

# 浅析NoSQL数据库

卢冬海 何先波

( 西华师范大学计算机学院, 四川 南充 637002 )

**摘要:** NoSQL数据库打破了传统的关系模型, 以一种模式自由的方式存储数据, 提供了新型的访问接口, 并克服了传统RDBMS的缺点。NoSQL数据库可部署在廉价的硬件之上, 支持分布式存储, 能透明地扩展节点。本文介绍了NoSQL数据库的基本特点与设计思想, 列举了几种流行的NoSQL数据库产品, 分析了其应用方向、优缺点及发展前景。

**关键词:** NoSQL; SQL; 关系型数据库

DOI: 10.3969/j.issn.1671-6396.2011.02.008

The Analysis of NoSQL Database

LU Dong-hai, HE Xian-bo

(School of Computer Science, China West Normal University, Nanchong Sichuan 637002)

**Abstract:** NoSQL database that breaks the traditional relational model, stores data as a free style, provides a new type of access interface, and overcomes the shortcomings of RDBMS database to design to be deployed on inexpensive hardware to support distributed storage and to transparent extension node. This article described some common ideas of NoSQL database, listed kinds of popular NoSQL database products, and analyzed their applications, advantages, disadvantages and prospects.

**Key words:** NoSQL; SQL; RDBMS

## 1 关系型数据库面临的挑战

### 1.1 数据库高并发读写需求

在Web2.0时代, 网站通常要根据用户的个性化定制实时生成页面, 例如现在流行的SNS网站, 微博网站等。网站几乎要实时地为用户提供信息。该类应用对数据库提出了很高的并发负载要求, 传统的RDBMS面临很大的挑战。

### 1.2 海量数据的高效存储需求

在Web2.0时代, 网站信息的提供者由传统的网站信息管理员变成了普通的用户, 用户提供的信息是海量的。类似facebook, qq空间等SNS类型的网站, 可能每天都会产生千万级的数据。如果在RDBMS里的一张存有亿级记录的数据表里作SQL查询, 耗时间巨大。虽然可通过分库、分表等方法切分数据, 部分地解决查询问题, 但也带来了诸如加重程序开发的复杂度和数据备份以及数据库扩容的复杂度等问题。

### 1.3 数据库高扩展性和高可用性需求

在云计算时代, 一项很重要的任务就是存储交由云端, 云计算供应商需面对存储海量数据的挑战。如果用传统的RDBMS来保证存储的海量性和高可用性, 云计算供应商必须花费巨额的资金去购置高性能高可靠性的机器。同时, RDBMS的无缝、不宕机扩容实现难度也大大增加。

## 2 NoSQL介绍

NoSQL数据库指那些非关系性的、定义不是很明确的数据存储仓库。NoSQL数据库不再使用关系模型的概念, 放弃了SQL数据库操作语句。NoSQL数据库克服了RDBMS的缺点, 可部署在廉价的硬件之上, 支持分布式存储, 能透

明地扩展节点。典型的NoSQL数据库以key-values的形式存储数据, 具有模式自由的特点。

### 2.1 key-values

key-values是指一个键名对应一个键值, 可以通过键名访问键值。例如一条员工的记录信息如图1和图2所示, 有Name、Age、Profession等键名, 各个键名对应着一个键值。

```
employeeA
{
  Name "tom" ,
  Age : 13 ,
  Profession "teacher" ,
  Birth {year 1990 , month 9 : day : 10} ,
  Email "myemail@gmail.com"
}
```

图1

```
employeeB
{
  Name "tom" ,
  Age : 13 ,
  Profession "teacher" ,
  Birth {year 1990 , month 9 : day : 10} ,
}
```

图2

### 2.2 模式自由

模式自由是指使用数据库前不再预先定义数据模型。在传统的RDBMS中, 如果想要存储某一员工的信息, 必须先定义一张员工表, 表里有各项与员工相关的字段。如果日后需求有变更, 要增加员工的信息就必须去修改原先定义的数据模型。模式自由的数据库没有预先定义要存储的数据的数据模型。仍以员工信息为例, 并不是所有员工的记录信息里都有name, age, profession, email这些key, 有可能员工B的

