第 40 卷 第 19 期 2016 年 10 月 10 日
电力系统自动化
Automation of Electric Power Systems
Vol.40 No.19 Oct.10,2016
DOI:10.7500/AEPS20150928001

智能电网调度控制系统中历史数据服务优化方案

于宏文[1],郑春伟[2],汪 洋[2],汤卫东[1],马志斌[2],徐建航[2]
(1. 国家电网华中电力调控分中心,湖北省武汉市 430077;2. 南瑞集团公司(国网电力科学研究院)(北京),北京市 100192)

摘要: 在智能电网调度控制系统中,历史数据存储规模呈爆发式增长态势,海量历史数据的冲击,影响了数据访问效率,为满足历史数据高效访问的需求,针对历史数据服务提出了全新的解决方案。文中提出增加历史数据缓冲区,缓存管理电网热点数据,提高数据处理效率,减轻历史库负载;优化任务分配器,协调数据缓存区与关系库分工作业,对历史数据服务进行有效的管理和调度;设计历史数据会话监听机制,实时监视数据库会话状态,保障历史数据服务的安全性和稳定性。实践证明,基于缓存的历史数据服务能够满足爆发式海量数据场景下的数据访问需求,保障电网调度控制系统的安全稳定运行。

关键词: 数据缓存;服务管理;缓冲区;曲线模版;任务分配器;会话监视

0 引言

在智能电网调度控制系统(简称“D5000 系统”)中,母线的电压值、频率值,交流线段端点、变压器绕组、断路器等的有功功率值、无功功率值、电流值等大量设备的历史数据值,均由历史数据服务提供,再以曲线或表格的形式展现历史数据变化,方便调度员直观地监测电网运行状况。随着电网快速发展和规模不断扩大[1-3],电网调度控制系统记录更加全面的电力设备状态,导致历史数据存储规模持续增长,当前历史数据服务访问效率已经成为 D5000 系统快速发展的瓶颈[4-5],为适应电网数据爆发式增长的需求,研究安全、高效、稳定的新型历史数据服务迫在眉睫。

历史数据服务是 D5000 系统上层应用与底层关系库交互的中间件,它通过服务总线[6-7]接收上层应用发送的数据查询请求,连接关系库,通过 DCI (database call interface)原语从关系库海量历史数据中筛选目标数据,通过服务总线返回给上层应用,继续等待下次数据查询请求。当前历史数据服务主要存在以下问题:面对海量历史数据,历史数据服务缺少稳定高效的数据查询机制;D5000 系统的应用在访问历史数据时,缺少数据库会话状态的实时监视机制。

在 D5000 系统中,现有的历史数据服务很难解决上述所有问题,需要一种高效稳定的历史数据服务,满足海量历史数据的快速查询请求,降低关系库负载。因此,本文提出了一种基于缓存管理的历史数据服务优化方案,相对数据库中间件技术[8]对系统兼容性和可扩展性方向的优化,本方案更侧重于数据检索效率提升和数据库负载降低等方向的研究。新型历史数据服务在历史数据服务和关系库之间建立数据缓存区,缓存电网热点数据,减小历史数据服务对关系库的直接访问频率,并通过最近最少使用(LRU-2)算法[9-10]管理数据缓存区,提高数据缓存区命中率。建立历史库访问监听机制,监视历史库会话状态,对异常会话提供告警推送功能。通过对历史数据服务的优化,满足了 D5000 系统对历史数据服务的高效性和稳定性需求。

1 总体设计

历史数据服务由曲线模版、任务调度、数据访问 3 个模块组成,为 D5000 系统应用程序提供历史数据交互服务。其中曲线模版负责简化服务协议,任务调度负责任务分配与应答,数据访问负责与关系库数据交互。此架构的优点是服务协议简洁,客户端无需编写结构化查询语言(SQL)即可访问关系库。缺点是数据服务均需操作关系库,随着数据点量增加,业务量增加,会持续不断地增加数据库负载,从而降低历史数据服务效率。

针对历史数据服务的缺点,设计了历史数据缓存管理机制,提高服务查询效率,减轻关系库负载;

