

The background is a collage of four images: top-left shows a stack of books in shades of purple and blue; top-right shows a clock face in shades of pink and purple; bottom-left shows a stack of books in shades of green and blue; bottom-right shows a clock face in shades of yellow and orange.

第十一章 商务智能

北京大学计算机系电子商务实验室



11.1 第一代电子商务

11.2 第二代电子商务

11.3 代理 (Agent) 技术

11.4 数据仓库、联机分析处理
与数据挖掘

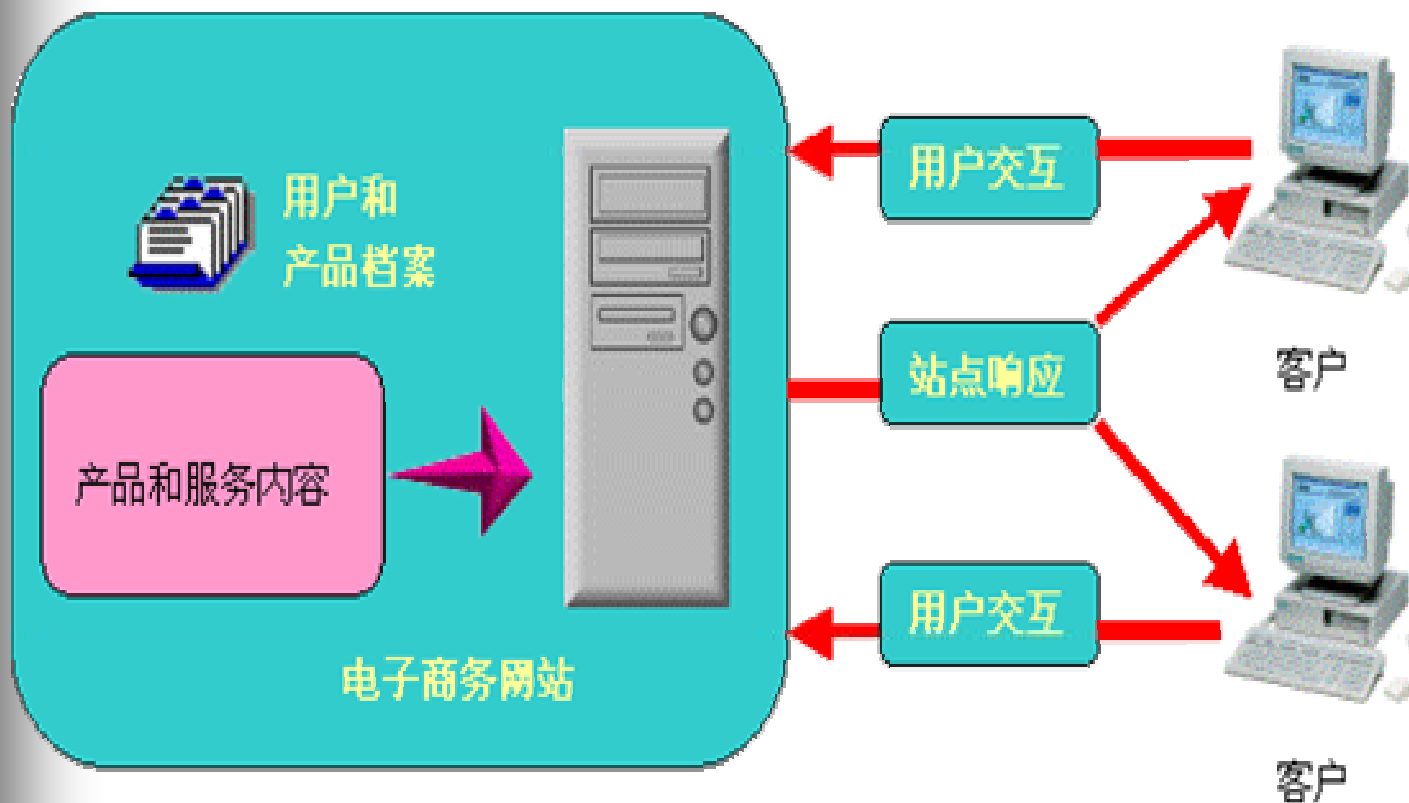
11.5 基于代理技术的电子商务

11.6 第三代电子商务

11.7 潜在问题和挑战

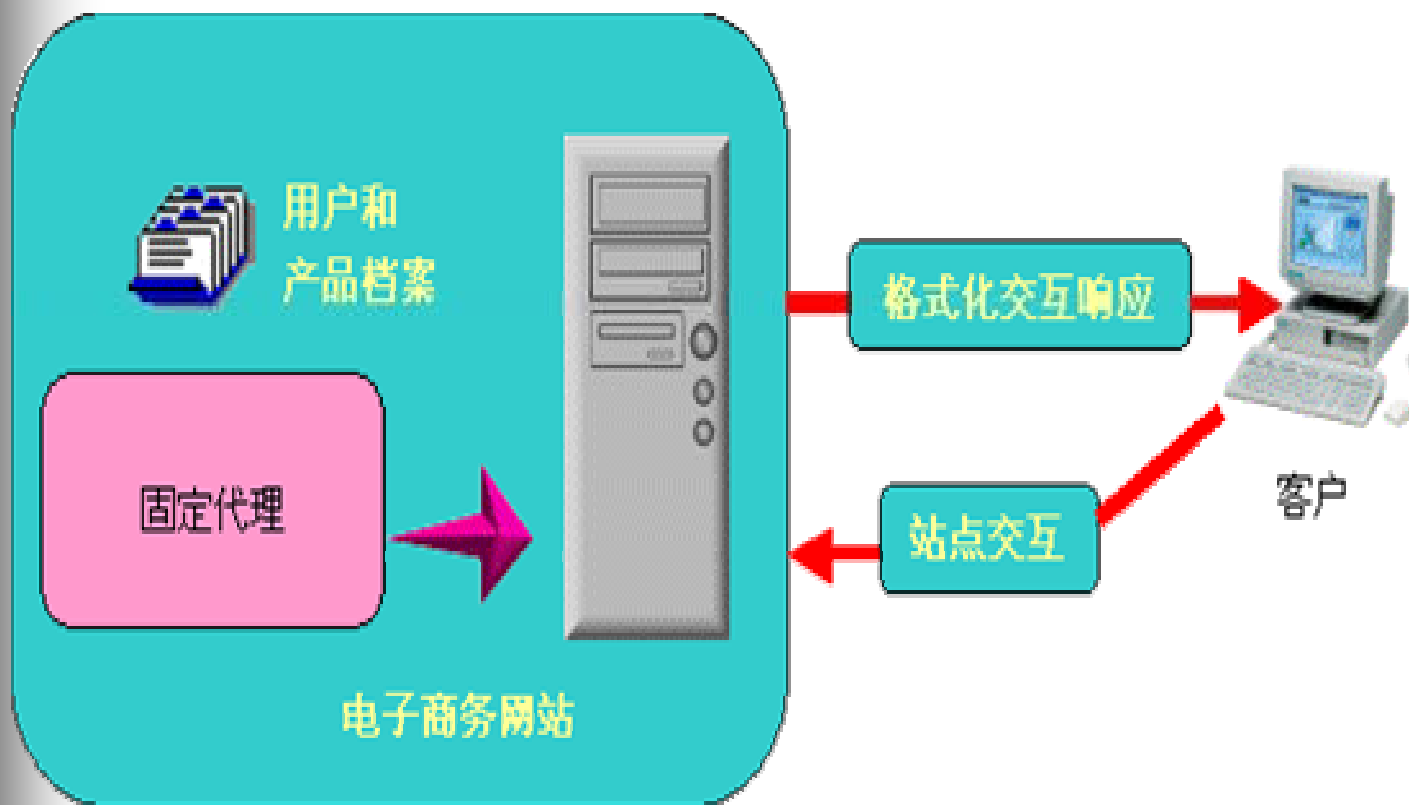
11.1 第一代电子商务

- 商务费用的明显减少
- 对多媒体的大量使用



11.2 第二代电子商务

固定代理被用作助手实现基于网络的集中用户市场策略





11.3代理 (Agent) 技术

■ 11.3.1代理的定义

在研究领域，关于代理的准确的定义几乎没有一致意见。

然而，有一种定义说代理是微小的软件实体，它通过代表用户(机器和人)利益的自主行为来操作。



The MuBot Agent

<http://www.crystaliz.com/logicware/mubot.html>

- *"The term agent is used to represent two orthogonal concepts. The first is the agent's ability for autonomous execution. The second is the agent's ability to perform domain oriented reasoning."*



The AI/MA(Artificial Intelligence: a Modern Approach) Agent

[Russell and Norvig 1995, page
33]

- *"An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors."*



The Maes Agent [Maes 1995, page 108]

- *"Autonomous agents are computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realize a set of goals or tasks for which they are designed."*



The Hayes-Roth Agent [Hayes-Roth 1995]

- *Intelligent agents continuously perform three functions: perception of dynamic conditions in the environment; action to affect conditions in the environment; and reasoning to interpret perceptions, solve problems, draw inferences, and determine actions.*



The IBM Agent

[\[http://activist.gpl.ibm.com:81/WhitePaper/ptc2.htm\]](http://activist.gpl.ibm.com:81/WhitePaper/ptc2.htm)

- *"Intelligent agents are software entities that carry out some set of operations on behalf of a user or another program with some degree of independence or autonomy, and in so doing, employ some knowledge or representation of the user's goals or desires."*



The Wooldridge;Jennings Agent [Wooldridge and Jennings 1995, page 2]

- *Agent is "... a hardware or (more usually) software-based computer system that enjoys the following properties:*
 - *autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;*



- *social ability: agents interact with other agents (and possibly humans) via some kind of agent-communication language;*
- *reactivity: agents perceive their environment, (which may be the physical world, a user via a graphical user interface, a collection of other agents, the INTERNET, or perhaps all of these combined), and respond in a timely fashion to changes that occur in it;*
- *pro-activeness: agents do not simply act in response to their environment, they are able to exhibit goal-directed behaviour by taking the initiative."*



- The definitions of the previous section seem to derive from one or both of two common uses of the word agent:
 - 1) one who acts, or who can act, and
 - 2) one who acts in place of another with permission.
- As these definitions make clear, there's no general agreement as to what constitutes an agent, or as to how agents differ from programs.



