



Information Management

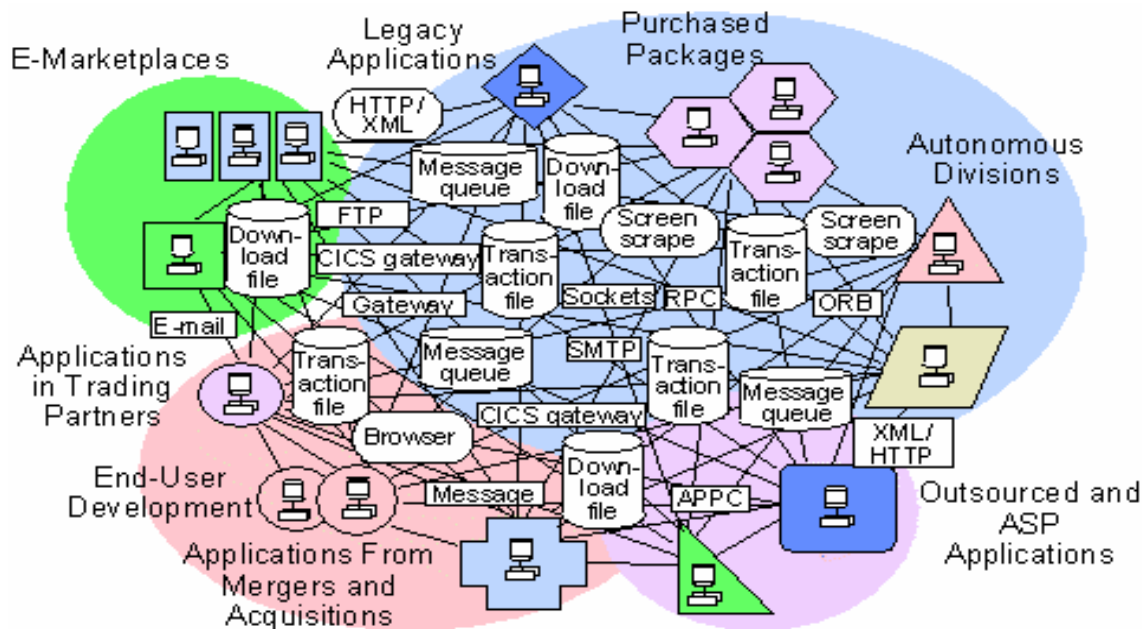
# 整合企业的信息资产 ——Information Integration Solution

朱丹 IBM软件部  
zhuddan@cn.ibm.com  
010-65391188-2275



**ON** DEMAND BUSINESS™

# 企业面临的挑战



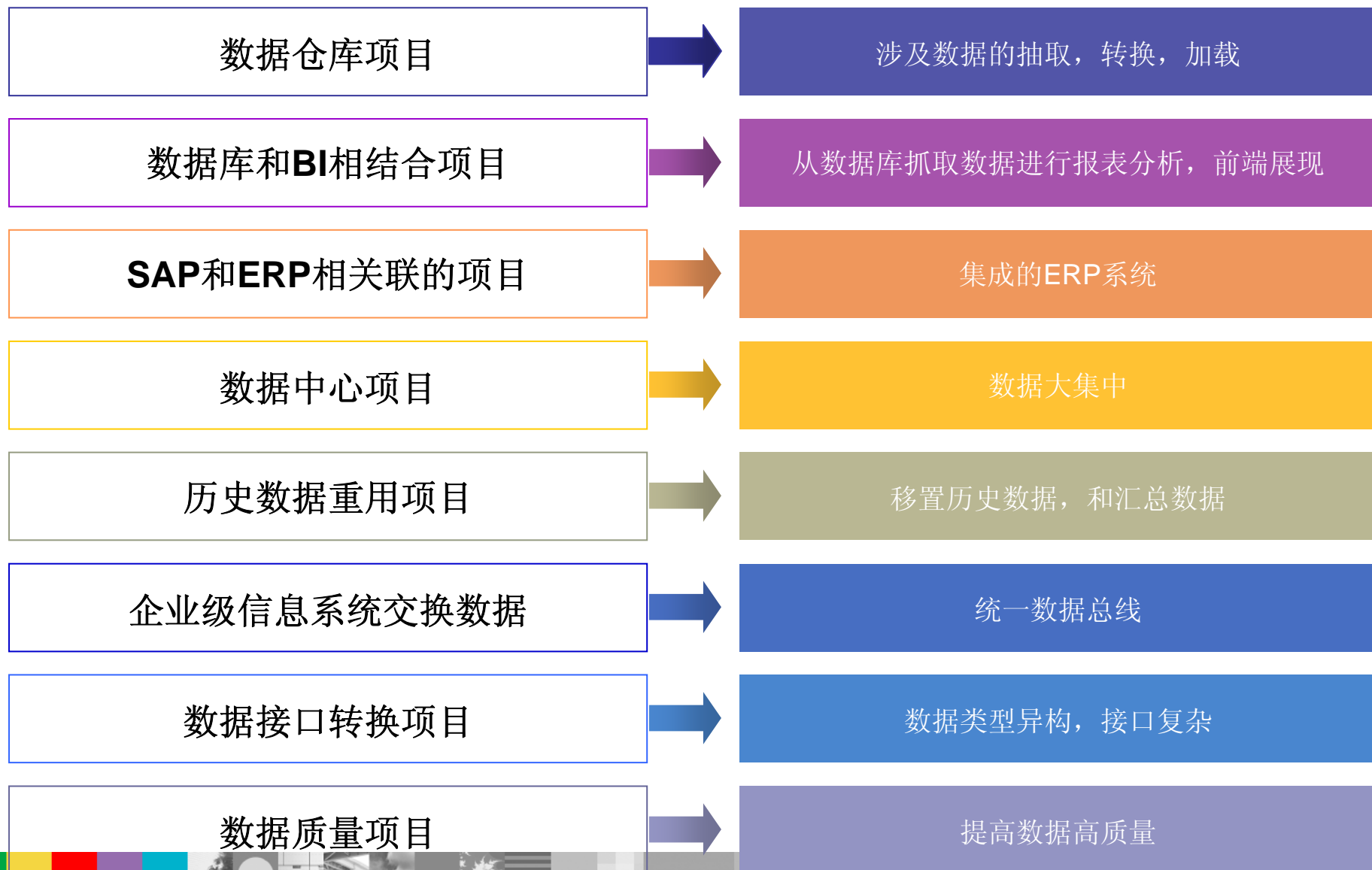
技术问题包括:

- 数据孤岛：未知数据丢失，相关数据被截留；忽视；浪费
- 响应时间限制了可处理的数据量
- 恶劣的数据质量 & 有瑕疵的假设
- 访问源数据的颗粒度不细
- 客户化整合使得应用延期

带来的相应结果:

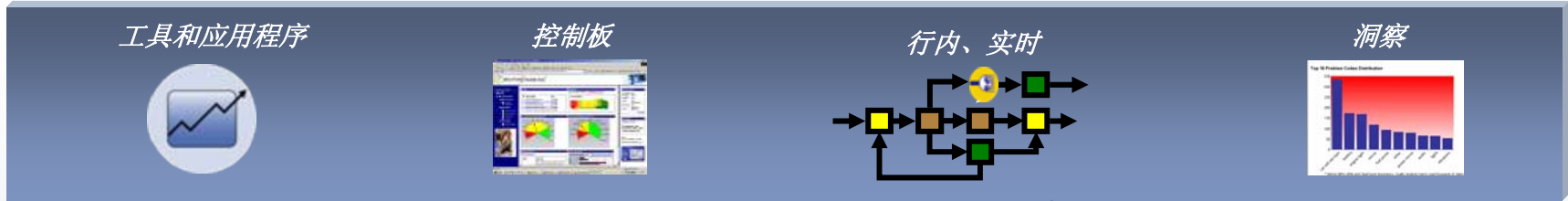
- 信息的可共享度降低
- 没有数据的解决方案
- 项目成本超支
- 不能及时地作出战略决定
- 技术与业务逻辑不能重新利用或重新组合业务

## 可以实施信息整合的项目



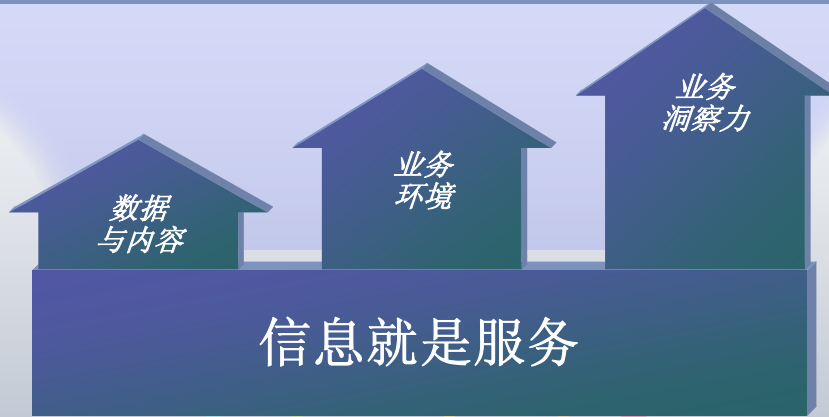
# 信息就是服务

## 从基于项目转向一个灵活的体系架构 (SOA)



基于各种标准: 例如,  
XQuery, JSR170, JDBC,  
Web 服务...

