

# 全球最给力的10大数据库

转载

weixin\_30527323

于 2011-05-03 09:48:00 发布

1399

收藏 1

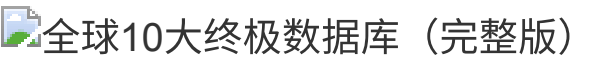
版权

文章标签：

数据库

### 1. 世界气候数据中心（WDCC）

如果你有一台价值3500万欧元的超级计算机，你会把它用来做什么？炒股？搭建自己的“因特网”？WDCC就有一台这样的机器，又逢全球变暖，正好用它来做气候研究。WDCC由马克思·普朗克气象研究院和德国气候计算机中心共同运作，其数据库是全球最大的。



WDCC有220TB的网络数据，包括气候研究、预测气候变化和110TB（24500张DVD）的气候模拟数据。尤为重要的是，它还有保存在磁带上的6PB的其他信息。你会问，这6PB的数据有多大呢？这个数据是全美所有学术研究图书馆内容总和的3倍。

WDCC数据库概括：

- 220TB的网络数据
- 6PB的其他数据

### 2. 美国国家能源研究科学计算中心（简称NERSC）

第二大的数据库应属加州奥克兰的NERSC。NERSC由“劳伦斯伯克利能源实验室”和“美国能源部”共同运作。它数据库保存的信息包括了原子能研究、高能物理实验和早期宇宙模拟等数据。如果想看宇宙过去的模样，启动NERSC的超级计算机就可以观看“大爆炸”时期的宇宙形态。



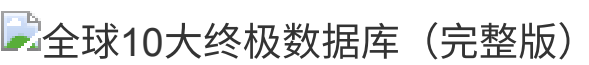
NERSC数据库由2000多名计算机科学家操作和维护，其数据量高达2.8PB。（注：1 PB = 1024 TB；1 TB = 1024 GB；）

NERSC数据库概括：

- 信息量高达2.8PB
- 由2000多名计算机科学家共同运作

### 3. AT&T 美国电话电报公司

和Sprint一样，美国的老牌通信公司AT&T的数据库也能进入排名。从结构上来说，AT&T的最大数据库是百里挑一的，因为它有2个重要“头衔”，总量最大的独立数据库（312TB）和行数第二大的独立数据库（1.9万亿行，这个数据库是处理呼叫记录的）。



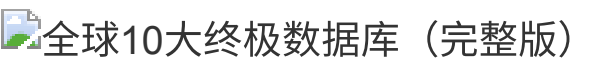
1.9亿条呼叫记录，包括了主叫和被叫号码、呼叫时间和通话时长以及其他各种账单目录信息。AT&T的工作做得非常细致，他们业保存了10年前的呼叫信息。要知道，10年前可没有保存数百TB级数据的技术。（或者说技术还不成熟。）

AT&T数据库概括：

- 信息量为323TB
- 1.9万亿的通话记录

### 4. Google

尽管目前外界没有太多关于Google的数据库真实大小的信息（Google一直能严守信息，这让诺克斯堡军事基地情何以堪。），但我们已知道Google所收集信息的数量和种类。



Google平均每天处理9100万次查询（Google的处理海量请求的能力是基于其庞大且强大的数据中心），这将近占整个互联网每日查询总量的50%。Google把用户的每一次搜索保存在数据库中。所以，Google平均一年要保存33万亿条查询记录。根据Google数据的结构类型，一年的查询总量将折合成数百TB。

除了查询记录，Google还收集用户的个人信息。（怎么收集？）Google把用户的查询请求和用户电脑上Cookie保留的信息整合到一起，创建虚拟的个人信息。

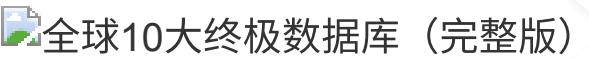
尤为重要的是，除了搜索，Google帝国的疆土还在不断扩大，现有疆土已有数字媒体（Google视频和YouTube）、广告（Google Ads）和邮箱等。归根结底，Google的疆土越大，其数据库处理的信息量就越大。就互联网范围内的数据库而言，Google是No 1。

Google数据库概括：

- 平均每天9100万查询请求
- 查询数量占所有互联网查询量的50%
- 不计其数的用户虚拟资料

### 5. Sprint 斯普林特

Sprint是全球最大的通信公司之一，向5300万客户提供服务。在和Nextel合并之前，提供本地和长途通信服务。（Sprint Nextel公司现已是全美第3大通信公司，仅次于Verizon和AT&T。）



