



您的信息 您的智慧

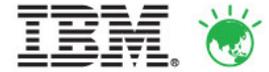
2011 IBM 信息管理与业务分析论坛

IBM 数据仓库软件和ISAS7700

周雄志

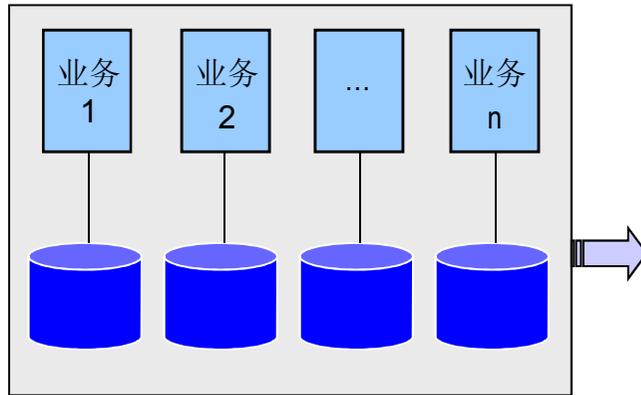
IBM高级工程师





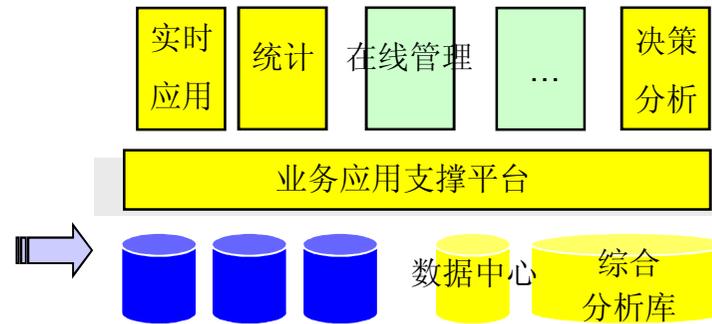
企业信息利用的变革

现状



- 应用系统垂直分割
- 数据存储相对独立
- 跨部门信息共享困难
- 业务关联日益密切

目标



- 面向要素组织数据
- 数据深度整合支持业务持续发展
- 数据深度利用服务决策
- 数据利用结果促进业务不断优化

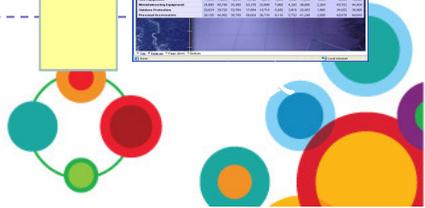
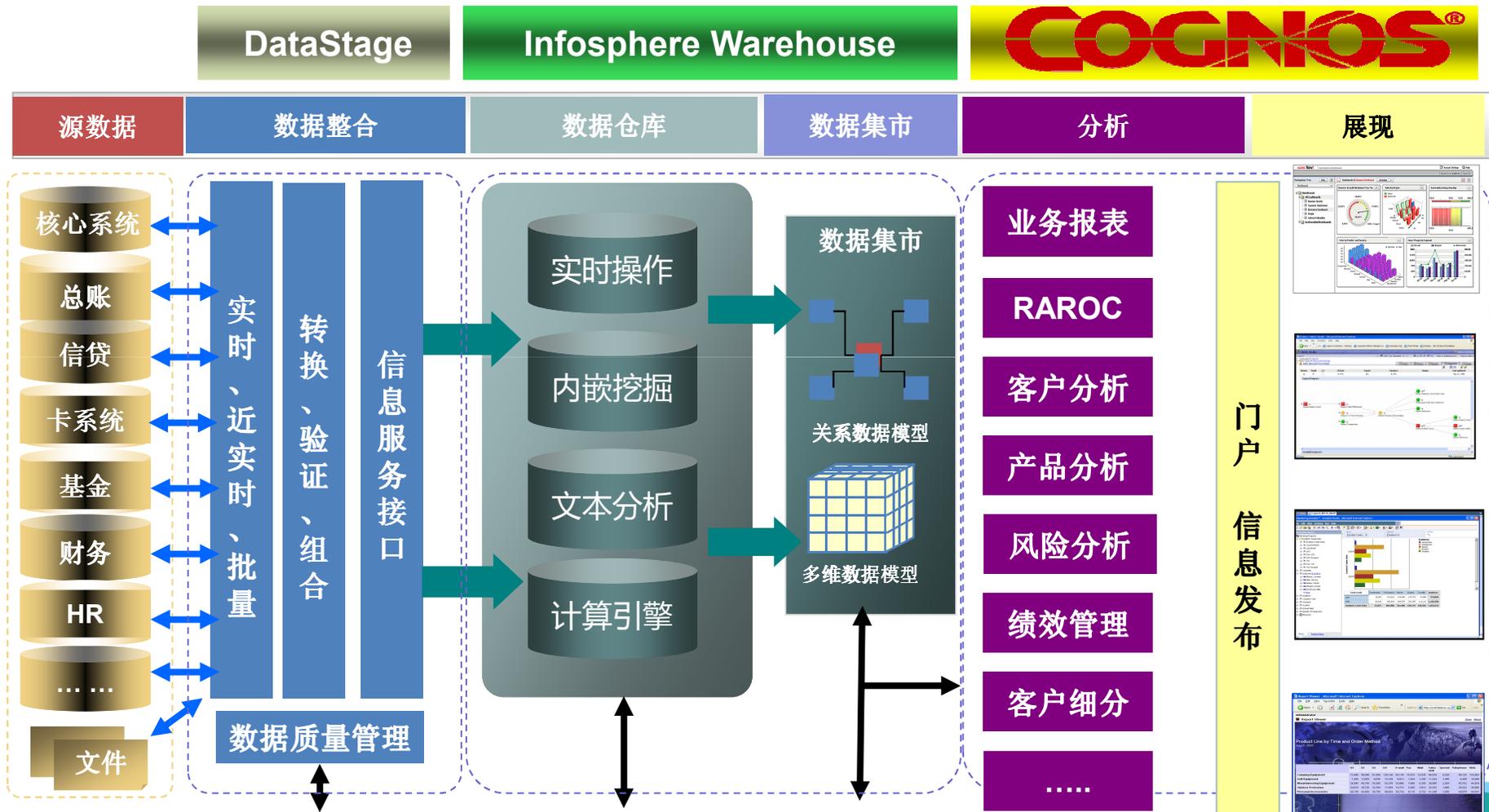
业务自动化

