基于图数据库的继电保护整定计算数据存储与应用

王镜毓[1],石东源[1],陈金富[1],曾次玲[2]

(1. 华中科技大学 电气与电子工程学院 强电磁工程与新技术国家重点实验室,湖北 武汉 430074;
2. 国网湖南省电力公司,湖南 长沙 410007)

摘要:电网继电保护整定计算数据涉及多个门类相互关联的信息,数据实体间的关系复杂而密集,更适合采用图数据库存储。对基于图数据库和传统关系数据库存储继电保护整定计算数据时的查询性能进行对比。将关系数据库中现存的继电保护整定计算数据准确迁移到图数据库中,既是进行对比测试前的一项准备工作,也是电力系统实际应用中升级数据存储方式的前提。在介绍图数据库和属性图模型基本概念的基础上,提出了一种利用离线交换文件实现数据迁移的通用方法,并通过 2 类数据库查询性能的对比测试,验证了基于图数据库存储和应用整定计算数据的可行性与高效性。

关键词:图数据库;属性图模型;Neo4j;数据迁移;性能对比;继电保护

中图分类号:TM 761 **文献标识码**:A DOI:10.16081/j.issn.1006-6047.2017.09.031

0 引言

长期以来,关系数据库凭借其成熟的技术、稳定的性能和通用的结构化查询语言(SQL),成为存储电力数据的首选方案[1-3]。继电保护整定计算数据面向实际电网物理对象,描述与电网继电保护分析计算相关的各类数据,不仅包括一次设备参数及其运行方式数据,还涉及与继电保护配置相关的二次设备数据等,因此其数据实体间具有复杂而密集的关联关系。目前主流整定计算软件普遍采用关系数据库存储继电保护整定计算数据。然而,由于关系数据库将数据实体间的关联关系以主键和外键的形式存储在实体属性中,这种关联关系仅在收到查询请求时才临时建构,导致其在执行复杂连接、嵌套查询时效率低下。随着电网规模的日益扩大和历史数据的不断积累,查询效率问题将更为突出。尤其是随着特高压联网的发展,传统的互联电网间仅交换边界等值参数的模式已无法适应准确开展继电保护整定计算的需要,国家电网和南方电网都在积极推进各级调控中心继电保护整定计算数据一体化的工作,通过各级调度逐级上报电网详细拓扑及参数形成更准确完整的电网基础数据,以此提升继电保护整定计算的准确度。对于数据一体化工作形成的数万节点规模电网,探索效率更高的数据存储模式具有重要意义[4-5]。

随着云计算、大数据等计算机新技术不断取得突破,一系列非关系型、分布式的 NoSQL 数据存储系统开始在某些场合替代关系数据库,以获得在具体应用下的性能优势[6-9]。图数据库就是一类重要的 NoSQL 数据存储系统。它将内部存储的数据组织为属性图逻辑模型,将关系数据库隐含在属性值中的关联关系独立成属性图模型中直接连接实体的边,并采用拓扑运算替代传统的关系运算,从而大幅提高了复杂连接、嵌套查询的执行效率。目前,图数据库已经引起了电力系统领域的关注[10-13]。众所周知,电力系统具有复杂网状拓扑结构,采用图数据库建模和存储,与电力系统的拓扑属性更为吻合,充分利用图数据库所具备的拓扑运算功能,将为电力系统应用的开发提供极大便利;另一方面,图数据库将实体间的关联关系固化存储,特别适合灵活高效地存储和管理关系密集的电力系统数据,为继电保护整定计算数据的存储和应用提供了新思路。

