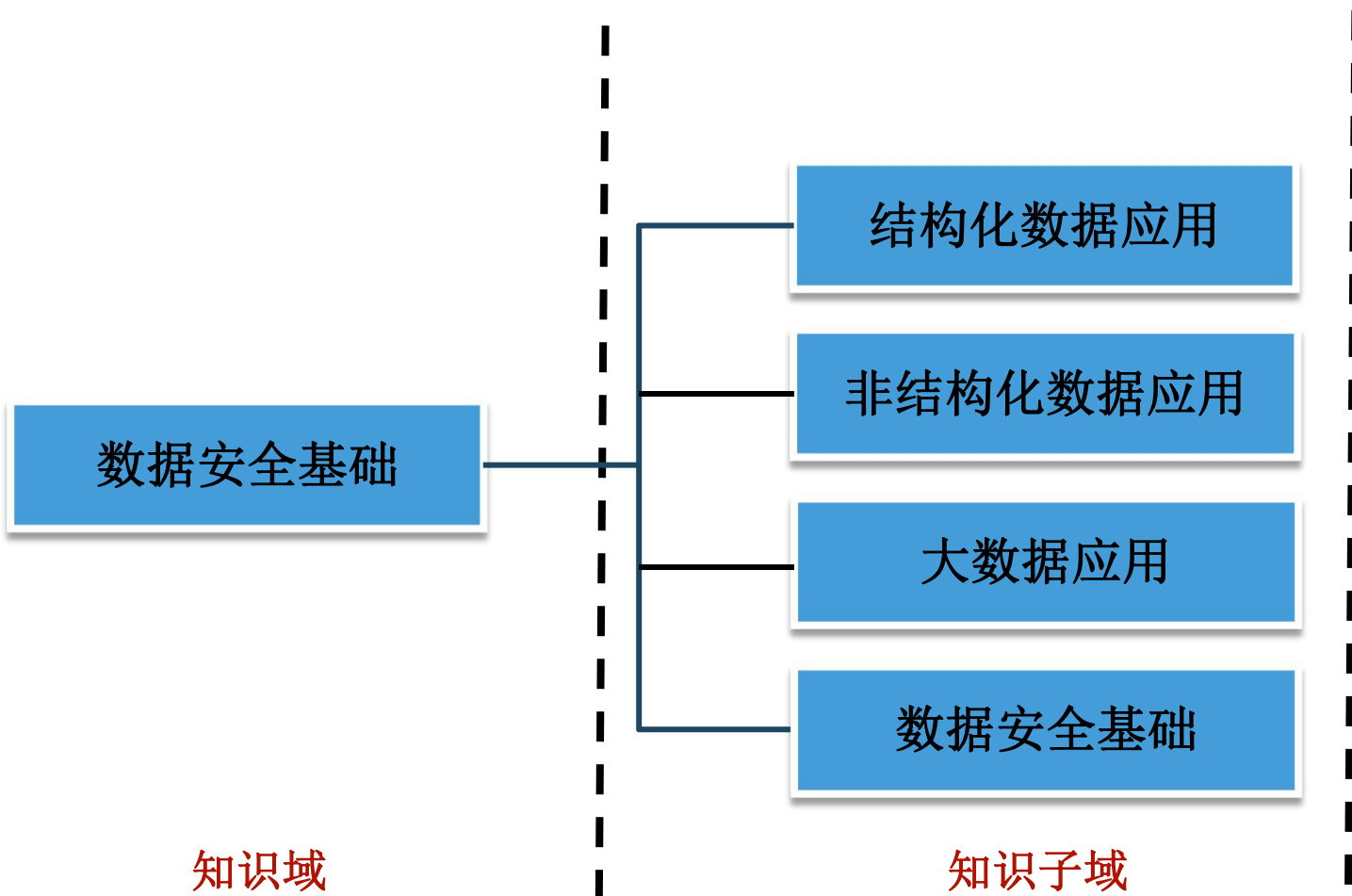
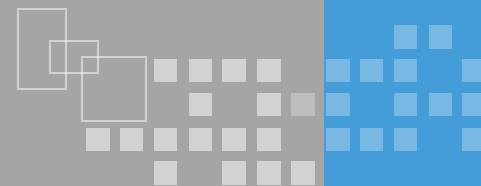




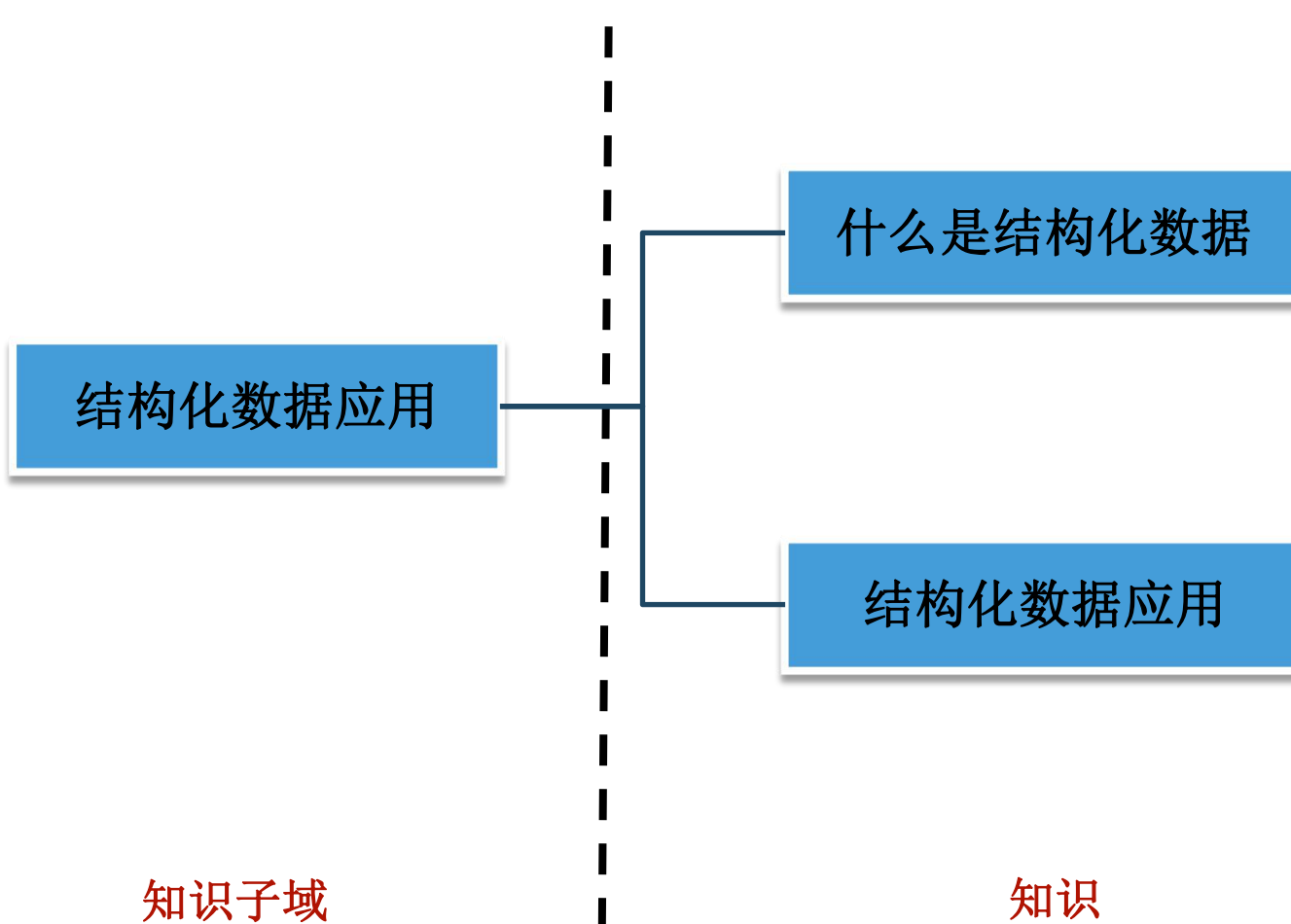
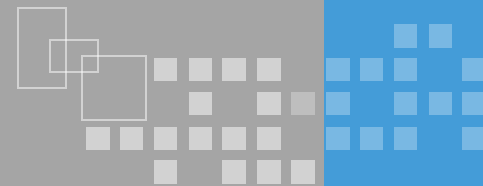
# CISP-DSG 数据安全基础

版本：1.0

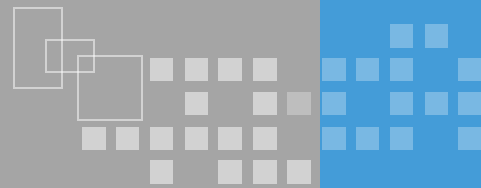
讲师姓名      机构名称



# 知识子域：结构化数据应用

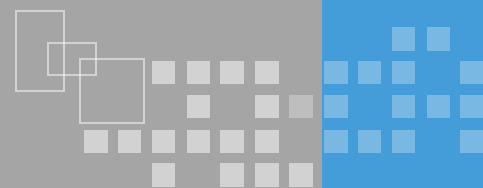


# 知识：什么是结构化数据



结构化数据，简单来说就是数据库。结合到典型场景中更容易理解，比如企业ERP、财务系统；医疗HIS数据库；教育一卡通；政府行政审批；其他核心数据库等。

结构化数据：行数据，存储在数据库里，可以用二维表结构来逻辑表达实现的数据。



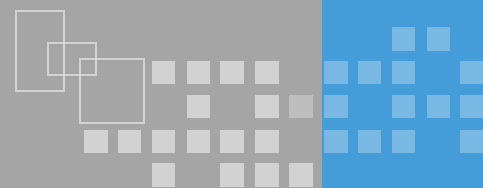
## 信息、数据与数据处理

在数据处理中，我们最常用到的基本概念就是数据和信息，信息与数据有着不同的含义。

### 一、信息 (Information)

1、定义：信息是关于现实世界事物的存在方式或运动状态的反应的综合，具体说是一种被加工为特定形势的数据，但这种数据形式对接收者来说是有意义的，而且对当前和将来的决策具有明显的或实际的价值。

如：“2019年硕士研究生将扩招30%”，对接收者有意义，使接收者据此作为决策。

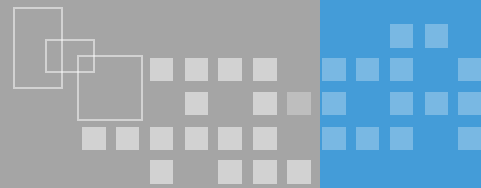


## 2、信息的特征

信息源于物质和能量，它不可能脱离物质而存在，信息的传递需要物质载体，信息的获取和传递要消耗能量。如信息可以通过报纸、电台、电视、计算机网络进行传递。

信息是可以感知的，人类对客观事物的感知，可以通过感觉器官，也可以通过各种仪器仪表和传感器等，不同的信息源有不同的感知形式。

信息是可存储、加工、传递和再生的。动物用大脑存储信息叫做记忆。计算机存储器、录音、录像等技术的发展，进一步扩大了信息存储的范围。借助计算机，还可以对收集到的信息进行取舍和整理。



## 二、数据 (Data)

### 1、数据的定义

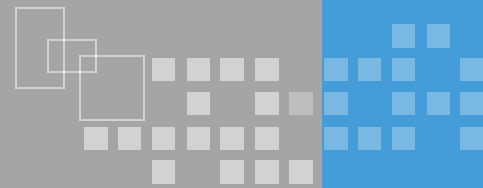
数据是用来记录信息的可识别的符号，是信息的具体表现形式。

### 2、数据的表现形式

可用多种不同的数据形式表示同一信息，而信息不随数据形式的不同而改变。

数据的概念在数据处理领域中已大大地扩宽了，其变现形式不仅包括数据和文字，还包括图形、图像、声音等。这些数据可以记录在纸上，也可以记录在各种存储器中。

# 信息和数据的联系



数据是信息的符号表示或载体，信息则是数据的内涵，是对数据的语义解释。

数据处理是讲数据转换成信息的过程，包括对数据收集、存储、加工、检索、传输等一系列活动。其目的是从大量的原始数据中抽取和推导出有价值的信心，作为决策的依据。

可用下面简单的表示信息、数据域数据处理的管理：

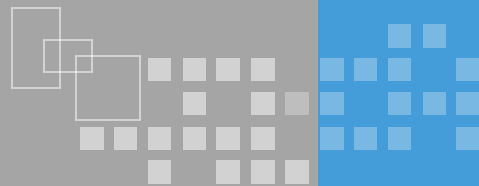
$$\text{信息} = \text{数据} + \text{数据处理}$$

数据是原料，是输入，而信息是产出，是输出结果。

“信息处理”的真正含义应该是为了产生信息而处理数据。



# 知识一：数据库系统概述



## 数据：

用来记录信息的可识别的符号，是信息的载体和具体表现形式。数据可以记录在纸上，也可以记录在各种存储器中

## 数据库：

为实现一定目的按某种规则组织起来的“数据”的“集合”

## 数据库管理系统(DBMS)：

专门用于管理数据库的计算机系统软件。数据库管理系统能够为数据库提供数据的定义、建立、维护、查询、统计等操作功能，并具有对数据的完整性、安全性进行控制的功能。

## 数据库系统：

带有数据库并利用数据库技术进行数据管理的计算机系统。一个数据库系统应由计算机硬件、数据库、数据库管理系统、数据库应用系统和数据库管理员5部分构成。



图 6.1 数据库系统

# 数据库系统的功能



## 数据定义功能

提供数据定义语言（Data Description Language, DDL），并提供相应的建库机制

## 数据操纵功能

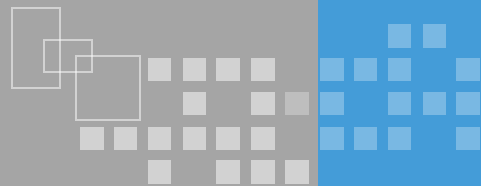
实现数据的插入、修改、删除、查询、统计等数据存取操作的功能

## 数据库的建立和维护

数据的载入、转储、重组织功能及数据库的恢复功能致性

## 数据库的运行和管理

并发控制、数据的存取控制、数据完整性条件的检查和执行、数据库内部的维护等

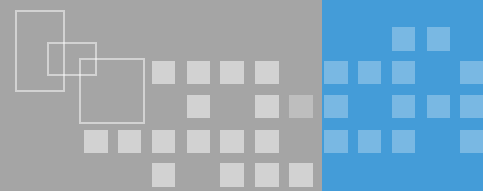


## ❖ 什么是数据库？

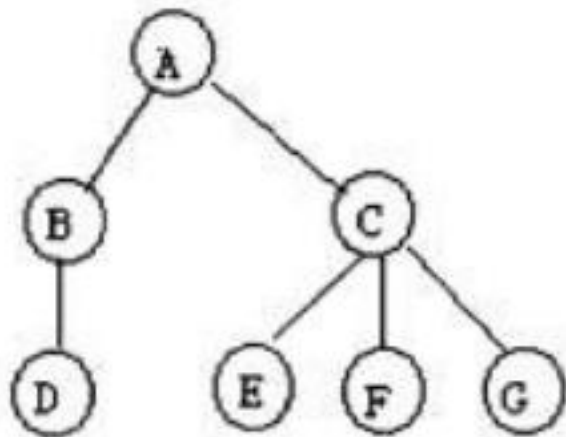
- 数据库是按照数据结构来组织、存储和管理数据的仓库

## ❖ 数据库组织结构

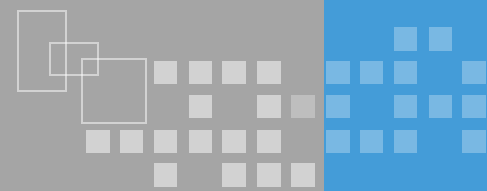
- (1) 层次式数据库
- (2) 网络式数据库
- (3) 关系式数据库



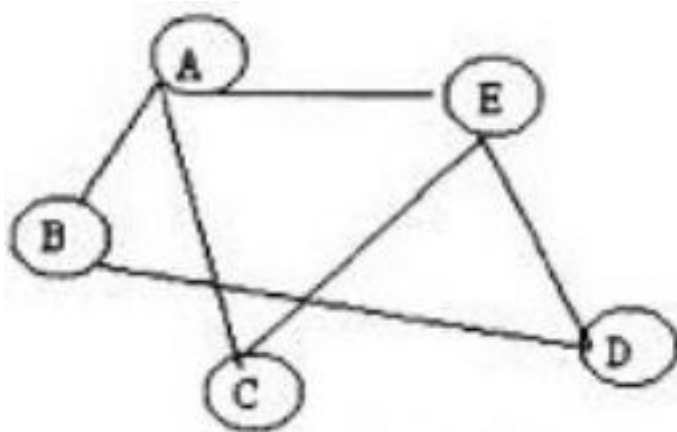
## 一、层次数据结构



- 举例说明：
  - 大学分为系
  - 系下面分为专业
  - 专业下面分为班级



## 二、网状数据结构

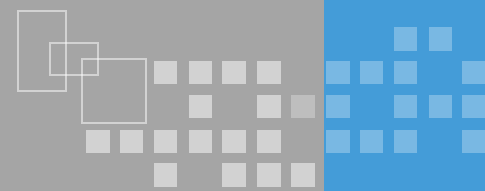


- 举例说明
  - 一个老师管理多个学生
  - 一个学生面对多个老师

# 数据库组织结构—关系型数据结构

## 三、关系型数据结构

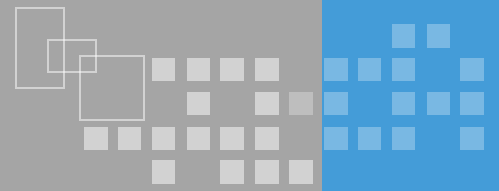
雇员号	雇员姓名	性别	出生日期	电话	所在部门
1001	李建华	男	56-5-8	2201	家电部
1002	黄玲	女	66-9-5	2201	家电部
1003	王晓新	男	70-5-26	2202	家电部
1004	刘燕	女	71-10-18	2205	服装部
1005	宋健辉	男	69-5-6	2206	服装部
1006	张海燕	女	68-9-23	2207	服装部
1007	何晓	男	68-5-21	2210	文化用品部
1008	张杨	男	70-1-4	2211	文化用品部



## ❖ 概念

- DBMS (DataBase Management System) , 用于管理数据库, 提供数据库的引擎。 类型
- 规模大小分类
  - 小型数据库管理系统 (Excel、Access、FoxPro)
  - 中型数据库管理系统 (SQL Server、MySQL)
  - 大型数据库管理系统 ( Oracle 、 Informix、 DB2)
- 存储类型分类
  - 文件存储类型 ( Excel、 Access、 FoxPro )
  - 结构存储类型 (SQL Server、 Oracle等等)

# SQL Server 的特点



- ❖ 与Windows无缝集成（win认证方式）
- ❖ 全图形界面操作环境，便于使用
- ❖ 提供Web和电子商务功能。对XML和Internet标准的支持。
- ❖ 可靠的安全机制



# 知识二、SQL语句及常见函数的用法

## ❖ SQL语句介绍-Select 语句语法

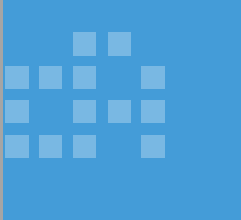
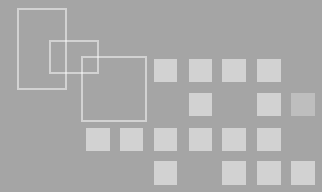
语法：

**Select** 字段1,字段2,字段3 **From** 数据表名称 **Where**  
条件

举例：

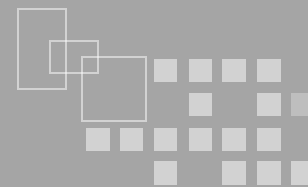
在科目表中查询所有末级科目，显示科目代码、科目名称、科目末级标志

```
SELECT cCode, cCode_Name, bEnd FROM code  
WHERE bEnd=1
```



## ❖ Select 查询条件

- 1) = /</>/>/< : **Where** 字段 = 值  
字符型( ' ) 、数值型、日期型(#)
- 2) like: **Where** 字段 like 字段%  
字符型通配符 % 、 ?
- 3) Between: **Where** 字段 **Between** 最小值  
**And** 最大值
- 4) In: **Where** 字段 **In** (值1, 值2, 值3)
- 5) Not: **Where** **Not** 字段 = 值

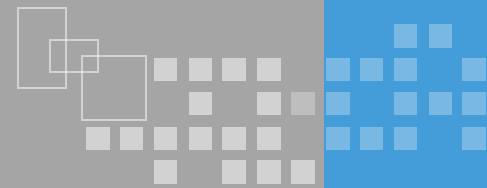


## ❖ Group By

- 语法：Select 统计函数(字段1) From 数据表  
Group By 字段2
- 分组统计函数
  - AVG()、COUNT()、MAX()、MIN()、SUM()

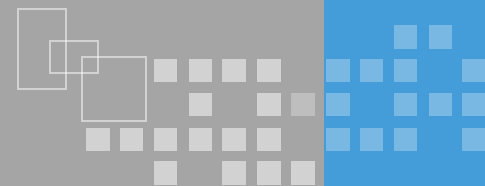
举例：按部门统计查询每个部门的工资总和

**SELECT** sum(工资) **FROM** 人员工资表 **GROUP BY**  
部门



## ❖ Order By

- 语法：Select 字段1、字段2、字段3 From 数据表  
Order By 字段
- 举例：查询存货的编号、名称、计量单位，并按照  
存货名称排序
- SELECT cInvcCode, cInvName, cInvM\_Unit FROM  
dbo.Inventory ORDER BY cInvName



## ❖ Case关键字

- 语法：字段 **Case When** 字段 = 值1 **Then** 显示1  
**When** 字段 = 值2 **Then** 显示2 **End AS** 别名

举例： 查询总帐模块是否结账

```
SELECT CASE WHEN iperiod=0 THEN '期初' WHEN  
iperiod<>0 THEN '第' + ltrim(str(iperiod)) +  
'期' END AS 期间, CASE WHEN bflag=1 THEN '  
已结账' WHEN bflag=0 THEN '未结账' END AS 总  
帐模块 FROM gl_mend
```

# SQL语句介绍—SELECT语句多表查询



	cPersonCode	cPersonName	cDepName
▶	001	林同	厂办
	002	李钢	厂办
	003	王宝扇	厂办
	004	王秋霞	厂办
	005	张晴	财务科
	006	李一凡	财务科
	007	张成	财务科
	008	薛明	财务科
	009	李大兵	金工车间
	010	张仪	金工车间
	011	周迪生	金工车间
	012	朱建	金工车间

❖ 举例：查询职员表中的职员编号，职员姓名以及相关联的部门名称

❖ 语句：SELECT

```
Person. cPersonCode, Person. cPersonName, Department. cDepName FROM Person LEFT OUTER JOIN  
Department ON Person. cDepCode =  
Department. cDepCode
```

# SQL语句介绍—INSERT INTO插入语句

## ❖ 语法：

- Insert Into 数据表 (字段1, 字段2, 字段3) Values (值1, 值2, 值3)
- Insert Into 数据表  
Select 字段1, 字段2, 字段3 From 数据表 Where 条件

## ❖ 举例：

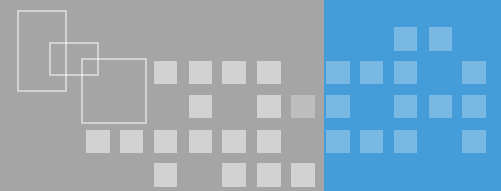
在“用户表”中插入记录，记录内容：姓名=张三，年龄=29，性别=男

- Insert Into 用户 (姓名, 年龄, 性别) Values ( '张三', 29, '男' )

