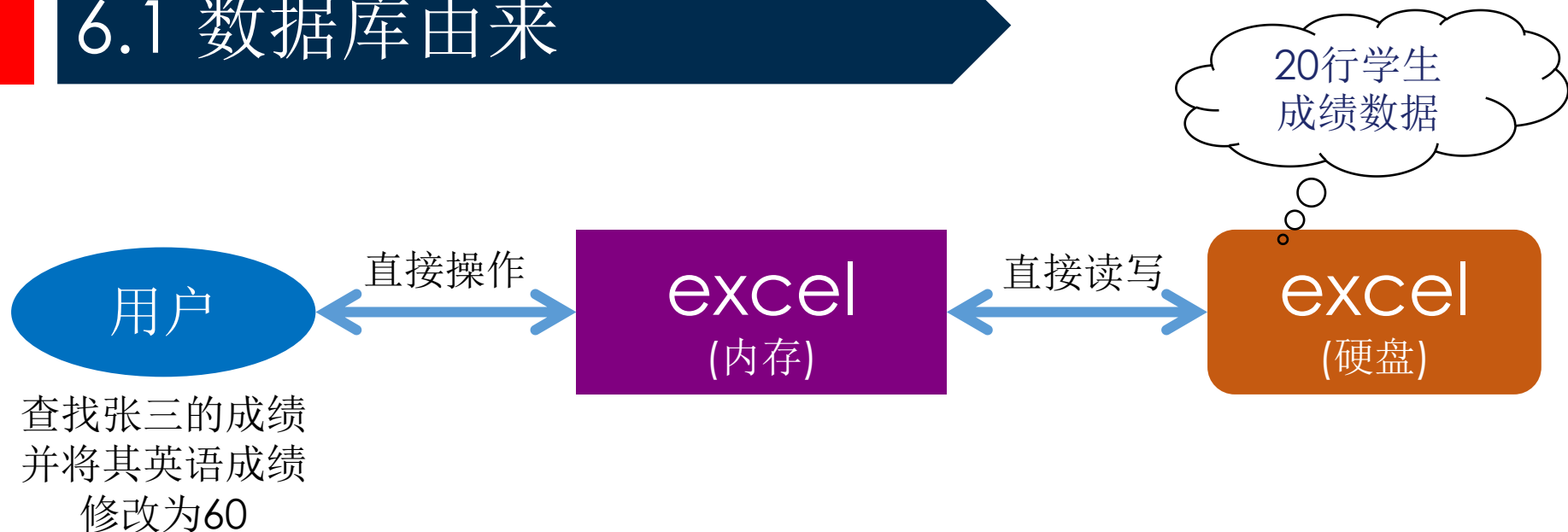




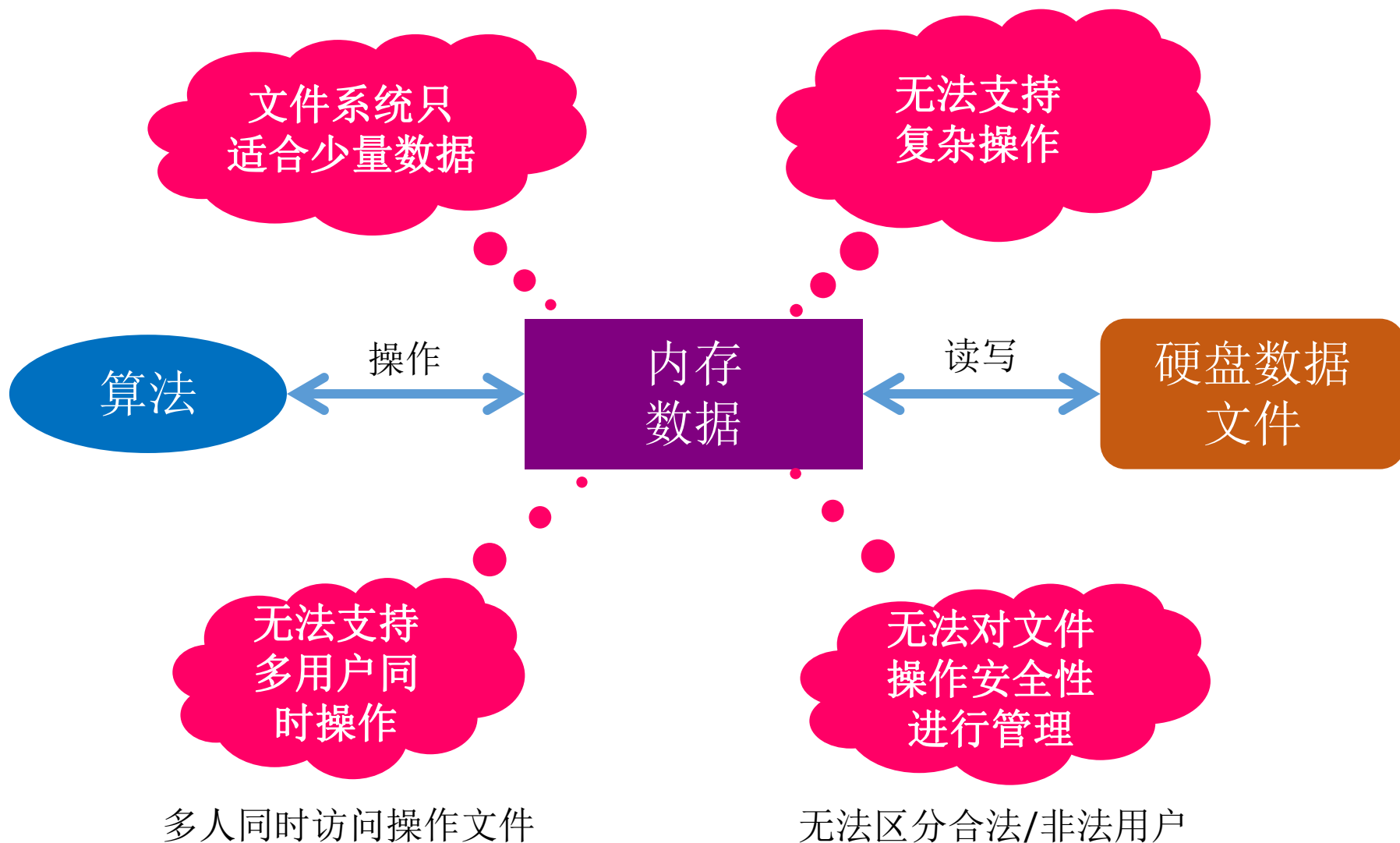
第6章 千里江陵一日还—数据库理论基础

6.1 数据库由来

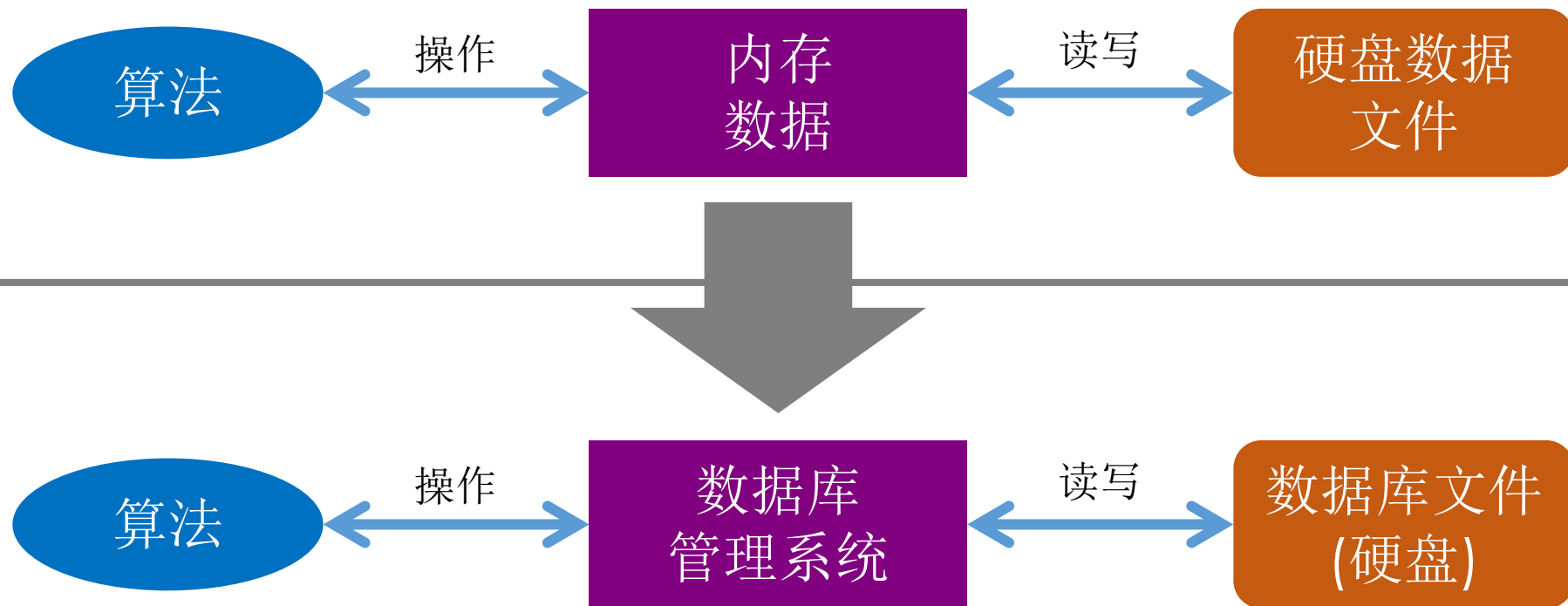


6.1 数据库由来

查找英语分数在40-50之间且高数分数在70分以上的学生

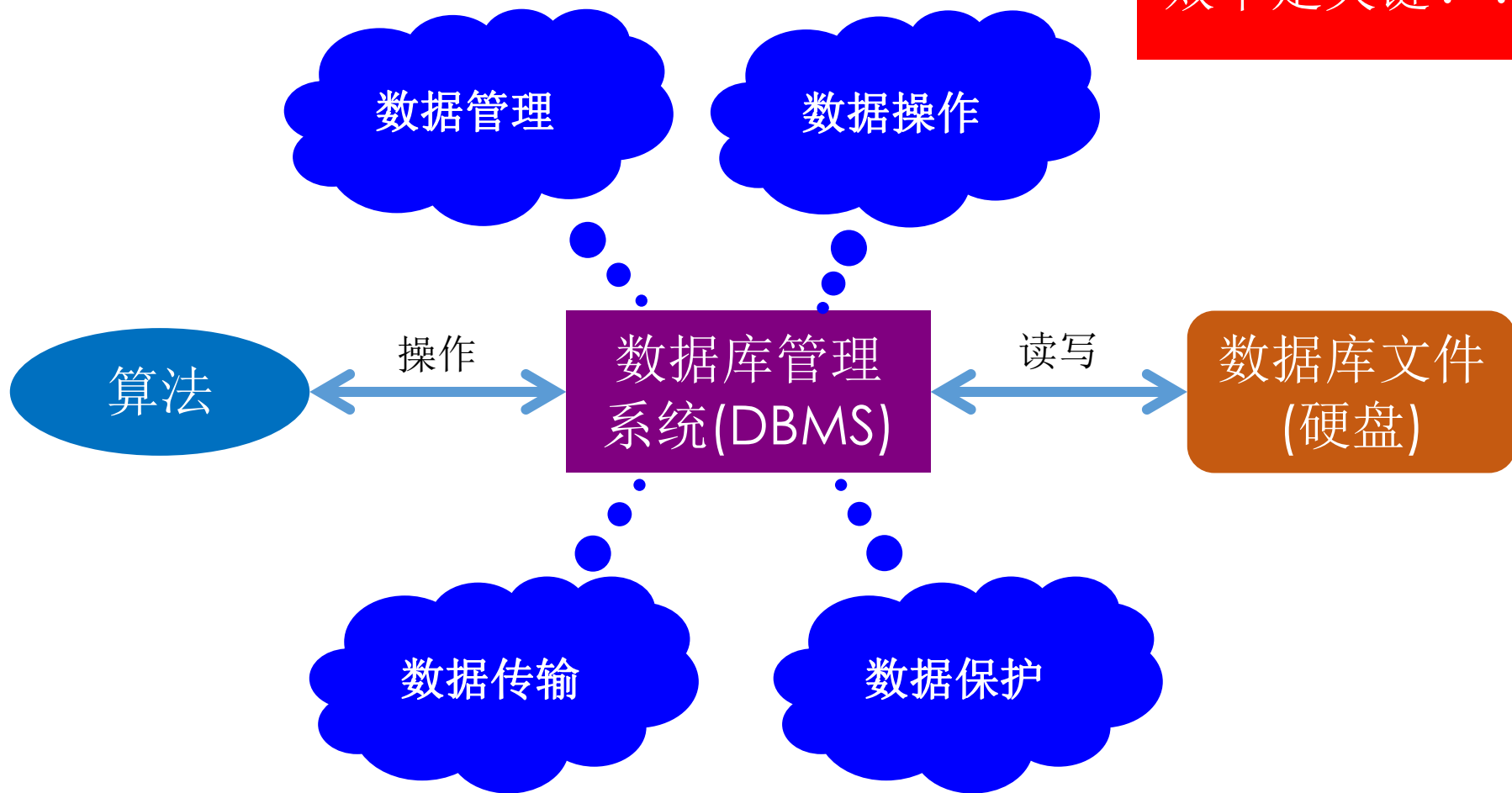


6.1 数据库由来



6.1 数据库由来

效率是关键！！



6.2 数据库发展历程

20世纪50年代,随着计算机的诞生和成熟,计算机开始运用于数据管理,与此同时,数据管理技术也迅速发展。传统的文件系统难以应对数据增长的挑战,也无法满足多用户共享数据和快速检索数据的需求。在这样的背景下,20世纪60年代,数据库应运而生。

而数据库管理系统(Database Management System,DBMS)的设计与实现取决于数据的存储结构(物理结构),即数据是如何存储的?数据之间的关系是如何表达的?