本文旨在对基于图数据库和传统关系数据库存储和管理继电保护整定计算数据时的查询性能进行对比,主要从以下 3 个方面进行深入探讨。

a. 本文介绍了图数据库与属性图模型的基本概念,指出了采用图数据库系统 Neo4j 存储关系密集型数据的优势。

b. 现有主流整定计算软件普遍采用关系数据库存储数据,为实现基于图数据库的数据存储与应用,需要研究可行的数据迁移方法。本文提出了一套基于离线交换文件的数据迁移方法,保障了迁移过程的准确性和通用性。

c. 结合继电保护整定计算工作的实际需要,本文对传统关系数据库和图数据库的查询性能进行了对比测试,验证了基于图数据库存储和应用继电保护整定计算数据的可行性和高效性。

1 图数据库概述

1.1 图数据库与属性图模型

图数据库是一类新兴的 NoSQL 数据存储系统,采用属性图模型组织数据。属性图是一种特殊的有向图,由一组节点和有向边构成。每条边的首末端

收稿日期:2016-08-06;修回日期:2017-05-17

必须与一个节点相连，任何一端都不允许悬空。节点之间通常通过一组边连接成复杂网络结构，但允许出现没有任何边与之相连的孤立节点。将现实系统建模为属性图时，节点一般用来表示数据实体，边用来表示实体间的关系，在不引起混淆的前提下，节点和边可分别用实体和关系代指。属性图模型如图 1 所示，属性图节点分别与实体和关系节点通过记录数据关系相连，关系节点与实体节点与通过组织关系相连，同时实体和关系节点均通过包含关系与属性节点相连。图数据库通过对属性图中节点与边的遍历、搜索与匹配代替关系数据库中针对关系表和元组的关系运算，从而实现数据查询操作。

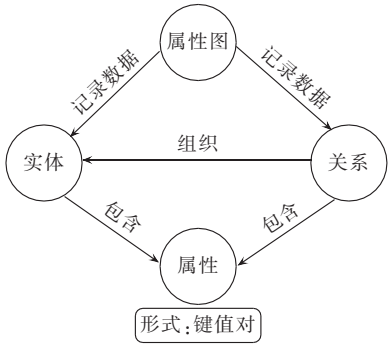


图 1　属性图模型
Fig.1　Property graph model

属性图中，节点和边均可包含若干属性，如图 1 中的属性节点包含一条形式：键值对属性。属性的引入是属性图与一般有向图的区别所在，也是属性图得名的原因。属性以键值对的形式存储，将现实系统中依附于实体和关系存在的信息映射到属性图中，为模型赋予复杂、丰富的语义。在实现层面，主流图数据库都支持向实体和关系赋予节点标签、关系类型等元属性，用来划定查询条件；同时，通过在属性上建立索引，加速以属性值为判据的查询操作。

图 1 中的属性图模型共包含 4 个实体与 5 条关系。事实上，在现实系统中事物之间的关联关系往往是复杂而密集的，属性图模型中边的数量通常是节点数量的指数倍[14]。直观来看，属性图模型就像实体与实体间的关系网络。传统关系数据库通过数据表中存储的主键与外键表达实体间的关联关系，这种关系仅在数据库收到查询请求时才会临时建构，从而降低了查询效率；而图数据库则利用属性图中的边，将数据实体间的关系固化在内部存储模型中，因而在存储和管理关系密集型数据方面具有效率优势。

1.2　图数据库系统 Neo4j

Neo4j 是一款基于 Java 语言开发的图数据库系统。与大多数底层采用关系数据库或基于键值对的 NoSQL 数据存储系统、仅在顶层引入图分析功能的非原生图数据库不同，Neo4j 作为一款原生图数据库系统，针对属性图的底层存储和查询分析都进行了大量优化。它可以嵌入到 Java 应用程序中，也可作为独立进程部署在单台服务器或主从集群上，并采用遵循原子性、一致性、独立性、持久性原则即 ACID 原则的事务机制保障数据一致性，从而满足不同应用场合对高性能、可扩展、高吞吐与冗余容错的不同需要。文献[15]对包括 Neo4j 在内的 4 款主流图数据库系统进行了性能测试，结果表明 Neo4j 在查询、遍历属性图模型方面具有绝对的效率优势。