从“用户1”表中查找年龄大于20岁的人员插入到“用户”表中

- Insert Into 用户 Select 姓名, 年龄, 性别 From 用户1  
Where 年龄 > 20

# SQL语句介绍—UPDATE更新语句



❖ 语法: **Update** 数据表 **Set** 字段1=值1, 字段2=值2  
**Where** 条件

❖ 举例

在Code表中, 将科目'1001'的末级标志改为1

- **Update Code Set Bend=1 Where cCode = '1001'**

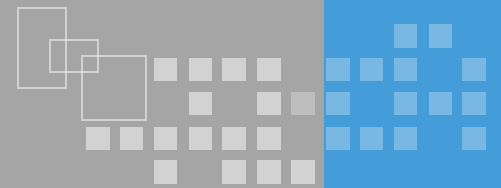
在“用户”表中, 将所有人的年龄增加1岁

- **Update 用户 Set 年龄=年龄+1**

注意: 要加入适当的Where条件, 否则将更新表中的所有记录



# SQL语句介绍—DELETE删除语句



❖ 语法: Delete From 数据表 Where 条件

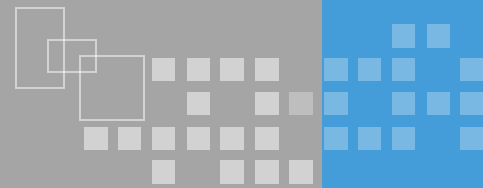
❖ 举例

删除总帐模块的科目互斥记录

■ DELETE FROM GL\_mccontrol

注意: 要加入适当的Where条件, 否则将更新表中的所有记录

# SQL语句介绍—常用函数介绍(1)



## ❖ Sum 求和函数

语法: sum(统计字段)

举例: 计算所有人员的工资综合

```
SELECT SUM(工资) FROM 人员工资
```

## ❖ Min 最小值函数

语法: min(统计字段)

举例: 得到所有人员中工资最少的工资数

```
SELECT MIN(工资) FROM 人员工资
```

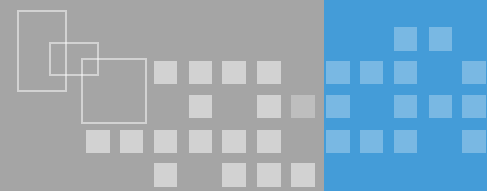
## ❖ Max 最大值函数

语法: max(统计字段)

举例: 得到所有人员中工资最多的工资数

```
SELECT MAX(工资) FROM 人员工资
```

# SQL语句介绍—常用函数介绍(2)



## ❖ Count 记录数量函数

语法: count(统计字段)

举例: 得到用户人数

```
SELECT COUNT(姓名) FROM 人员工资
```

## ❖ Len 字符串长度函数

语法: len(字符串)

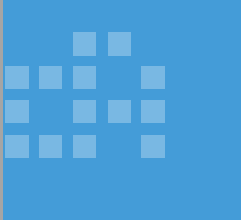
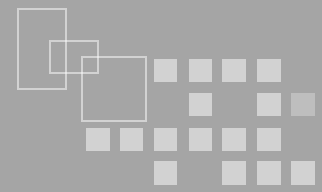
```
SELECT LEN('abc') 返回值: 3
```

## ❖ Left 从左边截取字符串函数

语法: left(字符串, 截取长度)

```
SELECT LEFT('abc', 1) 返回值: a
```

# SQL语句介绍—常用函数介绍(3)



## ❖ Right 从右边截取字符串函数

语法: right(字符串, 截取长度)

SELECT **RIGHT**( 'abc' , 1) 返回值: c

## ❖ Replace 字符串替换函数

语法: replace(字符串, 原字符串, 替换字符串)

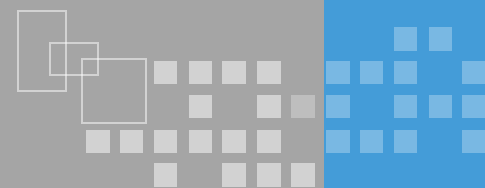
SELECT **REPLACE**( 'abc' , 'a' , 'b' ) 返回值:  
bbc

## ❖ Substring 字符串截取函数

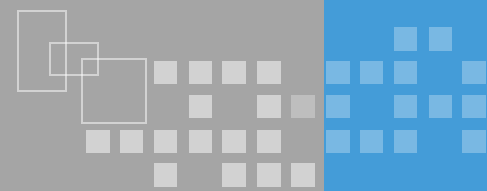
语法: substring(字符串, 开始位置, 截取长度)

SELECT **SUBSTRING**( 'abc' , 1, 2) 返回值: ab

# SQL语句介绍—常用函数介绍(4)



- ❖ **IsNull** 如果字段的值是空字符，那么使用指定的值替换  
语法： `IsNull (字段名称, 替换值)`  
举例：在“人员工资”表中，如果人员工资为空，那么该人员的工资为0  
`SELECT ISNULL(工资, 0) FROM 人员工资`
- ❖ **Isdate** 判断字符是否为有效日期，如果是有效日期，返回1；如果不是有效日期，返回0  
语法： `isdate(字段名称/字符串)`  
举例：判断“用户”表中的出生日期是否为有效日期格式  
`SELECT ISDATE(出生日期) FROM 用户`
- ❖ **Isnumeric** 判断字符是否为有效数值，如果是有效数值，返回1；如果不是有效数值，返回0  
语法： `isnumeric(字段名称/字符串)`  
举例：判断“人员工资”表中的工资是否为有效数值格式  
`SELECT ISNUMERIC(工资) FROM 人员工资`



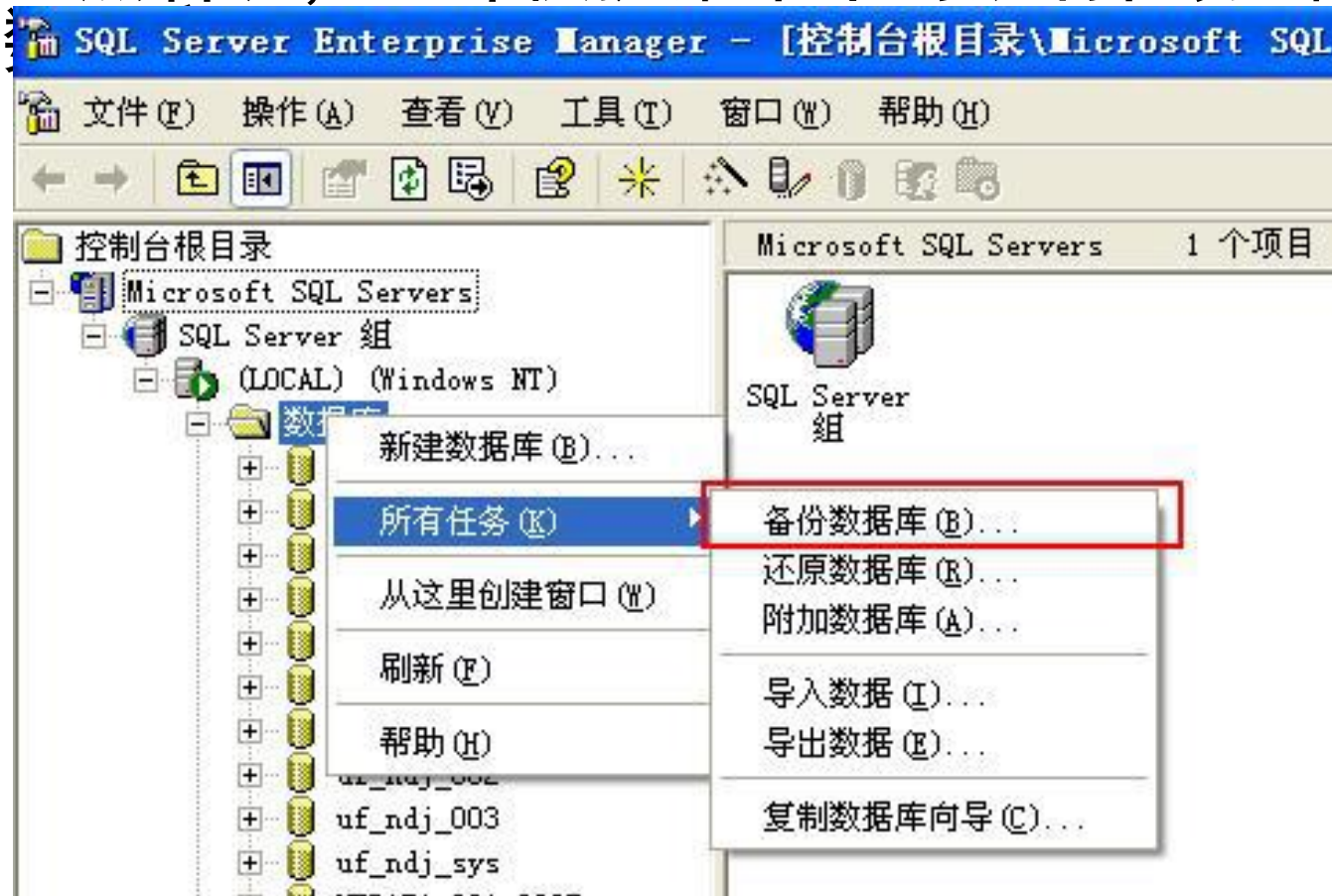
## 管理工具的使用—企业管理器

### ❖ 企业管理器

- 备份数据库
- 还原数据库
- 附加数据库（只有.mdf、.ldf文件）
- 导入、导出数据表记录（SQL之间、Excel文件）
- 查询及修改数据
- 字段类型
  - int、numeric、float                      数值型
  - bit    字节型 1或0
  - char、varchar                              字符型
  - Datetime                                      日期型

# 企业管理器—备份数据库(1)

- ❖ 打开SQL Server企业管理器。在“数据库”的位置点右键，选择快捷菜单中的所有任务/备份



# 企业管理器—备份数据库(2)

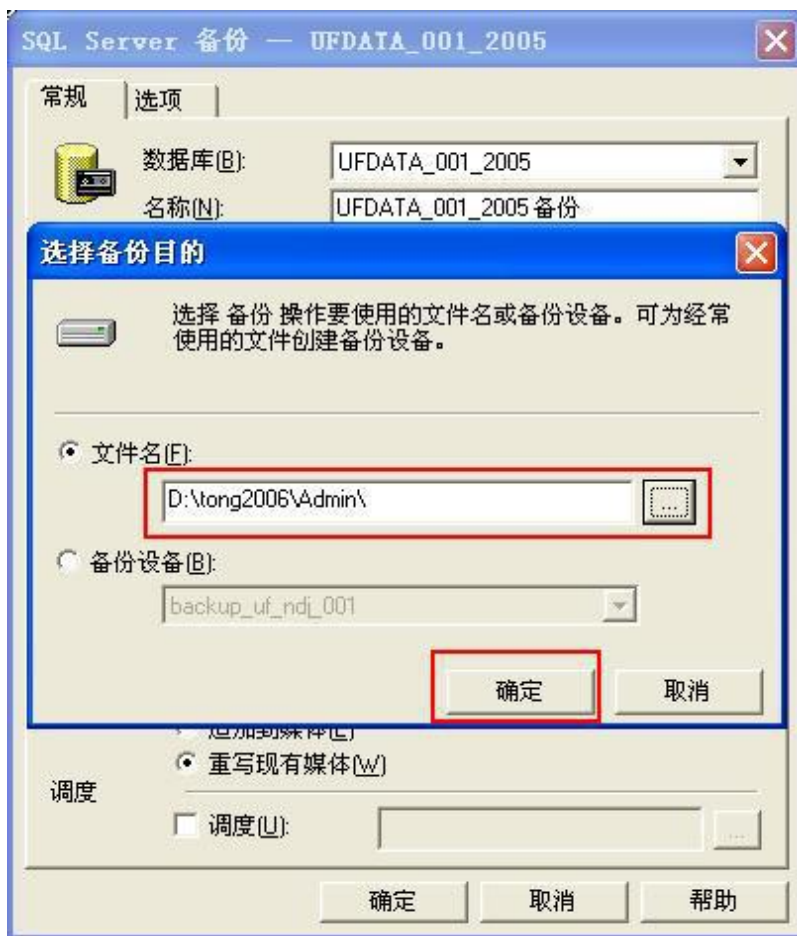
- ❖ 在打开的备份窗口中，选择需要备份的帐套。然后选择“重写现有媒体”。点“添加”按钮，选择数据库的备份位置以及备份文件名称。





# 企业管理器—备份数据库(3)

❖ 录入备份文件名称或点旁边的选择按钮选择。



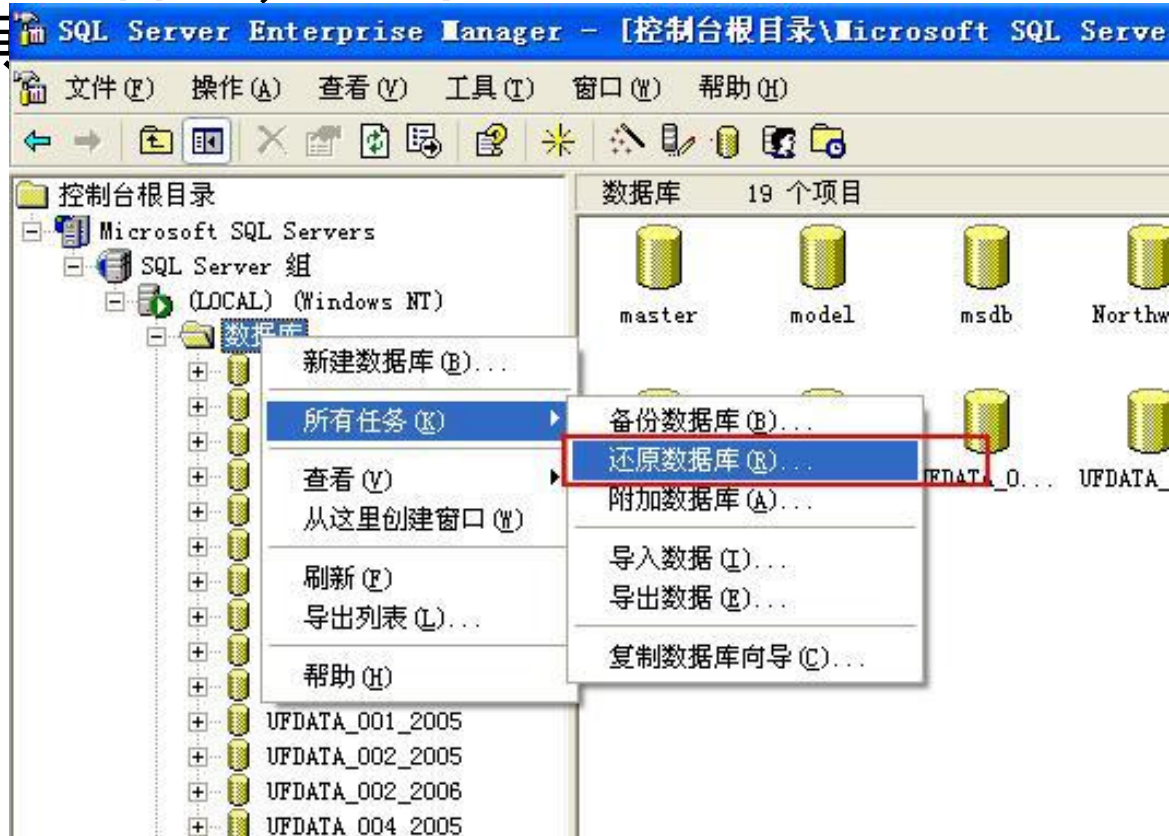
# 企业管理器—备份数据库(4)

- ❖ 录入文件名称之后，点“确定”按钮，系统会出现备份数据库的进度状况，一段时间后，系统会提示



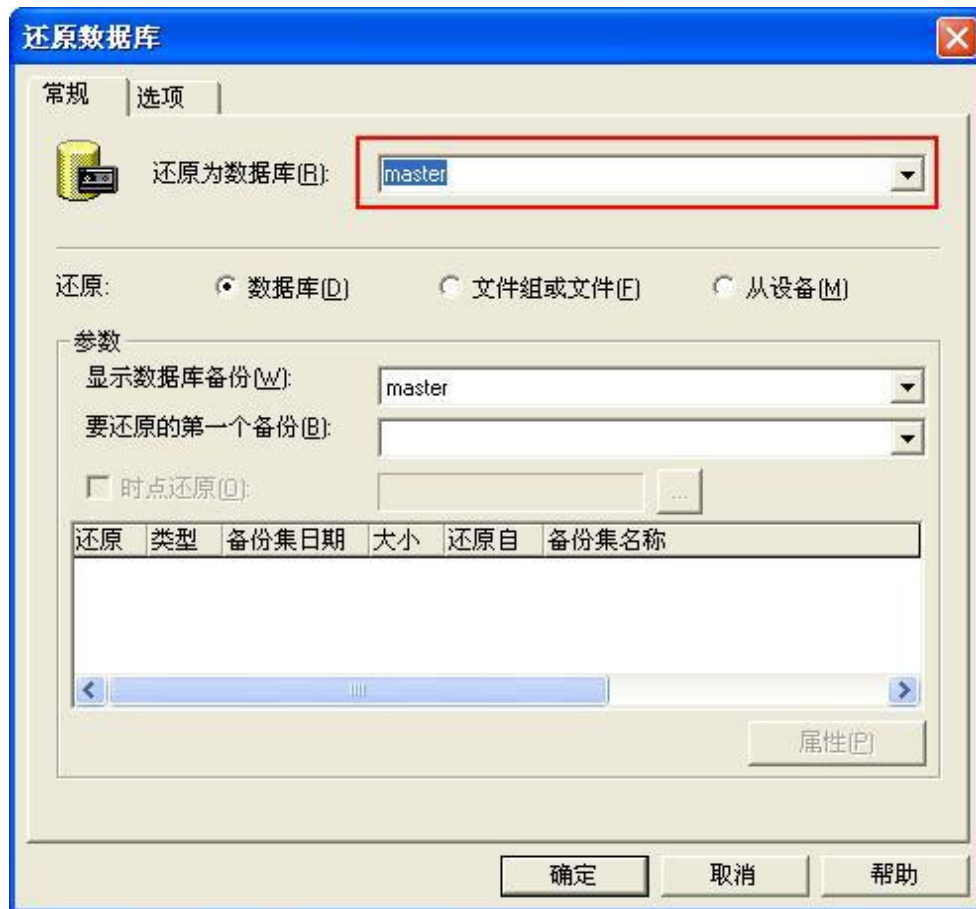
# 企业管理器—还原数据库(1)

- ❖ 打开SQL Server企业管理器，在“数据库”的位置点右键，选择快捷菜单中的“所有任务/还原”



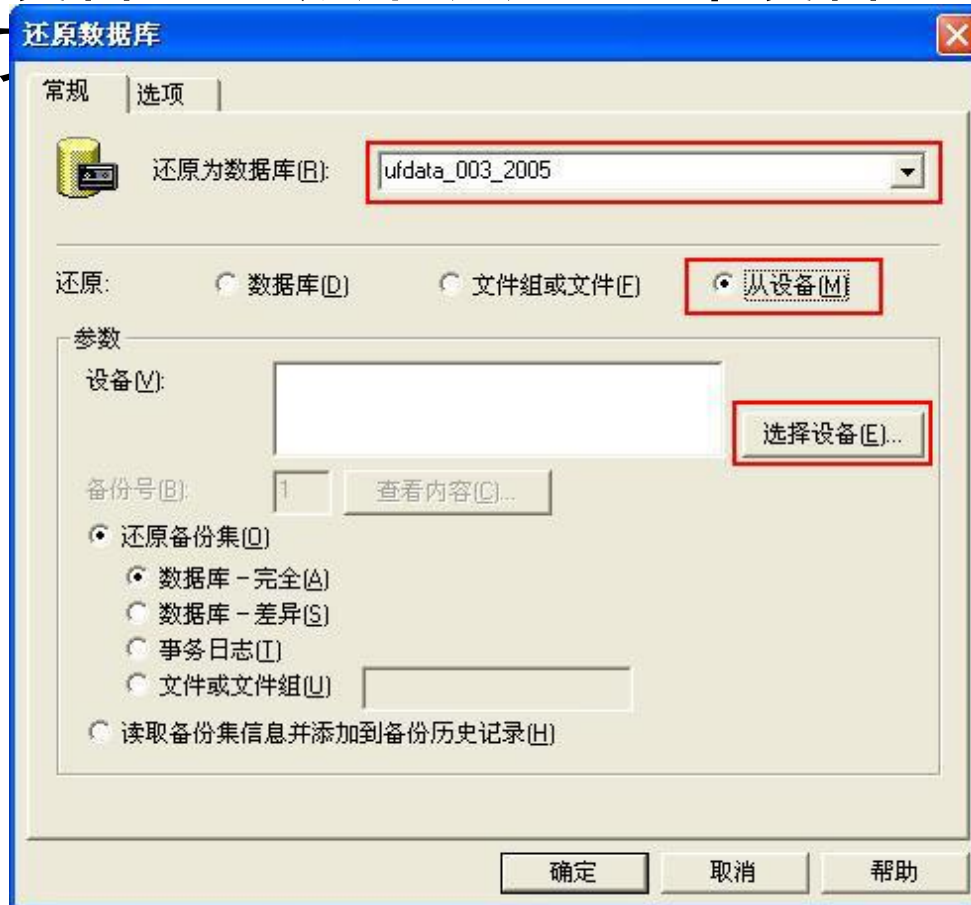
# 企业管理器—还原数据库(2)

- ❖ 录入“还原为数据库”，切记：一定要录入这个红色框里面的数据库名称，这个文件名称可以是任意名称。但是如果名称和已有数据库的名称重名，那么会给出提示，并不予还原。



# 企业管理器—还原数据库(3)

- ❖ 录入数据库还原后的名称之后，选择还原方式为“从设备”。然后点“选择设备”按钮。选择还原：



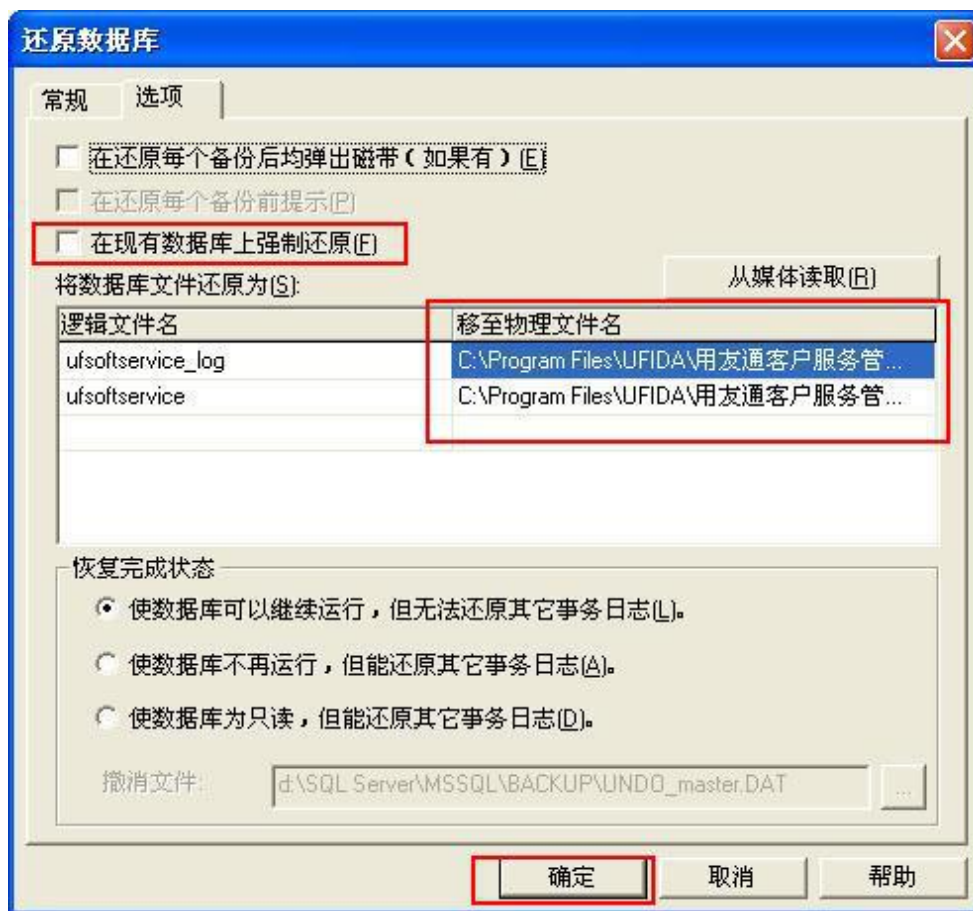
# 企业管理器—还原数据库(4)

