

文章编号:1001-9081(2011)06-1681-04

doi:10.3724/SP.J.1087.2011.01681

电力系统动态信息数据库关键技术

黄海峰¹, 张珂珩², 张 鸿¹, 季学纯¹, 陈 鹏¹

(1. 国网电力科学研究院 国电南瑞科技股份有限公司, 南京 210061; 2. 国网电力科学研究院 江苏瑞中数据股份有限公司, 南京 210003)
(huanghaifeng@vip.sina.com)

摘要:在介绍了基于时间序列的动态信息数据库结构的基础上,结合电力系统数据特点,分析了并发数据处理、内存映射文件、磁盘缓存、关联数据存储等构建动态信息数据库的关键技术,并着重研究了数据采集流程和混合压缩算法。经实际应用案例表明,采用该关键技术构建的动态信息数据库满足了电力系统海量数据高速存储的要求。

关键词:动态信息数据库; 实时数据处理; 并发数据处理; 海量数据; 电力系统

中图分类号: TP311 **文献标志码:**A

Key technologies of dynamic information database for power systems

HUANG Hai-feng¹, ZHANG Ke-heng², ZHANG Hong¹, JI Xue-chun¹, CHEN Peng¹

(1. NARI Technology Development Company Limited, Power Grid Automation Research Institute, Nanjing Jiangsu 210061, China;
2. China Realtime Database Company Limited, Power Grid Automation Research Institute, Nanjing Jiangsu 210003, China)

Abstract: In the paper, on the basis of analyzing the structure of dynamic information database, and in combination with the feature of the power system, the key technologies of concurrency data processing, memory-mapped file, disk cache management mechanism and associated data storage were discussed, and the data sampling flow and hybrid compression algorithm were also introduced in detail. The application case in the automatic system of power grid dispatching was introduced and the result proves that the dynamic information database can meet the performance requirement of high-speed data processing.

Key words: dynamic information database; real-time data management; concurrency data processing; mass data; power system

0 引言

数据库技术在电力系统电网调度自动化领域已得到普遍应用, 利用数据库可保存开关状态等离散的遥信量, 以及表征设备电压、电流、频率等连续变化的遥测数据等。数据库中保存的历史数据将应用于电网的事故反演等功能, 供电网发生事故或扰动后研究分析使用^[1], 因此历史数据保存的完整性将是分析的关键。目前针对低频率采样周期数据采用商用关系型数据库可以满足需要。

随着电网规模的快速发展, 大电网的安全稳定运行需要更多的信息。担负电网调度任务的能量管理系统(Energy Management System, EMS)将面临着多达百万级的数据采集规模, 同时反映电网运行状态信息的实时运行数据已由过去秒级周期性采样, 发展为变化采样, 即按照实际时间序列变化连续存储, 数据存储规模也将从目前的GB级增加到TB级。

此外, 随着相量测量单元(Phasor Measurement Unit, PMU)采集装置的普及, 海量电力信息数据存储将日益重要。与常规远端测控单元(Remote Terminal Unit, RTU)相比, PMU采集的特点是采集频率高, 达到每秒25、50甚至100帧, 且对所有数据必须完整保存。因此在相同采集点的情况下, 数据存储规模将为常规采集量数百到上千倍。根据理论测算, 对

于25帧/秒采集频率的PMU装置, 存储1000个向量一年所需的存储容量约为9.3 TB。因此, 无论对写入速度还是查询效率而言, 采用常规关系数据库存储海量信息将很难满足应用需求。

动态信息数据库的出现使得上述应用实现成为可能。动态信息数据库是专门处理具有时间序列特性的数据库管理系统, 具有极高的I/O速度和查询检索效率。同时为了存储较长时间的历史数据, 采用了高效的压缩算法, 与关系数据库相比, 容量大大减少^[2]。典型实时数据库管理系统产品有美国OSIsoft公司PI(Plant Information)和美国Instep公司eDNA^[3-5]。