The FAQ Agent

[\[http://www.ee.mcgill.ca/~belmarc/agent_faq.html\]](http://www.ee.mcgill.ca/~belmarc/agent_faq.html)

- *"This FAQ will not attempt to provide an authoritative definition ..."*
- It does provide a list of attributes often found in agents: **Autonomous, goal-oriented, collaborative, flexible, self-starting, temporal continuity, character, communicative, adaptive, mobile**, [Etzioni and Weld].



Definition of Agent

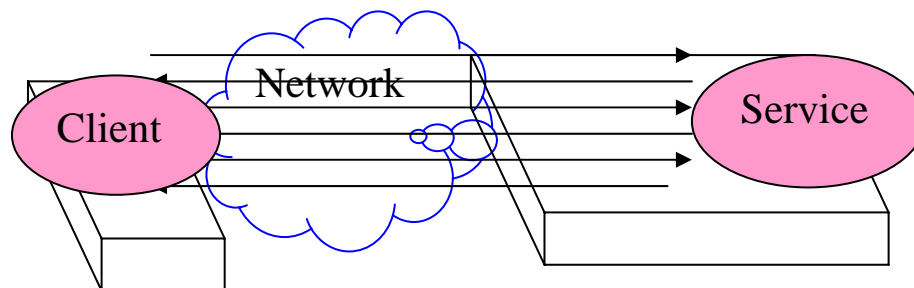
- An **autonomous agent** is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.

[Stan Franklin & Art Graesser, Institute for Intelligent Systems, University of Memphis, Proceeding of the Third International Workshop on Agent Theories, Architectures, and Languages, Springer-Verlag, 1996]

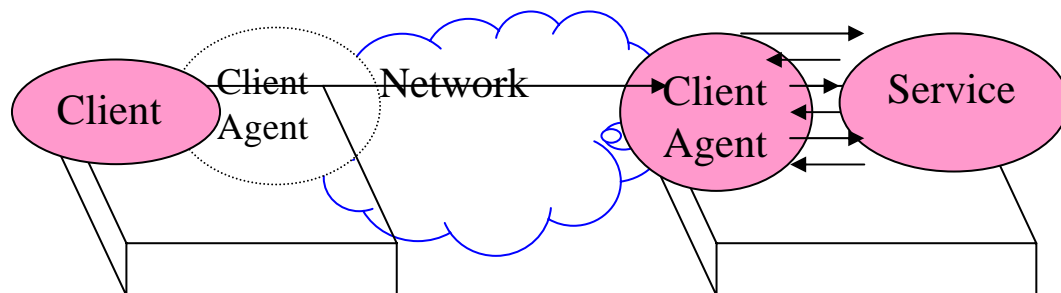
■ 11.3.2代理的分类

- 现在用于第二代系统中的代理是静态的
- 移动代理系统还是一种发展的技术

Static
Agent
System



Mobile
Agent
System



固定代理与移动代理之比较



■ 11.3.3代理的类型

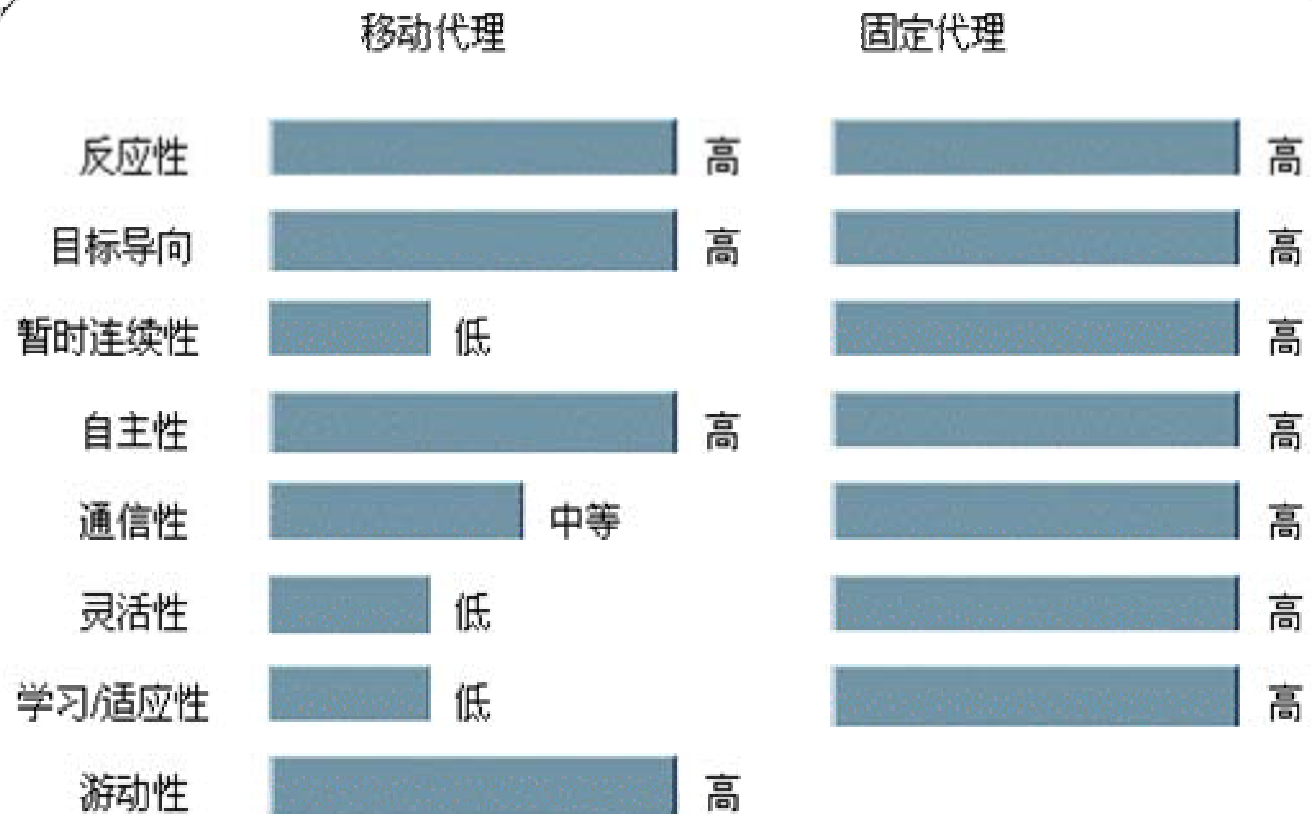
- 指导—为执行任务提供信息和指导方针
- 顾问—提出建议
- 提醒—用来通知异步事件的后台代理
- 助手—与向导相似，但是会把任务指派给它们(如：搜索，谈判，过滤等等)
- 分析—分析用户小档案，用户不能直接对它们进行处理



■ 11.3.4代理的特性

- 反应性
- 目标导向
- 暂时连续性
- 自主性
- 移动性 / 游动性
- 通信性(多代理或与用户)
- 灵活性
- 学习性 / 适应性

移动和固定代理特性的比较





11.4 数据仓库、联机分析处理与数据挖掘

- 包括了三个方面的内容：
 - 数据仓库技术(Dataware House , DW)
 - 联机分析处理技术(Online Analytical Processing , OLAP)
 - 数据挖掘技术(Data Mining , DM)
- 三者形成一种决策支持系统构架：

DW+OLAP+DM



■ 11.4.1 数据仓库

1. 数据仓库定义

- Devlin和Murphy在1988年的论文中第一次提出数据仓库的概念

- William. H. Inmon在1993的书“Building the Data Warehouse”:

数据仓库是一个面向主题的
(Subject Oriented)、集成的、随
时间变化的非易失性数据的集合, 用
于支持管理层的决策过程。

- 从大量的事务型数据库中抽取数据, 并将其清理、转换为新的存储格式, 为进一步的决策提供底层支持。



- 2. 数据仓库特点
- 面向主题
 - 主题：指决策时关心的重点问题。
 - 将信息按主题组织
- 集成的
 - 将数据经过系统地加工整理，形成相互一致的、具有代表性的数据
- 随时间变化
 - 存储的是一个时间段（5~10年）的数据，以便进行时间趋势分析
- 信息本身相对稳定
 - 信息一旦进入数据仓库则会长期保留而变化较少



