

Research on Data Storage Design of Signal Centralized Monitoring Based on Real-time Database

Yu Yao

Beijing Gengdun Data Technology Co., Ltd., Beijing, 100085, China

Abstract

In the current railway signal centralized monitoring data storage systems generally use traditional relational database to store data, there are data throughput is low, not timely response, short storage time, slow query to review and issue such as difference of data correlation analysis ability, lead to important signal monitoring data storage capacity and relational database technology can not match the current rapid development of railway system, unable to correctly perform field intelligent diagnosis and analysis. Signal centralized monitoring data is based on the time continuous switch or analog, and has a large amount of data and data types, strong timeliness, change intervals are more big characteristics of real-time database with high concurrency, high polymerization, high availability, high security, high throughput and distributed characteristics, suitable for processing huge amounts of continuous real-time data change, the design of signal centralized monitoring data storage based on real-time database can clean and write the data of signal centralized monitoring system, so as to realize the rapid storage and reading of data with large capacity, and can better complete the data analysis and summary work. This paper analyzes the application of real-time database in railway signal centralized monitoring data storage design in detail, which has been proved to be feasible and reliable by practice.

Keywords

centralized signal monitoring; data storage design; real-time database

基于实时数据库的信号集中监测数据存储设计探究

姚羽

北京庚顿数据科技有限公司, 中国·北京 100085

摘要

当前在铁路中, 信号集中监测数据存储系统一般使用传统关系型数据库存储数据, 主要存在数据吞吐量低、响应不及时、存储时间短、查询调阅速度慢以及数据关联分析能力差等问题, 导致重要信号监测数据存储能力与关系型数据库技术无法匹配当前高速发展的铁路系统, 无法正确做出现场智能诊断与分析。信号集中监测数据是基于时间的连续的开关量或模拟量, 具有数据量大、数据类型多、时效性强、变化间隔相差大等特点, 实时数据库具有高并发、高聚合、高可用、高安全、高吞吐以及分布式等特点, 适用于处理海量的连续实时变化的数据, 基于实时数据库的信号集中监测数据存储设计能够针对信号集中监测系统的数据进行清洗和写入, 从而实现数据的大容量快速存储与读取, 能够更好地完成数据分析汇总工作。论文详细分析实时数据库在铁路信号集中监测数据存储设计中的应用, 经实践证明具有较高的可行性与可靠性。

关键词

信号集中监测; 数据存储设计; 实时数据库

1 实时数据库

1.1 概念

实时数据库是用于处理不断更新的快速变化的时序数据及具有时间限制的事务处理, 实时数据库技术是实时系统和数据库技术相结合的产物。实时数据库最初是基于先进控制和优化控制而出现的, 对数据的实时性要求比较高, 因而实时、高效、稳定是实时数据库关键的指标。实时数据库将时间视作一等公民, 运用各种先进技术和优化架构设计, 通过提高效率来处理大规模时序数据, 并带来性能

的提升, 包括: 更高的容纳率、更快的大规模查询以及更好的数据压缩。

基于实时数据库所开展的信号集中监测系统能够更好的实现大量数据长期且有效存储与处理。通过计算我们可以发现, 铁路的一个中等站每天能够产生 300 万条以上的数据, 虽然传统关系型数据库 (oracle 理论单表数据) 能够记录的数据条数没有上限设置, 但是官方推荐单表数据一般不得超过 500 万条记录, 一旦记录条数超过 1 亿条, 数据查询效率便会快速下降, 由此可见基于实时数据库建立信号集中监测系统的现实意义^[1]。

1.2 实时数据库的特点及应用优势

实时数据库的特点及应用优势主要体现在数据写入、数据读取及数据存储三个方面。

【作者简介】姚羽 (1986-), 男, 中国湖北荆门人, 本科, 高级工程师, 从事实时数据库管理系统研究。

首先是数据写入，能够按照指定时间间隔完成数据不间断的写入，且能够实现实时写入与高并发写入，写入后无需更新或删除；其次是数据读取，要求具备实时聚合，虽然读取只能在数据写入后进行，且次数远不及写入次数，但是能够进行多时间粒度与指定维度读取；最后是数据存储，不同于以往的行式数据库，利用列数据库进行数据存储，在读取时能够以列为单位，避免先读取行，再从其中读取需要的列，这种能够直接读取指定维度列的方式，大大提高了数据检索效率。

庚顿实时数据库管理系统（Golden Real-time Database Management System，简称庚顿实时数据库）是北京庚顿数据科技有限公司推出的具有自主知识产权的工业基础软件产品。其具有高可靠、高可用、高安全、跨平台、可伸缩的特性，可实现对海量时序数据的采集、存储、计算、传输、展现、分析等各个环节的高效管理，特别适用于轨道交通、智能电网、智慧城市、物联网等领域。