与关系数据库普遍支持 SQL 不同，目前图数据库操作语言方面还没有形成统一的标准。Neo4j 通过其独有的描述性图形查询语言 Cypher 操作存储于其中的属性图模型。Cypher 使用类似 SQL 的原语，不需要编写复杂的图形算法即可直观而高效地实现属性图模型的增删改查。例如，语句“MATCH(n:Bus)--(:VoltageLevel{value:500kV})--(:ControlArea{name:华中})RETURN n”，表示查询华中区域内电压等级为 500 kV 的所有母线节点。而在关系数据库中出于一致性约束，往往需要将母线、电压等级和调控区域分别存储在 3 张数据表中，实现上述简单操作需要对 3 张数据表进行两重嵌套的连接查询，不仅编写的 SQL 语句较为繁琐，而且执行效率相对低下。此外，Neo4j 还将深度优先搜索算法、Dijkstra 算法、A* 算法等成熟图论算法集成进图数据库中，通过调用相应的应用程序编程接口(API)即可编写服务器端过程实现对属性图模型的拓扑分析，进一步挖掘数据间的隐含信息，这在传统关系数据库中是难以直接实现的。

2　整定计算数据迁移

2.1　迁移方式

目前主流整定计算软件均采用关系数据库存储整定计算数据，将其准确高效地迁移到图数据库中，是基于图数据库存储和应用整定计算数据的前提。

数据迁移方式主要有在线式和离线式 2 种。在线式数据迁移，即迁移程序同时连接源端关系数据库和目标端图数据库，实时从关系数据库中查询所需信息，并在图数据库中创建等价模型；离线式数据迁移，即将关系数据库中的数据通过整定计算软件导出为指定格式的离线交换文件，并通过对交换文件的解析和转换，在图数据库中创建相应的模型。与在线式数据迁移相比，离线式数据迁移不要求源端与目标端数据库服务器在同一网络中，且目前主流整定计算软件均支持导出符合国家电网相关标准的离线数据交换文件，具有更强的灵活性和通用性，因而本文采用离线交换文件实现整定计算数据由关系数据库向图数据库的迁移。

2.2　离线交换文件

国家电网公司已出台相关标准规范整定计算数

据交换文件的格式[4]。目前主流整定计算软件均支持导出CIM/E、CIM/G格式的交换文件。CIM/E与CIM/G是在IEC61970标准定义的能量管理系统公用信息模型的基础上发展起来的一套面向对象的数据描述格式。CIM/E将继电保护整定计算所涉及的厂站、设备、方式、保护等系统运行模型数据以数据表的形式给出,CIM/G则定义了设备图元,并给出了系统、厂站两层电网接线图。

CIM/E运行模型数据中的数据表可分为2类:一类反映电力系统实际物理对象,如发电机表,除包含用来表征发电机对象固有参数(如发电机名称、额定电压、电抗参数等)的字段外,还包含若干标识码字段,用来表征对象间的1:1或1:n关系,类似关系数据库中的外键,如发电机所属厂站字段等;另一类数据表则仅用于反映不同数据表元组间的m:n关系,如系统方式与厂站方式关系表等。

CIM/G电网图形数据则依托可扩展标记语言(XML)扩展,每个数据标签表示电网接线图中的一个图形元素。对于CIM/G中的区域、厂站和一次设备元素,通过一个全局标识码属性与CIM/E数据表中相应元组建立关联,实现图形模型一体化。在CIM/G中,图形元素间通过线路和连接线建立拓扑连接。不同的是,线路属于电力系统一次设备,对应于CIM/E线路表中的元组,而连接线信息在CIM/E中不存在,仅用来表示变压器绕组等设备与母线间的拓扑连接关系。

