

# 三十五个非主流开源数据库 MongoDB 领衔主演

更新时间:2010-08-12 作者: 51cto.com 来源:51cto.com

几乎每个 Web 开发人员都有自己喜欢的数据库，或自己最熟悉的数据库，但最常见的无外乎以下几种：

MySQL

PostgreSQL

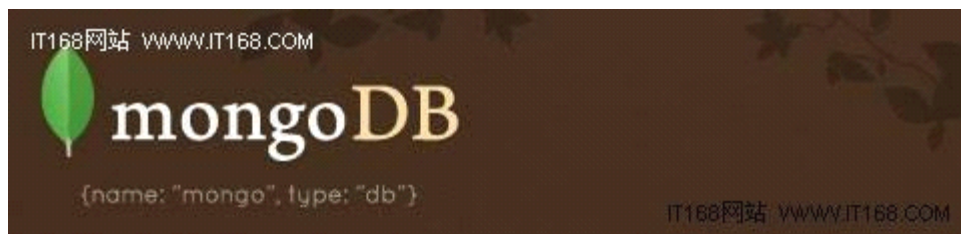
MSSQL Server

SQLite

MS Access

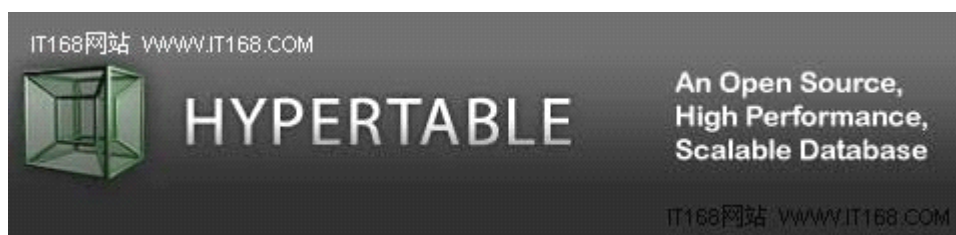
或是更简单的 XML，文本文件等。这些数据库有优秀的文档，背后有强大的社区支持，大部分流行的 CMS 都使用了其中之一或多个，它们都易于使用，大多数托管服务供应商都提供了相应的产品，因此它们的使用量很多，名气也很大。但除了这些主流的数据库外，还有很多其它非主流数据库存在，其中有一些也开始受到人们的高度重视，下面我们就一起来看看吧，注意本文只介绍开源数据库，说不定在下一个项目中，你就有尝试它们的冲动。

## 1、MongoDB



MongoDB 是一款开源，高性能，可扩展，无模式，面向文档(与 JSON 类似的数据模式)的数据库，它为时下最流行的编程语言提供了驱动，如 PHP, Python, Perl, Ruby, JavaScript, C++等，支持全文索引，自动分片，跨 LAN 或 WAN 扩展，采用 Key/Value 方式存储数据。MongoDB 服务端可运行在 Linux、Windows 或 OS X 平台，支持 32 位和 64 位应用。世界上最大的单词收录网站 Wordnik 就从 MySQL 转向了 MongoDB。

## 2、Hypertable



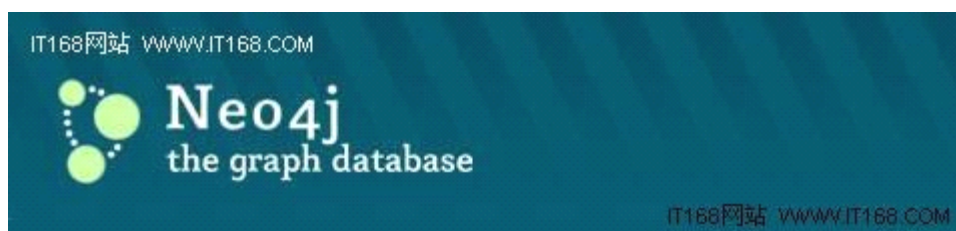
Hypertable 是一款高性能分布式数据存储系统，旨在为应用程序提供最好的性能，可扩展性和可靠性，它建立在 Google 的 BigTable 之上，主要面向大规模数据集应用，其目标是要成为世界上最好的大规模并发高性能数据库平台。百度目前也使用了 Hypertable，也是它的赞助商。