收稿日期:2015-09-28;修回日期:2016-03-16。
上网日期:2016-07-20。

设计了数据库会话监视机制，提高服务稳定性；优化服务调度，协调数据缓存区与关系库分工作业；并引用原服务曲线模板模块，保持服务协议不变，无需修改客户端源码。

优化后的历史数据服务由管理控制模块、数据缓存区模块、关系库交互模块3个部分构成。其中管理控制模块接收客户端请求，解析协议，分派任务、应答请求；数据缓冲区模块缓存热点数据，提供热点数据查询和修改功能，并通过LRU-2置换算法管理电网热点数据；关系库交互模块提供关系库数据交互接口，处理来自管理控制模块与缓冲区模块的数据请求，并监视关系库会话状态，提供会话异常告警推送功能[11]。优化后的历史数据服务结构如图1所示。

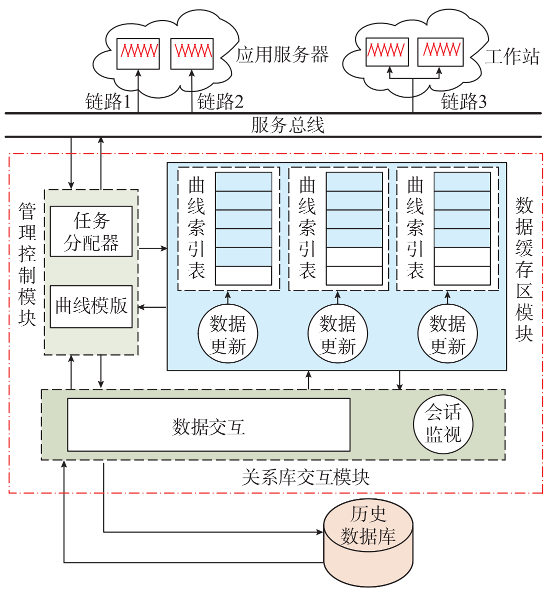

图1　历史数据服务结构图

Fig.1　Logical structure of historical data service

管理控制模块是历史数据服务的中枢模块，它由任务分配器和曲线模版组成。任务分配器负责任务接收、任务分派、服务统计、数据应答等功能；曲线模版负责协议编码、协议解码。

数据缓冲区模块是历史数据服务的技术核心，它由数据缓冲区、数据索引表、缓存数据同步、数据点置换算法组成。数据缓冲区缓存热点数据；数据索引表维护数据点信息，提供数据快速查询、更新功能；缓存数据同步模块实时监听缓存区数据点的归档信息，同步关系库和缓存区数据，确保数据一致性；数据点置换算法提供热点数据的置换策略。

关系库交互模块是历史数据服务中的关系库交互层，它由数据交互和会话监视组成。数据交互负责访问关系库；会话监视负责监视历史数据服务与关系库建立的会话状态，会话状态异常将推送会话异常告警。

2　关键技术

2.1　服务管理调度

历史数据服务提供曲线结构和表格结构的历史数据访问接口，结构化数据方便人机绘制曲线、表格，方便维护设备历史数据值。根据电网业务数据存储特点，抽象出曲线模版，简化服务协议，任务分配器根据曲线模版规则，逆向回溯出完整请求，并对请求任务进行分配、调度，以及应答请求。

2.1.1　曲线模版

历史数据源来自关系库，需通过标准SQL语句对关系库进行读取操作。为增强人机用户体验，简化用户操作，设计了曲线模版。

曲线模版是一组信息的集合，记录电网设备历史数据相关的存储信息，形成一种抽象的SQL模版协议，且每个模版对应一个唯一的曲线模版ID。通过曲线模版ID可定位其所属的曲线模版，根据SQL模版协议规则解析曲线模版内容，可生成标准SQL语句，使客户端传入曲线模版ID，可替代传入标准SQL语句，达到简化用户操作，简化服务协议的目的。

