

基于 MongoDB 的前置通信平台 大数据存储机制

史玉良¹, 王相伟², 梁波², 朱伟义², 吕梁¹

(1. 山东大学计算机科学与技术学院, 山东省济南市 250101;

2. 国网山东省电力公司, 山东省济南市 250001)

Big Data Storage Mechanism Based on MongoDB in Front Communication Platform

SHI Yuliang¹, WANG Xiangwei², LIANG Bo², ZHU Weiyi², LÜ Liang¹

(1. College of Computer Science and Technology, Shandong University, Ji'nan 250101, Shandong Province, China;

2. State Grid Shandong Electric Power Company, Ji'nan 250001, Shandong Province, China)

ABSTRACT: To deal with challenges of large electricity information data storage and processing, a cloud storage architecture mixing relational database with MongoDB database is built. According to demands of electricity data storage capacity and rate, large data storage mechanism for front communication platform based on MongoDB is presented, and key technology and optimization mechanism are analyzed in four aspects, i.e. data storage, availability, balancing distribution and read/write splitting. For shortcomings in MongoDB data equilibrium strategy and load difference between data nodes, equilibrium is optimized to realize dynamic equilibrium of data volume and load. Practical application confirms that big data storage mechanism based on MongoDB applied in front communication platform has good feasibility and high concurrent read and write performance.

KEY WORDS: power electric energy data acquire system; cloud storage; MongoDB

摘要: 为应对用电信息大数据存储和处理的挑战, 构建了关系数据库与 MongoDB 数据库混合存储的云存储架构。首先, 根据用电数据存储容量和速率的需求, 提出基于 MongoDB 的前置通信平台大数据存储机制, 并从数据存储、可用性、均衡策略和读写分离 4 个方面分析了其中的关键技术和优化机制。然后, 针对目前 MongoDB 数据均衡策略的不足, 结合数据节点负载的差异, 对均衡策略进行优化, 实现了数据量和负载的动态均衡。前置通信平台的实践应用分析验证了基于 MongoDB 的大数据存储机制具有高并发读写性能和良好的可行性。

关键词: 用电信息采集系统; 云存储; MongoDB

DOI: 10.13335/j.1000-3673.pst.2015.11.024

基金项目: 国家自然科学基金项目(61272241)。

Project Support by National Natural Science Foundation of China (61272241).

0 引言

电力数据的采集和存储是支撑智能电网的基础^[1], 随着电力用户数量和终端数量的增大, 用电数据成为典型的行业大数据^[2]。用电数据具备大数据的几个重要特征^[3], 具体表现在: 1) 数据规模巨大。以山东电网为例, 2014 年采集系统源码数据日增长量达到 900GB。2) 数据类型多样。用电数据包括档案数据、原始数据帧、业务数据以及统计分析数据等结构化、半结构化以及非结构化数据, 各类数据的存储需求存在较大差异。用电信息采集系统建设面临着大规模数据存储的挑战性问题, 即如何针对不同规模、类型、用途的数据选取合适的存储模型。

在大数据时代背景下, 根据 Brewer 的一致性、可用性和分区容错性 (consistency, availability, partition-tolerance, CAP) 理论^[4], 传统的并行数据库追求数据一致性和系统可用性, 限制了其扩展能力。大数据管理需求催生了以 NoSQL (非关系型数据库) 为代表的大数据存储技术, 并且凭借高效的读写效率、灵活支持复杂查询、负载均衡的自适应性等突出优势, 在多个领域发挥了积极作用。

在交通领域, 文献[5]基于 HBase 集群构建了交通流数据实时存储系统, 提高了多来源、高速率、体量大的交通流数据的存储性能。在医疗领域, 文献[6]设计实现了基于 MongoDB 的临床文档存储方案, 并应用于居民电子健康档案的存储和检索。在电力行业, 智能电网数据多样且复杂, 需要针对其特点选取合适的大数据存储技术^[7]。文献[8]提出一种基于分布式键值对存储 EMS 数据的