收稿日期: 2010-11-20 修回日期: 2010-12-17

作者简介: 卢冬海 (1986-), 男, 汉族, 浙江三门籍, 研究生, 研究方向为嵌入式系统。

信息就没有Email(如图2所示)。

### 3 NoSQL数据库的产品及应用现状

#### 3.1 Dedis

Dedis是一种高性能的key-values型的内存数据库,对数据库的操作都在内存中进行,并定期把数据更新到硬盘上以实现数据的持久存储。因为读写操作是在内存中进行的,所以Dedis的速度非常快,每秒可以处理超过10万次的读写操作。Dedis支持丰富的数据类型,它支持存储的value类型有strings(字符串)、lists(链表)、sets(集合)和zsets(有序集合)。strings可以用来存储一般的文本。使用get和set命令来存取值,可以使用INCR, DECR等命令进行加减操作。lists类型支持从两端插入,取lists区间,排序等操作。利用Dedis的lists类型做一个fifo双向列表,可以实现一个轻量级的高性能消息队列服务。sets类型支持对集合的交并操作,可以用来实现高性能的tags系统。Dedis的主要缺点是受到内存容量的限制,不能对海量数据作高性能的读写。如果突发掉电,Dedis来不及把数据flush到硬盘上,可能会出现丢失数据的现象。Dedis主要应用在较小数据量的高性能读写操作上。

#### 3.2 Tokyo Tyrant/Cabinet

Tokyo Tyrant/Cabinet是日本最大的SNS社交网站mixi.jp开发的key-values型的数据库,其中TC是一个NoSQL数据库,用来做持久化数据,TC的读写速度非常快,写入100万条数据只需要0.4秒,读取100万条数据只需要0.33秒。TC除了支持key-values存储外,还支持保存哈希表。TT则是TC的网络接口,它使用简单的基于tcp/ip二进制协议通信。TC/TT在mixi的实际应用中,存储了千万级的数据,支撑了上万个并发连接,表现出了良好的性能。

#### 3.3 MongoDB

Mongodb是一种非常优秀的面向文档存储的数据库,它主要解决海量数据的存储和访问效率的问题。数据以一种类似json格式的bson格式组织成一个文档,存储在一个集合里。根据官方测试,当数据量达到50Gb以上的时候,Mongodb的访问速度是MySQL的10倍以上,支撑的并发可以达到每秒0.5~1.5万次。Mongodb自带了一个很出色的分布式文件系统gridfs,用来支持海量数据存储。此外Mongodb还支持复杂的数据结构,有很强的数据查询功能,基本上可以完成关系型数据库要完成的任务。

### 4 NoSQL数据库的前景

#### 4.1 NoSQL数据库的优势

##### 4.1.1 灵活的数据模型

在传统的RDBMS的领域,存储数据库前必须分析数据,构建数据模型。构建数据模型就是建立数据表,确定表里的各个字段,字段的数据类型以及表之间字段之间的关联。应用随着时间的推移,需求必将,即数据模型将发生改变。然而,在生产环境中更改数据模型是非常困难的问题。即使是增加一个字段,或者更改一个字段的类型,都需要付出极大的代价,甚至是停机或者降低服务级别。

NoSQL数据库的共同思想就是要打破这种数据模型的限制。NoSQL数据库允许应用在一个数据单元中存入其想要的任何结构,数据单元间一般没有模型的限制,他们之间的联系是扁平的。如果要添加或删除数据单元里的数据项,不会影响到其他的数据单元。理论上可以提高应用的迭代速度。然而,如果不注意管理数据的完整性,那么模式自由也会带来灾难。

##### 4.1.2 弹性扩展

随着数据库的负载增加,RDBMS的管理员一般采用向上扩展的方式提高性能。即购买性能更好的服务器代替旧服务器,这针对负载缓慢增加的情况是较有效的。然而负载可能一直增加,不可能每次都更换服务器。这样,水平扩展的思想被提出来,把负载均衡的分到不同的主机上。RDBMS提供了水平扩展的功能,但对程序来说是半透明的,在水平扩展的时候,会大量的修改程序,甚至会停服。NoSQL数据库在设计之初就考虑到了水平扩展,它对程序来说是透明的,可以随时添加节点,删除节点。

##### 4.1.3 经济性

RDBMS通常被部署在高性能机器上,甚至是专用硬件上,价格比较昂贵。在云计算时代,数据量急速暴增,云提供商为了提供可靠的服务不得不大量采购昂贵的机器。而NoSQL数据库通常使用廉价服务器集群来管理暴增的数据与事务规模。廉价服务器集群的方案,相对高性能机器的RDBMS的集群有更多的数据节点,提供了更廉价、更可靠、更多备份的服务。