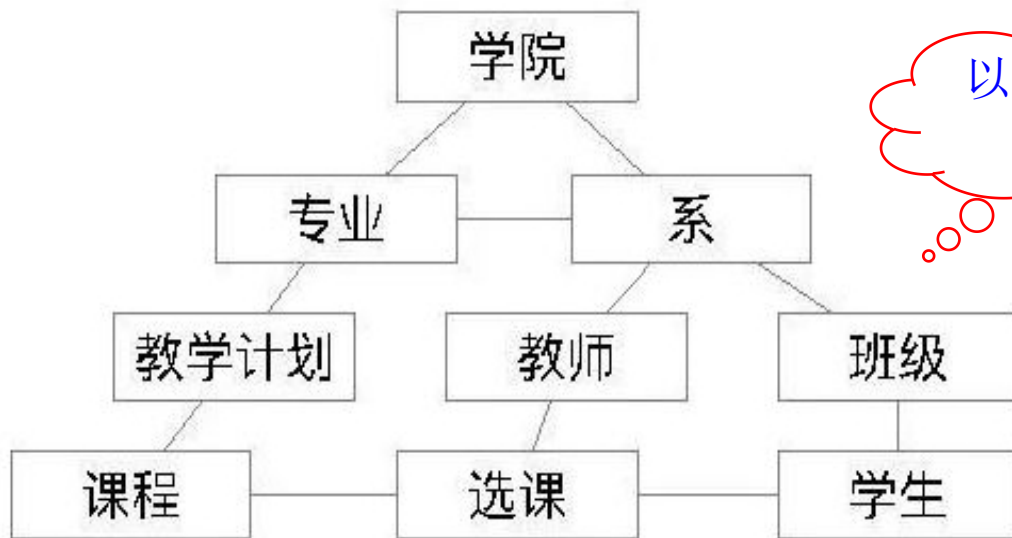


6.2 数据库发展历程

● 网状数据库

网状数据库是数据库历史上的第一代产品,它成功地将数据从应用程序中独立出来并进行集中管理。网状数据库基于网状数据模型建立数据之间的联系,能反映现实世界中信息的关联,是许多空间对象的自然表达形式。

1964年,世界上第一个数据库系统—**IDS(Integrated Data Storage, 集成数据存储)**诞生于**通用电气公司**。5年后,美国数据库系统语言协会(Conference on Data Systems Languages, CODASYL)下属的数据库任务组(Database Task Group, DBTG)发布了一份报告,阐述了网状数据库系统的许多概念、方法和技术,成了网状数据库的代表。



以教学管理
为例

6.2 数据库发展历程

● 网状数据库

在20世纪70年代与80年代初,网状数据库系统十分流行,在数据库系统产品中占据主导地位。例如,配备在富士通公司M系列机上的**AIM**(Advanced Information Manager)系统和配备在UNIVAC(UNIVersal Automatic Computer)上的**DMS1100**系统都是网状数据库系统。