方法,并采用混合索引策略提升多维数据的查询性能。文献[9]提出了负荷分布式控制的云计算平台架构,量化多种可控负荷的云存储任务量。文献[10]将用户位置信息和用电信息存储于面向文档数据库中,并采用 MapReduce 实现了 2 种数据的融合。

NoSQL 技术以非关系型、分布式存储、可水平扩展为主要特点,一般分为:1) 基于键值对存储技术,如 Redis、Voldemort 等;2) 基于数据列分组存储技术,如 Cassandra、HBase 等;3) 基于文档存储技术,如 CouchDB、MongoDB 等;4) 基于图存储,如 Neo4J、InfoGrid 等^[11]。相较于其他 NoSQL 技术,基于文档存储技术不仅存储效率更高而且给数据库设计者极大的灵活性对数据进行建模。MongoDB^[12-14]作为典型的面向文档数据库,支持复杂数据类型、索引、自动分片、自动故障转移等功能,成为 NoSQL 数据库领导者之一。

针对用电大数据的存储容量、存储速率等问题,MongoDB 的自动分片机制^[15]增强了集群水平扩展能力,解决了基本的用电大数据的存储问题;高并发读写性能可对高速采集数据流进行实时缓存,进而解决数据流到达速度与生产库写入速度不匹配的问题;自动故障转移机制保证了平台的高可用性;模式自由的键值对文档存储结构方便存储模式灵活、时效性低且利用率相对较低的通信源帧、采集状况等数据;内置的分布式文件存储结构方便存储时效性高、利用率高的数据。

因此本文提出基于 MongoDB 的前置通信平台大数据存储机制,并对其中的若干关键技术进行讨论,对档案类数据、信道类数据等具有严格数据结构且基本无变化的数据,采用关系数据库存储,弥补 MongoDB 在关联查询、事务支持等方面的不足,同时针对目前 MongoDB 数据均衡策略^[16]的不足提出优化机制。

1 用电信息系统大数据存储架构

用电信息采集系统^[17]对电力用户的用电信息进行采集、处理和实时监控,实现用电信息的自动采集、计量异常和电能质量监测、用电分析和管理工作。系统整体逻辑架构见图 1。

为适应大规模用电信息数据的存储要求,本节提出面向用电信息大数据的存储架构,见图 2。从功能和存储数据种类角度,数据平台可划分为前置通信平台数据库、生产数据库、分析数据库。

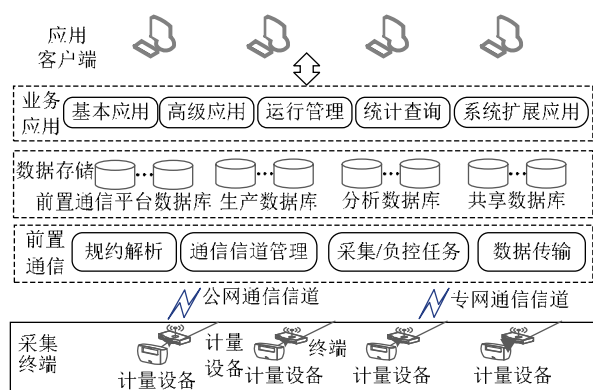


图 1 电力用户用电信息采集系统逻辑架构

Fig.1 Logical structure of power consumption information acquisition system

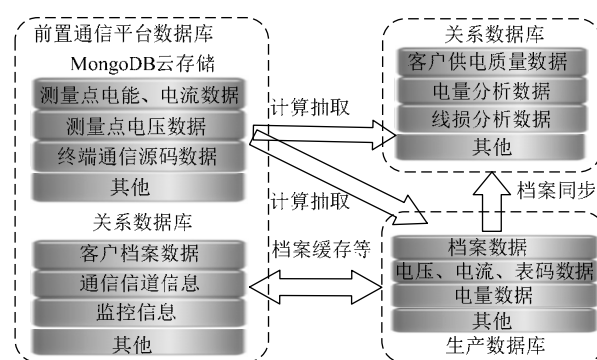


图 2 面向大数据的存储架构

Fig.2 Storage architecture for large data

2 基于 MongoDB 的前置通信平台大数据存储机制

2.1 采集数据存储过程

前置通信平台负责把二进制原始数据帧转换为 JSON(JavaScript object notation)格式^[18]的数据,并转换为业务数据存储至应用系统数据库,见图 3。

