

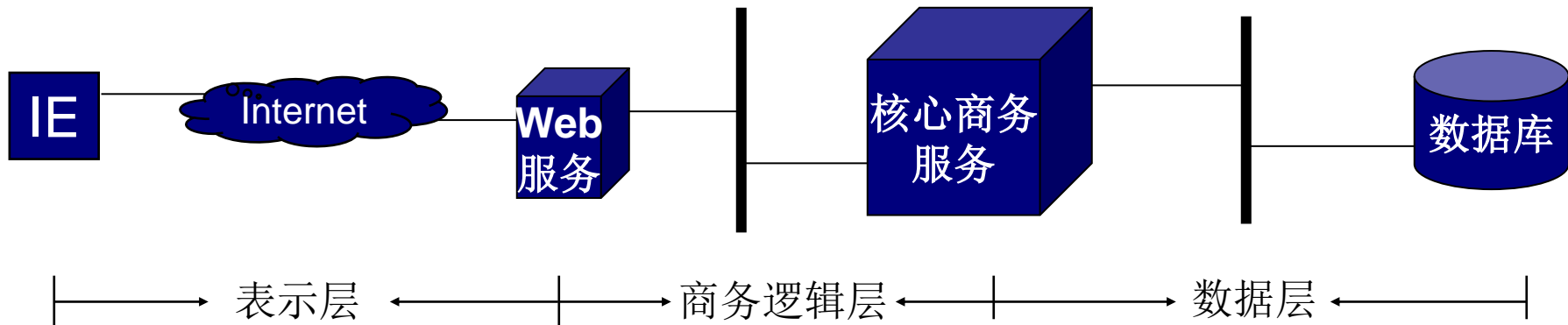


电子商务技术基础

第四章 商务数据层及其技术

第四章 商务数据层及其技术

商务数据层的任务和核心



□ 数据层的任务

- 负责对数据的管理，也就是对数据进行分类、组织、编码、存储、检索和维护等

□ 数据层的核心

- 数据库系统是数据层的核心，通过数据访问接口与表达层、逻辑层进行交互。

第四章 商务数据层及其技术

- 第一节 数据管理技术的发展
- 第二节 电子商务系统中的数据管理技术
- 第三节 电子商务系统数据层实现
- 第四节 电子商务个性化推荐技术



第一节 数据管理技术的发展

- 数据管理是指对数据的分类、组织、编码、存储、检索和维护，而数据库技术正是数据处理技术发展比较成熟后的产物。
- 数据库技术是电子商务的一项支撑技术，在电子商务的建设中占有重要的地位。
- 数据管理技术发展的四个阶段
 - 手工管理 — 20世纪50年代以前
 - 文件系统 — 20世纪50~60年代
 - 数据库系统 — 20世纪70~90年代
 - 数据仓库与商务智能 — 20世纪90年代至今

第一节 数据管理技术的发展

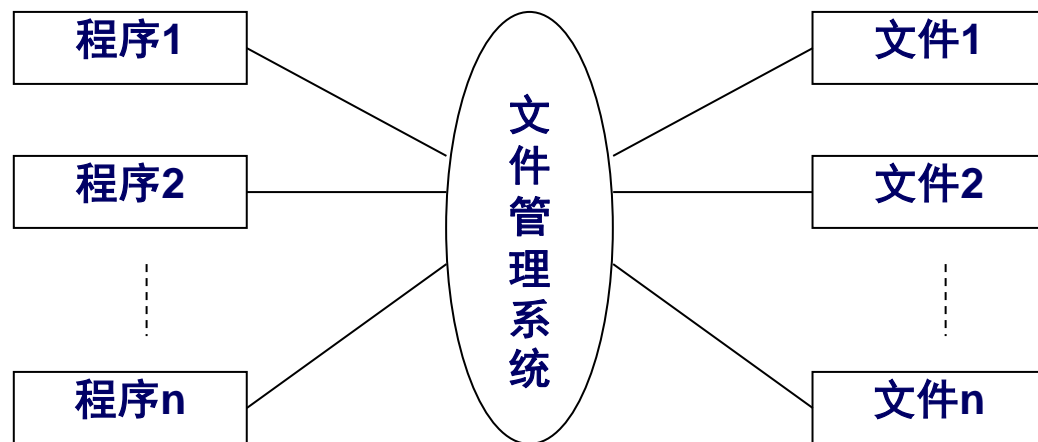
1、手工管理阶段

- 软硬件条件：无DMA(Direct Memory Access, 直接内存存取)设备、无OS
- 应用目的：科学计算
- 数据组织方式：手工
- 特点：数据不保存、数据无结构、数据无独立性、数据不能共享。

第一节 数据管理技术的发展

2、文件管理阶段

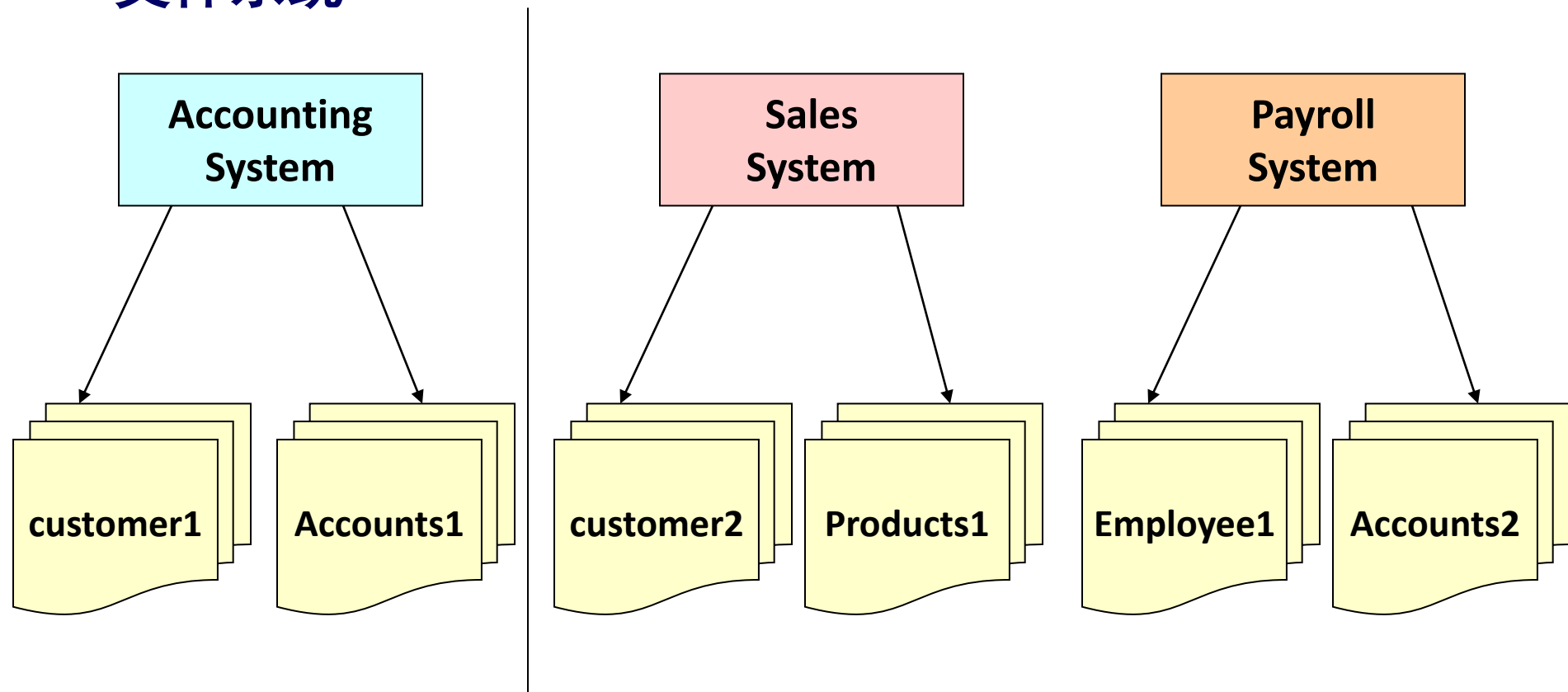
- 软硬件条件：有DMA设备、有OS
- 应用目的：科学计算、数据管理
- 数据组织方式：文件系统
- 特点：数据可保存、结构化不强、独立性较差、数据不能共享。