■ 数据仓库的组织和管理方法

- 依据决策要求，只从数据库中抽取那些需要的数据并进行一定的处理
- 数据仓库是多维的，即组织方式有多层的行和列
- 支持决策处理，不同于普通的事务处理

注：

维是指客观实体的一种抽象表示，如客户，时间，产品，提供商，地点，渠道等



● 3.数据仓库与数据库

数据库系统	数据仓库系统
数据库：操作型数据，增、删、改操作频繁	数据仓库：分析型数据，极少有更新操作
数据库核心：功能强大，面向OLTP应用	数据仓库管理系统：因极少有更新操作，故功能简单。
数据库工具：以查询工具为主	数据仓库工具：以分析工具为主



数据仓库需要以下数据库技术的支持

■ 并行数据库技术

- 数据量达到GB或TB级

■ 高性能的数据库服务器

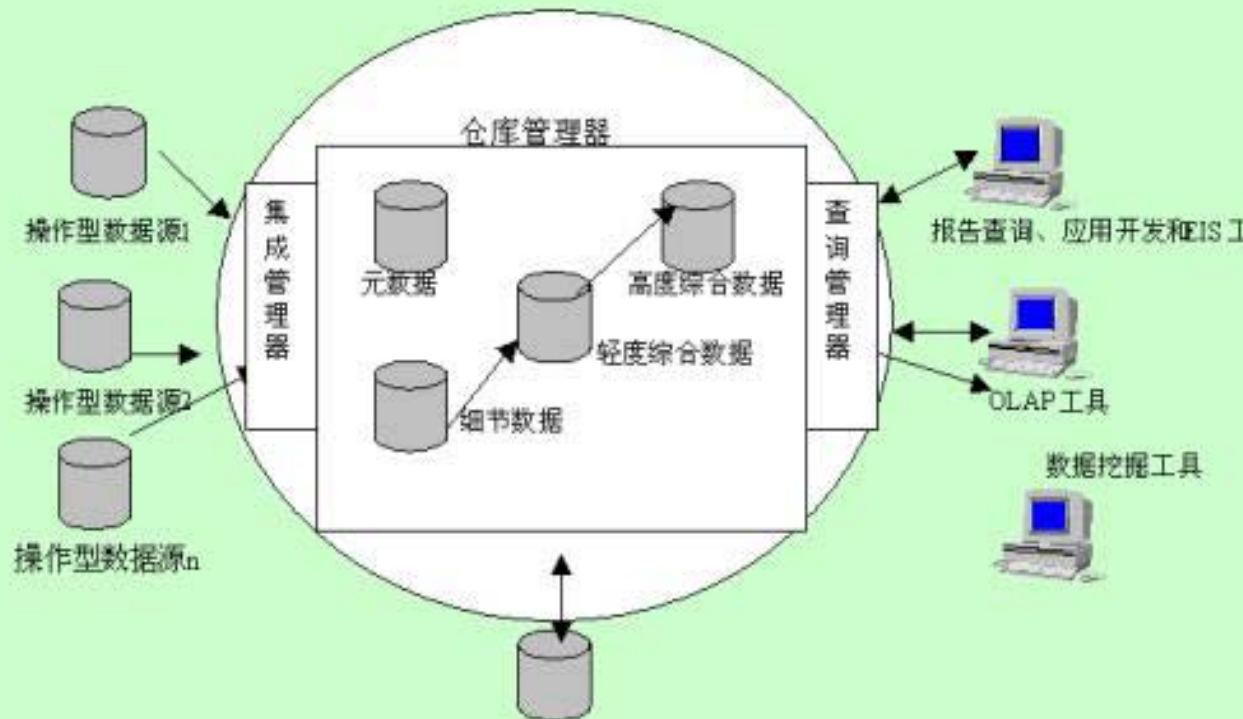
- 要求高性能的DBMS核心

■ 数据库互操作技术

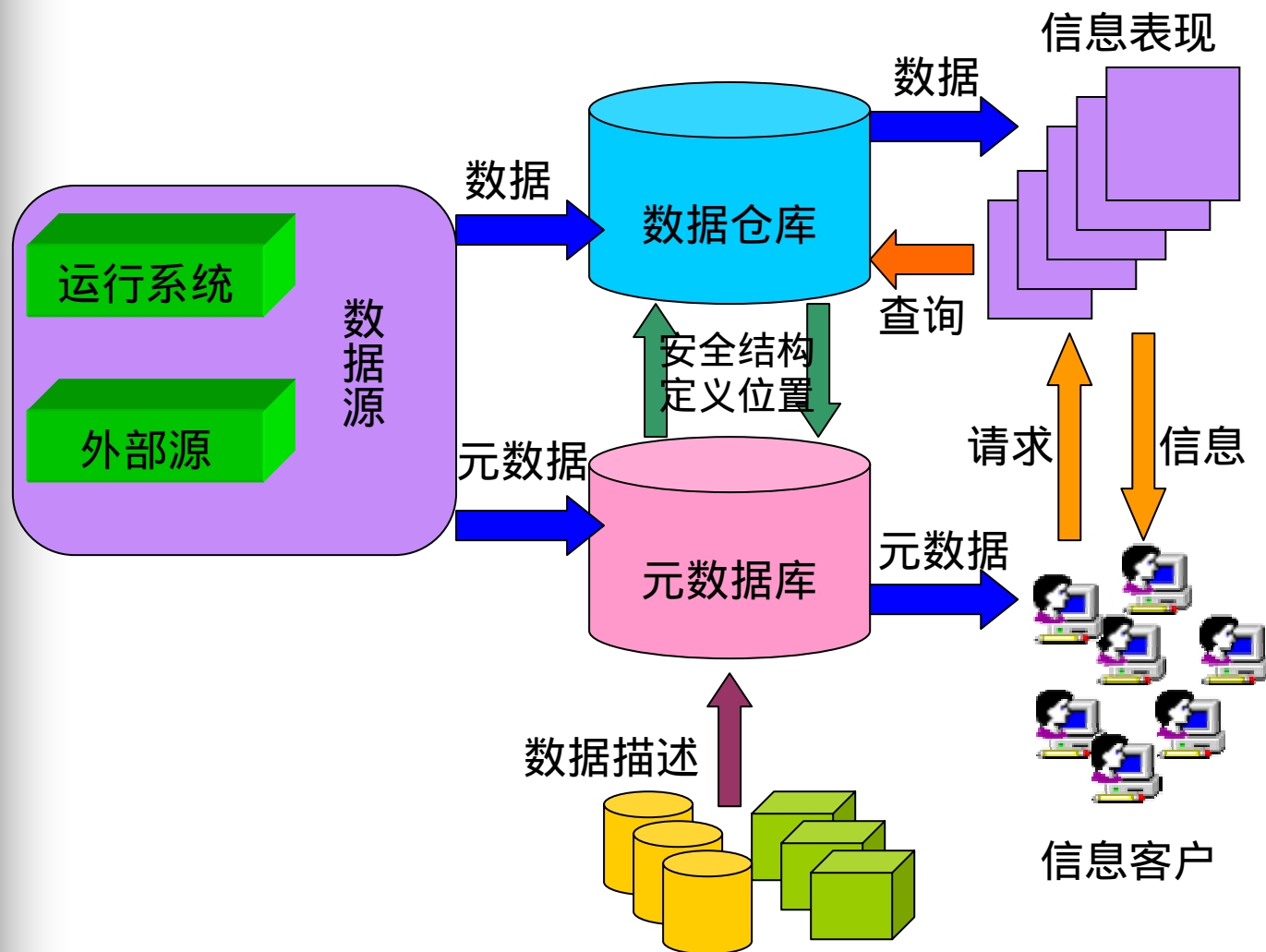
- 处理异构的、多来源的（文件系统或数据库）

● 4. 数据仓库体系结构

在数据仓库中的数据分为三个级别：细节级(有时候也分为早期细节级和当前细节级)、轻度综合级、高度综合级。



数据仓库概念图



数据仓库和数据模型设计



数据仓库的应用

- 风险分析
- 财务分析
- 欺诈分析
- 营销分析
- 资产管理
- 呼叫行为分析

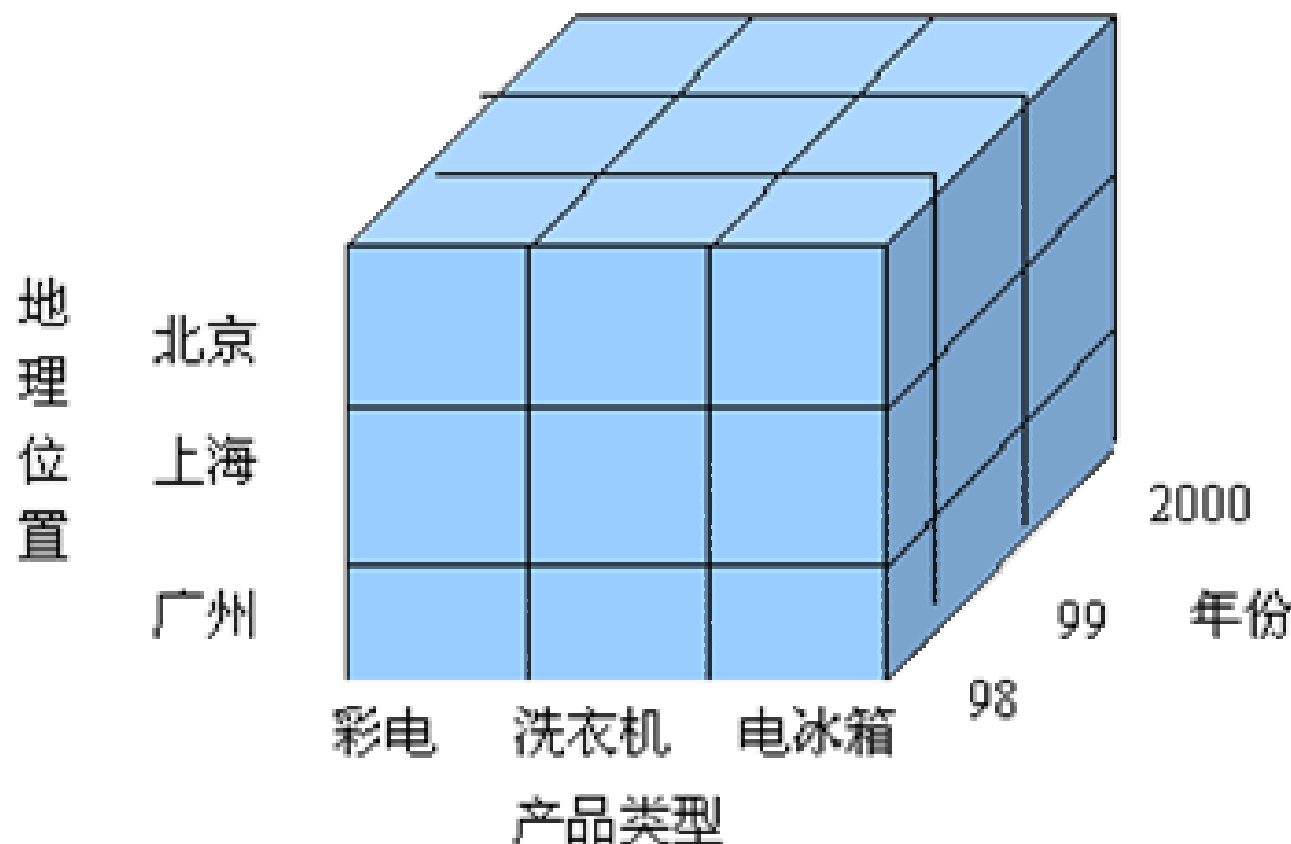


■ 11.4.2 联机分析处理 (OLAP)

● 1. 联机分析处理概述

- 联机分析处理(OLAP)是针对特定问题的联机数据访问和分析,通过对信息进行快速、稳定、一致和交互式的存取,对数据进行多层次、多阶段的分析处理,以获得高度归纳的分析结果。
- 提供自上而下、不断深入的分析工具,并将结果以较为直观的方式提供给用户。

多维数据库是以多维方式组织数据的