### 4.2 NoSQL数据库的缺陷

#### 4.2.1 成熟度

RDBMS已经存在了30多年了。经过实践的证明,RDBMS满足了过去30年来的软件需求,为数据存储作出了巨大的贡献,无数的应用以RDBMS作为后端的存储。NoSQL的实际应用刚刚起步,但对企业来说,稳定压倒一切,关系型数据库的成熟度是可以让人放心的,它运行稳定,功能也很丰富。相比而言,大部分NoSQL的产品都还处于实验阶段。

#### 4.2.2 商业支持

RDBMS有着强大的商业支持。Oracle、Microsoft和IBM等世界级的数据库厂商提供了完善的服务。相比而言,大部分的NoSQL数据库都是开源项目,虽然也有一些商业公司的支持,但支持的效果不可预计。如果出现故障,通常只能自己解决。

#### 4.2.3 专业技能

软件开发人员已经习惯了RDBMS,大部分的开发人员的思想已经被固定在关系模型上,想要改变他们的思想是一件比较难的事情。使用NoSQL数据库的开发人员在短时间内很难快速增加。

### 5 结束语

NoSQL数据库作为一种新型的数据库,打破了传统数据库存储的观念,提出了创新的存储方式,在海量存储,快速访问,分布式上很好满足了云计算时代数据存储的要求。各种NoSQL的产品不断涌现,NoSQL的实际应用也在迅速增加。但新生的NoSQL不太完美,并不能完全取代RDBMS,在应用中

(下转第14页)

循环流化床固硫渣除了具有一般燃煤灰渣的火山灰活性外,其重要的一个特性就是它还具有自硬性。根据GB5947-86定义,水硬性是指一种材料磨成细粉与水混合成浆体后,能在潮湿空气和水中硬化并形成稳定化合物的性能<sup>[10]</sup>。而循环流化床固硫渣自身与水拌合后,能在潮湿空气和水中硬化并形成稳定的化合物。这主要是由于固硫剂燃烧产生的游离态CaO激发了固硫渣中的活性SiO<sub>2</sub>和活性Al<sub>2</sub>O<sub>3</sub>,生成了C-A-H和C-S-H凝胶,使得固硫渣表现出一定的水硬性,而溶解的硬石膏与C-A-H反应生成的钙矾石,进一步增加了体系的强度。此外,二水石膏的生成也会产生强度。因此,固硫渣具有一定的自硬性。

#### 4.5 水化膨胀性

固硫渣不能像煤粉锅炉灰那样得到广泛应用的一个重要原因就是固硫渣具有“高硫”、“高钙”这个特殊的性质,导致其体积稳定性差,使制品后期表面开裂,强度大幅度下降,尤其是在养护环境水分充足的情况下。由于固硫渣是燃煤加入一定的固硫剂(如石灰石等)在循环流化床锅炉中燃烧后收集得到的燃煤灰渣,在固硫过程中,原煤中的硫被固定下来,使得固硫渣中的SO<sub>3</sub>含量较普通煤粉锅炉灰的高,且由于固硫剂的加入,使固硫渣中的CaO含量也较高,同时为了提高脱硫效率,固硫剂和原煤混合时,钙硫比必须高于脱硫反应的理论值,一般在2~2.5之间,导致固硫渣中还含有固硫剂残留并分解的f-CaO,因此固硫渣中的CaO、SO<sub>3</sub>、f-CaO含量较普通煤粉锅炉灰高。而且由于固硫渣中的硫是以在850~900℃烧成的-CaSO<sub>4</sub>形式存在,具有与天然硬石膏和400℃煅烧石膏不同的性质,溶解速度相比要慢得多。固硫渣中f-CaO被-CaSO<sub>4</sub>包裹,它们的存在既是固硫灰渣利用的有利因素,对于激发灰渣的活性有利<sup>[11]</sup>,又是影响灰渣及其胶凝材料体积稳定性的不利因素。CaSO<sub>4</sub>溶解后会与水化铝酸钙反应生产钙矾石,一方面有利于体系强度的提高,同时也会产生膨胀,而由于灰渣中CaSO<sub>4</sub>溶解速度较慢,在水化后期当-CaSO<sub>4</sub>溶解度达到二水石膏的饱和溶解度才能结晶析出,此时体系已经具有一定的强度,其结晶膨胀是导致灰渣安定性不好的原因之一。