2 信号集中监测数据存储设计

2.1 概念

信号集中监测系统利用现场总线、传感器技术、计算机网络通讯、数据库及软件工程等技术，通过监测并记录信号设备的主要运行状态，为电务部门掌握设备的当前状态和进行事故分析提供科学依据的系统。在铁路运营中，信号集中监测系统是加强信号设备综合管理、确保行车安全、监测铁路信号设备运用质量的重要行车设备。此外，该系统还具有较好的逻辑判断能力，一旦信号设备工作出现异常或偏离预定界限，能够第一时间报警，最大程度规避因设备故障导致列车无法安全前进与准点到达。由此可见，信号集中监测系统是现代化铁路装备的重要组成。

信号集中监测系统需要监测并存储的数据主要如下。

2.1.1 模拟量数据

- ①外电网输入相电压、线电压、电流、频率、相位角、功率；
- ②电源屏输入电压、电流、输出电压、电流；
- ③ 25Hz 电源输出电压、频率、相位角；
- ④电动转辙机道岔转换过程中转辙机动作功率、电流、动作时间、转换方向；
- ⑤道岔表示交、直流电压；
- ⑥电缆绝缘测试数据；
- ⑦电源对地漏泄电流检测数据；
- ⑧集中式有、无绝缘移频自动闭塞区间移频发送器发送电压、电流、载频、低频，区间移频接收器轨入（主轨、小轨）电压，轨出 1、轨出 2 电压、载频、低频，区间移频电缆模拟网络电缆侧发送电压、接收电压、发送电流；
- ⑨环境状态的模拟量温度、湿度、民用空调电压、电流、功率；

防灾系统与列控系统分界面处接口直流电压；

⑩站（场）间联线路直流电压、场间联系电压、自闭方向电路电压、区间监督电压。

2.1.2 开关量数据

- ①按钮状态、控制台表示状态、关键继电器状态等；
- ②列车信号主灯丝断丝状态；
- ③环境监控开关量。

另外，信号集中监测系统需要提供数据接入接口，可以接入多种具备自诊断功能的信号设备的状态信息和报警信息，信号设备主要包括：计算机联锁、列控中心、TDCS/CTC、智能电源屏、ZPW2000、有源应答器、道岔缺口等。

2.2 基于实时数据库的信号集中监测数据存储设计

庚顿实时数据库管理系统是一款非常适合于信号集中监测系统的数据库，基于此产品进行信号集中监测数据存储设计^[1]。设计原则主要有三点：一是广泛高效的数据接入；二是快速查询；三是实时分析展现，基于这些原则所开展的系统设计，其数据存储能够更加有效地对上层分析进行支持，并同时保证对底层数据吞吐量的监测，进而实现快速灵活查询与分析。

就当前的信号集中监测数据存储设计结构，我们主要是利用采集适配器 + 实时数据库 + 监控组态平台的方式完成数据汇总与分析展现，具体如下：

第一步，利用采集适配器采集数据，采集适配器采用插件的方式管理所有的采集服务，每个采集服务负责安全可靠的完成现场数据源的采集工作，数据源可以是自动化设备也可以第三方系统中的数据文件或关系数据库系统。

第二步，采集适配器将现场数据汇总写入实时数据库，实时数据库将对数据完成整理、整合、清洗、计算、存储等一系列工作，并将数据对上层应用进行转发。

第三步，利用监控组态平台构建一套美观、易用的数据展现界面，实现系统总貌、动态流程、设备状态、报警状态的显示，方便用户及时掌握设备的当前状态，提供各种 BI 图表、趋势曲线、报表、历史回放、事故追忆等多种分析手段，便于用户对监测数据进行可视化分析。

信号集中监测系统中的模拟量和开关量数据都是跟时间强相关的时间序列数据，即统一表示系统、过程或行为随时间变化的数据。本系统的时序数据主要有以下三个共同点：

①抵达的数据几乎总是作为新条目被记录（也有少量补写的历史数据）。

②数据通常按照时间顺序抵达（但受传输影响，部分数据延迟抵达是常见现象，甚至不是按顺序抵达）。

③时间是一个主坐标轴（既可以是规则的时间间隔，也可以是不规则的，通常是不规则的）。

这类数据的处理是伴随数据的抵达事件而触发进行的，处理过程必须满足严格的时间限制，并且处理方案需要周全

的考虑到数据纠错、数据延迟、数据无序等特例情况。

针对时序数据的上述特点，庚顿实时数据库在存储结构上做了以下设计。

2.2.1 数据组织方式

数据的最小管理单元为标签点，是物理世界中传感器的数字双胞胎，是一对一映射关系，在实时数据库中自动为其分配唯一 ID，作为访问与定位的主键^[3]。

对于传感器，这里指监测单一指标的最小单位传感器，有一些信息是标识它在物理世界唯一性，比如名字、位置、数据类型、量程上下限、工程单位等等，这些确定了它的存在唯一性；还有一些信息是它在运行中在某时刻观察物理世界时所产生的数字信号，这个数字信号具有时间唯一性，除了值 Value，还有时间戳 Timestamp、数据质量或状态 Quality。

