

第一周作业1529721 冯励：

1、阅读《数据库系统教程》第一章，遇到不懂的地方，可以对照其它书一起来看，看后按照自己的思路写出读书小结，上传EOL。要求：小结中需要将你认为重要的新名词、新术语一一列举出来，并给出你对这个名词或术语的理解。罗列出阻碍你深入阅读或深入理解的问题，或者阅读过程中想到的问题；写出你的收获 按照规范的参考文献形式，写出你参考的图书或者文章列表 上传截至时间：2017.2.27早8：15之前 每天阅读一小时，每天进步一点点。

基本术语

1. 数据库(database/ db)

db 是长期存储在计算机内、有组织的、统一管理的相关数据的集合。db 能为各种用户共享，具有较小冗余度、数据间联系紧密而又有较高的数据独立性等特点。

2. 数据库管理系统(database management system / 简称为dbms)

dbms 是位于用户与操作系统之间的一层数据管理软件，它为用户或应用程序提供访问 db 的方法，包括 db 的建立、查询、更新及各种数据控制。

3. 数据库系统 (database system/dbs)

dbs是实现有组织地、动态地存储大量关联数据、方便多用户访问的计算机硬件、软件和数据资源组成的系统，即它是采用数据库技术的计算机系统

4. 数据库技术

数据库技术是研究数据库的结构、存储、设计、管理和使用的一门软件课程.是一门综合性较强的课程。

数据库的发展历史

数据管理技术的发展经历了以下四个阶段:人工管理阶段、文件系统阶段、数据库阶段和高级数据库阶段。

人工管理阶段

1. 数据不保存在计算机内。计算机主要用于计算
2. 没有专用的软件对数据进行管理。每个应用程序都要包括存储结构、存取方法、输入，输出方式等内容。

3. 只有程序(program)的概念, 没有文件(file)的概念。数据的组织方式必须由程序员自行设计与安排。
4. 数据面向程序。即一组数据对应一个程序

文件系统阶段

类似全国大学都有的c语言大作业--学生信息管理系统, csv格式储存, 自己实现增改删查。。

1. 数据以“文件”形式可长期保存在外部存储器的磁盘上。
2. 数据的逻辑结构与物理结构有了区别, 但比较简单。程序与数据之间具有“设备独立性”, 即程序只需文件名就可与数据打交道, 不必关心数据的物理位置。由操作系统的文件系统提供存取方法。
3. 文件组织已多样化。有索引文件、链接文件和直接存取文件等。但文件之间相互独立、缺乏联系。数据之间的联系要通过程序去构造。
4. 数据不再属于某个特定的程序, 可以重复使用。

但与此同时, 文件系统也有三个明显的缺点。

1. 数据冗余
2. 数据不一致
3. 数据联系弱

由于这些原因, 人们只好研究新的数据库理论...

数据库阶段

20世纪 60 年代末, 磁盘技术取得重要进展, 具有数百兆字节容量和快速存取的磁盘陆续进入市场, 成本也不高, 这就为数据库技术的产生提供了良好的物质条件。数据管理技术进入数据库阶段的标志是 20 世纪 60 年代末的三件大事:

1. 1968年美国 IBM 公司推出层次模型的 IMS 系统;
2. 1969年美国 CODASYL 组织发布了 DBTG 报告。总结了当时各式各样的数据库, 提出网状模型;
3. 1970年美国 IBM 公司的 E.f.Codd 连续发表论文, 提出关系模型, 奠定了关系数据库的理论基础。

数据库阶段的数据管理具有以下特点:

1. 采用数据模型表示复杂的数据结构。数据模型不仅描述数据本身的特征, 还要描述数据之间的联系。
2. 有较高的数据独立性。数据的逻辑结构与物理结构之间的差别可以很大。用户以简单的逻辑结构操作数据而无需考虑数据的物理结构。
3. 数据库系统为用户提供了方便的用户接口。
4. 增加了系统的灵活性:对数据的操作不一定以记录为单位, 可以以数据项为单位。从文件系统发展到数据库系统是信息处理领域的一个重大变化。数据结构的设计成为信息系统首先关心的问题, 而

