

Chapter 1

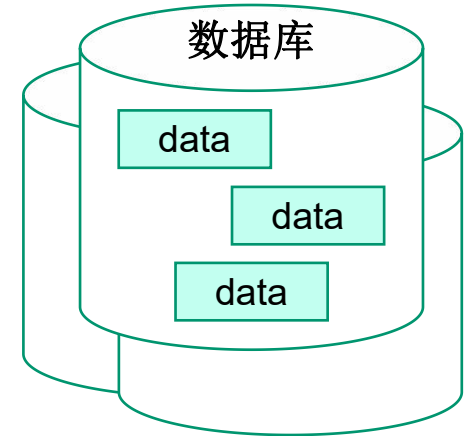
Introduction

(2021.8.30)

Ch1. Introduction

□ 什么是数据库？

✧ 长期存储在计算机内、有组织的、大量的、共享的数据集合。



□ 什么是数据库管理系统？

➤ 位于用户与操作系统之间的一层数据管理软件，为用户或应用程序提供访问数据库的方法和接口。

□ 什么是数据库技术？

➤ 研究数据库的结构、存储、设计、管理和使用的一门软件学科。

Ch1. Introduction

□ what we will learn in this course?

✧ concepts & Skills you need to make a database work

✧ not how to build one, but how to set one up

Terminology

□ DataBase Management System (DBMS)

☞ 数据库管理系统，简称 DBMS

☞ http://en.wikipedia.org/wiki/Database_management_system

□ DataBase (数据库)

☞ <http://en.wikipedia.org/wiki/Database>

□ DataBase User (数据库用户)

Ch1. Introduction

□ DataBase Management System

(数据库管理系统, 简称 **DBMS**)

✧ is a program product for keeping computerized records (on disk) about an enterprise.

✧ Example of records

- Records about sales
- Records about students
- Records about library inventory and loans

Database

Ch1. Introduction

□ Database Users

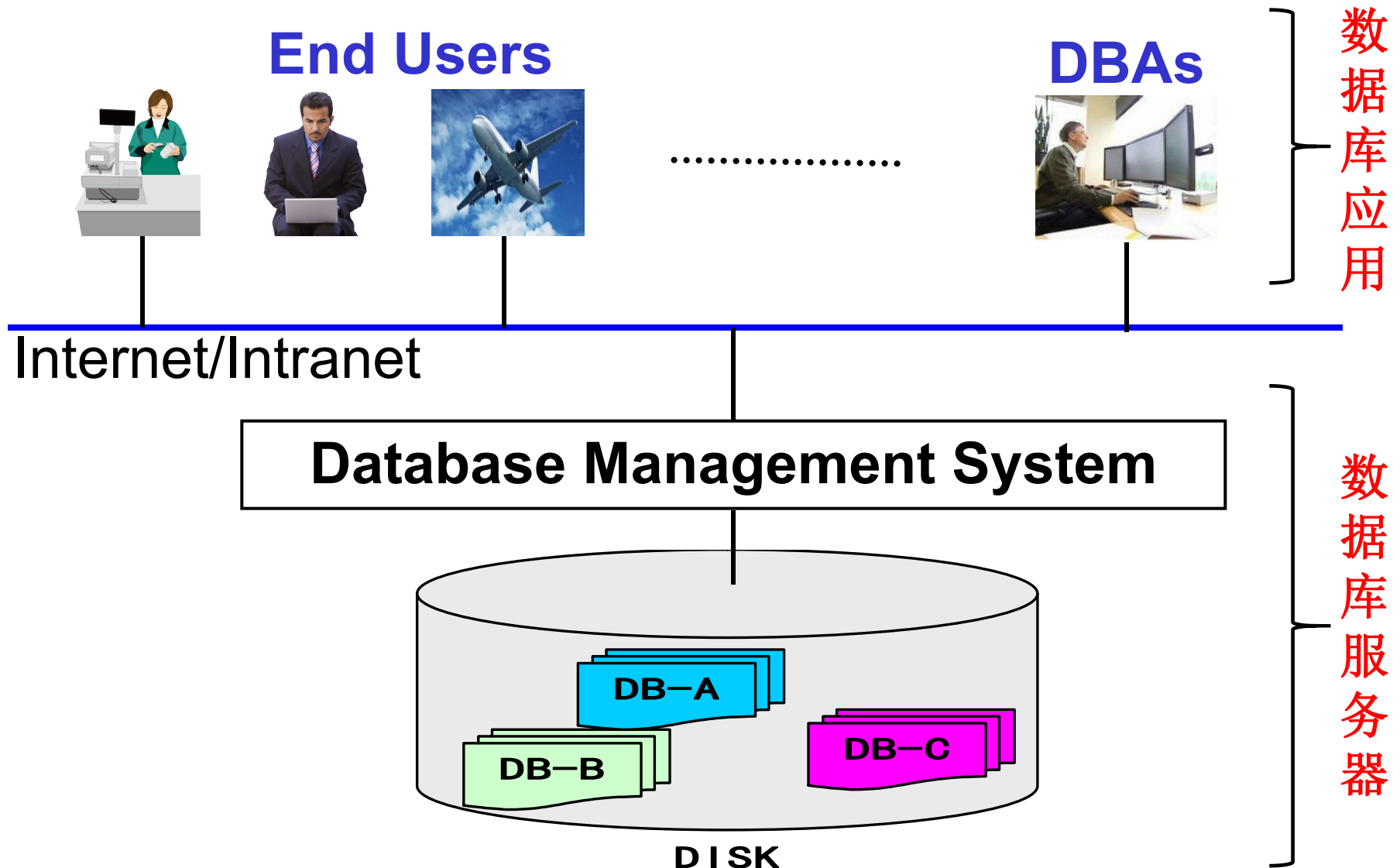
∞ End users

- Casual users ✓
- Naive users

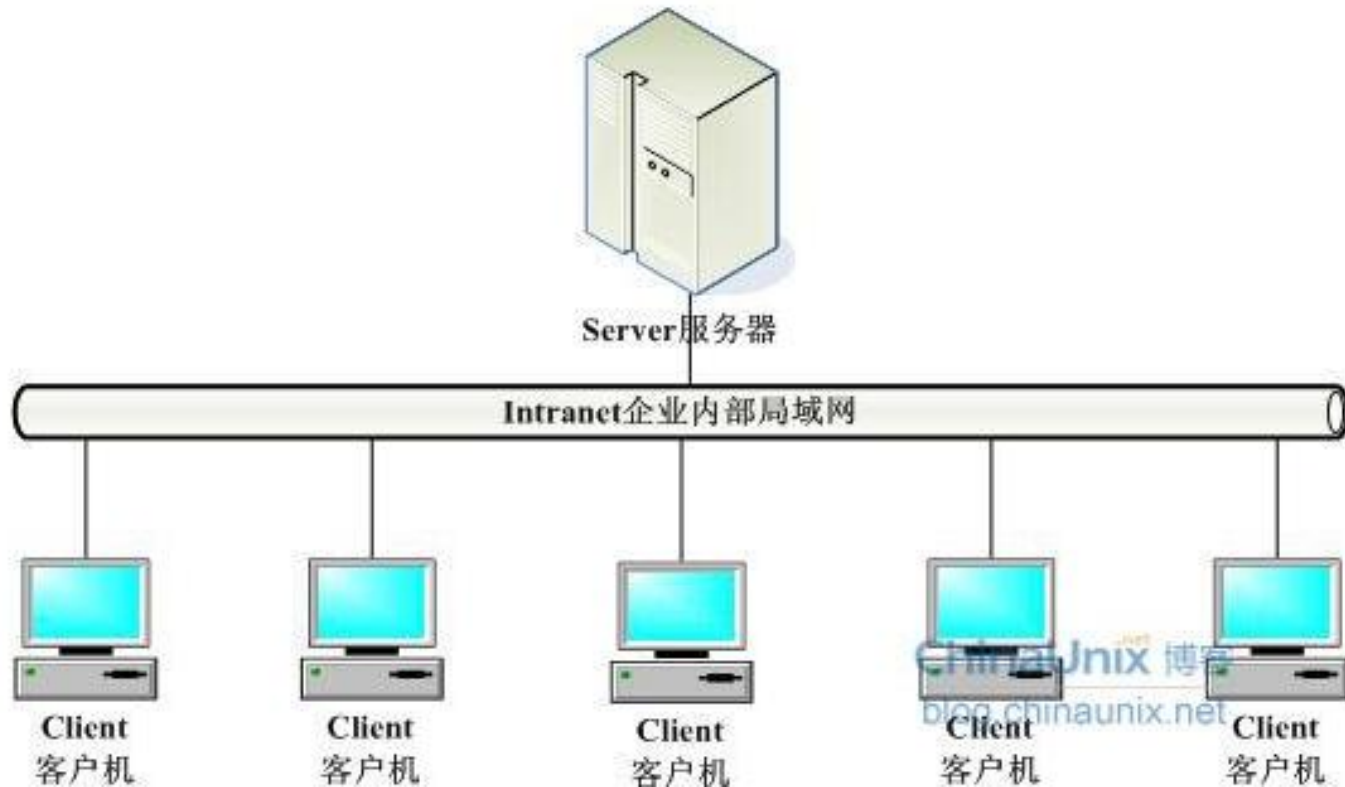
∞ Application programmers ✓

∞ DataBase Administrators (DBA) ✓

Architecture of Database System



Two-tier Architecture vs. Three-tier Architecture



(两层C/S架构)