■ OLAP工具分为两类

- 基于多维数据库
- 基于关系数据库

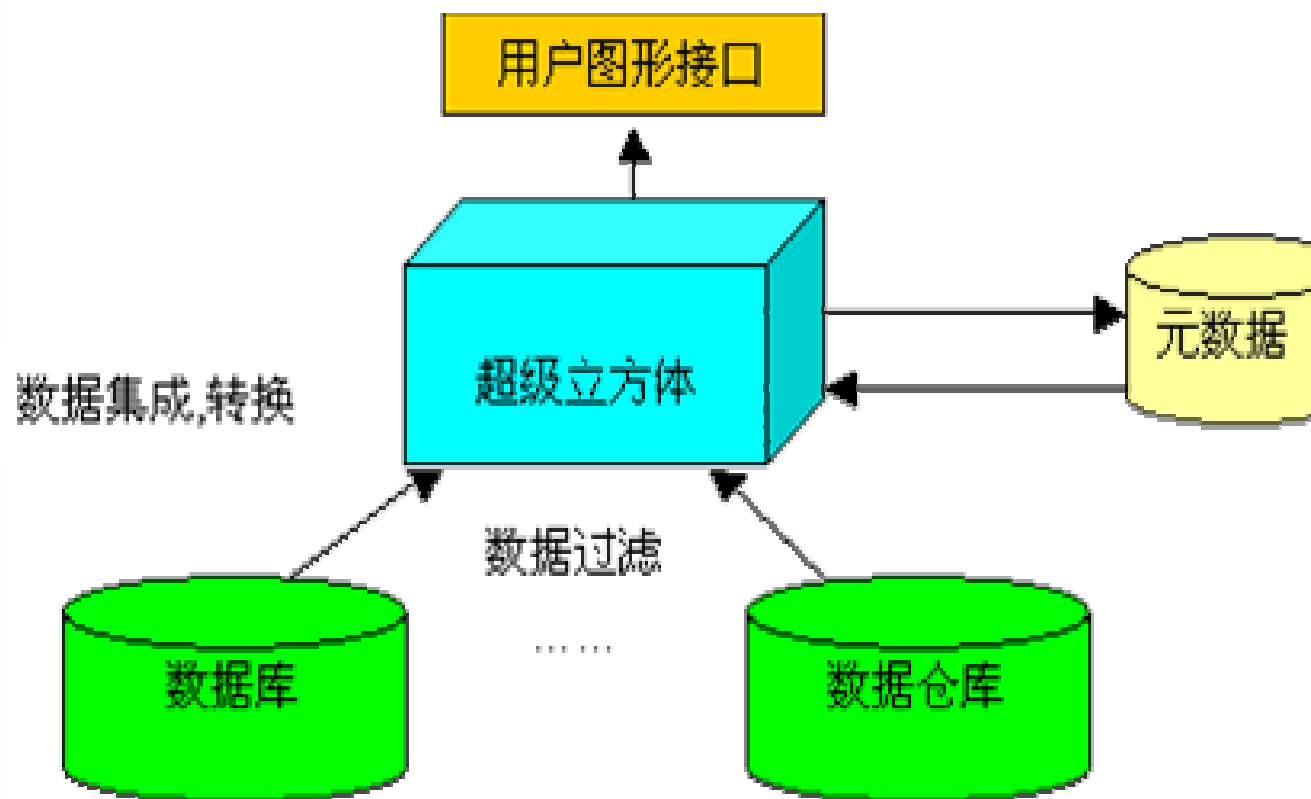
■ 相同点：

- 基本数据源为数据库和数据仓库
- 基于关系数据模型
- 向用户显示多维数据视图

■ 不同点：

- 基于多维数据库的方法是将分析所需的数据从数据仓库中抽取出来，物理地组织成多维数据库
- 基于关系数据库的方法是利用关系表来模拟多维数据，而不是物理地生成多维数据库。

● 2. 联机分析处理体系结构





OLAP的特性(Dr. Codd)

- 多维概念:将技术结构与业务并列起来
- 直观数据管理：GUI
- 可访问性：以合适的方式表示数据
- 解释性批处理提取
- OLAP分析模型
- 客户机/服务器结构
- 透明性和开放性
- 多用户：更新时保持数据的完整性
- 处理非正规数据：松耦合，强联系



OLAP的特性(Dr. Codd) (续)

- 存储OLAP结果：决策数据应与源数据分离
- 提取丢失数值：即空值 (NULL)
- 处理丢失值
- 弹性报告：
- 一致性能报告
- 对物理层自动调整
- 通用维：允许在不同的维之间进行比较和操作
- 无限制维与聚合层
- 无限制跨维操作



■ 11.4.3 数据挖掘

■ 定义：

- 数据挖掘就是这样一个过程，它从数据库的数据中识别出隐含的、未知的、有效的、新颖的、具有潜在效用的并最终可理解的信息(如规则、约束等)的非平凡过程。

■ 技术：

- 神经网络
- 基因算法
- 决策树

■ 方法：

- 使用一种数学模型，捕获一些特定的特征，同时忽略一些其他不关心的因素。

数据挖掘的用途

- CRM：发现并确定盈利客户的因素或促使客户转向竞争对手的因素；
- 医学：确定哪些治疗过程更为有效；哪些病人更适合于作外科手术；
- 市场营销：确定哪些客户更感兴趣于哪些特定商品或确定增加销售收入的方法；
- 制造业：通过数据构模来确定哪些过程参数最能影响产品的质量。