元数据 管理



实时的: 例如, 在线帮助,  
同步的主数据...

提取: 例如, 巴塞尔资本  
协议 (Basel II)、业务优  
化...

异类应用程序和信息



DB2

External Supplier /  
Business Partner

abc...



Point  
Content Management  
Application

IBM  
Content  
Manager



Internally  
Developed

xyz...



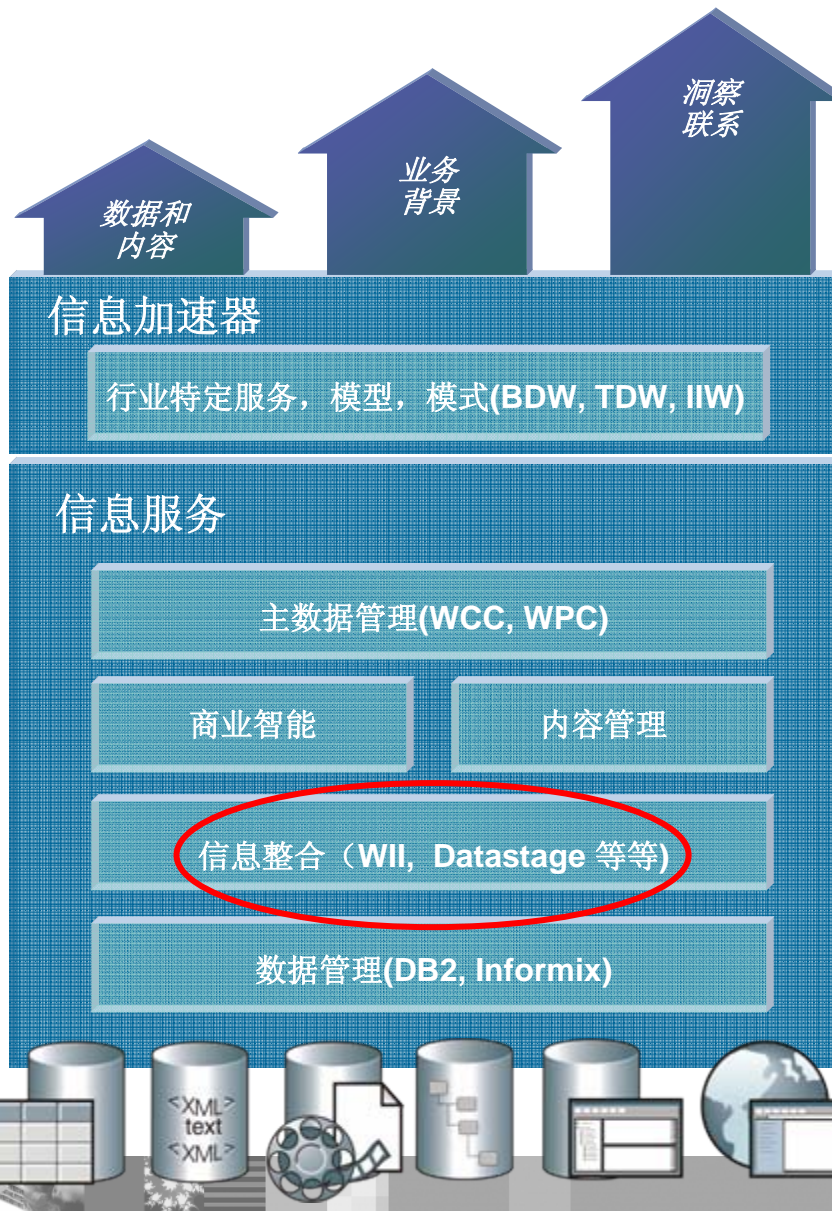
PeopleSoft.

Oracle



以及  
更多...

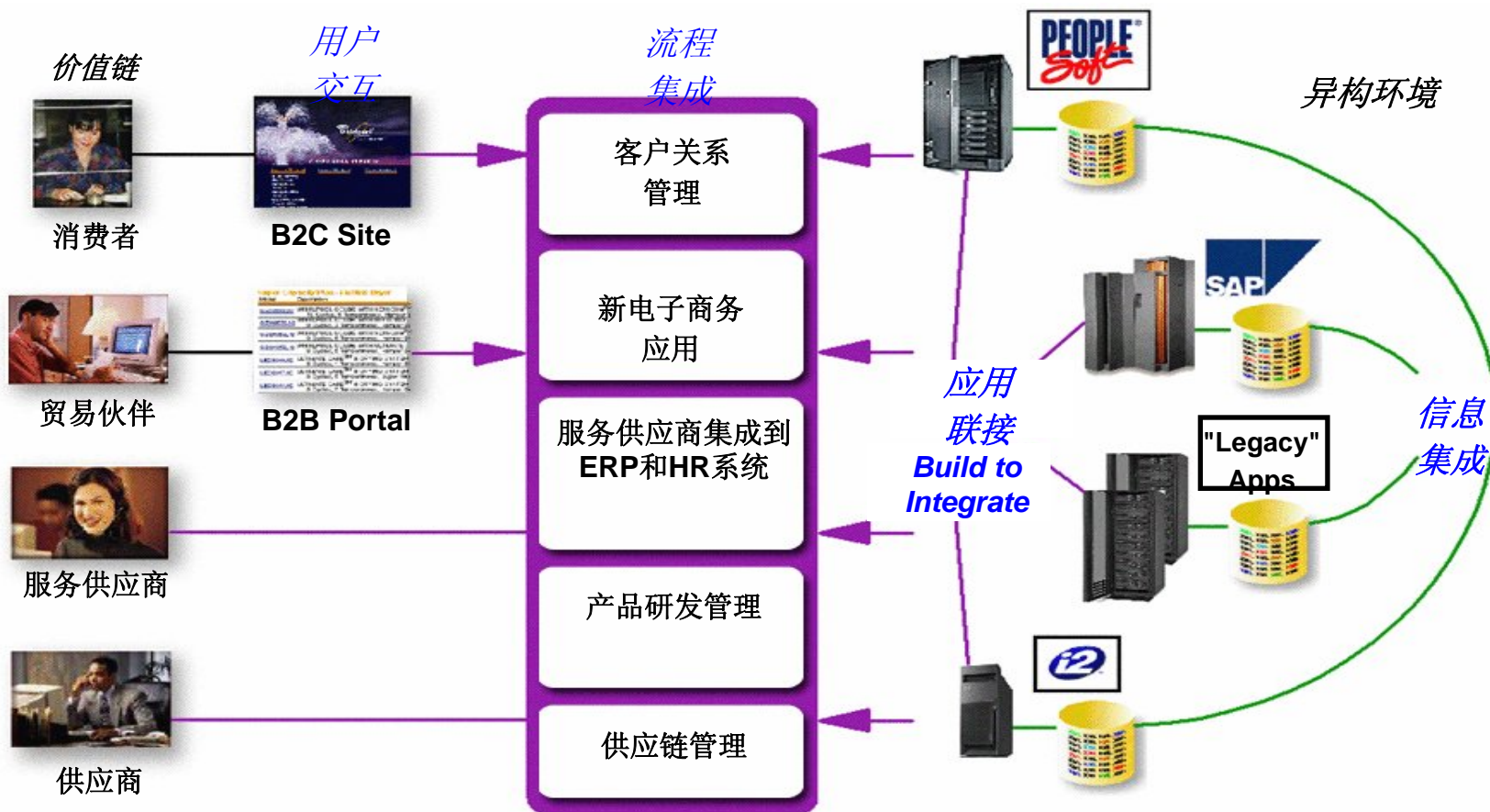
# IBM信息管理架构



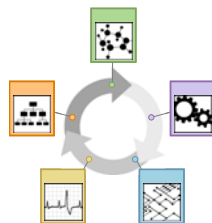
- 虚拟化
- 背景相关 (**context**)
  - 安全
  - 角色
  - 任务

- 基于各种标准
  - SQL, XQuery, JSR170, XML, Web 服务, JDBC, UIMA...