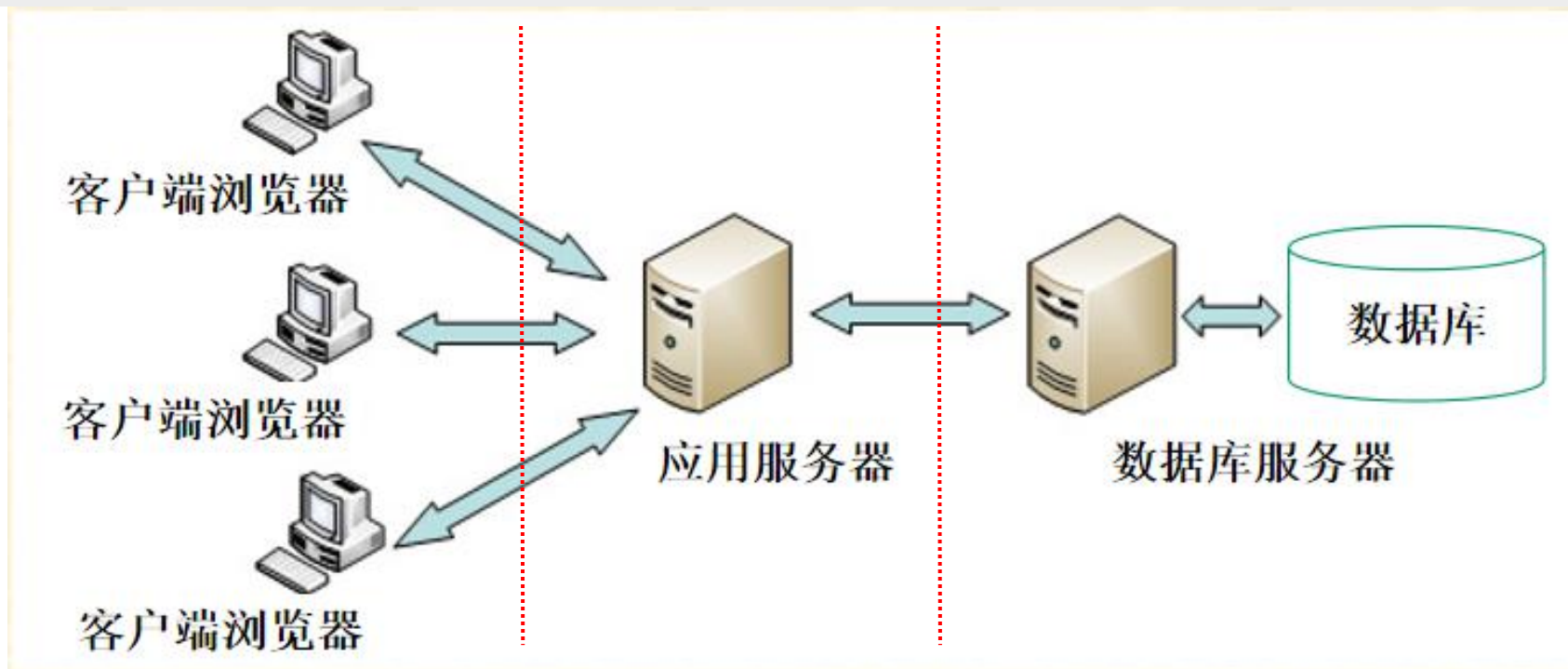
- 在两层C/S架构中，数据库和DBMS运行在数据库服务器中，数据库应用程序运行在客户机中，两者之间通过局域网实现数据访问。

Two-tier Architecture vs. Three-tier Architecture

- 但更多的时候，数据库服务器及其应用程序可能分布在距离遥远的不同地方（如下图所以），他们相互之间无法通过企业内部的局域网相连，只能通过更广阔的互联网来实现数据访问和数据传输。



Two-tier Architecture vs. Three-tier Architecture



(三层B/S架构)

- 在三层B/S架构中，数据库和DBMS运行在数据库服务器中，数据库应用程序运行在应用服务器（也称“Web服务器”）中，用户客户端只需要安装常用的浏览器，负责接收用户输入和结果展示。

Ch1. Introduction

□ DataBase Management System

(数据库管理系统, 简称 DBMS)

☞ is a program product for keeping computerized records (on disk) about an enterprise.

☞ Example of records

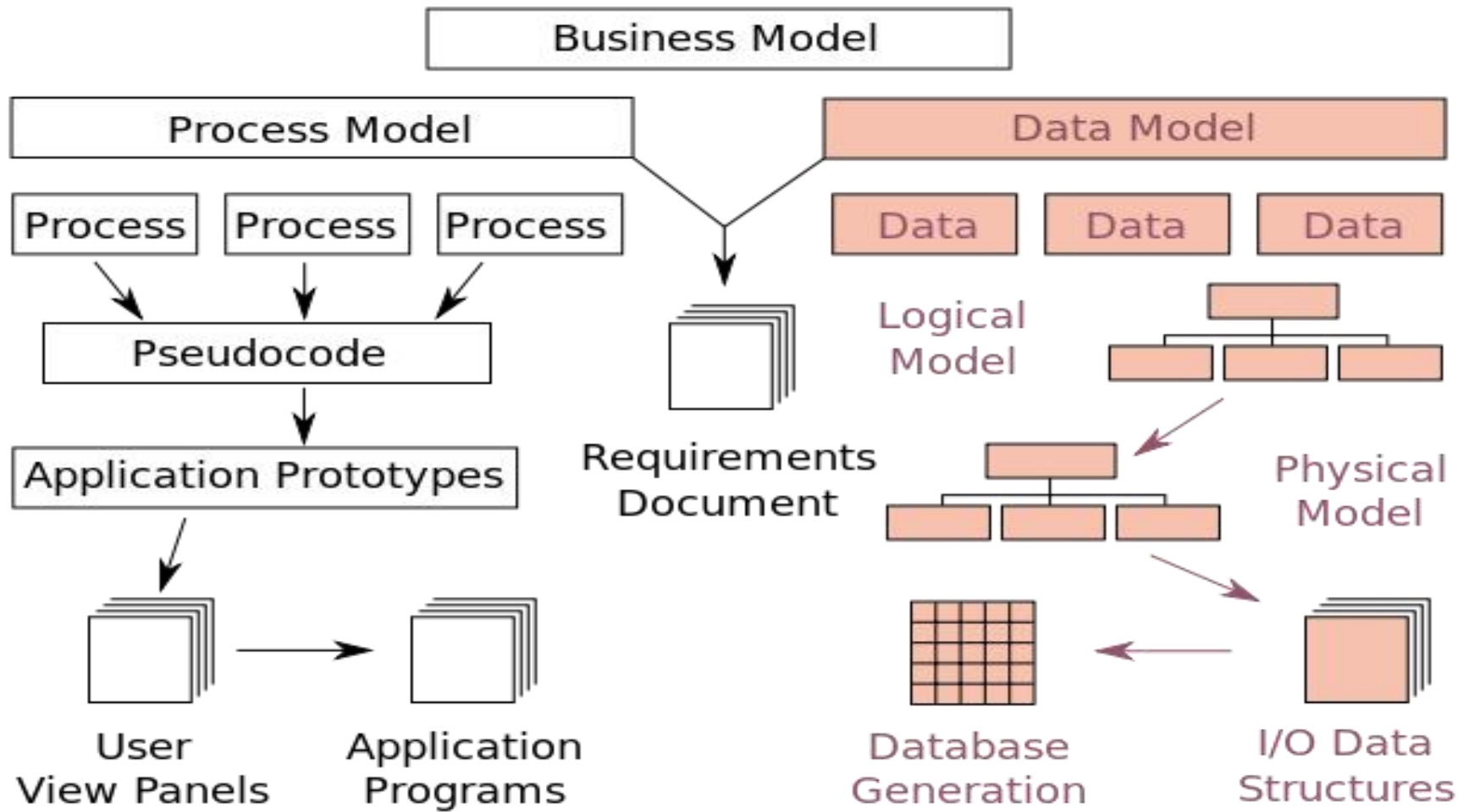
- Records about sales
- Records about students
- Records about library inventory and loans

