

数据库

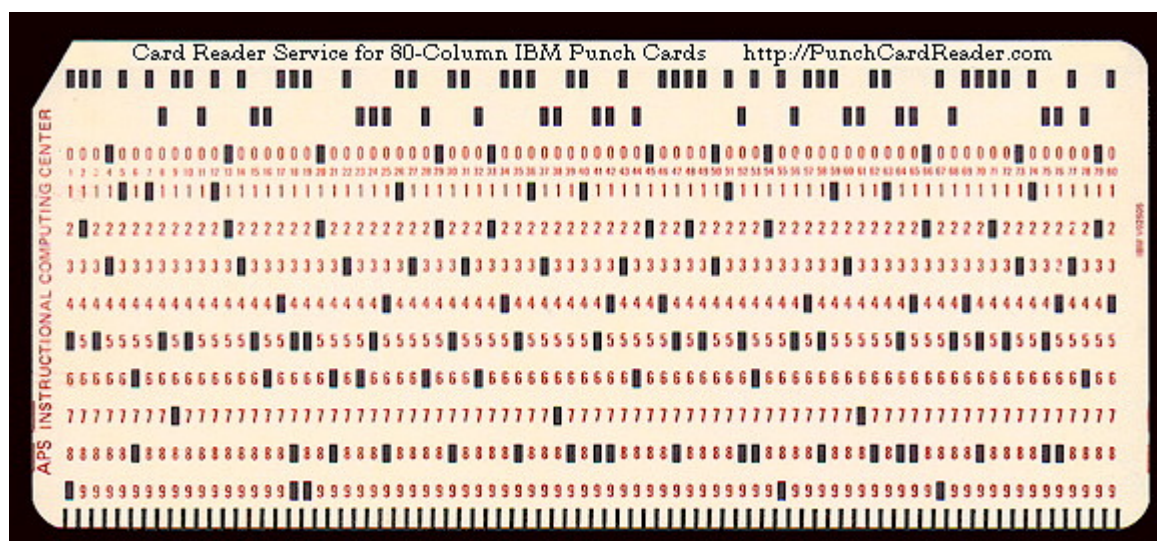
概念

数据库：按照数据结构来组织、存储、管理数据的仓库。

诞生

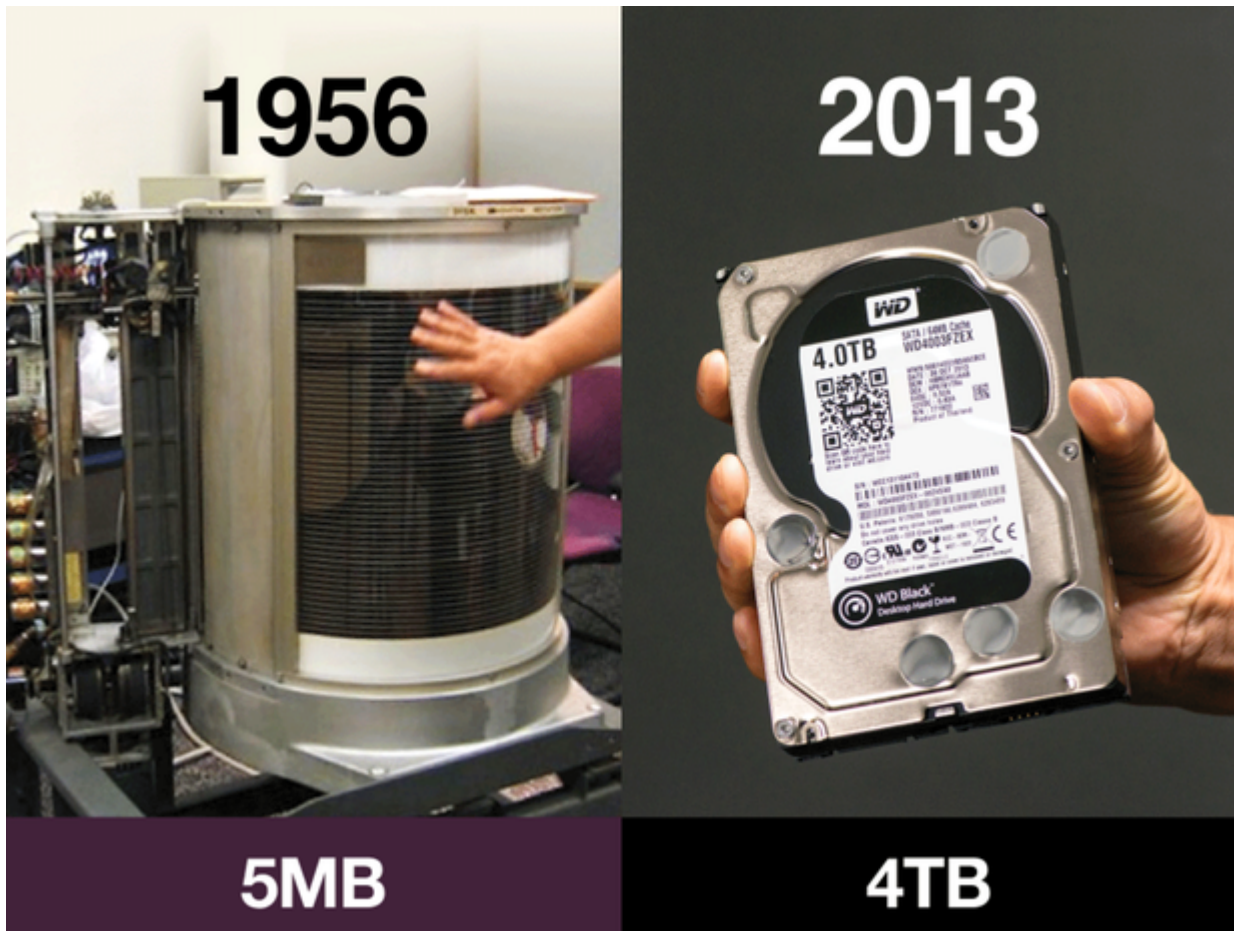
计算机的发明是为了做科学计算的，而科学计算需要大量的数据输入和输出。

早期，可以使用打孔卡片的孔、灯泡的亮灭来表示数据输入、输出。



1940年，数据可以存储在磁带上，顺序的读取、写入磁带。