2.1.2　任务分配器

任务分配器负责历史数据服务的请求调度和服务统计功能。

任务分配器接收客户端请求，解析协议内容，获取曲线模版ID与设备点ID等信息，根据曲线模版ID得到曲线模版信息。解析、拆分、重组请求内容，将请求信息结构化封装。当请求是历史数据查询时，任务分配器询问数据缓冲区是否缓存了查询的数据，若已缓存数据，则将封装的请求信息传递给缓冲区，缓冲区将请求数据应答给任务分配器；若缓冲区未缓冲该设备点数据，任务分配器将封装的请求信息传递给数据交互，数据交互访问关系库将请求的数据应答给任务分配器，同时缓冲区利用LRU-2算法记录该设备点信息，按照算法规则决定是否将该设备点定为热点数据。最后，任务分配器将查询结果应答给客户端。当请求是历史数据更新时，任务分配器将封装的请求信息传递给数据交互，数据交互将历史数据更新至关系库。若更新关系库成功，任务分配器询问数据缓冲区是否缓存了该设备点的数据，若已缓存数据，则将封装的请求信息传递给缓冲区，缓冲区将更新数据同步至缓冲区，否则缓存区不做更新。最后，任务分配器获取执行结果并应答给客户端。

任务分配器以接口的形式提供请求类型统计、请求次数统计、缓冲区命中率统计以及数据库会话状态监视和缓冲区数据信息监视等功能,通过该服务统计功能,可使系统监控员掌握服务运行状况,提高服务可用性。

2.2　数据缓存控制

数据缓冲区模块由数据索引表、数据缓冲区、缓存数据同步机制和数据点置换算法组成。数据索引表中存储了 D5000 系统的部分数据点信息,通过 LRU-2 算法对数据索引表进行管理和维护,提高数据索引表中缓存热点数据检索的命中率;数据索引表中数据点对应的数据存储在相应的数据缓冲区中,通过缓存数据同步机制,实现数据索引表中数据点和关系库中对应数据点信息的同步操作,确保数据缓冲区和历史库中数据的实时一致性。

2.2.1　数据索引表

数据缓存区主要存储 D5000 系统热点数据,为增强缓存区中数据检索效率,优化数据管理机制,在缓存区中增加数据索引表的设计。

数据索引表是一种数据管理结构,对缓存区中数据建立数据关键字和数据详细信息之间的映射关系[12]。历史数据服务通过数据索引表进行缓存区数据访问,可快速定位缓存区中数据点地址,获取被检索数据的详细信息;通过 LRU-2 算法管理数据索引表中关键字信息的更替,进而控制缓存区中数据点信息的增加和删除,简化缓存区维护成本。

数据索引表的个数和单个数据索引表存储数据点的最大数目可自由配置,根据各地 D5000 系统中关系库数据点查询的频率和数量,配置适当大小的数据索引表,满足现场的实际需求。

2.2.2　数据缓冲区

根据数据索引表中可存储的最大数据点数,系统分配固定大小的内存空间作为数据缓冲区,用来存储数据索引表中数据点对应的具体数据。按照智能电网业务特点,电网曲线展示的数据源主要为历史库中遥测数据,昨日和今日曲线占绝大部分,所以在数据缓冲区中,每个数据点对应存储两天的数据,开始时间为前一天的零点,结束时间为当前时间,间隔为 1 min,即每个数据点在数据缓冲区中最多存储 2 880 个数据,数据类型为 float 型。数据缓冲区的总大小为 SIZE=INDEX_COUNT 乘以 DEV_COUNT 乘以 LEN,其中 SIZE 表示数据缓冲区所占字节数,INDEX_COUNT 表示索引表数量,DEV_COUNT 表示索引表中数据点个数,LEN 表示 float 类型数据的长度。

2.2.3　缓存数据同步机制

历史数据服务监听历史数据归档消息,对归档成功的缓存区中存在的数据点信息实行数据同步操作。

在 D5000 系统中,曲线展示的数据多为遥测数据,关系库中遥测数据以整数分钟为周期向历史库归档。由此,根据数据索引表中的数据点信息,基于“订阅–发布”的消息管理模式[13],定制化监听历史数据归档成功的触发式消息,对于归档成功的索引表中存在的数据点信息,实现数据由关系库向缓存区同步的操作,确保缓存区和关系库之间数据实时一致性。