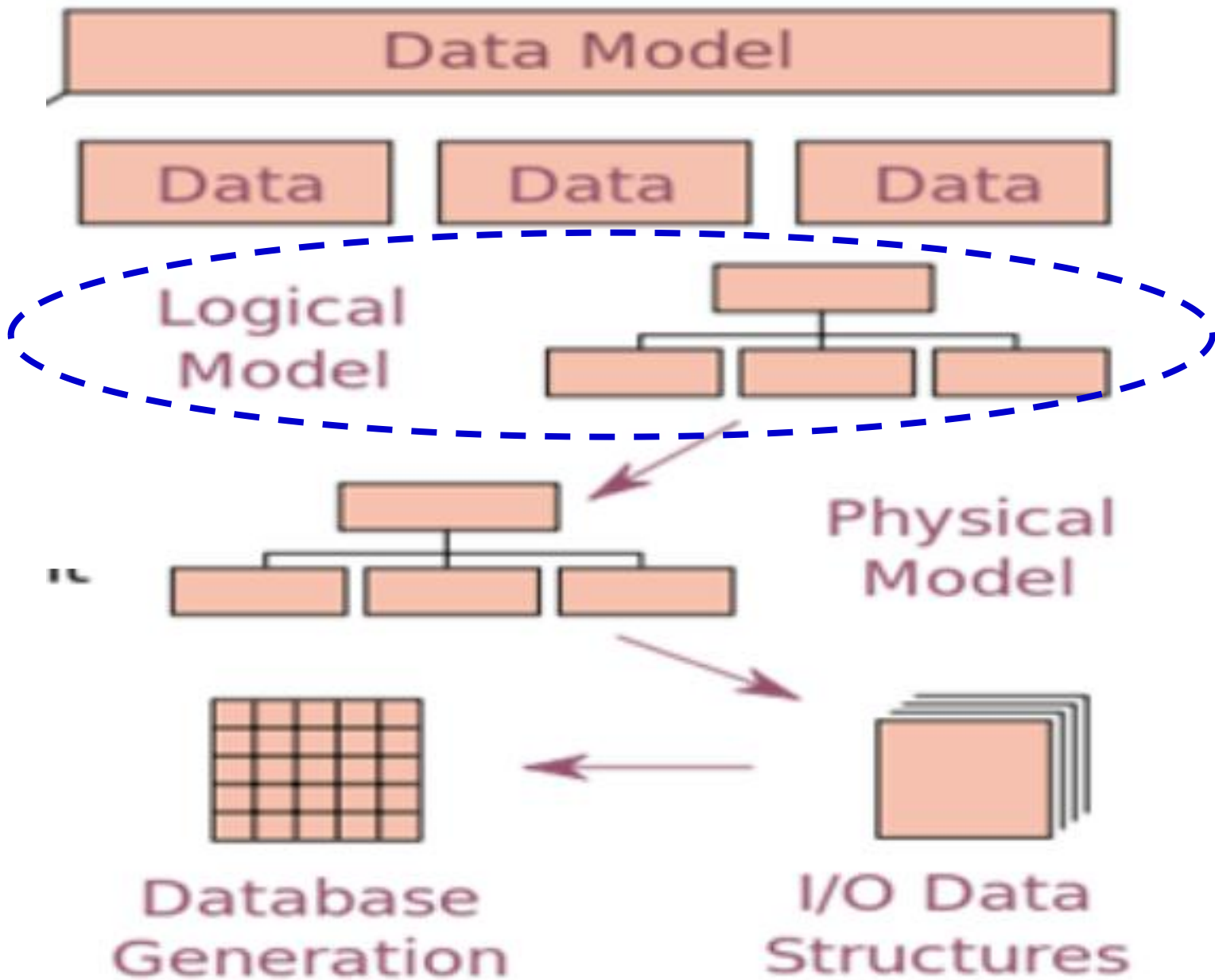
Data Model

Database

Data Model

Business Model Integration





Terminology

❑ Data Model (数据模型)

🌀 http://en.wikipedia.org/wiki/Data_model

- Hierarchical Data Model (层次数据模型)
- Network Data Model (网状数据模型)
- Relational Model (关系模型)
- Object-Oriented Model (面向对象模型)



Object-Relational Model
(对象关系模型)

数据管理系统的发展历史

- 第一阶段：层次/网状数据库（20世纪60年代 ~ 70年代）
 - 数据库从无到有，提供简单的共享数据的‘读 / 写’
- 第二阶段：关系数据库（20世纪70年代 ~ ）
 - 数据库 从有到 大规模商业化应用
 - 面向事务处理型应用：OLTP
- 第三阶段：数据仓库（20世纪90年代 ~ ）
 - 面向数据分析型应用：OLAP
- 第四阶段：新型数据管理系统（21世纪 ~ ）
 - NoSQL数据库 / 大数据管理系统

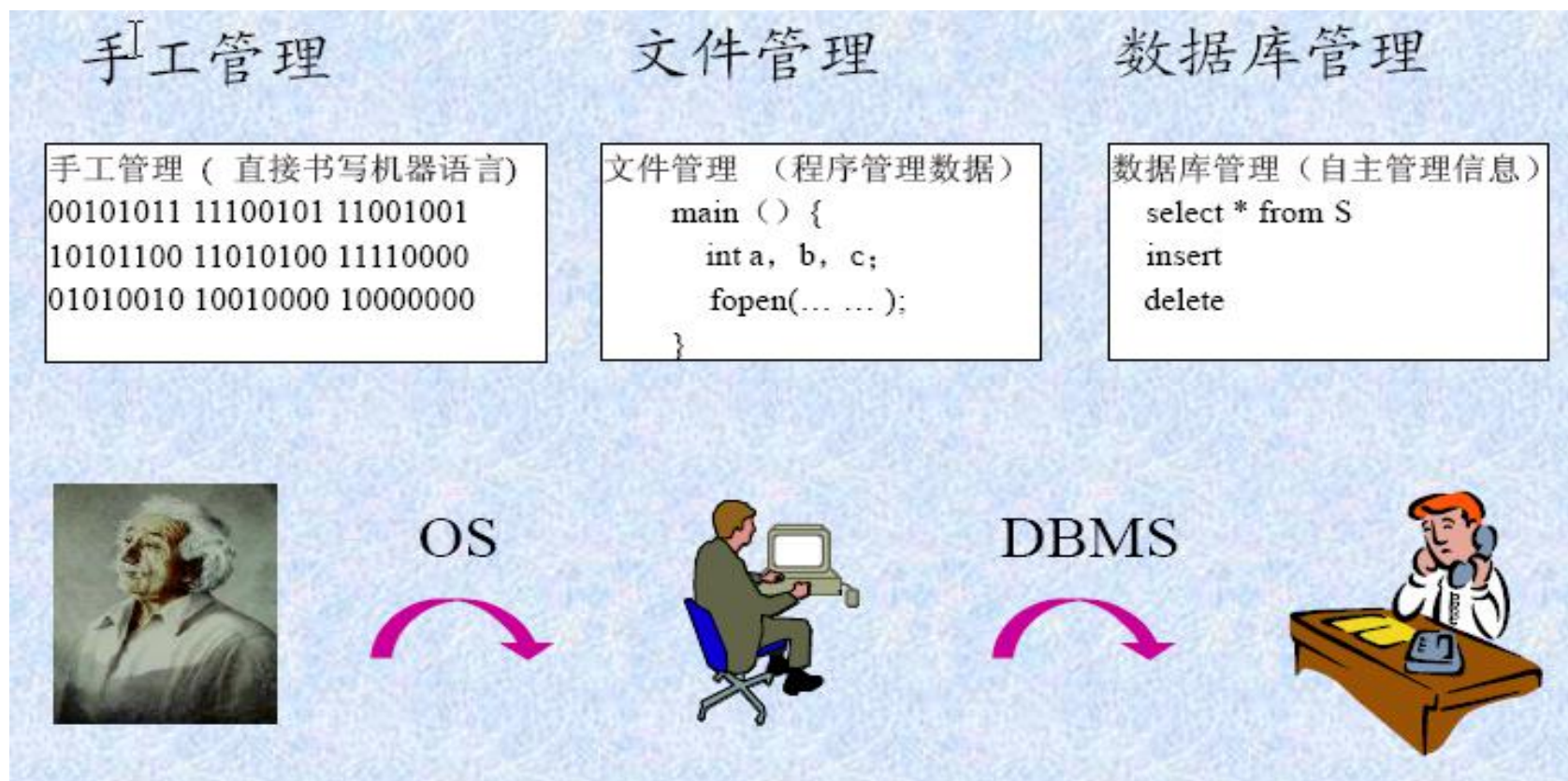
Ch1. Introduction

□ 数据管理技术的三个发展阶段

✧ 人工管理（50年代以前）

✧ 文件系统管理（60 ~ 70年代）

✧ 数据库系统管理（70年代至今）



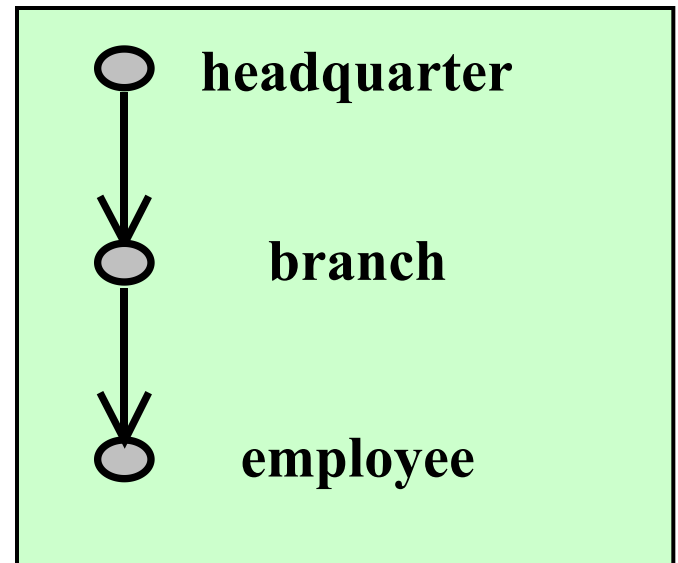
Ch1. Introduction

□ History of Database Systems

☞ Hierarchical Data Model (层次数据模型)