第一节 数据管理技术的发展

2、文件管理阶段

文件系统



第一节 数据管理技术的发展

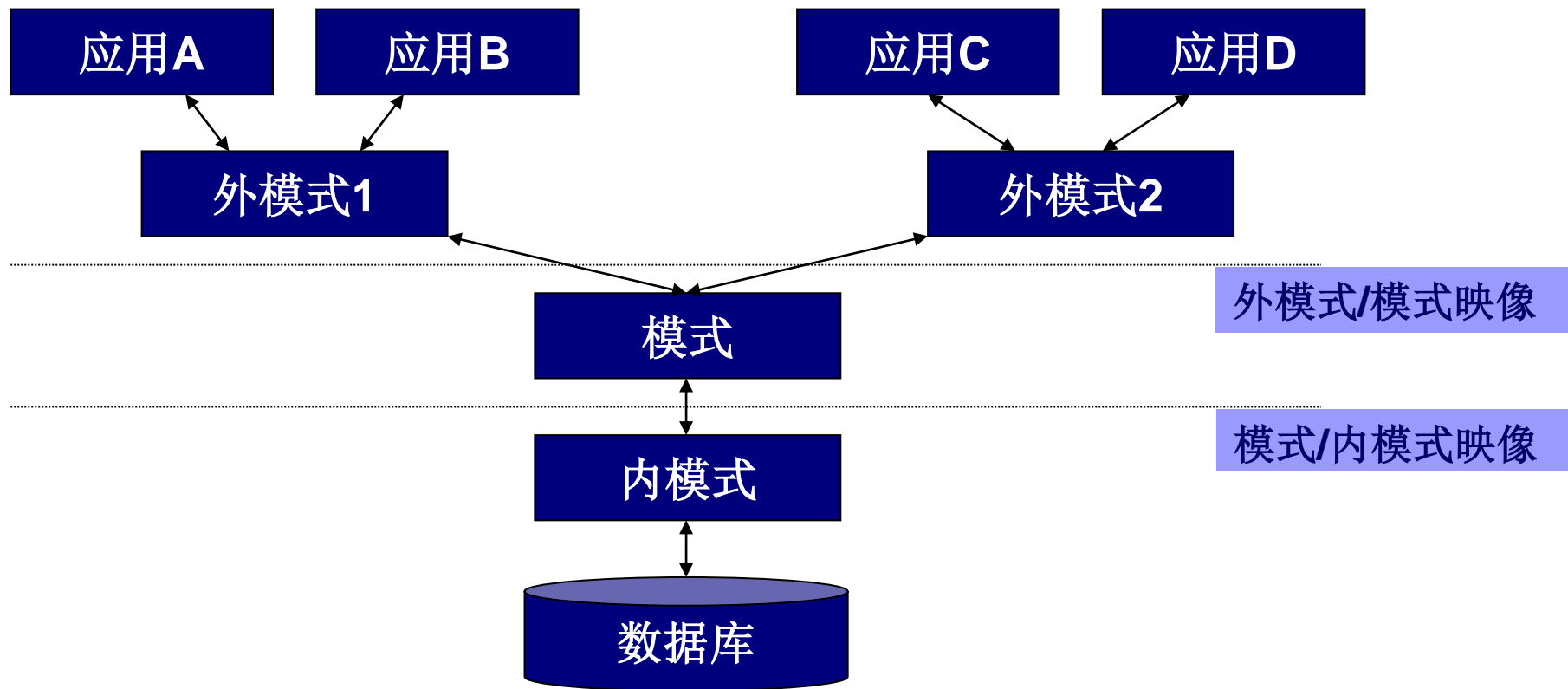
3、数据库系统阶段

- 软硬件条件：大容量快速存储设备、多用户多任务OS
- 应用目的：科学计算、OLTP（联机事务处理）、多媒体应用
- 数据组织方式：数据库三级模式
- 特点：数据保存在数据库、数据高度结构化、数据高度独立、数据共享、冗余小、数据统一控制。
- 缺点：不能满足OLAP（联机分析处理）的需求。

第一节 数据管理技术的发展

3、数据库系统阶段

数据库的三级模式——提高数据的逻辑独立性和物理独立性



第一节 数据管理技术的发展

3、数据库系统阶段

模式

- 也称**逻辑模式**，是数据库中全体数据的逻辑和特征的描述，是所有用户的公共数据视图。
- 全体用户数据的最小并集。
- 定义模式时不仅要定义数据的逻辑结构。
- 要定义数据之间的联系，定义数据有关的安全性、完整性要求。
- 它是数据库系统模式结构的中间层，一个数据库只有一个模式。
- 不涉及数据的物理存储细节。

第一节 数据管理技术的发展

3、数据库系统阶段

外模式

- 也称子模式(Subschema)或用户模式，它是数据库用户(包括应用程序员或最终用户)能够看见和使用的局部数据的逻辑结构和特征的描述。
- 是数据库用户的数据视图。
- 是与某一应用有关的数据的逻辑表示。
- 外模式通常是模式的子集，一个数据库可以有多个外模式。
- 外模式是保证数据库安全性的有力措施。

第一节 数据管理技术的发展

3、数据库系统阶段

内模式

- 也称存储模式(Storage Schema)，一个数据库只有一个内模式。
- 是数据物理结构和存储方式的描述，是数据在数据库内部的表示方式，例如数据是顺序存储还是树结构存储、是否压缩、是否加密等。
- 一个数据库只有一个内模式。
- 关系模型中，用户一般不需要描述内模式。

第一节 数据管理技术的发展

3、数据库系统阶段

■ 采用三级模式结构必要性：

- 减少了数据的冗余，实现了数据的共享 (模式)
- 定义完整性约束条件，保证数据正确、相容 (模式)
- 保证数据的安全性 (外模式)
- 实现了程序与数据的独立性

■ 数据库系统的三级模式是数据的三个抽象级别，把数据的具体组织留给数据库管理系统管理。数据库管理系统在三级模式之间提供两层映象：

- 外模式 / 模式映象
- 模式 / 内模式映象

第一节 数据管理技术的发展

3、数据库系统阶段

- 两层映象保证了数据库中的数据能够具有较高的逻辑独立性和物理独立性。
- 程序与数据的逻辑独立性
 - 定义外模式时，要定义外模式/模式映象；
 - 当模式改变时，修改变外模式/模式映象；
 - 外模式保持不变，从而不必修改应用程序。
- 程序与数据的物理独立性
 - 数据库中只有一个模式和一个内模式；
 - 模式/内模式映象只有一个；
 - 当内模式改变时，可改变模式/内模式映象；
 - 使得模式保持不变，从而不必修改应用程序。

第一节 数据管理技术的发展

4、数据仓库与商务智能

- 软硬件条件：C/S技术成熟。
- 应用目的：OLAP、决策支持、数据挖掘。
- 数据组织方式：多维数据。
- 特点：数据冗余大，资源占用多。

第一节 数据管理技术的发展

4、数据仓库与商务智能

- **联机事务处理**(Online transaction process, **OLTP**): 对数据库进行联机的日常事务处理，即利用数据库系统，对日常业务的记录数据进行维护，如进行记录的增加、删除、修改及简单查询。
- **联机分析处理**(Online analysis process, **OLAP**): 从企业已有的大量数据中提取决策信息并对之进行管理的服务。
- 例：售货员通过POS机等前端设备处理顾客的购买活动，属于事务处理；决策者通过分析大量的顾客购买信息决定商品的摆放位置，则属于分析处理。

第一节 数据管理技术的发展

4、数据仓库与商务智能

■ 事务处理与分析处理的区别

	事务处理	分析处理
应用领域	面向应用，日常管理和控制	面向分析，决策支持分析
主要功能	各部门的OLTP	OLAP（长期趋势分析）
需求特征	确定的应用需求	基于决策分析的主题
操作特点	操作频率高、处理时间短，增、删、改、查询均有	运行时间长、资源消耗大，以查询为主
数据特征	当前数据、细节数据	历史数据、综合数据
数据来源	组织外部	系统内部
数据组织	按业务，分散	按主题，集成

第一节 数据管理技术的发展

4、数据仓库与商务智能

- 基于OLTP的传统数据库系统无法满足OLAP需求，表现在以下三方面：
 - 无法满足对大量历史数据和综合数据的有效管理。
 - 辅助决策信息涉及许多部门的数据，而不同系统的数据难以集成。
 - 由于访问数据的能力不足，所以它对海量数据的访问性能明显下降。