Charles W. Bachman

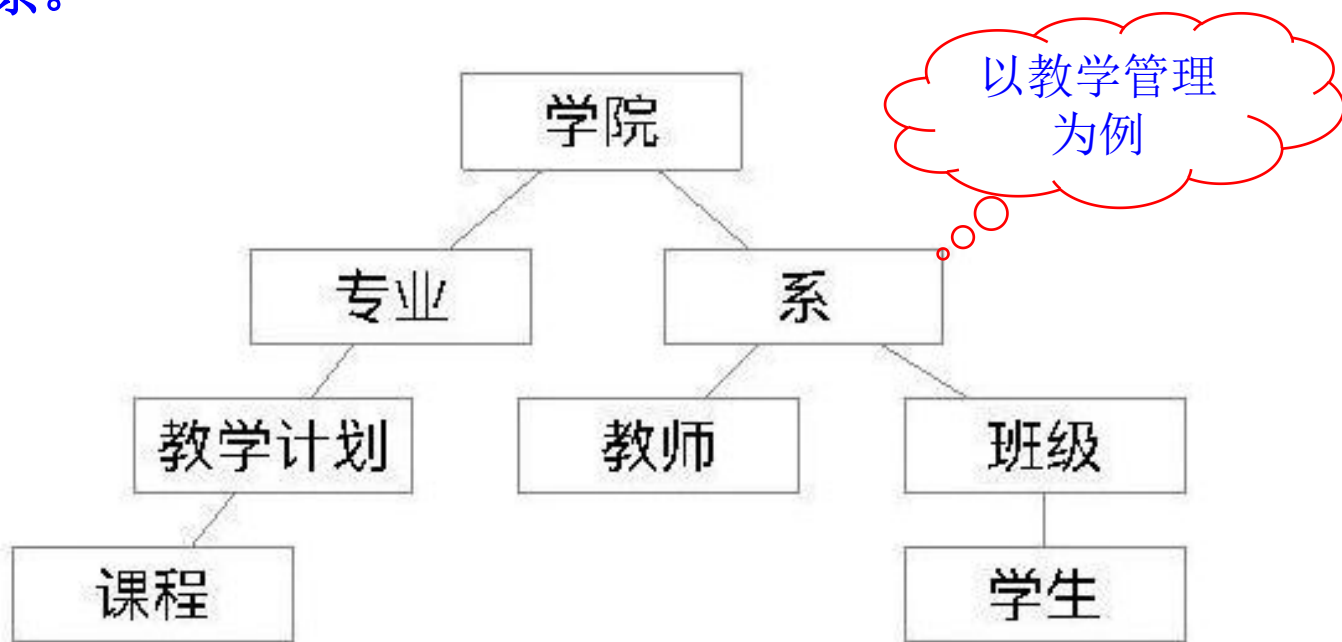
1973年ACM图灵奖
数据库之父



6.2 数据库发展历程

● 层次数据库

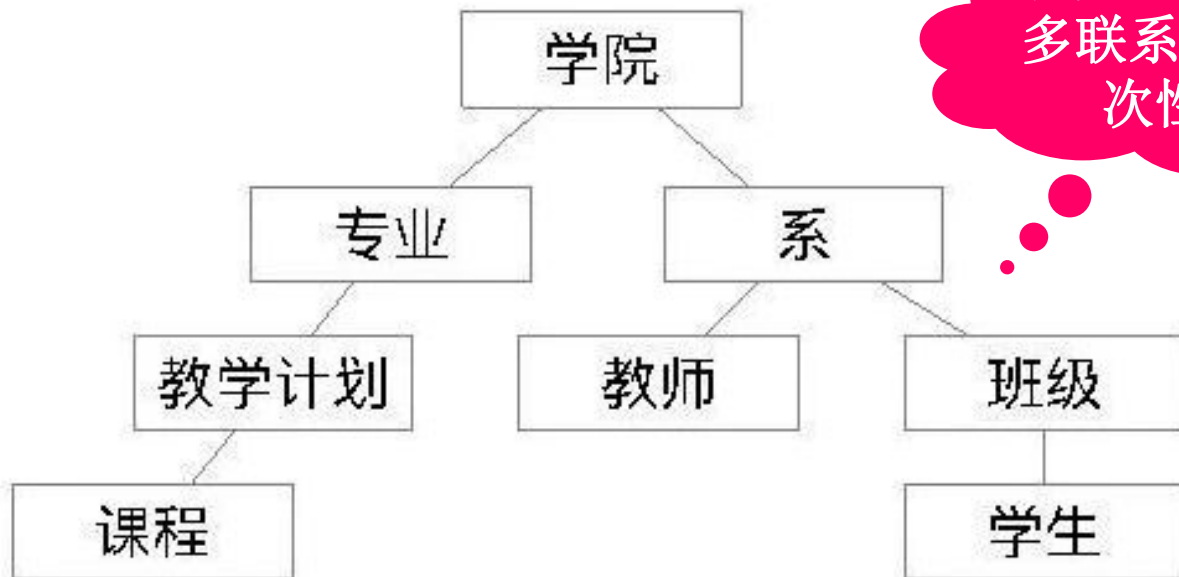
紧随网状数据库后出现的是层次数据库,其数据模型是层次数据模型,即使用树结构来描述实体及其之间关系的数据模型。在这种结构中,每一个记录类型都用节点表示,记录类型之间的联系则用节点之间的有向线段来表示。每一个子节点只能有一个父节点,但是每一个父节点可以有多个子节点。**这种结构决定了采用层次数据模型作为数据组织方式的层次数据库系统只能处理一对多的实体联系。**



6.2 数据库发展历程

● 层次数据库

1968年,世界上第一个层次数据库系统—**IMS**(*Information Management System*,信息管理系统)诞生于IBM 公司,这也是世界上第一个大型商用的数据库系统。

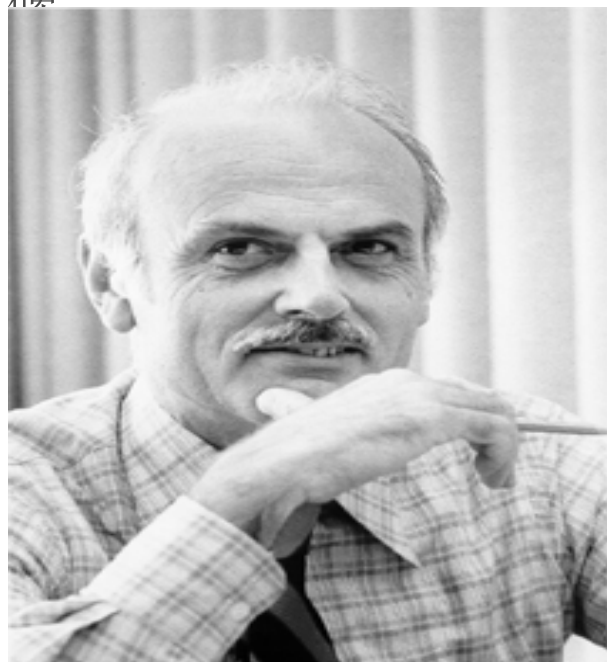


现实世界中很多联系是非层次性的

6.2 数据库发展历程

● 关系数据库

1970年,IBM 的研究员Edgar F.Codd发表了*A Relational Model of Data for Large Shared Data Banks* 论文,提出了关系数据模型的概念,奠定了关系数据模型的理论基础,这是数据库发展史上具有划时代意义的里程碑



Edgar F.Codd

1981年ACM图灵奖
关系数据库之父

教师信息表

工号	姓名	职称	学院
2001	李三	教授	经管
2002	张四	讲师	软件
2003	王五	教授	金融

课程信息表

课程号	课程名	学时	学分
X2001	经济学	64	4
X2002	金融学	48	3
X2003	软件工程	64	4

开课目录表

工号	课程号	开课学期	最多人数
2001	X2001	春季	100
2002	X2002	秋季	120
2003	X2003	秋季	100

6.2 数据库发展历程

● 关系数据库

随后,Edgar F.Codd又陆续发表了一篇文章,论述了范式理论,用数学理论奠定了关系数据库的基础,为关系数据库建立了一个数据模型——**关系数据模型**。关系数据模型的概念非常简单,结构特别灵活,能满足所有布尔逻辑运算和集合**运算**规则形成的查询要求;可以搜索、比较和组合不同类型的数据;使用关系数据模型进行数据**增加**和**删除**操作非常方便,关系数据模型具有较高的**数据独立性**和更好的**安全保密性**。

教师信息表

工号	姓名	职称	学院
2001	李三	教授	经管
2002	张四	讲师	软件
2003	王五	教授	金融

课程信息表

课程号	课程名	学时	学分
X2001	经济学	64	4
X2002	金融学	48	3
X2003	软件工程	64	4

开课目录表

工号	课程号	开课学期	最多人数
2001	X2001	春季	100
2002	X2002	秋季	120
2003	X2003	秋季	100

6.2 数据库发展历程

● 关系数据库

IBM

押错宝!

1970年

开始关系数据库项目SystemR的研发
(押宝层次数据库IMS, 进展缓慢)

1974年

SQL语言
关系数据库
标准语言

1980年

SystemR产品化

ORACLE

眼光敏锐

1978年

Oracle 1.0
数据库行业巨头
由此诞生

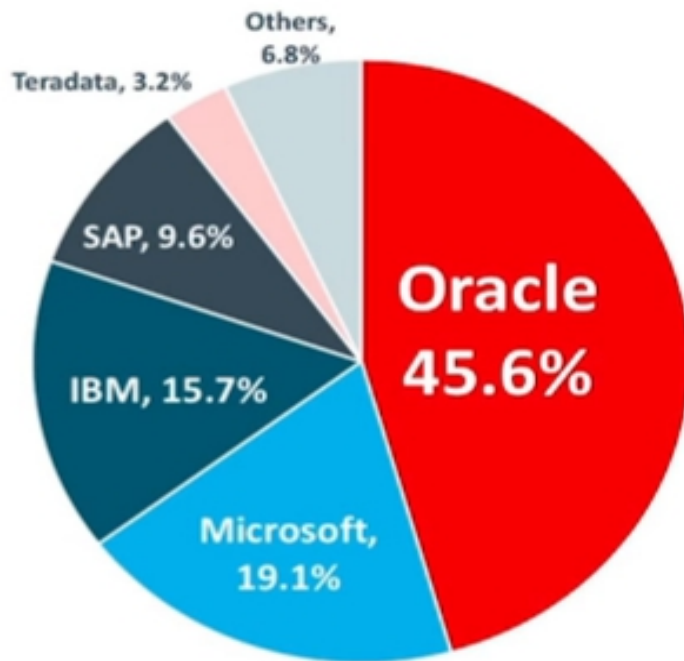


Larry Ellison

甲骨文创始人
财富¥4980亿

6.2 数据库发展历程

● 关系数据库



虽然关系数据库系统的技术很成熟,但随着市场和信息技术的发展,其局限性也逐渐暴露出来,即它能很好地处理所谓的“表格型数据”,却无法处理当前出现的越来越多的复杂类型数据(如文本、图像、视频等)。