# 需要不同类型的集成



# 集成: 关键能力需求



## Information

应用访问标准化  
 动态适应工作环境  
 支持任何时间,任何地点的访问

流程建模  
 应用整合  
 对外接口  
 流程监控  
 管理业务结果

有效利用数据及内容资源  
 分布式数据访问  
 数据集中  
 数据转换  
 管理数据分布

安全性和完整性





# 我们需要的是一致、快速、可重复的跨系统信息集成解决方案

## 理解

数据源是否以你想象的方式存在？

哪一个数据源包含项目所需的数据？

数据源提供的数据是否正确表达你期望的业务含义？

## 清洗

如何匹配相似的不同数据记录？

如何修正和提高数据质量？

如何标准化数据存在格式？

## 转换

如何及时地在系统间更新和提交所有的数据？

如何保持系统间数据的同步？

是否需要根据事件和数据内容为用户发送数据？

## 联邦

你的业务人员是否可以及时看到客户的当前活动和系统间的信息联系？

如何让新产品快速推向各个渠道？原有的方式有何不同？

遍布各地的供应商，如何获得全局的各类信息，比如费用？

时间价值



# The IBM WebSphere 信息集成平台

## Service Oriented Architecture

理解



探查, 定义, 模型, 管理  
信息, 质量和结构

清洗



标准化, 合并,  
协调信息

转换



转换和丰富信息

联邦



异构信息的虚拟化访问

元数据管理

并行处理

数据



连接



访问, 发布, 复制信息



内容



# The IBM WebSphere 信息集成平台

## Service Oriented Architecture

理解



WebSphere ProfileStage™

WebSphere MetaStage™

- 数据来源鉴别
- 数据分析
- 数据来源元数据确认
- 探索数据质量、内容和结构
- 业务规则确认

联邦



WebSphere Information Integrator Standard Edition

WebSphere Information Integrator Classic Federation

WebSphere Information Integrator Content Edition

元数据管理

并行处理

数据

连接

内容



WebSphere Information Integrator Replication Edition

WebSphere Information Integrator Event Publisher Edition

# 数据探索



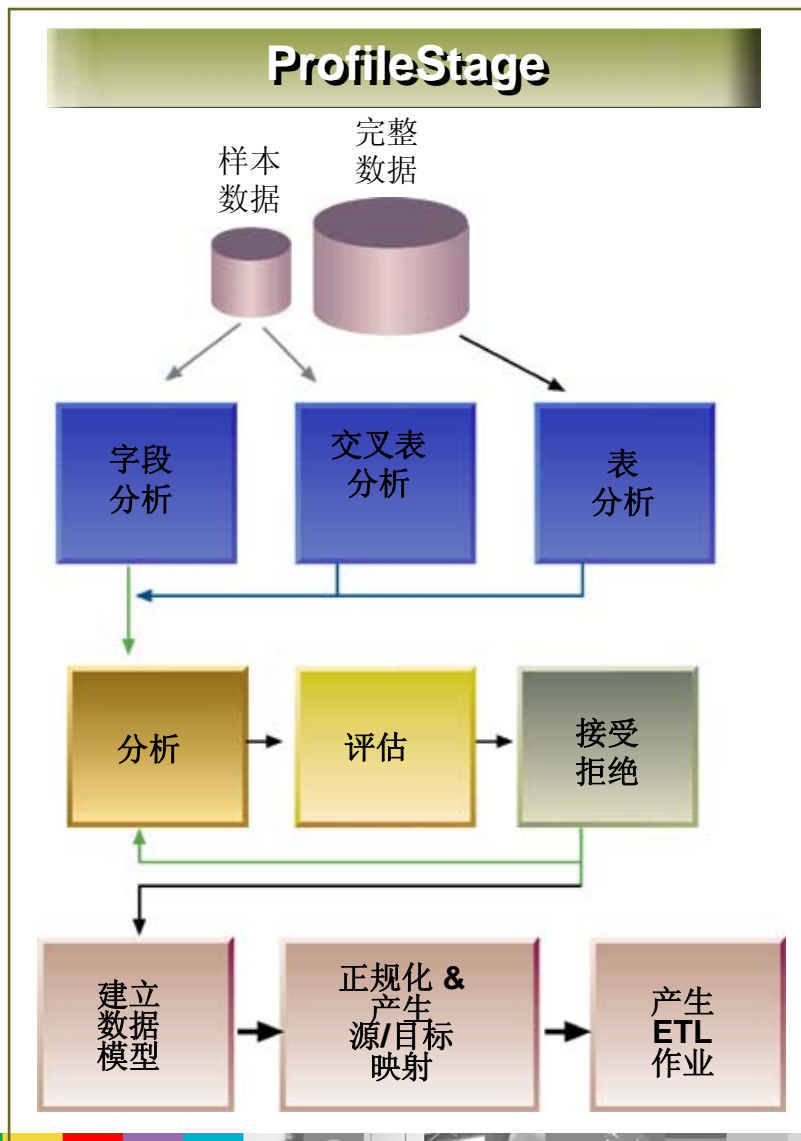
## 关键问题

- 相同数据在不同源系统中的数据格式和数据值不同
- 数据不完善或缺少文件
- 数据源不是静态的并且在没有任何警报的情况下频繁地改变

## 可供选择的方法

- 简单忽略，导致了生产错误，将引起严重的风险。
- 手工的、劳动力密集的、浪费资源的处理

# 数据探索: ProfileStage



## 为什么选择ProfileStage?

- 鉴别数据内容、结构和质量
- 在开发之前确认源数据是否支持目标需要
- 建立最佳数据模型& 自动映射数据源到目标
- 建立DataStage整合步骤
- 多CPU透明并行开发
- 多国语言支持
- 商务智能的元数据转换

## 业务收益

- 对数据源的分析减少70%的时间
- 及时发现问题，降低数据源分析的成本
- 自动定义ETL作业，不需要程序调试时间
- 均衡有价值的遗留数据，增加访问应用的可靠性

# The IBM WebSphere 信息集成平台

## Service Oriented Architecture

理解



WebSphere ProfileStage™

WebSphere MetaStage™

清洗



WebSphere QualityStage™

- 标准化、匹配、更正数据
- 清除重复数据
- 数据抽象概念鉴别
- 数据质量保障

元数据管理

并行处理

数据



连接











内容

WebSphere Information Integrator Replication Edition

WebSphere Information Integrator Event Publisher Edition



# 数据准备

Data Sources	
	Kent Fried Chick
	KFC
	Molly Talber DBA KFC
	Kentucky Fried Chicken
	Kentucky Fried
	Mrs. M. Talber
	John & Molly Talber
	Talber, KFC, ATIMA

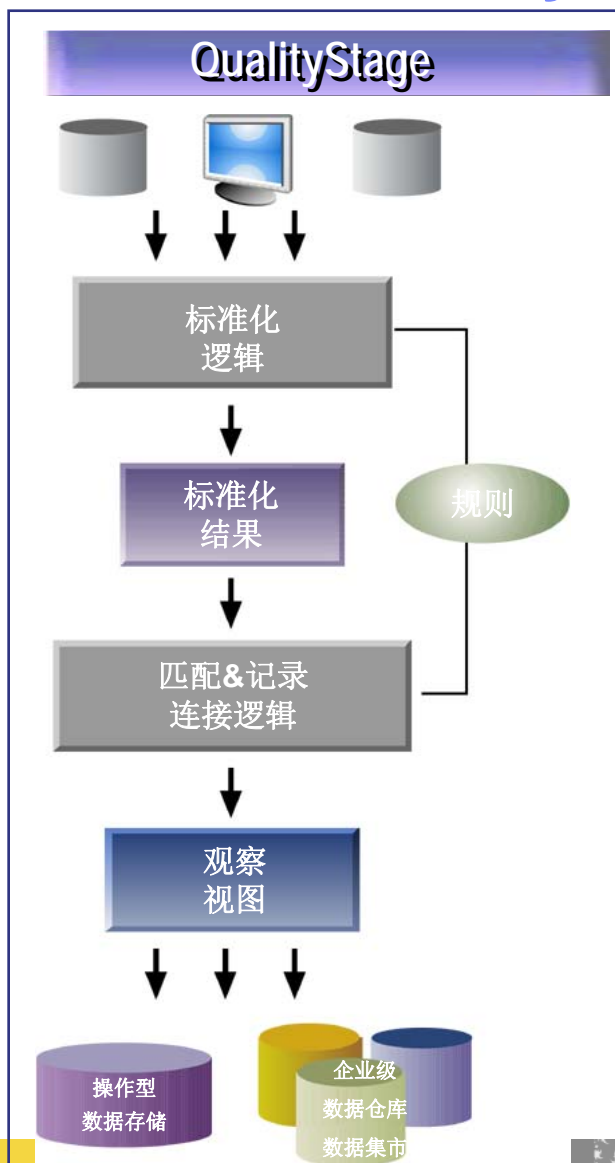
## 关键问题

- 数据的重复性问题
- 在字段中存储意外的值
- 字段发展 – 用于多种目的
- 数据不符合规则 – 尤其是来自于外部的数据

## 可选择方法

- 拒绝 & 拖延 – 问题理解错误和忽略问题直到已经太晚；加载和推翻
- 手工编码； 进行例外处理
- 简单化黑盒子清洗应用， 缺少灵活性和实际使用价值

# 数据准备: QualityStage



## 为什么选择 QualityStage?

- 更加准确、完整地对数据进行更加快速的匹配
- 最灵活的分析/标准化能力
- 处理复杂的自由格式的字段
- 全球 – 可以确认**200**多个国家地址，完全的 **NLS/Unicode** 支持
- 透明的并行能力，充分利用多**CPU**
- 面向**BI**的元数据转换

## 商业优势

- 一致的、可靠的竞争信息
- 不断增加的收入
  - ▶ 在企业应用&BI中，更快、更完全的ROI，改善营销& 客户服务
- 不断减少的成本
  - ▶ 高效率操作、预防过失、不断调整