第一节 数据管理技术的发展

4、数据仓库与商务智能

■ 数据仓库定义

- 数据仓库是支持管理决策过程的、面向主题的、集成的、随时间变化的、但信息本身相对稳定的数据集合。
- 包括了数据仓库生成技术、联机分析处理(从数据仓库中提取决策数据)和数据挖掘技术(从数据库和数据仓库中进行深度数据分析, 以实现决策支持的各种要求)三个方面的内容。

■ 数据仓库工具有两类:

- 联机分析处理工具, 用户可以从数据仓库中提取决策数据。
- 数据挖掘工具, 可以从数据库和数据仓库中进行深度数据分析, 实现决策支持的各种要求。

第一节 数据管理技术的发展

4、数据仓库与商务智能

■ 数据仓库特点

- **数据仓库是面向主题的：**其中的数据按照一定的主题域进行组织。主题是指用户使用数据仓库进行决策时所关心的重点方面，是与传统数据库的面向应用相对应的，是在较高层次上将企业信息系统中的数据综合、归类并进行分析利用的抽象。每一个主题对应一个宏观的分析领域。数据仓库排除对于决策无用的数据，提供特定主题的简明视图。
- **数据仓库是集成的：**数据仓库的数据有来自于分散的操作型数据，将所需数据从原来的数据中抽取出来，进行加工与集成，统一与综合之后才能进入数据仓库。
- **数据仓库是不可更新的：**数据仓库主要是为决策分析提供数据，所涉及的操作主要是数据的查询。

第一节 数据管理技术的发展

4、数据仓库与商务智能

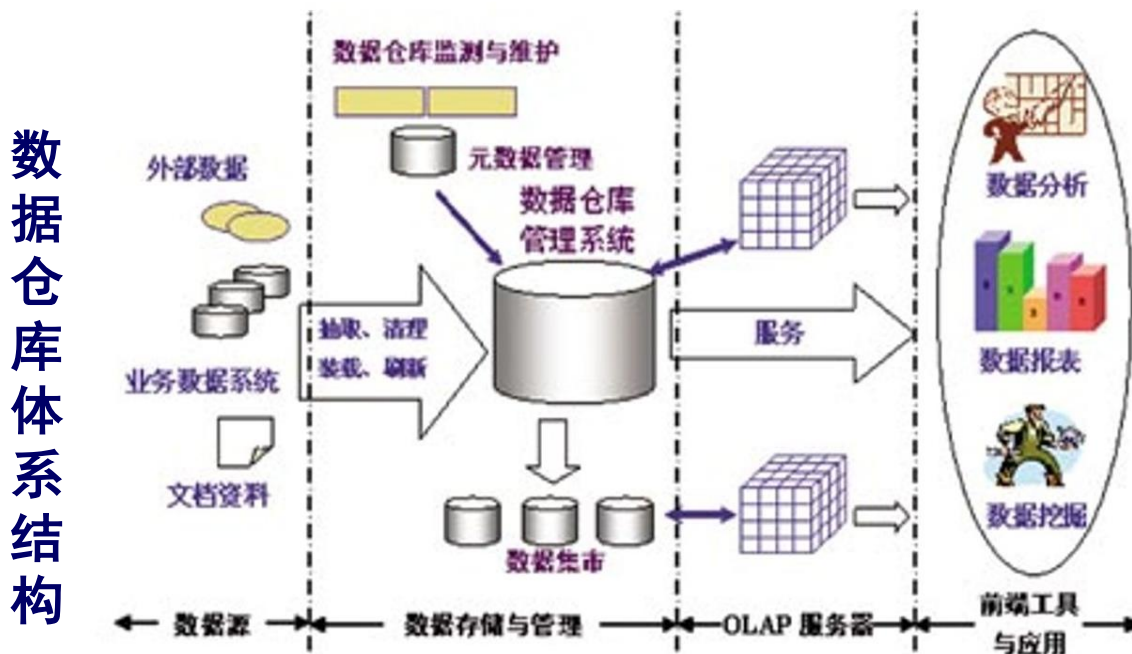
■ 数据仓库特点

- **数据仓库是随时间而变化的：**传统的关系数据库系统比较适合处理格式化的数据，能够较好的满足商业商务处理的需求。稳定的数据以只读格式保存，且不随时间改变。
- **汇总的：**操作性数据映射成决策可用的格式。
- **大容量：**时间序列数据集合通常都非常大。
- **非规范化的：**数据仓库的数据可以是而且经常是冗余的。
- **元数据：**将描述数据的数据保存起来。
- **数据源：**数据来自内部的和外部的非集成操作系统。

第一节 数据管理技术的发展

4、数据仓库与商务智能

- 数据库是一种通用平台，而数据仓库不是，仍建立在数据库管理系统基础上，只是一个新概念，在此概念下进行的构造过程称为数据仓库处理，所以数据仓库系统是一种解决问题的过程，而不是可买的现成产品。



第一节 数据管理技术的发展

5、数据库技术发展各阶段综合比较

	手工处理	文件系统	数据库系统	数据仓库
软硬件技术条件	无存储设备、操作系统	有存储设备、操作系统	技术的成熟，在管理领域的大量使用	C/S技术成熟
信息需求	无	具体业务操作	企业内信息共享各部门的OLTP	辅助决策信息
应用目标	科学计算	单项业务管理	面向应用，日常管理和控制	面向分析，决策支持分析
数据特征	极少	作业数据	当前数据、细节数据	历史数据、综合数据
数据组织方式	手工、纸制，不存储	直接面向应用	数据的结构化 数据库的三级模式	基于主题的、集成的多维或关系数据库
数据处理功能	无	数据的增、删、改	对日常作业数据的增、删、改、查询	运行时间长、资源消耗大，以查询为主
数据管理技术特征	无	数据对于程序的强烈依赖性	通用数据平台（数据逻辑性与处理能力	DW、OLAP、DM技术为主的解决方案
优点		提供简单的数据管理和复制功能	数据的集成、共享，冗余少，具有独立性	决策支持能力强
缺点		结构化和逻辑性不强，不能共享	难以满足联机分析处理	数据冗余少，资源占用多

第四章 商务数据层及其技术

- 第一节 数据管理技术的发展
- **第二节 电子商务系统中的数据管理技术**
- 第三节 电子商务系统数据层实现
- 第四节 电子商务个性化推荐技术



第二节 电子商务系统中的数据管理技术

1、电子商务系统中的数据管理需求

- 电子商务应用的推广和发展对数据库技术提出了新的要求，推动了数据库技术与Web技术的全面结合，在数据库核心技术和数据库访问技术等方面得到了发展。

第二节 电子商务系统中的数据管理技术

1、电子商务系统中的数据管理需求

- **数据**：原始记录，符号或数字形式。
- **信息**：数据经加工后转变为可使用的形式；如超文本、文本、图形、图像、音频/视频等。
- **知识**：人类索取、结构化和使用信息的能力；主要由企业文化、社会网络及解决问题的模型构成，只有模型可存储。

