

基于 TDengine 的智能电网监控系统数据存储方法研究

董雪^{1, 2}, 高远^{1, 2}, 敖炳^{1, 2}

(1. 南瑞集团(国网电力科学研究院)有限公司, 江苏 南京 2100061;

2. 国电南瑞南京控制系统有限公司, 江苏 南京 2100061)

摘要: 随着泛在电力物联网建设, 智能电网监控范围和接入设备规模急剧增加, 对数据存储提出了更高要求。针对智能电网设备建模需求, 介绍了智能电网监控系统中运用关系数据库和 TDengine 时序数据库在设备建模的设计细节和实现方法, 并详细描述了时序数据库接口设计的几个关键问题和实现方法, 最后验证了该方法的有效性。该方法已成功应用于多个智能电网变电站集控系统, 为历史反演和事故分析提供了有力的数据支撑。

关键词: 时序数据库; 时序数据建模; 提交缓存机制; 查询进程池

Research on data storage method of smart grid monitoring system based on TDengine

DONG Xue^{1, 2}, GAO Yuan^{1, 2}, AO Bing^{1, 2}

(1. NARI Group(State Grid Electric Power Research Institute) Co., Ltd., Nanjing 210061, China;

2. Nari-Tech Control System Co., Ltd., Nanjing 210061, China)

Abstract: With the construction of the ubiquitous power internet of things, the scope of smart grid monitoring and the scale of access equipment have increased dramatically, placing higher requirements on data storage. For smart grid equipment modeling requirements, the design details and implementation methods of using relational database and TDengine time series database sequence in smart grid monitoring system to model equipment are introduced, and several key issues and implementation methods of time series database interface design are described in detail. Finally, The effectiveness of this method is verified. This method has been successfully applied to multiple smart grid substation centralized control systems, providing powerful data support for historical inversion and accident analysis.

Key words: time series database; time series data model; submit cache mechanism; query process pool

0 引言

推进泛在电力物联网建设, 需要全方位对智能电网设备运行状态、用电信息等进行实时监测、分析, 数据采集点和采集的频率会大幅增加。智能电网数据采集及监控系统(SCADA)

作为电力物联网的一部分, 不仅需要实时监控, 也需要历史数据分析, 如历史 PDR 反演和事故分析。现阶段广泛使用的关系型数据库在海量数据存储、数据访问效率方面难以达到实时性要求。如何实现智能电网数据的快速存取, 为智能电网安全高效运行提供数据决策支撑,

将是智能电网建设和电力物联网建设的关键^[1]。

目前电力系统内的普遍做法是使用商业实时数据库软件,如 OSIsoft PI、eDNA 等作为时序数据库的存储和读取工具^[2]。随着大数据技术的发展,近年来也有一些基于 HBase、MongoDB 等大数据平台的研究和应用^[3-4]。

但这些实时数据库并非为电力系统设计,仅提供单个测点的数据存取接口,无法将设备模型数据与实时数据库中离散测点结合,难以支撑历史潮流分析、故障分析和 WAMS 等智能电网高级应用。

本文针对智能电网设备建模需求,提出将关系数据库中的电网设备模型与时序数据库结合,基于 TDengine 时序数据库的智能电网监控数据存储方法,并对时序数据库的存储模型、数据库状态监视和数据存取等接口进行详细设计。

1 TDengine 时序数据库特点

时序数据库全称为时间序列数据库,主要用于处理带时间标签(按照时间的顺序变化,即时间序列化)的数据,带时间标签的数据也称为时间序列数据^[5-6]。时序数据库一般都采用高效的压缩存储算法,如业界知名的 PI 数据库,采用了数据压缩和过滤技术,极大地节省了磁盘空间^[7]。

TDengine 数据库是一款面向物联网大数据的时序数据库,具有轻量化、高性能、高压缩、分布式和全栈时序数据处理引擎。与其他通用的时序数据库相比,有以下几个关键特点^[8]:

1) 多列模型。对于同一设备采集的多个量测,只要时标一致,就可以作为一张表的多个列存储在同一张表中。

2) 一个数据采集点一张表。一个数据采集点代表一个采集设备,如一台保护测控装置、一块电能表等。

3) 超级表。一个数据采集点一张表的建模方式导致表的数量巨大,引入超级表(STable)使表的管理和聚合操作变得简单。

TDengine 时序数据库采用关系数据库模型,使用者可以根据使用场景需要,创建相应的数据库和表。更为重要的是,TDengine 支持标准用户可以根据各自应用场景自建数据库和表结构,非常适合用于智能电网监控系统数据存储方案。