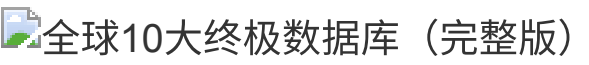
像Sprint这样的大型通信公司，均有庞大的数据库，以追踪客户呼叫请求。Sprint的数据库每天要处理多达3.65亿次呼叫记录。其数据库已扩展到2.85万亿行。在顶峰期间，每秒有7万次呼叫详细记录新增到数据库中。

Sprint数据库概括：

1. 数据库有2.85万亿行
2. 每日平均处理3.65亿个呼叫详细记录
3. 顶峰期间，每秒插入7万次呼叫详细记录

6. ChoicePoint

设想一下：在一本10亿页的电话本上找一个电话号码。这是一项什么的工作。当你有求于ChoicePoint的时候，ChoicePoint 的员工就得干这活。如果把ChoicePoint的数据库资料全部打印出来，在不这些纸张连起来，其长度足以在地月之间往返77回。



ChoicePoint是一家信息服务机构，你可以通过它查询美国的人口数量、地址和电话号码、驾驶记录和犯罪记录等。在大多数情况下，Choicepoint数据库中的数据是只卖出出价最高的买主，包括美国政府。

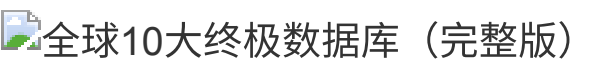
ChoicePoint到底知道多少呢？2002年，借助ChoicePoint数据中的DNA和个人记录数量，政府成功破获了费城和柯林斯堡的一系列强/奸案。2001年，世贸中心有很多遇难者已无法辨别身份，在提取遇难者的DNA后，通过比对ChoicePoint数据库中的遇难者家属的数据，才得以确认遇难者身份。

ChoicePoint数据库概括：

1. 个人数据总量为250TB
2. 信息覆盖2.5亿人

7. YouTube

在经过这些年的积累发展，YouTube已经成为全球最大的视频库（视频网站）。如今，YouTube上每日视频播放次数高达1亿次，60%的视频是在线观看的。



2006年8月，《华尔街日报》称YouTube的视频总量已有45TB。相对互联网上数据量来说，这个数字听起来不是很高，但自从这个数字公布以来，YouTube正经历了一段大幅增长时期，（每天有65000段视频上传至YouTube，）所以在过去五个月中，YouTube的数据库大小很可能已经翻倍了。

因为YouTube上的每段视频的大小不一，所以要估计YouTube的数据库大小并非易事。不过，也可以大胆来预测一下：

每天上传6500段视频，故每个月共有 65000 \* 30 = 1,950,000段视频；

假设每段视频大小为1MB，每月将有 1.95TB；

假设每段视频大小为10MB，每月将有19.5TB。

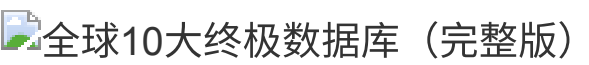
所以，YouTube下个月将增加近20TB数据。

YouTube数据库概括：

1. 每日的视频播放量为1亿次
2. 每日新增65000段视频
3. 60%的视频是在线观看的
4. 视频总量大小至少是45TB

8. Amazon 亚马逊

Amazon，全球最大的零售网站，“收藏”了5900万活跃用户的信息，包括个人信息（电话号码等）、收据/发票、清单和还有其他能从用户那提取的数据。Amazon也有25万多的在线书籍（可以全文阅读的），并允许用户在其各个页面进行评论和交互。因此，Amazon跻身为全球最大的在线社区之一。



（不知你有注意Amazon的Logo上的箭头么？ 有2层意思： 1.从A到Z是指其物品繁多； 2.笑脸。）

上述这些数据和Amazon每年的数百万巨额销售量，再加上Amazon合伙人的巨额销售量，造就了一个超大数据库。Amazon最大的2个数据库的数据量加起来有42TB，这还只是开始。如果Amazon公开其所有的数据库数量和每个数据库的数据量，这就不仅是42TB了。

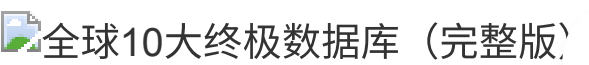
尽管如此，我们暂且就算它42TB吧。42TB数据，如果换算成论坛帖子的话，应该是37万亿个帖子。

Amazon数据库概括：

1. 5900万活跃用户
2. 数据量超过42TB

9. 美国中央情报局（CIA）

CIA的任务（之一）就是收集和散发各种信息和资料，所以CIA能上榜就不足为奇了。虽然CIA数据库的总量外界不得而知，但下面这个说法是十分肯定的。这个机构一定收集了大量的公共信息和私人信息。