## 3、Apache CouchDB



Apache CouchDB 是一款面向文档的数据库，可以使用 JavaScript 通过 MapReduce 方法进行查询和索引，它提供了一个 RESTful JSON API，因此可以在任何环境中通过 HTTP 访问，CouchDB 内置了 Web 管理控制台，支持通过浏览器管理数据库。CouchDB 使用 Erlang 编写，Erlang 是一种健壮的函数式编程语言，非常适合于构建并发的分布式系统，Erlang 的设计非常灵活，其可伸缩性与可扩展性都非常棒。

## 4、Neo4j



Neo4j 是一个嵌入式，基于磁盘的，支持完整事务的 Java 持久化引擎，它在图像中而不是表中存储数据。Neo4j 提供了大规模可扩展性，在一台机器上可以处理数十亿节点/关系/属性的图像，可以扩展到多台机器并行运行。相对于关系数据库来说，图形数据库善于处理

大量复杂、互连接、低结构化的数据，这些数据变化迅速，需要频繁的查询——在关系数据库中，这些查询会导致大量的表连接，因此会产生性能上的问题。Neo4j 重点解决了拥有大量连接的传统 RDBMS 在查询时出现的性能衰退问题。通过围绕图形进行数据建模，Neo4j 会以相同的速度遍历节点与边，其遍历速度与构成图形的数据量没有任何关系。此外，Neo4j 还提供了非常快的图形算法、推荐系统和 OLAP 风格的分析，而这一切在目前的 RDBMS 系统中都是无法实现的。

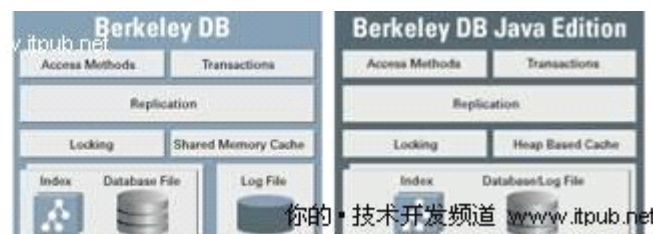
## 5、Riak



你的 • 技术开发频道 [www.itpub.net](http://www.itpub.net)

Riak 是一款非常适合于 Web 应用程序的数据库，它提供了去中心化的 Key/Value 存储，灵活的 map/reduce 引擎和友好的 HTTP/JSON 查询接口。它是一个真正的容错系统，不会出现单点故障，在 Riak 世界中，没有哪台机器是特殊的或属核心服务器，它们都是对等的。

## 6、Oracle Berkeley DB



你的 • 技术开发频道 [www.itpub.net](http://www.itpub.net)

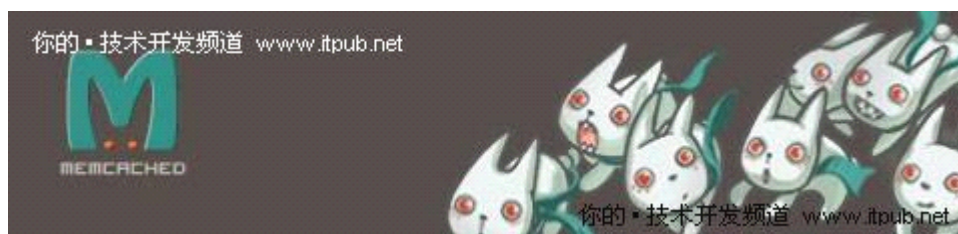
Oracle Berkeley DB 是一系列开源的**嵌入式数据库**，使开发人员能够将一个快速、可伸缩、具有工业级别的可靠性和可用性的事务处理数据库引擎结合进他们的应用程序中。Berkeley DB 最先由伯克利加州大学为了移除受到 AT&T 限制的程式码，从 BSD 4.3 到 4.4 时所改写的软件。Berkeley DB 运行在大多数的操作系统中，例如大多数的 UNIX 系统，和 windows 系统，以及实时操作系统。

