

1. NoSQL

1.1 什么是NoSQL?

NoSQL，指的是非关系型的数据库。NoSQL有时也称作Not Only SQL的缩写，是对不同于传统的关系型数据库的数据库管理系统的统称。NoSQL用于超大规模数据的存储。（例如谷歌或Facebook每天为他们的用户收集万亿比特的数据）。这些类型的数据存储不需要固定的模式，无需多余操作就可以横向扩展

1.2 为什么使用NoSQL?

今天我们可以通过第三方平台（如：Google,Facebook等）可以很容易的访问和抓取数据。用户的个人信息，社交网络，地理位置，用户生成的数据和用户操作日志已经成倍的增加。然而，传统的关系型数据库难以存储这些非结构化或半结构化数据；传统的关系型数据库更适用于结构化较强的交易性数据。我们如果要对这些用户数据进行挖掘，那SQL数据库已经不适合这些应用了，NoSQL数据库的发展也却能很好的处理这些大的数据。

1.3 RDBMS与NoSQL的比较

RDBMS

- 高度组织化结构化数据
- 结构化查询语言（SQL）(SQL)
- 数据和关系都存储在单独的表中。
- 数据操纵语言，数据定义语言
- 严格的一致性
- 基础事务

NoSQL

- 代表着不仅仅是SQL
- 没有声明性查询语言
- 没有预定义的模式
- 键 - 值对存储，列存储，文档存储，图形数据库
- 最终一致性，而非ACID属性
- 非结构化和不可预知的数据
- CAP定理
- 高性能，高可用性和可伸缩性

1.4 NoSQL简史

NoSQL一词最早出现于1998年，是Carlo Strozzi开发的一个轻量、开源、不提供SQL功能的关系数据库。

2009年，Last.fm的Johan Oskarsson发起了一次关于分布式开源数据库的讨论，来自Rackspace的Eric Evans再次提出了NoSQL的概念，这时的NoSQL主要指非关系型、分布式、不提供ACID的数据库设计模式。

2009年在亚特兰大举行的"no:sql(east)"讨论会是一个里程碑，其口号是"select fun, profit from real_world where relational=false;"。因此，对NoSQL最普遍的解释是"非关联型的"，强调Key-Value Stores和文档数据库的优点，而不是单纯的反对RDBMS。

1.5 NoSQL的优点/缺点

优点:

- 高可扩展性
- 分布式计算
- 低成本
- 架构的灵活性，半结构化数据
- 没有复杂的关系

缺点:

- 没有标准化
- 有限的查询功能（到目前为止）
- 最终一致是不直观的程序

1.6 谁在使用？

现在已经有很多公司使用了 NoSQL：

- Google
- Facebook
- Mozilla
- Adobe
- Foursquare
- LinkedIn
- Digg
- McGraw-Hill Education
- Vermont Public Radio

2. MongoDB

2.1

MongoDB 是由C++语言编写的，是一个基于分布式文件存储的开源数据库系统。

在高负载的情况下，添加更多的节点，可以保证服务器性能。

MongoDB 旨在为WEB应用提供可扩展的高性能数据存储解决方案。

MongoDB 将数据存储为一个文档，数据结构由键值(key=>value)对组成。MongoDB 文档类似于JSON 对象。字段值可以包含其他文档，数组及文档数组。

2.2 主要特点

MongoDB 是一个面向文档存储的数据库，操作起来比较简单和容易。

- 你可以在MongoDB记录中设置任何属性的索引 (如: FirstName="Sameer",Address="8 Gandhi Road")来实现更快的排序。
- 你可以通过本地或者网络创建数据镜像，这使得MongoDB有更强的扩展性。
- 如果负载的增加（需要更多的存储空间和更强的处理能力），它可以分布在计算机网络中的其他节点上这就是所谓的分片。
- Mongo支持丰富的查询表达式。查询指令使用JSON形式的标记，可轻易查询文档中内嵌的对象及数组。
- MongoDB 使用update()命令可以实现替换完成的文档（数据）或者一些指定的数据字段。
- MongoDB中的Map/reduce主要是用来对数据进行批量处理和聚合操作。Map和Reduce。Map函数调用emit(key,value)遍历集合中所有的记录，将key与value传给Reduce函数进行处理。Map函数和Reduce函数是使用Javascript编写的，并可以通过db.runCommand或mapreduce命令来执行MapReduce操作。
- GridFS是MongoDB中的一个内置功能，可以用于存放大量小文件。
- MongoDB允许在服务端执行脚本，可以用Javascript编写某个函数，直接在服务端执行，也可以把函数的定义存储在服务端，下次直接调用即可。
- MongoDB支持各种编程语言:RUBY, PYTHON, JAVA, C++, PHP, C#等多种语言。
- MongoDB安装简单。

2.3 历史

- 2007年10月，MongoDB由10gen团队所发展。2009年2月首度推出。
- 2012年05月23日，MongoDB2.1 开发分支发布了! 该版本采用全新架构，包含诸多增强。
- 2012年06月06日，MongoDB 2.0.6 发布，分布式文档数据库。
- 2013年04月23日，MongoDB 2.4.3 发布，此版本包括了一些性能优化，功能增强以及bug修复。
- 2013年08月20日，MongoDB 2.4.6 发布。
- 2013年11月01日，MongoDB 2.4.8 发布。

.....