- ❖ 点“添加”按钮，选择还原数据库的文件，然后点“确定”按钮



# 企业管理器—还原数据库(5)

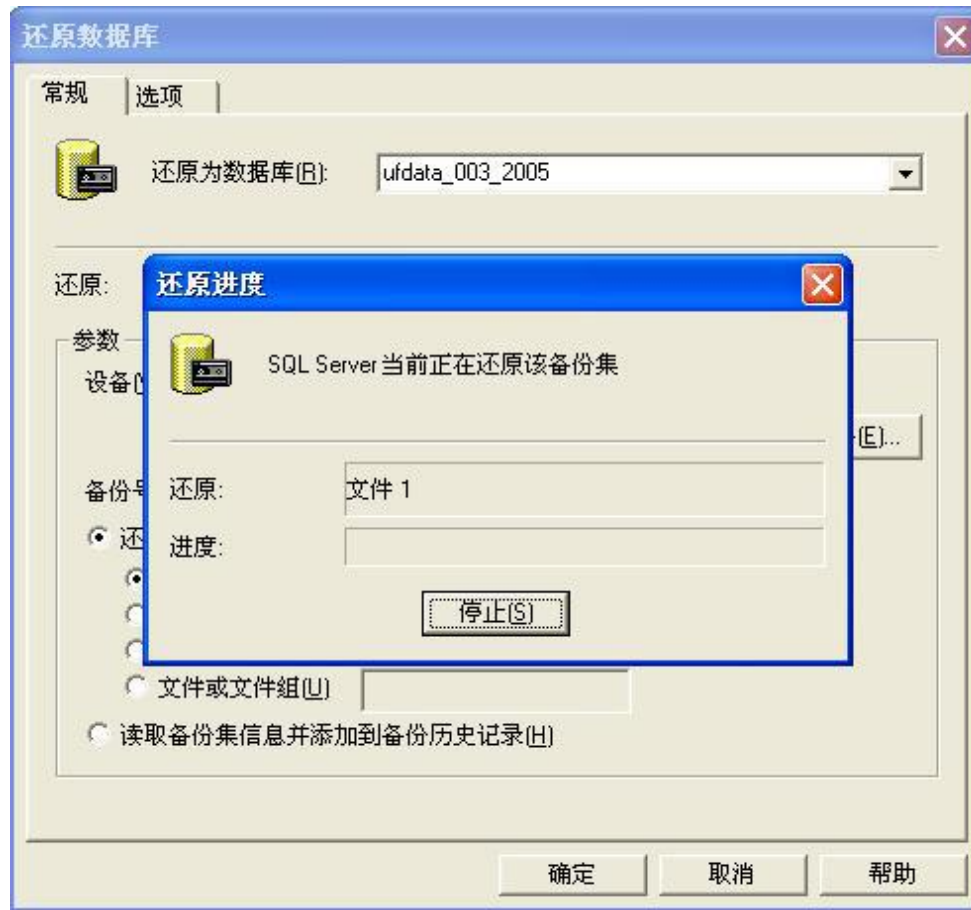
- ❖ 选择“选项”这个选项卡，会看到“在原有数据库上强制还原”的选项，如果这个选项打勾，则系统会在还原的数据库的名称与已有数据库的名称重名时，覆盖原有数据库。所以这个选项要慎用。
- ❖ 移至物理文件名的列表中显示的是还原文件的保存路径。如果在磁盘上面没有这个路径的话，系统会报错。所以这里必须选择实际存在的路径。





# 企业管理器—还原数据库(6)

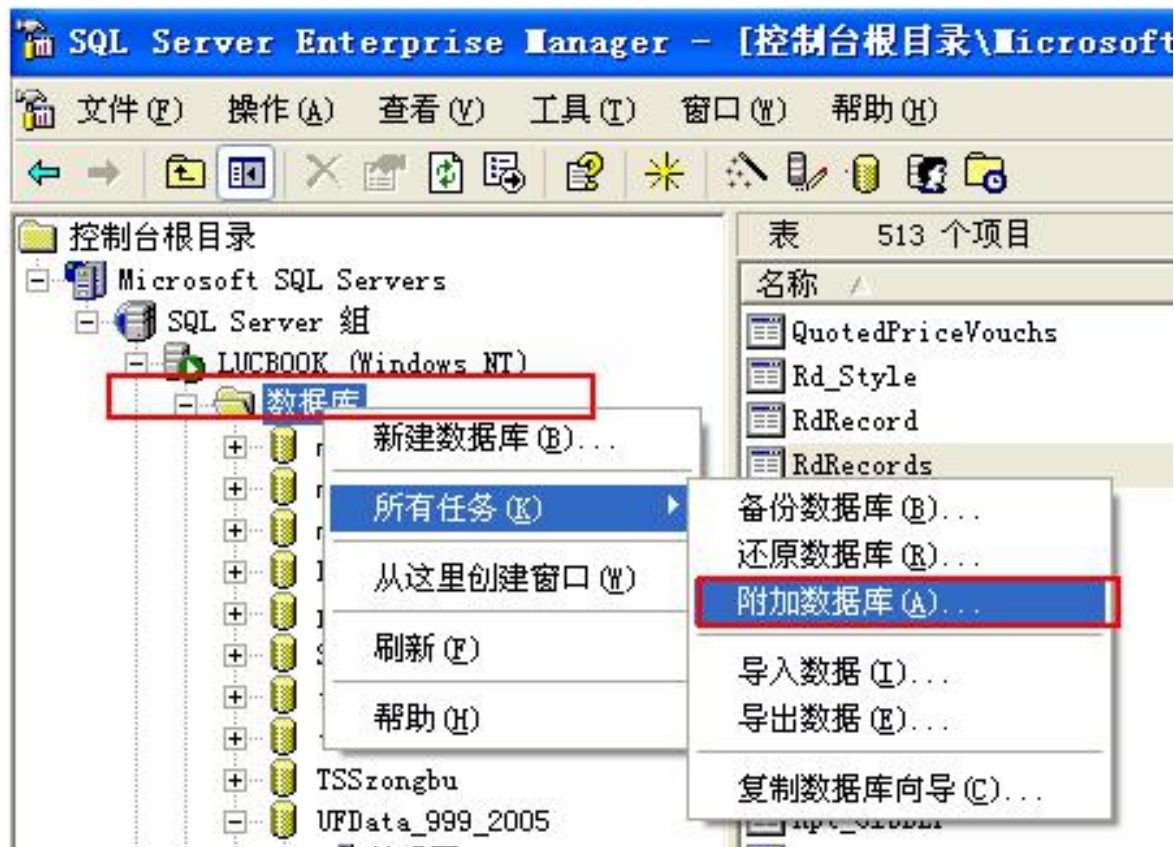
- ❖ 设置完毕后，点“确定”按钮，系统将还原数据库。并有还原进度提示。还原成功后，会有提示。





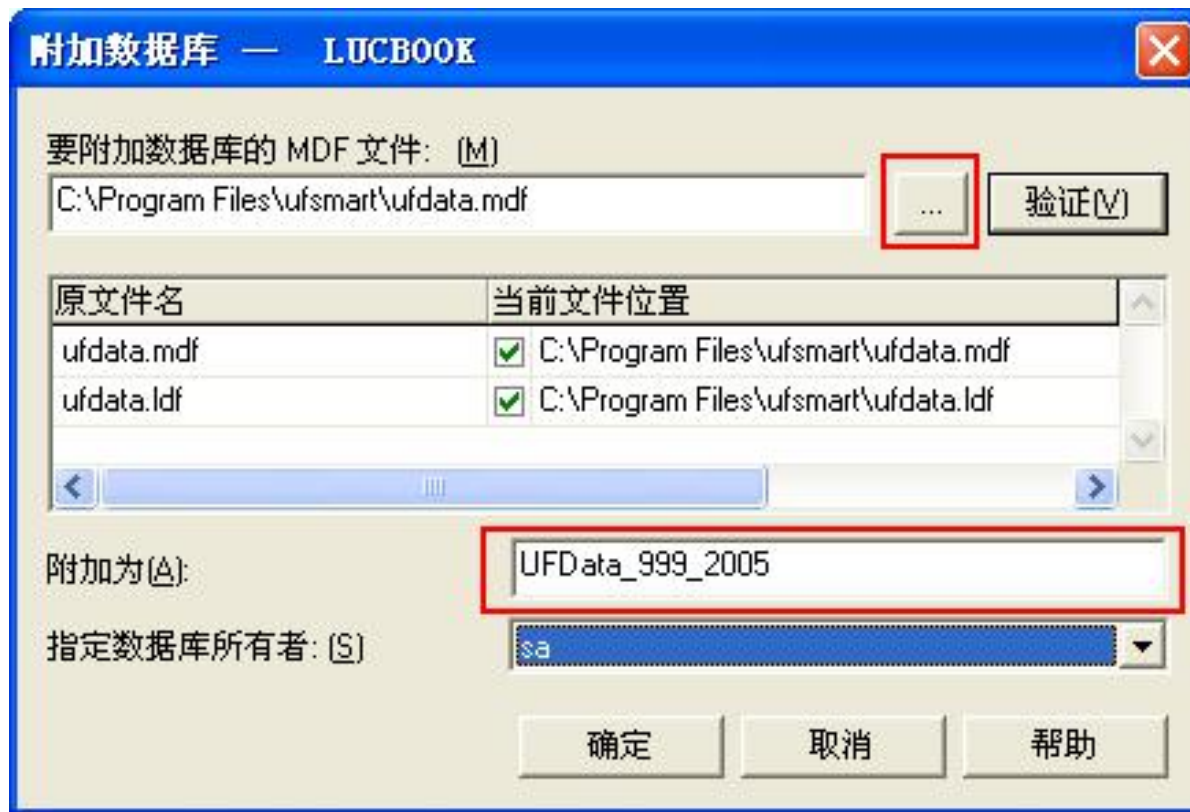
# 企业管理器—附加数据库(1)

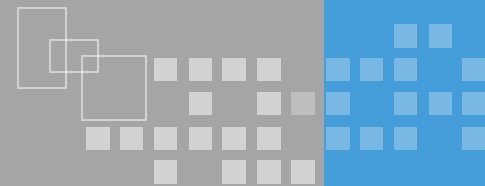
- ❖ 在“数据库”的位置点右键，选择所有任务/附加数据库



# 企业管理器—附加数据库(2)

- ❖ 选择mdf文件，录入附加为数据库名称，然后点击确定。

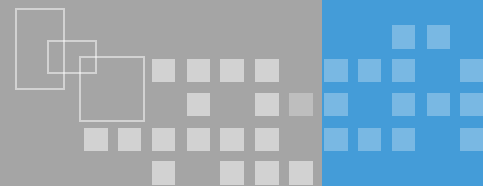




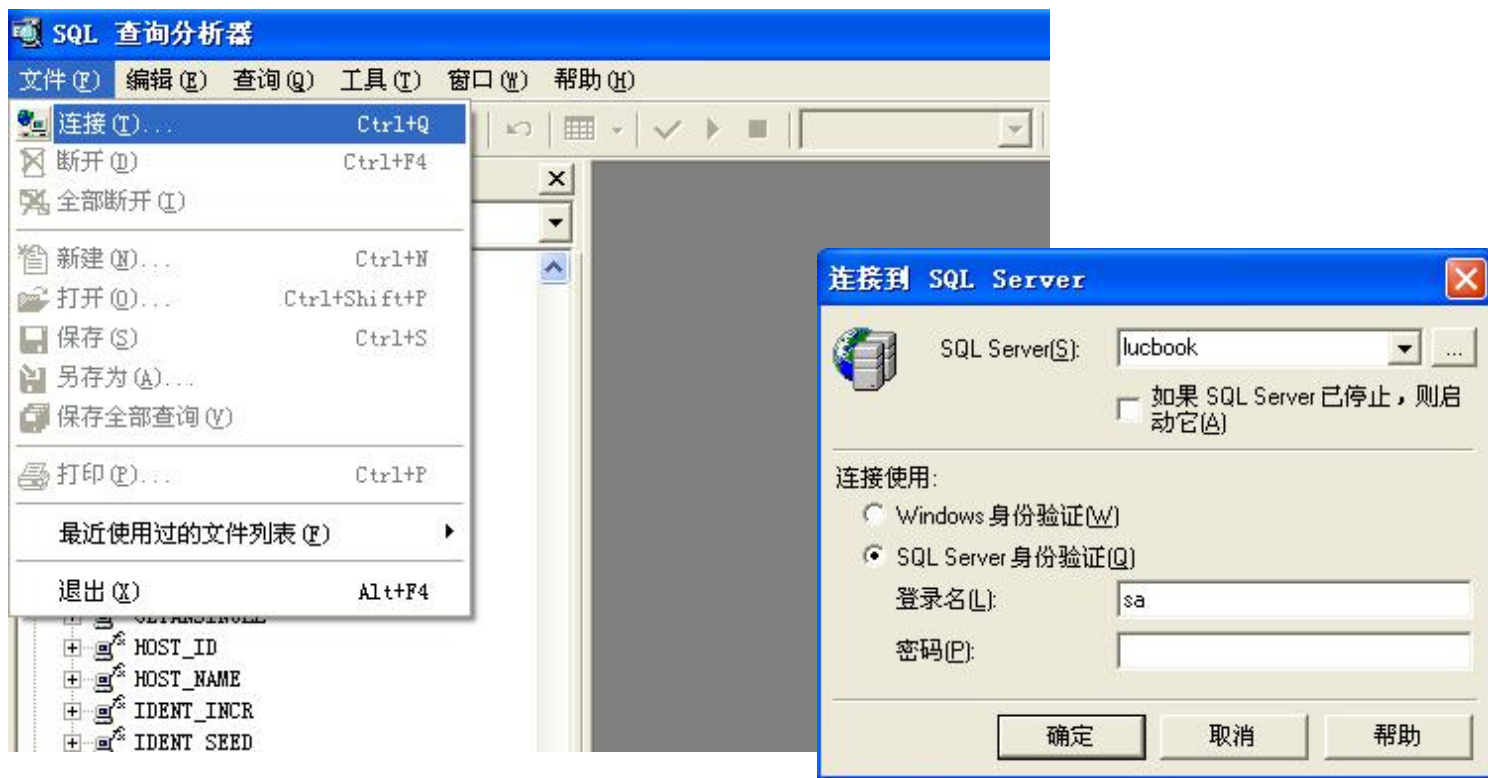
## ❖ 查询分析器

- SQL语句的执行
- 查询结果导入Excel（.csv）

# 查询分析器(1)

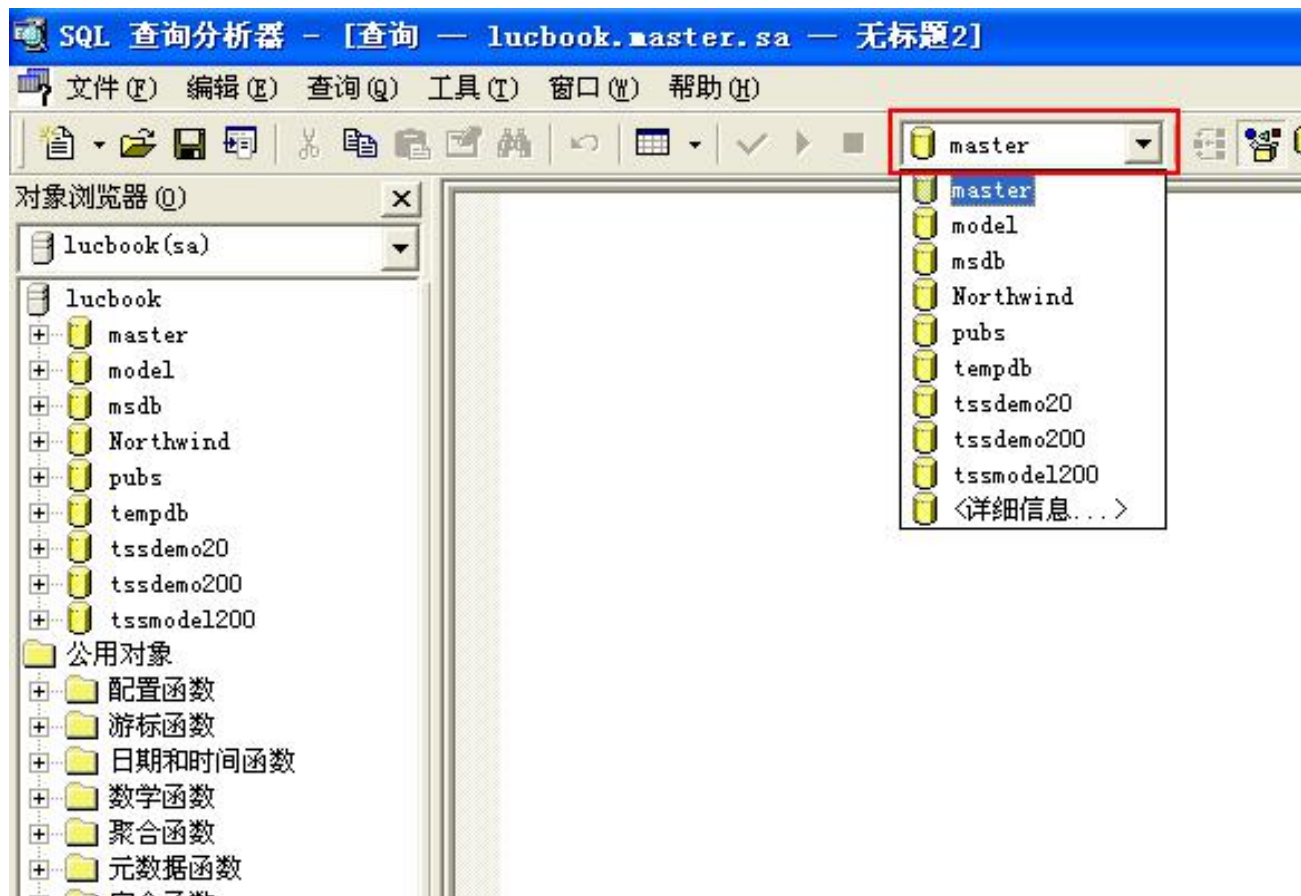


- ❖ 点击菜单文件/连接，打开“连接到SQL Server”窗口，选择SQLServer服务器，录入登录名称和密码



# 查询分析器(2)

❖ 在下拉框中选择需要操作的数据库名称



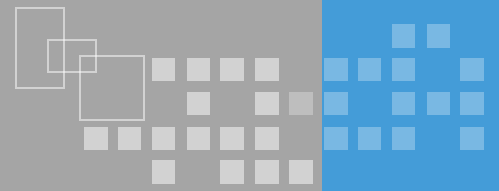
# 查询分析器(3)

- ❖ 录入SQL语句，点击“执行”按钮，显示执行结果列表。

The screenshot shows the SQL Query Analyzer interface. The title bar reads "SQL 查询分析器 - [查询 - lucbook.pubs.sa - 无标题2\*]". The menu bar includes "文件(F)", "编辑(E)", "查询(Q)", "工具(T)", "窗口(W)", and "帮助(H)". The toolbar contains various icons, with the "Execute" button (a green play icon) highlighted by a red box. The "Object Explorer" on the left shows the database structure, with "pubs" expanded to show "用户表" (User Tables), including "dbo.authors". The main query window contains the SQL statement: `select * from authors`. Below the query window, the execution results are displayed in a table grid. The status bar at the bottom indicates "批查询完成。" (Batch query completed) and shows details for the "pubs" database, including "23 行" (23 rows) and "行 1, 列 15" (Row 1, Column 15).

	au_id	au_lname	au_fname	phone	address
1	172-32-1176	White	Johnson	408 496-7223	10932 Bigge Rd.
2	213-46-8915	Green	Marjorie	415 986-7020	309 63rd St. #411
3	238-95-7766	Carson	Cheryl	415 548-7723	589 Darwin Ln.
4	267-41-2394	O'Leary	Michael	408 286-2428	22 Cleveland Av. #1
5	274-80-9391	Straight	Dean	415 834-2919	5420 College Av.
6	341-22-1782	Smith	Meander	913 843-0462	10 Mississippi Dr.
7	409-56-7008	Bernet	Abraham	415 658-9932	6223 Bateman St.
8	427-17-2319	Dull	Ann	415 836-7128	3410 Blonde St.
9	472-27-2349	Gringlesby	Burt	707 938-6445	PO Box 792
10	486-29-1786	Locksley	Charlene	415 585-4620	18 Broadway Av.
11	527-72-3246	Greene	Morningstar	615 297-2723	22 Greather House R.