## 7、Apache Cassandra



Cassandra 是一款高可扩展性第二代分布式数据库，属于混合型的非关系的数据库，类似于 Google 的 BigTable，支持的数据结构非常松散，类似于 JSON 的 BJSON 格式，因此可以存储比较复杂的数据类型。Cassandra 最初由 Facebook 开发，后转变成了开源项目。Cassandra 的主要特点就是它不是一个数据库，而是由一堆数据库节点共同构成的一个分布式网络服务，对 Cassandra 的一个写操作，会被复制到其他节点上去，对 Cassandra 的读操作，也会被路由到某个节点上面去读取。对于一个 Cassandra 群集来说，扩展性能是比较简单的事情，只管在群集里面添加节点就可以了。Facebook，Digg，Twitter 和 Cisco 等大型网站都使用了 Cassandra。

## 8、Memcached



Memcached 是开源的分布式 cache 系统，现在很多的大型 web 应用程序包括 facebook，youtube，wikipedia，yahoo 等等都在使用 memcached 来支持他们每天数亿级的页面访问。通过把 cache 层与他们的 web 架构集成，他们的应用程序在提高了性能的同时，还大大降低了数据库的负载。

Memcached 处理的原子是每一个 key/value 对，key 会通过一个 hash 算法转化成 hash-key，便于查找、对比以及做到尽可能的散列。同时，memcached 用的是一个二级散列，通过一张大 hash 表来维护。

## 9、Firebird



Firebird 是一个关系数据库，可以运行在 Linux，Windows 和各种 Unix 平台上，Firebird 相对 MySQL 和 PostgreSQL 来说比较小，这也使其可以称得上是理想的嵌入式数据库，可用于与其它应用程序服务器和应用程序捆绑。Firebird 具有大部分成熟数据库所具有的功能，比如支持存储过程、SQL 兼容等。

## 10、Redis



Redis 是一款快速的 Key/Value 数据库引擎，它在保持键值数据库简单快捷特点的同时，又吸收了部分关系数据库的优点，从而使它的位置处于关系数据库和键值数据库之间。Redis 不仅能保存 Strings 类型的数据，还能保存 Lists 类型(有序)和 Sets 类型(无序)的数据，而且还能完成排序(SORT)等高级功能，在实现 INCR，SETNX 等功能的时候，保证了其操作的原子性，除此以外，还支持主从复制等功能。Redis 使用 C 语言编写，可以想 memcached 那样使用，放在传统数据库的前端，它支持许多编程语言，受到许多流行的项目使用，如 GitHub 和 Engine Yard，有一个用 PHP 编写的客户端叫做 Rediska，专门来管理 Redis 数据库。

## 11、HBase



HBase 是一个分布式，面向列存储的数据库引擎，也可以叫做 Hadoop 数据库，因为它



是 Hadoop 的子项目，HBase 的目标是托管数十亿行，数百万列的大表，它提供了一个 REST 风格的 Web 服务器网关，支持 XML，Protobuf 和二进制数据编码选项。

## 12、Keyspace



Keyspace 是一家叫做 Scalien 的创业公司开发的高可靠 key/value 存储系统，Keyspace 强调的技术点是高可靠性，有以下一些特点：

**Key/Value 存储：**一个 key/value 数据存储系统，只支持一些基本操作，如：SET(key, value) 和 GET(key) 等；

**分布式：**多台机器(nodes)同时存储数据和状态，彼此交换消息来保持数据一致，可视为一个完整的存储系统。为了更可靠，Keyspace 推荐使用奇数个 nodes，比如：3，5，7 等；

**数据一致：**所有机器上的数据都是同步更新的、不用担心得到不一致的结果，Keyspace 使用著名的 Paxos 分布式算法；

**冗余：**所有机器(nodes)保存相同的数据，整个系统的存储能力取决于单台机器(node)的能力；

**容错：**如果有少数 nodes 出错，比如重启、当机、断网、网络丢包等各种 fault/fail 都不影响整个系统的运行；

**高可靠性：**容错、冗余等保证了 Keyspace 的可靠性。

## 13、4store



4store 是一个容纳 RDF 数据的数据库存储和查询引擎，它使用 ANSI C99 编写，可在类 Unix 系统上运行，提供一个高性能，可扩展和稳定的平台。4store 专门为无共享集群进行优化，最大可支持 32 节点集群，导入性能最大可以达到 120kT/s，它的查询性能也相当出众。