# The IBM WebSphere 信息集成平台

## Service Oriented Architecture

理解



WebSphere. Profile

WebSphere. MetaS

- 规范化、转换、丰富数据
- 唯一键管理
- 开发装载作业
- 加工规则和参照数据管理

连接

转换



WebSphere. DataStage®

WebSphere. DataStage® TX

联邦



WebSphere. Information Integrator Standard Edition

WebSphere. Information Integrator Classic Federation

WebSphere. Information Integrator Content Edition

元数据管理

并行处理

数据



连接



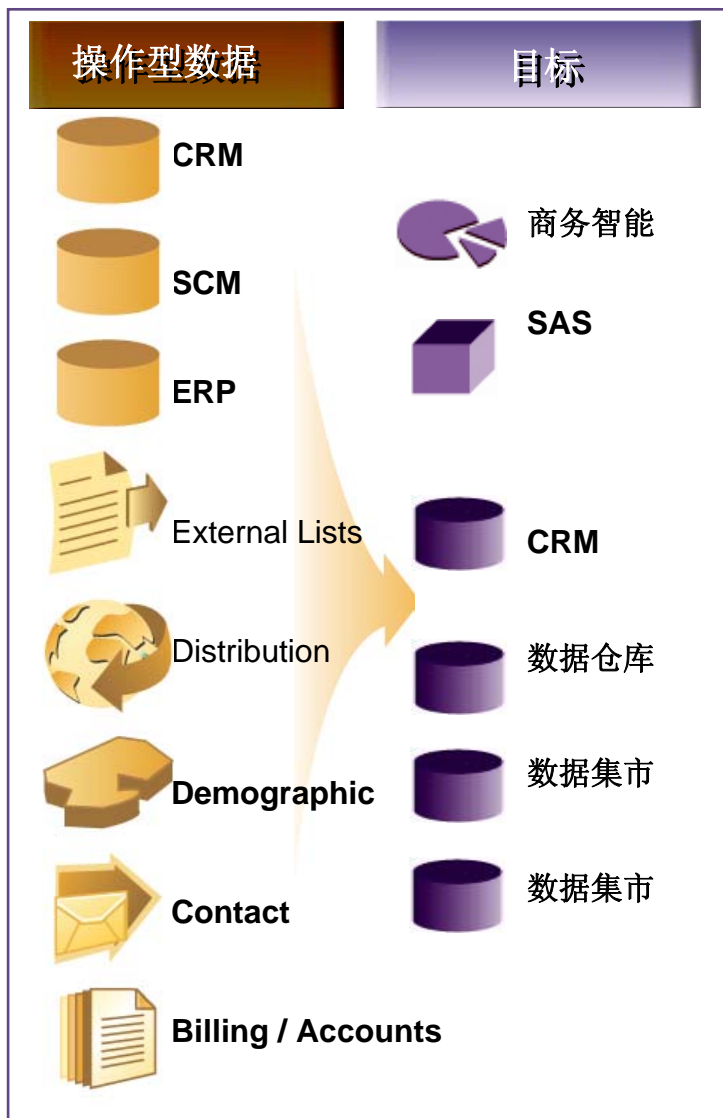
内容

WebSphere. Information Integrator Replication Edition

WebSphere. Information Integrator Event Publisher Edition



# 数据转换



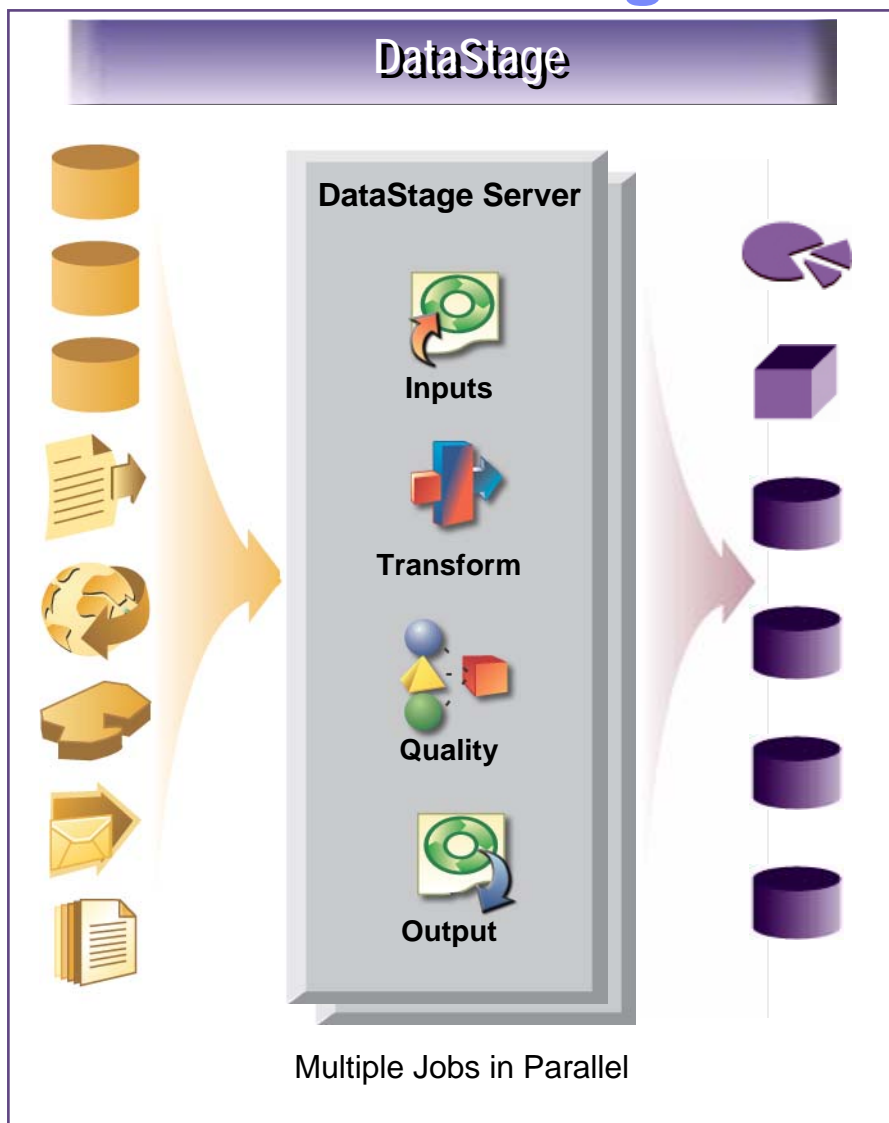
关键问题:

- 需求总是在变化
- 传统的代码方式是耗时、不能扩展和维护费用高的
- 通过自动化和流程化处理来提高处理能力
- 需要不断调整来符合需求，需要对数据进行整合

可选择的方法:

- 使用手工、费力费资源的处理
- 花时间和费用但实施的是不能扩展的数据整合
- 重新建立相同的转换逻辑和元数据通过不同工具访问

# 数据转换: DataStage



为什么选择DataStage?

- 图形化的、非代码的设计环境
- 可扩展的转换平台，均衡已存在的逻辑
- 建立在可升级和调整的处理平台上
- 企业级平台提供被证明的**ROI**
- 管理开发到配置的无缝升级

商业优势

- 卓越的产品功能
- 准确、准时、一致的产品发布
- 统一的访问规则应用
- 数据和处理可检查

# 例子: 从上到下的GUI方法

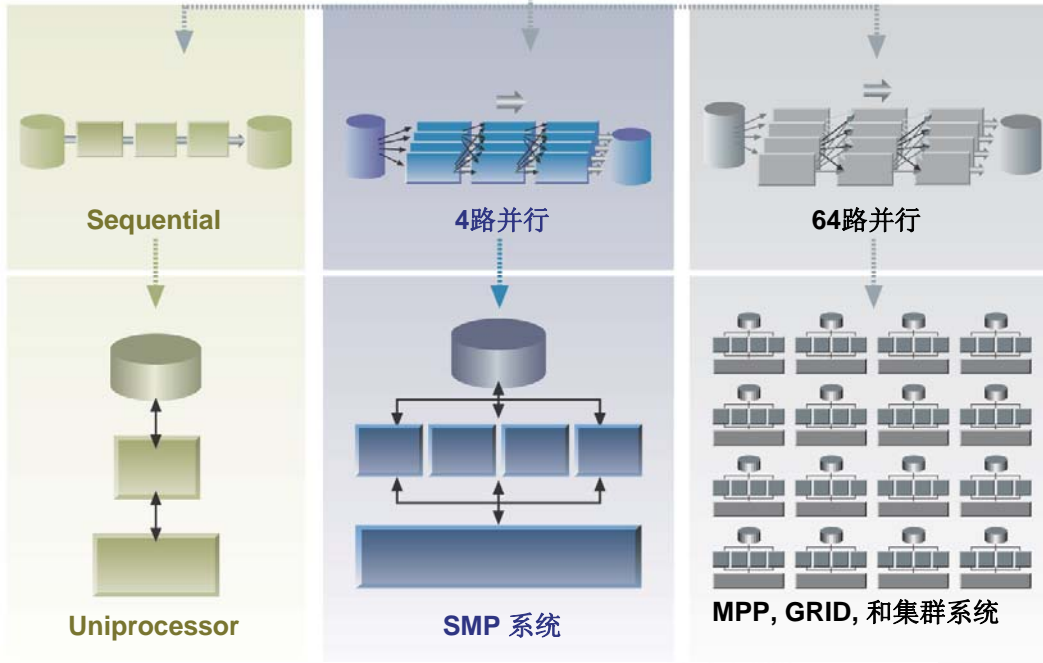
The image displays two overlapping windows of the Ascential DataStage Designer software. The top window shows a 'Server' view with a tree structure on the left containing folders like 'Jobs', 'Examples', and 'CompanyDemo'. A yellow box highlights the text: 'Typical DataStage \* Visual process flow \* Self-Documenting'. The bottom window shows a 'Sequence' view with a complex workflow diagram. The workflow starts with 'WaitForFile' and 'FTPget', leading to 'ToWPSB08M1' and 'ToWPSB08P2', which then connect to 'WPSB08M1' and 'WPSB08P2'. These lead to 'Sync1' and 'ToSync', which connect to 'ToCleanse' and 'cond\_CLEANSE'. From 'cond\_CLEANSE', there are two paths: 'no\_Clean' leading to 'Sync2' and 'yes\_Clean' leading to 'WPSB08Q1'. 'Sync2' connects to 'ToWPSB08L1', and 'WPSB08Q1' connects to 'FromWPSB08Q1' and 'Notify\_Complete'. A red box at the bottom of the workflow diagram is titled 'Exception Error Processing' and shows a flow from 'Exceptions' to 'CheckError', then to 'cond\_Error', which branches to 'ToMgr' (leading to 'Notify\_Manager') and 'ToAdmin' (leading to 'Notify\_Admin').