当数据采集服务和数据处理单元启动时,从前置通信平台关系数据库加载档案数据,用于数据帧解析和原始数据转存。此后每隔 15 min 同步更新一次。采集数据的存储过程为:1) 数据采集服务将终端上送的原始数据帧按其信道类型、规约进行解析得到原始数据,原始数据以文件形式存入本地磁盘缓存,并在平台配置库中写入文件相关的配置信息,原始数据帧上传至 MongoDB 私有云。2) 检测本地磁盘文件的数量、大小和最近一次写入时间,若文件大小达到设定阈值或文件在设定时间阈值内未更新,数据采集服务将文件以 GridFS 文件形式上传至 MongoDB 私有云,并更新平台配置库中的文件上传时间。3) 大数据管理引擎定时监测获取配置库内未进行数据转存计算的文件信息,并参照这部分文件信息,将文件从 MongoDB 私有云中提取,同时更新配置库内文件提取时间以保证同一

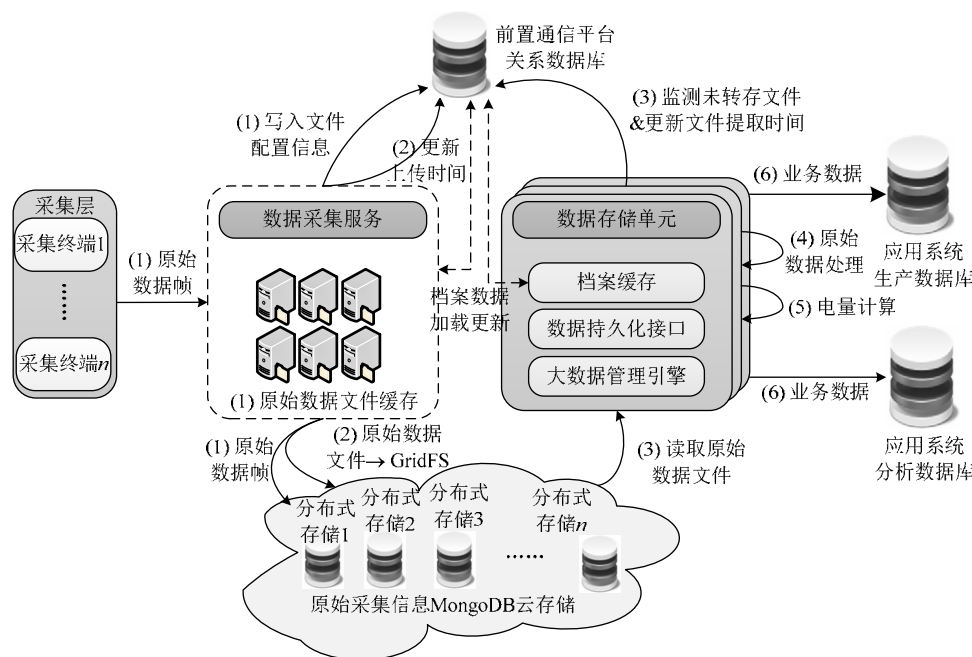


图3 前置通信平台采集数据存储

Fig.3 Acquisition information storage of front communication platform

文件只被1个数据存储单元处理存储,避免资源浪费。4)提取的文件数据转化为数据字符流,并进行相关的费率校验、数据校验和综合倍率的计算,得到业务数据。5)判断业务数据是否需计算电量,若需要则进行电量计算。6)数据持久化接口将业务数据提交给生产库持久化通道和分析库持久化通道,交由各自的持久化通道进行业务数据持久化。