6.2 数据库发展历程

● 分布式数据库

随着互联网的不断普及,特别是移动互联网的兴起,数据规模爆炸式增长,**单机数据库越来越难以满足用户需求**。进入大数据和移动互联时代后,因为数据的特性和应用场景的变化,注定着不论是传统的关系数据库,还是新型的NoSQL和NewSQL数据库都会向着分布式的发展方向发展,**分布式数据库也成了数据库领域的主流方向之一**。但是分布式数据库也存在一些问题。例如,众多节点之间通信会花费大量时间;数据的安全性和保密性在众多节点之间会受到威胁;在分布式系统复杂的存取结构中,原本在集中式系统中有效存取数据的技术可能不再适用;分布式的数据划分、负载均衡、分布式事务处理和分布式执行技术缺乏新的突破。



6.2 数据库发展历程

● 云数据库

云数据库将传统的数据库系统配置在“云上”,有专门的云服务提供商进行这些“云上”数据库系统的管理和部署工作,用户只需要通过付费的方式就能获取数据库服务。不同于传统数据库,云数据库通过计算存储分离、存储在线扩容、计算弹性伸缩来提升数据库的可用性和可靠性。代表性的云数据库是亚马逊的**Aurora**和阿里巴巴的**OceanBase**,它首先提出了日志即是数据库的理念,减少了网络消耗,提升了系统的可用性。

2019年6月,Gartner发布*The Future of the Database Management System(DBMS)Market Is Cloud* 报告,明确提出传统的部署数据库的方式已经过时,云是未来,所有组织,无论大小,都将越来越多地使用云数据库。但是,云数据库中存在的问题也不可忽略,云计算中最值得关注的是安全问题,云计算对数据安全带来了极大威胁,数据极易泄露,存在意外丢失的风险。



6.2 数据库发展历程

新型数据库
管理系统

NOSQL
2000s



NEWSQL
2010S



HATP
2010S



6.2 数据库发展历程

支付宝
OceanBase
数据库这十年

InfoQ 

6.3 数据库设计过程

- 业务→实体关系(ER)图



6.3 数据库设计过程

● 业务→实体关系(ER)图

1 软件工程系制定专业教学计划(课程)

2 教师确定讲授课程

3 学生选择课程（选课）

4 教师给出学生课程成绩

6.3 数据库设计过程



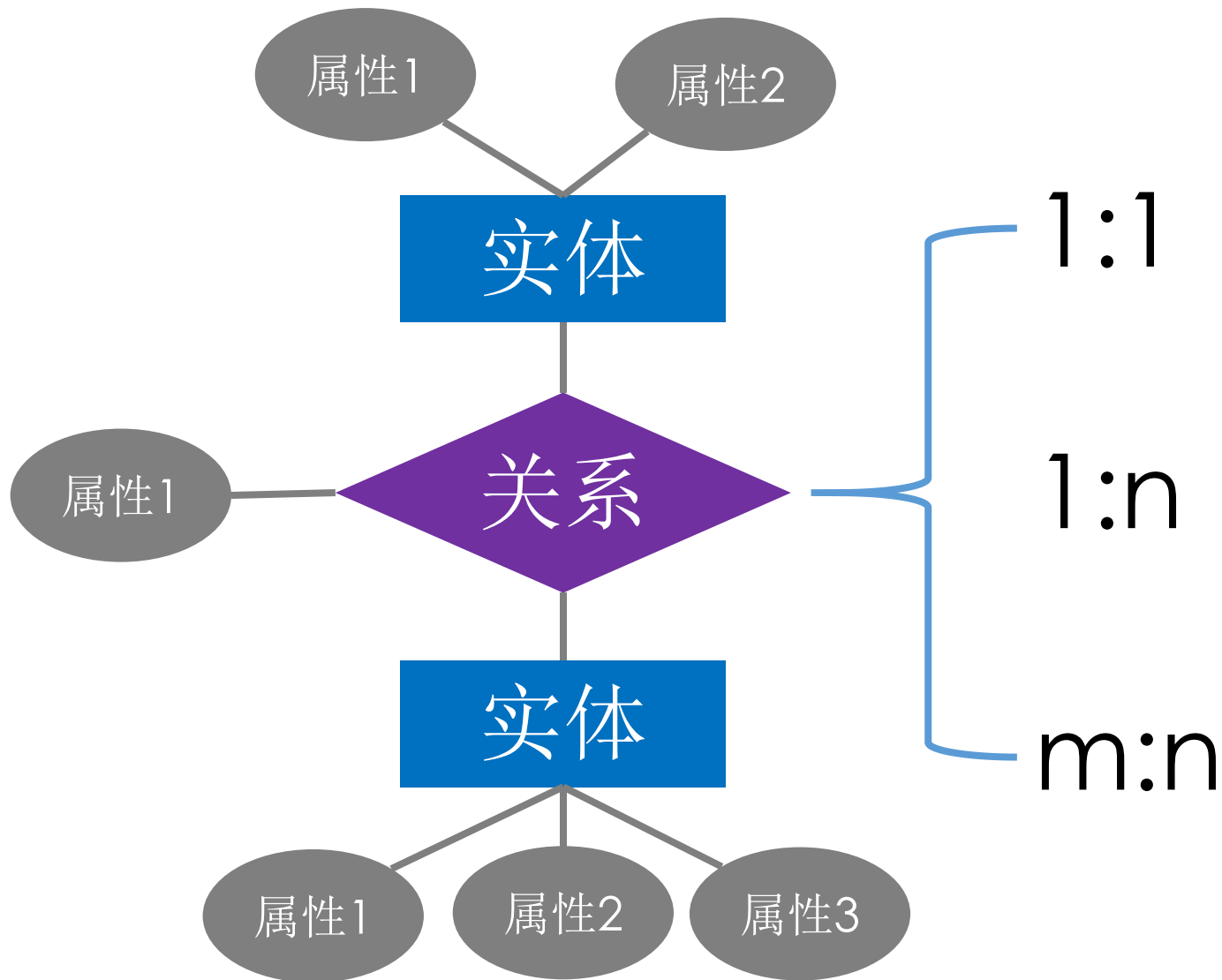
● 业务→实体关系(ER)图

实体关系模型(*Entity-Relationship Model, E-R Model*), 它是陈品山(*Peter P.S Chen*)博士于1976年提出的一套数据库的设计工具, 他运用真实世界中事物与关系的观念, 来解释数据库中抽象的数据架构。实体关系模型利用图形的方式来表示数据库的概念设计, 有助于设计过程中的构思及沟通讨论。*E-R*模型的基本元素是: 实体、属性和联系。