2.2.4　数据点置换算法

为确保数据缓存区的高命中率,在有限的数据缓存区中,需有效利用缓存区空间,提高缓存区中热点数据占有率。基于计算机系统中广泛应用的局部性原理分析[14],D5000 系统中,之前被查询的曲线数据很可能在之后的时间内也会被查询,换言之,已经很久没有被查询的曲线数据很可能在未来较长的一段时间内不会被用到。LRU-*K* 算法可以很好地满足历史数据服务对缓冲区的管理要求,LRU-*K* 中的 *K* 代表最近使用的次数,本文中 *K* 值为 2。

LRU-2 算法根据数据的历史访问记录来进行淘汰数据,其核心思想是“如果数据最近被访问过,那么将来被访问的几率也更高”。LRU-2 算法需要维护两个队列,分别为“访问历史队列”和“数据缓存队列”。对于 LRU-2 算法在历史数据服务中应用的具体实现方式如图 2 所示。

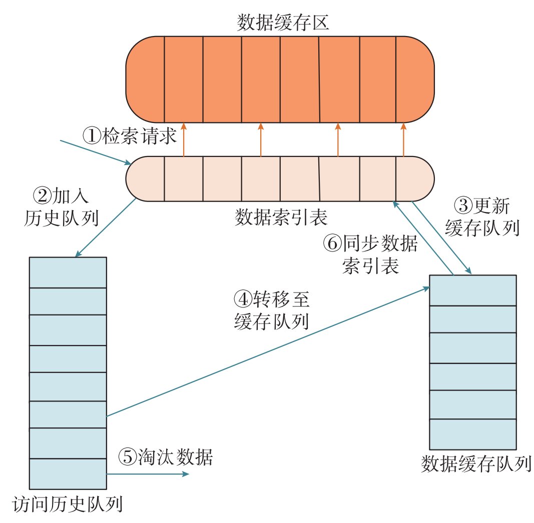

图 2　LRU-2 算法详解图
Fig.2　Detailed diagram of LRU-2 algorithm

图 2 中各序号含义如下。

①:数据点检索请求。

②:如果“数据索引表”中没有当前查询数据点的记录,将该数据点索引存入“访问历史队列”。

③：如果“数据索引表”中存在当前查询数据点的记录，按照访问时间更新“数据缓存队列”中数据点索引存储序列。

④：将“访问历史队列”中访问次数≥2 次的数据点移动至“数据缓存队列”，“数据缓存队列”重新按照时间排序，并将新插入的数据点索引同步至数据索引表中。

⑤：如果数据点在“访问历史队列”中没有达到 2 次访问，则按照 LRU 规则选择淘汰。

⑥：需要淘汰数据时，淘汰“数据缓存队列”中排在末尾的数据，即淘汰“倒数第 2 次访问离现在最久”的数据，同时将“数据索引表”中对应的该数据点索引删除，保持“数据索引表”和“数据缓存队列”的一致性。

2.3　数据交互控制

数据交互控制模块主要由数据交互和会话监视两部分组成，在同关系库进行数据交互的同时，实时监视自身会话状态，及时发现并处理异常会话，确保数据交互的安全性和稳定性。

2.3.1　数据交互

数据交互模块是历史数据服务的重要组成部分，其连通管理控制模块、数据缓存区模块和关系库，建立缓存区与关系库和数据管理模块与关系库之间的数据通道，完成历史数据服务对关系库的统一访问和关系库会话的统一管理，按需分配数据库连接资源，提升系统资源的有效利用率。

2.3.2　会话监视

在 D5000 系统中，历史数据服务同关系库之间的数据交互频繁进行，建立的数据库会话繁多，对于系统运行过程中出现的异常会话，缺乏及时有效的定位方法，严重影响 D5000 系统的稳定性和高效性。由此，增加会话监视模块，监听每条数据库会话的实时执行状态，通过告警服务推送并记录异常会话信息，及时准确地发现并处理异常会话，提高系统稳健性。

会话监视模块包含两个队列，会话等待队列和会话结束队列，分别记录正在执行的数据库会话信息和已经执行结束的数据库会话信息。会话监视模块以 5 s 为周期，循环遍历每条正在进行数据交互的数据库会话，即会话等待队列内所有会话，检查其执行时间，筛选访问关系库超时的会话，对其中的异常会话通过告警服务向 D5000 系统发送告警信息，并在关系库中永久保存。对于会话监视模块的详细监听机制如图 3 所示。