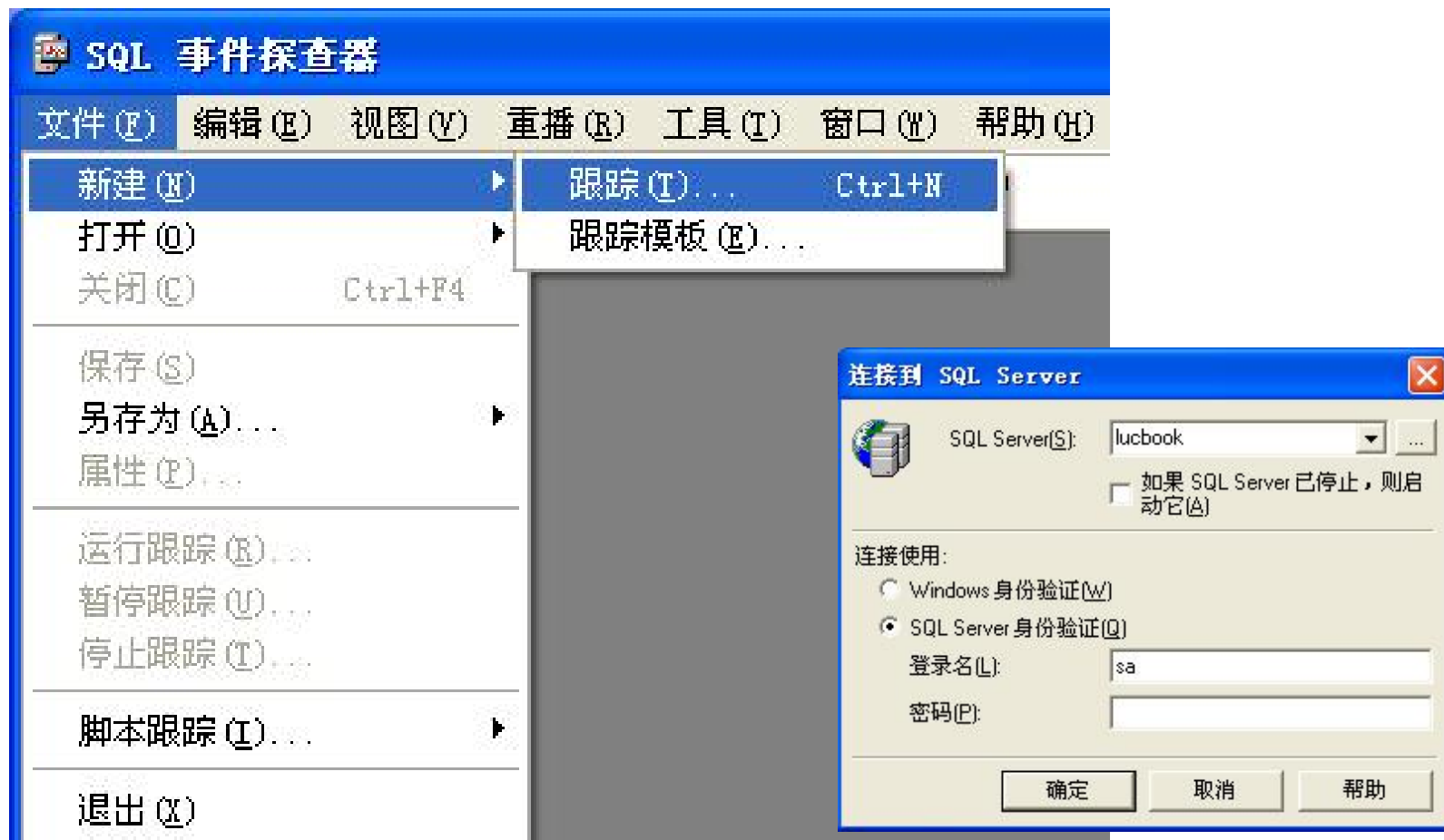


## ❖ 事件探查器

- 建立一个新的跟踪（SQL 和存储过程）
- 结合案例跟踪数据库的运行

# 事件探查器(1)

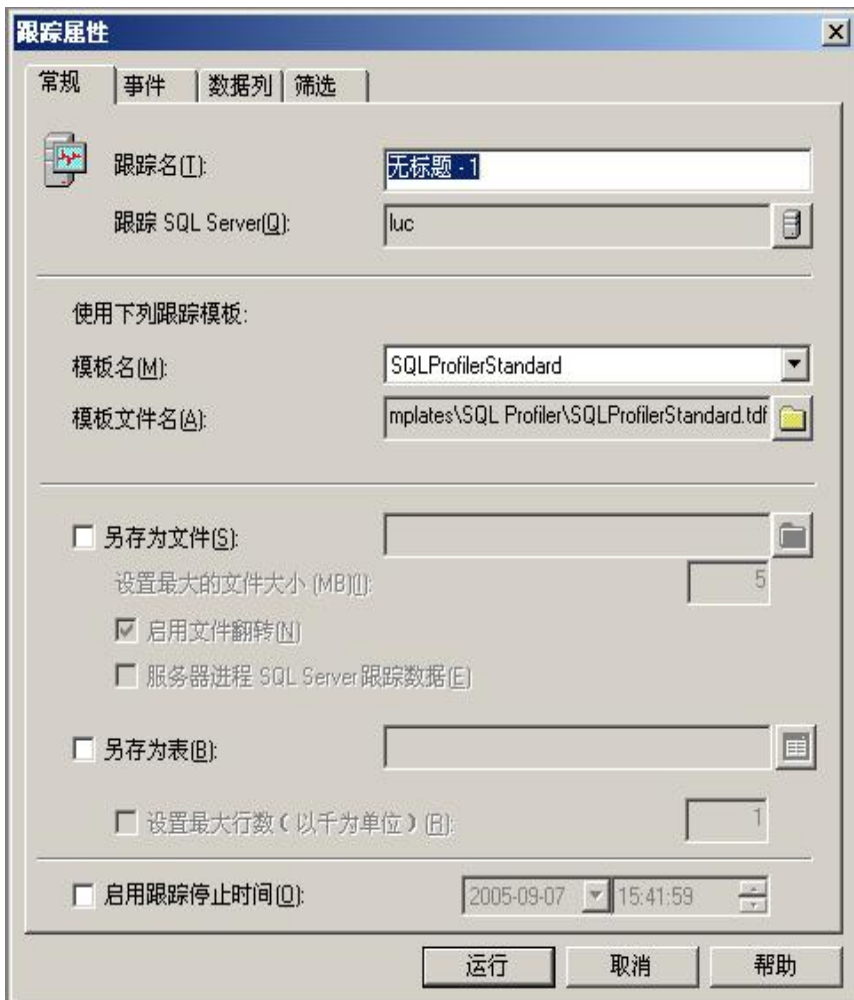
- ❖ 点击菜单文件/新建/跟踪，打开“连接到SQL Server”窗口，选择SQLServer服务器，录入登录名称和密码





# 事件探查器—建立跟踪

- ❖ 在左边窗口点击“运行”按钮，会弹出右边的窗口，选择需要跟踪的事件类（一般选择TSQL和存储过程），点击“运行”按钮开始跟踪。



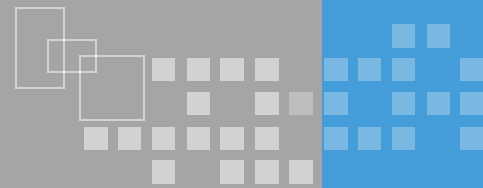
# 事件探查器

- ❖ 点击执行、暂停、停止按钮可以控制跟踪语句的起始结束时间。跟踪到的语句全部显示在列表中。可以复制语句到查询分析器中执行

EventClass	TextData	ApplicationName	NTUserName	LoginName	CPU	Reads	Writes
TraceStart							
RPC:Completed	exec sp_reset_connection	.Net SqlCli...	Admini...	sa	0	0	0
SQL:BatchCompleted	SELECT name, dbid From sysdatabases...	.Net SqlCli...	Admini...	sa	0	30	0
SQL:BatchCompleted	SELECT * FROM sysobjects WHERE id =...	.Net SqlCli...	Admini...	sa	20	357	0
SQL:BatchCompleted	SELECT cDBVersion, dDate FROM sysInf...	.Net SqlCli...	Admini...	sa	0	29	0
SQL:BatchCompleted	SELECT cid FROM pl_Person1 WHERE cN...	.Net SqlCli...	Admini...	sa	0	61	0

SELECT cid FROM pl\_Person1 WHERE cName = 'system' AND cPWD = 'ogvu{u' AND cState = '0' AND isAdmin = '1'

跟踪正在运行 行 6, 列 1 行: 连接



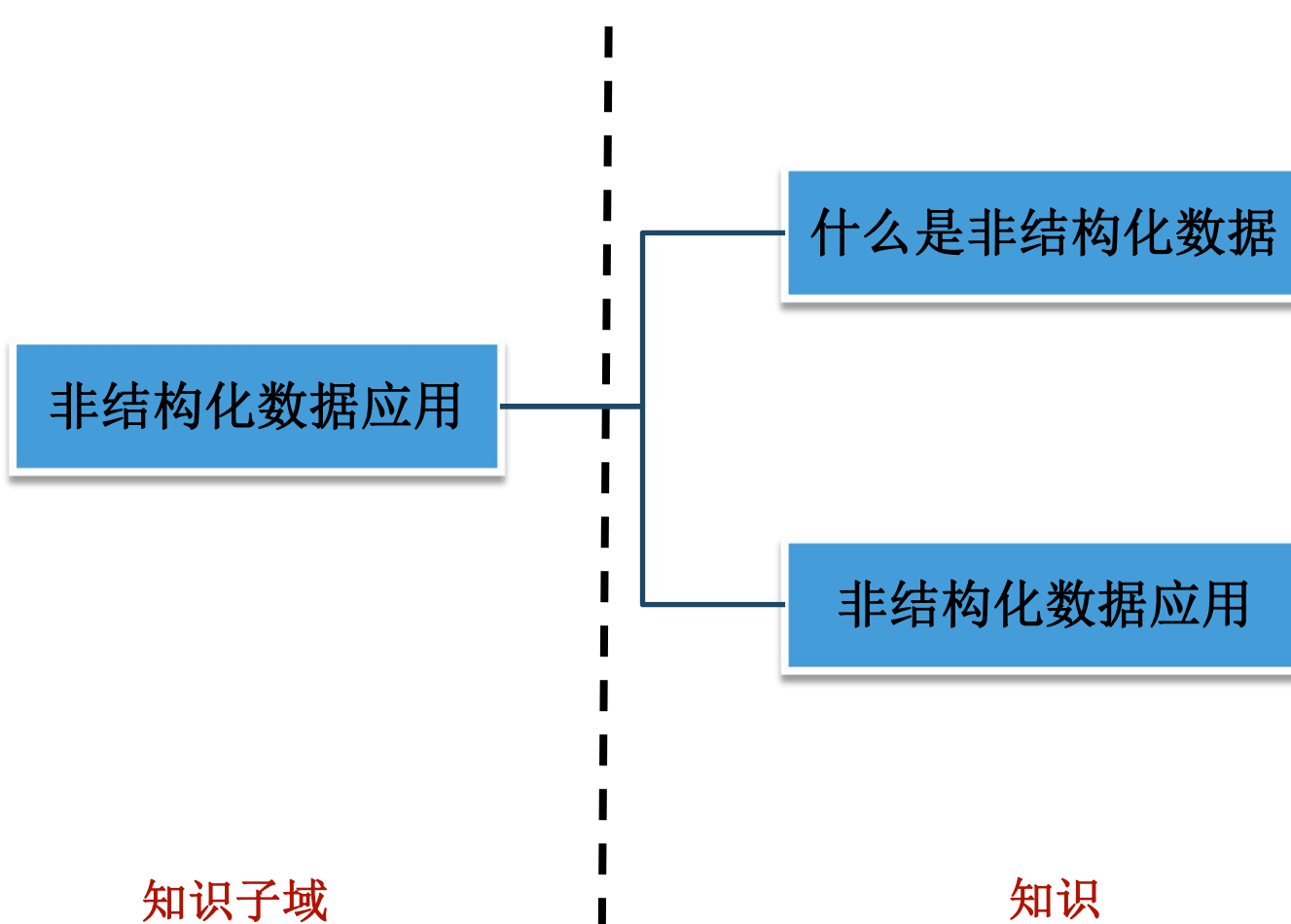
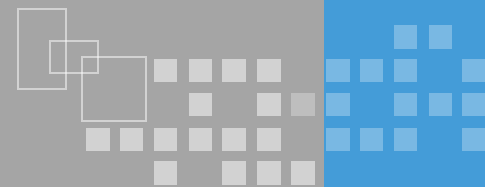
## 如航空售票系统，包括的数据项：

- 系统工作：座位预定信息：座位分配、座位确认等；
- 航班信息：航班号、飞机型号、机组号、起飞地、目的地、起飞时间、到达时间等；
- 机票信息：票价、折扣、有无等。

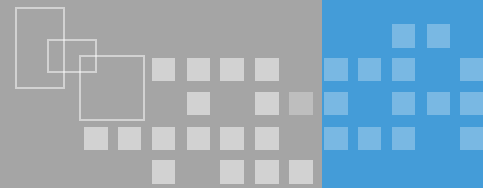
## 系统工作：

- 查询：某一时间内从某个特定地到另一指定地的航班，是否有可以选择的座位、飞机型号、票价、折扣等。
- 更新：登记航班、分配座位等；
- 统计：统计经常乘坐某一航班的乘客信息等。

# 知识子域：非结构化数据应用

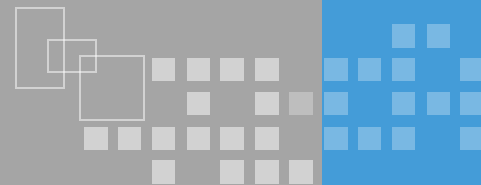


# 什么是非结构化数据

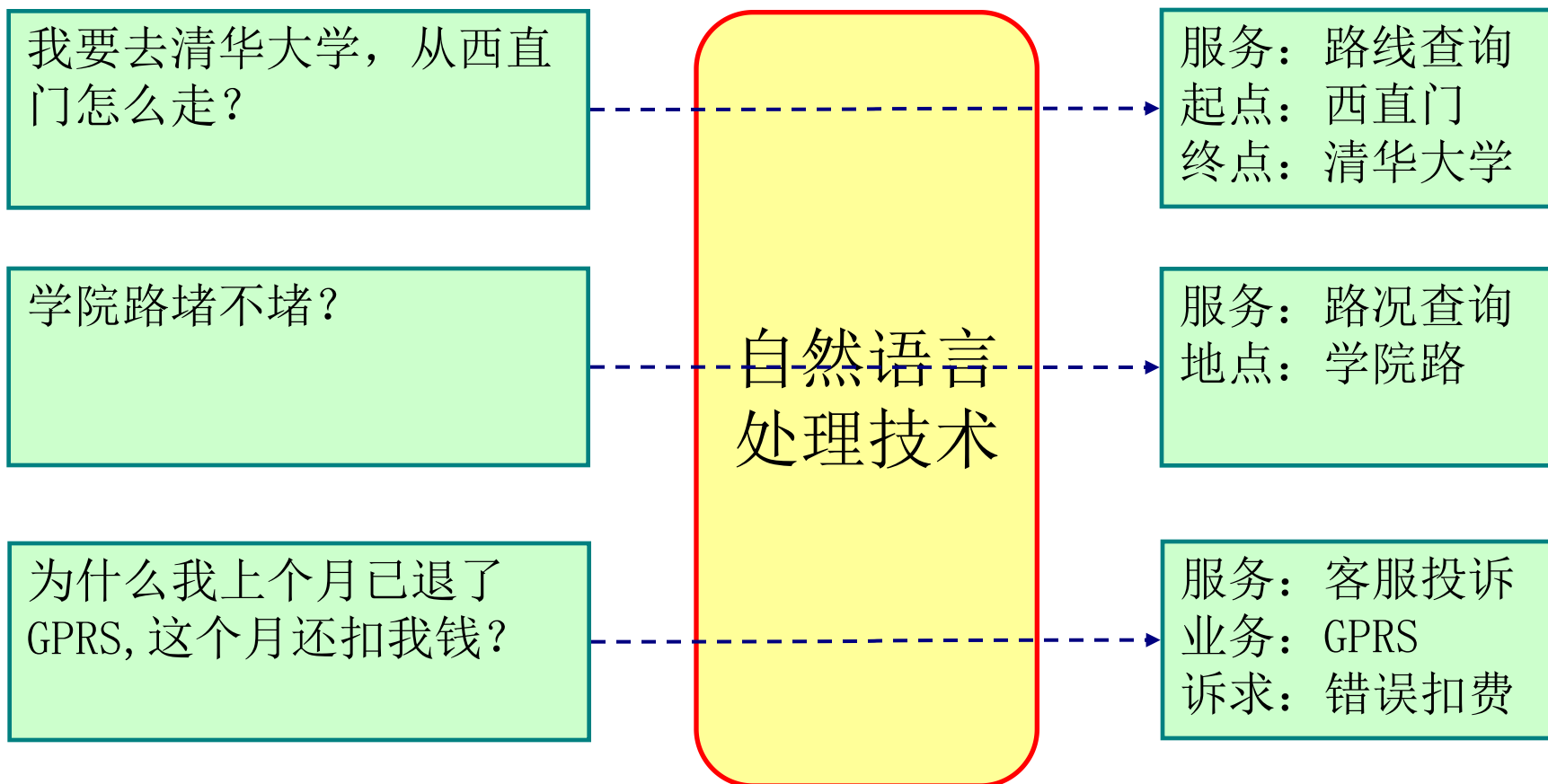


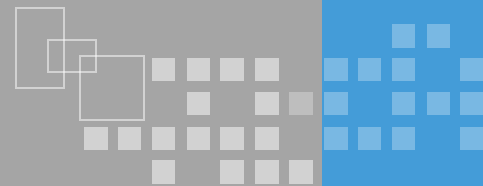
非结构化数据：不方便用数据库二维逻辑表来变现的数据，包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像和音频/视频信息等。

# 非结构化数据处理的技术

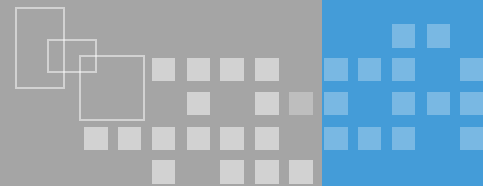


- 非结构化数据处理的主要技术是自然语言处理技术。用来对非结构化数据进行各种层次的理解。





- 什么是**自然语言**
  - 语言是人类交际的工具，是人类思维的载体
  - 人造语言：编程语言，包括C++，BASIC等
  - 世界语
  - 自然语言：
    - 形式：口语、书面语、手语
    - 语种：汉语、英语、日语、法语…
- **语言学**是研究语言规律的科学



“昨晚，我的JJ带着他的青蛙BF到我家来吃饭。在饭桌上，JJ的BF一个劲儿地对我妈妈PMP，说她年轻的时候一定是个漂亮MM，那酱紫真是好BT，7456……”

JJ：姐姐

酱紫：这样子

BF：boy friend

青蛙 B F：长相不好的男朋友

PMP:拍马屁

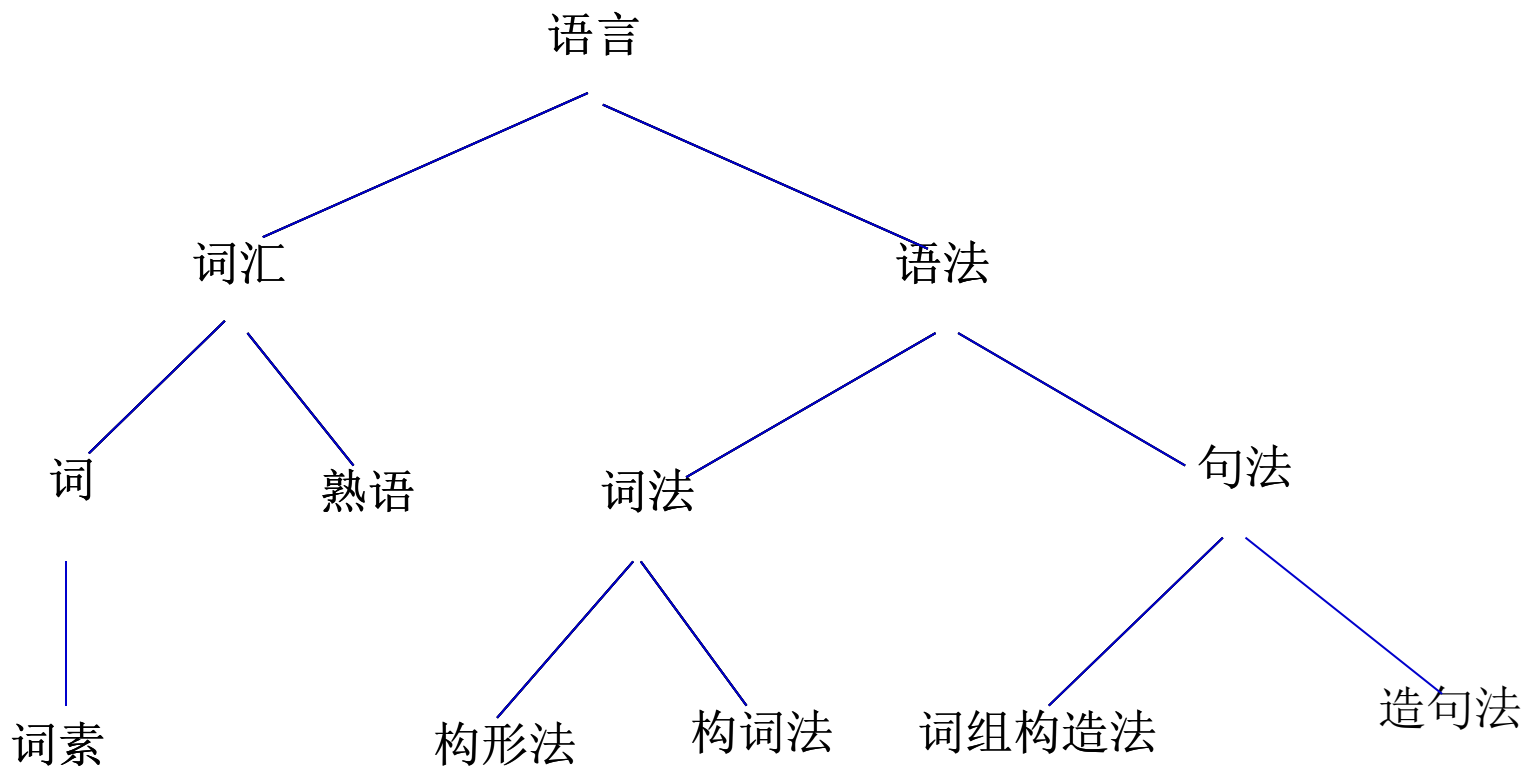
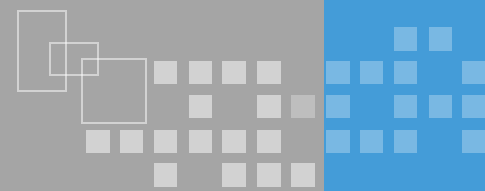
MM：妹妹

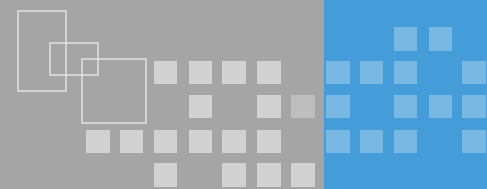
BT：变态

7456：气死我了

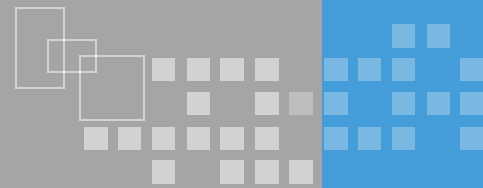


# 语言的构成



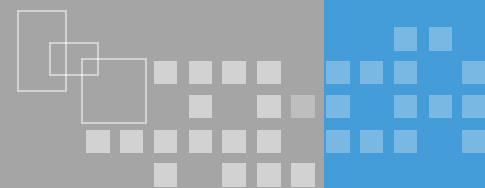


- **语音分析**：从语音流中区分出一个一个声音单元——音素
- **词法分析**：从句子中切分出单词、找出词汇的各个词素，确定单词的词性、词义等。
- **句法分析**：对句子和短语的结构进行分析，找出词、短语等的相互关系及在句子中的作用等。
- **语义分析**：识别一句话所表达的实际意义。
- **语用分析**：研究语言所在的外界环境对语言使用所产生的影响。

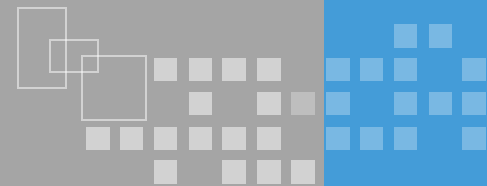


- 同一词语在不同的“语境”中具有不同“语义”
  - 例如：中国奥运史上十大女杰的精彩“转身”
- 病毒
  - 计算机领域：计算机病毒
  - 医学领域：生物学病毒

# 自然语言处理的概念



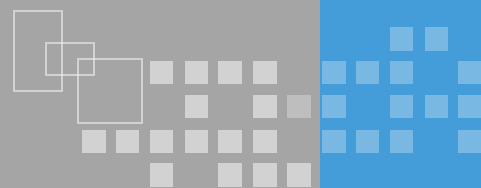
- **自然语言处理** ( Natural Language Processing, NLP )
  - 也称**自然语言理解**或**计算语言学**；
  - 主要研究如何让机器进行自然语言信息处理，即人类语言活动中，信息成分的**发现、提取、存储、加工与传输**。
  - NLP是计算机科学、语言学、人工智能与数学等学科的交叉学科和边缘学科。
- **计算语言学**是从计算角度处理语言
  - 将人们对语言的结构规律的认识用精确的、形式化的、可计算的方式（计算模型）加以表示。



- 自然语言具有多样性（不同语种、不同地域、不同人群）
- 自然语言具有进化性
- 自然语言的模糊性
- 自然语言的歧义性

处理歧义问题是NLP的**核心问题**。自然语言处理过程就是各种歧义现象的消解过程。

# 自然语言理解的困难



## 食堂卫生检查报告

检查时间：2010-4-30

参加人员：校医院：邹晓丽 闫梅青 南食堂：曹天明 北食堂：苗经理

南食堂卫生状况良好，进步较大，主要表现在：

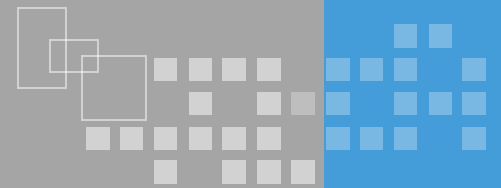
- 1、地面清洁无积水。
- 2、销售间多数人员着装标准戴口罩。
- 3、生熟食品按器皿标识盛放。

北食堂卫生状况一般，需要改进。

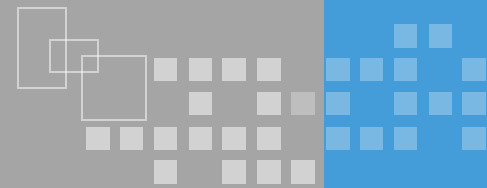
友情提示：天气渐热，又到了肠道疾病高发季节。师生就餐前要洗手，洗手后不再触摸钱币等物品；注意擦拭就餐卡；高温加工过的食品更为安全可靠。