- **实体**是一个数据对象(类型相同数据元素集合), 指应用中可以区别的客观存在的事物。在E-R模型中, 实体用方框表示, 方框内注明实体的命名。
- **属性**, 实体的某一特性称为属性。在一个实体中, 能够唯一标识实体的属性或属性集为“实体标识符”。
- **联系**, 在现实世界中, 每一个实体都不是孤立存在的, 实体和实体之间都是存在联系的。在E-R图中, 联系用菱形框表示, 并用线段将其与相关的实体联系起来。实体间的联系分为一对一(1:1)、一对多(1:N)和多对多(M:N)。

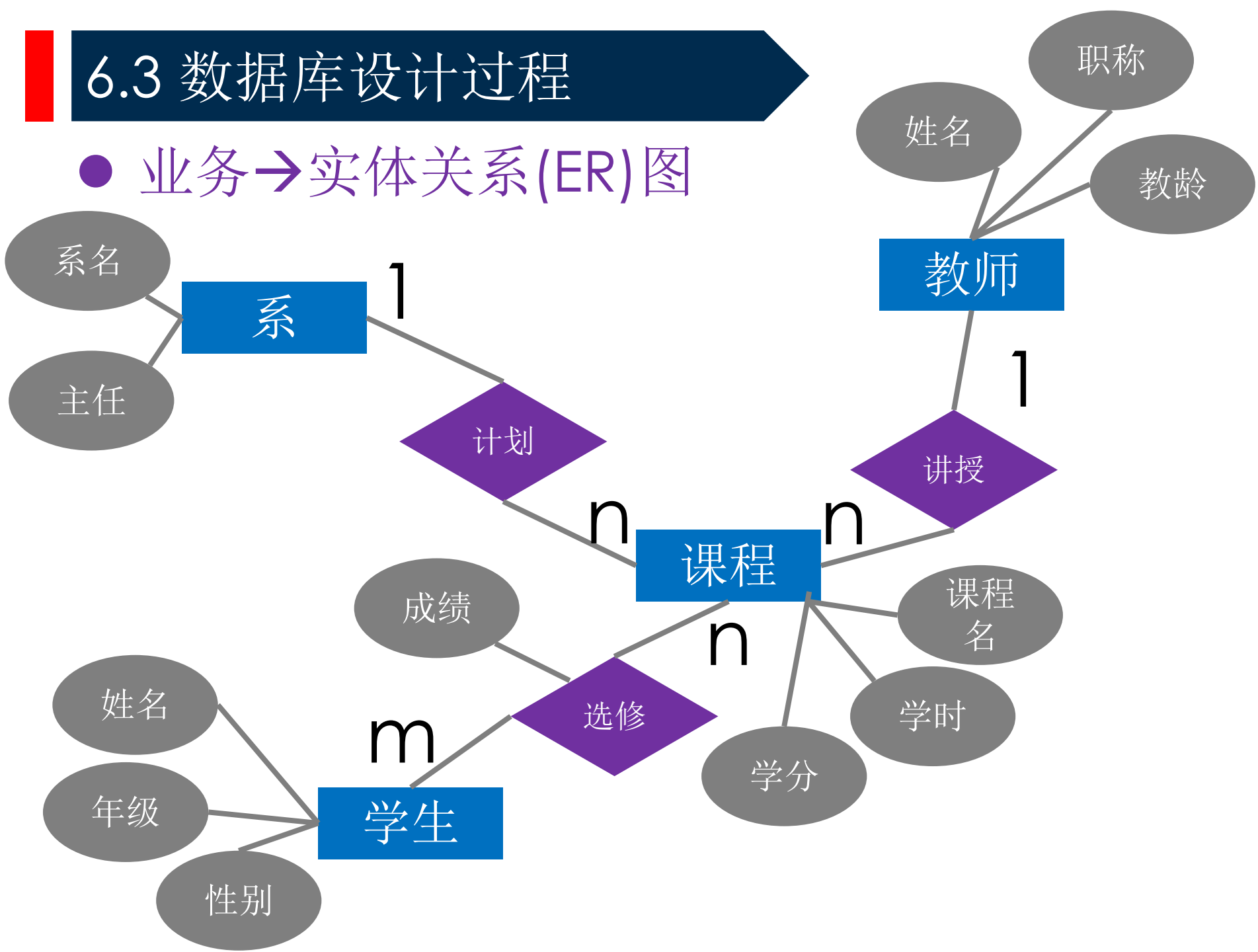
6.3 数据库设计过程

● 业务→实体关系(ER)图



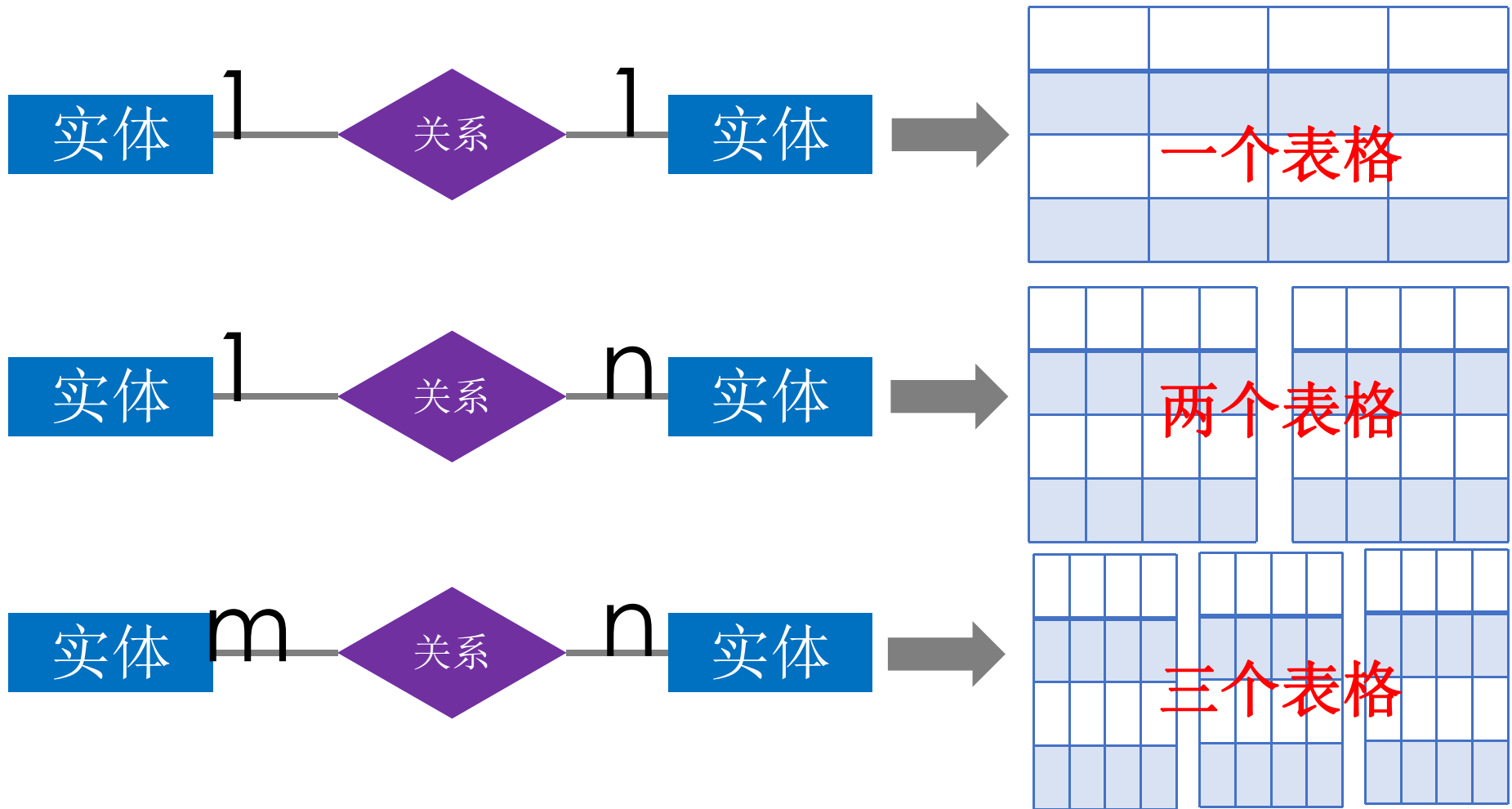
6.3 数据库设计过程

● 业务→实体关系(ER)图



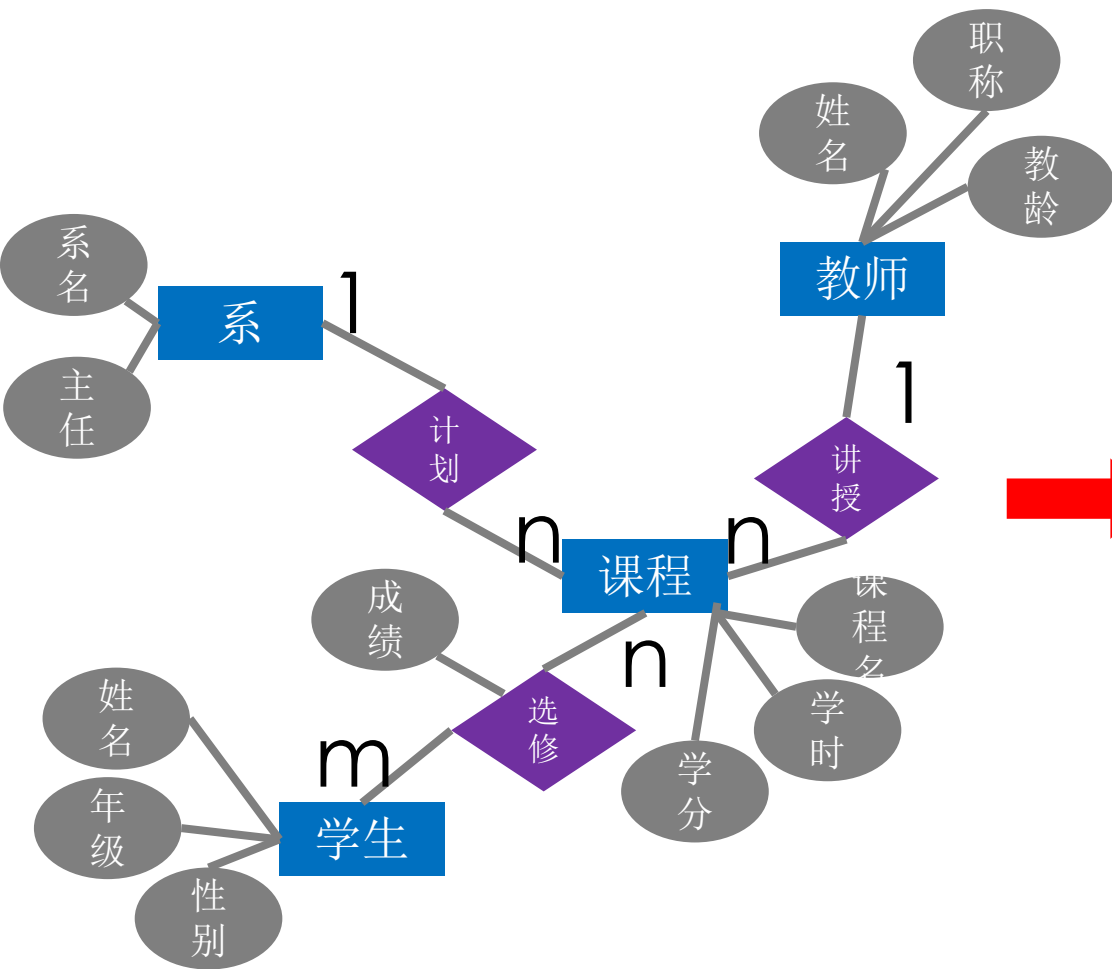
6.3 数据库设计过程

- 实体关系(ER)图→数据库结构



6.3 数据库设计过程

● 业务→实体关系(ER)图



系信息表

教师信息表

课程信息表

学生信息表

学生课程成绩表

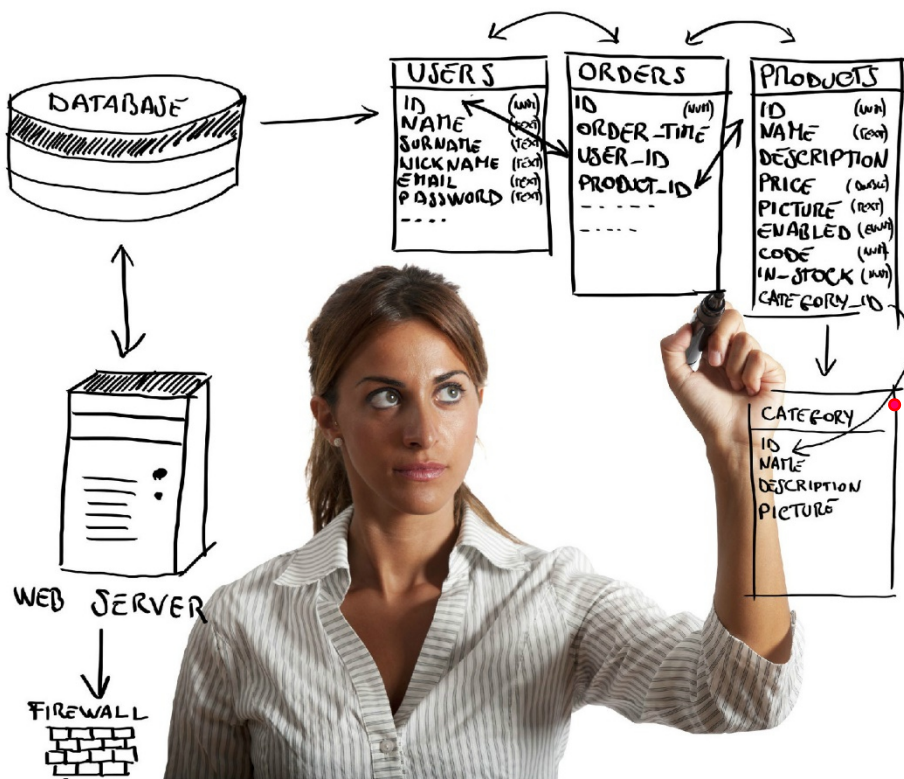
6.4 数据库管理员



6.4 数据库管理员

数据库管理员（Database Administrator，简称DBA），是从事管理和维护数据库管理系统(DBMS)的相关工作人员的统称，属于运维工程师的一个分支，主要负责业务数据库从设计、测试到部署交付的全生命周期管理。