## 14、MariaDB



MariaDB 是一个向后兼容的，旨在替换 MySQL 数据库的 MySQL 分支，它包括所有主要的开源存储引擎，另外也开发了属于自己的 Maria 存储引擎。MariaDB 是由原来 MySQL 的作者 Michael Widenius 创办的公司所开发的免费开源数据库服务器，与 MySQL 相比较，MariaDB 更强的地方在于：

Maria 存储引擎

PBXT 存储引擎

XtraDB 存储引擎

FederatedX 存储引擎

更快的复制查询处理

线程池

更少的警告和 bug

运行速度更快

更多的 Extensions (More index parts, new startup options etc)

更好的功能测试

数据表消除

慢查询日志的扩展统计

支持对 Unicode 的排序

## 15、Drizzle



# A Lightweight SQL Db

Overview

Branches

Bugs

Blueprint

你的 • 技术开发频道 [www.itpub.net](http://www.itpub.net)

Drizzle 是从 MySQL 衍生出来的一个数据库，但它的目的不是要取代 MySQL，它的宗旨是构建一个“更精练、更轻量、更快速”的 MySQL 版本，它的扩展性和易用性与 MySQL 相当，但为了提高性能和扩展性，它从原来的核心系统里移除了部分功能。Drizzle 是一种为云和网络程序进行了特别优化的数据库，它是为在现代多 CPU/多核架构上实现大规模并发而设计的。

## 16、HyperSQL



**hsqldb - 100% Java Database**

你的 • 技术开发频道 [www.itpub.net](http://www.itpub.net)

HyperSQL 是用 Java 编写的一款 SQL 关系数据库引擎，它的核心完全是多线程的，支持双向锁和 MVCC(多版本并发控制)，几乎完整支持 ANSI-92 SQL，支持常见数据类型，最新版本增加了对 BLOB 和 CLOB 数据的支持，最高支持达 64T 的数据量。同时，HyperSQL 也是一个不错的嵌入式数据库。

## 17、MonetDB



MonetDB 是一个高性能数据库引擎，主要用在数据挖掘，OLAP，GIS，XML Query，文本和多媒体检索等领域。MonetDB 对 DBMS 的各个层都进行创新设计，如基于垂直分片的存储层，为现代 CPU 优化的查询执行架构，自动和自助调整索引，运行时查询优化，以及模块化的软件架构。MonetDB/SQL 是 MonetDB 提供的关系数据库解决方案，MonetDB/XQuery 是 XML 数据库解决方案，MonetDB Server 是 MonetDB 的多模型数据库服务器。



## 18、Persevere



Persevere 是针对 Javascript 设计的基于 REST 的 JSON 数据库，分布式计算，持久对象映射的框架，提供独立的 web 服务器，主要用于设计富客户端应用，可以用在任何框架和客户端上。Persevere Server 是一个基于 Java/Rhino 的对象存储引擎，在交互式的客户端 JavaScript 环境中提供持久性的 JSON 数据格式。

## 19、eXist-db



eXist-db 是使用 **XML 技术**构建的数据库存储引擎，它根据 XML 数据模型存储 XML 数据，提供高效的，基于索引的 XQuery 查询。eXist-db 支持许多 Web 技术标准，使得它非常适合 Web 应用程序开发：

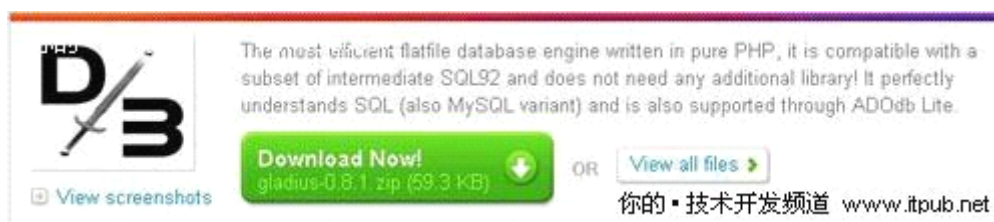
XQuery 1.0 / XPath 2.0 / XSLT 1.0 (使用 pache Xalan)或 XSLT 2.0