利用这些数据的 应用程序设计则退居到以既定的数据结构为基础的外围地位。

高级数据库技术阶段

这一阶段的主要标志是 80 年代的分布式数据库系统、90 年代的对象数据库系统和 21 世纪初的网络数据库系统的出现。

数据描述

在数据处理中，数据描述将涉及到不同的范畴。从事物的特性到计算机中的 具体表示，实际上经历了三个阶段——概念设计中的数据描述、逻辑设计中的数据 描述和物理设计中的数据描述

概念设计中的数据描述

entity

客观存在，可以相互区别的事物称为实体。实体可以是具体的对象，例如一名男学生，一辆汽车等。也可以是抽象的对象，例如一次借书，一场足球比赛等。

entity set

性质相同的同类实体的集合称为实体集。例如所有的男学生，全国足球锦标赛的所有比赛等。

attribute

实体有很多特性，每一个特性称为属性。每个属性有一个值域，其类型可以是整数型、实数型、字符串型等。例如学生有学号、姓名、年龄、性别等属性。

identifier

能惟一标识实体的属性或属性集的标识符称为实 体标识符。有时也称为关键码(key)，或简称为键。例如学生的学号可以作为学生 实体的标识符。

逻辑设计中的数据描述

field

标记实体属性的命名单位称为字段，或数据项。它是可以命 名的最小信息单位，所以又称为数据元素或初等项。字段的命名往往和属性名相 同。例如学生有学号、姓名、年龄、性别等字段。

record

字段的有序集合称为记录。一般，用一个记录描述一个实 体，所以记录又可以定义为能完整地描述一

一个实体的字段集。例如一个学生记录，由有序的字段集组成:(学号，姓名，年龄，性别)。

file

同一类记录的集合称为文件

key

能惟一标识文件中每个记录的字段或字段集，称为记录的 关键码(简称为键)。

物理设计中的数据描述

数据库的物理设计是设计数据库的物理结构

熟悉的：

bit , byte, word , block ,

bucket

外存的逻辑单位，一个桶可以包含一个物理块或多个在空间 上不一定连续的物理块

volume

一个输入@输出设备所能装载的全部有用信息，称为“卷”。 例如磁带机的一盘磁带就是一卷，磁盘的一个盘组也是一卷。

数据联系的描述

一对一，一对多， 多对多 ， 一元，二元，三元 ...n 元

数据模型的定义

这里的定义不是用DDL定义，说的是是数学定义。。。

逻辑数据模型应包含数据结构、数据操作和数据完整性约束三个 部分.

entity relationship model (er图)

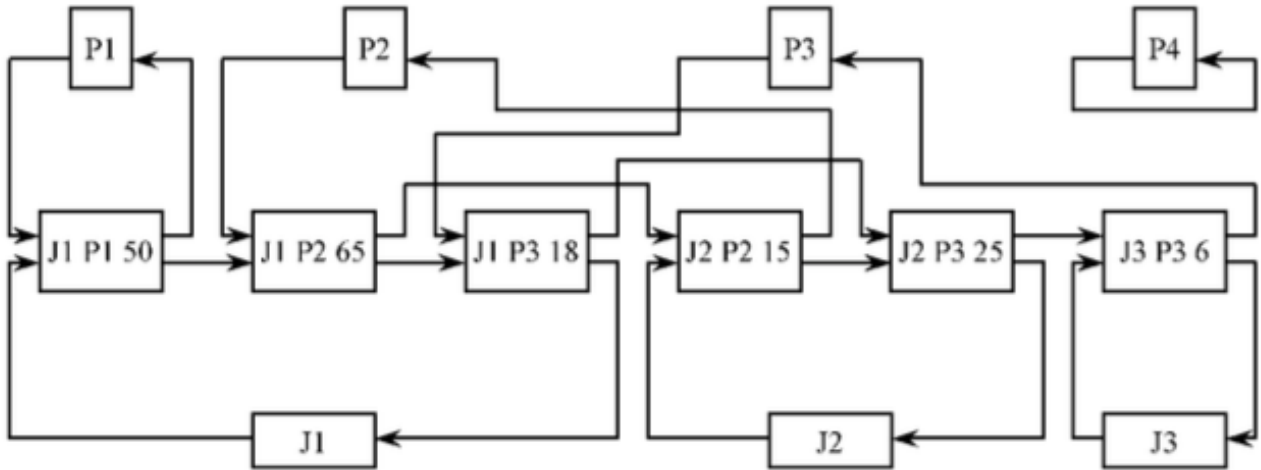
没什么可说的。。。这东西还有个发明人。。。

层次模型

层次模型面对的是逻辑数据 而不是物理数据，用户不必花费大量的精力考虑数据的物理细节

只能表示 1:n 联系，虽然系统有多种辅助手段 实现 n:m 联系，但较复杂，用户不易掌握;且由于层次顺序的严格和复杂，引起数据的查询和更新操作很复杂，因此应用程序的编写也比较复杂。