1956年IBM发明了磁盘驱动器这个革命性产品，支持随机访问。



随着信息化时代的到来，有了硬件存储技术的发展，有大量的数据需要存储和管理。

数据库DBMS发展

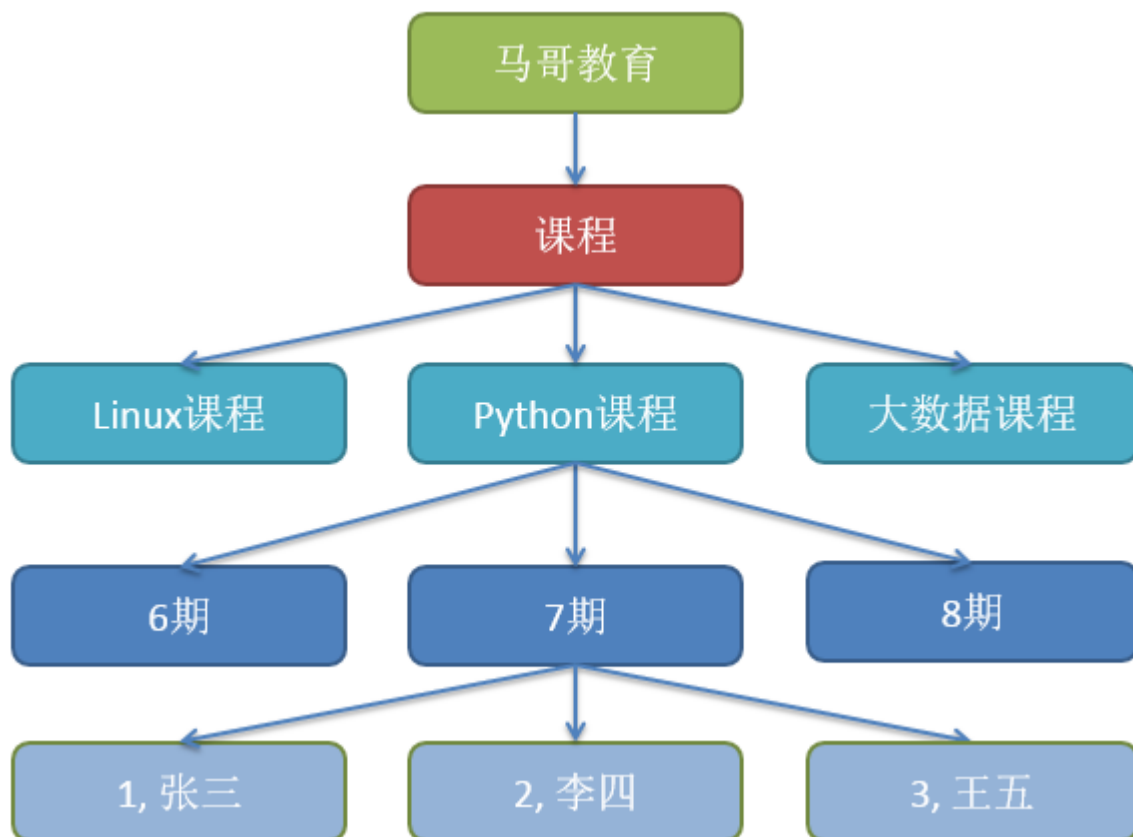
- 萌芽期：文件管理
- 第一代：层次数据库、网状数据库
- 第二代：SQL、关系型数据库
- 第三代：面向对象的DBMS（OODBMS）、对象关系的DBMS（ORDBMS）