2.3 从离线交换文件建立属性图模型

从CIM/E、CIM/G格式的整定计算数据离线交换文件建立属性图模型的过程如下。

a. 通过全局标识码匹配CIM/G中的电力设备元素与CIM/E相应一次设备表中的元组,并将其固有参数合并,构造电力设备图形、模型一体化实体。

b. 将该电力设备实体表达为属性图中的节点,将实体的图形、模型参数表达为节点属性,并将CIM/G中的连接线表达为属性图中各电力设备实体间的电气连接关系。值得注意的是,由于CIM/G中的线路既对应一次设备实体,又表达拓扑连接关系,在属性图建模时需要对其进行模型解耦,将其对应于一次设备图形参数的属性归并入设备实体中,并新增虚拟的连接线表达其拓扑连接关系。

c. 将CIM/E中除一次设备表外的其他数据表中的元组构造为属性图中的节点,将利用独立的表或依附于字段建立的元组间关系表达为属性图中的边,并为节点和边赋予相应属性。

d. 对基本完成的属性图模型进行优化,针对继电保护整定计算工作的需要,在频繁查询的属性上建立索引,以提高图数据库的查询效率。

整定计算数据离线交换文件的属性图建模的过程如图2所示。以厂站宝月变内的双绕组变压器宝月1号变(图中为宝月1#变)及部分相关对象的建模为例,通过全局标识码将CIM/E中的PowerTransformer表相应元组与CIM/G中的变压器图形元素建立联系,并合并对应属性,实现图形、模型一体化,继而通过全局标识码将CIM/E中TransformerWinding和Substation表中的相关元组与PowerTransformer中的宝月1号变元组建立关联,最终在图数据库中建立起包含4个实体和3条关系的属性图模型片段。仿照这一过程,针对CIM/E、CIM/G中的全部数据进行建模,即可在图数据库中得到完整的整定计算数据属性图模型。

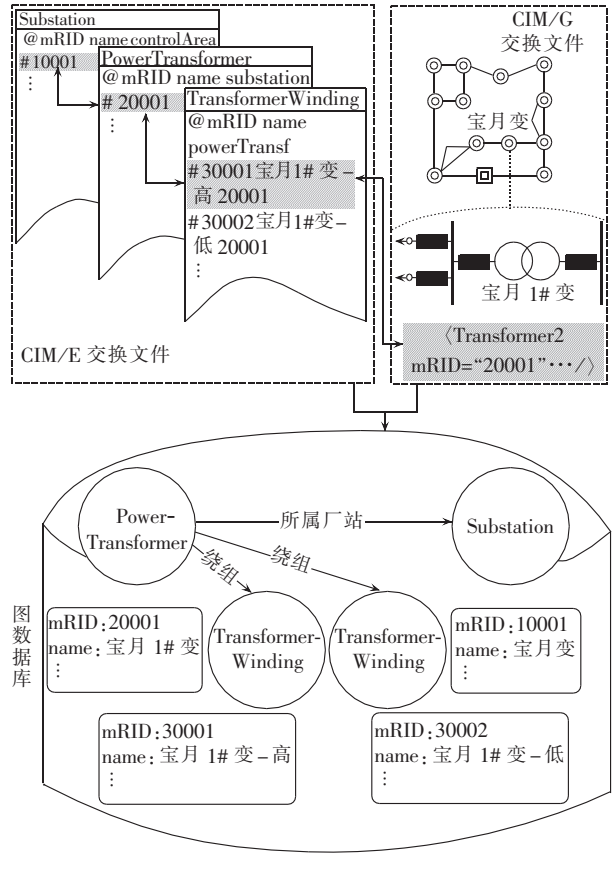


图2 离线数据交换文件的属性图建模过程
Fig.2 Modeling of property graph based on offline data exchange files

2.4 迁移结果

由于CIM/E和CIM/G格式的整定计算数据通过一次设备建立联系,实现了图形、模型数据的一体化表达,且CIM/E中各数据表都直接或间接与电力设备产生关联,故最终存储于图数据库中的属性图模型是一张围绕电力设备节点展开的复杂关系网络。图数据库中存储的部分整定计算数据如图3所示,在与宝月1号变节点相关的属性图中,宝月1号变节点居于中心位置。特别地,由电力设备连接而