固硫渣与水混合后,-CaSO<sub>4</sub>除可水化为二水石膏之外,还可与活性Al<sub>2</sub>O<sub>3</sub>、游离CaO发生火山灰反应而生成钙矾石,游离CaO可水化为Ca(OH)<sub>2</sub>,这些水化都会引起体积明显膨胀。CaO水化为Ca(OH)<sub>2</sub>体积增大为原来的1.98倍,-CaSO<sub>4</sub>水化结晶为CaSO<sub>4</sub>·2H<sub>2</sub>O体积增大到2.26倍,-CaSO<sub>4</sub>溶于水与活性Al<sub>2</sub>O<sub>3</sub>和Ca(OH)<sub>2</sub>反应生成钙矾石体积增大到2.22倍<sup>[12]</sup>,因此固硫灰水化后会引起的明显的体积膨胀。文献[13]、[14]研究结果表明固硫渣具有很强的膨胀性,标

准养护28d龄期时的自由膨胀值可高达万分之三百以上。

#### 5 结语

循环流化床锅炉在燃烧过程中添加固硫剂(如石灰石等),可以把含硫煤燃烧产生的SO<sub>2</sub>固定下来,同时不降低其燃烧效率。不仅大大降低了SO<sub>2</sub>的排放量,而且还符合国家产业政策和环保要求,因此这种燃煤技术无疑是将来的发展方向。

由于流化床燃煤固硫渣是一种相对较为特殊的废渣,其成分波动较大,处理和用起来较为困难,一直缺少有效的利用方式,从现有资料分析,关于流化床燃煤固硫渣的研究成果不多,对其综合利用的研究显得很缺乏。通过基本特性研究可以发现,固硫渣是一种具有较高火山灰活性的燃煤灰渣,它在建材中的应用是可能的,具有广阔的前景。

参考文献:

- [1] Shah N D. Fluidized bed combustion-a review[J]. Chemical Engineering World, 1985, 20(11): 37~45.
- [2] Fox E C, Krishnan R P, Daw C S, et al. A review of fluidized bed combustion technology in the United States[J]. Energy, 1986, 11(11-12): 1183~1200.
- [3] Topper J M, Cross P J I, Goldthorpe S H. Clean coal technology for power and cogeneration[J]. Fuel, 1994, 73(7): 1056~1063.
- [4] Beer J M. Combustion technology developments in power generation in response to environmental challenges[J]. Progress in Energy and Combustion Science, 2000, 26(4-6): 301~327.
- [5] Everett A S, Steven A B, John P H, et al. Review of advances in combustion technology and biomass cofiring[J]. Fuel Processing Technology, 2001, 71(1-3): 7~38.
- [6] 刘妮, 路春美, 骆仲决等. 石灰石颗粒固硫反应特性的模型研究[J]. 环境科学学报, 2001, 21(2): 172~177.
- [7] Cille E, Belens L W. 陆华, 武洞明译. 石灰[M]. 北京: 中国建筑工业出版社, 1981: 151~160.
- [8] 杨娟. 固硫灰渣特性及其作水泥掺合料研究[D]. 重庆大学学位论文, 2006, 10.
- [9] 宋远明. 流化床燃煤固硫灰渣水化研究[D]. 重庆大学学位论文, 2007, 10.
- [10] Iribarne A P, Iribarne J V, Anthony E J. Reactivity of calcium sulfate from FBC systems[J]. Fuel, 1997, 76(4): 321~327.
- [11] 孙恒虎, 郑娟荣. 低温煤渣火山灰活性的机理研究[J]. 煤炭学报, 2006(12): 664~667.
- [12] 侯天明. 循环流化床锅炉环保技术[J]. 石油化工环境保护, 2002, 25(3): 44~47.
- [13] 梁文泉. 固硫渣混合水泥的研究及钙矾石相的定量分析[J]. 武汉水利电力大学学报, 1994, 27(5): 492~499.
- [14] 刘燕芳. 用沸腾炉燃煤固硫渣制作膨胀水泥的研究[D]. 北京: 清华大学, 1988: 33~39.

(上接第16页)可根据实际需要作为RDBMS很好的补充。

参考文献:

- [1] 杨卫华. 一切为了分布式——2009年Web后端技术回顾[J]. 程序员, 2010, (2): 71~73.
- [2] Ricky Ho. NoSQL的模式[J]. 程序员, 2010, (2): 95~99.

- [3] 潘凡. 从MySQL到MongoDB——视觉中国的NoSQL之路[J]. 程序员, 2010(6): 79~81S
- [4] 黄贤立. NoSQL非关系型数据库的发展及应用初探[J]. 福建电脑, 2010, (7): 30~32.
- [5] 范凯. NoSQL数据库综述[J]. 程序员, 2010, (6): 76~78.