# 性能: 并行执行

集合应用: 一个数据流视图



应用执行: 顺序或并行

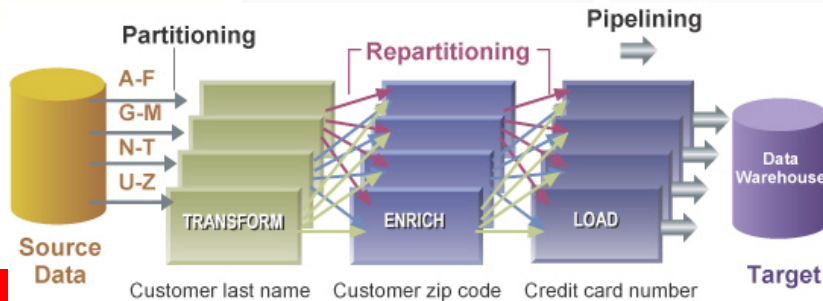


## 并行处理模式

- 顺序设计, 并行配置
- 线性扩展
- 动态数据分区和可变数据分区
- 支持 **SMP**, 集群, **GRID**, 和 **MPP**并行 平台
- 支持并行RDBMS, 包括 **IBM DB2 UDB**, **Oracle**, & **Teradata**
- 非代码级并行

## 商务优势

- 通过高产品化, 符合商务需求
- 最佳硬件应用
- 灵活的执行选项



# The IBM WebSphere 信息集成平台

## Service Oriented Architecture

理解



WebSphere. ProfileStage™

WebSphere. MetaStage™

- 异构数据连接
- 透明访问
- 实时数据访问
- 异构数据复制
- Q复制

联邦



WebSphere. Information Integrator Standard Edition

WebSphere. Information Integrator Classic Federation

WebSphere. Information Integrator Content Edition

元数据管理

并行处理

数据



连接



内容

WebSphere. Information Integrator Replication Edition

WebSphere. Information Integrator Event Publisher Edition



# 信息集成体系结构

## 应用服务平台

- 全局统一的数据视图
- 丰富的特色功能
- 灵活的访问方式
- 对于标准的支持

存取

Web服务

XML

SQL

功能

存储

查询

联邦

缓存

转换

复制

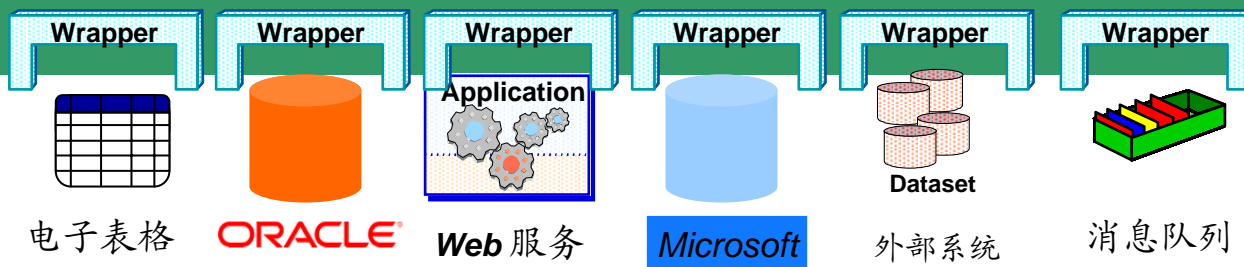
DB2 Information Integrator

基础

高可用

易管理

可扩展

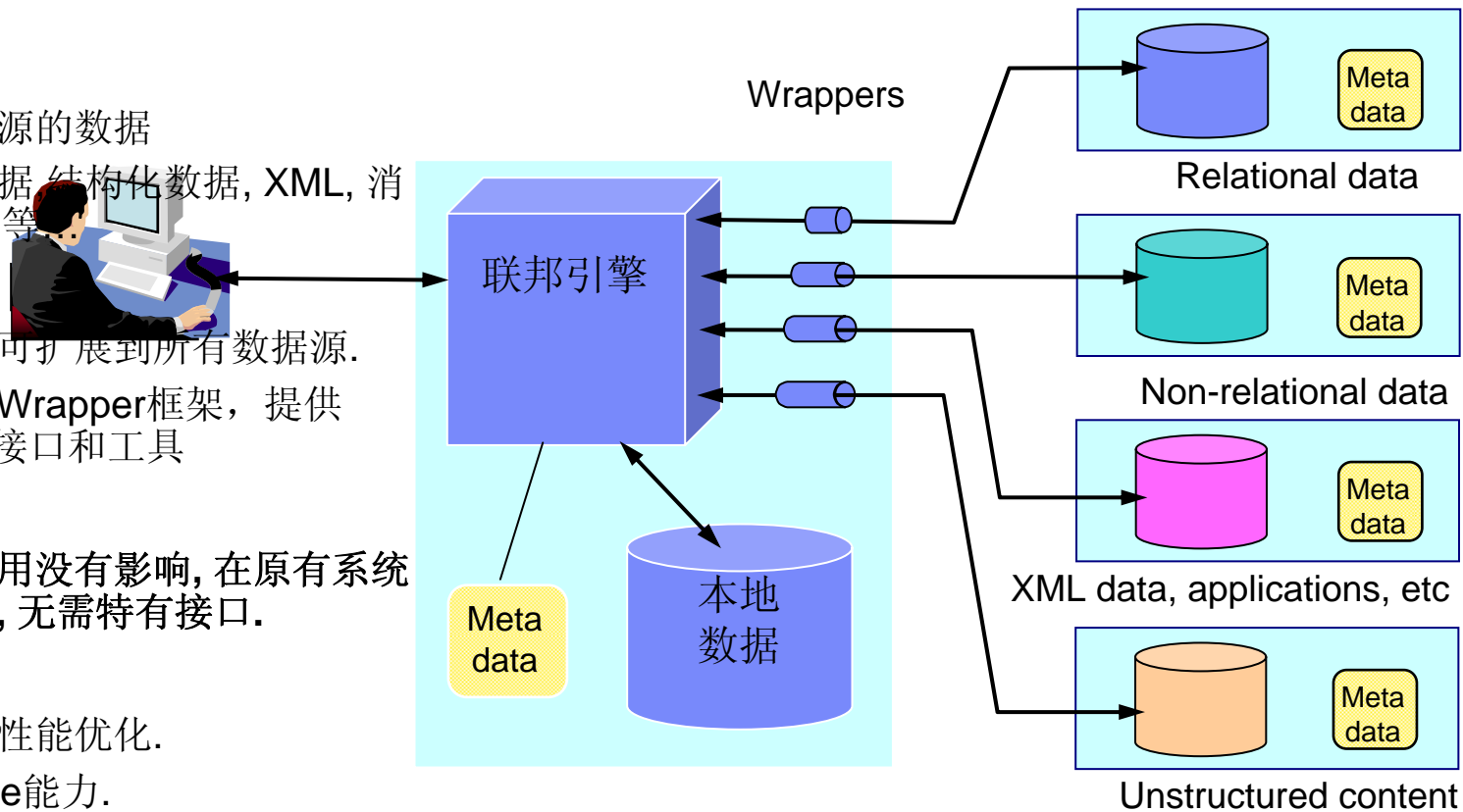




## 数据联邦提供单点数据访问

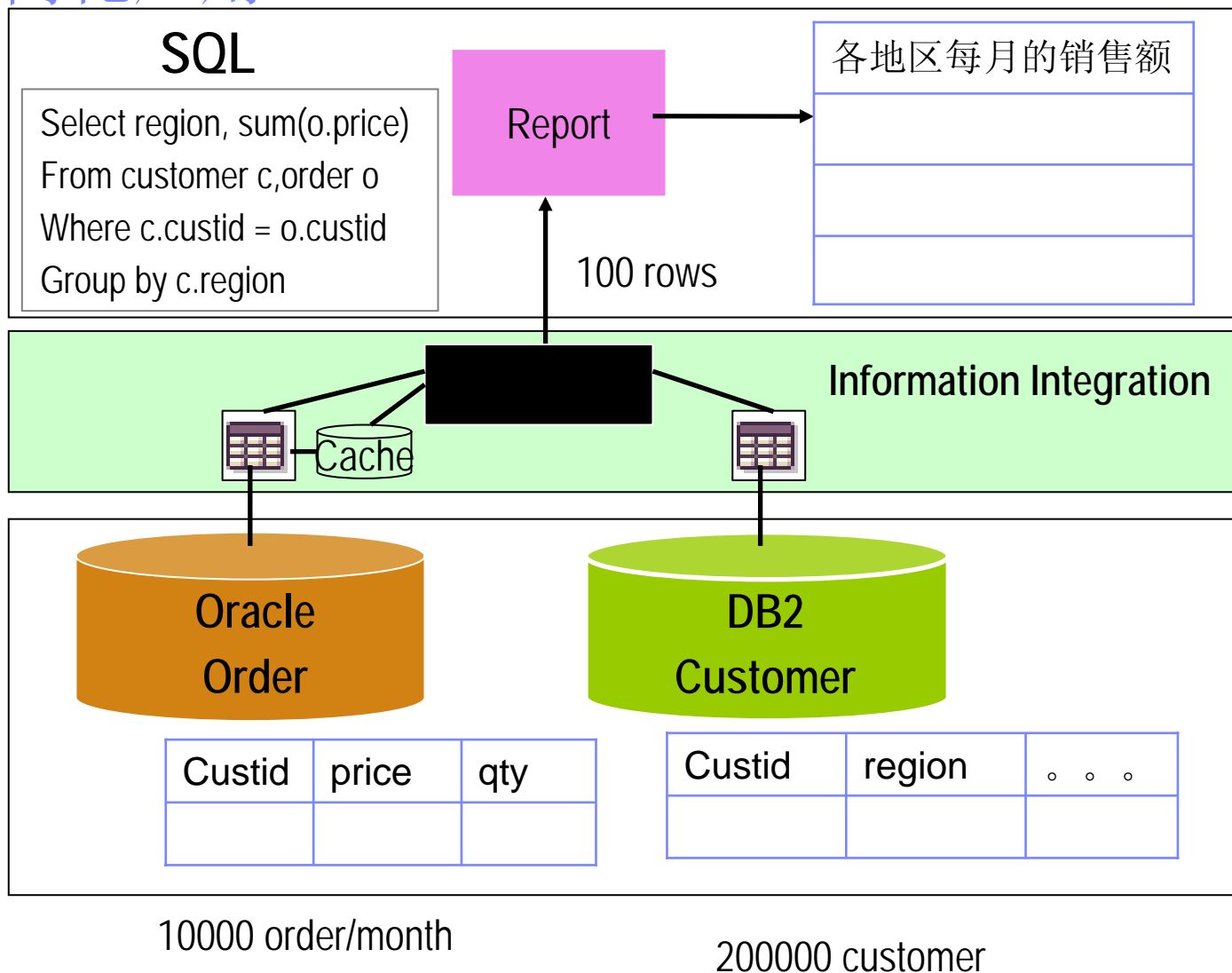
数据源

- 透明
  - ▶ 对应用而言，所有数据如同在同一数据库中。
- 异构
  - ▶ 集成各类数据源的数据
  - ▶ 包括关系型数据、结构化数据、XML、消息队列、Web，等
- 可扩展
  - ▶ 联邦技术几乎可扩展到所有数据源。
  - ▶ 基于开放式的Wrapper框架，提供Wrapper开发接口和工具
- 自治
  - ▶ 对原有系统应用没有影响，在原有系统无需安装软件，无需特有接口。
- 性能优化
  - ▶ 分布式的查询性能优化。
  - ▶ 提供数据cache能力。