电力系统调度领域中数据测点采用双元表述方法, 测点采用实测值以及质量码关联表示。而通用的动态信息数据库只能支持单个测点的数值存储, 缺乏采集点相互关联性, 不支持测点的4字节数据质量码关联存储, 在实际应用中造成很大困难。

新应用发展迫切需要研究适合电力系统特点和实际需求的基于时间序列的动态信息处理技术, 以适应超大容量数据高效、快速存储和检索的需要, 以满足大电网静态、动态、暂态三位一体的信息处理与分析需求^[6-8]。目前研究主要集中采用合理的压缩算法提高数据存储的效率, 包括:利用小波包分

收稿日期:2010-11-13;修回日期:2011-02-22。 基金项目:国家科技支撑计划项目(008BAA13B06)。

作者简介:黄海峰(1969-),男,江苏苏州人,高级工程师,博士,主要研究方向:分布式对象、电力系统调度自动化基础软件平台; 张鸿(1980-),男,吉林通化人,工程师,主要研究方向:调度自动化、动态数据存储; 张珂珩(1973-),男,湖北武汉人,高级工程师,主要研究方向:数据库管理、数据存储; 季学纯(1977-),男,江苏江都人,工程师,主要研究方向:数据库管理、数据存储; 陈鹏(1979-),男,江苏徐州人,工程师,主要研究方向:数据库管理。

解方法去掉干扰数据并压缩数据^[9];采用小波双正交滤波器组的压缩算法^[10];提升小波和熵编码的数据压缩^[11];利用哈夫曼方法压缩数据^[12];利用在线无损压缩进行数据处理^[13]等。以上讨论均集中在压缩算法上,但均缺乏从动态信息数据库对数据管理的整体过程进行效率的分析^[14]。

因此,本文在分析电力系统对动态信息需求的基础上,着重讨论构建电力系统基于时间序列的动态信息库基础框架、解决动态数据高效存储和管理的关键技术以及实现方法,最后结合具体应用事例,说明了该技术在电力系统领域的发展和应用情况。

1 基于时间序列的动态信息数据库架构

电力系统动态信息数据库分为三层,如图 1 所示。底层为数据存储层,包括文件系统管理和磁盘缓存管理模块,负责对海量动态数据的存储;中间为数据处理层,包括网络通信和并行化处理、数据压缩、数据索引、数据解压、数据查询和数据统计模块,主要负责对需存储的数据进行压缩、建立索引等前期处理,及进行查询、统计、解压等后期处理;顶层是应用程序接口(Application Programming Interface, API)接口服务层,供二次开发。



图 1 动态信息数据库的系统架构

动态信息数据库系统实际应用中由数据处理服务器、命名服务器和数据访问客户端组成。

其中,数据处理服务器是动态信息数据库的核心,完成时序数据的压缩存储和查询功能;命名服务器则主要完成服务名转换和数据库授权控制功能。数据访问客户端由两部分组成:1) 动态信息数据库数据管理维护工具;2) 数据存储和查询的二次开发接口。

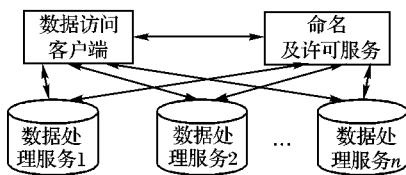


图 2 动态信息数据库的组成

数据处理服务器在系统启动时,向命名服务器注册其服务名称和服务地址(包括服务 IP 和端口号);客户端在访问数据处理服务器之前,先与命名服务器建立连接,查询取得需要访问的数据处理服务器的服务地址,在建立其与数据处理服务器之间的服务连接后,可以进行数据存储和访问。

2 关键技术

电力系统的动态信息数据库关键在于研究对基于时间序列的动态数据进行高效存储、查询和管理的方法。关键技术包括并发数据处理机制、内存映射文件方法和磁盘缓存机制、关联数据存储方法、适应电力系统的数据采集处理流程和混合压缩算法等。