网状模型



特点是记录之间联系通过指针实现,m:n 关系容易实现

缺点是数据结构复杂和编程复杂

由于层次系统和网状系统的天生缺点，因此从 20 世纪 80 年代中期起其市场已被关系模型取代

关系模型

relational model 的主要特征是用二维表格表示实体集。与前两种 模型相比，关系模型数据结构简单，容易为初学者理解

PART 关系

P #	PNAME	COLOR	WEIGHT
P1	SCREW	BLUE	14
P2	BOLT	GREEN	17
P3	NUT	RED	12
P4	SCREW	RED	19

PROJECT 关系

J #	JNAME	DATE
J1	JA	89.1
J2	JB	90.5
J3	JC	91.3

SUPPLIER 关系

S #	SNAME	SADDR
S1	PICC	SHANGHAI
S2	FADC	BEIJING

P_P 关系

J #	P #	TOTAL
J1	P1	50
J2	P2	15
J3	P3	6
J1	P2	65
J2	P3	25
J1	P3	18

P_S 关系

P #	S #	QUANTITY
P1	S1	100
P2	S2	200
P2	S1	150
P3	S2	300
P4	S2	100

图 1.22 关系模型的具体实例



对象模型

对象数据库是面向对象概念与 数据库技术相结合的产物。

object model中基本的概念是对象和类。

继承，重用 等等 对象数据库尚未达到关系数据库的普及程度。

	层次模型	网状模型	关系模型	对象模型
创始	1968 年 IBM 公司的 IMS 系统	1969 年 CODASYL 的 DBTG 报告 (71 年通过)	1970 年 E.F.Codd 提出关系模型	20 世纪 80 年代
数据结构	复杂 (树结构)	复杂 (有向图结构)	简单 (二维表)	复杂 (嵌套, 递归)
数据联系	通过指针	通过指针	通过表间的公共属性	通过对象标识
查询语言	过程性语言	过程性语言	非过程性语言	面向对象语言
典型产品	IMS	IDS/ II , IMAGE/3000, IDMS, TOTAL	Oracle, Sybase, DB2, SQL Server, Informix	ONTOS DB
盛行期	20 世纪 70 年代	70 年代至 80 年代中期	80 年代至现在	90 年代至现在