GIGO&NINO

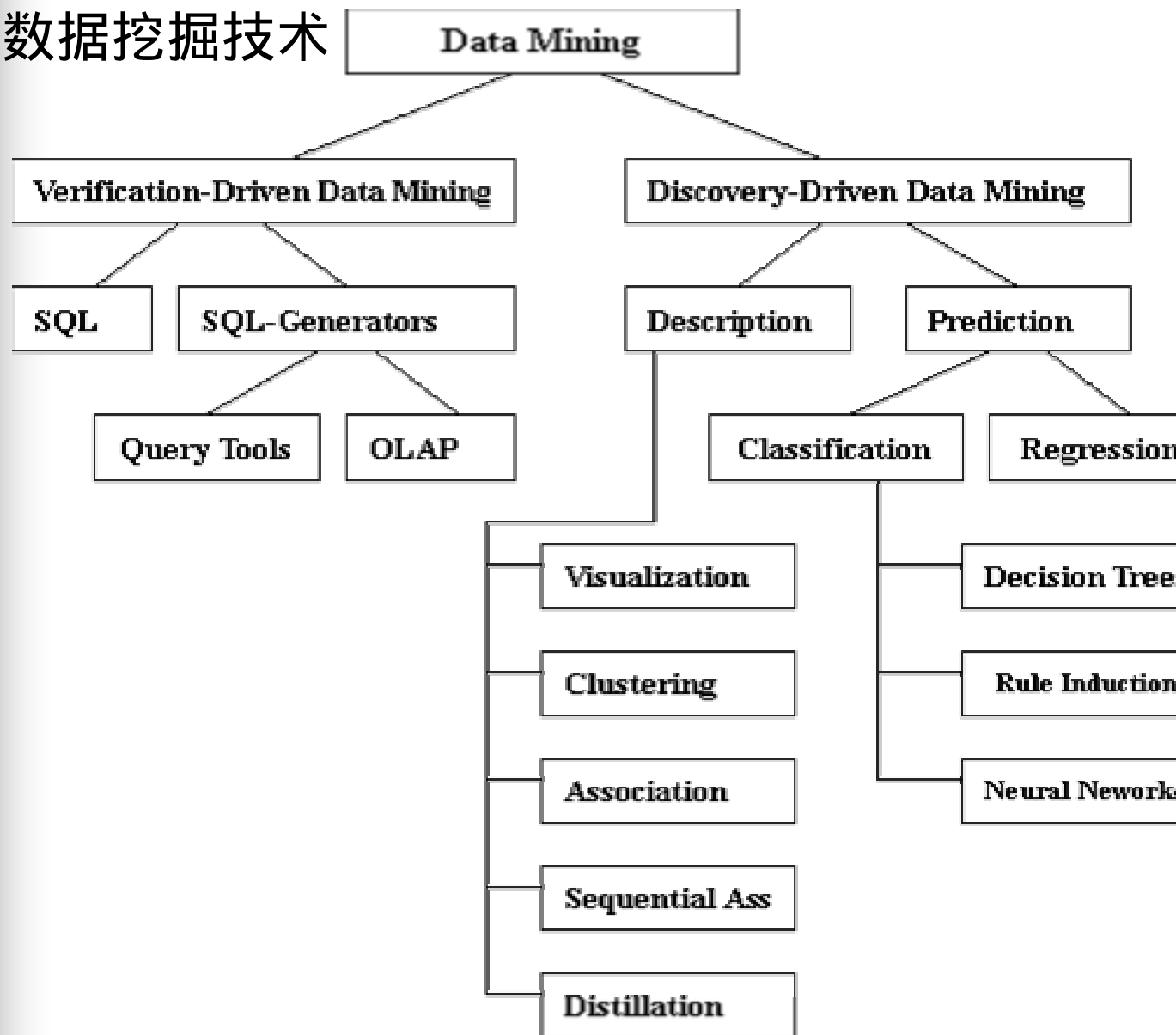
- 数据挖掘工作不会产生比数据或假设更好的结果，即：
 - GIGO - garbage in, garbage out , 或
 - NINO - nothing in, nothing out
- 数据质量非常重要
- 如何识别出“噪声”



数据挖掘的过程

- 了解数据（数据定义和理解）
 - 通过使用或创建元数据来完成，包括数据选择，计算，编码和分组。
- 数据清理（数据质量改进）
 - 是数据质量改进程序的步骤之一
- 数据准备
 - 是数据挖掘最重要的组成部分
- 模型开发（挖掘数据，解释和外推）
 - 通过模型来挖掘数据
- 模型的调整
 - 根据新的情况加以调整

数据挖掘技术






■ OLAP与数据挖掘的区别

● OLAP

- 一种传统的决策支持方法
- 是一种验证型的分析，即在某个假设前提下通过数据查询和分析来验证或否定该假设。其缺陷是：
 - 以假设为基础，从而对用户的假设能力有要求
 - 要求对用户的需求有全面深入的了解，而实际上有时用户的需求是不确定的
 - 抽取信息的质量依赖于用户对结果的解释，容易产生误导
- 用户驱动
- 分析层次较浅

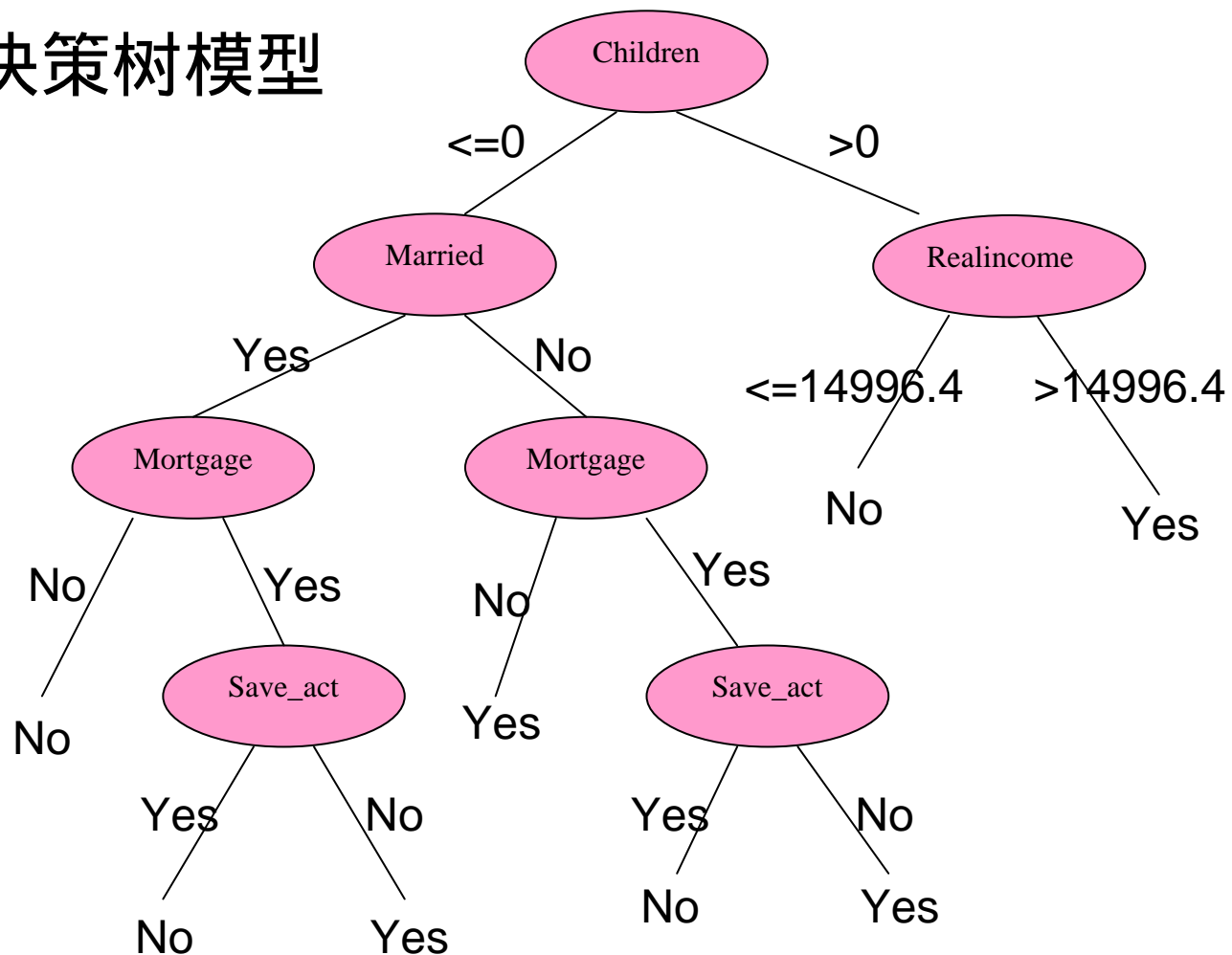


■ OLAP与数据挖掘的区别（续）

● 数据挖掘

- 数据驱动，是一种真正的知识发现方法
- 自动进行数据分析，即用户不必提出确切的要求，系统可以根据数据本身的规律性，自动地挖掘出数据潜在的模式，或通过联想，建立新的业务模型，并做出正确的决策
- 分析层次较深