2.1 并发数据处理机制

动态信息数据库高效运行的关键是提高处理客户端请求并与磁盘文件交互的效率。常规情况下,客户端通过网络链接向动态信息数据库系统提交数据,服务端的处理线程与之对应,由于多路多核 CPU 技术已普遍应用,计算处理能力很高,故物理磁盘读写速度将是整个系统的处理瓶颈。因此必须考虑较慢的磁盘 I/O 与高速数据吞吐间的平衡。数据处理主要包括数据写操作和数据读取操作。

为了达到高速处理的要求,采用线程池技术实现并行处理数据写入请求。通过使用多个 CPU(多核)并行处理来自一个 TCP 链接上的多个请求报文,即将一个 TCP 链接的报文处理任务并行化,从而达到高效处理的目标。并行化处理技术在写入数据时有效地提升服务端数据处理的速度,经过实际测试使用单个客户端(单线程)向服务端提交数据,数据处理的速度可以达到 300 万事件/秒。图 3 为并行处理写数据的过程。

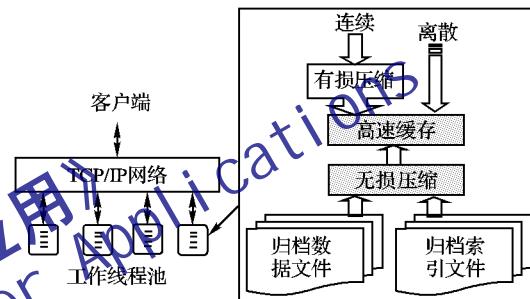


图 3 并行化处理写数据方法

同样,数据读取处理也采用线程池技术并行处理不同客户端的数据请求。系统设置高速缓存区,采用类似 ORACLE 数据库的“最近最少使用”(Least Recently Used, LRU)的策略,将最常用的数据保持在高速缓存中,在并发查询中实现数据共享,从而达到并发高效访问的目标。并行化处理技术在读取数据时有效地提升了服务端的响应速度,经过实际测试系统在 20 个并发查询时,每个客户端的读取速度大约为 10 万事件/秒。图 4 为并行化处理数据读取操作的方法。

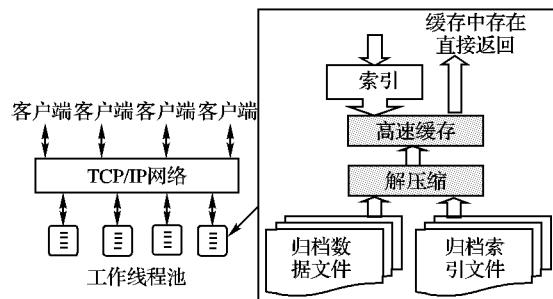


图 4 并行化处理读取数据操作的方法

2.2 内存映射文件方法和磁盘缓存机制

动态信息数据库系统管理着 TB 级别的海量数据文件,提高文件磁盘 I/O 处理速度也是提高动态海量信息数据库效率的关键。通过利用统一的磁盘缓存管理技术,采用分块内存映射文件方法,多个工作线程并发地将数据文件需要的部分进行映射,以达到高速存取数据的要求。

2.2.1 内存映射文件方法

内存映射文件方法与虚拟内存相似,通过内存映射文件

保留一个地址空间的区域,同时将物理存储器提交给此区域。内存文件映射的物理存储器来自一个已经存在于磁盘上的文件,即非系统的页文件。对该文件操作之前须首先对文件映射,类似于将整个文件从磁盘加载到内存。使用内存映射文件处理存储于磁盘上的文件时,无须对文件执行 I/O 操作和为文件申请并分配缓存,所有的文件缓存操作均由系统直接管理,这样节省了将文件数据加载到内存、数据从内存到文件的回写以及释放内存块等时间,因此,对于大数据量的文件处理的效率极高。

