

基于 NoSQL 数据库的结构化存储设计与应用

杨磊

(河南电力试验研究院, 河南郑州 450052)

[摘要] NoSQL 数据库是最近几年才逐渐兴起的, 它没有严格的结构及类型方面的限制, 具有良好的扩展性和灵活性, 并且在海量存储及并发等方面, 也远远高于我们传统使用的关系型数据库。本文作者结合在工作中的实际应用, 以发电设备的技术规范结构化存储为例, 介绍了在设计方面遇到的一些问题, 以及基于 NoSQL 数据库之一的 MongoDB 分布式文档型数据库的解决思路。通过与传统设计方式进行比较, 使用 NoSQL 数据库能够在满足客户需求的同时, 很大程度上简化系统设计及开发等方面的工作, 提高了生产效率。

[关键词] NoSQL 数据库; MongoDB; 结构化存储

1 NoSQL 数据库简介

近年来, 随着 Web2.0 技术的兴起, NoSQL 数据库得到了迅速的发展, 并且逐步走进了我们的视线。与传统的关系型数据库不同, NoSQL 数据库所拥有的高效率、高并发、高扩展等特性是传统关系型数据库不可比拟的, MongoDB 数据库 (以下简称为 MongoDB) 就是众多出色 NoSQL 数据库中的一员。MongoDB 属于分布式文档型数据库, 具有高性能、易部署、易使用、易存储等特点, 拥有面向集合存储、模式自由等特性。它采用类似 JSON 的 BSON 数据格式, 能够存储各种复杂类型数据, 并且, 提供了高效的海量数据存储, 以及强大的查询功能以及索引机制。

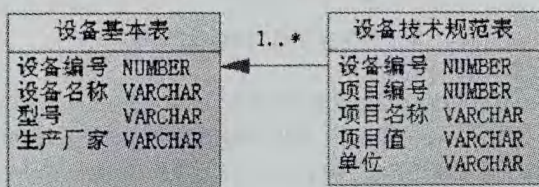
2 发电设备技术规范存储的现状

随着电力工业的快速发展, 大量先进设备的引进和投运, 对技术监督工作提出了更多的要求。在配合技术监督工作的顺利进行的同时, 为了加快知识积累和提高服务的及时性, 技术监督系统需要对大量文档型数据 (包括设备技术规范、试验报告等), 以及种类繁多的报表资源进行管理和监视。其中, 有很大一部分是采用文档格式存储的 (如: doc、xls 等文件格式)。为了能够更好地利用这些数据, 对其进行统计分析、纵/横对比等操作, 我们需要对其进行结构化存储。

以设备技术规范为例, 电厂设备种类繁多, 每类设备都拥有大量的技术规范, 甚至同种类的不同设备之间的差别也很大。如何有效、便捷地对这些数据资源进行结构化存储和管理, 成为众多软件设计人员需要面临和解决的难题之一。

3 基于 MongoDB 文档数据库的结构化存储方案

在应用 MongoDB 之前, 我们先讨论一下, 在使用传统关系型数据库时, 是如何对设备技术规范进行结构化存储的。通常会采用以下的数据结构, 建立两个数据表, 分别用来存储设备基本信息和设备技术规范信息, 数据结构设计如下:



这种设计方式当然可行, 并且我们之前也是这样做的。但是, 它存在着一个非常棘手的问题, 那就是设备技术规范表中“项目值”的数据类型无法统一, 它可能出现字符、数值等类型, 甚至还有可能会是一个范围。在设计时, 我们只好将该字段的类型设置为“varchar”, 以便能够存储可能出现的各类型信息。这样做解决了数据存储的问题, 但并没有真正地实现结构化存储, 在后期的统计分析时同样无法有效利用这些信息。

那么, 问题该如何解决呢? 我们在大量实验中遇到了一个 NoSQL 数据库, 它就是 MongoDB, 它面向集合存储、模式自由的特性为我们提供了直接而有效的解决方案。

在 MongoDB 中, 每行记录的数据结构都可以不一样, 包括值的类型等, 这里, 笔者继续以设备技术规范结构化存储为例, 介绍一下

MongoDB 是如何解决上述问题的, 我们设计了以下的存储结构及示例:

表格 1 设备信息 (基于 MongoDB 的设计)