CIA数据库对公众开放的部分有：“信息自由法”电子阅览室、“世界概况”和“世界情报相关出版物。在“信息自由法”电子阅览室中，公众可以看到成千上万的美国政府官方文档（也有已解密的文档），电子阅览室每月新增100份文档，文档主题内容范围很广，从巴基斯坦核发展到“韩战”（西方皆称“韩战”，大陆称“抗美援朝”）中的啤酒种类。“世界概况”收藏了各个国家和地区的信息，包括地图、人口数量和军事能力等。

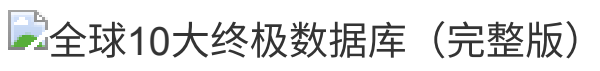
CIA数据库概括：



- 1.电子阅览室每个月新增100份文档；
- 2.覆盖全球250多个国家和地区的统计数据；
- 3.数目不详的机密资料

10. 美国国会图书馆

不管是不是在数字时代，全球最大的图书馆——美国国会图书馆都可以荣登这一列表。国会图书馆的藏书数量之高——1.4亿多册，范围之广——从烹饪书籍到殖民地时期的报纸到美国政府刊物。据估计，如果把国会图书馆的文字以数字形式存储，其总量将达20TB。



如果你（特指老美）在互联网上找东西又找不到时，那么国会图书馆应是你要去的地方。对于研究美国历史的用户来说，国会图书馆是个必去之处，用户访问国会图书馆网站的“美国记忆”目录，该目录下大约有5百万册相关资料。

但不幸的是，国会图书馆并无计划把所有的馆藏内容都数字化，并且有些馆藏是限制查阅人群，只对最高法院法官、国会议员、馆内工作人员和其他政府官员。但是，只要你有国会图书馆的借阅卡，就可以借阅（大部分）资料。（编者注：该图书馆的资料不能外借，读者只能在馆内查阅。）

国会图书馆数据库概括：

- 1.馆藏资料总量1.3亿份（包括书籍、照片和地图等）
2. 2900万册书
- 3.每天新增1万馆藏
- 4.书架长度共有530英里
5. 5百万份数字文档
- 6.文字数据总量20TB

转载于:https://www.cnblogs.com/Wolves/archive/2011/05/03/2034945.html

🔍 文章知识点与官方知识档案匹配，可进一步学习相关知识

MySQL入门技能树 > 数据库组成 > 表 41724 人正在系统学习中

📖 相关资源：浅析MongoDB数据库的海量数据存储应用\_mongodb应用案例-其它代码...