DBA的核心目标是保证数据库管理系统的**稳定性**、**安全性**、**完整性**和**高性能**。



数据库
运维

6.4 数据库管理员

数据库管理员（*Database Administrator*，简称DBA），是从事管理和维护数据库管理系统(DBMS)的相关工作人员的统称，属于运维工程师的一个分支，主要负责业务数据库从设计、测试到部署交付的全生命周期管理。



6.4 数据库管理员

数据库管理员（*Database Administrator*，简称DBA），是从事管理和维护数据库管理系统(DBMS)的相关工作人员的统称，属于运维工程师的一个分支，主要负责业务数据库从设计、测试到部署交付的全生命周期管理。

目标
很明确

难度
却很大



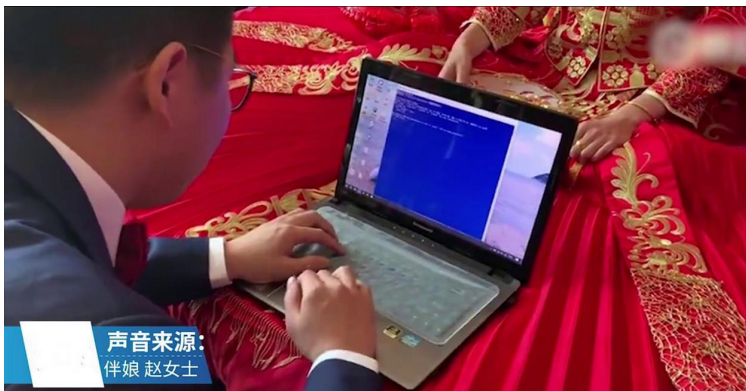
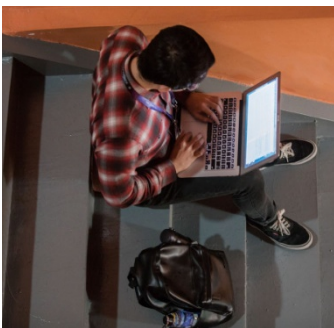
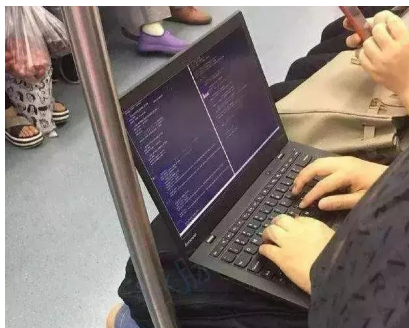
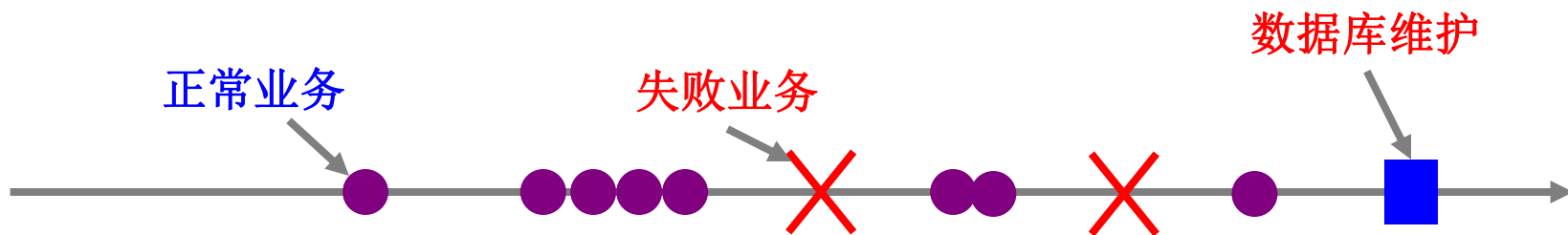
6.4 数据库管理员

企业对数据应用水平不同，因而造成BDA工作重心及岗位需求也有所不同：

- **层次1** 是以**数据库维护**为主，常见表现是**"救火队员"**型。数据库维护基本靠人，随着运维体量的增加，需要线性增加人员,整体数据应用水平，基本处于简单、粗放型。
- **层次2** 仍是以数据库维护为主，但已形成较为**完善的运维体系**。除了基础运维之外，甚至可以考虑一些预防性的措施，提高整体的运维效益。但此阶段尚未从更高角度去考虑数据问题，仍仅限于运维层面。
- **层次3** 数据设计应用阶段，企业已不满足数据简单的"存取类"需求，而是从更高的应用角度，考虑如何**提高整体数据应用水平**。这个阶段会增加数据库架构、设计，加强业务端数据优化工作。
- **层次4** 数据架构治理阶段，企业不单从某个应用、某条业务线去考虑数据问题，而是公司**整体层面做数据的顶层设计**。考虑建立专门的机构（如数据委员会）或岗位-首席数据官（*CDO*）。近些年来，颇为火热的"数据中台"，正是为迎合这一需求而产生的。



6.4 数据库管理员



最新研究进展

序号	刊物名称	刊物全称	地址
1	VLDB	International Conference on Very Large Data Bases	http://dblp.uni-trier.de/db/conf/vldb/
2	ICDE	IEEE International Conference on Data Engineering	http://dblp.uni-trier.de/db/conf/icde/
3	SIGMOD	ACM Conference on Management of Data	http://dblp.uni-trier.de/db/conf/sigmod/
4	SIGIR	International Conference on Research on Development in Information Retrieval	http://dblp.uni-trier.de/db/conf/sigir/