文件系统管理

- 磁盘上一个一个文件，数据孤立，数据冗余
- 格式不统一，很难统一管理
- 无法高效查询，无法灵活查询

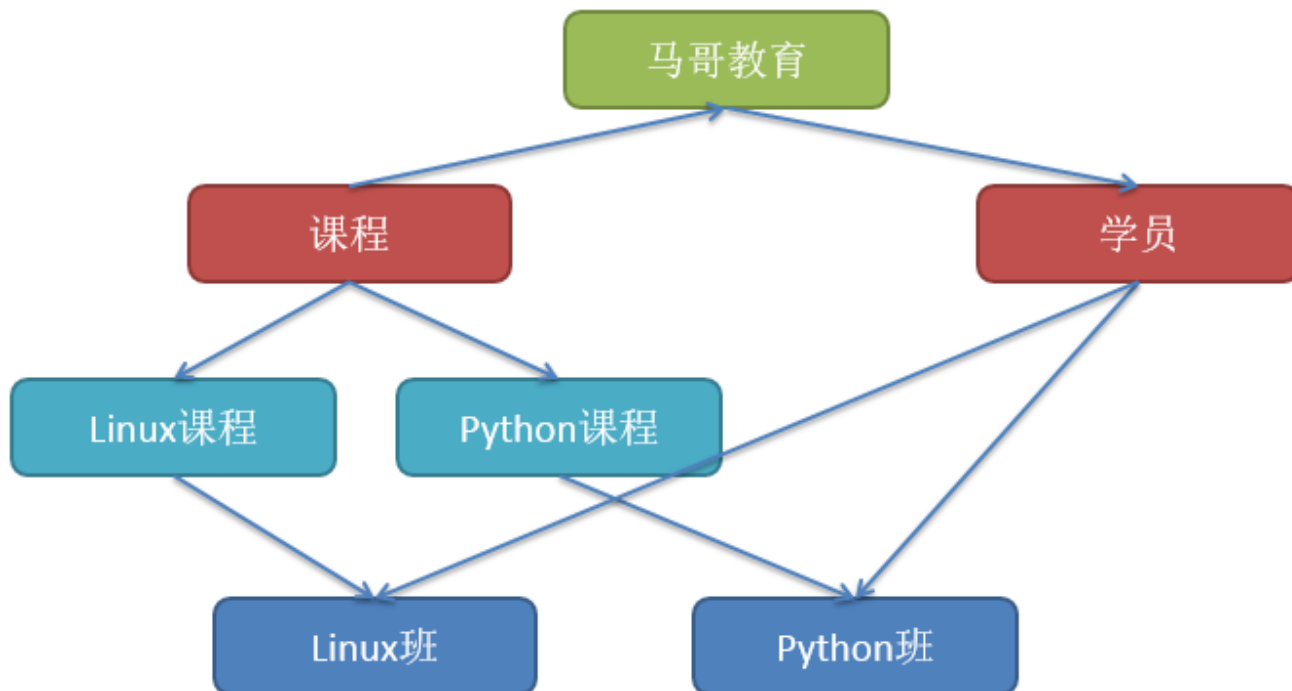
层次数据库

以树型结构表示实体及其之间的联系。关系只支持一对多。
代表数据库IBM IMS。



网状数据库

通用电气最早在1964年开发出网状数据库IDS，只能运行在GE自家的主机上。

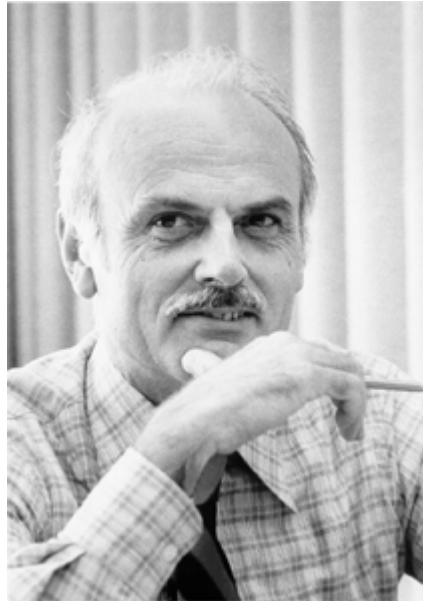


结点描述数据，结点的联系就是数据的关系。

能够直接描述客观世界，可以表示实体间多种复杂关系，而这是层次数据模型无法做到的。比如，一个结点可以有多个父结点，结点之间支持可以多对多关联。

关系数据库

使用**行、列**组成的**二维表**来组织数据和关系，表中**行（记录）**即可以描述数据**实体**，也可以描述实体间**关系**。关系模型比网状模型、层次模型更简单，不需要关系数存储的物理细节，专心于数据的逻辑构建，而且关系模型有论文的严格的数学理论基础支撑。



1970年6月，IBM的研究员E.F.Codd发表了名为“A Relational Model of Data for Large Shared Data Banks”的论文，提出了关系模型的概念，奠定了关系模型的理论基础。

1976年，IBM实验室System R项目，通过实现数据结构和操作来证明关系模型实用性，并直接产生了**结构化查询语言SQL**。1987年，SQL被ISO组织标准化。

关系模型，有严格的数学基础，抽象级别较高，简单清晰，便于理解和使用。

经过几十年的发展，关系数据库百花齐放，技术日臻成熟和完善。

基于关系模型构建的数据库系统称为RDBMS(Relational DataBase System)。

IBM DB2、Oracle的Oracle和Mysql、微软的MS SQL。以前的Infomix、Sybase等。

Oracle的发展