<div>全球国家数据数据库详细</div> <div>MySQL版本的全球各个国家的数据，可用，详细，放心下载（修改下就能引用到其他数据库）</div>	01-23
<div>企业财务管理升级，智慧税务和数据可视化打造新标准 <div>最新发布</div></div> <div>可视化将各类数据存入各级数据库，同时将数据库统一管理，提高分析和决策的效率，使大量数据库和信息存储用于税收的事务管理、信...</div> <div>二三维数据可视化/数字孪生 349</div>	
<div>Rexsee API介绍:CallLog通话记录_dblvx22880的博客</div> <div>rexseeCallLog对象,实现自定义数量的最近通话记录,自定义查询最近通话,读取记录数据库表的URI地址。。【函数】 JsonObjectArray get...</div>	10-10
<div>获取iPhone通话记录call_history.db(需越狱)_ios开发 获取通话记录_胖次...</div> <div>越狱后的手机的数据库文件可以自由访问,通话记录通常保存在call_History.db这个文件中,只要读取这个文件,我们就能知道目前手机的通话...</div>	3-7
<div>Android通话记录数据库表分析</div> <div>数据库 实现在ContactsProvider中，具体是packages/providers/ContactsProvider/src/com/android/providers/contacts/CallLogProvider...</div> <div>firedancer0089的专栏 3272</div>	
<div>最快的数据库查询工具、最方便的数据迁移工具</div> <div>最快的数据库查询工具 最快的数据库查询工具，可独立运行。 有单表查询、透视表查询、树_列表查询、图表查询、多表级联查询，归档...</div> <div>runsheng678的博客 350</div>	
<div>【Android】Android 联系人数据库浅析之通话记录_moto0421的博客-CSD...</div> <div>1. 删除通话记录 2. 负责存放呼叫记录的内容提供者源码在 ContactsProvider 项目下: 源码路径: com/android/providers/contacts/CallLog...</div>	1-21
<div>Hummer TimeSeries DB (蜂鸟时序数据库)技术介绍_白乔的博客-CSDN博 ...</div> <div>面对如此大量的时序数据,传统数据库既“存不下”,也“查不出”(原因见 2.2/2.3/2.4)。因此当前业界普遍采用的妥协方案是:或者只存近期数据...</div>	3-11
<div>一个app开发需要多少钱</div> <div>app开发费用</div>	
<div>史上最强数据库总结，看完绝不后悔！！</div> <div>数据库总结 一、数据库简介 1、数据库简介 数据库（Database，DB）是按照数据结构来组织，存储和管理数据的仓库。典型特征：数据...</div> <div>He Xin Fan 925</div>	
<div>数据库索引</div> <div>数据库索引用于加速查询 虽然哈希索引是O(1)，树索引是O(log(n))，但SQL有很多“有序”需求，故数据库使用树型索引 InnoDB不支持哈希...</div> <div>ciap37959的博客 34</div>	
<div>Android 通讯录开发之通话记录_sp_StudyAndoird的博客</div> <div>删除通讯录记录 ![]这个 是通讯录数据库contacts的通话记录calls表[https://img-blog.csdn.net/20150504105101656] 其中: 这里是删除某个...</div>	2-28
<div>分析通话记录数据库_pan_blog的博客</div> <div>数据表字段分析 问题:calls表中记录的是每一条通话记录,但是当我通过手机删除一条通话记录,或者使用代码删除通话记录发现表格中的记...</div>	2-18
<div>全球前10大数据库产品厂家</div> <div>南大通用分析型数据库GBase 8a是支撑大数据快速分析的新型数据库，作为列存储的分析型数据库，南大通用Gbase还是有其独到之处...</div> <div>喜欢打酱油的老鸟 2715</div>	
<div>盘点 10大 数据库！</div> <div>大家好，我是小猿???DB-Engines 最近发布了 2021 年 9 月份的数据库排名。该网站根据数据库管理系统的受欢迎程度对其进行排名，实...</div> <div>程序IT圈 1187</div>	
<div>2016全球最强数据库大盘点</div> <div>数据库排名年度盘点 说到盘点，首先肯定得看看DB-Engines的全球数据库排名。下表是2017年1月份前20名数据库引擎最新排名。 DB-En...</div> <div>weixin_34081595的博客 798</div>	
<div>关于世界上的超大数据库</div> <div>&amp;#13;手边有一份 2005 Winter TopTen Award Winners的报告，包含了一些关于世界上排名前几位的VLDB 的信息。VLDB，超大数据库...</div> <div>Feng's Blog 376</div>	
<div>信息安全的 CIA 三要素 <div>热门推荐</div></div> <div>互联网就像一个虚拟的社会。在它诞生的初期，互联网的应用相对简单，使用互联网的人数较少，人们对安全的设计与考虑都比较少。经...</div> <div>天天向尚的专栏 3万+</div>	
<div>Android 7.1.1 通话记录数据库详解</div> <div>在拨打或接听来电时，都会在Dialer中看到新增的通话记录，前两天遇到了个通话记录显示得不对的问题，就跟踪了下Android中C...</div> <div>Do Better Every Day 1万+</div>	
<div>全球主流数据库</div> <div>简单介绍全球主流数据库。</div> <div>喝醉酒的小白 3414</div>	
<div>SQLlist数据库入门（安卓通讯录填充）--未完</div> <div>SQLlist工具是我们手机上的数据库，存放联系人、通话记录之类的，使用的也是SQL语言，我们导入一个公网联系人试试 从终端导出的...</div> <div>测试领域 举报 领域测试技术、管理、工具都略懂... 1163</div>	
<div>世界上最大的十个数据库:</div> <div>世界上最大的十个数据库: 1.全球气象数据中心： 220千兆网络数据，6个petabytes的其它数据。 2.全美能源研究科技计算中心： 2.8个pe...</div> <div>weixin_34026276的博客 886</div>	
<div>2018年12月全球数据库排行榜：Oracle惨不忍睹！</div>	代码技巧 1660

有哪些一般人不知道的数据获取方式

DataCastle 3万+

想要对一个新的领域进行研究和探索，拥有这个领域的数据库那都是有十分重要的意义的。 在这里给大家推荐一些能够用上数据获取方式...

全世界各个国家地区城市 数据库

07-20

本字典库为世界各国地区城市的数据，等级关系为 洲-国家-省（州）-市

中国能源数据库—年度数据（2006-2016）.xls

10-17

中国能源数据库—年度数据（2006-2016），省级的，包括能源类别，可以下载做数据分析

“相关推荐”对你有帮助么？

- 非常没帮助
- 没帮助
- 一般
- 有帮助
- 非常有帮助