校 医 院

# 机器能够理解人的语言吗？



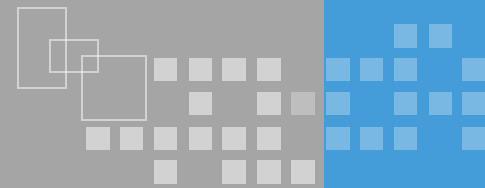
- **很难！**
- 什么是理解？
  - 结构主义：机器的理解机制与人相同（白盒）
    - 问题：人类语言理解机理尚未清楚
  - 功能主义：机器的表现与人相同即可（黑盒）
    - 图灵测试
      - 如果通过自然语言的问答，一个人无法识别和他对话的是人还是机器，那么就应该承认机器具有智能



- 给计算机输入一段自然语言文本，如果计算机能
  - 问答(question-answering) — 机器能正确地回答输入文本中的有关问题；
  - 文摘生成(summarizing) — 机器有能力产生输入文本的摘要；
  - 复述(paraphrase) — 机器用不同的词语和语句复述输入文本；
  - 翻译(translation) — 机器把一种语言(源语言)翻译为另一种语言(目标语言)

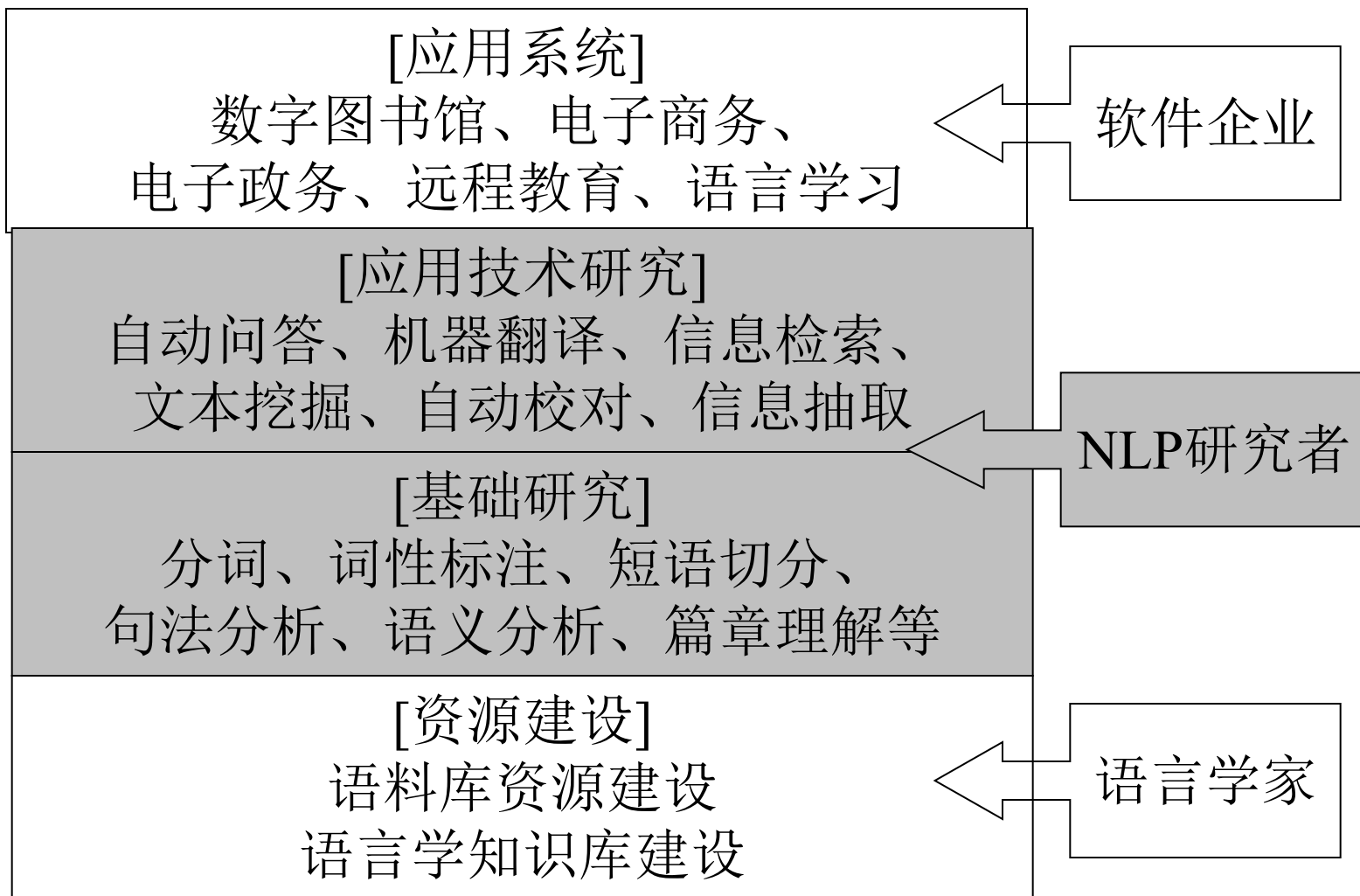
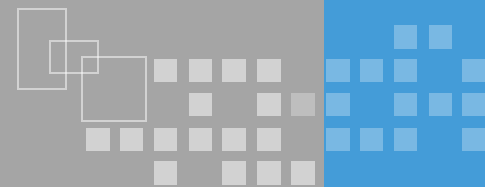


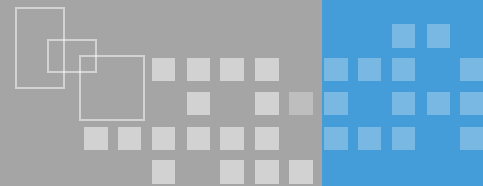
# 自然语言处理的研究目标



- **弱人工智能目标**：建立一个足够精确的语言数学模型使计算机通过编程来完成自然语言的相关任务。如：听、读、写、说，释义，翻译，回答问题等；
- **强人工智能目标**：让用户能通过自然语言与计算机自由对话；

# 自然语言处理的研究内容



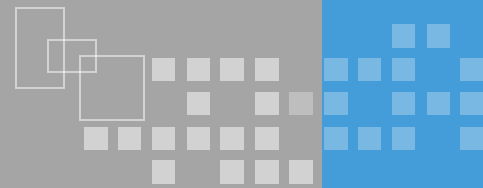


- NLP应用前景

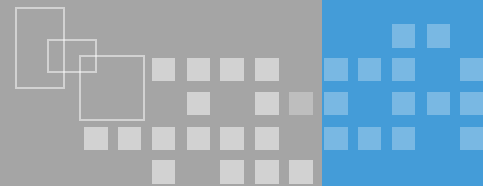
- 据统计，日常工作中80%的信息来源于语言，处理文本的需求在不断增长
- 文本是人类知识最大的存储源，并且文本的数量在不停地增长
  - 电子邮件、新闻、网页、科技论文、 用户抱怨信

- NLP典型应用

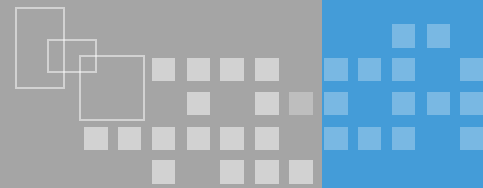
- 智能搜索引擎、自动问答、信息获取、语义网
- 语音识别，文字识别、输入法
- 机器翻译，自动文摘，跨语言检索
- 文本分类、文本聚类、文本分析（结构、内容、情感）、文本挖掘（主题跟踪：人物跟踪，企业跟踪）、文本过滤



- **智能接口**
  - **功能：**
    - 把现实世界中的信息送入电子世界
  - **主要成果**
    - 拼音输入、手写输入、语音合成、语音输入，手机输入
- **知识处理**
  - **功能：**
    - 对于已进入电子世界中的信息进行加工处理获得知识
  - **知识经济的时代已经到来！**
  - **知识就是力量——知识就是财富**
  - **百度爱问、百度文库、新浪爱问、VC知识库**

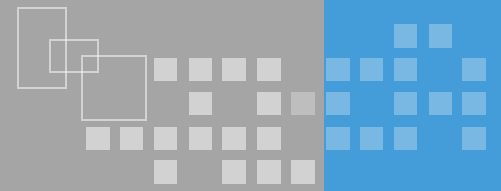


- 词法分析
- 句法分析
- 语义分析

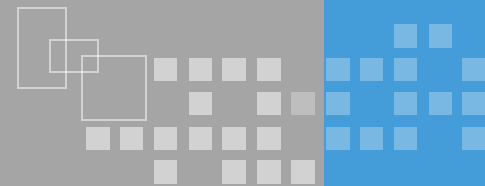


- 词法分析目的是从句子中切分出单词，找出词汇的各个词素，从中获得单词的语言学信息并确定单词的词性
  - 词性：名词、动词、形容词、介词等
  - 词的构成：动宾，动补，偏正，主谓  
如：开学，生病，加深，认清，原油，火热，头痛，人造
- 自动分词：汉语处理的难题之一
  - 用程序从句子中切分出单词

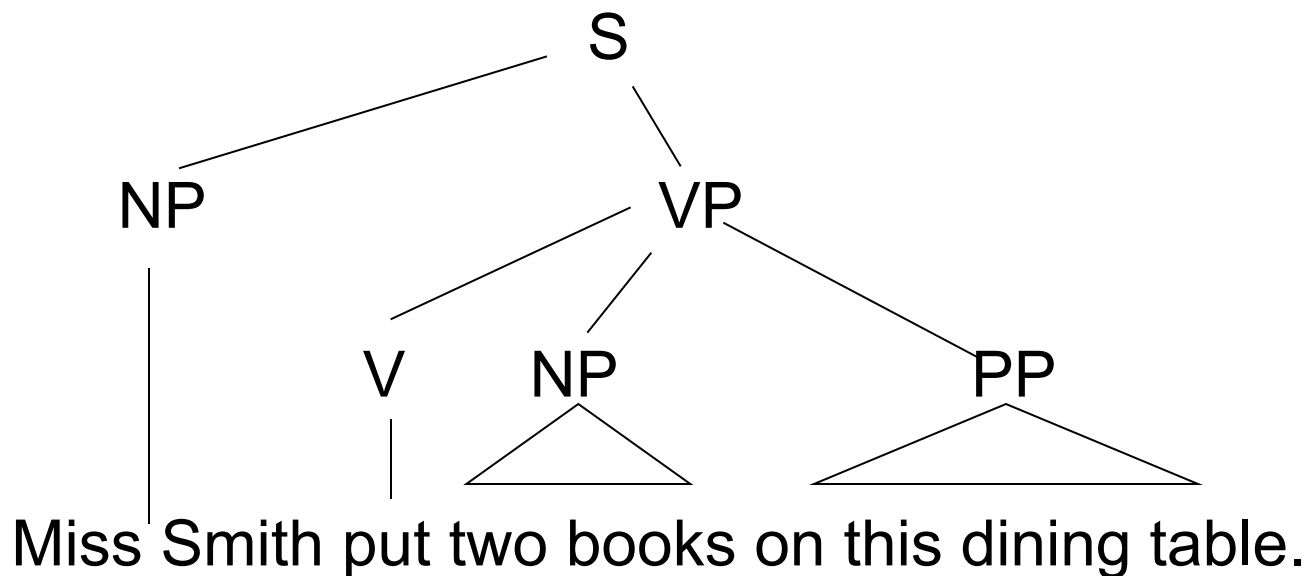
# 词法分析的分词歧义



- 例：南京市长江大桥  
    南京|市长|江大桥  
    南京市|长江|大桥
- 例：我们研究所有东西
  - 我们--研究所--有--东西 （交叉歧义）
  - 我们--研究--所有一东西  
    把手放在桌上
  - 把--手--放在--桌上 （组合歧义）
  - 把手--放在--桌上



- 目的：分析句子结构，找出词, 短语在句中的相互关系以及各自的作用，并用层次结构来加以表示



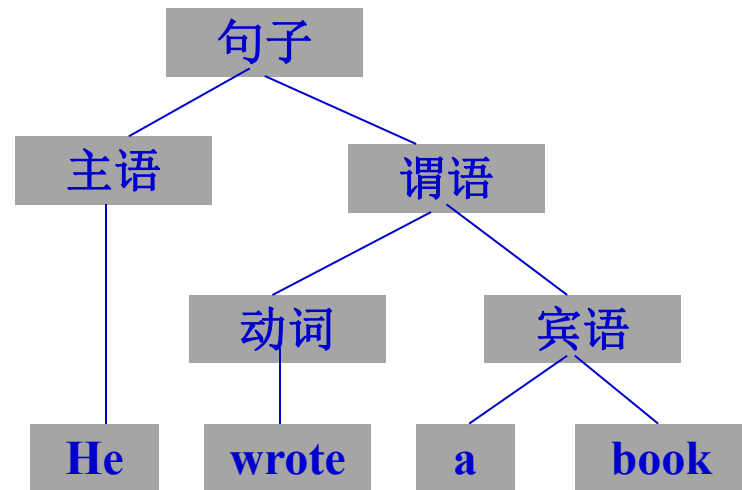


# 句法分析过程

一个句子是由各种不同的句子成分组成的。这些成分可以是单词、词组或从句。句子成分还可以按其作用分为主语、谓语、宾语、宾语补语、定语、状语、表语等。这种关系可用一棵树来表示，如对句子：

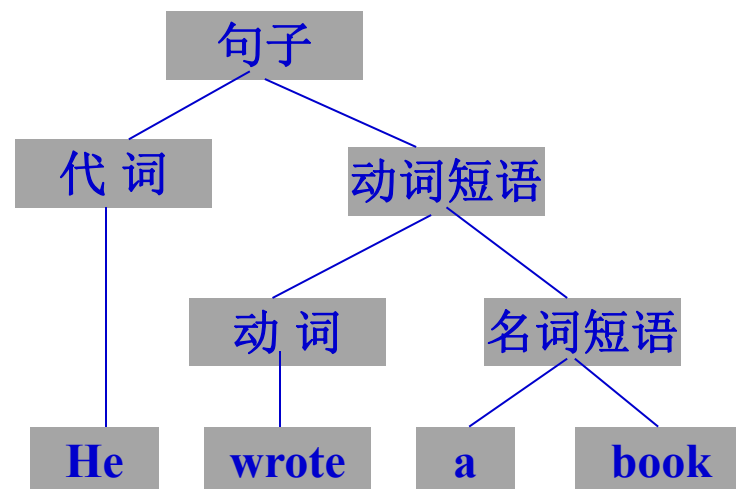
**He wrote a book.**

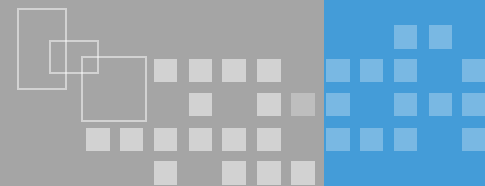
可用图示的树形结构来表示。



# 句法分析过程

一个句子又是由若干  
个词类构成的，如名词、  
动词、代词、形容词等。  
若从句子的词类来考虑，  
一个句子也可用一棵树  
来表示，这种树称为句  
子的分析树，如图所示：





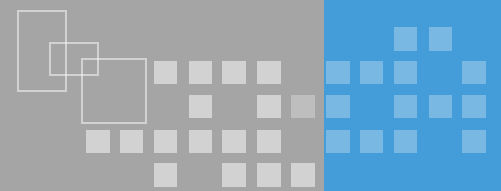
- 语义分析就是要识别一句话所表达的实际意义。即弄清楚“干什么了”，“谁干的”，“这个行为的原因和结果是什么”以及“这个行为发生的时间、地点及其所用的工具或方法”等。
- 语义分析仅是在句法范围内根据词性信息来分析自然语言中句子的文法结构的，由于它没有考虑句子本身的含义，也就不能排除像

The paper received the professor.

这种在语法结构上正确，但实际意义上错误的句子

。

# 语义分析过程



- **Agent**（施事）：指行为的施动者；
- **Object**（受事）：指行为作用的对象；
- **Co-Agent**（共施事）：指行为施动者的合作者；
- **Instrument**（工具）：指施事者或共施事者实现行为中所使用的对象；
- **Time**（时间）：指行为发生的时间；
- **Source**（来源）：指行为作用对象移出的位置；
- **Goal**（目标）：指行为作用对象到达的位置；
- **Trajectory**（轨迹）：指从来源到目标所经过的路径。

John gave the book to Sally

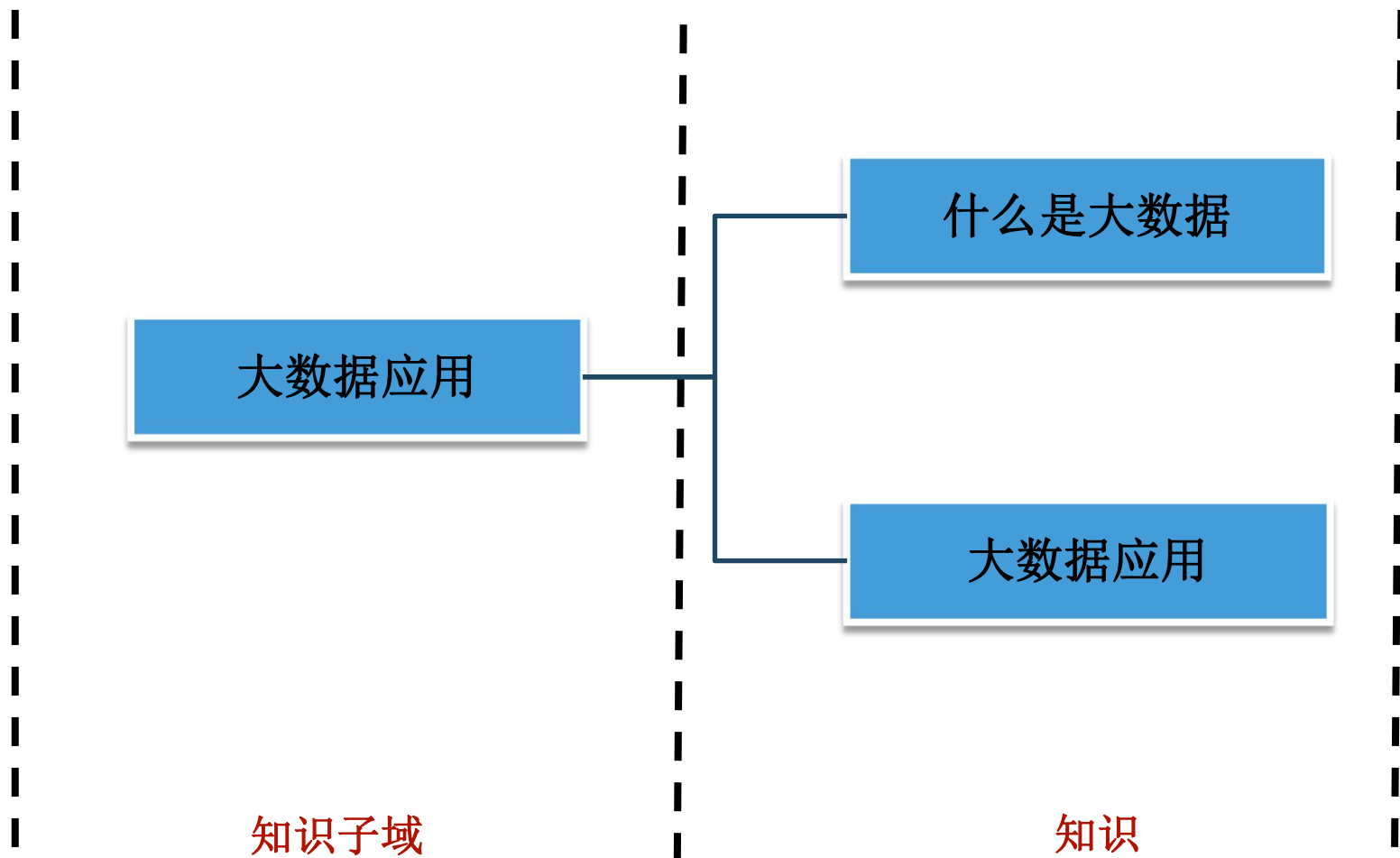
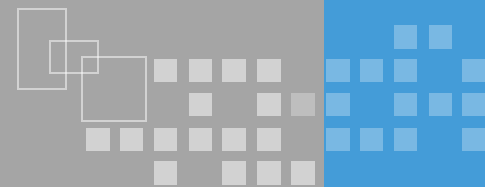


[GAVE

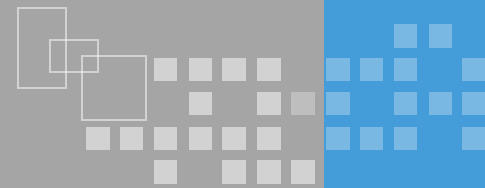
Agent: John  
Object: the book  
Source: John  
Goal: Sally

]

# 知识子域：大数据应用



# 什么是大数据



何为大？—数据度量

1Byte = 8 Bit

1KB = 1,024 Bytes

1MB = 1,024 KB = 1,048,576 Bytes

1GB = 1,024 MB = 1,048,576 KB = 1,073,741,824 Bytes

1TB = 1,024 GB = 1,048,576 MB = 1,099,511,627,776 Bytes

1PB = 1,024 TB = 1,048,576 GB = 1,125,899,906,842,624 Bytes

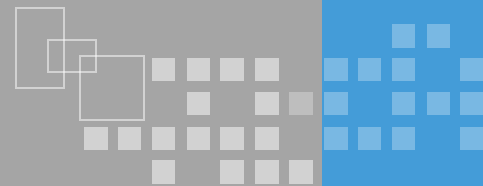
1EB = 1,024 PB = 1,048,576 TB = 1,152,921,504,606,846,976 Bytes

1ZB = 1,024 EB = 1,180,591,620,717,411,303,424 Bytes

1YB = 1,024 ZB = 1,208,925,819,614,629,174,706,176 Bytes

**数据没有办法在可容忍的时间下使用常规软件方法完成存储、管理和处理任务**

# 什么是大数据



## 大数据定义：

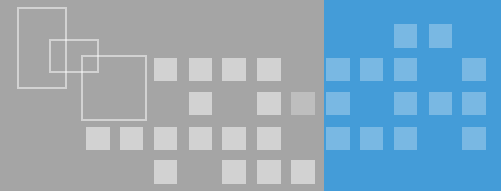
所涉及的资料量规模巨大到无法通过目前主流软件工具，在合理时间内达到撷取、管理、处理、并整理成为帮助企业经营决策更积极目的的资讯。

## 主要的特点：

- Volume 数据体量
  - Velocity 数据速度
  - Variety 数据多样
  - Veracity 数据真实
- 4V特征



# 什么是大数据



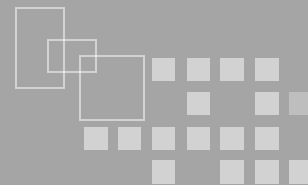
大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

大数据就是“未来的新石油”。





# 大数据的特征



## Volume

### 数据量体

- 非结构化数据的超大规模和增长
- 总数据量的80~90%
- 比结构化数据增长快10倍到50倍
- 是传统数据仓库的10倍到50倍

## Big Data 大数据



## Variety

### 数据多样

- 大数据的异构和多样性
- 很多不同形式（文本、图像、视频、机器数据）
- 无模式或者模式不明显
- 不连贯的语法或句义

## Value

### 数据价值

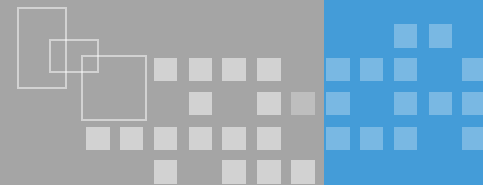
- 大量的不相关信息
- 对未来趋势与模式的可预测分析
- 深度复杂分析（机器学习、人工智能Vs传统商务智能）