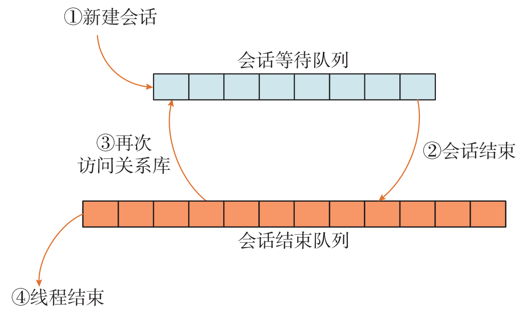

图 3　会话监视机制
Fig.3　Session monitoring mechanism

图 3 中各序号含义如下。

①：对于新建立的数据库会话，将其存入会话等待队列中，记录会话详情。

②：会话执行结束后，将会话信息由会话等待队列转存入会话结束队列。

③：对于结束的会话，如果再次访问关系库，则将其会话信息加载入会话等待队列，继续进行会话监听。

④：使用该会话的线程结束时，将会话信息从会话结束队列中移除。

3　性能测试

为了验证基于缓存历史数据服务的实际性能，按照华中网调现场配置搭建测试环境，在该测试环境中，分别运行优化前后的历史数据服务，对比测试关系库数据访问的效率。

历史数据服务对关系库的访问包括数据增加、删除、查询和修改等功能，以曲线查询为例，记录关系库数据查询在历史数据服务优化前后的性能变化。

关系库中昨日和今日分钟采样数据表记录数皆为 35 万条，通过历史数据服务对分钟采样数据表进行曲线查询操作，曲线查询开始时间为昨日零点整，结束时间为当前时间(12:00:00)，单次查询一条曲线，查询 10 次，每次查询耗时如图 4 所示。

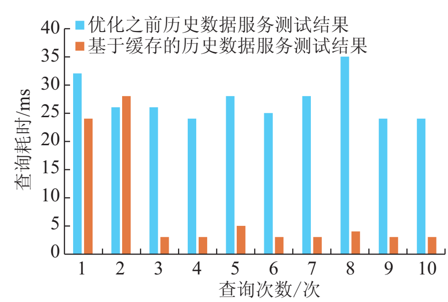

图 4　历史数据服务性能测试结果
Fig.4　Test results of historical data service performance

通过图4可以清晰地看出，在前2次曲线查询时，优化前后的历史数据服务在性能上无明显差别，但是自第3次开始，基于缓存的历史数据服务在数据检索性能上的优势远远高于优化之前的历史数据服务，数据检索速度提升10倍左右，大幅度提升了关系库数据检索性能。

设定曲线的刷新频率为100 ms，分钟采样数据表记录数为35万条，通过历史数据服务每隔100 ms对分钟采样表完成一次曲线查询操作，持续测试2 h，统计数据库平均占用的CPU和内存等资源比例，如图5所示。相比优化前的历史数据服务，优化之后数据库整体CPU使用率有明显下降，内存无明显变化，数据库运行压力在一定程度上得到缓解。

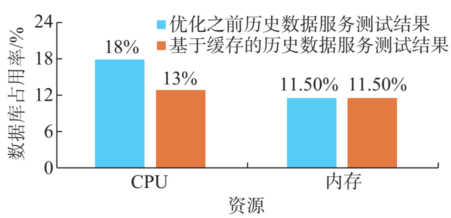

图5　历史数据库负载测试结果
Fig.5　Test results of historical database load

为验证基于缓存的曲线服务在处理海量历史数据时的性能，创建分钟采样表1和分钟采样表2，其中分钟采样表1记录数为35万条，分钟采样表2记录数为350万条。使用基于缓存的历史数据服务，分别对分钟采样表1和分钟采样表2进行同样的曲线查询操作，单次查询一条曲线，查询10次，每次查询耗时如图6所示。通过图6中的测试结果分析，在历史存储表数据的10倍增长态势下，基于缓存的历史数据服务在数据检索性能上无明显变化，可以满足爆发式海量数据场景下的数据访问需求。