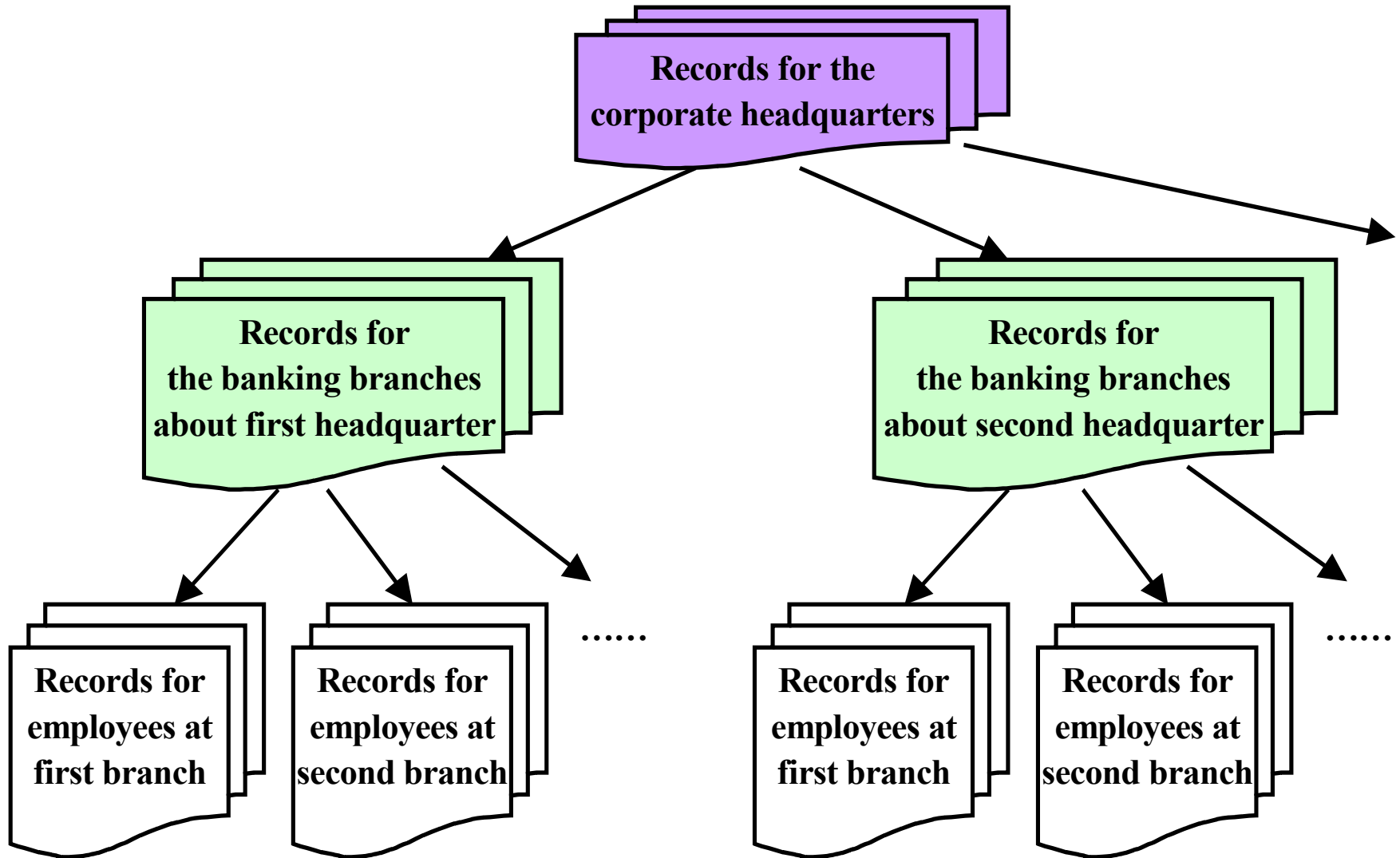
- 1968, IMS (Information Management System)
- Different kinds of records relate to one another in a hierarchical form.

➤ A directed tree (有向树)