数据采集服务首先对原始数据进行本地缓存,然后将文件上传至 MongoDB 私有云,而没有把原始数据直接以键值对形式写入 MongoDB 私有云,一方面可以有效保证原始数据的安全性,在很大程度上避免因网络中断及机房断电等异常造成数据丢失;另一方面,缓存后的文件写入 MongoDB,与键值对写入 MongoDB 相比,前者速度快得多,能更好地适应高速采集数据流的大数据处理。

2.2 MongoDB 私有云存储关键技术

2.2.1 存储模式

结合当前采集数据种类和 MongoDB 不同存储模式的特点,对采集数据采取不同的存储模式,确保大规模数据存储和查询效率。由于目前 MongoDB 仅提供数据库级别的全局锁,会对原始数据读写性能造成一定影响^[19],采用分库分表存储策略可在一定程度上缓解锁争用的问题,即把不同种类的数据表划分到不同的数据库下。

除分库分表策略外,为进一步提高 MongoDB 集群读写性能,采用多集群的部署模式,并根据各地市区划的终端数据量和地理位置划分区划为若

干组,以对应不同的集群。如图4所示,每个 MongoDB 集群由9台服务器组成,其中包含3个路由服务,3个配置服务以及3组数据分片,每组数据分片均由副本集构成。经大量测试发现,路由服务和配置服务对内存的依赖较小,因此将其置于数据分片所在的服务器上,从而充分利用云资源。

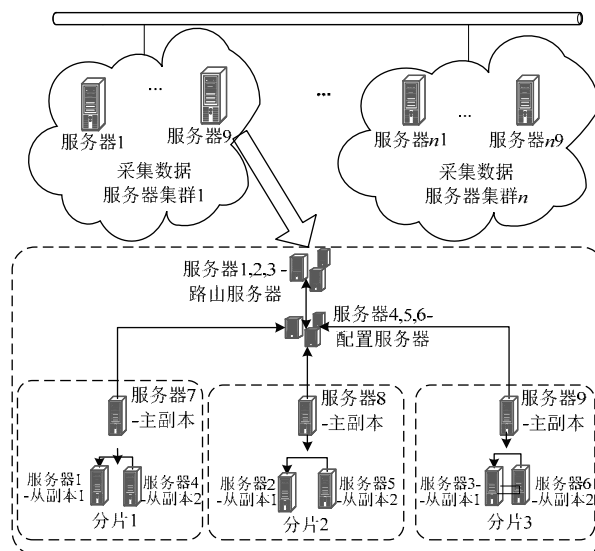


图4 MongoDB 私有云部署

Fig. 4 MongoDB private cloud storage

2.2.2 可用性

为保证云中数据的高可用性,实现自动故障转移,采用复制策略。每3个Mongo服务构成一个副本集,包含1个主副本和2个从副本,如图4所示,写入主副本调度的数据被异步同步到从副本,实现最终一致性。各副本间彼此互相进行“心跳”检测,

一旦某个节点失去了响应,在不需要人工干预的情况下,副本集就会迅速采取相应措施。

2.2.3 数据均衡策略

MongoDB 提供的自动分片机制为分布式存储的水平扩展提供了便利,均衡器在一定程度上保证了各分片数据量的均衡。MongoDB 将数据按用户指定的分片键进行范围划分,形成一系列 chunk(均衡器进行数据迁移的基本单位)。目前,均衡器检查各分片内的 chunk 数,若拥有 chunk 最多的分片和拥有 chunk 最少的分片的 chunk 数之差超过某个阈值(例如 8),均衡器将迁移前者的 chunk 至后者。该均衡策略仅考虑各分片内 chunk 数目,未能考虑分片所在数据节点负载的差异,无法从根本上解决云存储数据访问的动态均衡。为此,我们对均衡策略进行优化,通过路由 mongos 获取分片所在节点负载,并在迁移限制条件判断、迁移源分片与目标分片选取等环节加入负载因素的考虑,从数据量和负载 2 方面实现均衡。均衡策略算法具体如下。

输入:集合命名空间 ns ;分片约束信息映射表 $shardToLimits$;

分片分块映射表 $shardToChunks$;分片数 n ;