## Velocity

### 数据速度

- 实时分析而非批量式分析
- 数据输入、处理与丢弃
- 立竿见影而非事后见效

# 大数据的4V特征—Volume



一般情况下，  
大数据是以PB、  
EB、ZB为单  
位进行计量的

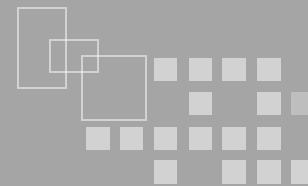
1PB相当于50%的全美学术研究图书馆藏书  
信息内容

5EB相当于至今全世界人类所讲过的话语

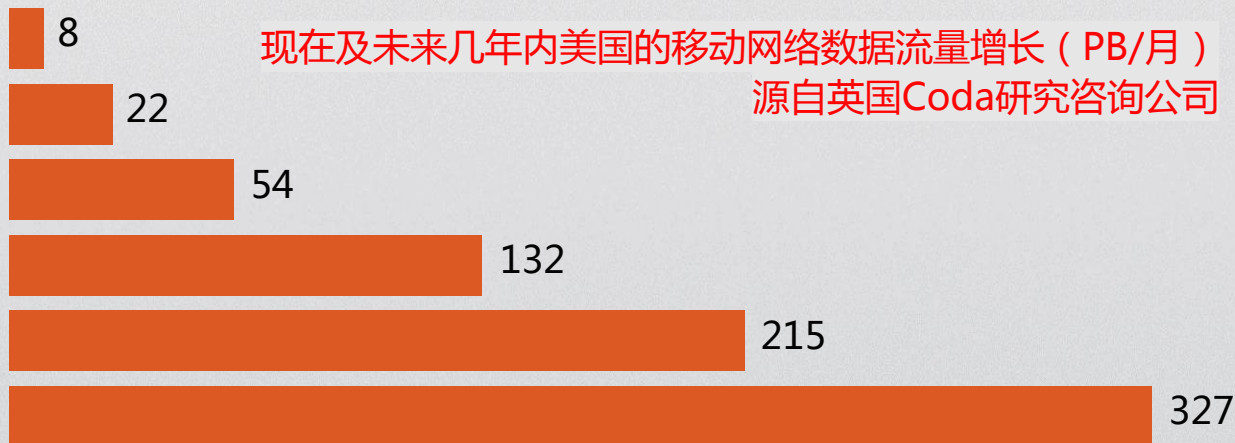
1ZB如同全世界海滩上的沙子数量总和

1YB相当于7000位人类体内的微细胞总和

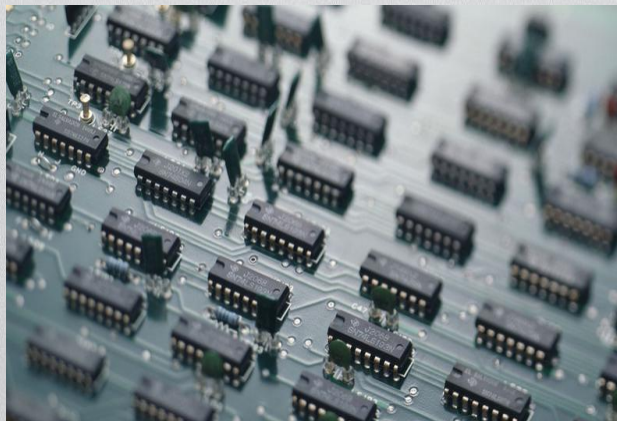
# 大数据的4V特征—Velocity



大数据的增长速度快



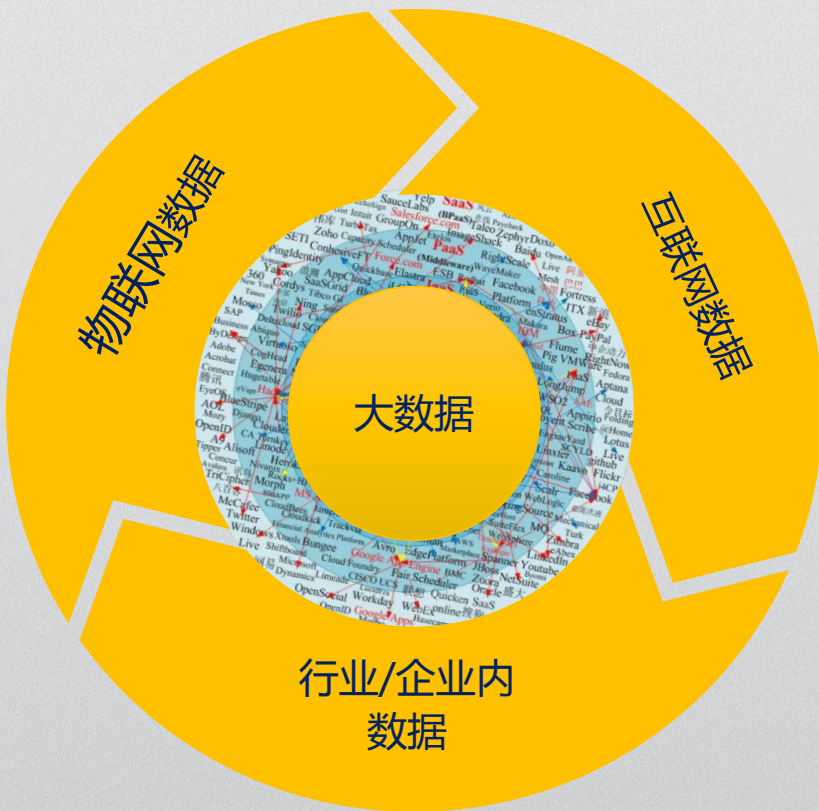
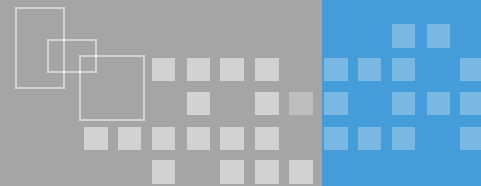
大数据的处理速度快



- 实时数据流处理的要求，是区别大数据引用和传统数据仓库技术，BI技术的关键差别之一；
- 1s 是临界点，对于大数据应用而言，必须要在1秒钟内形成答案，否则处理结果就是过时和无效的；



# 大数据的4V特征—Variety



## 数据来源多

企业内部多个应用系统的数据、互联网和物联网的兴起，带来了微博、社交网站、传感器等多种来源。

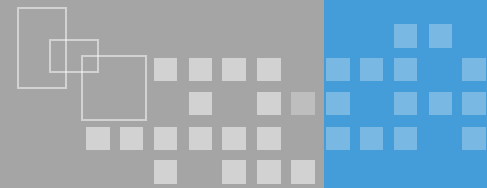
## 数据类型多

保存在关系数据库中的结构化数据只占少数，70~80%的数据是如图片、音频、视频、模型、连接信息、文档等非结构化和半结构化数据。

## 关联性强

数据之间频繁交互，比如游客在旅行途中上传的图片和日志，就与游客的位置、行程等信息有了很强的关联性。

# 大数据的4V特征—Value

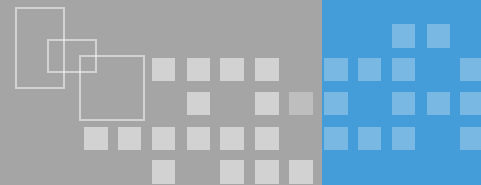


- 价值密度低(Value)
- 价值密度的高低与数据总量的大小成反比。以视频为例，一部1小时的视频，在连续不间断的监控中，有用数据可能仅有一二秒。
- 如何通过强大的机器算法更迅速地完成数据的价值“提纯”成为目前大数据背景下亟待解决的难题。



**挖掘大数据的价值类似沙里淘金，从海量数据中挖掘稀疏但珍贵的信息**

# 大数据技术简介



## 数据可视化



- 2D法
- 时间可视化

- 多维法
- 层次法

## 数据分析挖掘



- 实时处理
- Spark
- Storm

- 机器语音
- R语言
- 关联分析

- 用户画像构建
- 推理预测
- 知识图谱

## 数据存储



- 分布式架构
- Hadoop
- MapReduce

- 数据库体系
- NoSQL
- NewSQL

- MPP混合架构
- Lambda架构

## 数据预处理



- 数据清理
- 遗漏值处理
- 噪音数据

- 数据集成
- 实体识别
- 数据冗余

- 数据归约
- 维度归约
- 数值归约

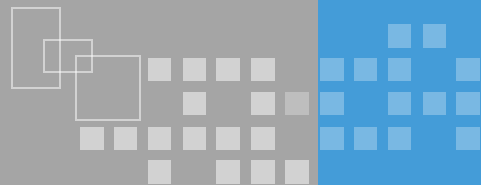
## 数据采集



- 硬件采集 传感技术
- RFID
- 

- 软件采集
- 系统日志抓取
- 企业特定API、网络众包

# 大数据技术现况



## 采集端

### 多源数据融合

- 社交数据、IOT数据 等
- 外部数据与企业内 部数 据融合拉通

### 运营化发展

- 通过长期采集积累 海量 数据资源的企 业实现数 据联合运 营



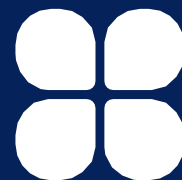
## 分析端

### 处理实时化

- 针对网络视频化潮流， 更 加强化对实时流数 据的处 理能力

### 高效化

- 数据量的不断增加， 云端 处理API结合后 台支撑将 使数据处理 更加高效



## 应用端

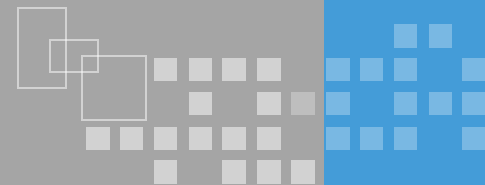
### 智能化

- 以机器学习作为核心， 将 实现对用户需求的自 我理 解和智能迭代

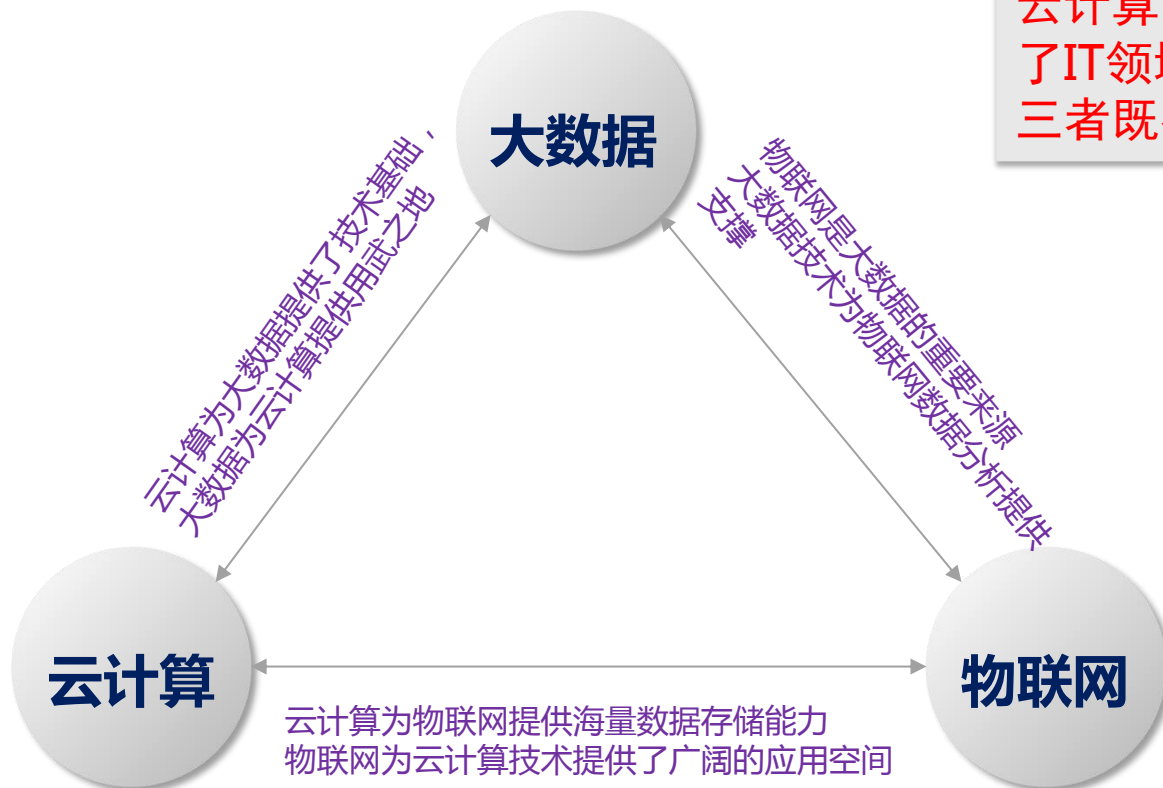
### 云端化

- 云计算和移动互联网 将促 进大数据应用从 2B市场迅速推广到 2C市场

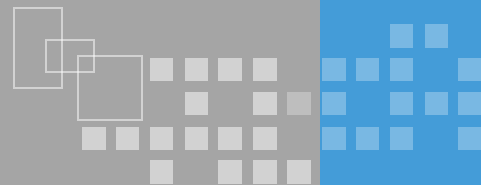
# 大数据及其关联技术



云计算、大数据和物联网代表了IT领域最新的技术发展趋势，三者既有区别又有联系





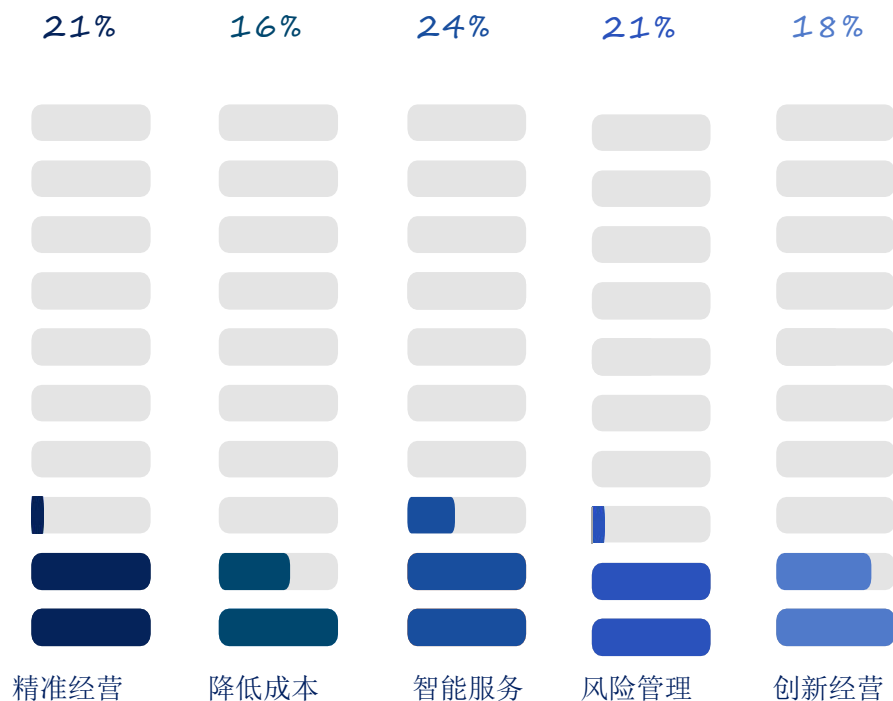


## 大数据主要用途

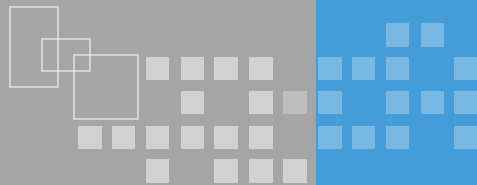
- 24%的大数据企业认为在为用户实施智能服务方面有较大帮助

- 数据价值 体现在三个方面

- 增加收入
- 减少支出
- 降低风险



# 大数据主要应用行业



## 大数据主要应用行业

### 能源行业

随着智能电网的发展，电力公司可以掌握海量的用户用电信息，利用大数据技术分析用户用电模式，可以改进电网运行，合理设计电力需求响应系统，确保电网运行安全。

### 物流行业

利用大数据优化物流网络，提高物流效率，降低物流成本。

### 城市管理

可以利用大数据实现智能交通、环保监测、城市规划和智能安防。

### 生物医学

大数据可以帮助我们实现流行病预测、智慧医疗、健康管理，同时还可以帮助我们解读DNA，了解更多的生命奥秘。

### 体育娱乐

大数据可以帮助我们训练球队，决定投拍哪种题材的影视作品，以及预测比赛结果。

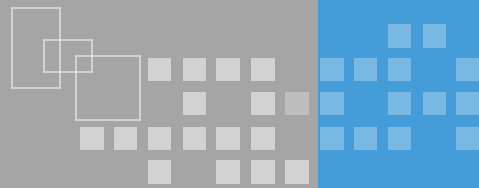
### 安全领域

政府可以利用大数据技术构建起强大的国家安全保障体系，企业可以利用大数据抵御网络攻击，警察可以借助大数据来预防犯罪。

### 个人生活

大数据还可以应用于个人生活，利用与每个人相关联的“个人大数据”，分析个人生活行为习惯，为其提供更加周到的个性化服务。

# 大数据主要应用行业



## 大数据主要应用行业

### 制造业

利用工业大数据提升制造业水平，包括产品故障诊断与预测、分析工艺流程、改进生产工艺，优化生产过程能耗、工业供应链分析与优化、生产计划与排程。

### 金融行业

大数据在高频交易、社交情绪分析和信贷风险分析三大金融创新领域发挥重大作用。

### 汽车行业

利用大数据和物联网技术的无人驾驶汽车，在不远的未来将走入我们的日常生活。

### 互联网行业

借助于大数据技术，可以分析客户行为，进行商品推荐和针对性广告投放。

### 餐饮行业

利用大数据实现餐饮O2O模式，彻底改变传统餐饮经营方式。

### 电信行业

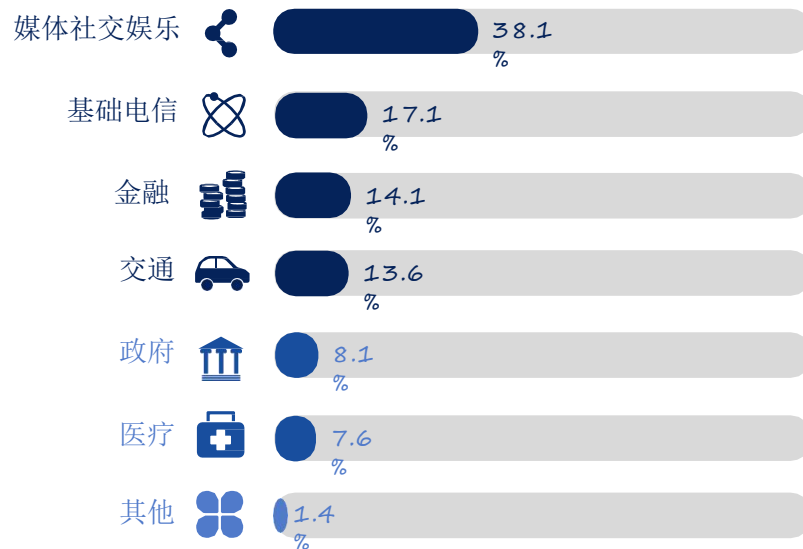
利用大数据技术实现客户离网分析，及时掌握客户离网倾向，出台客户挽留措施

# 大数据应用领域不断丰富

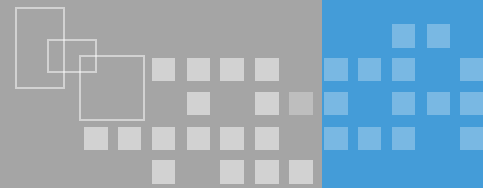


大数据应用领域不断丰富，  
从互联网、电信、金融开始向医疗、交通、政府领域深入

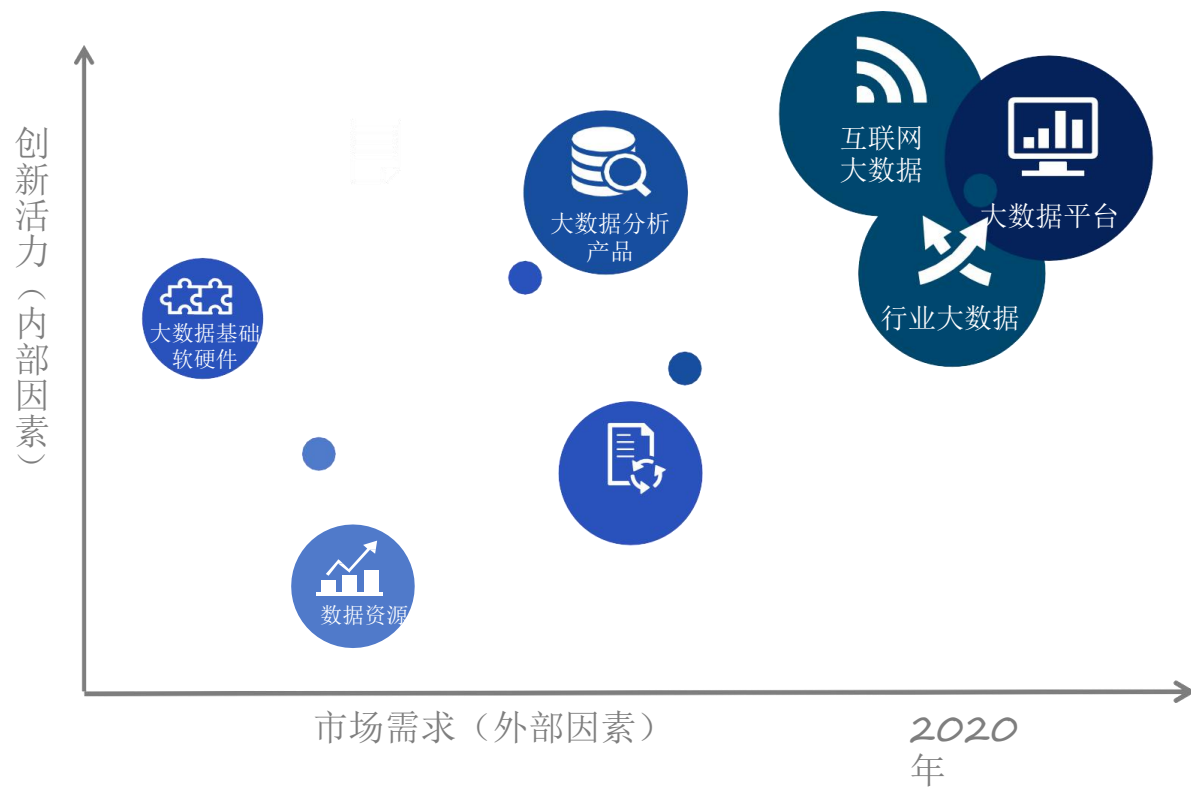
2018年中国大数据市场行业结构预



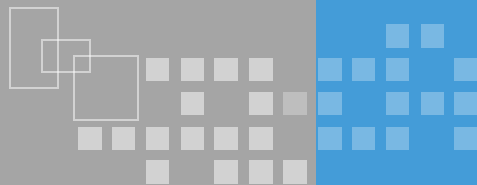
# 大数据产业发展



与互联网和传统行业融合创新将成为大数据产业爆发点



# 大数据发展趋势



01

## 机器学习成为智能分析核心技术

近年来，机器学习已经开始渗透到生活各个领域：客服机器人、垃圾邮件过滤、人脸识别、语音识别、个性化推荐……随着大数据分析能力的不断提高，2018年机器学习将继续在智能分析方面发挥重要作用。

02

## 多种科技和学科交叉融合

大数据技术的发展不仅能够将网络计算中心、移动网络技术和物联网、云计算等新型尖端网络技术充分地融合成一体，促进不同科学技术的交叉融合，同时还能够促进多学科的交叉融合，充分发挥出交叉学科和边缘学科在新时代的新功能与效用。

03

## 政府大数据将迅速发展

近日，国家相关部门就实施国家大数据战略进行第二次集体学习，指出将推动实施国家大数据战略，加快完善数字基础设施，推进数据资源整合和开放共享，保障数据安全，加快建设数字中国，更好服务我国经济社会发展和人民生活改善。因此，2018年政府将步入大数据建设快速发展的新阶段。