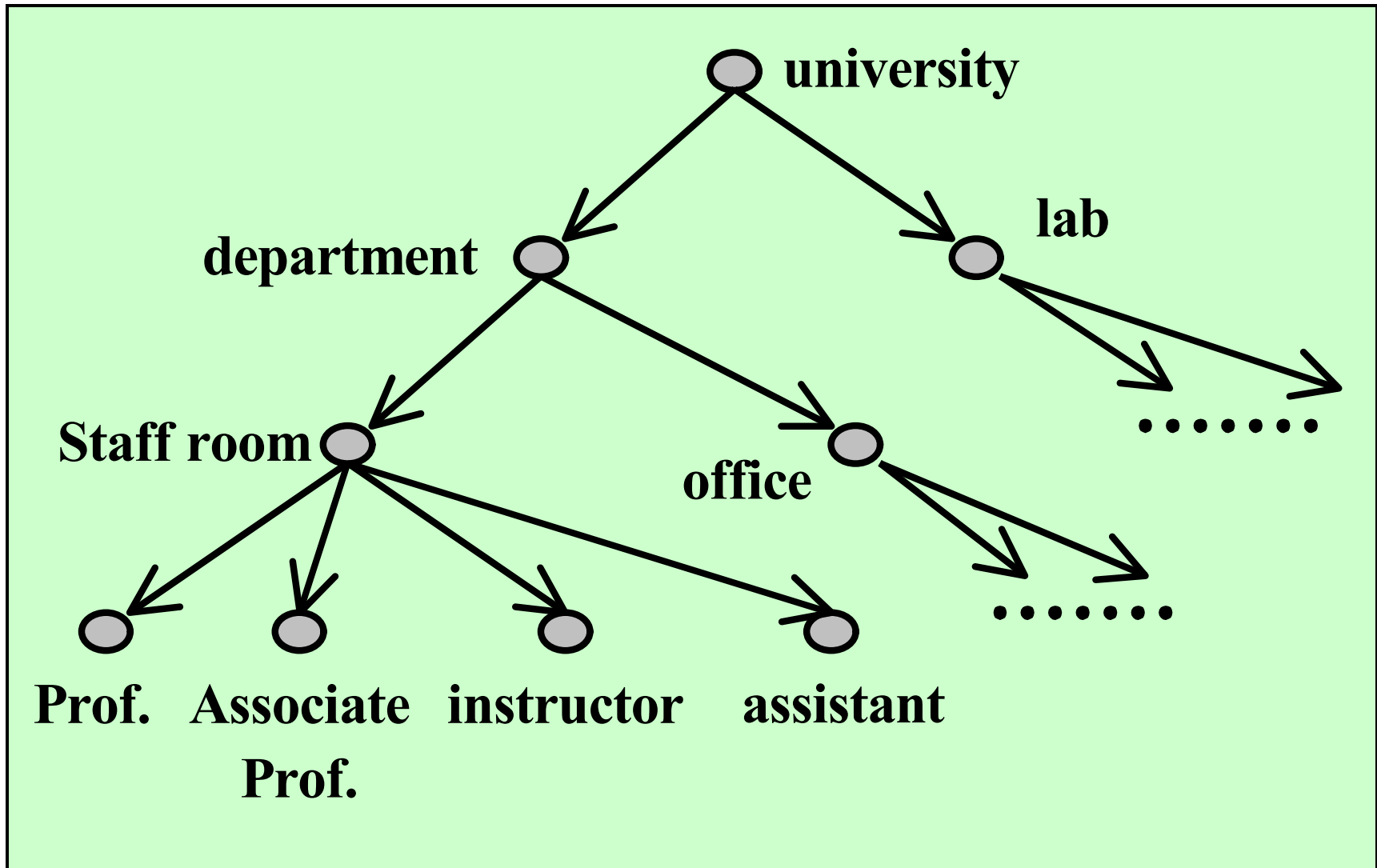


- Example: a hierarchical datamodel for a bank

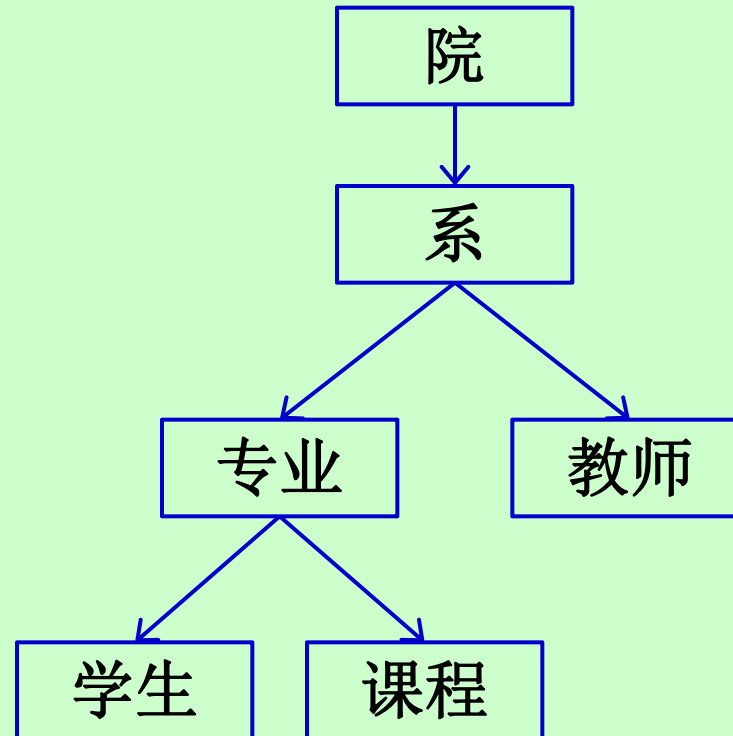
a database for a bank



Another example of hierarchical data model



Another example of hierarchical data model



Ch1. Introduction

□ History of Database Systems (cont.)

✧ **Network Data Model** (网状数据模型)

■ 1970, IDMS

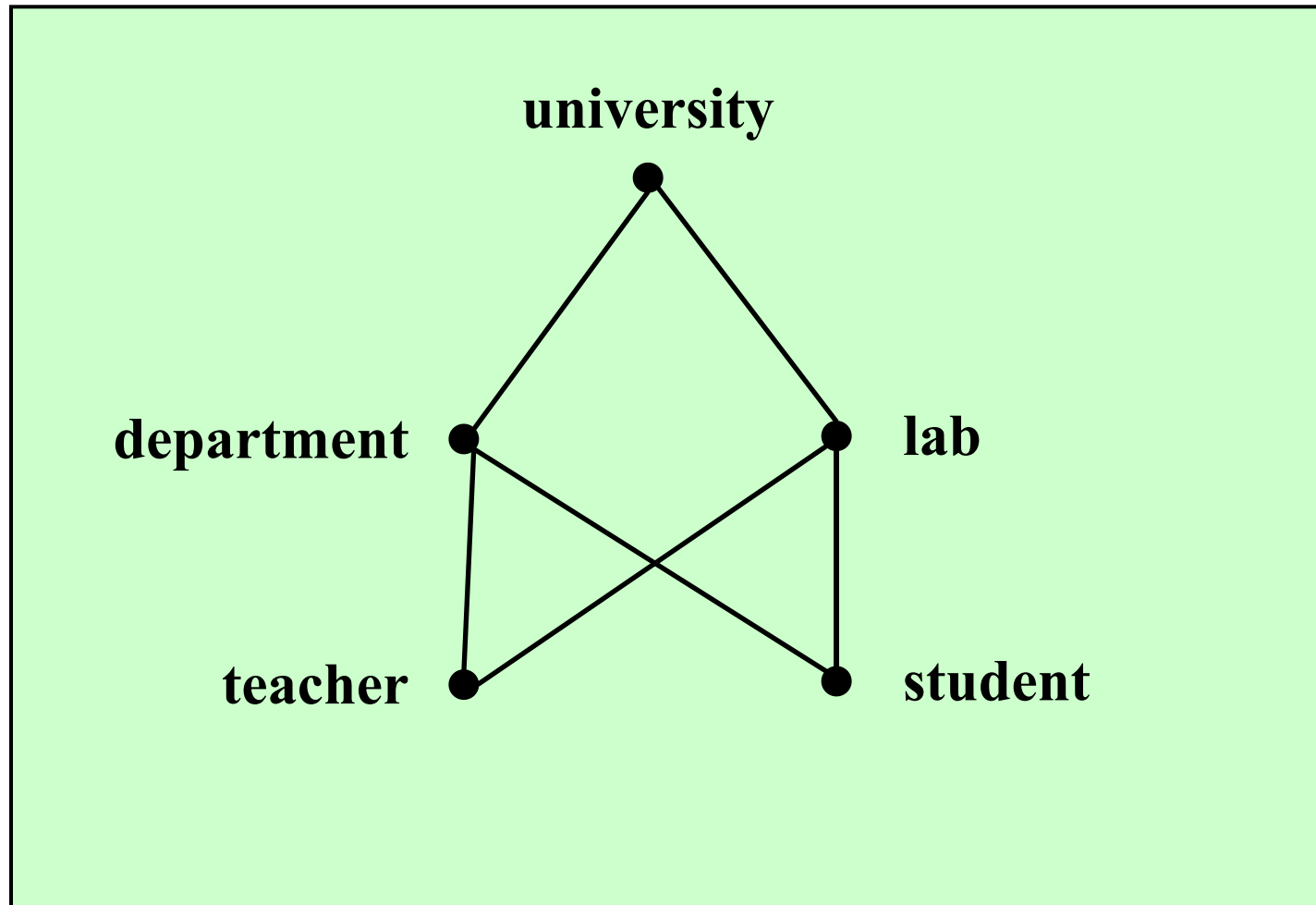
➤ was conceived as a result of the 1971 CODASYL report

■ A generalization of the hierarchical model where a set of records in one layer might have two different containing hierarchies at the next layer up.

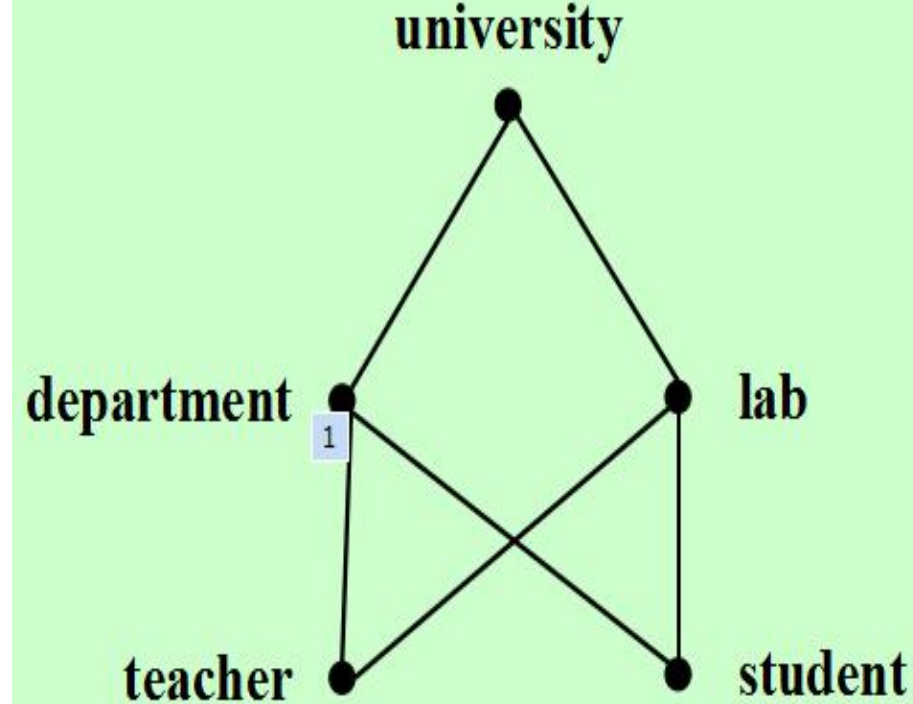
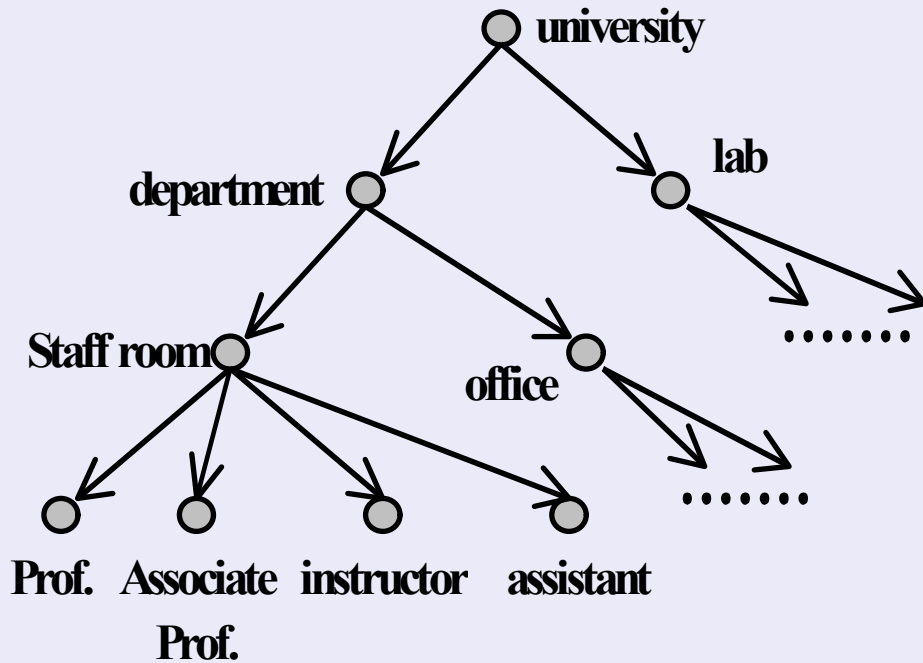
➤ A directed graph without circuits
(有向无环图)