块均衡阈值 α ,负载选择比例 β ,负载均衡阈值 γ

输出:迁移源分片 $from$;迁移目标分片 to

1. $minList \leftarrow \emptyset, drainingShards \leftarrow \emptyset$;

/*确定集合中 chunk 数最大的分片, chunk 数小&负载最低的分片,正移除的分片*/

2. FOR each shard $s \in ns$ DO

3. $shardLimits \leftarrow shardToLimits.find(s)$ //获取 s 信息

4. $load \leftarrow getLoad(shardLimits)$ //获取所在节点负载

5. $sumLoad \leftarrow sumLoad + load$;

6. IF $s.load > maxLoad.load$ THEN

7. $maxLoad \leftarrow s$; //负载最高的分片

8. IF $!isSizeMaxed(shardLimits)$

AND $!isDraing(shardLimits)$ THEN

9. $minList.add(s)$; //加入候选 min 列表

10. IF $s.size > max.size$ THEN

11. $max \leftarrow s$; //chunk 数最大的分片

12. IF $isDraing(shardLimits)$ THEN

13. $drainingShards.add(s)$; //待移除的分片

14. $avgLoad \leftarrow sumLoad/n$

15. $minSortedList \leftarrow Sort(minList)$;

16. FOR each shard $s \in minSortedList$ & $s.rank < n \times \beta$ DO

17. IF $s.nodeload > min.nodeload$ THEN

18. $min \leftarrow s$; // chunk 数最小的分片

19. IF $max.size - min.size > n \times \alpha$ THEN

20. $from \leftarrow max; to \leftarrow min$;

21. ELSEIF $!drainingShards.isEmpty()$ THEN

22. $from \leftarrow drainingShards.get(); to \leftarrow min$;

23. ELSE IF $(maxLoad - avgLoad) / avgLoad > \gamma$ THEN

24. $from \leftarrow maxLoad; to \leftarrow min$;

25. ELSE return NULL;

上述算法分为 2 个阶段:阶段 1(第 1—18 行)收集集合所有分片的相关信息,确定负载最高的分片、chunk 数最大的分片、待移除的分片以及 chunk 数较小的分片中负载最低的分片;阶段 2(第 19—25 行)检查数据迁移条件是否满足,若满足则确定迁移源分片和目标分片。具体来说,先检查迁移条件 1:若分片内 chunk 之差超过阈值,则确定源分片为 chunk 数最大的分片;否则检查条件 2:若存在等待删除的分片,则选取一个带删除分片作为源分片。若条件 2 不满足,则检查条件 3:若存在过载的分片则确定源分片为负载最高的分片,否则无需迁移。

2.2.4 读写分离

通过云存储和副本集技术的应用,从副本能很好地适用于读扩展。路由服务将查询请求路由到从副本上执行,从而达到将“读写请求”按实际负载情况进行均衡的效果,增大了吞吐量。路由选择执行请求的从副本时采用响应速度均衡策略,对从副本发出探测请求,并将请求分发给响应时间最快的从副本,以较准确地反映节点的运行状态。

3 应用效用分析

3.1 基准测试

本节以零点采集终端上送数据为基准测试场景验证基于 MongoDB 的大数据存储机制的性能。

如图 5 所示,零点采集终端主动上送数据至前置通信平台,原始数据经本地缓存、上传至 MongoDB 私有云,大数据处理引擎读取数据文件并处理后,转存应用系统数据库。在此过程中,位于 MongoDB 私有云中的目录表记录终端采集状况,待终端不主动上数(或主动补招开始),前置通