成的实际电网物理接线图是该属性图的一个子图,如图 3 中"⇢"所示。

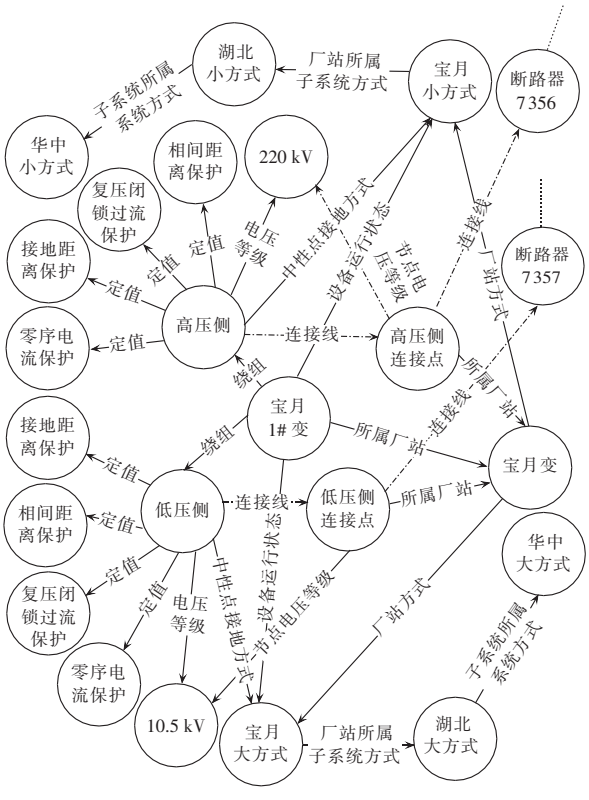


图 3 图数据库中存储的部分整定计算数据

Fig.3 Partial relay protection setting calculation data stored in graph database

3 性能比较

结合继电保护整定计算中常见的数据查询需求,本文使用一台 2.30 GHz Intel Core i5 CPU、4 G 内存、运行 Windows 7 64 位操作系统的计算机对传统的关系数据库和图数据库进行了性能测试。关系数据库选用业界广泛使用的 Microsoft SQL Serever 2008(下文简称 MSSQL 2008)图数据库选用本文中介绍的 Neo4j 3.0。关系数据库中的测试数据是通过目前电力调控部门广泛采用的主流整定计算软件 Relay CAC 对华中电网(7232 个节点)建模得到的,数据库文件大小为 607 MB。图数据库中的测试数据按照前文介绍的数据迁移方法,采用离线交换文件建立的等价模型,数据库文件大小为 2.42 GB。为了使测试结果对生产环境更具参考价值,2 类数据库均已结合实际需要进行了合理的优化,如在 MSSQL 2008 数据库中为数据表合理设置关键字、索引,并针对复杂操作建立存储过程等。测试程序采用 Java 语言编写,分别通过 Java 数据库连接(JDBC)和 Neo4j Driver 访问 2 类数据库,并测量执行查询操作所需时间开销。待测试的查询操作共 8 组:

a. 根据全局标识码 mRID 查询指定厂站;

b. 查询指定厂站内的所有一次设备;

c. 查询指定电压等级的全部线路;

d. 查询连接 2 个指定厂站的全部线路;

e. 查询包含指定厂站方式的全部系统方式;

f. 查询所有与变压器高压绕组相连的断路器;

g. 查询所有母联断路器;

h. 查询全部线路保护记录。

为了排除偶然因素对测试结果的干扰,保证测试的准确性,对 **a**—**e** 组操作以变换查询条件的形式各自随机生成 50 条查询语句,对 **f**—**h** 组操作重复执行 50 次,分别向 2 类数据库发起查询请求,并将测得的查询操作执行时间绘制在图 4 所示的盒形图中。由图 4 可以明显看出,图数据库 Neo4j 在执行诸如操作 **a** 的单条索引查询时,可以保证与关系数据库 MSSQL 2008 相近的高效率。当待查询问题变得复杂时,Neo4j 的执行时间均远少于传统关系数据库 MSSQL 2008。由于在实际开展整定计算和数据管理工作的过程