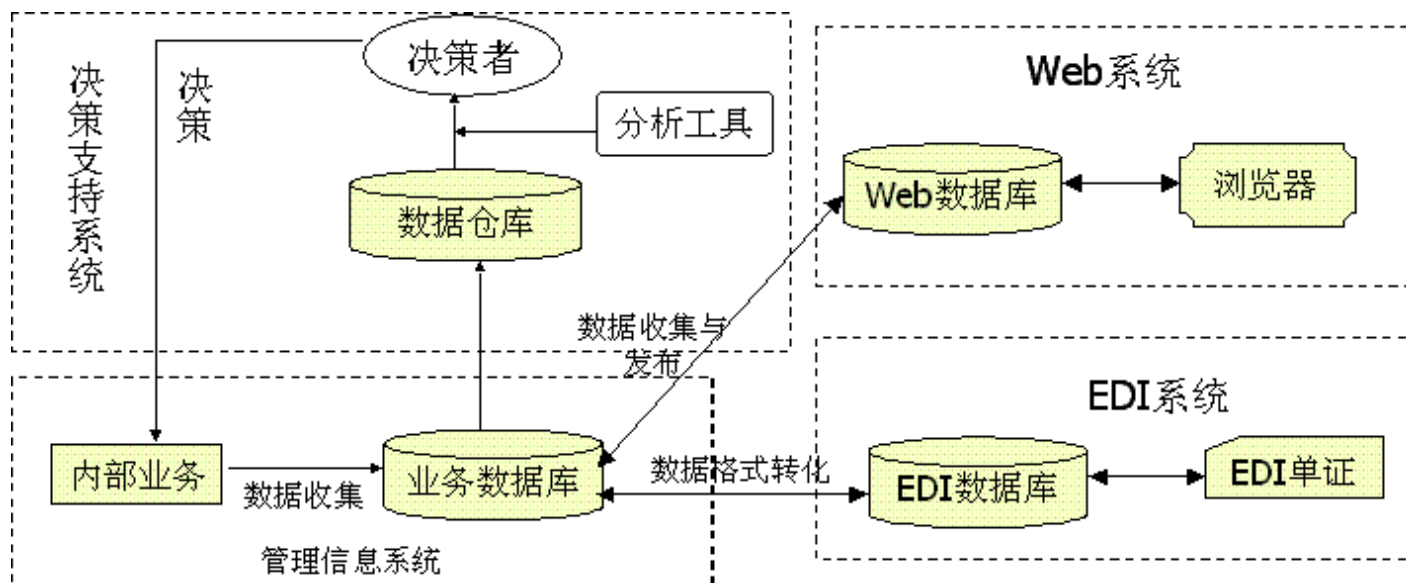
结构化和半结构化的组织存储

存储类型	数据	信息	知识
结构化	数据库、数据仓库	带索引网站、报表、图表、手册	虚拟团队、文档数据库、专家系统、常见问题、新闻组
半结构化	未组织的列表	网页、E-mail	公告版、聊天组

第二节 电子商务系统中的数据管理技术

2、数据库技术对电子商务的支持

- 数据的收集、存储和组织
- 决策支持
- 对EDI的支持
- Web与数据库的结合



第二节 电子商务系统中的数据管理技术

3、电子商务对数据库技术的挑战

- 管理内容：信息种类繁多，结构化/半结构化并存。
 - 数据模型：关系模型力不从心，关系对象模型大行其道。
 - 性能方面：高度的可靠性和极高的响应速度。
 - 体系结构：需分担一部分应用逻辑处理功能，支持更多数据及流量。
-
- 电子商务系统建立在Web基础上，只有Web技术与数据库有机地结合在一起，才能满足电子商务系统中的数据管理需求，发挥二者各自的优势，弥补它们各自的缺陷。

第四章 商务数据层及其技术

- 第一节 数据管理技术的发展
- 第二节 电子商务系统中的数据管理技术
- **第三节 电子商务系统数据层实现**
- 第四节 电子商务个性化推荐技术



第三节 电子商务系统数据层实现

1、数据模型

- 电子商务系统中数据为平台建设主要采用已有的数据库技术方法，**数据模型是数据库组织和管理的基础。**
- **数据模型**是描述一个组织中的数据、数据之间的关系，及对数据约束的一组完整的概念，是对数据库的结构与语义的描述，对现实世界的抽象，是**数据库系统的核心和基础。**
- 主要的数据库模型有：
 - 层次模型(Hierarchical Model)
 - 网状模型(Network Model)
 - 关系模型(Relational Model)
 - 面向对象模型(Object-Oriented Model)

关系型数据库常用平台：Oracle, Sybase, DB2, SQL Server

第三节 电子商务系统数据层实现

1、数据模型

■ 关系数据模型

- 组成要素：数据结构+数据操作+数据完整性

■ 完整性类型有4类：

- 实体完整性：是为了唯一的确定表中的一条记录，因此实体完整性要求充当主码的字段值应唯一且不能为空。
- 参照完整性：要求参照字段（参照码）的取值只能取被参照的字段（某个表中的主码）中的值，或者索性为空。
- 域完整性：(通过CHECK约束、DEFAULT约束)限制字段的数据类型、格式、取值范围等。
- 用户自定义完整性。

第三节 电子商务系统数据层实现

1、数据模型

例1：数据结构定义与SQL查询

- 在SQL Server中建立一个数据库mytest，包含5个表：教师Teacher、学生Student、课程Course、授课TC、选课SC，其关系模式如下：
 - **Teacher** (教师号Tno，姓名Tname，年龄Tage，性别Tsex，系名Tdept，电话号码Ttel，邮件地址E-mail,家庭地址Taddr)；
 - **Student** (学号Sno，姓名Sname，年龄Sage，性别Ssex，系名Sdept)
 - **Course** (课程号Cno，课程名Cname，学分数Ccredit，课时数Chour)
 - **TC** (教师号Tno，课程号Cno，任课学年Ayear，任课学期Semester，评教等级Rating)
 - **SC** (学号Sno，课程号Cno，成绩Grade)
- 将Teacher.Tno、Student.Sno、Course.Cno设为主码；将TC.Tno、TC.Cno、SC.Sno、SC.Cno设为外码，分别与Teacher、Student、Course建立参照关系。

第三节 电子商务系统数据层实现

1、数据模型

例1：数据结构定义与SQL查询

■ 用SQL查询语句完成下列操作：

- 查询所有不姓李的教师的姓名、年龄、性别和系名；

```
SELECT Tname, Tage, Tsex, Tdept  
FROM teacher  
WHERE Tname like '[!李]*'
```
- 查询选修“1”号和“4”号课程且成绩介于80分和90分之之间的学生的学号；

```
SELECT sno  
FROM sc  
WHERE cno='001' or cno='004' and (grade between 80 and 90)
```
- 查询会计系男生的平均成绩；

```
SELECT avg(grade) AS 平均成绩  
FROM student INNER JOIN sc ON student.sno=sc.sno  
WHERE student.sdept='会计系' and student.ssex='男'
```

第三节 电子商务系统数据层实现

1、数据模型

例1：数据结构定义与SQL查询

■ 用SQL查询语句完成下列操作：

- 将一个教师记录（教师号：0899；姓名：岳枫；年龄：28；性别：女；系名：经贸系；E-mail：yf@citiz.net）插入到Teacher表中；

```
INSERT INTO teacher ( Tno, Tname, Tage, Tsex, Tdept, [E-mail] )  
VALUES ('0899', '岳枫' , 28, '女' , '经贸系' , 'yf@citiz.net ')
```
- 将一个系学生的平均成绩存入数据库的某个表中；

```
SELECT student.sdept,avg(sc.grade) as 平均成绩,sc.cno into 平均成绩  
FROM student,sc  
WHERE student.sno=sc.sno group by student.sdept,sc.cno
```
- 将岳枫教师的E-mail改为yf@yahoo.com；

```
UPDATE teacher set [E-mail]='yf@yahoo.com' WHERE tname='岳枫'
```
- 删除某个学生（譬如张三）的选课信息

```
DELETE  
FROM student, sc  
WHERE student.sno=sc.sno and sname='张三'
```

SC

第三节 电子商务系统数据层实现

2、数据完整性声明

- 声明数据完整性可以使用**约束、规则和默认实现**。
- 约束类型有：
 - DEFAULT约束
 - CHECK约束
 - PRIMARY KEY约束
 - UNIQUE约束
 - FOREIGN KEY约束

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 约束

- **PRIMARY KEY**: 是通过唯一索引对给定的一列或多列强制实体完整性的约束。对于每个表只能创建一个PRIMARY KEY约束。
- **CHECK**: 是通过限制可输入到一系列或多列中的可能值强制域完整性的约束。
- **FOREIGN KEY [(column[,...])] REFERENCES ref_table [(ref_column[,...])]**: 是为列中的数据提供引用完整性的约束。
 - FOREIGN KEY约束要求列中的每个值在被引用表中对应的被引用列中都存在。
 - FOREIGN KEY约束只能引用被引用表中为PRIMARY KEY或UNIQUE约束的列或被引用表中在UNIQUE INDEX内引用的列。
ref_table: 是FOREIGN KEY约束所引用的表名。
- **UNIQUE**: 是通过唯一索引为给定的一系列或多列提供实体完整性的约束。一个表可以有多个UNIQUE约束。

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 约束

- 完整性约束三种形式：
 - PRIMARY KEY子句：实体完整性。
 - FOREIGN KEY子句：参照完整性。
 - CHECK子句：用户定义完整性。
- 完整性约束条件涉及到该表的多个属性，则必须定义在表级上，否则既可以定义在列级也可以定义在表级。
- 可以在创建表时建立约束。