图 1.24 四种逻辑数据模型的比较

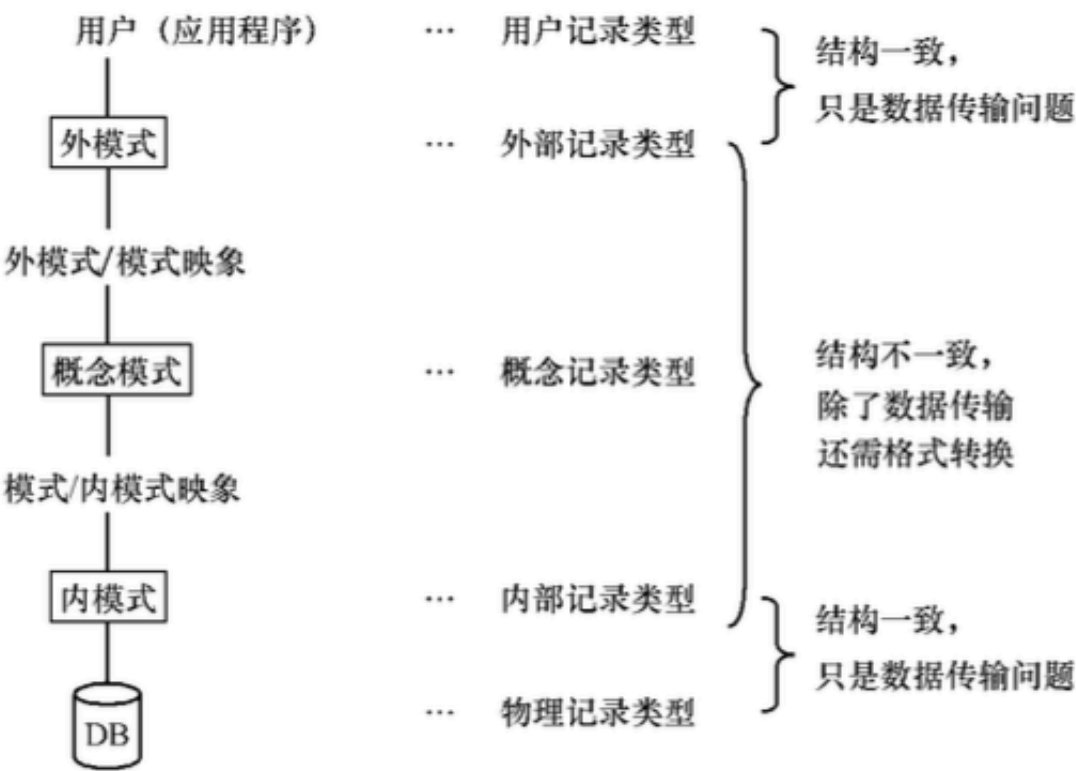
数据库的体系结构

三级体系结构 五大要素 两级数据独立性

external , conceptual , internal ..通俗易懂..

	数据模型	用数据定义语言 描述后的称呼	DBTG 报告 中的称呼
外部级	外模型	外模式	子模式
概念级	概念模型	概念模式	模式
内部级	内模型	内模式	物理模式