HTTP 接口：REST，WebDAV，SOAP，XMLRPC，Atom 发布协议

XML 数据库规范：XMLDB，Xupdate，XQuery 更新扩展

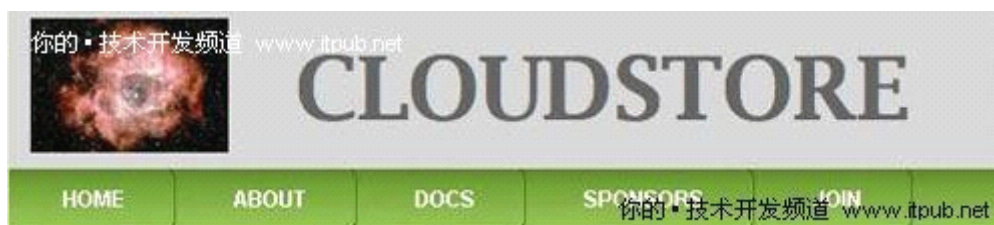
最新的 1.4 版本还增加了基于 Apache Lucene 的全文索引，轻量级 URL 重写和 MVC 框架，以及对 XProc 的支持。eXist-db 与 XQuery 标准高度兼容(目前 XQTS 的得分是 99.4%)。

## 20、Gladius



Gladius 是用**纯 PHP 编写的平面文件数据库引擎**，它的 SQL 语法与 SQL92 的一个子集兼容，它捆绑了一个轻量级的 adoDB 驱动。

## 21、CloudStore



CloudStore(以前叫做 Kosmos 文件系统)是一个开源的高性能**分布式文件系统**，它是用 C++编写的，CloudStore 可以和 Hadoop 以及 Hypertable 集成，这样就允许应用程序构建在那些系统上，而底层数据存储无缝地使用 CloudStore。CloudStore 支持 Linux 和 Solaris，主要用来存储 Web 日志和 Web 爬行数据。

## 22、OpenQM



[About OpenQM](#)

[Sales and Downloads](#)

# Taking Multivalue

你的·技术开发频道 www.itpub.net

OpenQM 是唯一一款同时有商业支持和免费的开源多值数据库，基于 GPL 协议发布，多值数据库对 NoSQL 运动起到了推动作用，它自身也因速度快，体积小，比关系数据库便宜而很快得到了认可。名称 OpenQM 中的 Open 表示开源版本，QM 表示商业闭源 QM 数据库。商业版本支持 Windows，Linux(RedHat, Fedora, Debian, Ubuntu)，FreeBSD，Mac OS

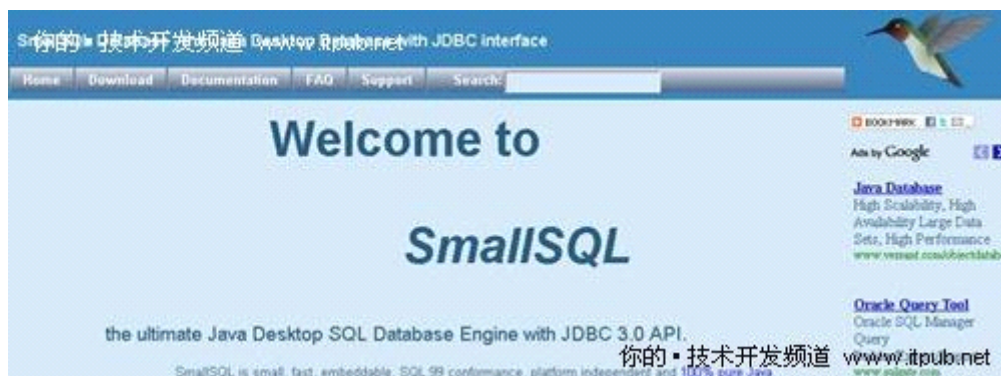
X 和 Windows Mobile，其列表价格还不到其它多值产品的 1/5，商业版本还包括一个 GUI 管理界面和终端模拟器，开源版本仅包括核心多值数据库引擎，主要是为开发人员准备的。