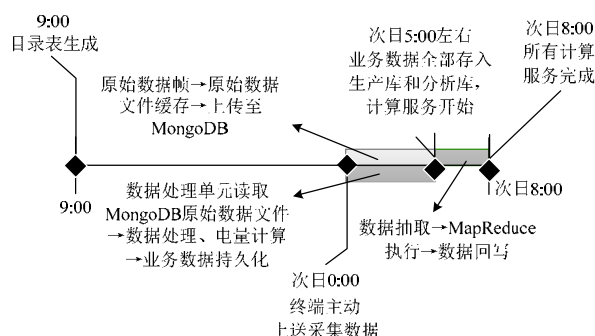


图 5 用电信息采集流程

Fig.5 Acquisition process of power information

信单元查询目录表中的未采集数据，并向相应终端下发命令进行数据采集。目录表应按照生产数据库中设置的任务提前生成并存储于 MongoDB 中。

MongoDB 集群搭建在 2*8 核 2.9GHz CPU、32GB 内存、200GB 硬盘的服务器集群上，MongoDB 采用分片模式，开启 3 个 Mongos 路由服务；生产数据库采用 Oracle RAC，搭建在一台 31*4 核 2.9GHz CPU、256GB 内存服务器上；Oracle 分析数据库搭建在一台 2*4 核 2.0GHz CPU、32GB 内存、4.5TB 硬盘的服务器上。

为模拟生产环境，测试数据均参照真实采集数据格式随机生成，并根据前期建设经验，模拟 3000 万个测量点在零点上送的采集数据量，包括高压、低压居民和低压非居民 3 种类型。每个测量点上送数据包括日冻结电能示值、电压曲线、电流曲线、功率曲线等多条采集数据，共计约 3.2 亿条。

3.2 采集数据存储性能分析

本文基于 MongoDB 优化的均衡策略实现用电信息数据的高并发、大容量存储，首先对数据表建立查询所需要的索引，提高查询的速度。MongoDB 没有 SQL 语言，无法使用 SQL 语言进行高效的查询操作，如连接查询等。但是由于 MongoDB 拥有高效的索引机制，可以自由为存储的数据集合添加想要的索引，一样完成高效的数据查询。在创建数据索引时，考虑到索引带来的系统维护成本，首先对数据的查询需求进行分析，然后建立查询所需要的合理索引，优化数据查询过程。在进行用电大数据读写时，路由 mongos 将分请求分发到具体节点，节点根据请求条件进行命名空间的遍历并判断所使用的键值是否有索引。如果有索引则根据索引的命名空间进行遍历查找，否则对整个命名空间进行遍历。数据检索完成后将数据缓存至内存，便于下次查询使用，同时将查询结果返回 mongos。

实验首先验证 MongoDB 集群的高并发读写性能，对 3000 万条基础数据量的 MongoDB 内单条数据查询，与同等条件下 Oracle 内单条数据查询时间进行对比，并测试 MongoDB 内置 GridFS 文件读写性能。写入采集文件数量为 3617 个，每个 10M 左右，每个文件中数据有 3 万条或 10 万条左右。表 1 显示了读写性能测试结果，可以看到基于 MongoDB 的私有云平台提供了良好的读写性能。

在上述实验的基础上，测试分别由 11 个数据处理单元并行处理 11 个区划的数据采集文件，并

表 1 MongoDB 集群读写性能
Tab.1 Read and write performance of MongoDB cluster

性能指标	实验结果
Oracle 千万级数据中单条数据查询	21.6ms
MongoDB 千万级数据中单条数据查询	4ms
采集数据文件写入 MongoDB	10~20MB/s
采集数据文件读出 MongoDB	40~60MB/s

将结果转存生产数据库和分析数据库。由图 6 可知，整个存储过程耗时共分为 2 个阶段：阶段 1 为从 MongoDB 私有云中读取表码文件、处理转存生产库的时间；阶段 2 包括加载计量点档案表缓存以及等待向生产库中存储的时间。3.2 亿条采集数据在 90min 内可快速存储至生产数据库和分析数据库。