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 约束

例2：创建学生表S， Sno、 Sname、 Ssex、 Sage、 Sdept五个属性组成。 Sno为主码， Sname取值不能为空且要唯一， Ssex只能取男或女。

```
CREATE TABLE S (Sno CHAR(6) PRIMARY KEY,  
                 Sname CHAR(8) NOT NULL UNIQUE,  
                 Ssex CHAR(2) CHECK (Ssex IN ('男', '女')),  
                 Sage SMALLINT,  
                 Sdept CHAR(14))
```

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 约束

例3：创建选修表SC，由Sno 、 Cno 、 Grade三个属性组成。其中(Sno, Cno)为主码，Grade取值为空或在0到100之间。

表S，C要先定义

```
CREATE TABLE SC
```

```
(Sno CHAR(6) FOREIGN KEY REFERENCES S(Sno),
```

```
  Cno CHAR(6) FOREIGN KEY REFERENCES C(Cno),
```

```
  Grade SMALLINT CHECK (Grade IS NULL OR Grade  
  BETWEEN 0 AND 100 ),
```

```
  PRIMARY KEY (Sno, Cno)
```

```
)
```

主码两个属性（表级）

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 规则

- **规则**是数据库的对象，是用来对表中的字段值（或用户自定义的数据类型中的值）进行限制的一种手段，用于执行一些与 CHECK 约束相同的功能。
- 不同点：
 - CHECK约束在表创建时创建，与表共生存。
 - 规则是单独的对象进行创建而存在，通过绑定到某个表的字段上而生效，不依赖表而存在。

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 规则

□ 创建规则命令

Create rule 规则名 as 定义规则的条件

- 规则条件可以是WHERE子句中任何有效的表达式，可包含算术、关系、逻辑运算符和谓词（IN、LIKE、BETWEEN）。
- 规则不能引用列或其他数据库对象。
- 每个局部变量的前面都有一个@。
- 在创建规则时可以使用任何名称或符号表示值，但第一个字符必须是@字符。

- 规则创建后，必须通过执行系统存储过程sp_bindrule将其绑定于列或用户自定义的数据类型上：

Exec sp_bindrule 规则名 表名.字段名

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 规则

例4 在SQL Server中进行规则的创建与绑定。

(1)创建一个规则，用以限制Teacher表中Tage取值范围。

- ☐ **USE mytest**
- ☐ **GO**
- ☐ **CREATE RULE rule1**
- ☐ **AS**
- ☐ **@range >= 1 AND @range < 120**
- ☐ **GO**
- ☐ **EXEC sp_bindrule 'rule1', 'Teacher.Tage'**

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 规则

例4 在SQL Server中进行规则的创建与绑定。

(2)创建一个规则，用以限制Teacher表中Tdept取值范围。

- ☐ USE mytest
- ☐ GO
- ☐ CREATE RULE rule2
- ☐ AS
- ☐ @list IN ('经贸系', '会计系', '金融系', '计算机系')
- ☐ GO
- ☐ EXEC sp_bindrule 'rule2', 'Teacher.Tdept'

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 默认值

- **默认值**是一种数据库对象，在数据库中只需定义一次后，就可以被一次或多次应用于任意表中的一列或多列，还可以用于用户定义的数据类型。
- 默认值对象的作用类似于建表时的DEFAULT约束，对未赋值的字段赋予初值，其管理和使用与规则比较相似，**但优先级低于规则**。

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 默认值

- 在默认被创建后，用户必须通过执行系统存储过程 `sp_binddefault` 将其定于列或用户自定义的数据类型上，从而将默认用于数据库中任意表的一列或多列，以及用于用户自定义的数据类型。

- 创建默认命令：

Create default 默认名 as 默认的常数值

- 执行默认绑定命令：

Exec sp_binddefault 默认名 '表名.字段名'

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 默认值

例5 定义一个默认值，使教师表中的Ttel初值为unkown。

- ☐ USE mytest
- ☐ GO
- ☐ **CREATE DEFAULT** def1 AS 'unknown'
- ☐ GO
- ☐ **EXEC sp_bindefault** def1, 'teacher.Ttel'
- ☐ 步骤如同规则创建。

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 存储过程

- **存储过程**是存放在服务器上的预先编译好的SQL语句。
- 是SQL语句和可选控制流语句的预编译集合，以一个名称存储并作为一个单元处理。
- 在第一次执行时进行语法检查和编译，编译好的版本在调整缓存中用于后续调用，使存储过程执行时带回迅速，更加高效。
- 存储过程存储在数据库内，可由应用程序通过一个调用执行，而不是由SQL Server自动执行。
- 允许用户声明变量、有条件执行以及其它强大的编程功能。存储过程可包含程序流、逻辑以及对数据库的查询。它们可以接受参数、输出参数、返回单个或多个结果集以及返回值。
- SQL Server支持的存储过程分两类：**系统提供的存储过程、用户自定义的存储过程。**

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 存储过程

□ 优点

- 单个存储过程中可执行一系列SQL语句。
- 从一个存储过程中可调用其他存储过程，可简化一系列复杂语句。
- 提供了安全机制、改进了执行性能、减少了网络流量、增加了代码的重用性。
- 模块化程序设计，便于移植、便于共享，提高开发效率。

□ 存储过程操作主要包括：

- 创建存储过程、设置用户执行存储过程权限、指定存储过程中的参数、查看存储过程、删除存储过程、执行存储过程、重命名存储过程。

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 存储过程

例6 创建一个能查询出所有不姓李的教师的姓名、年龄、性别和系名的存储过程。

- ☐ USE mytest
- ☐ IF EXISTS (SELECT name FROM sysobjects
- ☐ WHERE name = 'pro1' AND type = 'P')
- ☐ DROP PROCEDURE pro1
- ☐ GO
- ☐ CREATE PROCEDURE pro1
- ☐ AS
- ☐ SELECT teacher.Tname, teacher.Tage, teacher.Tsex, teacher.Tdept
- ☐ FROM teacher
- ☐ WHERE (((teacher.Tname) Like '[^李]%'));
- ☐ GO
- ☐ EXEC pro1

P是类型，表示存储过程

//创建过程

第三节 电子商务系统数据层实现

2、数据完整性声明

■ 存储过程

例7 创建一个能查询出选修了“001”号和“004”号课程且成绩介于80分和90分之间的学生的姓名的存储过程。

- ☐ USE mytest
- ☐ IF EXISTS (SELECT name FROM sysobjects
- ☐ WHERE name = 'pro2' AND type = 'P')
- ☐ DROP PROCEDURE pro2
- ☐ GO
- ☐ CREATE PROCEDURE pro2 //创建过程
- ☐ @score1 int, //变量名、参数名前须以符号 @为前缀
- ☐ @score2 int
- ☐ AS
- ☐ SELECT sc.Sno FROM sc
- ☐ WHERE (((sc.Cno)='1' Or (sc.Cno)='4') AND ((sc.Grade) Between @score1
- ☐ And @score2))
- ☐ GO
- ☐ EXEC pro2 80, 90 //执行过程

第三节 电子商务系统数据层实现

2、数据完整性声明

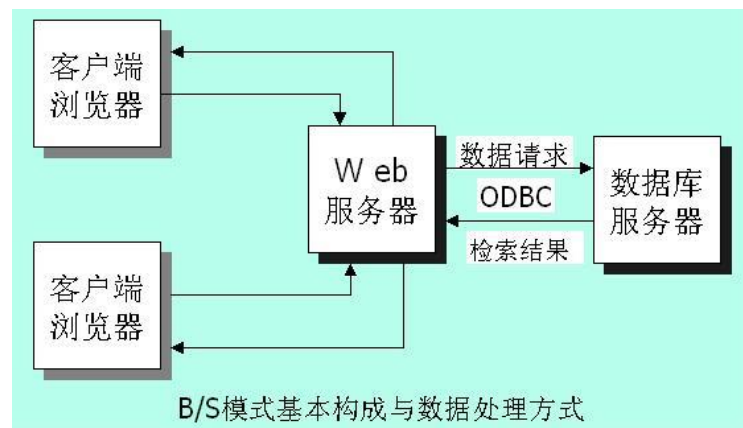
■ 触发器

- 触发器是一种特殊类型的存储过程，是一种在基表被修改时自动执行的内嵌过程，当使用UPDATE、INSERT或DELETE命令在指定表中对数据进行修改时，用来防止对数据进行的不正确或不一致的修改。
- 优点
 - 强制比CHECK约束更复杂的数据完整性。
 - 使用自定义的用户错误信息：通过使用触发器，用户可以捕获破坏数据完整性的操作，并返回自定义的错误信息。
 - 实现数据库中多张表的级联修改：通过触发器对数据库中的相关表进行级联修改。如在student表的sno列上写入一个删除触发器，以使其他表中的各匹配行采取删除操作。
 - 比较数据库修改前后数据的状态。
 - 维护非规范化数据。