## 23、ScarletDME



ScarletDME 也是一个开源多值数据库，它是 OpenQM 的社区分支版，最初由 Ladybridge 开发，这个项目创立于 2008 年 11 月 28 日，它既在独立开发自己的功能，也在为 OpenQM 贡献代码。这个项目最初的名字叫做 Ladybridges GPL OpenQM，现在正式改为 ScarletDME，其中的 DME 是 Data Management Environment(数据管理环境)的首字母缩写。

## 24、SmallSQL



SmallSQL 是一个 100%纯 Java 编写的轻量级数据库，一般用于嵌入式领域，兼容 SQL 99 标准，支持 JDBC 3.0 API，定位于高端 Java 桌面 SQL 数据库。支持所有能运行 Java 的平台，可直接嵌入到应用程序中。不过它也有一些不足，如没有网络接口，必须安装 Java 运行时，同一时间不能在多个应用程序之间共享数据库，没有用户管理。

## 25、LucidDB



LucidDB 是唯一一款专注于数据仓库和商务智能的开源 RDBMS，它使用了列存储架构，支持位图索引，哈希连接/聚合和页面级多版本，大部分数据库最初都注重事务处理能力，而分析功能都是后来才加上去的。相反，LucidDB 中的所有组件从一开始就是为满足灵活的需求，高性能数据集成和大规模数据查询而设计的，此外，其架构设计彻底从用户出发，操作简单，完全无需 DBA。

LucidDB 对硬件要求也极低，即使不搭建集群环境，在单一的 Linux 或 Windows 服务器上也能获得极好的性能。最新版本还加入了对 Mac OS X 和 Windows 64 位的支持，官方网站上的文档和教程也非常丰富，非常值得你体验一下。

## 26、HyperGraphDB



HyperGraphDB 是一种通用的，可扩展的，可移植的，分布式，嵌入式和开源数据存储机制，它是一个图形数据库，专门为人工智能和语义 Web 项目而设计，它也可用于任意规模的嵌入式面向对象的数据库。正如其名，HyperGraphDB 是用来存储超图的，但它也属于一般图形数据库家族，作为一个图形数据库，它不施加任何限制，相比其他图形数据库它的功能更丰富。

HyperGraphDB 非常稳定，已经应用在多个生产环境，包括一个搜索引擎和 Seco scripting IDE。它支持\*nix 和 Windows 平台，需要 Java 5+。

## 27、InfoGrid



The Internet Graph Database

InfoGrid 是一个互联网图形数据库，它提供了许多额外的组件，使得在图像基础上开发 RESTful Web 应用程序变得更加容易。InfoGrid 是开源的，包括一系列项目：

InfoGrid 图形数据库项目 – InfoGrid 的心脏 GraphDatabase，可以独立使用，也可以附加到其它 InfoGrid 项目。

InfoGrid 图形数据库网格项目 – 在 GraphDatabase 基础上增加了复制协议，因此多个分布式 GraphDatabase 就可以在一个非常大的图像管理环境中协作。

InfoGrid 存储项目 – 象 SQL 数据库和分布式 NoSQL 哈希表那样，为存储技术提供一个抽象的通用接口，这样 InfoGrid GraphDatabase 就可以使用任何存储技术持久化数据。

InfoGrid 用户接口项目 – 将 GraphDatabase 中的内容以 REST 风格映射成浏览器可访问的 URL。

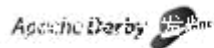
InfoGrid 轻量级身份识别项目 – 实现以用户为中心的身份识别技术，如 LID 和 OpenID。

InfoGrid 模型库项目 – 定义一个可复用对象模型库，作为 InfoGrid 应用程序的模式使用。

InfoGrid Probe 项目 – 实现 Probe 框架，它允许开发人员将任何互联网上的数据源当作一个图像对象看待。

InfoGrid Utilities 项目 – 收集 InfoGrid 使用的常见对象框架和实用代码。

## 28、Apache Derby



Home Quick Start Download Community Documentation Resources

你的·技术开发频道 www.itpub.net

Apache Derby 是 Apache DB 的子项目，它完全用 Java 编写，是一个开源关系数据库，它的体积非常小，基础引擎加上 JDBC 驱动只有 2.6MB，它支持 SQL 标准，它提供了一个嵌入式 JDBC 驱动，因此可以嵌入到任何基于 Java 的应用程序中，Derby 也支持常见的客户端/服务器模式，它也易于安装和使用。