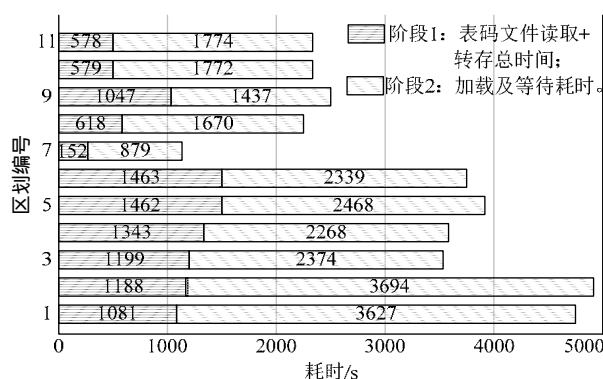


图 6 采集数据存储性能分析

Fig.6 Storage performance analysis of acquisition information

为测试数据处理单元生成目录表存入 MongoDB 私有云的性能，使用 3 个数据处理单元分别生成区划 1、2、3 的目录表，平均耗时见图 7。耗时主要包括 2 个阶段：阶段 1 为查询生产数据库中相关目录表生成任务；阶段 2 为生成目录表存储于 MongoDB。总耗时为 124min，与 Oracle 模式下的 300min 相比，目录表生成速率提升了 142%。

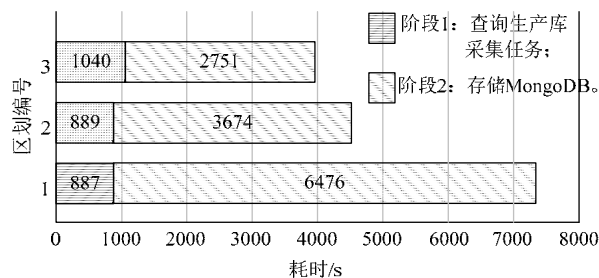


图 7 目录表生成性能分析

Fig.7 Analysis of directory table generation performance

4 结论

针对用电大数据存储需求，本文依托云存储技术，构建关系数据库与 MongoDB 数据库混合存储架构，提出前置通信平台大数据存储机制，并针对 MongoDB 现有负载均衡策略未考虑实际负载的不

足,提出改进的均衡策略。通过以零点采集终端上送数据为基准测试场景,验证了基于 MongoDB 的前置通信平台大数据存储性能较原有 Oracle 模式下的存储性能有大幅提升,具有可行性和高效性。