_id	Data
1	{ "设备名称": "某吸风机", "型号": "Y4", "生产厂家": "厂家 A", "设备种类": "吸风机", "技术规范": { "电机型号": { "v": "YB800-5-B/10" }, "转速": { "v": { "L": 594, "H": 745, "U": "x/min" }, "风量": { "v": { "L": 45.0319, "H": 90.0638, "U": "万 m³/h" } } } }
2	{ "设备名称": "某阀门", "型号": "J61Y-32", "生产厂家": "厂家 B", "设备种类": "阀门", "技术规范": { "公称直径": { "v": 20, "u": "mm" } } }

这次, 我们运行 MongoDB 面向集合存储, 以及模式自由的特性, 真正意义上实现了对设备技术规范进行了结构化存储。并且, 它提供了强大的检索功能, 能够很快速地查询这些数据, 例如“查找转速达到 600r/min 的吸风机”:

```
Coll.find("设备信息.设备种类": "吸风机",
"设备信息.技术规范.转速.V.L": "$gt;600")
);
```

同时, MongoDB 也为我们提供了更强大的 Map/Reduce 分布式计算框架。能够提高大规模数据计算的效率, 并且它提供一套机制, 能够将计算扩展到无限规模的机器集群上进行。

4 总结

本文以设备技术规范结构化存储为例, 介绍了笔者在实际工作中所遇到的问题, 以及基于 MongoDB 数据库的解决方案。MongoDB 数据库灵活的存储方式和强大的检索功能, 不仅能够很好地应用在对文档型数据的结构化存储中, 还可以满足其它多种场景下数据存储的需求。MongoDB 只是众多优秀 NoSQL 数据库中的一员, 还有其它很多 NoSQL 数据库也能够帮助我们解决实际问题中所遇到的问题, 并且大部分都是开源免费的, 例如: Redis、CouchDB 等。相信, 随着 NoSQL 数据库的不断深入推广和应用, 它们将逐渐成为主流的数据存储解决方案。

【参考文献】

- [1] (美)霍多罗夫, (美)迪洛尔夫著, 程显峰译. MongoDB 权威指南. 人民邮电出版社, 2011.
- [2] (美)鲁比, (美)托马斯, (美)汉森著, 林芷薰译. Web 开发敏捷之道——应用 Rails 进行敏捷 Web 开发 (第 3 版). 电子工业出版社, 2010.
- [3] (印)沃哈著, 祁飞, 何菁译. Ruby on Rails 实践之路. 科学出版社, 2010.

作者: [杨磊](#)
作者单位: [河南电力试验研究院, 河南郑州, 450052](#)
刊名: [科技风](#)
英文刊名: [Technology Trend](#)
年, 卷(期): 2011(18)

本文读者也读过(10条)

1. [韩君易](#) [应用NoSQL数据库解决方案的思考](#)[期刊论文]-[科技信息](#)2011(22)
2. [吕明育](#), [李小勇](#) [NoSQL数据库与关系数据库的比较分析](#)[期刊论文]-[微型电脑应用](#)2011, 27(10)
3. [韩君易](#) [基于电子商务平台的NoSQL解决方案应用研究](#)[期刊论文]-[科技信息](#)2011(20)
4. [张华强](#), [ZHANG Hua-qiang](#) [关系型数据库与NoSQL数据库](#)[期刊论文]-[电脑知识与技术](#)2011, 07(20)
5. [甲骨文NoSQL数据库第一印象](#)[期刊论文]-[硅谷](#)2011(23)
6. [NoSQL正在替代关系型数据库](#)[期刊论文]-[硅谷](#)2011(20)
7. [徐娟娟](#), [朱成亮](#), [Xu JuanJuan](#), [Zhu ChengLiang](#) [NOSQL在WEB日志分析中的应用](#)[期刊论文]-[中国新技术新产品](#)2011(10)
8. [韩君易](#) [NoSQL数据库解决方案Tair浅析](#)[期刊论文]-[电子商务](#)2011(9)
9. [姚林](#), [张永库](#), [YAO Lin](#), [ZHANG Yong-ku](#) [NoSQL的分布式存储与扩展解决方法](#)[期刊论文]-[计算机工程](#)2012, 38(6)
10. [卢冬海](#), [何先波](#), [LU Dong-hai](#), [HE Xian-bo](#) [浅析NoSQL数据库](#)[期刊论文]-[中国西部科技](#)2011, 10(2)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_kjf201118084.aspx