图 1.26 数据抽象的术语



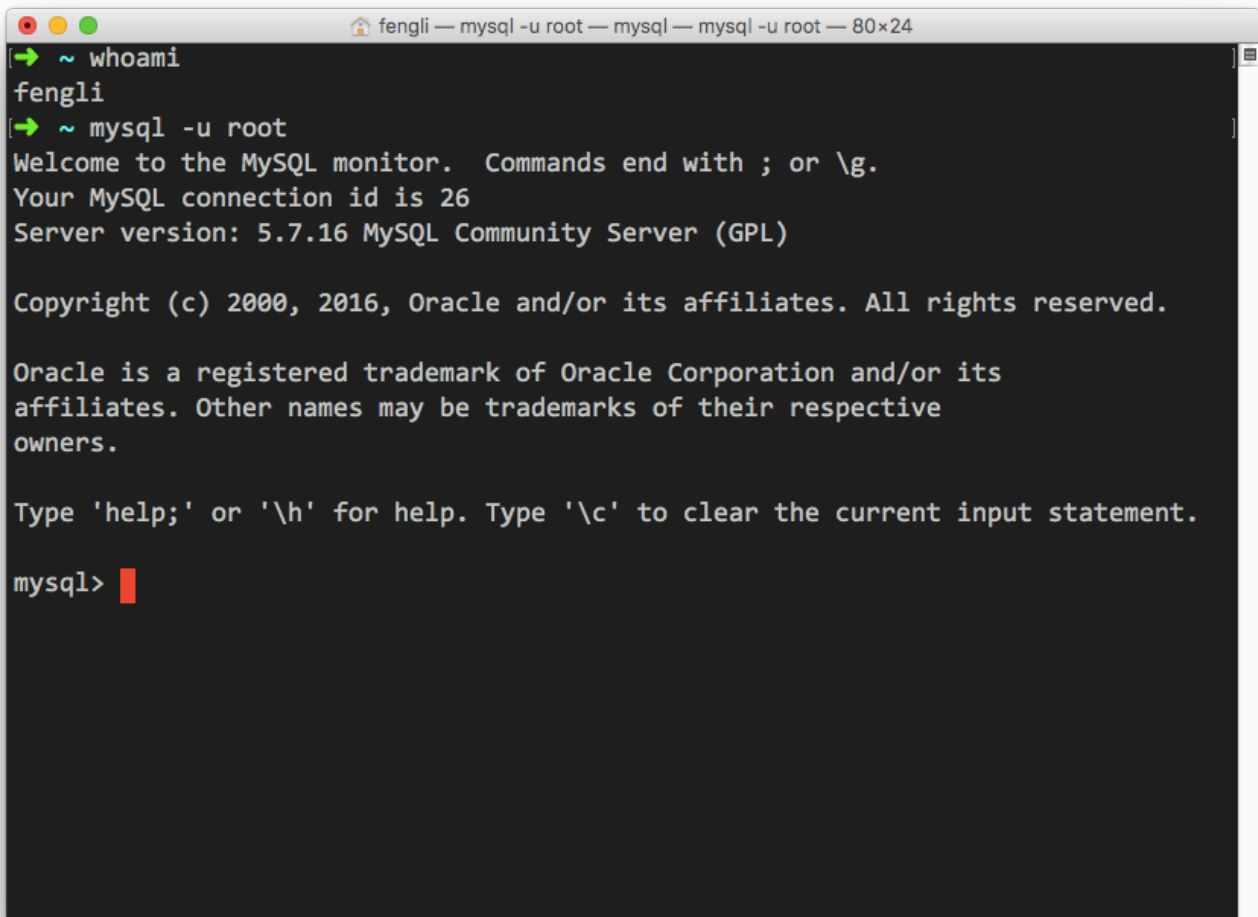
DBMS

DBMS 的功能：

1. 定义功能， 提供DDL
2. 操纵功能， DML

3. 保护功能
4. 维护功能
5. 数据字典,对数据库的操作都要通过 DD 才能实现。DD 中还存放数据库运行时的统计信息,例如记录个数、访问次数等。管理 DD 的子系统称为“DD 系统” DBMS 由两大部分组成:查询处理器和存储管理器

终端mysql:



```
fengli ~ whoami
fengli
fengli ~ mysql -u root
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 26
Server version: 5.7.16 MySQL Community Server (GPL)

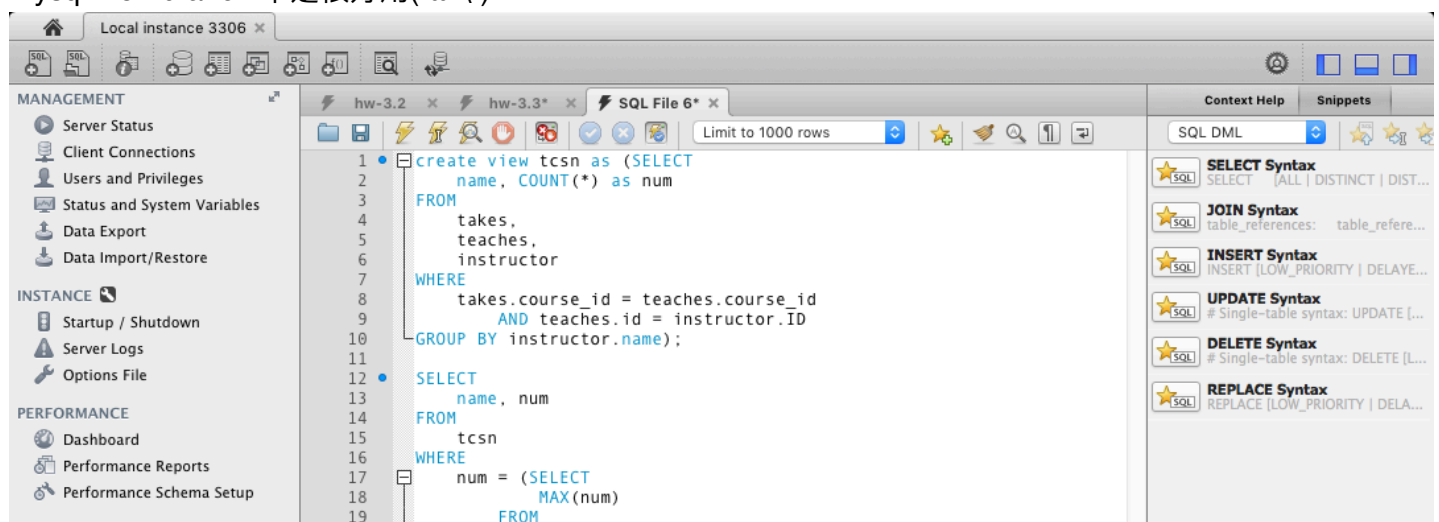
Copyright (c) 2000, 2016, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
```

mysql-workbranch 不是很好用(/ω\)