2.2.2 部分映射的缓存管理机制

由于需要加载的数据源的数据文件为 TB 级别,因此,无法一次性将海量数据文件全部映射为内存映射文件,而通过缓存管理采用部分映射的方法实现大容量数据文件的内存文件映射。

缓存管理采用固定块内存映射文件的方法,并发地将数据文件的各个部分依次映射到数据处理服务中的一个较小的地址空间,以达到高速存取数据的要求。

动态信息数据库系统通过索引确定需要的数据块,磁盘缓存管理负责将数据从物理文件加载到高速缓存区,并按照 LRU 策略对高速缓存区统一管理。高速缓存区分为数据、索引两类,并设置专门的读取缓存。部分映射缓存管理机制如图 5 所示。

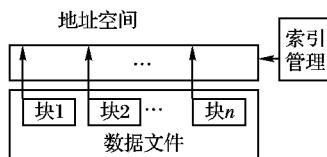


图 5 部分映射缓存管理机制

该机制在测试中,采用单台服务器启动两个数据处理服务,每个服务可达 300 万事件/秒的吞吐率。测试中仅使用无损压缩,压缩比例约为 3:1,按照一个事件 16 字节计算,此时磁盘吞吐量可以达到 30 MBps($2 \times 3000000 \times 16/3$)。

2.3 关联数据存储方法

根据电力系统数据的特点,需要存储的带时标历史数据(事件)由三个部分组成:时标、数据值以及质量码。符合电力系统特点的数据格式如下:时标 8 字节,数据值 4 字节,质量码 4 字节。其中时标由两个 4 字节整数(hours 和 usecs)表示, hours 表示的是当前时刻到公元元年的小时数,usecs 则表示当前时刻到上一个整点时刻的微秒数,时标的表示精度为 1 微秒;数据值由单精度浮点数(4 字节)表示;质量标志由一个 4 字节整数表示。因此,一个完整的事件点占用的存储空间为 16 字节。

2.4 适应电力系统的数据采集处理流程

动态海量数据库将数据点分为两类:离散数据点和连续数据点,并根据数据性质不同,采用不同的数据处理方式。

离散数据点指电力系统中的变化不连续的量,例如遥信量、设备运行状态等;对于离散数据点采用无损压缩处理,在查询时也只返回所存储的原始值,不进行插值处理。

连续数据点对应连续变化的量测量,如电压、电流、有功、相角等物理上连续变化的数据。动态海量数据库根据设置的

有损压缩允许误差,对历史数据序列进行有损压缩,压缩处理后的结果再经过无损压缩处理后存储在磁盘上。为了达到存储空间循环利用的目的,在存储压缩数据前,需要检查相应数据点历史数据的实际保留时间是否超过设定值,否则将删除该数据点最早的历史数据,然后再将压缩数据存储到磁盘文件中。对于连续数据点,数据处理流程如图 6 所示。

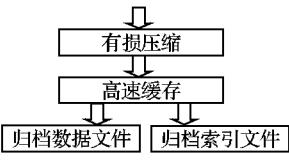


图 6 数据处理流程

2.5 混合压缩算法

动态信息数据库整合有损(线性带宽)和无损两类(哈夫曼)主流压缩算法,形成了混合压缩算法。具体而言:

1) 有损压缩。主要采用线性带宽压缩算法,压缩率为 8~10 倍。

2) 无损压缩。将浮点数按照 IEEE-754 的表示方法分解为符号位(1 位)、指数位(7 位)和小数位(23 位),对于小数位则进一步按照压缩存储精度要求将其截断,最后将这三部分分别采用 Huffman 压缩算法进行压缩;随机序列压缩率为 3~5 倍,对于质量位直接采用哈夫曼算法压缩处理;而连续量则采用混合压缩法,压缩率为有损压缩和无损压缩率的乘积,典型情况下压缩率为 25~30 倍。