2.4 MongoDB 下载

你可以在mongodb官网下载该安装包，地址为：<https://www.mongodb.com/download-center#community>。MongoDB支持以下平台：

- OS X 32-bit
- OS X 64-bit
- Linux 32-bit
- Linux 64-bit
- Windows 32-bit
- Windows 64-bit
- Solaris i86pc
- Solaris 64

2.5 MongoDB 工具

有几种可用于MongoDB的管理工具。

- 监控
 - MongoDB提供了网络和系统监控工具Munin，它作为一个插件应用于MongoDB中。
 - Gangila是MongoDB高性能的系统监视的工具，它作为一个插件应用于MongoDB中。基于图形界面的开源工具 Cacti, 用于查看CPU负载, 网络带宽利用率,它也提供了一个应用于监控 MongoDB 的插件。
- GUI
 - Fang of Mongo – 网页式,由Django和jQuery所构成。
 - Futon4Mongo – 一个CouchDB Futon web的mongodb山寨版。
 - Mongo3 – Ruby写成。
 - MongoHub – 适用于OSX的应用程序。
 - Opricot – 一个基于浏览器的MongoDB控制台, 由PHP撰写而成。
 - Database Master – Windows的mongodb管理工具
 - RockMongo – 最好的PHP语言的MongoDB管理工具, 轻量级, 支持多国语言。

2.6 MongoDB 应用案例

下面列举一些公司MongoDB的实际应用：

- Craigslist上使用MongoDB的存档数十亿条记录。
- FourSquare，基于位置的社交网站，在Amazon EC2的服务器上使用MongoDB分享数据。
- Shutterfly，以互联网为基础的社会和个人出版服务，使用MongoDB的各种持久性数据存储的要求。
- bit.ly，一个基于Web的网址缩短服务，使用MongoDB的存储自己的数据。
- spike.com，一个MTV网络的联营公司，spike.com使用MongoDB的。

- Intuit公司，一个为小企业和个人的软件和服务提供商，为小型企业使用MongoDB的跟踪用户的数据。
- sourceforge.net，资源网站查找，创建和发布开源软件免费，使用MongoDB的后端存储。
- etsy.com，一个购买和出售手工制作物品网站，使用MongoDB。
- 纽约时报，领先的在线新闻门户网站之一，使用MongoDB。
- CERN，著名的粒子物理研究所，欧洲核子研究中心大型强子对撞机的数据使用MongoDB。

3. MongoDB安装

3.1 Windows安装和启动

1. 下载软件的时候注意系统是32还是64位
2. 安装好之后，将bin路径加入到系统路径
3. 构建数据库文件夹（d:\mongodb\data\db）和log文件夹（d:\mongodb\log）
4. 启动mongodb: mongod --dbpath "d:\mongodb\data\db" (如果提示engine是错的，则在以上语句最后附上example提示内容)
*注意：在win32中，默认引擎为WiredTiger，但32位置支持MMAPv1引擎，所以在启动服务时应加上storageEngine参数: * mongod --dbpath "d:\mongodb\data\db" --storageEngine=mmapv1
5. 注册mongodb服务: mongod --dbpath "d:\mongodb\data\db" --logpath "d:\mongodb\log\mongodb.log" --install --serviceName "MongoDB" --journal
注意：在win32中，默认引擎为WiredTiger，但32位置支持MMAPv1引擎，所以在启动服务时应加上storageEngine参数: mongod --dbpath "d:\mongodb\data\db" --storageEngine=mmapv1 --logpath "d:\mongodb\log\mongodb.log" --install --serviceName "MongoDB" --journal
6. 启动mongodb服务: net start mongodb
或者 在服务列表当中更改mongodb服务的属性
7. 进入mongodb: mongo

3.2 导入、导出collection:

导出

语法: mongoexport -d 数据库名称 -c 集合名称 -o 文件路径与名称 --type json

注意: 文件路径与名称不能加文件后缀, 文件类型可以为json或csv

- 示例: `mongoexport -d cloudfunding -c projects -o /Users/xiaoyu/Downloads/projects --type json`

导入

语法: `mongoimport -d 数据库名称 -c 集合名称 --file 文件路径与名称 --type json`

- 示例: `mongoimport -d test -c projects --file /Users/xiaoyu/Downloads/projects --type json`

3.3 备份、还原database:

备份

语法: `mongodump -h 服务器地址 -d 数据库名称 -o 目标备份存放文件夹`

- 示例: `mongodump -h 127.0.0.1 -d test -o /Users/xiaoyu/Downloads/`

还原

语法: `mongorestore -h 服务器地址 -d 数据库名称 --dir 数据库备份存放文件夹`

- 示例: `mongorestore -h 127.0.0.1 -d test2 --dir /Users/xiaoyu/Downloads/test/`