拉里·埃里森(Larry Ellison)仔细阅读了IBM的关系数据库的论文，敏锐意识到在这个研究基础上可以开发商用软件系统。他们几个创始人决定开发通用商用数据库系统Oracle，这个名字来源于他们曾给中央情报局做过的项目名。

1979年发布了ORACLE 2.0版本（实际上是1.0）。1983年，Oracle v3。1984年，Oracle v4。1985年，Oracle v5。1988年，Oracle v6引入了行级锁等新技术，然而这是个不稳定的版本。直到1992年的时候，Oracle7才逐渐稳定下来，并取得巨大成功。2001年的9i版本被广泛应用。

2009年4月20日，甲骨文公司宣布将以每股9.50美元，总计74亿美金收购SUN（计算机系统）公司。2010年1月成功收购。

2013年，甲骨文超过IBM，成为继微软之后的全球第二大软件公司。

Mysql发展

1985年几个瑞典人为大型零售商的项目设计了一种利用索引顺序存取数据的软件，这就是MyISAM的前身。1996年，MySQL 1.0发布，随后发布了3.11.1版本，并开始往其它平台移植。2000年MySQL采用GPL协议开源。MySQL 4.0开始支持MyISAM、InnoDB引擎。2005年10月，MySQL 5.0成为里程碑版本。

2008年1月被Sun公司收购。

2009年1月，在Oracle收购MySQL之前，Monty Widenius担心收购，就从MySQL Server 5.5开始一条新的GPL分支，起名MariaDB。

MySQL的引擎是插件化的，可以支持很多种引擎：
MyISAM，不支持事务，插入、查询速度快。
InnoDB，支持事务，行级锁，MySQL 5.5起的默认引擎

去IOE

它是阿里巴巴造出的概念。其本意是，在阿里巴巴的IT架构中，去掉IBM的小型机、Oracle数据库、EMC存储设备，取而代之使用自己在开源软件基础上开发的系统。传统上，一个高端大气的数据中心，IBM小型机、Oracle数据库、EMC存储设备，可以说缺一不可。而使用这些架构的企业，不但采购、维护成本极高，核心架构还掌握在他手中。