2 数据存储模型设计

在智能电网设备运行中,需要存储的数据包括状态量和模拟量。状态量反映设备的运行状态和告警状态,一般通过事件顺序(Sequence Of Event, SOE)记录可满足对历史数据的检索和分析需求,而电流、电压和功率等模拟量会随着设备运行时刻变化,具有明显的时序特性,需要进行时序存储。通过在选定的时间内查询 SOE 记录,可以将开关量和模拟量进行关联分析。

在智能电网监控系统中,已对断路器、隔离开关、变压器、负荷和联络线等电网设备进行建模,数据模型包括状态量和模拟量,其中模拟量包括电压、电流、有功、无功和频率等,不同的设备类型需要采集和存储的模拟量不同,如表 1 所示。

表 1 智能电网设备模型存储需求

Tab. 1 Storage requirements of smart grid equipment model

	I	U	P	Q	F
母联	√		√	√	
母线		√			√
变压器	√		√	√	
容抗器	√			√	
联络线	√		√	√	
负荷	√		√	√	
发电机	√		√	√	

2.1 时序数据建模方法

在电网运行过程中，模拟量的变化频率比较高，且同一个设备不同测点的变化频率不一致。面对秒级存储需求，有两种时序数据存储方法。

以单个测点对测点定义时序采样。按设备将每个设备的测量值属性分别定义到时序数据库中。这种存储方式数据定义、变化值提交和历史值查询都相对简单，但缺点同样明显，存储在时序数据库中的数据没有关联性，如需同时查询线路的有功和电流，则需要分两次查询，查询效率和性能较低，同时也不利于做数据关联性分析。

如图 1 所示，线路设备有电流、有功和无功三个量测属性，一个线路设备触发并在测点表派生三条量测记录，对应三个量测点。对三

个量测点分别定义时序数据采样，并在时序数据库中生成三个测点记录。三个测点记录在时序数据库中没有任何联系，无法做关联查询、聚合计算。

基于电网设备建模。在时序数据库中，将同一个设备的不同测点属性存储在一张表的不同域。这样只需要一次查询，即可得到一个设备的所有量测，查询效率和数据同步性都相对较高。

同样以线路为例，直接以线路设备为对象，定义时序采样，直接在时序数据库中保留了线路设备模型数据。

线路设备在时序数据库中的记录可定义为八元组： $\langle oid, timestamp, i_value, i_status, p_value, p_status, q_value, q_status \rangle$ ，如图 2 所示。

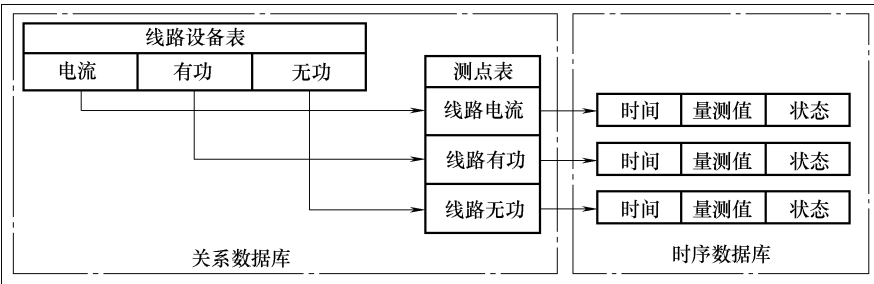


图 1 设备测点存储触发示意图

Fig. 1 Schematic diagram of storage trigger of equipment measuring point

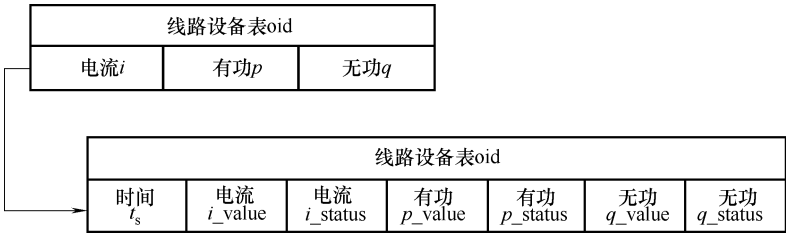


图 2 设备模型存储触发示意图

Fig. 2 Schematic diagram of device model storage trigger

图 2 中，oid 为设备在监控系统中的唯一对象编码； i_value 为电流值； i_status 为电流值的品质； p_value 为有功值； p_status 为有功值的品质； q_value 为无功值； q_status 为无功值的品质。