Ch1. Introduction

□ Example of network data model

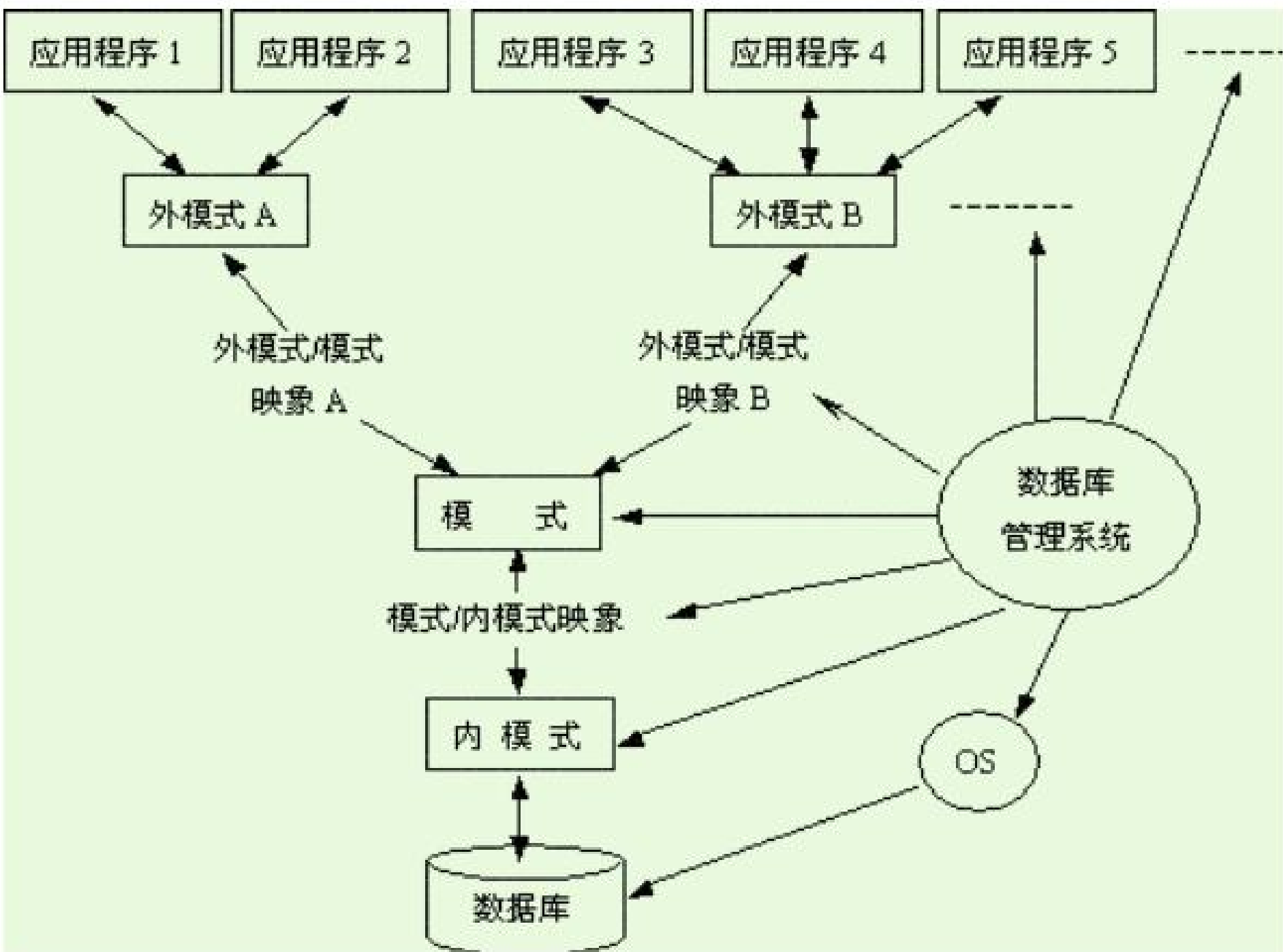


Hierarchical Data Model vs. Network Data Model



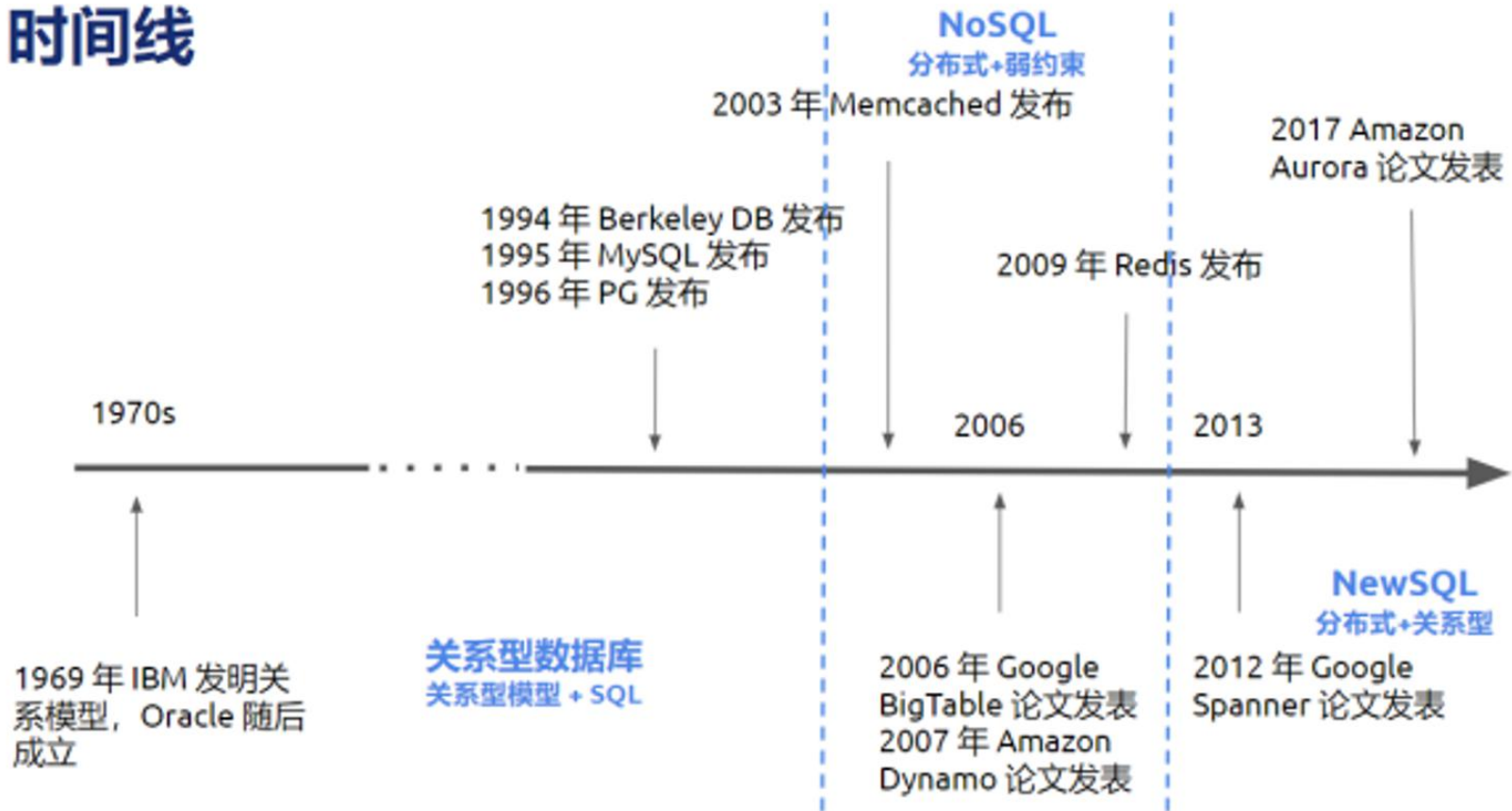
➤ 第一代数据库系统

- 查理士·巴赫曼（Charles William Bachman，73年图灵奖）
 - 数据库之父：第一个数据库系统IDS（64年）
 - 《DBTG报告》：三级模式



关系数据库系统的发展历史

时间线



□ 一张图清晰追溯数据库的发展历程（1962-2016年）

http://www.cbdio.com/BigData/2016-02/24/content_4651751.htm

Ch1. Introduction

□ History of Database Systems (cont.)