混合压缩算法将有损压缩和无损压缩相结合,同时考虑到数据的特点,既保证了对动态信息的压缩效率,也提高了对数据处理的速度。

3 应用实例

实际运用中为了保证电网的安全稳定运行,需采集和存储包含 PMU 的动态数据、数据采集和监控系统(Supervisory Control and Data Acquisition, SCADA)稳态数据、保护和故障录波等暂态数据,简称三态数据。通过基于扰动案例的三态数据集中式存储管理功能,将电网中的事故和故障描述成为扰动的案例,以扰动信息为案例事件索引,将扰动相关的各种类型数据进行整合,为全面的案例分析提供完整的数据信息,从而为构建现代电力系统安全稳定协调防御体系提供基础数据。

三态数据中稳态数据记录了扰动事件在事故发生前后的总体信息,由于采样频率为分钟级,故采用关系型数据库可以满足要求。暂态数据刻画了扰动事件的细节,在事故发生的时间,由保护故障录波或 PMU 故障录波以每秒 4 800 帧或更高的时间尺度详细记录事故发生的瞬时信息。动态数据描述扰动事件的发展过程,将来自于 PMU 的动态数据,以每秒 100 帧的时间尺度记录扰动发生、发展的整个过程。对于动态数据需要通过时间序列数据库存储。

图 7 和图 8 记载的是真实的电网扰动案例及其对应的动态数据。通过动态信息数据库记录了 2008 年 11 月 9 日 0 点 20 分 24 秒发生的一次低频振荡扰动事件,这个事件持续 2 分 50 秒。根据 PMU 数据列表,选择指定厂站的指定设备的量测,根据起始时间进行 PMU 数据查询。

图 8 显示指定时间段从动态数据库中查询得到的 PMU

数据的曲线形式结果,可以清楚观察到本次低频振荡事件功率波动的全部过程。

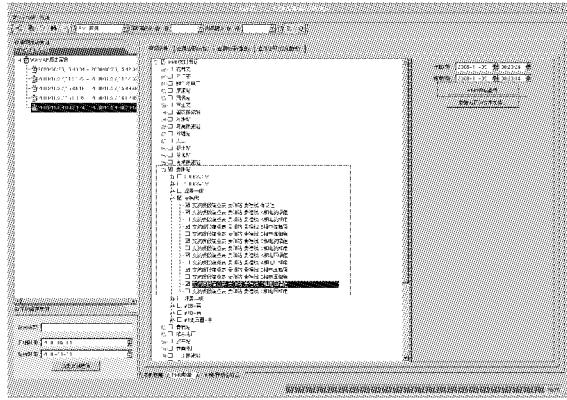


图 7 扰动案例选择

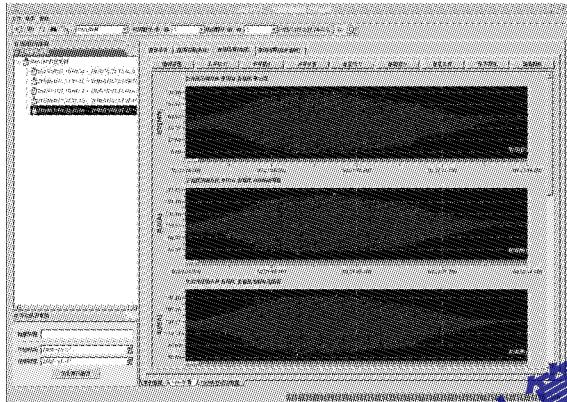


图 8 来自于动态信息数据库中查询的 PMU 曲线

4 结语

本文结合电力系统的具体应用需求,分析了基于时间序列的动态信息数据库。实践表明,动态信息数据库可以满足电力系统对高速、海量的数据存储和查询的需要,满足应用分析对电网动态数据的要求,从而为大电网安全稳定可靠的运行奠定基础。

(上接第 1674 页)