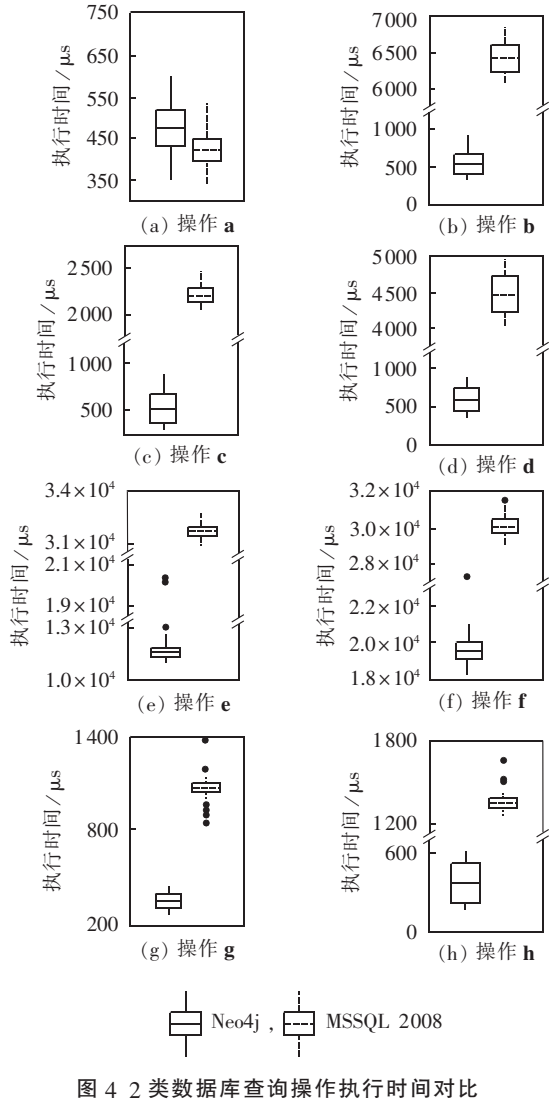


图 4 2 类数据库查询操作执行时间对比

Fig.4 Comparison of query operation execution time between two databases

中,涉及大量反复的数据库查询操作,因此采用Neo4j可以极大地提高整定计算软件的整体性能。

随着电网规模的逐渐扩大和历史数据的不断累积,数据库中存储的数据量将逐渐增多。为了比较数据量增长对2类数据库执行时间的影响,将2类数据库中的数据分别复制若干次,测量不同数据量下上述8组查询操作的平均执行时间,测量结果如图5所示。观察图5可以看出,2类数据库中查询操作的执行时间都随着数据量的增大而增长,但图数据库的增长速度远小于传统关系数据库。