Filter objects

a-small-university

information_schema

mysql

performance_schema

tysql

Object Info

Session

Schema: a-small-university

20

tcscn);

21

drop view tcscn;

22

23

24

25

SELECT

26

instructor.name, salary

27

FROM

28

instructor

29

WHERE

30

instructor.salary > (SELECT

31

MAX(instructor.salary)

32

FROM

33

instructor

34

GROUP BY dept_name

35

HAVING dept_name = 'english');

36

37

SELECT

38

dept_name, COUNT(*)

39

FROM

40

instructor

41

GROUP BY dept_name

42

HAVING COUNT(*) >= ALL (SELECT

43

COUNT(*)

44

FROM

45

instructor

46

GROUP BY dept_name);

47

48

49

SELECT

50

*

51

FROM

52

course

53

WHERE

54

course.course_id NOT IN (SELECT

55

takes.course_id

56

FROM

57

takes);

58

SELECT

59

*

60

FROM

61

takes

62

WHERE

63

takes.course_id NOT IN (SELECT

64

course.course_id

65

FROM

66

course);

100%

20:66

Result Grid

Filter Rows:

Search

Edit:

Export/Import:

ID	course_id	sec_id	semester	year	grade
00128	CS-101	1	Fall	2009	A
00128	CS-347	1	Fall	2009	A-
1000	239	1	Fall	2006	C
1000	319	1	Spring	2003	B+
1000	362	1	Fall	2005	B+
1000	493	1	Spring	2010	A-
1000	571	1	Spring	2004	C+
1000	642	1	Fall	2004	C-
1000	663	1	Spring	2005	C+
1000	696	1	Spring	2002	B
1000	748	1	Fall	2003	A
1000	802	1	Spring	2003	C+
1000	959	1	Fall	2006	B+
1000	962	1	Spring	2008	B-
1000	972	1	Spring	2009	B+
10033	242	1	Fall	2009	B
10033	334	1	Fall	2009	C-
10033	338	1	Spring	2007	A
10033	338	2	Spring	2006	C
10033	352	1	Spring	2006	A
10033	362	3	Spring	2008	A-
10033	408	1	Spring	2007	C-
10033	408	2	Spring	2003	B+
10033	443	2	Spring	2002	C-
10033	445	1	Spring	2001	C
10033	457	1	Spring	2001	C-
10033	486	1	Fall	2009	C
10033	493	1	Spring	2010	C-
10033	603	1	Fall	2003	B

Result 36

takes 37

Apply

Revert

Action Output

	Time	Action	Response	Duration / Fetch Time
✓ 140	11:02:24	select disti...	114 row(s) returned	0.031 sec / 0.000032...
✓ 141	11:03:12	select disti...	114 row(s) returned	0.032 sec / 0.00003...
✓ 142	11:03:12	select * fro...	1000 row(s) returned	0.0010 sec / 0.00045...
✓ 143	11:03:16	select disti...	114 row(s) returned	0.037 sec / 0.000023...
✓ 144	11:03:16	select * fro...	1000 row(s) returned	0.00067 sec / 0.0003...

Result Grid

Form Editor

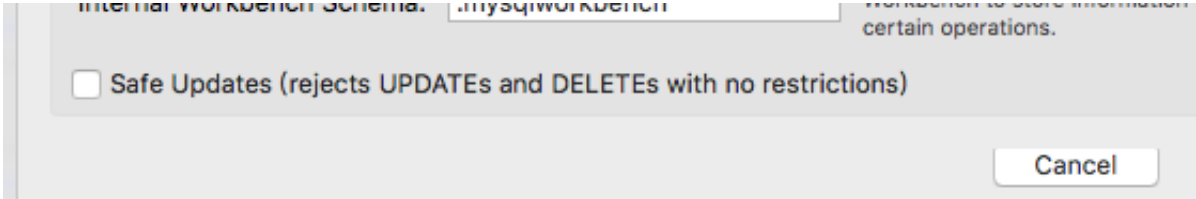
Field Types

Query Stats

Execution Plan

Formatted 1 statements.