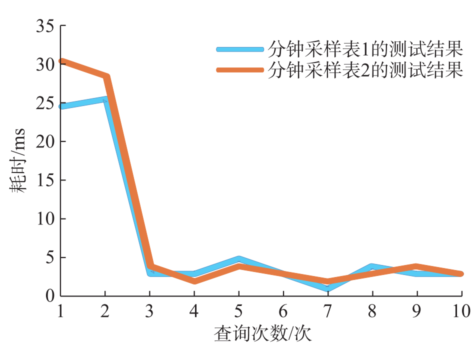

图6　爆发式海量数据性能测试结果
Fig.6　Test results of explosive mass data performance

基于缓存的历史数据服务已在华中电网工程现场部署运行，缓存管理机制与数据库会话监视机制的应用有助于大幅度提高历史数据查询效率，并减轻关系库负载，提高服务稳定性。

4　结语

本文提出的缓存管理方案，有效地提高了历史数据检索效率，减轻了数据库负载；提出的历史数据会话监听机制加强了历史数据服务预警能力，提高了历史数据服务的健壮性。通过性能测试与工程现场实践证明，基于缓存管理的历史数据服务可以适应电网的快速发展和规模连续的扩大，满足海量历史数据高效查询的需求，提升了系统的安全性和稳定性。

参考文献

[1] 刘纯，黄越辉，张楠，等.基于智能电网调度控制系统基础平台的新能源优化调度[J].电力系统自动化，2015，39(1)：159-163. DOI：10.7500/AEPS20141008018.
LIU Chun, HUANG Yuehui, ZHANG Nan, et al. Renewable energy dispatching based on smart grid dispatching and control system platform[J]. Automation of Electric Power Systems, 2015, 39(1): 159-163. DOI: 10.7500/AEPS20141008018.

[2] 马韬韬，郭创新，曹一家，等.电网智能调度自动化系统研究现状及发展趋势[J].电力系统自动化，2010，34(9)：7-11.
MA Taotao, GUO Chuangxin, CAO Yijia, et al. Current status and development trend of intelligent dispatching automation system of power system[J]. Automation of Electric Power Systems, 2010, 34(9): 7-11.

[3] 孙玉娇，周勤勇，申洪.未来中国输电网发展模式的分析与展望[J].电网技术，2013，37(7)：1929-1935.
SUN Yujiao, ZHOU Qinyong, SHEN Hong. Analysis and prospect on development patterns of China's power transmission network in future[J]. Power System Technology, 2013, 37(7): 1929-1935.

[4] 汪际峰，沈国荣.大电网调度智能化的若干关键技术问题[J].电力系统自动化，2012，36(1)：10-16.
WANG Jifeng, SHEN Guorong. Some key technical issues on intelligent power dispatching of bulk power grid[J]. Automation of Electric Power Systems, 2012, 36(1): 10-16.

[5] 辛耀中，石俊杰，周京阳，等.智能电网调度控制系统现状与技术展望[J].电力系统自动化，2015，39(1)：2-8. DOI：10.7500/AEPS20141008024.
XIN Yaozhong, SHI Junjie, ZHOU Jingyang, et al. Technology development trends of smart grid dispatching and control systems[J]. Automation of Electric Power Systems, 2015, 39(1): 2-8. DOI: 10.7500/AEPS20141008024.

[6] 王恒，辛耀中，尚学伟，等.智能电网调度控制系统数据总线技术[J].电力系统自动化，2015，39(1)：9-13. DOI：10.7500/AEPS20140930019.
WANG Heng, XIN Yaozhong, SHANG Xuewei, et al. A new data bus technology for smart grid dispatching control system[J]. Automation of Electric Power Systems, 2015, 39(1): 9-13.

DOI：10.7500/AEPS20140930019.

[7] 黄河清，宋晓华，张哲，等.一种面向多安全区的新型调度服务总线互联技术[J].电网技术，2011，35(11)：218-221.
HUANG Heqing，SONG Xiaohua，ZHANG Zhe，et al. A new dispatching interconnecting technology for multiple security zones based on service bus[J]. Power System Technology，2011，35(11)：218-221.