✧ Relational Model (关系模型)

- our data will look like **tables**.

✧ Relational Database (关系数据库)

- All Information is represented in the form of named **tables** with labeled **columns**.

✧ Relational DBMS (关系数据库管理系统)

Ch1. Introduction

students

sid	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

2021-8-19 **Figure 1.1a Relational Student Enrollment Database**

Ch1. Introduction

students

sid	lname	fname	class	telephone
1	Jones	Allen	2	555 1234
2				
3				
5				

courses

cno	cname	croom	time
101	French I	2-104	MW2
102	French II	2-113	MW3
105	Algebra	3-105	MW2
108	Calculus	2-113	MW4

2021-8-30 **Figure 1.1a Relational Student Enrollment Database**

Ch1. Introduction

students

sid	lname	fname
1	Jones	Allan
2	Smith	John
3	Brown	Harold
5	White	Edward

courses

cno	cname	credits
101	French I	2-
102	French II	2-
105	Algebra	3-
108	Calculus	2-

enrollment

sid	cno	major
1	101	No
1	108	Yes
2	105	No
3	101	Yes
3	108	No
5	102	No
5	105	No

2021-8-30 **Figure 1.1a Relational Student Enrollment Database**

Ch1. Introduction

students

sid	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

courses

cno	cname	croom	time
101	French I	2-104	MW2
102	French II	2-113	MW3
105	Algebra	3-105	MW2
108	Calculus	2-113	MW4

enrollment

sid	cno	major
1	101	No
1	108	Yes
2	105	No
3	101	Yes
3	108	No
5	102	No
5	105	No

2021-8-30 **Figure 1.1a Relational Student Enrollment Database**

Students-Courses

si d	Iname	fname	cname	majo r
1	Jones	Allan	French I	No
1	Jones	Allan	Calculus	Yes
2	Smith	John	Algebra	No
3	Brown	Harry	French I	Yes
3	Brown	Harry	Calculus	No
5	White	Edward	French II	No
5	White	Edward	Algebra	No

物理实现
结构

用户视角
中的结构

DISK

File B

File A

逻辑独立性

物理独立性

系统层次上
的结构

students

si d	Iname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

enrollment

si d	cn o	majo r
1	10 1	No
1	10 8	Yes
2	10 5	No
3	10 1	Yes
3	10 8	No
5	10 2	No
5	10 5	No

courses

cn o	cname	croom	time
10 1	French I	2-104	MW 2
10 2	French II	2-113	MW 3
10 5	Algebra	3-105	MW 2
10 8	Calculus	2-113	MW 4

三级模式 & 数据独立性

Ch1. Introduction

□ History of Database Systems (cont.)

✧ **Object-Relational Model** (对象关系模型)

✧ **Object-Relational Database**

✧ **Object-Relational DBMS**

Ch1. Introduction

students

sid	name		class	telephone	enrollment	
	lname	fname			cno	major
1	Jones	Allan	2	555-1234	101	No
					108	Yes
2	Smith	John	3	555-4321	105	No
3	Brown	Harry	2	555-1122	101	Yes
					108	No
5	White	Edward	3	555-3344	102	No
					105	No

Figure 1.1b Object-Relational Student Enrollment Database

Ch1. Introduction

□ 新一代数据管理技术

☞ 新型的应用模式

- **OLAP (On-Line Analytical Processing)**
- **Web应用**

☞ 新型应用的特点

- 大规模、海量数据
- 分析应用
- 高度实时&并发

☞ 解决方案: **CAP** 理论

- **C**onsistency, **A**vailability, **P**artition tolerance
- **NoSQL**

数据库系统分类

关系数据库 RDBMS	Oracle, DB2, Microsoft SQL Server, MySQL, PostgreSQL, ...
键值数据库 Key-value stores	Redis, Memcached, Amazon Dynamo DB, Microsoft Azure Cosmos DB, ...
文档数据库 Document-oriented stores	MongoDB, Couchbase, CouchDB, Amazon Dynamo DB, Microsoft Azure Cosmos DB, ...
图数据库 Graph-oriented DBMS	Neo4j, OrientDB, Microsoft Azure Cosmos DB, ...
列数据库 Wide column store	Cassandra, HBase, ...
其他	面向对象数据库 (object-oriented DBMS), XML数据库, RDF三元组数据库, ...

NoSQL
数据库

DB-Engines: 2020年7 月全球数据库排名



















358 systems in ranking, July 2020

Rank			DBMS	Database Model	Score		
Jul 2020	Jun 2020	Jul 2019			Jul 2020	Jun 2020	Jul 2019
1.	1.	1.	Oracle +	Relational, Multi-model i	1340.26	-3.33	+19.00
2.	2.	2.	MySQL +	Relational, Multi-model i	1268.51	-9.38	+38.99
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model i	1059.72	-7.59	-31.11
4.	4.	4.	PostgreSQL +	Relational, Multi-model i	527.00	+4.02	+43.73
5.	5.	5.	MongoDB +	Document, Multi-model i	443.48	+6.40	+33.55
6.	6.	6.	IBM Db2 +	Relational, Multi-model i	163.17	+1.36	-10.97
7.	7.	7.	Elasticsearch +	Search engine, Multi-model i	151.59	+1.90	+2.77
8.	8.	8.	Redis +	Key-value, Multi-model i	150.05	+4.40	+5.78
9.	9.	↑ 11.	SQLite +	Relational	127.45	+2.64	+2.82
10.	10.	10.	Cassandra +	Wide column	121.09	+2.08	-5.91