下一步工作主要是进一步提升系统数据存储效率:MongoDB 设计之初选择了 CAP 理论中的一致性和分区容忍性,需要对可用性进行优化。

参考文献

- [1] 宋亚奇,周国亮,朱永利.智能电网大数据处理技术现状与挑战[J].电网技术,2013,37(4):927-935.
Song Yaqi,Zhou Guoliang,Zhu Yongli.Present status and challenges of big data processing in smart grid[J].Power System Technology,2013,37(4):927-935(in Chinese).
- [2] 赵腾,张焰,张东霞.智能配电网大数据应用技术与前景分析[J].电网技术,2014,38(12):3305-3312.
Zhao Teng,Zhang Yan,Zhang Dongxia.Application technology of big data in smart distribution grid and its prospect analysis[J].Power System Technology,2014,38(12):3305-3312(in Chinese).
- [3] 马建光,姜巍.大数据的概念、特征及应用[J].国防科技,2013,34(2):10-17.
Ma Jianguang,Jiang Wei.The concept,characteristics and application of big data[J].National Defense Science & Technology,2013,34(2):10-17(in Chinese).
- [4] 林子雨,赖永炫,林琛,等.云数据库研究[J].软件学报,2012,23(5):1148-1166.
Lin Ziyu,Lai Yongxuan,Lin Chen,et al.Research on cloud database[J].Journal of Software,2012,23(5):1148-1166(in Chinese).
- [5] 陆婷,房俊,乔彦克.基于 HBase 的交通流数据实时存储系统[J].计算机应用,2015,35(1):103-107.
Lu Ting,Fang Jun,Qiao Yanke.HBase-based real-time storage system for traffic stream data[J].Journal of Computer Application,2015,35(1):103-107(in Chinese).
- [6] 刘愉,王立军.基于 MongoDB 的 EHR 存储方案研究与设计[J].中国数字医学,2013,8(6):20-24.
Liu Yu,Wang Lijun.Research and design of HER storage solution based on MongoDB[J].China Digital Medicine,2013,8(6):20-24(in Chinese).
- [7] 李芳,陈勇,张松树,等.大电网统一数据库建设相关技术研究[J].电网技术,2013,37(2):417-424.
Li Fang,Chen Yong,Zhang Songshu,et al.Research on construction of unified database of large-scale power grid[J].Power System Technology,2013,37(2):417-424(in Chinese).
- [8] 闫湖,周薇,李立新,等.基于分布式键值对存储技术的 EMS 数据库平台[J].电网技术,2012,36(9):162-167.
Yan Hu,Zhou Wei,Li Lixin,et al.Database platform of EMS system based on distributed key-value pair storage technology[J].Power System Technology,2012,36(9):162-167(in Chinese).
- [9] 刘萌,褚晓东,张文,等.负荷分布式控制的云计算平台构架设计[J].电网技术,2012,36(8):140-144.
Liu Meng,Chu Xiaodong,Zhang Wen,et al.Design of cloud computing architecture for distributed load control[J].Power System Technology,2012,36(8):140-144(in Chinese).
- [10] Ghidini G, Das S K.Improving home energy efficiency with E2Home: a web-based application for integrated electricity consumption and contextual information visualization [C]//Proceedings of the 3rd IEEE International Conference on Smart Grid Communications, China: SmartGridComm,2012:471-475.
- [11] 覃雄派,王会举,李芙蓉,等.数据管理技术的新格局[J].软件学报,2013,24(2):175-197.
Qin Xiongpai,Wang Huiju,Li Furong,et al.New landscape of data management technologies[J].Journal of Software,2013,24(2):175-197(in Chinese).
- [12] MiloradPantelijaStevic, BrankoMilosavljevic, BrankoRadePerisic. Enhancing the management of unstructured data in e-learning systems using MongoDB[J].Program:Electronic Library and Information Systems,2015,49(1),91-114.
- [13] Boicea A, Radulescu F, Agapin L. I. MongoDB vs Oracle [C]//Proceedings of the 3rd International Conference on Database Comparison Emerging Intelligent Data and Web Technologies. Bucharest: EIDWT,2013:330-335.
- [14] Liu Y, Wang Y. Research on the improvement of MongoDB auto-Sharding in cloud environment[C]//Proceedings of the 7th International Conference on Computer Science & Education. Melbourne: ICCSE,2012:851-854.
- [15] Banker K. MongoDB in action[M].New York: Manning Publications Co.,2011:184-217.
- [16] 金培权,郝行军,岳丽华.面向新型存储的大数据存储架构与核心算法综述[J].计算机工程与科学,2013,35(10):12-24.
Jin Peiquan,HaoXingjun,Yue Lihua.A survey on storage architectures and core algorithms for big data management on new storages[J].Computer Engineering & Science,2013,35(10):12-24(in Chinese).
- [17] 曹军威,万宇鑫,涂国煜,等.智能电网信息系统体系结构研究[J].计算机学报,2013,36(1):143-167.
Cao Junwei,Wan Yuxin,Tu Guoyu,et al.Information system architecture for smart grids[J].Chinese Journal of Computers,2013,36(1):143-167(in Chinese).
- [18] 高静,段会川.JSON 数据传输效率研究[J].计算机工程与设计,2011,32(7):2267-2270.
Gao Jing,DuanHuichuan.Research on data transmission efficiency of JSON[J].Computer Engineering and Design,2011,32(7):2267-2270(in Chinese).
- [19] Dede E, Govindaraju M, Gunter D, et al. Performance evaluation of a MongoDB and Hadoop platform for scientific data analysis [C]//Proceedings of the 4th ACM workshop on Scientific cloud computing. New York: Science Cloud,2013:13-20.



收稿日期:2015-07-06。

作者简介:

史玉良(1978),男,通信作者,博士,副教授,研究方向为服务计算、云计算、数据库,E-mail: shiyuliang@sdu.edu.cn;

王相伟(1971),男,高级工程师,研究方向为电力系统及其自动化;

梁波(1972),男,高级工程师,研究方向为电力系统及其自动化;

朱伟义(1971),男,高级工程师,研究方向为电力营销;

吕梁(1985),男,硕士,研究方向为大数据管理、云计算。

(责任编辑李兰欣)