2.2 TDengine 时序数据库数据模型实现

基于 TDengine 建表特性，依次构建超级表和设备采集表^[6]。不同设备需采样存储的模拟量不同，因此需要为所有的设备建立不同的设备类超级表。

同样以线路为例，首先建立超级表：

```
create stable stbl_line
( ts timestamp,
  i_value float,
  i_status int,
  p_value float,
  p_status int,
  q_value float,
  q_status int)
tags( name binary(64))
```

上述建表语句中，stable 表示建立超级表；stbl_line 表示超级表表名，具体为线路设备类超级表；value 和 status 域是线路设备需要采集的量测；tags 表示线路设备类超级表的静态属性标签，可以根据系统需要自定义。

针对每条线路设备，建立设备采集表：

```
create table <oid> using stbl_line tags(“xxx
线”)
```

设备采集表通过 using 继承超级表的动态属性域，静态属性域 tags 在建表时指定。<oid> 为线路设备在关系库中的对象 id。

3 时序数据库接口设计

3.1 接口总体设计方案

时序数据库主要对电网运行过程中重要的数据进行时序采样和存储，较多的是模拟量（遥测量）的存储和查询。

数据库的接口设计需要重点实现以下功能：

1) 数据采样定义和修改。通过采样定义界面，对系统中的重要运行参数进行采样的定义和采样属性的修改。

2) 采样查询。时序数据库的重要功能是满足对上层应用对历史数据的查询需要，如画面曲线、报表、统计数据和 PDR 反演等查询功能。

3) 采样提交。接收变化数据，并根据时标把数据按时标存入数据库中。同时，采样提交功

能需考虑数据库短时间离线后的数据补录。

4) 状态监视等功能。监视数据序的运行状态、连接状态监视等功能，当数据库因某些原因离线时，通过系统的告警服务提醒运维人员尽快处理。

为保证历史采样数据的安全可靠，需考虑双数据库主备模式，这种主备为热备用。这就需要设计一种主备库的监视手段，同时监视两个数据库的连接状态，并从中选取一个作为主数据库，其他的进程从当前选定的主库读写。

如图 3 所示，监视进程定时循环测试时序库的连接情况，并从所有可用的数据库中依据优先级的高低选定一个作为主库，并把数据库状态及主备选定结果更新至共享内存。其他进程定时读取共享内存，对数据库进行读写操作。为保证主备库数据的同步性，数据的提交保存则不依据主备库，直接写入进程对应的数据库。

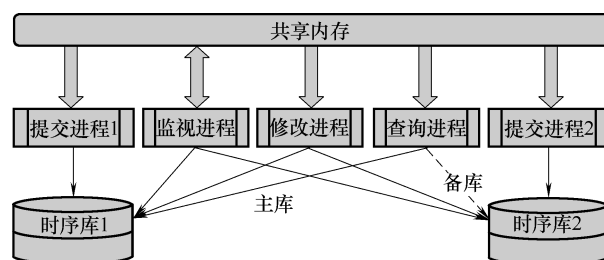


图 3 时序数据库接口设计方案

Fig. 3 Design scheme of sequential database interface

3.2 接口抽象

考虑后续有可能扩展兼容其他类型的时序数据库，有必要对时序数据库的接口进行抽象，应用层程序只需要调用对应的查询、提交和监视接口，不必关心具体的接口如何实现以及具体是哪种类型的时序数据库。

如图 4 所示，提交、查询等应用层进程调用时序数据库的抽象接口，不同类型的时序数据库依各自提供的 API，实现所有抽象接口。

3.3 数据关系映射

时序数据库虽然提供了历史数据的高速缓冲池，但本质上是一种纯文件系统的存储方

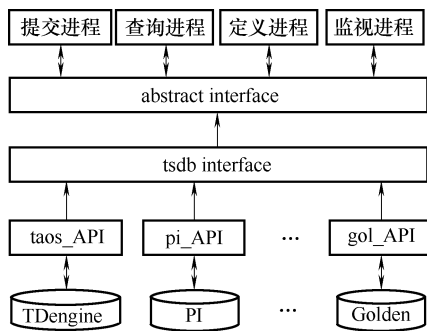


图4 时序数据库接口抽象示意图

Fig. 4 Schematic diagram of temporal database interface abstraction

式。关系型数据库如 Oracle 在存储、查询、分布式访问和管理上均好于纯文件系统。智能电网监控系统的设备模型参数、所有的量测定义等均保存于关系数据库，因此必须考虑同一量测在两套数据库之间的数据关系映射等配合问题。