决策树模型



数据挖掘的小例子

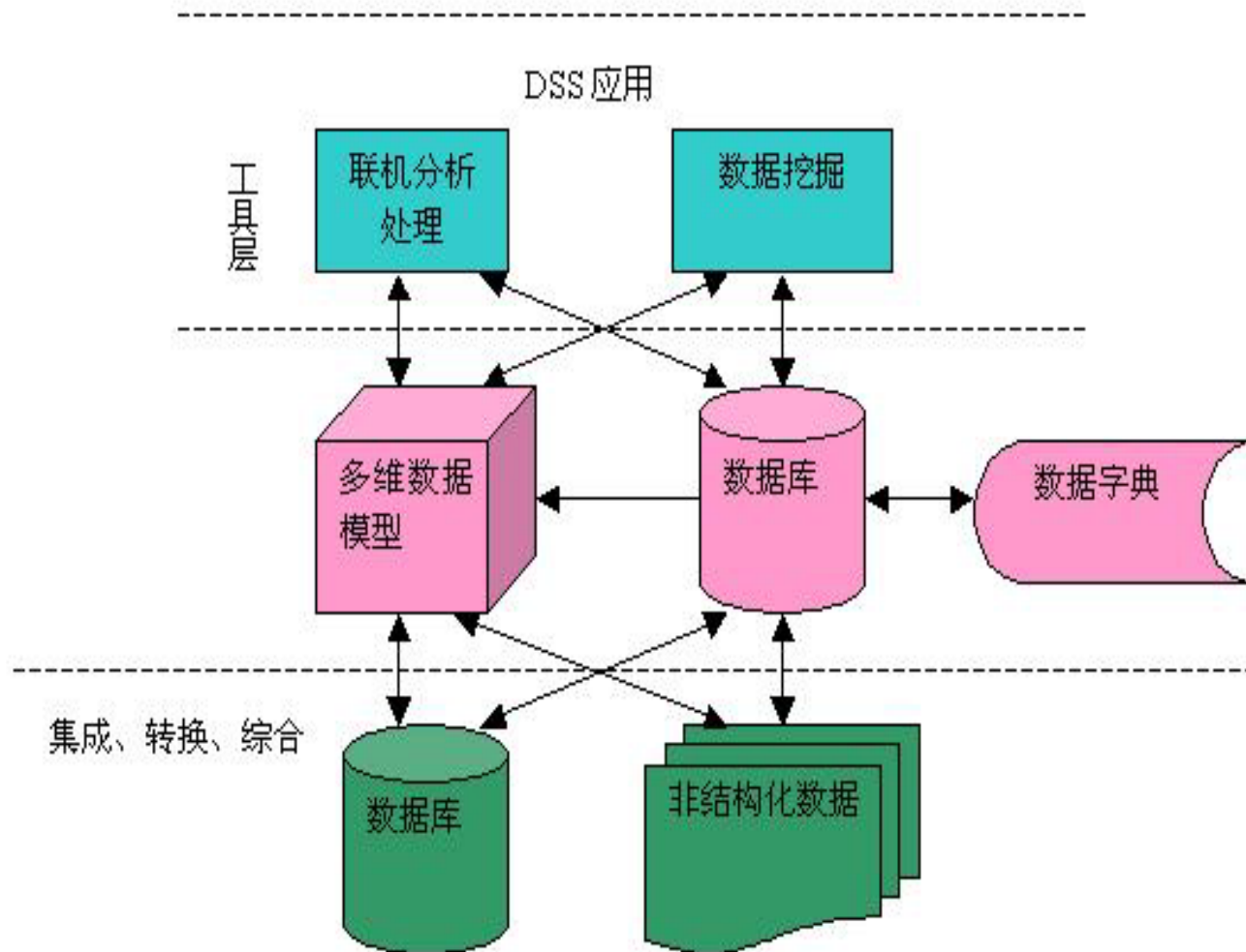
Table (300 records)

Table Generate

age	sex	region	income	married	children	car	save_act	current_act	mortgage	pop	realincome
48	FEMALE	INNER_CITY	17546.0	NO	1	NO	NO	NO	NO	YES	17546.0
40	MALE	TOWN	30085.1	YES	3	YES	NO	YES	YES	NO	10026.4
51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES	NO	NO	16575.4
23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES	NO	NO	6791.3
57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO	NO	NO	50576.3
57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES	NO	YES	18934.9
22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES	8877.07
58	MALE	TOWN	24946.6	YES	0	YES	YES	YES	NO	NO	24946.6
37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO	NO	NO	NO	12652.1
54	MALE	TOWN	24212.1	YES	2	YES	YES	YES	NO	NO	12106.1
66	FEMALE	TOWN	59803.9	YES	0	NO	YES	YES	NO	NO	59803.9
52	FEMALE	INNER_CITY	26658.8	NO	0	YES	YES	YES	YES	NO	26658.8
44	FEMALE	TOWN	15735.8	YES	1	NO	YES	YES	YES	YES	15735.8
66	FEMALE	TOWN	55204.7	YES	1	YES	YES	YES	YES	YES	55204.7
36	MALE	RURAL	19474.6	YES	0	NO	YES	YES	YES	NO	19474.6
38	FEMALE	INNER_CITY	22342.1	YES	0	YES	YES	YES	YES	NO	22342.1
37	FEMALE	TOWN	17729.8	YES	2	NO	NO	NO	YES	NO	8864.9
46	FEMALE	SUBURBAN	41016.0	YES	0	NO	YES	NO	YES	NO	41016.0
62	FEMALE	INNER_CITY	26903.2	YES	0	NO	YES	NO	NO	YES	26903.2

■ 11.4.4 决策支持方案

● 三种技术形成互补





■ 11.4.5 典型的数据仓库解决方案

● 1. IBM数据仓库解决方案

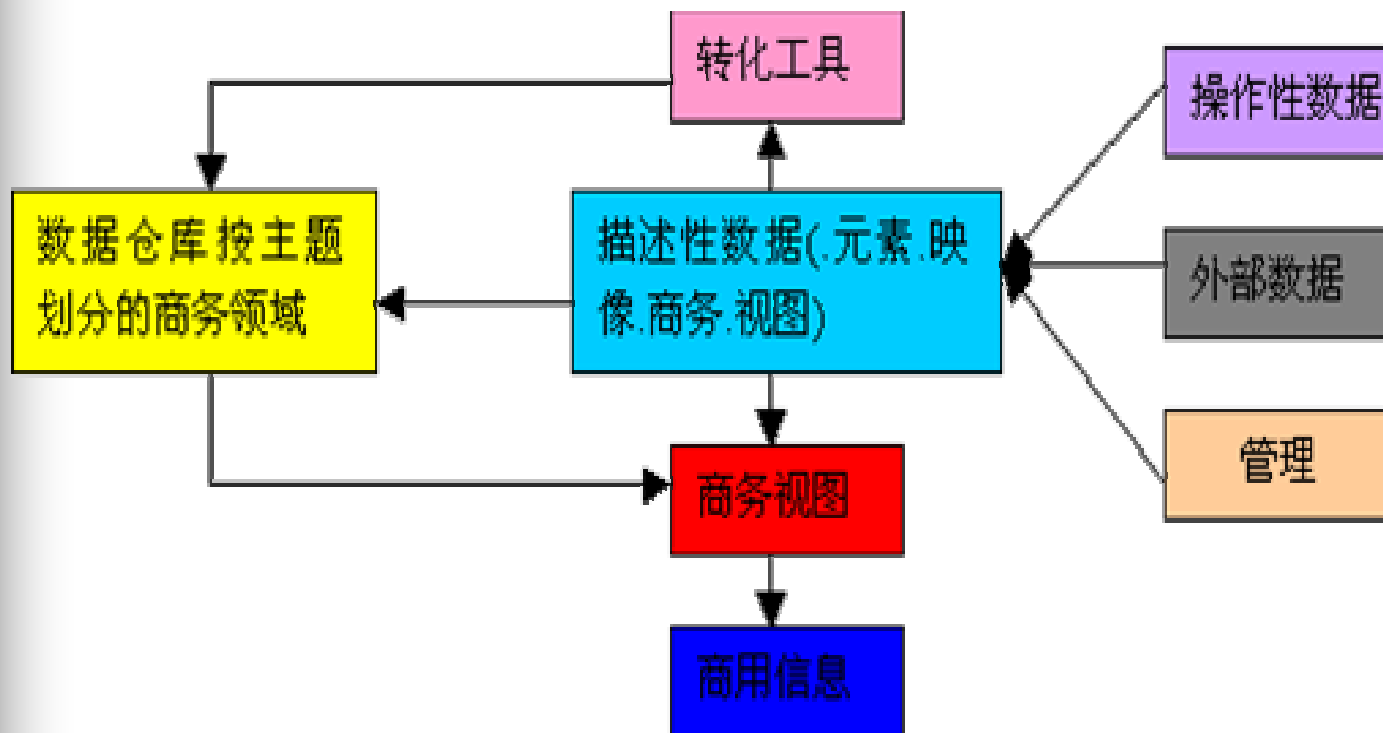
建立数据仓库包括以下阶段：

- 数据抽取阶段：
 - 完成对各种数据源的访问
- 数据转换阶段：
 - 完成对数据的清洗、汇总、整合
- 数据分布阶段：
 - 完成对结果数据存储的分配

此三阶段紧密结合在一个产品中：

- 数据仓库的存储：
 - 由DB2家族完成
- 数据的呈现：
 - 由不同产品完成不同的分析要求
- 管理和维护数据仓库中的数据：
 - 交给Visual Warehouse