完整排名请查看: <https://db-engines.com/en/ranking>

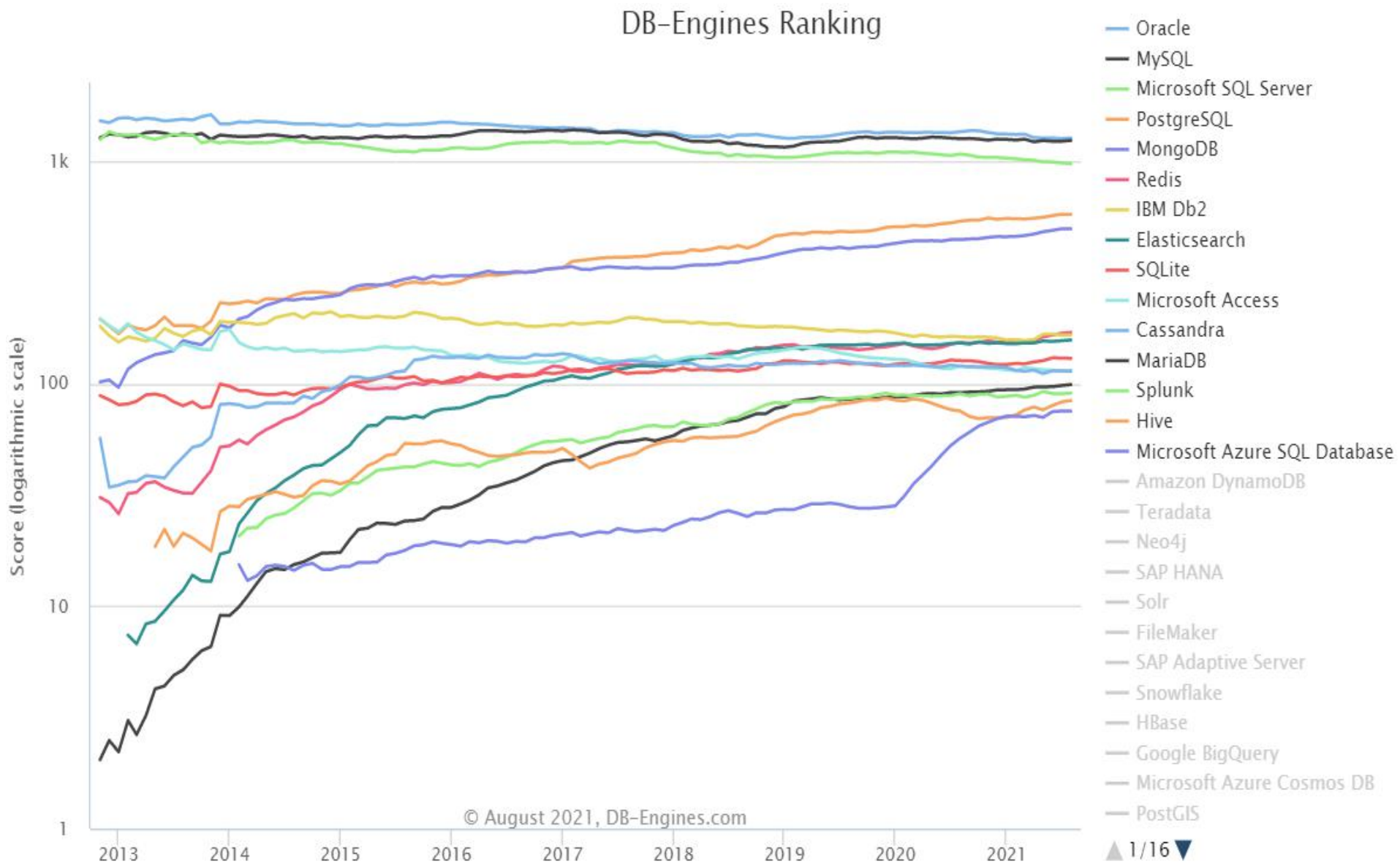
DB-Engines: 2021年8 月全球数据库排名

373 systems in ranking, August 2021

Rank			DBMS	Database Model	Score		
Aug 2021	Jul 2021	Aug 2020			Aug 2021	Jul 2021	Aug 2020
1.	1.	1.	Oracle 	Relational, Multi-model 	1269.26	+6.59	-85.90
2.	2.	2.	MySQL 	Relational, Multi-model 	1238.22	+9.84	-23.36
3.	3.	3.	Microsoft SQL Server 	Relational, Multi-model 	973.35	-8.61	-102.53
4.	4.	4.	PostgreSQL 	Relational, Multi-model 	577.05	-0.10	+40.28
5.	5.	5.	MongoDB 	Document, Multi-model 	496.54	+0.38	+52.98
6.	6.	 7.	Redis 	Key-value, Multi-model 	169.88	+1.58	+17.01
7.	7.	 6.	IBM Db2	Relational, Multi-model 	165.46	+0.31	+3.01
8.	8.	8.	Elasticsearch	Search engine, Multi-model 	157.08	+1.32	+4.76
9.	9.	9.	SQLite 	Relational	129.81	-0.39	+3.00
10.	 11.	10.	Microsoft Access	Relational	114.84	+1.39	-5.02

完整排名请查看: <https://db-engines.com/en/ranking>

数据库引擎排名变化趋势 (2013~2021)



国产数据库的现状

□ “核高基”

- “核心电子器件、高端通用芯片、基础软件产品”
- 是2006年国务院发布的《国家中长期科学和技术发展规划纲要(2006-2020年)》中的16个重大科技专项之一

□ 基础软件：操作系统，数据库，中间件

□ 国产数据库的研究历史

- **2000**年以前：高校、科研院所，研究型
- **2000**年后：高校、科研院所 尝试产业化
- **2010**年以后：以 企业/公司 为研究主体

国产数据库的现状

以高校为背景成立的数据库公司

公司	成立时间	背景
人大金仓KingbaseES系统	1999年	中国人民大学从事科研和教学的专家
武汉达梦数据库	2000年	华中科技大学数据库与多媒体研究所
神通数据库系统	2003年	中国航天和浙江大学
东软OpenBASE系统	2000年	东软承接863计划
南大GBase系统	2004年	南开大学背景


















国产数据库的现状（2010后）

以公司研发为主导的数据库产品

产品	公司	背景
Polar DB	阿里云	基于MySQL开发
OceanBase	蚂蚁金服	
GaussDB	华为	基于PostgreSQL开发
TDSQL/Tbase	腾讯	基于MySQL/PG开发
TiDB	PingCAP	基于Google论文自研
SequoiaDB	巨杉	

□ 标志：2009年，阿里云的成立（去IOE行动）

2020.07 国产数据库排行榜（数据来源： <https://www.modb.pro/>）

排行			名称	类型	国产	开源	得分			热度
本月	6月	5月					本月	6月	5月	
	1	1	TiDB +	关系型			23.02	17.21	15.26	<div><div></div></div>
	2	2	OceanBase +	关系型		-	12.95	13.01	13.36	<div><div></div></div>
	3	3	DM +	关系型		-	12.24	12.41	11.87	<div><div></div></div>
4	4	4	GaussDB +	关系型			11.66	12.13	11.57	<div><div></div></div>
5	5	5	PolarDB +	关系型		-	11.50	11.03	10.56	<div><div></div></div>
6	6	6	GBase +	关系型		-	10.53	9.11	8.74	<div><div></div></div>
7	7	7	TDSQL +	关系型			7.96	7.52	7.43	<div><div></div></div>
8	8	8	TBase +	关系型			6.93	6.75	6.85	<div><div></div></div>
9	9	↑ 10	Kingbase +	关系型		-	6.51	6.42	6.35	<div><div></div></div>
10	10	↑ 11	K-DB +	关系型		-	6.14	6.17	6.18	<div><div></div></div>

2021.08 国产数据库排行榜（数据来源： <https://www.modb.pro/>）

排行			名称	类型	开源	信通院 评测	得分		
本月	7月	6月					本月	7月	6月
	1	1	TiDB +	分布式			630.21	+24.38	+45.78
	2	2	OceanBase +	分布式			493.73	-27.22	+12.69
	↑ 4	↑ 4	达梦 +	关系型	-		403.04	+28.60	+68.05
4	↓ 3	↓ 3	PolarDB +	云原生			399.49	-38.52	-67.33
5	5	↑ 6	openGauss +	关系型			329.43	-9.98	+57.48
6	↑ 7	↑ 7	GaussDB +	云原生	-		307.74	+30.01	+65.01
7	↓ 6	↓↓ 5	GBase +	关系型	-		290.62	+9.25	-38.84
8	8	8	TDSQL +	分布式	-		250.30	+23.16	+57.01
9	9	9	AnalyticDB +	云原生	-		148.71	+5.09	+19.11
10	↑ 11	↑ 11	Kingbase +	关系型	-	-	133.79	+12.33	+24.37

相关阅读

❑ 数据库领域的图灵奖获得者

☞ **1973: Charles William Bachman**（数据库技术）

- 主要贡献：主持设计与开发了最早的网状数据库系统IDS；积极推动与促成数据库标准的制定（DBTG）

☞ **1981: E.F.Codd**（关系数据模型）

- 主要贡献：关系数据模型

☞ **1998: James Gray**（数据库与事务处理）

- 主要贡献：事务处理技术，解决了关系数据库管理系统从理论到产品实现过程中所碰到的一系列技术问题

☞ **2014: Michael Stonebraker**

- 主要贡献：在现代数据库系统概念和实践方面做出的奠基性贡献

数据库界的图灵奖获得者

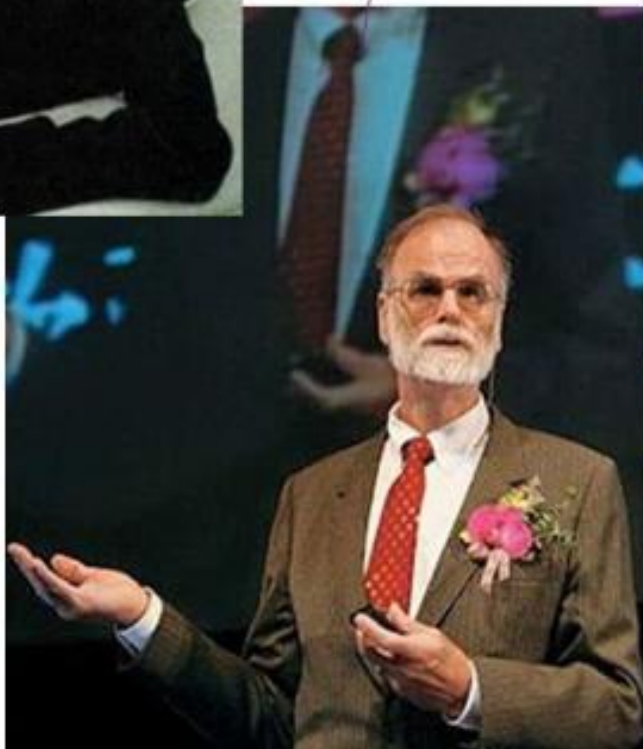


Charles W. Bachman
网状数据库，奠基者兼实践者



Jim Gray
事务处理，数据库系统实现

Michael Stonebraker
现代数据库概念与实践



E. F. Codd
关系数据库



数据库界的四位图灵奖得主