- [5] 刘怀军,车万翔,刘挺.中文语义角色标注的特征工程[J].中文信息学报,2007,21(1):79-84.
- [6] XUE N W, PALMER M. Calibrating features for semantic role labeling [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Barcelona, Spain: [s. n.], 2004: 88-94.
- [7] XUE N W, PALMER M. Automatic semantic role labeling for Chinese verbs [C]// IJCAI'05: Proceedings of the 19th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc., 2005: 1160-1165.
- [8] JIANG Z P, NG H T. Semantic role labeling of NomBank: A maximum entropy approach [C]// EMNLP'06: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2006: 138-145.
- [9] XUE N W. Semantic role labeling of nominalized predicates in Chinese [C]// HLT-NAACL'06 Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2006: 431-438.
- [10] 李军辉,周国栋,朱巧明,等.中文名词性谓词语义角色标注研究

参考文献:

- [1] 常乃超,兰州,甘德强,等.广域测量系统在电力系统分析及控制中的应用综述[J].电网技术,2005,29(10):47-52.
- [2] 张珂珩,季学纯,陈鹏.动态信息数据库在 EMS、WAMS 系统中的应用[J].电力系统自动化,2007,31(z):92-96.
- [3] OSisoft. PI server reference guide [EB/OL]. [2007-08-21]. <http://techsupport.osisoft.com>.
- [4] Instep Soft. White paper of eDNA [EB/OL]. [2007-08-23]. <http://support.instepsoftware.com>.
- [5] 蒙永萍,许勇,王俊永,等.基于 PI 数据库的实时动态监测系统的建设与应用[J].电力设备,2007,8(7):36-40.
- [6] 雷霆,黄太贵,袁林.动态信息数据库在调度自动化系统中的应用[J].电力系统自动化,2007,31(z):106-109.
- [7] 陆进军,高鑫,张力,等.配电网广域测量系统动态数据存储技术分析[C]//2007 年电力调度及场站自动化新技术交流会论文汇编.南京: [出版者不详], 2007: 186-189.
- [8] 林辉,余卓煌,刘晶.基于 SCADA 系统的面向海量数据的电力预测系统[J].计算机工程与设计,2010,31(5):1114-1117.
- [9] 刘志刚,钱清泉.基于多小波的电力系统故障暂态数据压缩方法[J].中国电机工程学报,2003,23(10):1-5.
- [10] 同常发,杨奇逊,刘万顺.基于提升格式的实时数据压缩和重构算法[J].中国电机工程学报,2005,25(9):6-10.
- [11] 周瑞,鲍文,于雷,等.基于提升小波和混合熵编码的数据压缩方法[J].电力系统自动化,2007,31(22):65-68.
- [12] HSIEH C T, HUANG S J. Disturbance data compression of a power system using the Huffman coding approach with wavelet transform enhancement[J]. IEE Proceedings Generation, Transmission & Distribution, 2003, 150(1): 7-14.
- [13] 齐文斌,李东平,杨东,等.广域测量系统数据在线无损压缩算法[J].电网技术,2008,32(8):86-90.
- [14] 余卓煌,林辉,刘晶.基于 SCADA 系统的面向海量数据的电力预测系统[J].计算机工程与设计,2010,31(5):32-36.

[J].软件学报,2011,22(8).

- [11] HACIOGLU K. Semantic role labeling using dependency trees [C]// COLING'04: Proceedings of the 20th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2004: 1273-1276.
- [12] XUE N W. Labeling Chinese predicates with semantic roles [J]. Computational Linguistics, 2008, 34(2): 225-255.
- [13] SUN WEIWEI, SUI ZHIFANG, WANG MENG, et al. Chinese semantic role labeling with shallow parsing [C]. EMNLP'09: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2009, 3: 1475-1483.
- [14] SUN W W, SUI Z. Chinese function tag labeling [C]// Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation. Hong Kong: [s. n.], 2009.
- [15] XUE N W. Annotating the predicate-argument structure of Chinese nominalizations [C]// Proceedings of the fifth International Conference on Language Resources Evaluation. Genoa, Italy: [s. n.], 2006: 1382-1387.