[8] 张海梁，袁荣湘，孙婉胜.数据库访问中间件技术在SCADA数据库系统中的应用[J].电网技术，2005，29(17)：58-62.
ZHANG Hailiang，YUAN Rongxiang，SUN Wansheng. Application of database access middleware technology in SCADA database system[J]. Power System Technology，2005，29(17)：58-62.

[9] 赵俊化，胡金霞.LRU页面置换算法的改进与实现[J].计算机工程，2012，38(17)：24-27.
ZHAO Junhua，HU Jinxia. Improvement and implementation of LRU page replacement algorithm[J]. Computer Engineering，2012，38(17)：24-27.

[10] 王涛，朱怡安，黄姝娟.基于改进LRU替换策略的共享Cache划分[J].微电子学与计算机，2012，29(1)：80-83.
WANG Tao，ZHU Yian，HUANG Shujuan. Shared Cache partition based on improved LRU replacement policy[J]. Microelectronics & Computer，2012，29(1)：80-83.

[11] 闪鑫，戴则梅，张哲，等.智能电网调度控制系统综合智能告警研究及应用[J].电力系统自动化，2015，39(1)：65-72.DOI：10.7500/AEPS20140930016.
SHAN Xin，DAI Zemei，ZHANG Zhe，et al. Research on and application of integrated smart alarm based on smart grid dispatching and control systems[J]. Automation of Electric Power Systems，2015，39(1)：65-72. DOI：10.7500/AEPS20140930016.

[12] 梅华威，米增强，吴广磊.基于MapReduce模型的间歇性能源海量数据处理技术[J].电力系统自动化，2014，38(15)：76-80.DOI：10.7500/AEPS20131215006.
MEI Huawei，MI Zengqiang，WU Guanglei. Massive data processing of intermittent energy based on MapReduce model[J]. Automation of Electric Power Systems，2014，38(15)：76-80. DOI：10.7500/AEPS20131215006.

[13] 潘国伟，宋玮，王相南，等.发布/订阅模式消息中间件在SCADA系统中的应用[J].电网技术，2008，32(18)：77-81.
PAN Guowei，SONG Wei，WANG Xiangnan，et al. Research on application of message oriented middleware based on publish/subscribe model in SCADA system[J]. Power System Technology，2008，32(18)：77-81.

[14] 侯昉，黄承慧，陆寄远.时间与空间局部性强弱的量化方法[J].微电子学与计算机，2014，31(6)：48-52.
HOU Fang，HUANG Chenghui，LU Jiyuan. Quantitative approaches to temporal and spatial locality strength[J]. Microelectronics & Computer，2014，31(6)：48-52.

于宏文(1967—)，男，硕士，高级工程师，主要研究方向：电力系统自动化。E-mail：yuhw@cc.sgcc.com.cn

郑春伟(1990—)，男，工程师，主要研究方向：电力系统自动化。E-mail：15134550125@163.com

汪　洋(1990—)，男，通信作者，工程师，主要研究方向：电力系统自动化。E-mail：wysea1990@163.com

（编辑　蔡静雯）

Historical Data Service Optimization Scheme for Smart Grid Dispatching and Control Systems

YU Hongwen[1], *ZHENG Chunwei*[2], *WANG Yang*[2], *TANG Weidong*[1], *MA Zhibin*[2], *XU Jianhang*[2]

(1. Central China Electric Power Dispatching and Control Sub-Center of State Grid, Wuhan 430077, China;
2. NARI Group Corporation (State Grid Electric Power Research Institute) (Beijing), Beijing 100192, China)

Abstract: In the smart grid dispatching and control system, the explosive growth of historical data storage scale and the onrush of massive historical data affect the efficiency of historical data access. To meet the demand for efficient access to historical data, an all-new solution for historical data service is proposed. The addition of historical data buffers is intended for buffer management of hot grid data to improve data processing efficiency and reduce historical database load. The task allocator is optimized for work coordination between data cache and historical database to effectively manage and schedule historical data services. The session monitoring mechanism of historical data is designed for real-time monitoring of database session state, to ensure the security and stability of historical data service. Practice has peroved that the cache based historical data service is able to meet the data access demand in the explosive mass data scenarios, while protecting the safety and stable operation of the power grid dispatching and control system.

Key words: data cache; service management; cache buffer; curve template; task allocator; session monitoring