第三节 电子商务系统数据层实现

3、数据库访问方式

- 电子商务系统中数据库访问方式是：浏览器/服务器的数据访问。用户不直接使用数据管理系统，而是通过表达层或逻辑层上的某种应用程序，利用访问接口技术间接地完成数据的操作。
- 目前主要有三种数据访问方式：
 - 简单式：基于浏览器的B/S模式，利用HTML页面在用户的计算机上表示信息。只限于使用HTML协议，也可用Java Applet, ActiveX, JavaBean来加强表达能力，给网页带来动态性。
 - 交互式：利用Web服务器或脚本对表单进行处理。



第三节 电子商务系统数据层实现

3、数据库访问方式

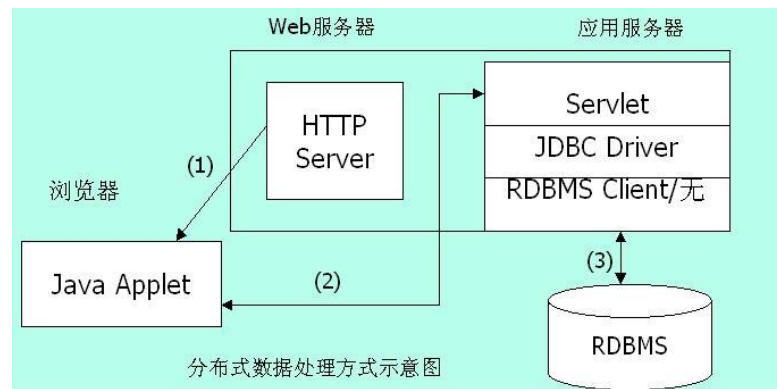
- 目前主要有三种数据访问方式：

- **分布式**：利用浏览器/Web服务器的应用程序借助组件技术与应用服务器建立连接，并通过其与数据库服务器相连接，并最终返回结果到浏览器，如Java的小程序(Applet)。

- HTTP服务器将含有小程序页面下载到浏览器。

- 小程序在浏览器中运行，并通过组件支持的通信协议(IIOP,DCOM)与应用服务器上的小程序(Servlet)通信会话。

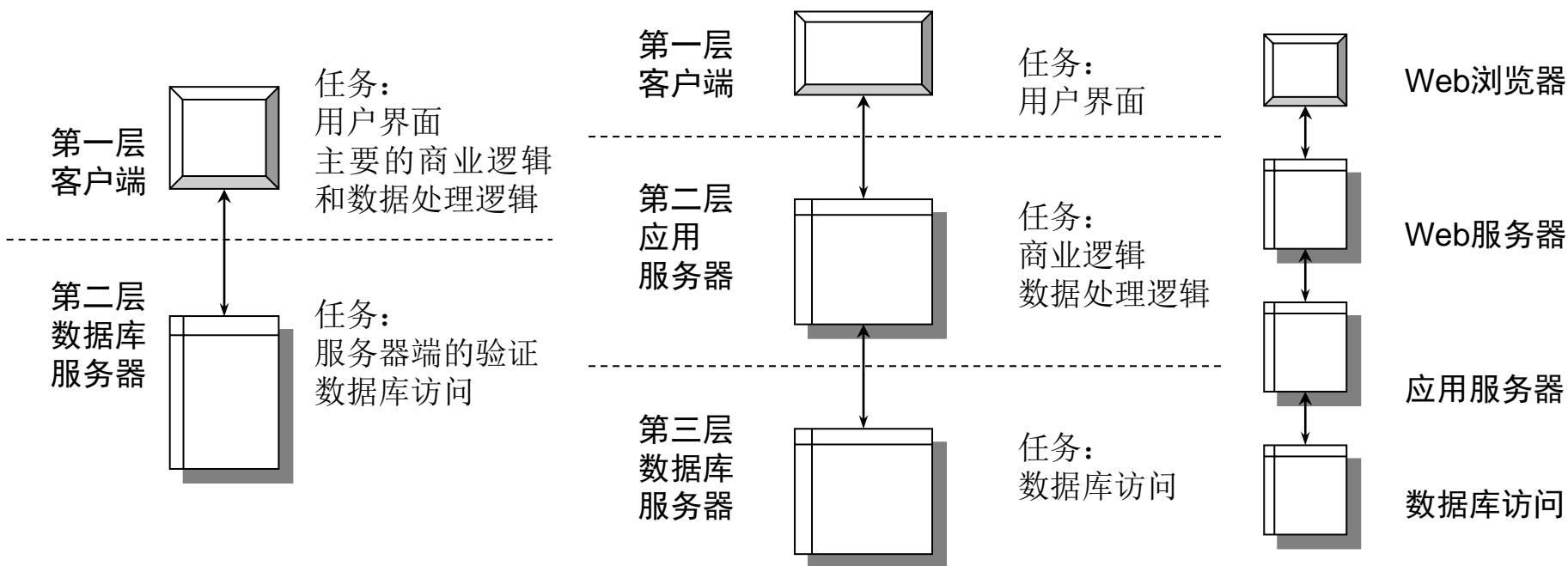
- 小程序收到信息后，经过JDBC、ODBC等向数据库服务器发出请求，对方接到命令处理后，将结果传送给Servlet，最后结果显示在浏览器中。



第三节 电子商务系统数据层实现

3、数据库访问方式

■ 应用体系结构与数据处理任务分担



第四章 商务数据层及其技术

- 第一节 数据管理技术的发展
- 第二节 电子商务系统中的数据管理技术
- 第三节 电子商务系统数据层实现
- 第四节 电子商务个性化推荐技术



第四节 电子商务个性化推荐技术

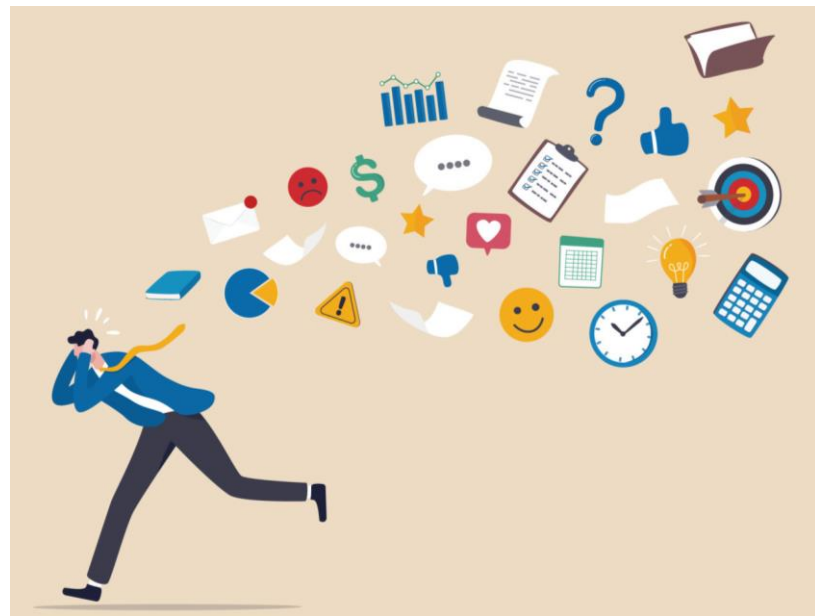
1、个性化推荐概述

■ 随着网络技术的发展，人们可以随时随地接触到海量的信息资源，网络信息量不断膨胀，出现了

“**信息过载**”的现象。

■ 信息过载问题表现形式：

- ✓ 用户对信息反应的速度远低于信息传播的速度。
- ✓ 大众媒介中的信息量大大高于用户需要的信息量。
- ✓ 大量无关的、冗余的数据信息干扰了用户对相关有用信息的准确性的选择。