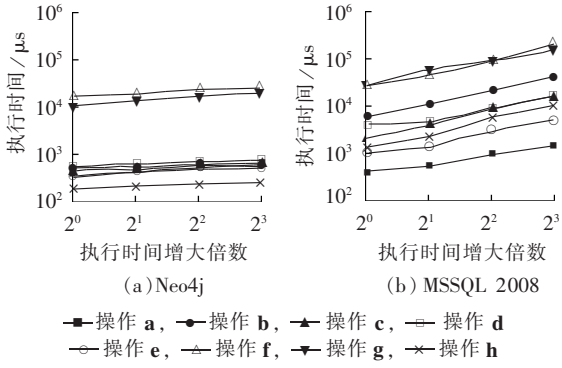


图5 数据量对2类数据库执行时间的影响

Fig.5 Effect of data amount on execution time between two databases

表1中给出了当数据量增大为原来的8倍时,图数据库与关系数据库平均执行时间增大的倍数。由表1可以看出,当数据量增大为原来的8倍时,图数据库Neo4j不管是对操作**a**这类简单的单条索引查询,还是对操作**f**、**g**这类涉及到复杂拓扑分析的查询操作,执行时间的增长均控制在2倍左右,与传统关系数据库相比可以更好地应对数据量增长对查询操作执行时间的影响,从而满足较长时期内继电保护整定计算数据规模增长的要求。

表1 数据量增大8倍时2种数据库执行时间增大倍数

Table1 Times of execution time magnification when data amount increases by 8 times for two databases

操作	执行时间增大倍数		操作	执行时间增大倍数	
	MSSQL 2008	Neo4j		MSSQL 2008	Neo4j
a	3.64	1.23	**e**	8.58	1.62
b	7.47	1.29	**f**	5.36	1.98
c	8.34	1.11	**g**	5.71	2.01
d	4.15	1.39	**h**	8.65	1.35

4 结语

结合计算机技术的发展,本文提出了基于图数据库存储和应用继电保护整定计算数据的新思路。由于现有主流整定计算软件普遍采用关系数据库存储数据,为保证现有数据从关系数据库向图数据库的准确迁移,本文给出了一种基于离线交换文件的整定计算数据迁移的通用方法,并在此基础上结合继电保护工作的实际需求,对2类数据库的查询效率进行了对比测试。测试表明,图数据库大幅提升了复杂连接、嵌套查询的执行效率,且查询时间受存储数据量增长的影响相对较小,与传统关系数据库相比,可以更好地保障互联大电网继电保护整定计算工作的开展。同时,本文提出的从CIM/E、CIM/G格式的离线交换文件建立属性图模型的方法,也可为能源互联网背景下电力系统各部门将现有基于关系数据库的应用移植到图数据库环境,以更高效地存储和应用海量关系密集型数据提供借鉴。

参考文献:

[1] 李济沅,钟一俊,王东举,等. 基于CIM和ESB的企业级电网准实时资源中心研究与应用[J]. 电力自动化设备,2015,35(8):148-155.
LI Jiyuan,ZHONG Yijun,WANG Dongju,et al. Research of enterprise-class quasi-real-time grid resource center based on CIM & ESB and its application[J]. Electric Power Automation Equipment,2015,35(8):148-155.

[2] 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.

[3] 曲广龙,杨洪耕,熊晓祎. 可在线编辑电网地理接线信息的电能质量监测系统设计与实现[J]. 电力自动化设备,2015,35(11):165-168,174.
QU Guanglong,YANG Honggeng,XIONG Xiaoyi. Design and realization of power quality monitoring system with online editable geographical wiring information of power grid[J]. Electric Power Automation Equipment,2015,35(11):165-168,174.

[4] 国家电网公司. 省级以上电网继电保护一体化整定计算技术规范:Q/GDW 11420—2015[S]. 北京:中国电力出版社,2015.

[5] 陈朝晖,周红阳,石东源,等. 大型互联电网继电保护整定计算数据一体化管理系统[J]. 电力系统自动化,2012,36(3):106-110.
CHEN Zhaohui,ZHOU Hongyang,SHI Dongyuan,et al. Data integrated management system for relay protection setting calculation of interconnected power system[J]. Automation of Electric Power Systems,2012,36(3):106-110.

[6] 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.

[7] 葛磊蛟,王守相,瞿海妮. 智能配用电大数据存储架构设计[J]. 电力自动化设备,2016,36(6):194-202.
GE Leijiao,WANG Shouxiang,QU Haini. Design of storage framework for big data of SPDU[J]. Electric Power Automation Equipment,2016,36(6):194-202.

