; 此外, 还可以考虑提供社会化服务, 比如规划数据、设计数据, 社会用电数据等。进而创造出新的盈利模式。

4 结语

本文分析了智能电网的海量数据分级存储的需求特征, 并以此为基础, 提出了智能电网海量数据分级存储系统的架构以及总体设计, 重点说明了几项海量数据分级存储管理的关键技术, 包括数据增量扫描技术、数据分级策略和分级存储中的缓存技术等。最后, 分析电网大数据潜在的应用场景, 为大数据技术落地探索前进的方向。

参考文献:

- [1] 陈树勇, 宋书芳, 李兰欣, 等. 智能电网技术综述 [J]. 电网技术, 2009, 33(8): 1-7.
CHEN Shuyong, SONG Shufang, LI Lanxin, et al. Survey on Smart Grid Technology [J]. Power System Technology, 2009, 33(8): 1-7.
- [2] 丁杰, 奚后玮, 韩海韵, 等. 面向智能电网的数据密集型云存储策略 [J]. 电力系统自动化, 2012, 36(12): 66-70.
DING Jie, XI Houwei, HAN Haiyun, et al. Data Dense Cloud Persistent Storage Strategy for Smart Grid [J]. Automation of Electric Power Systems, 2012, 36(12): 66-70.
- [3] 张海梅. 大规模智能电网中数据存储方法的研究 [D]. 沈阳: 辽宁大学, 2012.
- [4] 李乃湖, 倪以信, 孙舒捷, 等. 智能电网及其关键技术综述 [J]. 南方电网技术, 2010, 4(3): 1-7.
LI Naihu, NI Yixin, SUN Shujie, et al. Survey on Smart Grid and Relevant Key Technologies [J]. Southern Power System Tehnology. , 2010, 4(3): 1-7.
- [5] 张榆, 马友忠, 孟小峰. 一种基于 HBase 的高效空间关键字查询策略 [J]. 小型微型计算机系统, 2012, 33(10): 2141-2146.
ZHANG Yu, MA Youzhong, MENG Xiaofeng. Efficient Processing of Spatial Keyword Queries on HBase [J]. Journal of Chinese Computer Systems. 2012, 33(10): 2141-2146.
- [6] XU Chuncong, HUANG Xiaomeng, WU Nuo, et al. Using Memcached to Promote Read Throughput in Massive Small-File Storage System [C]//International Conference on Grid and Cooperative Computing (GCC), 2010: 22-29.

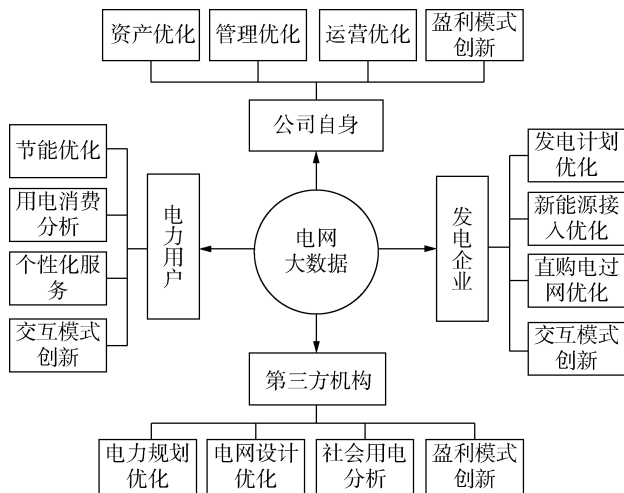


图 6 电网大数据应用前景

Fig. 6 The Prospect of Smart Grid Big Data Application

(责任编辑 黄肇和)