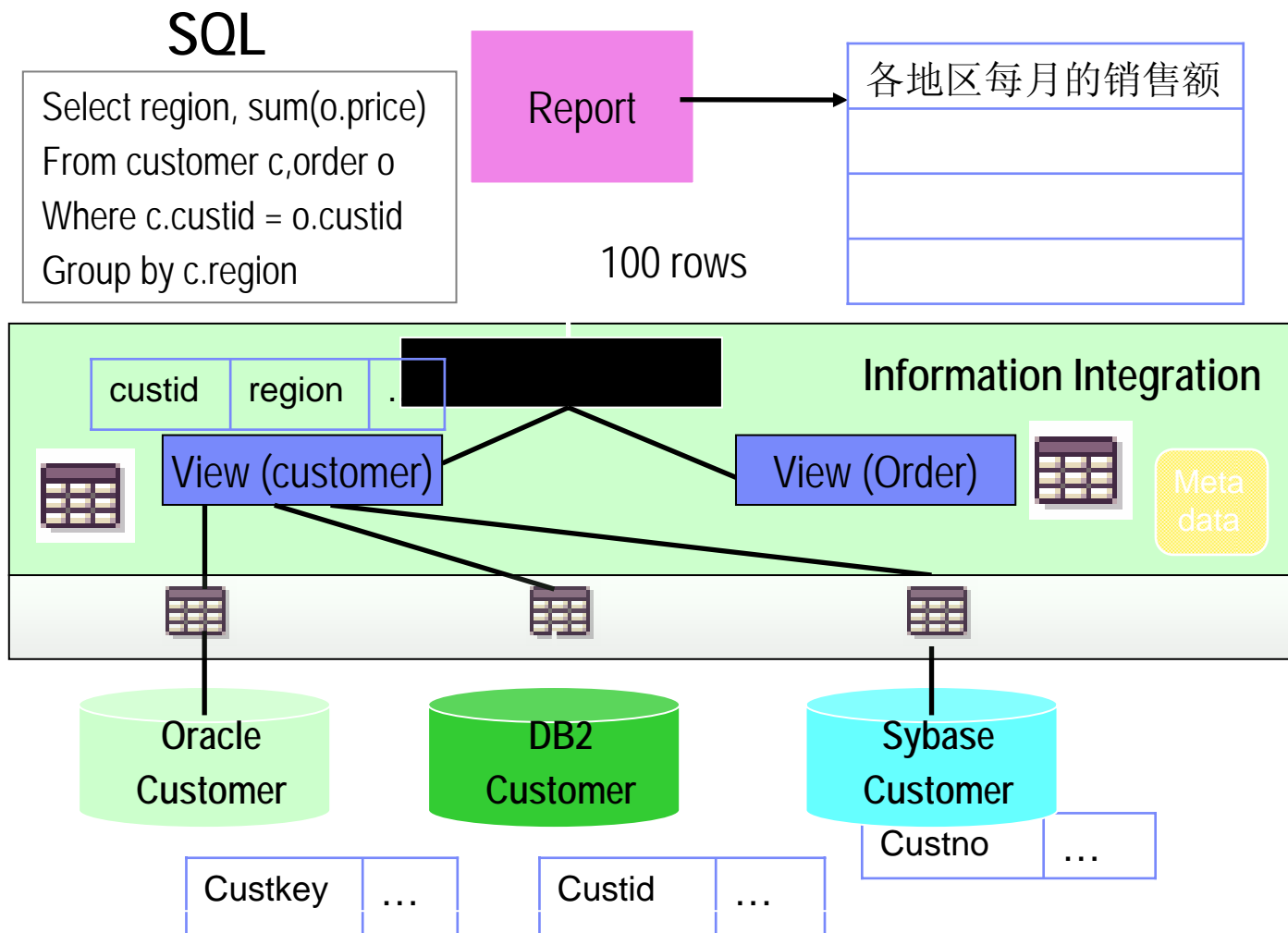
# 信息集成技术简化应用

- 性能优化
- 系统级的控制

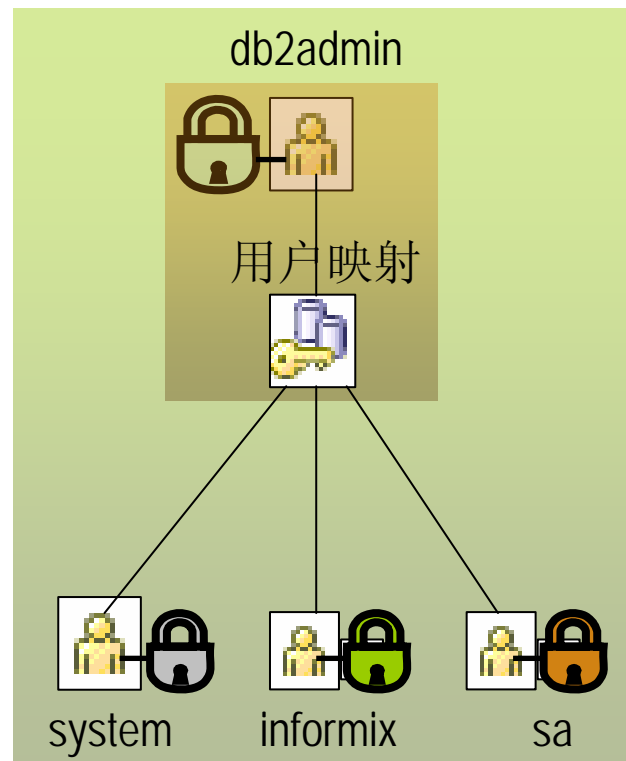
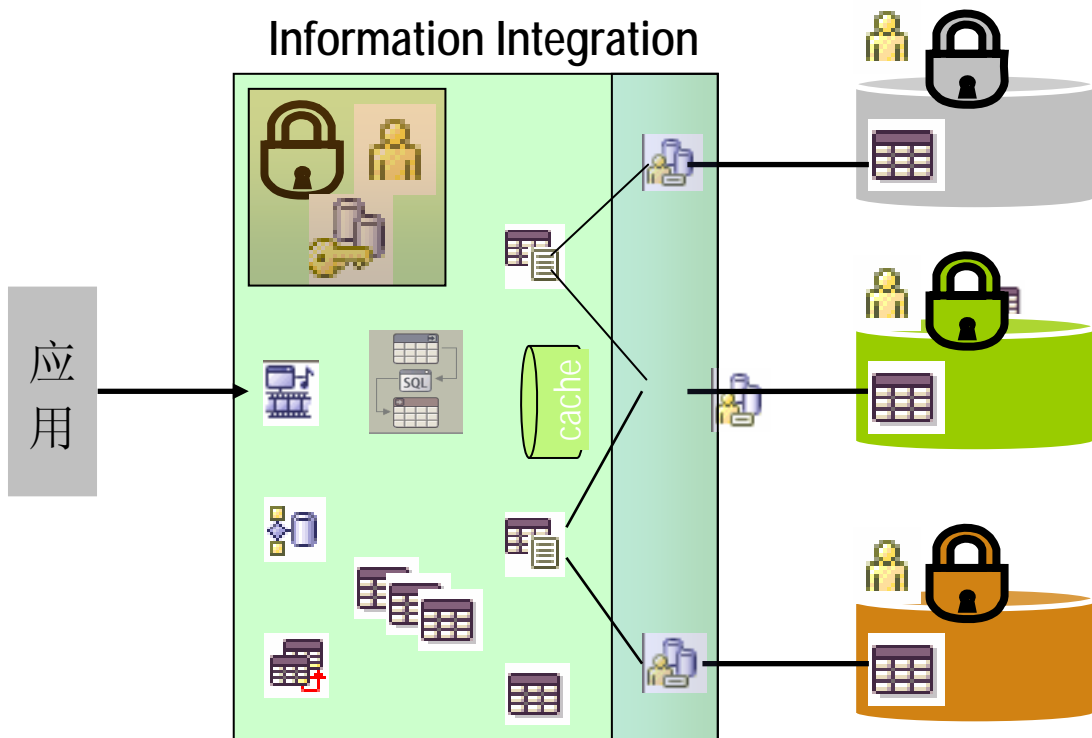


# 信息集成提供统一的信息视图

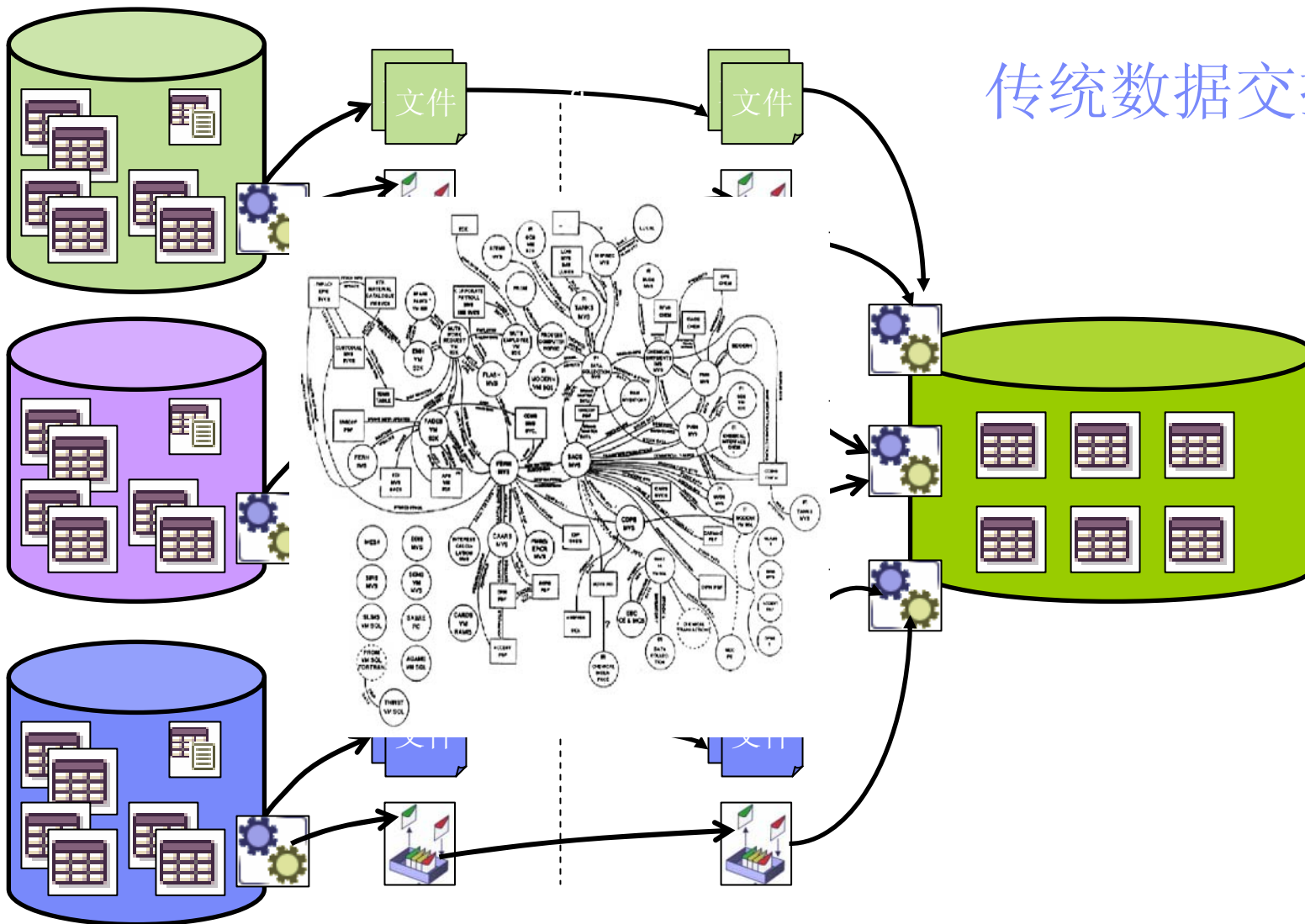
- 现有的应用不需要修改
- 通过修改元数据，新的数据源可以方便的增加进来



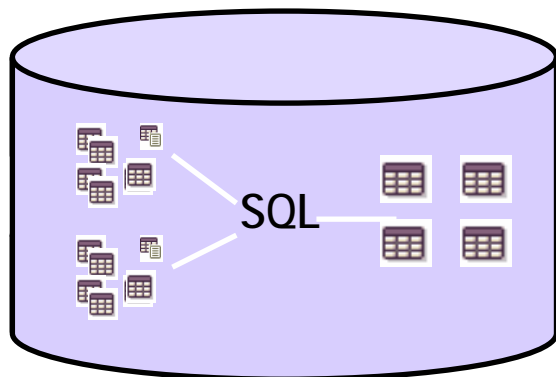