## 29、hamsterdb



Hamsterdb 是一个轻量级嵌入式 NoSQL Key/Value 存储引擎，它已经有 5 年历史，现在它的开发重点放在易用性，高性能，稳定性和可扩展性上。Hamsterdb 支持事务(同一时间只能处理一个事务)，支持内存数据库，支持基于 HTTP 服务器的嵌入式远程数据库，支持日志/恢复，AES 加密，基于 zlib 的压缩，支持 C++，Python，.NET 和 Java 编程语言。

## 30、H2 Database



H2 Database 是一个开源的 Java 数据库，它的速度很快，包括 JDBC API，支持嵌入式和服务器模式，内存数据库，提供了一个基于浏览器的控制台程序，它的体积也非常小，只有一个大小约 1MB 的 jar 文件，它还支持 ODBC 驱动和全文搜索。

## 31、EyeDB



EyeDB 是一款基于 ODMG 3 规范的面向对象数据库管理系统，为 C++ 和 Java 提供了编程接口，它功能非常强大，并且成熟，稳定和安全，实际上，它起源于 1992 年的 Genome View 项目，1994 年又进行了重写，广泛用于生物信息项目。

## 32、txtSQL



txtSQL 是一个面向对象的平面文件数据库管理系统，它使用 PHP 编写，支持对普通文本文件的操作，虽然是一个文本数据库，但同样支持 SQL 的一个子集，并且执行效率非常高，txtSQL 使用文件系统的方法与 MySQL 的表和数据库原理类似，它有一个类似于 phpMyAdmin 管理界面。

## 33、db4o



你的·技术开发频道 [www.itpub.net](http://www.itpub.net)

db4o 是一个面向对象的开源数据库，允许 Java 和 .NET 开发人员用一行代码存储和检索任何应用程序对象，无需预定义或维护一个独立的，僵化的数据模型，因为模型现在是由

db4o 根据需要自动创建和更新的。db4o 成功的秘密是因为它的易用性，它原生为 Java 和 .NET 设计，存储数据对象的方法直接在应用程序中定义，因此 db4o 很容易集成到应用程序中，由于只需要一行代码，因此执行效率非常高。

## 34、Tokyo Cabinet



Tokyo Cabinet 是一个 Key/Value 型数据库，每个 Key 和 Value 的长度都可以不同，Key 和 Value 既可以是二进制数据，也可以是字符串，无数据表和数据类型的概念，记录是以哈希表、B+树和固定长度数组形式组织的。Tokyo Cabinet 具有以下优点：

空间利用率高 – 数据文件尺寸更小；

执行效率高 – 更快的处理速度；

并发性能好 – 在多线程环境性能更好；

改善的可用性 – 简化的 API；

改善的可靠性 – 即使在发生灾难的情况下，数据文件也不会损坏；

支持 64 位架构 – 支持海量的存储空间和巨型数据库文件。

Tokyo Cabinet 是用 C 语言编写的，为 C，Perl，Ruby，Java 和 Lua 提供了 API。

## 35、Voldemort 项目



Voldemort 是一个分布式 Key/Value 存储系统，它具有以下特点：

数据自动在多个服务器之间复制；

数据自动分区，因此每个服务器只包括整体数据的一个子集；

服务器故障处理是透明的；

支持插入式序列化，允许丰富的 Key 和 Value 类型，包括列表和元组，也可以集成常见的序列化框架，如 Protocol Buffers, Thrift, Avro 和 Java Serialization

数据项支持版本化，即使在故障情况下，数据完整性也可以得到保障;

每个节点都是独立的，无需其他节点协调，因此也没有中央节点;

单节点性能优秀：根据机器配置、网络、磁盘系统和数据复制因素的不同，每秒可以执行 10-20k 操作;

支持地理分散式部署。

LinkedIn 目前就使用 Voldemort 解决了高可伸缩性存储问题。

不知道你是否一口气看完本文，我想你也一定会惊讶于这么多开源数据库吧，事实上，本文也尚未完全罗列，欢迎你的补充。