对于阿里巴巴这样大规模的互联网应用，应该采用开源、开放的系统架构。这种思路并不是阿里巴巴的新发明，国外的谷歌、Facebook、亚马逊等早已为之。只不过它们几乎一开始就没有采用IT商业公司的架构，所以他们也不用“去IOE”。

去IOE，转而使用廉价的架构，稳定性一定下降，需要较高的运维水平解决。

NoSQL

NoSQL是对非SQL、非传统关系型数据库的统称。
NoSQL一词诞生于1998年，2009年这个词被再次提出指非关系型、分布式、不提供ACID的数据库设计模式。

随着互联网时代的到来，数据爆发式增长，数据库技术发展日新月异，要适应新的业务需求。
随着移动互联网、物联网的到来，大数据的技术中NoSQL也同样重要。

数据库流行度排名 2017.12

Rank			DBMS	Database Model	Score		
Dec 2017	Nov 2017	Dec 2016			Dec 2017	Nov 2017	Dec 2016
1.	1.	1.	Oracle +	Relational DBMS	1341.54	-18.51	-62.86
2.	2.	2.	MySQL +	Relational DBMS	1318.07	-3.96	-56.34
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1172.48	-42.59	-54.17
4.	4.	4.	PostgreSQL +	Relational DBMS	385.43	+5.51	+55.41
5.	5.	5.	MongoDB +	Document store	330.77	+0.29	+2.09
6.	6.	6.	DB2 +	Relational DBMS	189.58	-4.48	+5.24
7.	7.	8.	Microsoft Access	Relational DBMS	125.88	-7.43	+1.18
8.	9.	9.	Redis +	Key-value store	123.24	+2.05	+3.34
9.	8.	7.	Cassandra +	Wide column store	123.21	-1.00	-11.07
10.	10.	11.	Elasticsearch +	Search engine	119.78	+0.37	+16.51
11.	11.	10.	SQLite +	Relational DBMS	115.19	+2.44	+4.36
12.	12.	12.	Teradata	Relational DBMS	74.74	-3.49	+1.37
13.	13.	14.	Solr	Search engine	66.30	-2.86	-2.70
14.	14.	13.	SAP Adaptive Server	Relational DBMS	65.68	-1.35	-4.74
15.	15.	16.	Splunk	Search engine	63.79	-1.08	+8.87
16.	16.	15.	HBase	Wide column store	63.41	-0.15	+4.79
17.	18.	20.	MariaDB +	Relational DBMS	56.73	+1.44	+12.64
18.	17.	17.	FileMaker	Relational DBMS	55.20	-3.64	+1.08
19.	19.	19.	Hive +	Relational DBMS	54.67	+1.42	+5.27

数据库流行度排名 2018.6

Rank			DBMS	Database Model	Score		
Jun 2018	May 2018	Jun 2017			Jun 2018	May 2018	Jun 2017
1.	1.	1.	Oracle	Relational DBMS	1311.25	+20.84	-40.51
2.	2.	2.	MySQL	Relational DBMS	1233.69	+10.35	-111.62
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1087.73	+1.89	-111.23
4.	4.	4.	PostgreSQL	Relational DBMS	410.67	+9.77	+42.13
5.	5.	5.	MongoDB	Document store	343.79	+1.67	+8.79
6.	6.	6.	DB2	Relational DBMS	185.64	+0.03	-1.86
7.	7.	9.	Redis	Key-value store	136.30	+0.95	+17.42
8.	9.	11.	Elasticsearch	Search engine	131.04	+0.60	+19.48
9.	8.	7.	Microsoft Access	Relational DBMS	130.99	-2.12	+4.44
10.	10.	8.	Cassandra	Wide column store	119.21	+1.38	-4.91
11.	11.	10.	SQLite	Relational DBMS	114.26	-1.19	-2.44
12.	12.	12.	Teradata	Relational DBMS	75.77	+1.36	-1.55
13.	14.	18.	MariaDB	Relational DBMS	65.85	+0.85	+12.95
14.	13.	16.	Splunk	Search engine	65.78	+0.68	+8.26
15.	15.	14.	Solr	Search engine	62.06	+0.55	-1.55
16.	16.	13.	SAP Adaptive Server	Relational DBMS	61.49	-0.02	-6.04
17.	17.	15.	HBase	Wide column store	59.70	-0.25	-2.17
18.	18.	20.	Hive	Relational DBMS	57.33	+0.36	+12.95
19.	19.	17.	FileMaker	Relational DBMS	56.18	+1.51	-0.90