# 信息集成的数据安全



# 传统数据交换方式



# DB2 II简化数据迁移



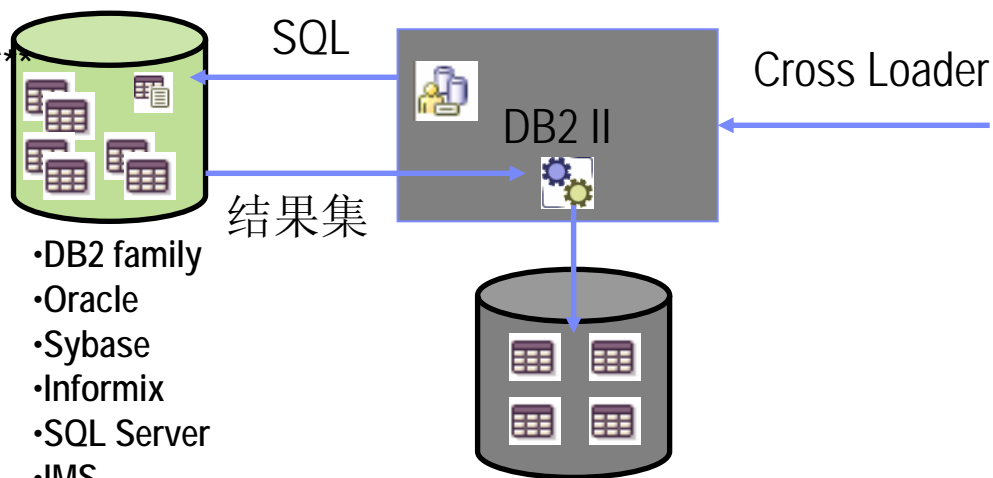
```
EXEC SQL
DECLARE C1 CURSOR FOR
  SELECT JOB,MAX(SALARY) AS MAX_SAL,MIN(SALARY) AS MIN_SAL
  FROM DSN8710.EMP
  GROUP BY JOB
ENDEXEC

LOAD DATA REPLACE
INCURSOR C1
INTO TABLE EMPSUMMARY
```