[8] 宋亚奇,刘树仁,朱永利,等. 电力设备状态高速采样数据的云存储技术研究[J]. 电力自动化设备,2013,33(10):150-156.
SONG Yaqi,LIU Shuren,ZHU Yongli,et al. Cloud storage of power equipment state data sampled with high speed[J]. Electric Power Automation Equipment,2013,33(10):150-156.

[9] 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.
[10] MA Yisong,WU Zhigang,GUAN Lin,et al. Study on the relationship between transmission line failure rate and lightning information based on Neo4j[C]//2014 International Conference on Power System Technology. Chengdu,China:IEEE,2014:474-479.
[11] GELLI R,KHAPARDE S A. CIM oriented graph database for network topology processing and applications integration[C]// 2015 50th International Universities Power Engineering Conference. Stanffordshire,UK:IEEE,2015:1-7.
[12] SRIKRISHNA P,AVINASH S B. Application of polyglot persistence to enhance performance of the energy data management systems[C]//2014 International Conference on Advances in Electronics Computers and Communications. Bangalore,India:IEEE,2014:1-6.
[13] 马义松,武志刚. 基于 Neo4j 的电力大数据建模及分析[J]. 电工电能新技术,2016,35(2):24-30.
MA Yisong,WU Zhigang. Modeling and analysis of big data for power grid based on Neo4j[J]. Advanced Technology of Electrical Engineering and Energy,2016,35(2):24-30.
[14] 程学旗,靳小龙,王元卓,等. 大数据系统和分析技术综述[J]. 软件学报,2014,25(9):1889-1908.
CHENG Xueqi,JIN Xiaolong,WANG Yuanzhuo,et al. Survey on big data system and analytic technology[J]. Journal of Software,2014,25(9):1889-1908.
[15] SALIM J,VALENTIN V. An empirical comparison of graph databases[C]//2013 International Conference on Social Computing(SocialCom). Alexandria,USA:IEEE,2014:1-8.

作者简介:


王镜毓

王镜毓(1993—),男,河北昌黎人,硕士研究生,主要研究方向为大数据与人工智能算法在电力系统中的应用(**E-mail**:jywang@hust.edu.cn);

石东源(1974—),男,湖南邵阳人,副教授,博士,主要研究方向为电力系统二次系统自动化、电力系统软件技术(**E-mail**:dongyuanshi@hust.edu.cn);

陈金富(1972—),男,福建莆田人,副教授,博士,主要研究方向为电网继电保护整定计算软件开发、信息技术在电力系统中的应用(**E-mail**:chenjinfu@hust.edu.cn);

曾次玲(1976—),女,湖南新化人,高级工程师,博士,主要研究方向为电力系统继电保护整定计算、故障信息系统运行与维护(**E-mail**:zengcl@hn.sgcc.com.cn)。

Storage and application of relay protection setting calculation data based on graph database

WANG Jingyu[1],SHI Dongyuan[1],CHEN Jinfu[1],ZENG Ciling[2]

(1. State Key Laboratory of Advanced Electromagnetic Engineering and Technology,School of Electrical and Electronic Engineering,Huazhong University of Science and Technology,Wuhan 430074,China;
2. State Grid Hunan Electric Power Company,Changsha 410007,China)

Abstract: The relay protection setting calculation data of power system involve the interrelated information of multiple categories. Since the relationships among data entities are complex and dense,they are more suitable to be stored in the graph database. The query performance is compared between the protection setting calculation data storage based on graph database and that based on relational database. The relay protection setting calculation data existed in the relational database are accurately migrated to the graph database,which is not only a preparation of the comparison,but also the premise of data storage mode update for actual power system applications. The basic concepts of graph database and property graph model are introduced and a general data migration method based on the offline data exchange files is proposed. By the comparison of query performance between relational database and graph database,the feasibility and efficiency of the storage and application of relay protection setting calculation data based on graph database are verified.

Key words: graph database; property graph model; Neo4j; data migration; performance comparison; relay protection