电网监控系统中以 oid 标识一个设备对象，TDengine 时序数据库设备建一张数据表，故可用设备 oid 作为表名，在 TDengine 中唯一标识一个设备采集单元。通过在共享内存中创建 $\text{map}\langle \text{oid}, \text{oid_name} \rangle$ 映射，建立了两套数据库之间的数据映射关系。在关系数据库中建立一张采样定义表，记录已经定义的采样点的 oid 和时序库表名 oid_name， $\text{map}\langle \text{oid}, \text{oid_name} \rangle$ 。基于此表进行共享内存的动态维护，如图 5 所示。

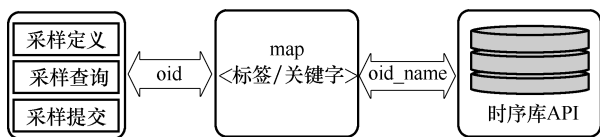


图5 标签/关键字映射

Fig. 5 Label/keyword mapping

3.4 时序库状态监视

根据总体功能设计，状态监视需要实现的主要功能有：

1) 读取主备库的基本参数，并写入共享内存。

2) 测试所有数据库的连接状态，并写入共享内存。

3) 根据配置的优先级，从所有可用的数据库中选定一个作为主库，并把结果写入共享内存。

4) 当任一个数据库发生离线/在线状态切换时，向系统的告警服务发送数据库状态切换告警消息。

5) 根据数据库的离线/在线切换，重新选定一个作为主库，并更新共享内存，向系统的告警服务发送主备切换告警消息。

3.5 数据提交

在智能电网监控系统中，模拟量以变化方式上送至系统前置预处理，最终由 SCADA 应用展示并存储^[5]。模拟量变化的频率取决于负荷变化率、间隔采样准确度和数据传送准确度等多种因素。在采样周期内未变化的数据以前向插值填充，形成完整的数据时间序列。

数据提交进程负责接收变化量测，并写入时序数据库。为保证变化数据采样存储时间有效性，提交进程采用提交缓存机制，即数据缓存队列和文件存储，同时用三个工作线程来实现二级缓存的数据存取，如图 6 所示。

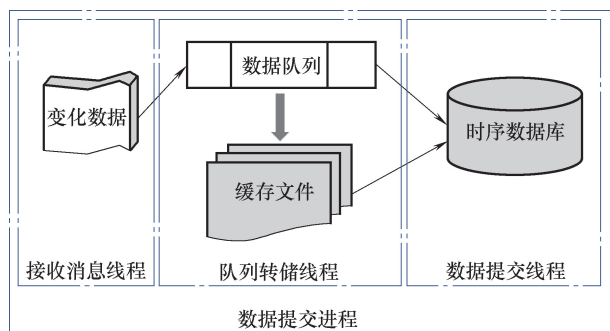


图6 数据提交进程

Fig. 6 Data submission process

1) 接收消息线程。从消息总线接收变化数据，并写入提交队列。

2) 队列转储线程。负责视提交队列，当提交队列的长度超过设定上限时，把队列中所有的数据写入本地磁盘。

3) 数据提交线程。负责数据的提交,如本地有转储文件,则优先把转储文件读入队列并优先提交,否则直接提交队列中的缓存数据。

3.6 数据查询

通过时序数据库的相应数据访问接口,可以快速高效地获取应用层所需要的数据,包括:①数据单点/多点时序查询;②数据单点/多点离散断面查询;③数据任意步长查询;④数据任意时间段任意步长查询。

数据查询过程中,经常会有多个上层服务请求历史查询,如画面、曲线、报表、历史反演 PDR 以及其他人工查询的方式。如何高效地处理多个应用同时查询并返回准确数据是必须要解决的问题。本设计中,通过查询服务进程池来解决此问题。进程池包括一个调度进程和 N 个查询进程,调度进程对进程池中其他查询服务进程进行调度管理,包括进程的选择、释放和进程状态的监视和管理等。查询进程池如图 7 所示。

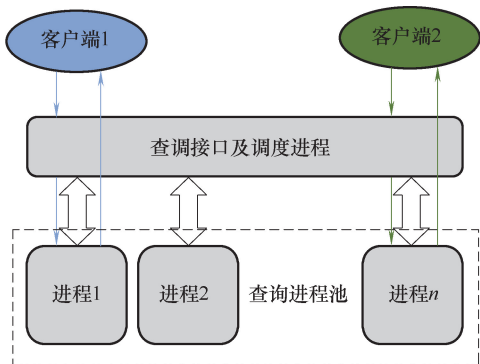


图 7 查询进程池

Fig. 7 Query process pool

调度进程将服务接口暴露给应用程序,当调度进程接收到一个应用的查询请求时,调度进程从当前为“空闲”状态的进程中选取一个进程为客户端进行查询服务,同时把此进程标记为“工作”状态,完成查询后再标记为“空闲状态”。其他同一时刻的查询请求则从其他为“空闲”状态的进程中选取。