第四节 电子商务个性化推荐技术

1、个性化推荐概述

■ 如何解决信息过载问题：

被动形式：信息检索

主动形式：信息推荐

- ✓ **信息检索**需要一个明确的目的/意图去搜索相关信息，受用户知识面宽度影响，信息检索结果存在局限性，有可能越流行的东西越被检索。
- ✓ **推荐系统**基于用户历史行为数据预测用户偏好进行主动推荐(无/模糊意图)，能实现信息推荐的准确性、新颖性、多样性等特点。
- ✓ 电子商务网站中的货架成本低廉，电商网站能够出售更多数量庞大的不热门商品，其规模甚至会超过热门商品。主流商品代表绝大多数用户需求，长尾商品代表一小部分用户个性化需求，而用户需求是个性化的（有可能是偏离主流的），如果要提升销售额，需要进一步挖掘用户个性化需求。
- ✓ **推荐系统可以满足用户个性化需求。**

第四节 电子商务个性化推荐技术

1、个性化推荐概述

推荐系统可以看成是一个函数

■ 输入：

- 用户反馈行为：**显示反馈**（评分等）、**隐式反馈**（点击、加购、收藏、点赞等）
- 商品信息，用户信息，好友偏好等

■ 计算/预测相关商品评分，用于商品排序

■ 输出：

- 为用户推荐相关商品（用户感兴趣、会购买、满意）

第四节 电子商务个性化推荐技术

1、个性化推荐概述

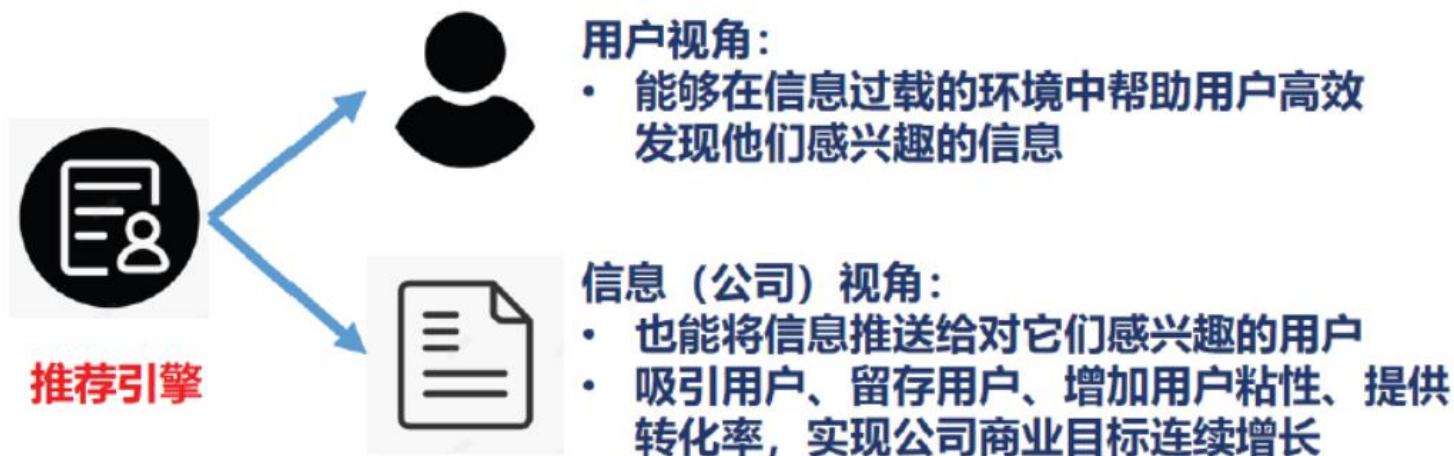
推荐系统目标

■ 定义：给用户推荐感兴趣的商品

□ 评分预测：预测用户对某给定商品的评分偏好

□ 排名预测：向特定用户推荐前N个(TopN)喜欢/点击商品

■ 原理(核心)：基于多源数据推断用户偏好/兴趣/行为



第四节 电子商务个性化推荐技术

1、个性化推荐概述

推荐系统目标

■推荐系统最终目标是提升商业利润

通过洞察用户需求，提高消费者满意度，进而提升用户忠诚度和重购行为。

相关性



最主要的要求，
推荐的商品必须和
用户兴趣相关。

新颖性



如果推荐的商品是用
户以前没见过的说明
推荐系统是有用的。
反复推荐流行商品可
能会降低多样性。

意外性



推荐的商品出乎意料
，与新颖性的区别在
于用户感到惊喜，通
常推荐与用户兴趣不
相关的商品。

多样性



推荐top-k列表都相似
时，很可能用户都不
喜欢，确保用户不会
对相似的商品反复推
荐感到厌烦。

第四节 电子商务个性化推荐技术

2、个性化推荐算法分类

(1) 基于协同过滤的推荐算法

■ 基于邻域的协同过滤

界定相似邻居，使用已有评分直接预测用户对其他商品的评分。

- 基于用户的协同过滤 (User-based CF)
- 基于商品的协同过滤 (Item-based CF)

■ 基于模型的协同过滤

利用历史评分构建预测模型，预测用户对其他商品的评分。一般采用机器学习算法，找出用户与商品的相互作用模型。

- 传统机器学习模型（神经网络、朴素贝叶斯、聚类等）
- 隐语义模型

(2) 基于内容的推荐算法

第四节 电子商务个性化推荐技术

2、个性化推荐算法分类

(1) 基于协同过滤的推荐算法

■ 基于邻域的协同过滤

Item-based CF

- ✓ 给用户推荐那些之前喜欢的商品相似的商品。
- ✓ 通过分析用户的行为记录，计算商品之间的相似度(而非商品的属性)。
- ✓ 能够提供可解释性(根据以前购买过的商品)。

User-based CF

- ✓ 当用户规模越来越大时，计算用户相似性矩阵越来越困难，运算时间复杂度和空间复杂度与用户数成近乎平方关系。
- ✓ 冷启动用户推荐准确度低。
- ✓ 考虑用户隐私原因，User-based CF不能给出充分的推荐解释。

第四节 电子商务个性化推荐技术

3、基于用户的协同过滤算法

- 基于用户和其他用户的兴趣/评分相似度预测用户对未评分商品的评分，主要包括两个步骤：
 - 通过计算相似度，找到和目标用户兴趣相似的用户集合。
 - 找到这个集合中的用户喜欢的，且目标用户没有评价过(交互过)的商品推荐给目标用户。
- 适用于商品更新频繁的情景。根据用户购买行为，找出兴趣相似的用户，将相似用户喜欢的商品推荐给目标用户。
 - 老张喜欢看的书有A, B, C, D；老王喜欢看的书有A, B, C, E。通过这些数据可以判断老张和老王的口味相似，于是给老张推荐E这本书，同时给老王推荐D这本书。

第四节 电子商务个性化推荐技术

3、基于用户的协同过滤算法

- 例如，想要确定Alice是否喜欢她还没有评过分的item5，可以采用基于用户的协同过滤算法。
- 首先，计算Alice和其他用户的兴趣相似度，常用的是皮

尔森相似度：

$$\text{sim}(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

P 为用户 a 和 b 共同评分过的商品集合。
计算平均评分时，只考虑两个用户共同评过商品的评分。

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1



sim = 0.85

sim = 0.71

sim = 0.00

sim = -0.79

第四节 电子商务个性化推荐技术

3、基于用户的协同过滤算法

- 例如，想要确定Alice是否喜欢她还没有评过分的item5，可以采用基于用户的协同过滤算法。

- 首先，计算Alice和其他用户的兴趣相似度，常用的是皮尔森相似度：

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

P 为用户 a 和 b 共同评分过的商品集合。
计算平均评分时，只考虑两个用户共同评过商品的评分。

- 然后，预测Alice对未评过分的item5的评分：

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

目标用户 a 的平均评分，计算时考虑该用户的所有评分(不包括item5)

N 为目标用户 a 的相似用户集合

相似用户 b 的平均评分，计算时考虑该用户的所有评分(包括item5)

第四节 电子商务个性化推荐技术

4、基于商品的协同过滤算法

- 基于用户评分过的商品和其他商品的评分相似度预测用户对未评分商品的评分，主要包括两个步骤：
 - 找到和目标商品相似的商品集合。
 - 找到这个集合中的用户评过分的商品，来预测用户未评分的商品。
- 适用于item的增长速度远小于user的增长速度的情况。
根据用户的喜好先找出相似的商品，如果同时喜欢两个商品的人比较多的话，就认为这两个商品相似。
 - 发现喜欢看书籍A的人大多喜欢看书籍B，则会给看了A的人推荐B，这就是“买了...的人也买了...”。