随着大数据技术的快速发展，企业和政府部门开始已经开始运用大数据来进行业务的分析、预测和决策。

04

## 物联网、云技术、大数据和网络安全深度融合

数据管理技术，如数据质量控制、数据准备、数据分析以及数据整合等方面的融合程度将在2018年达到新的高度。当我们将对智能设备的依赖程度增加时，互通性以及机器学习将会成为保护资产免遭网络安全危害的重要手段。

05

## 基于知识图谱的大数据应用将成为热门应用场景

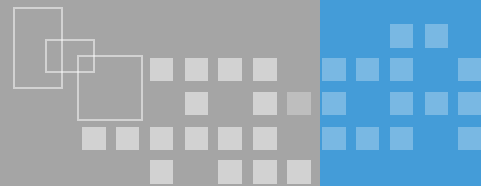
知识图谱的应用场景非常广泛，比如搜索、问答、推荐系统、反欺诈、不一致性验证、异常分析、客户管理等。2018年，基于知识图谱的大数据应用将衍生出更多热门应用场景。

06

## 隐私的保护与大数据的安全备受关注

大数据应用在带来便利的同时，也暴露了一系列问题，人们开始担心个人信息的安全，骚扰电话、账户盗用、地址泄露……如何保护隐私大数据也将提上日程。

# 大数据发展趋势



## 大数据平台将成为成长性最高的领域

各类互联网平台从原有的信息服务平台向“数据+”信息服务平台转变，大数据平台成为了信息通信领域成长性最快领域。

01



商贸、政务、金融、医疗卫生、工业等各行各业数据指导决策的需求日益迫切。

02

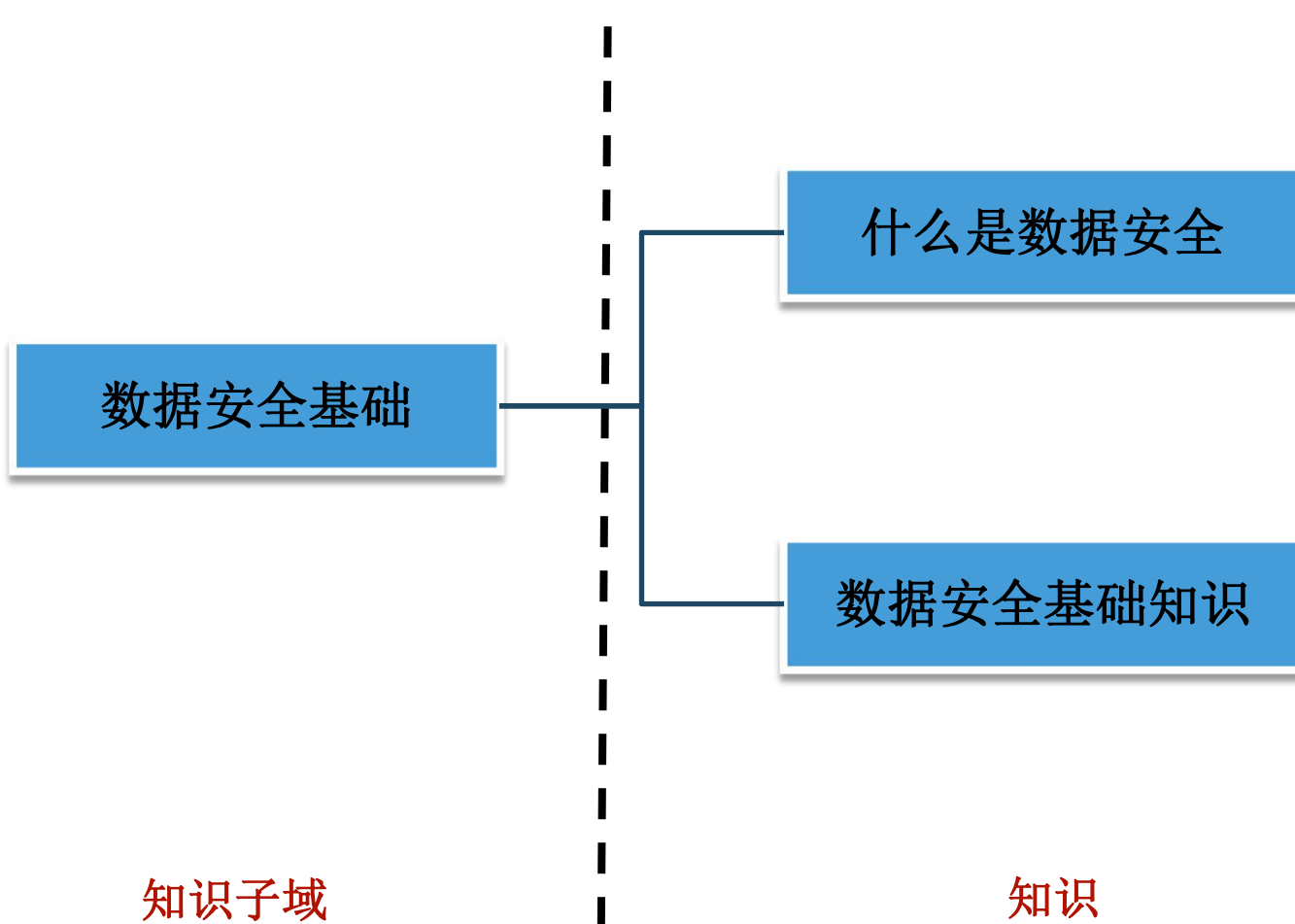
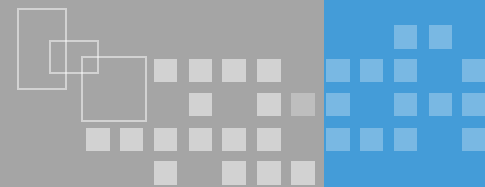


智能化、网络化、平台化加速了数据的采集和积累。

03

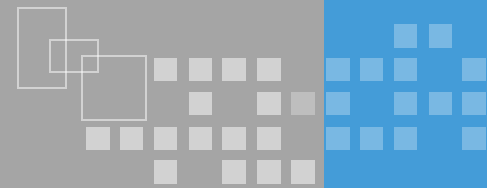


云计算技术发展为大数据存储和挖掘提供了重要的计算和存储资源保障。



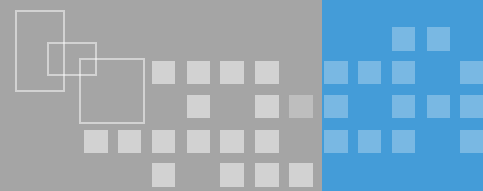


# 知识子域：什么是数据安全



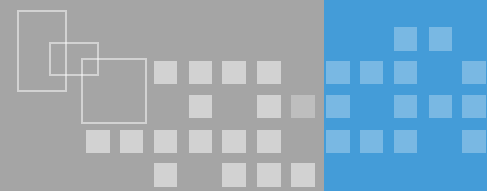
国际标准化组织（ISO）对计算机系统安全的定义是：为数据处理系统建立和采用的技术和管理的安全保护，保护计算机硬件、软件和数据不因偶然和恶意的原因遭到破坏、更改和泄露。由此计算机网络的安全可以理解为：通过采用各种技术和管理措施，使网络系统正常运行，从而确保网络数据的可用性、完整性和保密性。所以，建立网络安全保护措施的目的是确保经过网络传输和交换的数据不会发生增加、修改、丢失和泄露等。

# 数据安全-含义



信息安全或数据安全有对立的两方面的含义：一是数据本身的安全，主要是指采用现代密码算法对数据进行主动保护，如数据保密、数据完整性、双向强身份认证等，二是数据防护的安全，主要是采用现代信息存储手段对数据进行主动防护，如通过磁盘阵列、数据备份、异地容灾等手段保证数据的安全，数据安全是一种主动的包含措施，数据本身的安全必须基于可靠的加密算法与安全体系，主要是有对称算法与公开密钥密码体系两种。数据处理的安全是指如何有效的防止数据在录入、处理、统计或打印中由于硬件故障、断电、死机、人为的误操作、程序缺陷、病毒或黑客等造成的数据库损坏或数据丢失现象，某些敏感或保密的数据可能不具备资格的人员或操作员阅读，而造成数据泄密等后果。而数据存储的安全是指数据库在系统运行之外的可读性。一旦数据库被盗，即使没有原来的系统程序，照样可以另外编写程序对盗取的数据库进行查看或修改。从这个角度说，不加密的数据库是不安全的，容易造成商业泄密，所以便衍生出数据防泄密这一概念，这就涉及了计算机网络通信的保密、安全及软件保护等问题。

# 数据安全-特点



## 1) 机密性 ( Confidentiality )

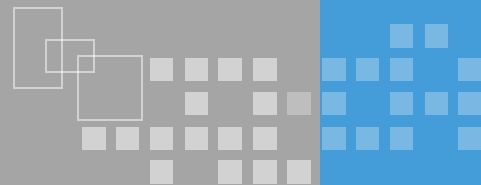
保密性 ( secrecy ) , 又称机密性 , 是指个人或团体的信息不为其他不应获得者获得。在电脑中 , 许多软件包括邮件软件、网络浏览器等 , 都有保密性相关的设定 , 用以维护用户资讯的保密性 , 另外间谍档案或黑客有可能会造成保密性的问题。

## 2) 完整性 ( Integrity )

数据完整性是信息安全的三个基本要点之一 , 指在传输、存储信息或数据的过程中 , 确保信息或数据不被未授权的篡改或在篡改后能够被迅速发现。在信息安全领域使用过程中 , 常常和保密性边界混淆。以普通RSA对数值信息加密为例 , 黑客或恶意用户在没有获得密钥破解密文的情况下 , 可以通过对密文进行线性运算 , 相应改变数值信息的值。例如交易金额为X元 , 通过对密文乘2 , 可以使交易金额成为2X。也称为可延展性 ( malleably ) 。为解决以上问题 , 通常使用数字签名或散列函数对密文进行保护。

## 3) 可用性 ( Availability )

数据可用性是一种以使用者为中心的设计概念 , 易用性设计的重点在于让产品的设计能够符合使用者的习惯与需求。以互联网网站的设计为例 , 希望让使用者在浏览的过程中不会产生压力或感到挫折 , 并能让使用者在使用网站功能时 , 能用最少的努力发挥最大的效能。基于这个原因 , 任何有违信息的 “可用性” 都算是违反信息安全的规定。因此 , 世上不少国家 , 不论是美国还是中国都有要求保持信息可以不受规限地流通的运动举行。



## Data Loss/Leakage Prevention/Protection

### 数据泄露防护、数据防泄漏

#### 国外定义：

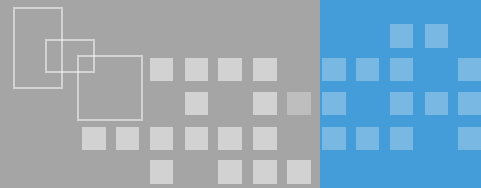
DLP是基于**深度内容识别**技术，对**传输中、存储中、使用中**的数据进行检测，**识别敏感数据**，依据预先定义的**策略**，实施特定**响应**，敏感数据泄漏达到有效防护效果的解决方案。（引用：Gartner报告）

#### 国内定义：

数据泄露防护产品通过使用关键字、正则表达式等方式，对**运行、存储**于主机内或者**网络**中传输的文件、数据进行**内容识别**，对数据的操作和移动过程进行**监视和控制**，实现对数据以非授权的形式流出安全域进行**防护**的功能；同时该类产品还应具有基本的身份鉴别、安全管理和审计功能。该类产品的部署可分为主机型、网络型或综合型（包含主机和网络设备）。（引用：数据泄露防护产品检测条件（JCTJ 008—2015））

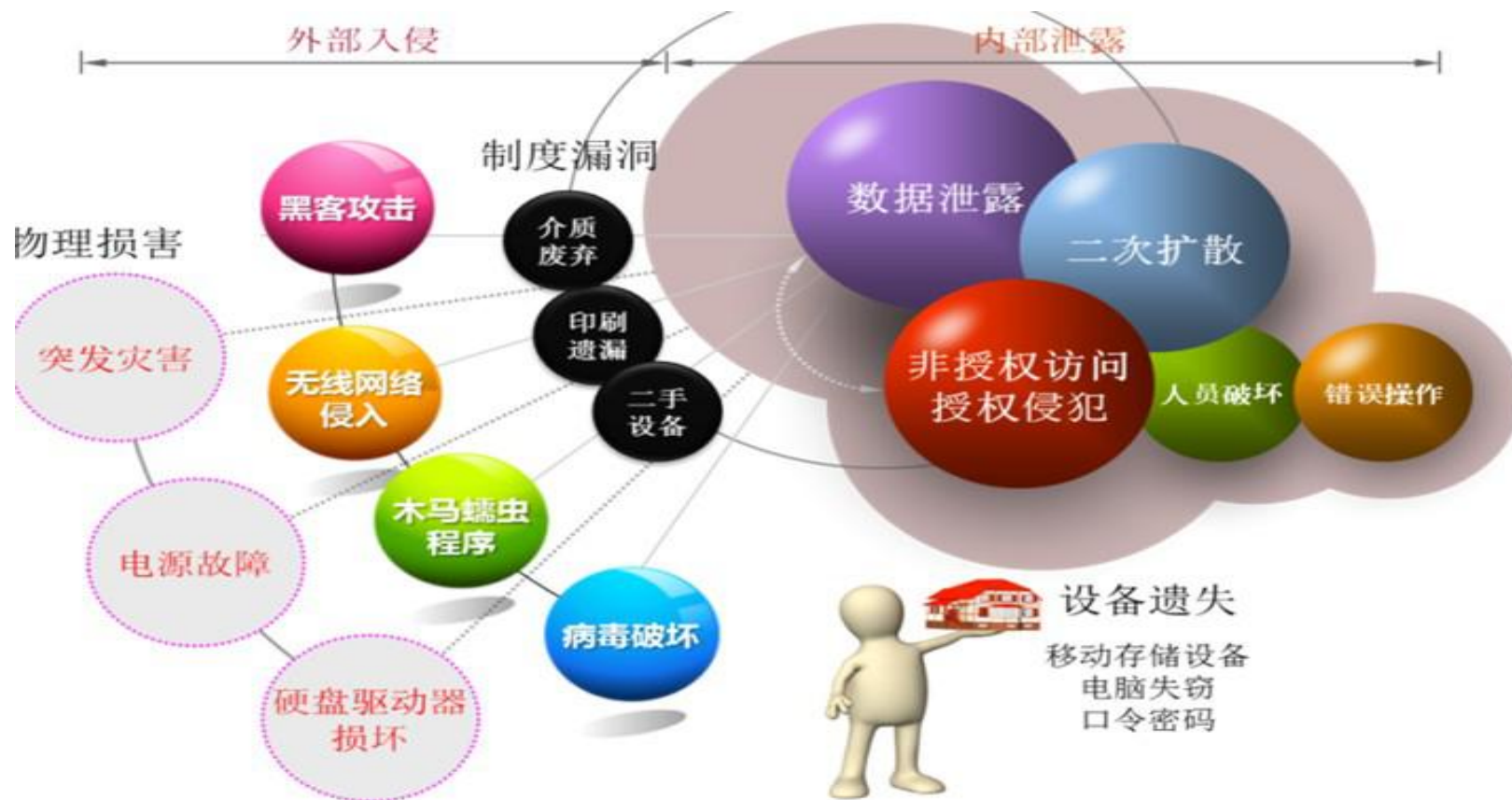
**DLP核心标志在于是否使用“深度内容识别技术”**

# 数据泄露途径、防范措施

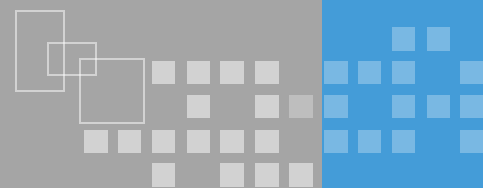


- ❖ 数据泄露途径
- ❖ 数据泄露防范措施
- ❖ 数据泄密统计分析表

# 数据泄密途径示意图

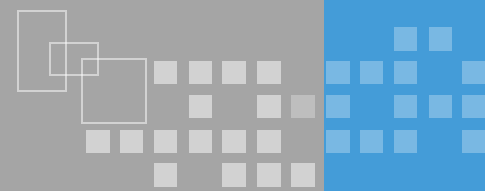


# 泄密途径具体分析



- ❖ 泄密的原因有内部泄密、内部失密和外部窃密。
- ❖ 1、内部人员离职拷贝带走资料泄密—这类情况发生概率最高。
- ❖ 2、内部人员无意泄密和恶意泄密—企业内部人员在上网时候不小心中了病毒或木马，电脑上存储的重要资料被流失的情况也非常多
- ❖ 3、外部竞争对手窃密—竞争对手采用收买方式，买通企业内部人员，让内部人员把重要信息发送竞争方，从而窃取机密的情况也非常多。

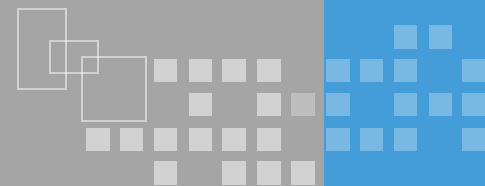
# 泄密途径具体分析



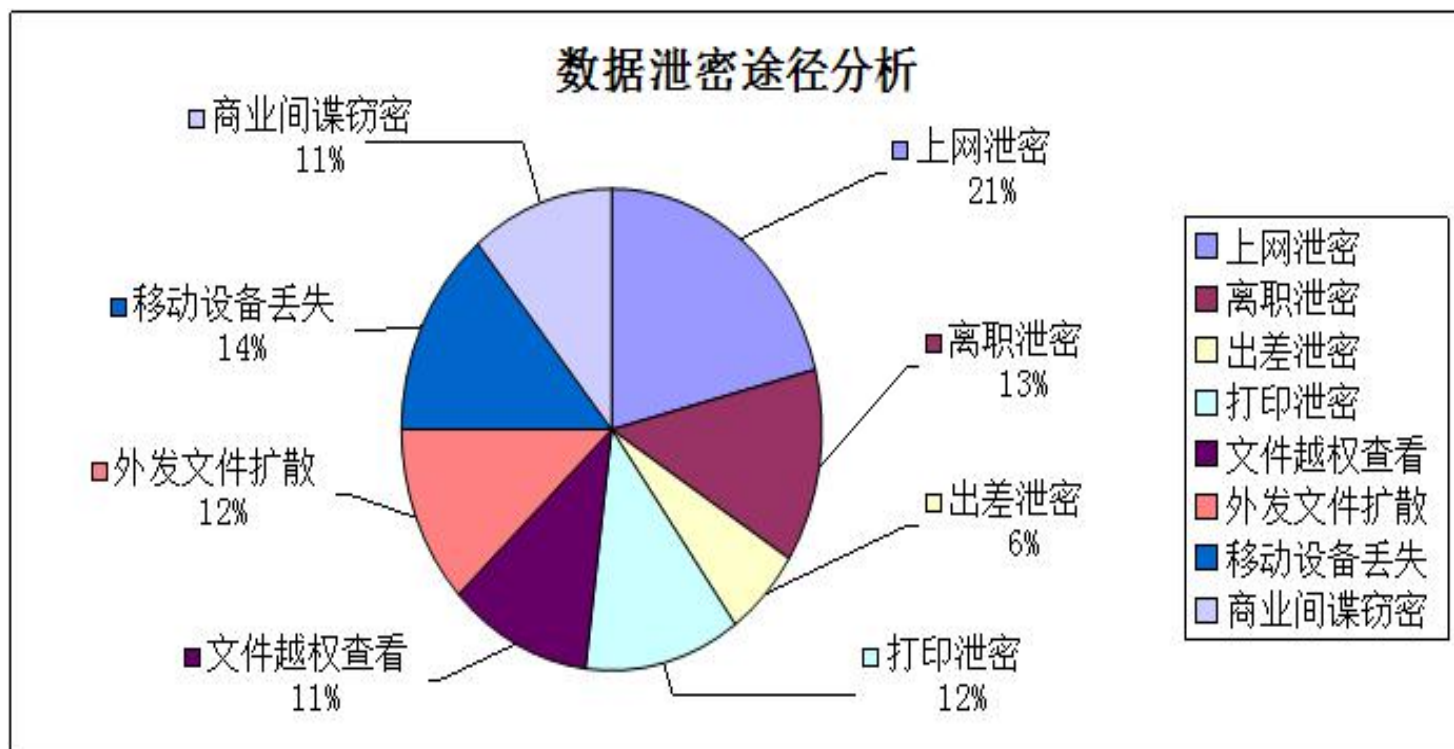
- ❖ 4、黑客窃密—目前国内许多黑客，通过层出不穷的技术手段，窃取国内各种重要信息。
- ❖ 5、存储设备丢失和维修失密 移动存储设备例如笔记本电脑、移动硬盘、手机存储卡、数码照相/摄录机等，一旦遗失、维修或者报废后，其存储数据往往暴露无遗。随着移动存储设备的广泛使用，家庭办公兴起，出差人员的大量事务处理等等都会不可避免地使用移动存储设备。因此，移动存储设备丢失和维修导致泄密也是当前泄密事件发生的主要原因之一。



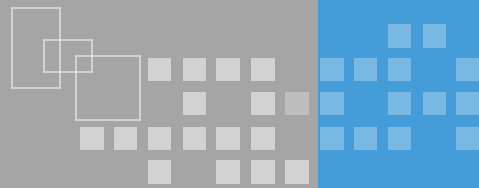
# 例如U盘泄露数据



# 数据泄密统计分析表

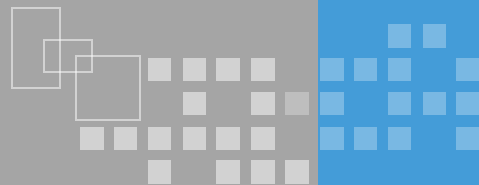


# 数据泄露防范措施



- 1、切断源头，设立信息级别制度——对公司的机密文件进行级别划分，确定机密文件传播的范围，让所有人了解信息的传播界限，员工由自身的岗位确定相应的信息级别，低层次的岗位级别不能查阅、了解、拷贝、接触到高层次的信息级别。
- 2、重视保密协议——无规矩不成方圆，无论作用大小，和员工签署清晰的保密协议还是必要的。保密协议的内容越详细越好，如果对方心胸坦白，自然会欣然同意。
- 3、责任分解——明确每个人对相关信息的安全责任。所有的机密文件如果出现泄露，可以根据规定找到责任人，追究是次要的，相互监督和防范才是责任分解的最终目的。

# 数据泄露防范措施

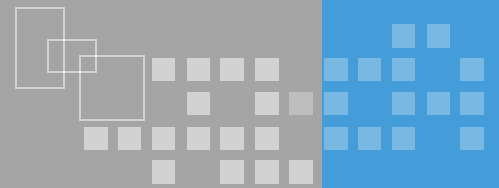


4、防病毒 计算机病毒一般都隐藏在程序和文档中。目前典型的防病毒技术就是对信息中的病毒特征代码进行识别和查杀。

5、VPN加密通道 虚拟专用网VPN需要通过不可信的公用网络来建立自己的安全信道，因此加密技术是重要的选择。

6、水印技术 水印技术是信息隐藏技术的一种。一般信息都是要隐藏在有一定冗余量的媒体中，比如图像、声音、录像等多媒体信息，在文本中进行隐藏比较少。水印技术是可以替代一般密码技术的保密。