3.7 数据可视化统计分析

对智能电网设备运行除了关注数据变化趋势外,还关注相关的统计值,如最大值、最小值和最新值等。

TDengine 提供类似标准 SQL 的查询语句,结合聚合函数,可实现对测点最大值、最小值、平均值和均方差等统计查询。除了通过 Shell 终端查询外,TDengine 还可通过 Grafana、Matlab 和 Python 等工具查询分析,通过可视化工具可以快捷地搭建可视化监控工具。

4 接口及性能测试

性能测试包括数据定义接口测试、数据提交性能测试和数据查询性能测试等,其中的重点在提交性能测试和查询性能测试。

测试环境:8 GB 内存,4 核 CPU,3.0 GHz,单线程。

4.1 数据提交性能测试

在时序库中定义 100 000 条记录的数据采样,使用模拟软件每秒模拟不同的变化数据提交。测试提交数据 5 s 一次,提交 100 次,记录平均耗时如图 8 所示。

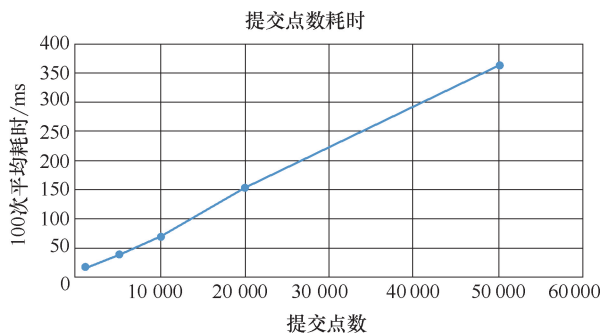


图 8 数据提交测试耗时

Fig. 8 Test time for data submission

从图 8 中得出,平均耗时和提交数据量保持线性增长关系。

4.2 数据查询性能测试

通过系统的历史查询工具,选取一定数量的历史采样数据进行时间段内采样点的查询,查询 100 次,得到每次查询的平均耗时,如表 2 所示。

表 2 数据查询耗时

Tab. 2 Data query time

序号	采样点个数	采样点记录数	平均耗时/ms
1	100	10 000	18
2	10	50 000	46
3	10	100 000	92
4	100	100 000	89
5	10	1 000 000	861

从表 2 结果看, 查询耗时和返回点的个数基本成线性关系, 同时时间跨度也在一定程度上影响了查询耗时。查询服务性能稳定, 在人机界面的查询操作方式下, 不影响使用体验。

5 结束语

基于 TDengine 时序数据库的智能电网监控系统, 解决了通用时序数据库无法对电网模型建模的问题, 并提供高密度数据的存储及查询的接口和性能, 为各类应用提供了准确的历史数据, 满足各类应用对数据存储实时性、海量性和安全性的要求, 为智能电网安全运行监控和事故分析提供了有力的数据支撑。EA

参考文献

- [1] 王远, 陶烨, 等. 一种基于 HBase 的智能电网时序大数据处理方法 [J]. 系统仿真学报, 2016 (3): 559-568.
- [2] 徐化岩, 初彦龙. 基于 influxDB 的工业时序数据库引擎设计 [J]. 计算机应用与软件, 2019 (9): 33-36.

- [3] 肖子达, 朱立谷, 冯东煜, 等. 分布式数据库聚合计算性能优化 [J]. 计算机应用, 2017, 37 (5): 1251-1256.
- [4] 王红涛, 王志超, 等. 基于时序数据库的工业大数据应用研究 [J]. 重型机械, 2020 (4): 17-21.
- [5] 戴则梅, 葛云鹏, 张珂珩, 等. 电网广域监测系统的数据库集成方案 [J]. 江苏电机工程, 2013 (1): 1-4.
- [6] 荣雪琴, 刘勇, 刘昊, 等. 基于时序数据库的电力运维系统关键技术研究 [J]. 电测与仪表, 2018 (9): 48-51.
- [7] 彭晖, 赵家庆, 等. 大型地区电网调度控制系统海量历史数据处理技术 [J]. 江苏电机工程, 2014 (9): 11-14.
- [8] TDengine 白皮书 [EB/OL]. 北京涛思数据科技有限公司, 2018.

作者简介



董 雪

董雪 (1987—) 男, 学士, 高级工程师, 研究方向为电力系统自动化、新能源发电控制。

(收稿日期: 2021-04-13)