第四节 电子商务个性化推荐技术

5、基于模型的推荐算法

- 基于模型的方法是在使用已有评分，训练预测模型，预测新评分。一般是使用机器学习算法建立用户和商品的相互作用模型，从而找出数据中的模式。
- 有许多不同的算法可用于构建模型并基于这些模型进行预测，例如，神经网络、贝叶斯网络、聚类、分类、回归、矩阵分解等。
- 基于矩阵分解的方法是一类较经典的基于模型的方法。

第四节 电子商务个性化推荐技术

5、基于模型的推荐算法

基于矩阵分解的方法

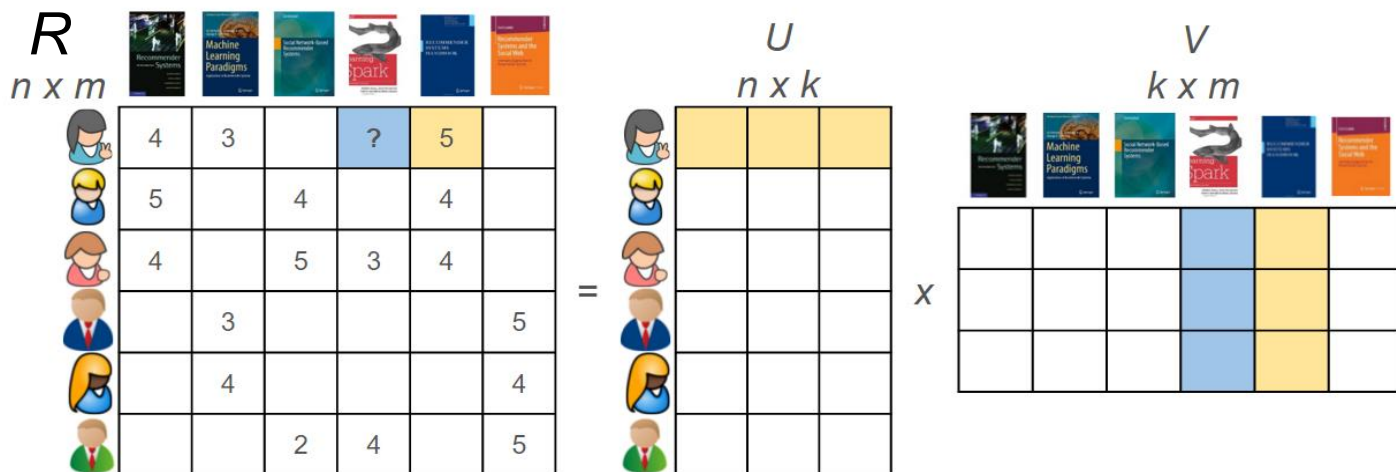
- **“兴趣”** 分类：基于低秩矩阵分解的方法从评分矩阵中抽取一组潜在的（隐藏的）因子，并通过这些因子向量描述用户和商品。
- **矩阵因子分解**（如奇异值分解）将商品和用户都转化成相同的潜在空间，即将用户评分矩阵分解成一个用户-隐因子矩阵乘以一个隐因子-商品矩阵，代表用户和商品之间的潜相互作用。
- 在电影领域，这些自动识别的因子可能对应一部电影的常见标签，比如风格或者类型（戏剧片或者动作片），也可能是无法解释的。

第四节 电子商务个性化推荐技术

5、基于模型的推荐算法

基于矩阵分解的方法

- 将用户评分矩阵分解为两个低秩的矩阵，这两个矩阵的乘积能对原始矩阵进行某种程度的复原，从而可以评估出缺失值。



第四节 电子商务个性化推荐技术

5、基于模型的推荐算法

基于矩阵分解的方法

■ 音乐评分示例

- ✓ 用户-隐因子矩阵 U : 表示不同的用户对于不同元素的偏好程度, 1代表很喜欢, 0代表不喜欢。
- ✓ 隐因子-音乐矩阵 V : 表示每种音乐含有各种元素的成分。如下表中, 音乐A是一个偏小清新的音乐, 含有小清新这个隐因子的成分是0.9, 电音的成分是0.1, 优雅的成分是0.2。

U	小清新	电音	优雅	伤感	民谣
张三	0.6	0.8	0.1	0.1	0.7
李四	0.1	0	0.9	0.1	0.2
王五	0.5	0.7	0.9	0.9	0

V	小清新	电音	优雅	伤感	民谣
音乐A	0.9	0.1	0.2	0.4	0
音乐B	0.5	0.6	0.1	0.9	1
音乐C	0.1	0.2	0.5	0.1	0
音乐D	0	0.6	0.1	0.2	0

第四节 电子商务个性化推荐技术

5、基于模型的推荐算法

基于矩阵分解的方法

■ 音乐评分示例

	小清新	电音	优雅	伤感	民谣
张三	0.6	0.8	0.1	0.1	0.7

	小清新	电音	优雅	伤感	民谣
音乐A	0.9	0.1	0.2	0.4	0

- ✓ 利用这两个矩阵，能得出张三对音乐A的喜欢程度是：

张三对小清新的偏好 \times 音乐A含有小清新的成分 + 对电音的偏好 \times 音乐A含有电音的成分 + 对优雅的偏好 \times 音乐A含有优雅的成分 + ……

- ✓ 即：张三对音乐A的喜欢程度

$$= 0.6 \times 0.9 + 0.8 \times 0.1 + 0.1 \times 0.2 + 0.1 \times 0.4 + 0.7 \times 0 = 0.68$$

- ✓ 每个用户对每首歌都由此计算，得到不同用户对不同歌曲的评分矩阵：

	音乐A	音乐B	音乐C	音乐D
张三	0.68	1.58	0.28	0.51
李四	0.31	0.43	0.47	0.11
王五	1.06	1.57	0.73	0.69

- ✓ 为张三推荐四首歌中得分最高的B，为李四推荐得分最高的C，为王五推荐B。

第四节 电子商务个性化推荐技术

6、基于内容的推荐算法

- 基于内容的推荐算法是为用户推荐那些与用户过去喜欢的商品类似的商品。通常使用用户模型的向量特征来描述用户的兴趣爱好，同样对于每个商品进行特征提取，作为商品模型的内容特征。然后计算用户模型的向量特征和候选商品模型的向量特征两者之间的匹配度，选择匹配度较高的商品推荐。
- 不同于协同过滤，**基于内容的推荐算法是基于商品自身信息，比较商品之间的相似度，没有考虑用户评分情况。**
 - 例如，如果一个用户喜欢电影“指环王：魔戒再现”和“指环王：双塔奇兵”，然后使用电影的标题信息，推荐系统可以向用户推荐电影“指环王：王者无敌”。
- 在基于内容的推荐中，将获取到商品的描述信息作为商品的特征向量。这些特征向量被用于创建一个反映用户偏好的模型。

第四节 电子商务个性化推荐技术

6、基于内容的推荐算法

■ 优势：

- 用户之间的独立性，不受其他人评分的影响。
- 好的可解释性，因为推荐的商品有某某属性，跟用户的品味很匹配等。
- 新商品可以立刻得到推荐，被推荐的机会和老的商品是一致的。

■ 不足：

- 商品的属性有限，很难有效的得到更多数据。
- 商品相似度的衡量标准只考虑到商品本身，有一定的片面性，新颖性较差。
- 需要用户的商品的历史数据，有新用户冷启动的问题。

第四节 电子商务个性化推荐技术

7、个性化推荐评价指标

(1) Top-N 推荐

■ **准确率 Precision:**
$$Precision = \frac{N_{tp}}{N_{tp} + N_{fp}} = \frac{|good\ movies\ recommended|}{|all\ recommendations|}$$

■ **召回率 Recall:**
$$Recall = \frac{N_{tp}}{N_{tp} + N_{fn}} = \frac{|good\ movies\ recommended|}{|all\ good\ movies|}$$

■ **F1:**
$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

待预测的商品可能的4种情况

用户喜好	系统推荐	系统不推荐
喜欢	Ture-Positive N_{tp}	False-Negative N_{fn}
不喜欢	False-Positive N_{fp}	Ture-Negative N_{tn}

第四节 电子商务个性化推荐技术

7、个性化推荐评价指标

(2) 评分预测

- 预测评分的准确度衡量的是算法预测的评分和用户的实际评分的贴近程度。

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - r_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2}$$

本章小结



- 数据管理的4个发展阶段
- 数据仓库的概念及其与数据库的区别
- 数据库的数据模型
- 数据完整性
- 常见个性化推荐算法类型
- 用户兴趣相似度计算



Thank You !