

数据库软件

Question：科普类型：美国的数据库软件在什么地方先进，我们差距在哪？有什么卡脖子的？现在全球竞争格局是怎样的？

1 数据类型

数据库是用来存储数据的，先简单介绍一下数据有哪些类型。

常见的数据类型有：

1.1 关系型数据

其实就类似于Excel表格通常所呈现的数据那样，每一行每一列整整齐齐，井井有条。同一个表之间有可能有唯一性关系，不同的表之间也可能存在一些关联。

例如：

序号	姓名	身高	体重
1	张三	160	50
2	李四	161	60
3	王五	162	70

1.2 非关系型数据

其他的数据基本就归类为非关系型数据，例如图片、视频、文档 等等

2 数据库

由此，为了针对不同的数据就需要专门设计对应的数据库进行存储。

数据库软件在不同的维度有不同的软件设计，按照**数据类型**，**数据量大小**，**存储位置** 等维度有不同的数据库软件。

按照数据类可以划分为关系型数据库和非关系型数据库。

2.1 关系型数据库软件

- 关系型数据库软件有：MySQL，Oracle，MariaDB，PostgreSQL

数据库软件	分类	简介
MySQL	开源软件	关系型数据库里面业界使用最广泛的软件，在互联网公司大量使用。 阿里巴巴有对其进行优化并部署在云端，出售服务给需要的公司或或者个人使用
Oracle	商业收费软件	关系型数据库的鼻祖和代表作，成就了甲骨文公司。 相对于MySQL而言，Oracle支持的数据量更大，对数据安全要求，周边各种设施服务也更完善，部署对硬件条件要求也更高。
MariaDB	开源软件	由MySQL发展而来，另一个分支发展。 使用广泛程度不如MySQL
PostgreSQL	开源软件	美国伯克利大学开发，数据库功能完整。 可能性能上较MySQL稍差，使用广泛程度不如MySQL
SQL Server	商业收费软件	微软开发的关系型数据库软件， 缺点是只能在Windows系统上运行。
SQLite	开源软件	存储功能简单，支持的数据量较小，无需部署到专门服务器上， 内嵌在手机或者电脑APP，使用一个本地文件存储数据， 主要用于本地存储一些简单的本地数据，例如微信的聊天记录等

备注：**开源软件** 就是源代码开放的软件，所有人都是可以看见代码，并进行自己修改和部署。代码托管在世界上最大的代码托管网站 **GitHub** 上面。 虽然是开源软件，但目前 **GitHub** 两年前被微软收购，虽然可能性较小，但未来发展还是由微软进行控盘。

2.2 非关系型数据库软件

数据库软件	分类	简介
MongoDB	开源软件	相比较MySQL而言，其优点在于可以存储复杂的嵌套的数据， 例如一篇文章（标题 作者 正文（正文里面有段落，段落里面有语句），评论等这种复杂结构信息， 并支持这种数据内部信息的查询，使用非常广泛， 成为非关系型数据库的事实标准
Redis	开源软件	严格来说，不算是数据库软件，更多地是用来作为数据缓存。 数据存储在内存中。相比较MySQL，它存取数据速度快， 但是不用作持久存储。

数据库软件	分类	简介
CouchDB	开源软件	跟MongoDB大体类似，一些细节特性上的区别

2.3 大数据

随着数据量的越来越大，越来越复杂，上面的一些软件渐渐的无法满足一些特定场景的需求，随后就发展出了分布式的大数据软件系统。这些大数据软件系统也基本都是开源软件。

因为大数据的特点是，数据密度价值低，数据量超大，需要进行挖掘发现其价值。而数据来自许多不同的领域，针对不同领域的特点，一些垂直领域专用的数据库软件、数据分析软件也有其市场。

随便列举一些如下：

数据库软件	类别	简介
Cassandra	开源软件	最开始由Facebook公司开发，后来捐赠给Apache开源社区成为开源软件。
HBase	开源软件	Hadoop是一个大数据系统，由一系列存储、分析、管理等软件构成，而HBase是其数据存储软件。
Neo4j	开源软件	Neo技术 公司开发的号称全球领先的图形数据库软件，目前开源在GitHub，但受Neo公司控盘
Lucene	开源软件	大数据 搜索引擎 。例如百度、谷歌等搜索引擎底层就是基于Lucene进行开发的
Wolfram Alpha	商业收费软件	Proessaywriting 公司开发的进行数据内容 关联搜索 的软件
flockDB	开源软件	Twitter公司为社会化网络而开发的图形数据库，虽是开源软件，但由Twitter控盘
GaussDB	开源软件	华为 开发的全球首款人工智能原生数据库，将人工智能能力植入到数据库内核算法中
Teradata	商业收费软件	来个来说不算是数据库软件，而是数据分析、数据仓库软件，由美国Teradata公司开发，并提供完整的解决方案，我国政企、大中华区已有超过十多个行业的客户，包含政府与公共服务、通信行业、银行、保险、证券、航空、地铁、交通运输、物流、快递行业、制造行业、汽车、零售、电子商务、互联网、电力能源等等。

3 总结

3.1 关于差距

总体来说，我国和美国的差距是很大的，主要表现在数据库软件数量上。我国很少有开发一些知名数据库软件。但目前主流的数据库软件基本都是开源软件，中国大部分公司使用的也都是这种类型的软件，我国程序员也渐渐越来越多参与到开源软件的开发当中，这个差距正在慢慢缩小，至少是在数据库技术能力上的差距在慢慢缩小。

3.2 关于卡脖子

而目前开源软件的协议是开放的，所有人均可使用。受美国卡脖子的概率较小，除非开源软件协议被强行修改。(有极小概率可能，但即使发生，问题也不大，已有代码还是可以使用，只是不能进行和社区同步升级而已。)

关系型数据库软件中，Oracle和SQL Server被MySQL等开源软件的可替代性很高，不用太担心这两个数据库软件的断供。

一些垂直领域的商业软件，更多的掌握在美国商业公司手里，被卡脖子的概率更大一些。但是这些软件大都是收费软件，使用并不算广泛。

3.3 全球格局

数据库软件已经比较成熟。特别是开源软件已经足够多，而要想开发一款新的能够脱颖而出的数据库软件，难度较大，成本很高，所以越来越少的公司专门投入开发新的付费数据库软件。

更多的是采用开源社区化形式进行开发。更多的是倾向于开发垂直领域的商业数据分析软件。