直观的数据库保护，会阻止不带where使用update。。。。



数据库系统

由数据库， 硬件， 软件， 管理员组成 DBMS和磁盘间还有操作系统

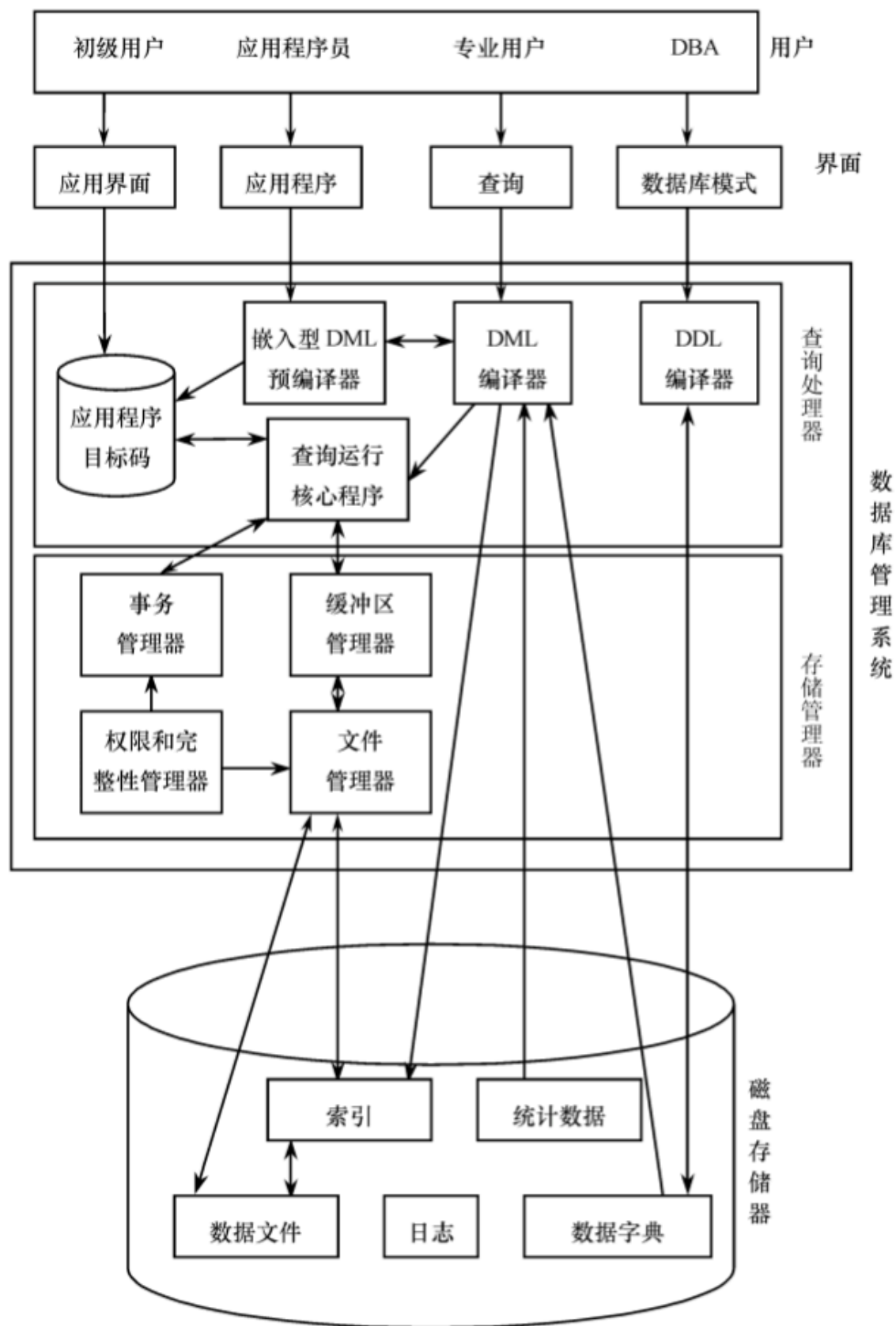


图 1.31 DBS 的全局结构

DBS 有 centralized, client/server, parallel, distributed 等等种类

sentiment

这一章应该只是打算给读者一个大致印象吧？难度应该在后面。。 总之，我认为这门课十分重要，希望这学期在老师指导下能打好基础