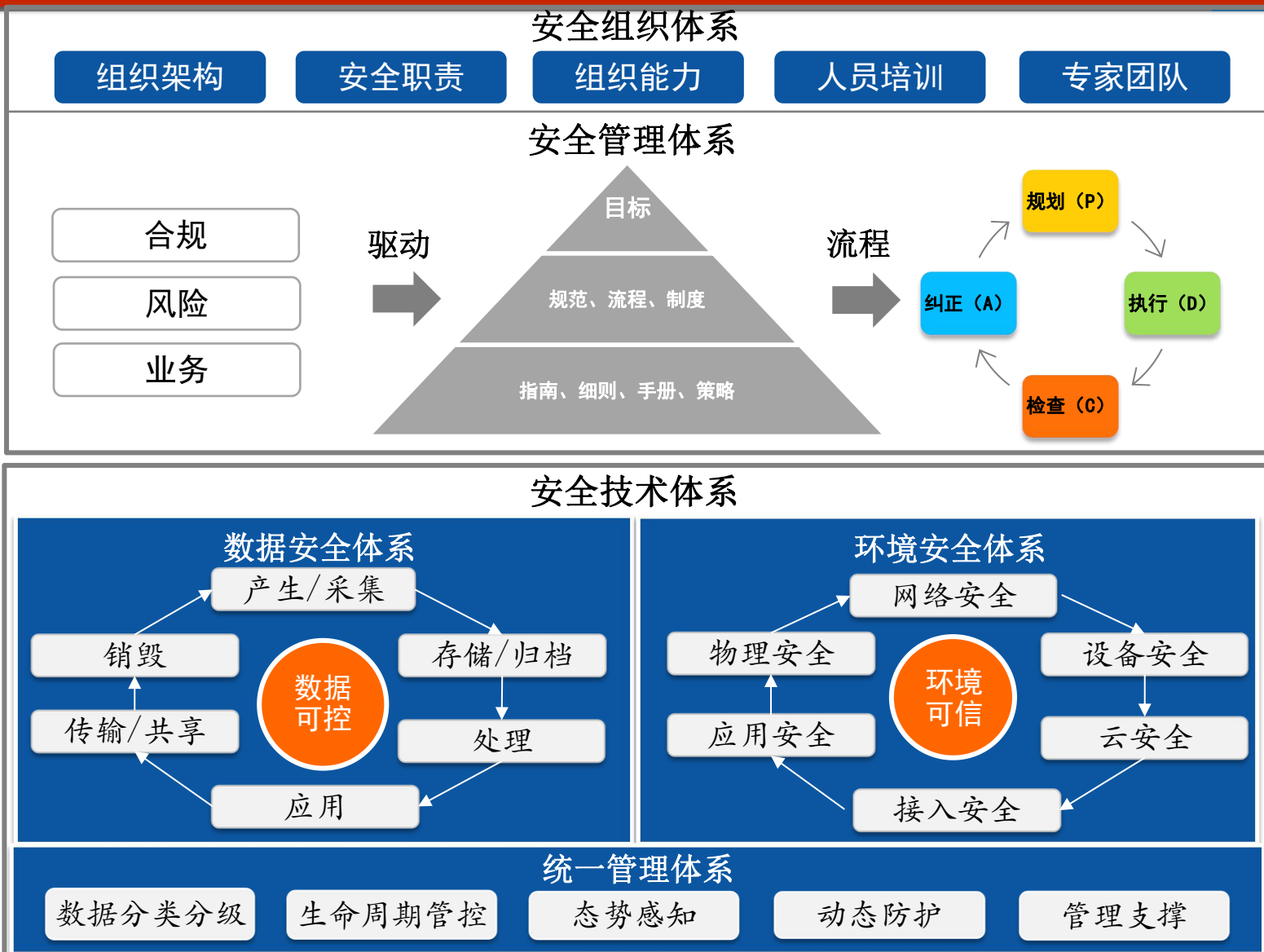
# 安全措施小例



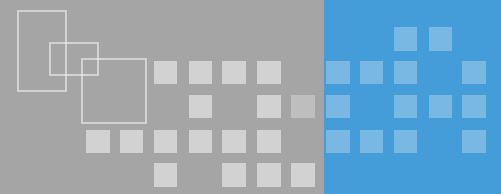
- 1、提示语：“请擦去黑板并将你所有的秘密文件处理掉，工业间谍在你之后预定了这一房间。”
- 2、办公区放置了碎纸机，员工需要将所有用过的废纸进行处理。

# 数据安全体系

## 以数据为中心的安全架构



分阶段分步骤建设 保证内外合规和持续改进



## 一、物理设备安全

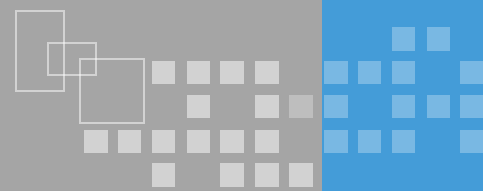
1.1、服务器、路由器、交换机、工作站、打印机等硬件实体和通信链路免受自然灾害、人为破坏、和搭线窃听攻击。

1.2、验证用户的身份和权限，防止越权操作；

1.3、建立完备的机房安全管理制度，妥善保管备份文件和文档资料，防止非法人员进入机房进行偷窃和破坏活动等。

## 二、环境安全

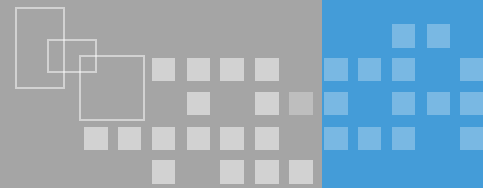
2.1、确保网络设备有一个良好的电磁兼容环境，物理位置选择、物理访问控制、防盗和防破坏、防雷击、防火、防水、防潮湿、防静电、电力保障、电磁防护抑制和防止电磁泄漏，可以采用屏蔽措施和伪噪声技术等来解决。



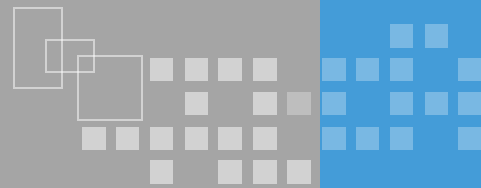
## 一、身份鉴别机制

- 1.1、建立用户和分配权限，定期检查用户实际权限与分配权限的符合性；
- 1.2、通过身份鉴别、访问控制等严格的规定限制远程管理账户的操作权限和登录行为；
- 1.3、明确各类用户的责任、义务和风险，对系统账户的登记造册、用户名分配、初始口令分配、用户权限及其审批程序、系统资源分配、注销等作出规定；

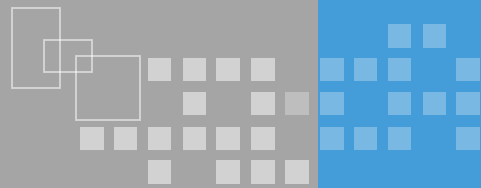




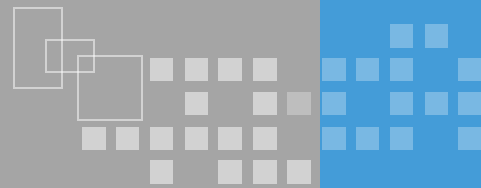
- ❖ 身份鉴别
- ❖ 访问控制
- ❖ 通信保密性
- ❖ 通信完整性
- ❖ 软件容错
- ❖ 资源控制



- ❖ 数据完整性
- ❖ 数据保密性
- ❖ 数据备份与恢复

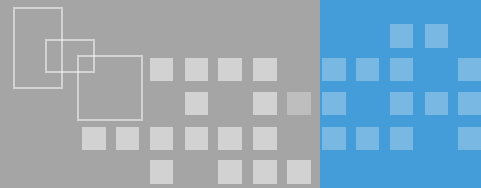


- ❖ 结构安全和网段划分
- ❖ 网络访问控制
- ❖ 网络入侵防范
- ❖ 网络设备防护
- ❖ 恶意代码防范



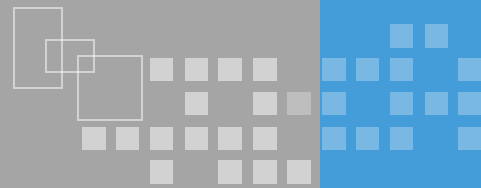
- ❖ 身份鉴别
- ❖ 自主访问控制
- ❖ 强制访问控制
- ❖ 系统保护
- ❖ 剩余信息保护
- ❖ 入侵防范
- ❖ 恶意代码防范
- ❖ 资源控制

# 企业数据安全防范



- 1、数据安全防范大纲
- 2、企业数据安全防范措施
- 3、企业职工数据安全防范措施

# 数据安全防范大纲



1、保证企业数据财富的安全，一定是人防与技防并举，指望靠一个硬件或者一个加密软件解决所有问题，是企业信息化过程中很容易犯的低级错误。

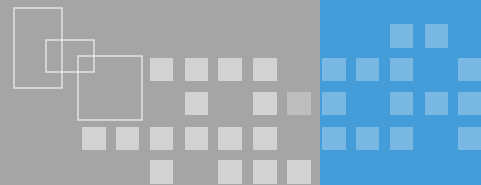
用好硬件和加密软件的主要要素，最少应该包括如下三点：

1.领导重视。要把企业数据看成财富和资产，有强烈的数据安全意识，建立制度并监督执行。

2.建立制度。用制度规范不同岗位职责和数据流向，时刻提醒员工扣紧数据安全这根弦。

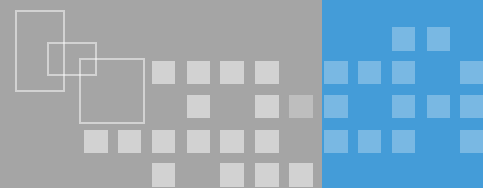
3.做好服务。出现使用问题及时处理，出现业务冲突及时解决，但必须保证数据是安全的。

# 数据安全防范措施



- 1、强化安全保护意识，加强计算机及系统本身实体的安全管理。
- 2、建立健全企业信息安全管理制度和操作规程制度建设是确保企业信息安全的关键。
- 3、结合自身硬软件、数据和网络等方面实际情况，提高工作人员安全意识通过设置身份限制，对不同设备设置访问权限，对于各子系统工作人员每人设置一个账号，并且每个账号设置口令，并要求定期进行更改口令。
- 4、企业数据安全简单说主要分为是防御和恢复两个部分，防御就是安装一些杀毒软件，服务器防火墙等等，恢复的话主要就是数据的备份，已经数据的恢复。

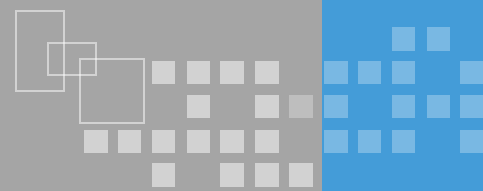
# 具体数据安全防范措施



- 1、打印机——打印时人必须在跟前，防止走开别别人复印泄露信息。
- 2、打印纸背面——提醒：重要文档不要戴着节约的帽子，就顺手纳入到废物利用的行列！
- 3、电脑易手——了解新公司最好的渠道,提醒：电脑部的管理人员必须将工作资料清空再格式，小心驶得万年船，况且这只是举手之劳。
- 4、共享——做好文件再通知窃取者,提醒：要么不用共享，要用的话，记住删除你的共享文件，或者提示一声“拷完就删”。
- 5、新进培训——无所保留只会有所失望,提醒：培训是好事，但对新员工而言，重要的是基础培训，而不是一切信息告知。

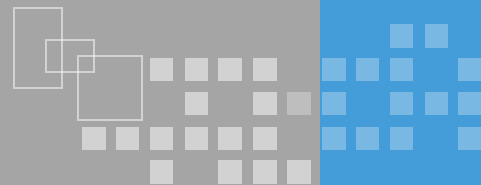


# 具体数据安全防范措施

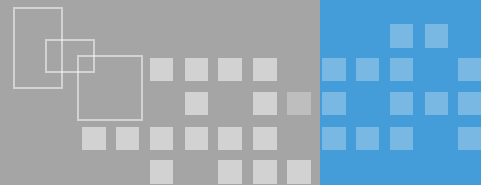


- 6、公用设备——不等于公用信息，例如移动硬盘公用，那数据可能会泄密。
- 7、光盘刻录——资料在备份过程中流失,要求重新备份，大多情况下，留在光驱里的“废盘”就可以在下班后大大方方带出公司。
- 8、会议记录——被忽视的黄金,提醒：行政要有安全意识。
- 9、应该按照一定的规则设置桌面系统的登录密码，并且定期修改，减少登录密码泄露或被破解的可能性。
- 10、定时清除IE浏览器的临时文件夹，可以防止部分敏感内容的泄露。

# 具体防范措施



- 11、为防止恶意代码植入系统，预防计算机病毒感染，在收到来历不明的电子邮件时，应注意不要随意打开邮件，并立即予以删除。
- 9、禁止在未经授权的情况下，私自修改系统的计算机命名标识和网络配置。
- 10、禁止在未经授权的情况下，私自安装任何应用软件，在获得授权后，只允许安装与工作有关的正版应用软件。
- 11、禁止访问互联网上与工作无关或来历不明的站点，禁止从互联网下载与工作无关的文件。



- ❖ ISO27001信息安全管理体系标准
- ❖ 数据安全应急方针
- ❖ 数据安全防范措施
- ❖ 数据安全硬件启用方针

# ISO27001 信息安全管理标准

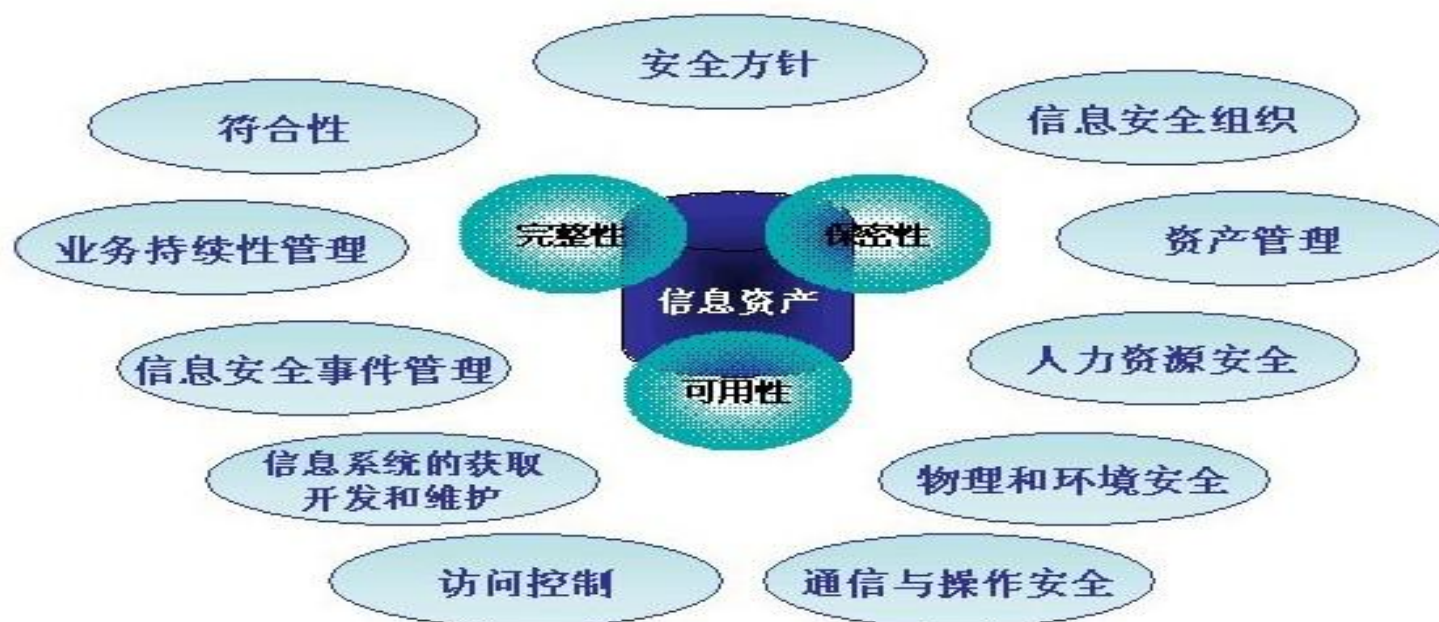
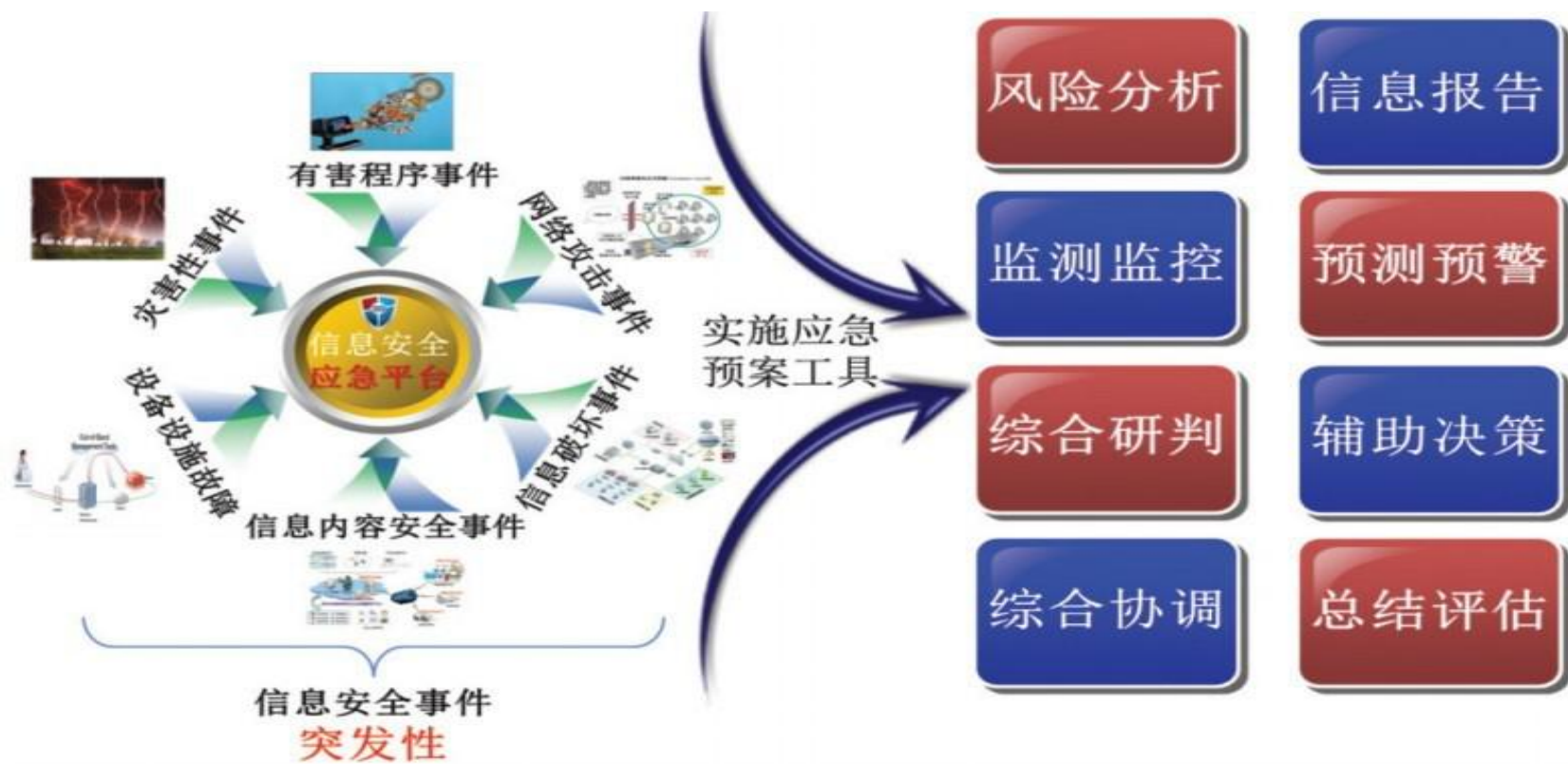
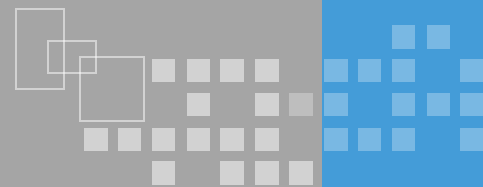
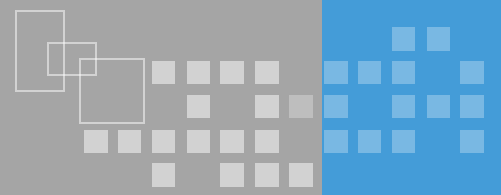


图 1：ISO27001 信息安全管理标准 11 大控制领域

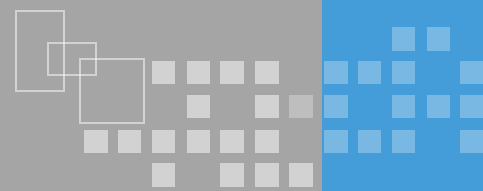
# 数据安全应急方针



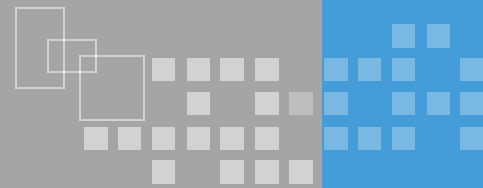
# 数据安全防范措施—运维



- 1、指定专人对网络进行管理，负责运行日志、网络监控记录的日常维护和报警信息分析和处理工作；
- 2、根据厂家提供的软件升级版本对网络设备进行更新，并在更新前对现有的重要文件进行备份；
- 3、进行网络系统漏洞扫描，对发现网络系统安全漏洞进行及时的修补；应保证所有与外部系统连接均应得到授权和批准；
- 4、建立网络安全管理制度，对网络安全配置、网络用户以及日志等方面作出规定；
- 5、对网络设备的安全策略、授权访问、最小服务、升级与打补丁、维护记录、日志以及配置文件的生成、备份、变更审批、符合性检查等方面做出具体要求；

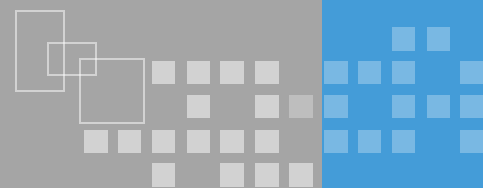


- 6、明确各类用户的责任、义务和风险，并按照机构制定的审查和批准程序建立用户和分配权限，定期检查用户实际权限与分配权限的符合性；
- 7、对日志的备份、授权访问、处理、保留时间等方面做出具体规定，使用统一的网络时间，以确保日志记录的准确；
- 8、通过身份鉴别、访问控制等严格的规定限制远程管理账户的操作权限和登录行为；
- 9、定期检查违反规定或其他违反网络安全策略行为。
- 10、指定专人对系统进行管理，删除或者禁用不使用的系统缺省账户；



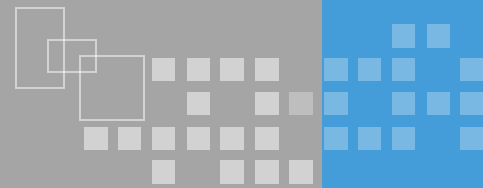
- 11、制定系统安全管理制度，对系统安全配置、系统帐户以及审计日志等方面作出规定；
- 12、对能够使用系统工具的人员及数量进行限制和控制；
- 13、定期安装系统的最新补丁程序，并根据厂家提供的可能危害计算机的漏洞进行及时修补，并在安装系统补丁前对现有的重要文件进行备份；
- 14、根据业务需求和系统安全分析确定系统的访问控制策略，系统访问控制策略用于控制分配信息系统、文件及服务的访问权限；
- 15、对系统账户进行分类管理，权限设定应当遵循最小授权要求；





- 16、对系统的安全策略、授权访问、最小服务、升级与打补丁、维护记录、日志以及配置文件的生成、备份、变更审批、符合性检查等方面做出具体要求；
- 17、提高所有用户的防病毒意识，告知及时升级防病毒软件；
- 18、在读取移动存储设备(如移动硬盘、光盘)上的数据以及网络上接收文件或邮件之前，先进行病毒检查，对外来计算机或存储设备接入网络系统之前也要进行病毒检查；
- 19、办公桌面系统必须安装防病毒软件，并且启动病毒监控和在线自动更新功能。

# 数据安全防范措施—运维



20、为了防止客户机在浏览互联网时，被含有恶意代码的程序所感染，应将IE浏览器的安全级别调整为“中”，并将隐私级别设置为“中高”。

21、应当安装操作系统最新的补丁软件包，弥补操作系统本身存在的安全漏洞，防止被攻击者或者计算机病毒所利用。

22、启动操作系统的自动更新服务，操作系统会定期自动升级并安装最新的补丁程序。

23、为防止意外情况造成文件损失，注意定期备份重要的文件，并保管好存储备份文件的介质。



**谢谢，请提问题！**