IBM数据仓库解决方案



- 
- 2. ORACLE数据仓库解决方案
 - Oracle Express Server是一个先进的计算引擎和数据高速缓存。
 - Oracle 的OLAP工具主要包括
 - Oracle Express Objects、 Oracle Express Analyzer和Oracle Discoverer。
 - 将分析能力和Web浏览器相结合
 - Oracle Express Web Agent
 - 针对财务和销售而预制的OLAP应用
 - Oracle Financial Analyzer , Oracle Financial Controller和Oracle Sales Analyzer

ORACLE数据仓库解决方案

OLAP 应用软件

Oracle
Sales
Analyzer

Oracle
Financial
Analyzer

Oracle
Financial
Controller

OLAP 工具

第三
方
前
端
工
具

Oracle
Express
Objects

Oracle
Express
Analyzer

Oracle
Express
Web
Agent

Oracle
Discoverer

OLAP Server

Oracle Express Server

数据源

其他传统的和外
部数据库

关系型数据库管
理系统




11.5 基于代理技术的电子商务

■ 11.5.1 基于代理的信息收集

● 1. 信息收集代理概述

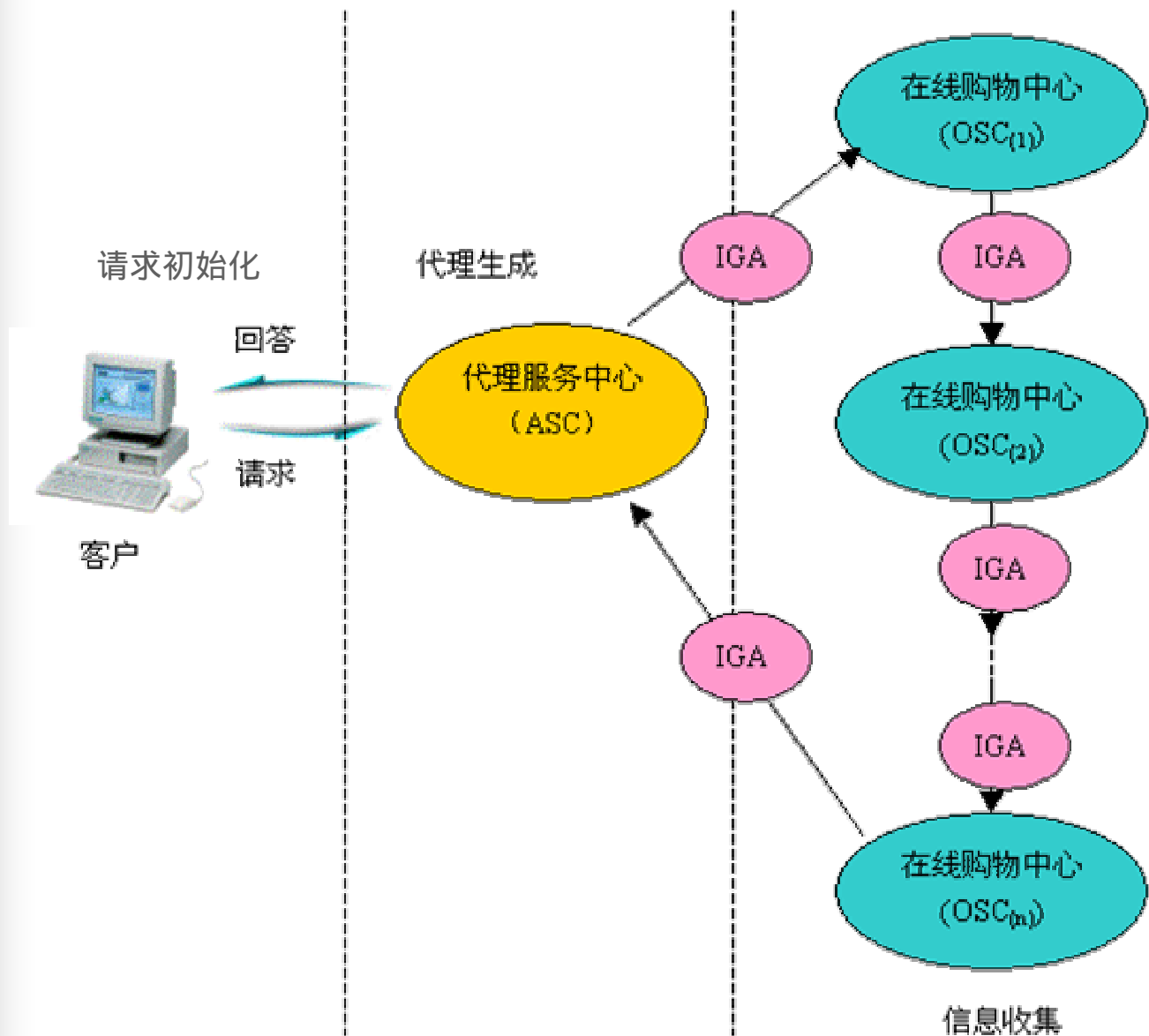
- 万维网已成为宝贵的信息资源
- 基于索引的搜索引擎，只是简单地给出许多与询问主题有关的文档。
- 通过个人信息代理主动地去寻找其他相关信息，而不是简单地对大量站点进行单个查询。



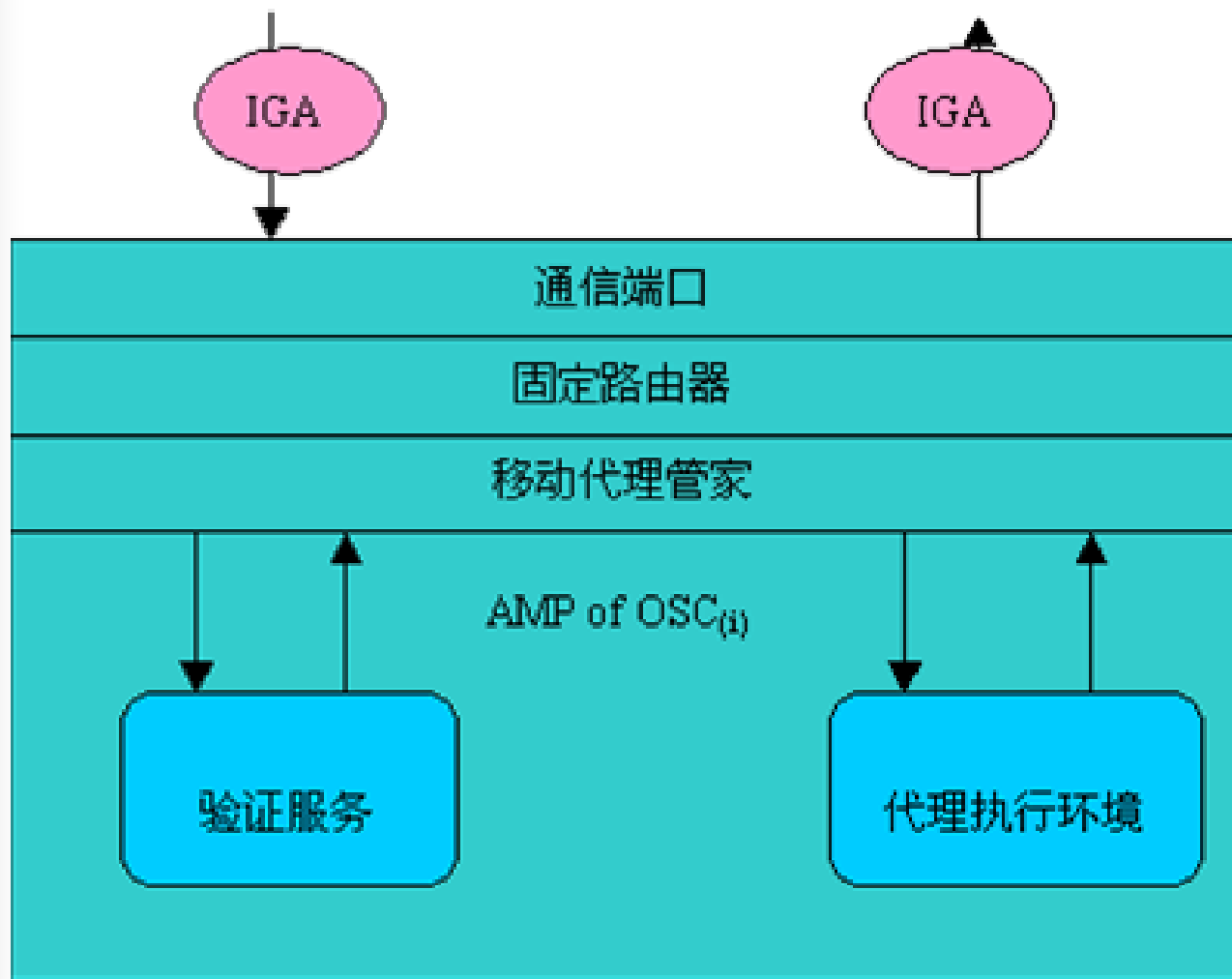
● 2.信息收集代理结构

- 客户(C)——发出收集信息请求
- 代理服务中心(ASC)——按照从客户来的不同请求生成IGA(信息收集代理), 发动IGA。
- 信息收集代理(IGA)——以ASC规定的目标为准去收集信息
- 在线购物中心(OSC)——每个在线购物中心都为IGA提供一个代理集会地(AMP), 使IGA在AMP上运行, 并按IGA的请求提供信息。