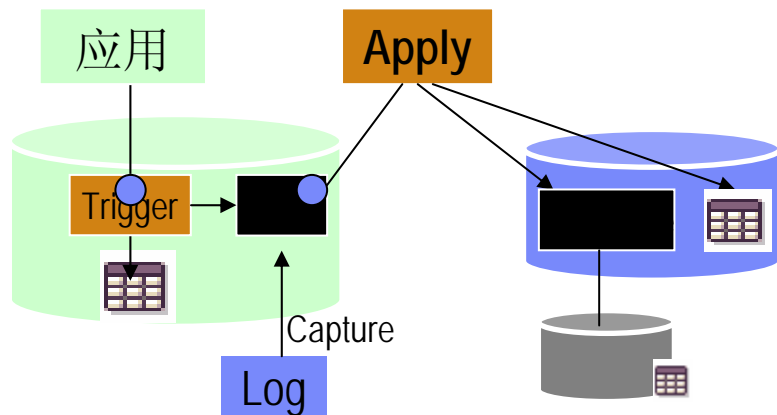
Insert into **target** select \* from **source** where \*\*

## •Cross Loader高速数据跨库加载

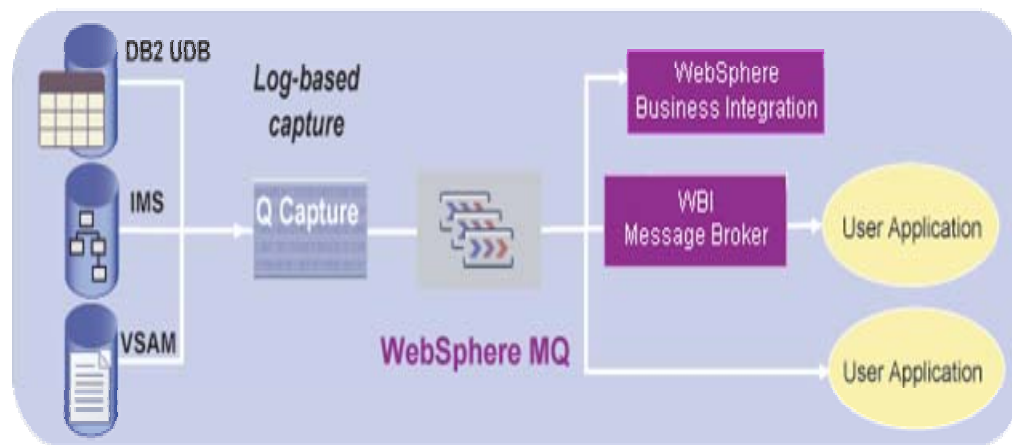
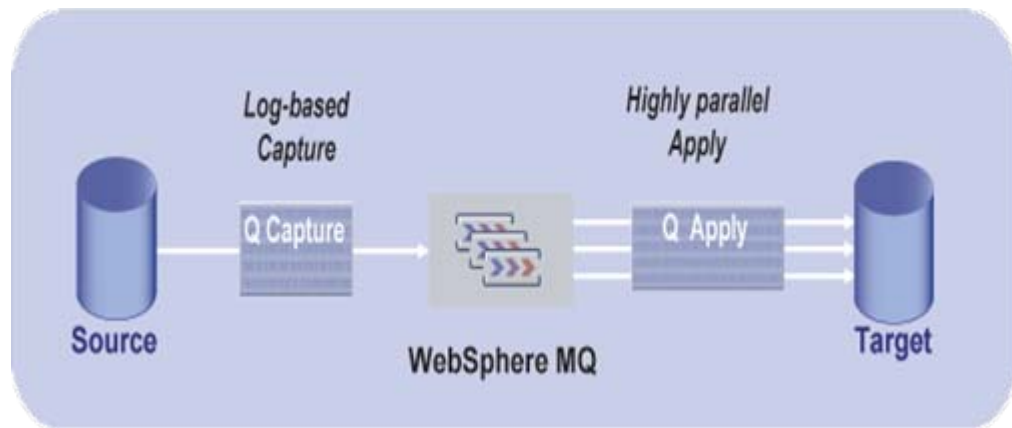
- 数据源由SQL定义,提供灵活性
- 数据无需中间缓存
- 在目标端数据加载可不记日志
- 实测速度可达50GB/小时



# 异构数据复制和 Q-replication



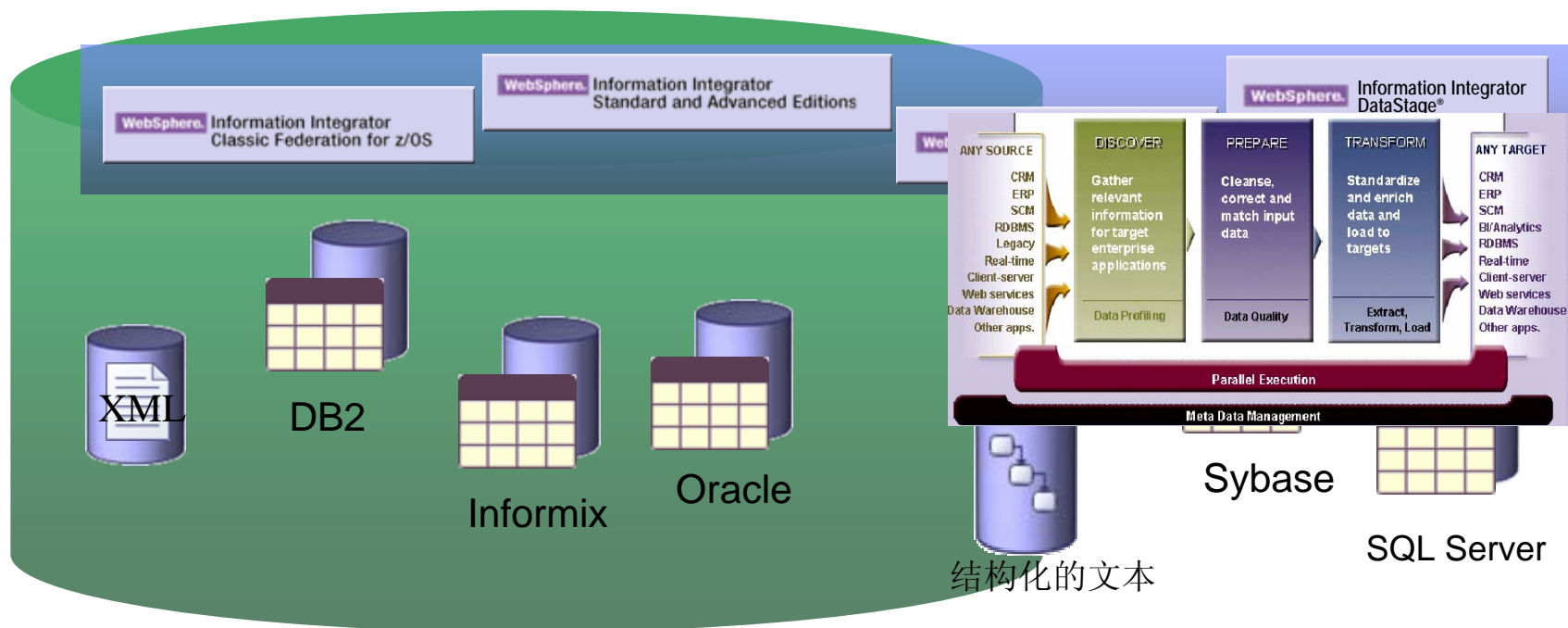
- 异构数据复制
- 增量数据的识别
- 增量数据的迁移
- 复制流程的管理和监控





# 企业统一信息平台

- 打破信息壁垒：虚拟化异构存储环境、提供网格计算能力。
- 信息交换与共享：在虚拟的数据库中实现分布式查询和数据的移动和交换。
- 数据抽取与转换：在统一的平台实现数据的抽取、清洗、转换以及元数据管理。



异构数据，自由发布—Information Services

# IBM 数据整合解决方案的优势

## 企业级平台

- 可扩展的技术平台
- 大数据量的并行处理
- 处理多数据源到多数据目标的数据 ETL
- 流程控制
- 高性能的数据处理
- 满足几何级大数据量增长挑战而不需要重新设计体系架构
- 处理多数据源到多数据目标的数据 ETL, 包括 SAP / PeopleSoft / Oracle Finance / SIEBEL 等企业级系统
- 复杂的转换逻辑

## 使用工具的好处

- 缩短企业应用开发时间
- 提高整体的数据质量和项目推行质量
- 降低风险和成本
- 加快预期的投资回报



# Websphere IIS 全球客户



# 中国区主要客户

- 中国银行总行；广东分行；香港中银数据仓库
- 交通银行总行
- 中国建设银行分行-历史数据分析、客户综合信息分析
  - ▶ 海南建行、四川建行
  - ▶ 河北建行、陕西建行
- 浦东发展银行  
SAP ERP、SWFTI实时整合
- 光大银行：国际业务、信贷分析
- 民生银行  
企业级集中数据平台、SAP ERP
- 农信、平安保险…
- 河北地税…
- China Mobile - 决策经营分析；网管优化  
上海、广东、浙江、江苏、天津、江西、福建、山东、吉林、辽宁、安徽、贵州、四川、云南、重庆、北京、总公司 等20 / 31个用户
- China Unicom - 运营支撑：上海、陕西、北京、浙江、辽宁联通. . .
- CNC：天津、河北； 电信：福建、重庆、安徽
- 中国远洋运输集装箱  
上海通用汽车
  - ▶ Center data Storage ; ODS
  - ▶ SAP; Oracle; Seibel ;XML …
- 苏州西门子;TOM.COM;吉林省烟草专卖局  
南洋烟草, 中国烟草销售公司, 广东中烟工业公司

# 谢谢!