那么，对于传感器的数字双胞胎标签点，它会记录上述两类信息。与存在唯一性相关的数据保存在标签点属性中，对于已知的物理世界、部署已知的传感器，这类数据几乎很少发生变化，将其定义为静态数据；与时间唯一性相关的数据，即 VTQ (Value+Timestamp+Quality 的简称)，随时间不断有新的产生，历史数据逐渐累积，故将其定义为动态数据。

2.2.2 数据存储方式

标签点属性数据，即静态数据，采用行存储，通过 ID

可以一次定位静态数据，可以指定常用的数据字段加载到内存以提高查询速度；

时序数据，即动态数据，采用列存储，通过 ID + Timestamp 的二级索引可以快速定位时序数据，可以将近期数据的索引加载到内存以提高查询速度^[4]。

2.2.3 数据模型

一个数据库实例包含多张表，每张表里面包含多个标签点，每个点包含多个属性列，如表 1 所示。

标签点的值、时间戳、质量码 3 列的数据在数据库专门开辟的空间中按列存储，上面的二维表格只是当前快照的一个切面，可以想象其后面有众多历史切面构成一个三维立体的数据空间。

2.3 测试结果

针对信号集中监测系统的应用特点，我们对庚顿实时数据库进行了详细的测试，由于篇幅有限无法将测试报告尽附于此，现摘录部分性能测试内容如下：

测试环境：

服务端：48 核 CPU、128G 内存、10T 磁盘。

客户端：8 核 CPU、16G 内存、120G 磁盘。

2.3.1 急速写入数据

多并发给 100 万标签点急速写入实时数据，写入 60 批数据，总数据量 6000 万条数据（见表 2）。

表 1 数据模型表

标签点名	ID	描述	值	时间戳	质量	数值类型	工程单位	设备标签	设备位址
1#_I	1	南站 1# 道岔输入电流	2.1	2021-09-26 12:00:00	Good	FLOAT32	A	1#	南站
1#_U	2	南站 1# 道岔输入电压	385	2021-09-26 12:00:00	Good	FLOAT32	V	1#	南站
2#_I	4	南站 2# 道岔输入电流	2.2	2021-09-26 12:00:00	Good	FLOAT32	A	2#	南站
2#_U	5	南站 2# 道岔输入电压	384	2021-09-26 12:00:00	Good	FLOAT32	V	2#	南站
3#_I	6	南站 3# 道岔输入电流	1.9	2021-09-26 12:00:00	Good	FLOAT32	A	3#	南站
3#_U	7	南站 3# 道岔输入电压	383	2021-09-26 12:00:00	Good	FLOAT32	V	3#	南站

表 2 急速写入数据表

并发数	总耗时（秒）	吞吐量（万条/秒）
5	9.68	619.85
10	5.31	1129.94
20	4.34	1382.49
40	3.61	1662.05
80	3.14	1910.83

2.3.2 混合读写实时数据

50并发给100万测点持续不断写入实时数据(100万条/秒),不同并发(5、10、20、40、80)查询100万点实时数据(见表3)。

表3 混合读写实时数据表

并发数	总耗时(秒)	吞吐量(万条/秒)
5	0.211	473.93
10	0.172	581.39
20	0.123	813.01
40	0.089	1123.59
80	0.080	1250.00

2.3.3 历史数据统计分析

并发查询多点12小时(4.32万条历史数/点)、1天(8.64万条历史数/点)、7天(60.48万条历史数/点)、30天(259.2万条历史数/点)、365天(3153.6万条历史数/点)历史统计值,每个连接查询1个点的统计数据(见表4)。

3 结语

基于实时数据库的信号集中监测数据存储设计,能够实现同步数据导入和清洗,从而将二进制流数据转化成以时间为关键维度的列数据,通过对列进行查找和预统计来进一步提高数据查询及分析效率,大大优化信号集中监测数据存储性能,为监测系统更进一步发展提供的可靠的数据支持。

表4 历史数据统计分析表

并发数	查询天数(天)	平均耗时(毫秒)				
		12小时	1天	7天	30天	365天
1		3.10	3.86	4.43	7.93	44.39
5		4.59	4.34	7.91	9.54	47.47
10		6.20	7.66	9.16	14.59	71.76
20		12.51	16.31	20.41	21.82	86.41
40		20.25	22.66	27.76	32.26	188.87
50		26.56	31.90	35.44	37.71	257.45

参考文献

- [1] 丁顺亭,蒋纲,王立延.信号集中监测智能预警及诊断系统的设计与实现[J].铁道通信信号,2019,55(2):21-24.
- [2] 李昕,王荣亮.简谈区间综合监控与信号集中监测接口调试[J].铁路通信信号工程技术,2019,16(11):44-47.
- [3] 白银涛.论铁路信号集中监测系统的应用与发展[J].商品与质量,2016(38):11-12.
- [4] 随鸿杰.铁路信号集中监测智能分析系统实现方案的探讨[J].科学之友,2018(12):159-160.