信息收集代理漫游的过程

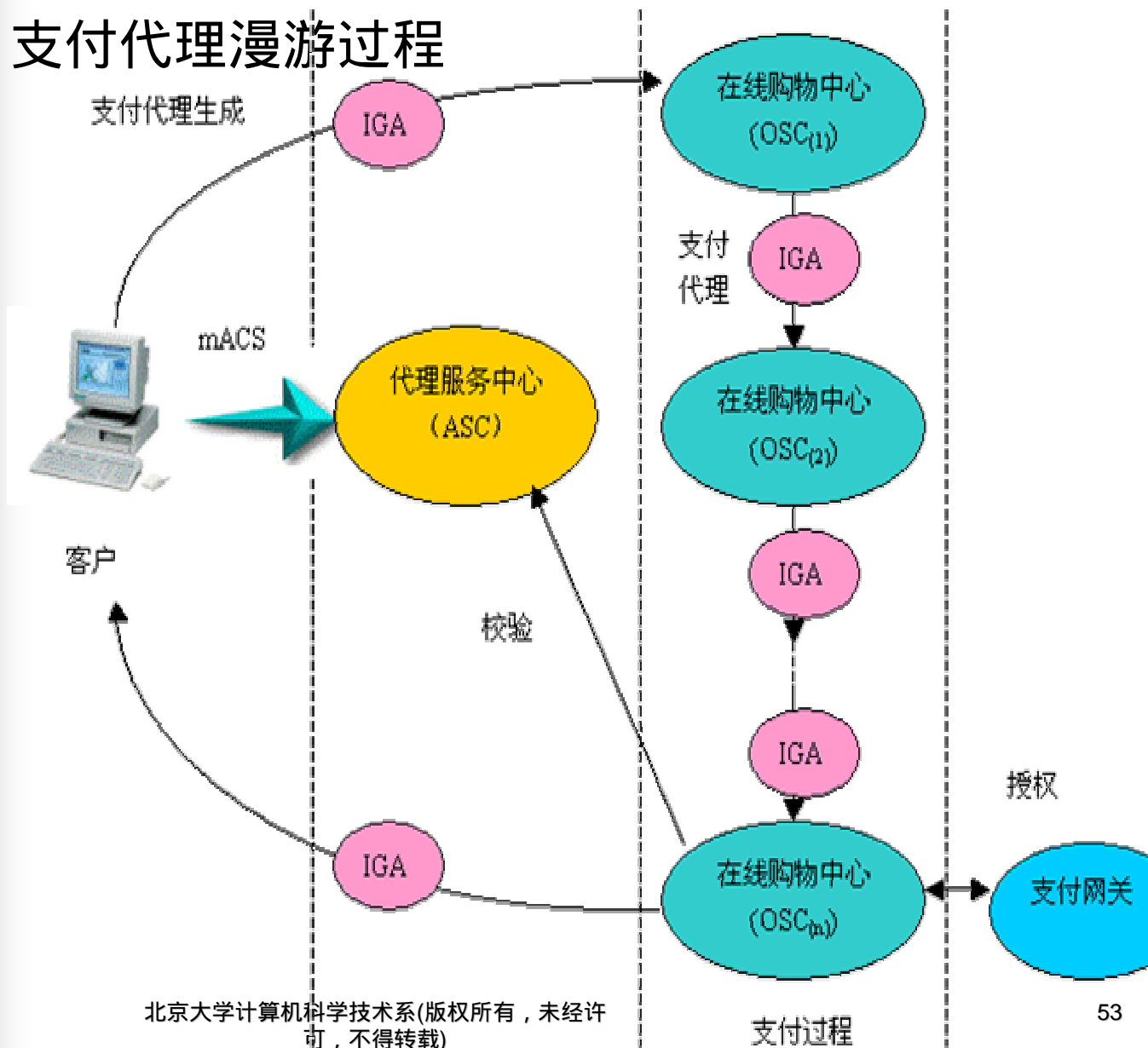


在OSC(i)的AMP上IGA的执行过程



■ 11.5.2基于代理的支付系统

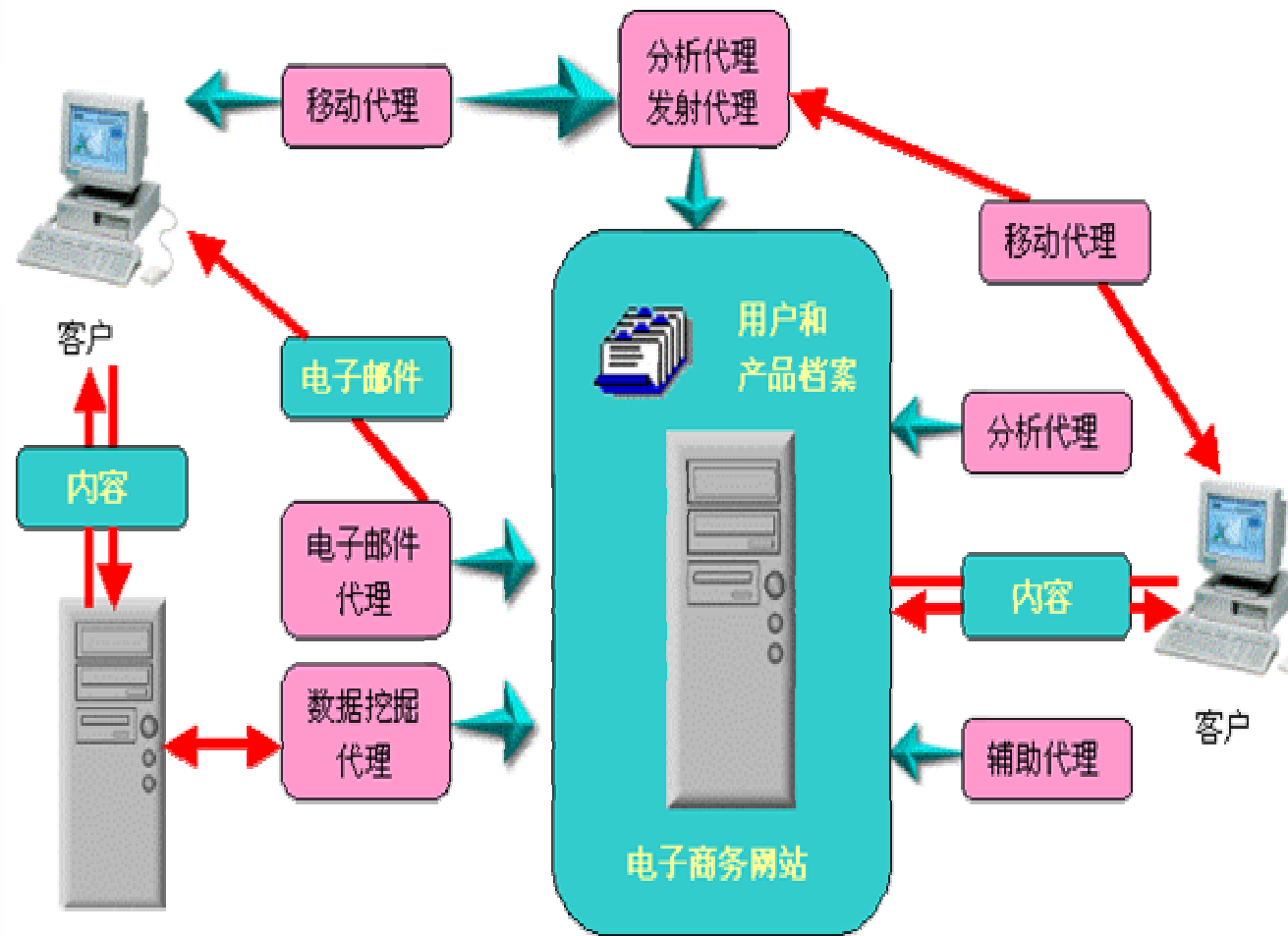
支付代理漫游过程



11.6 第三代电子商务

- 第二代模型的增强形式
- 运用具有交互功能并可随意选择的智能移动代理
- 以浏览器为中心
- 增加对多媒体应用的使用

第三代电子商务系统



11.7 潜在问题和挑战

- 最大挑战是相关技术
- 移动代理的使用可能危及系统完整和用户隐私
- 通过代理表现出的恶意(malicious intent)
- 隐含着大量社会和法律问题
- 有许多和信息分享相关的隐私问题



思考题

1. 代理有哪些类型？
2. 代理有哪些特性？
3. 什么是数据仓库？它有哪些特点？
4. 什么是联机分析处理？什么是数据挖掘？
5. 第三代电子商务面临哪些挑战和困难？



OVER Ch11



Thank You