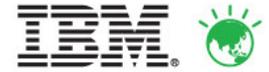


业务持续优化

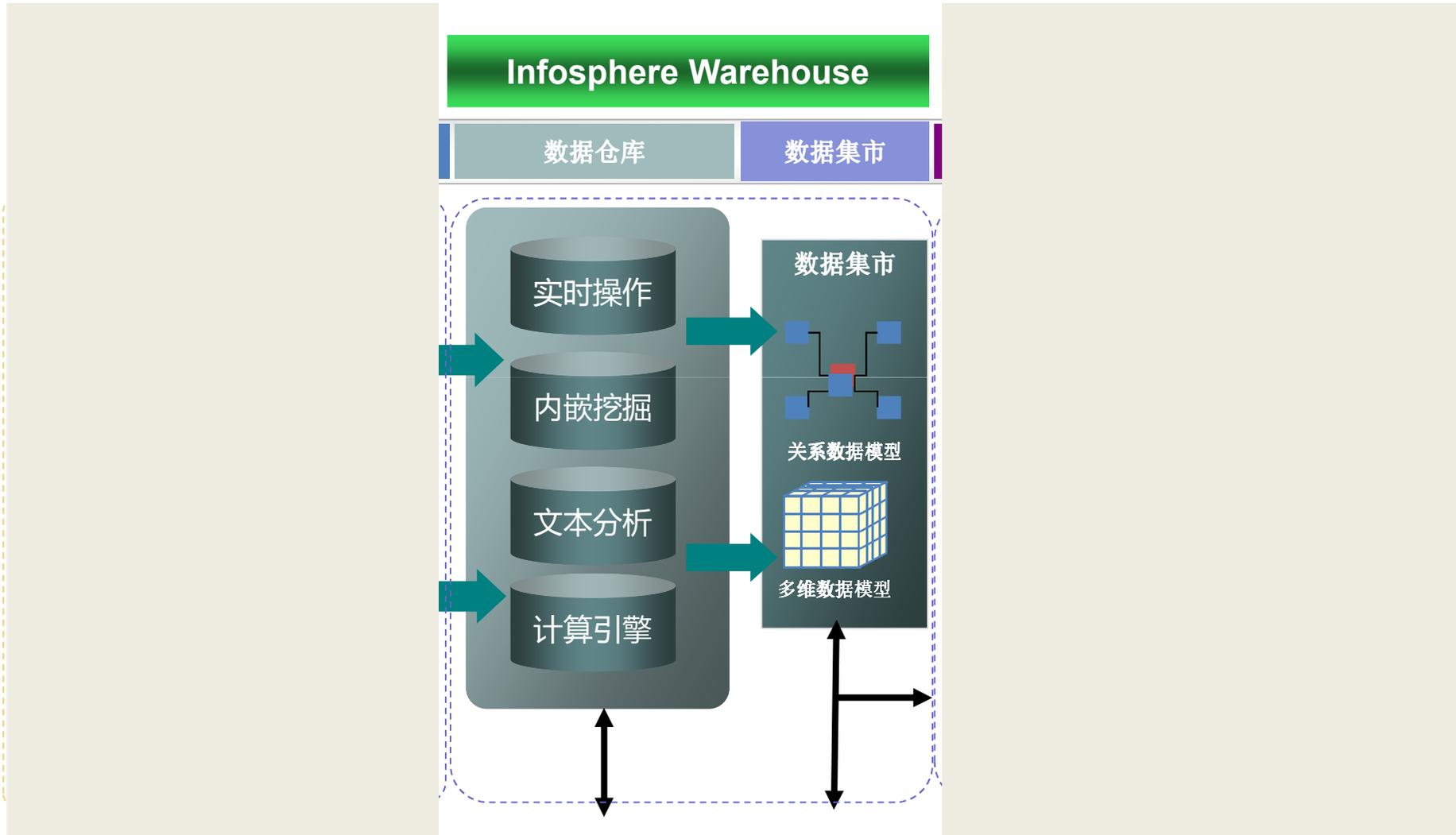


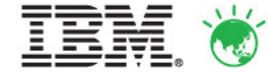
数据仓库系统软件架构



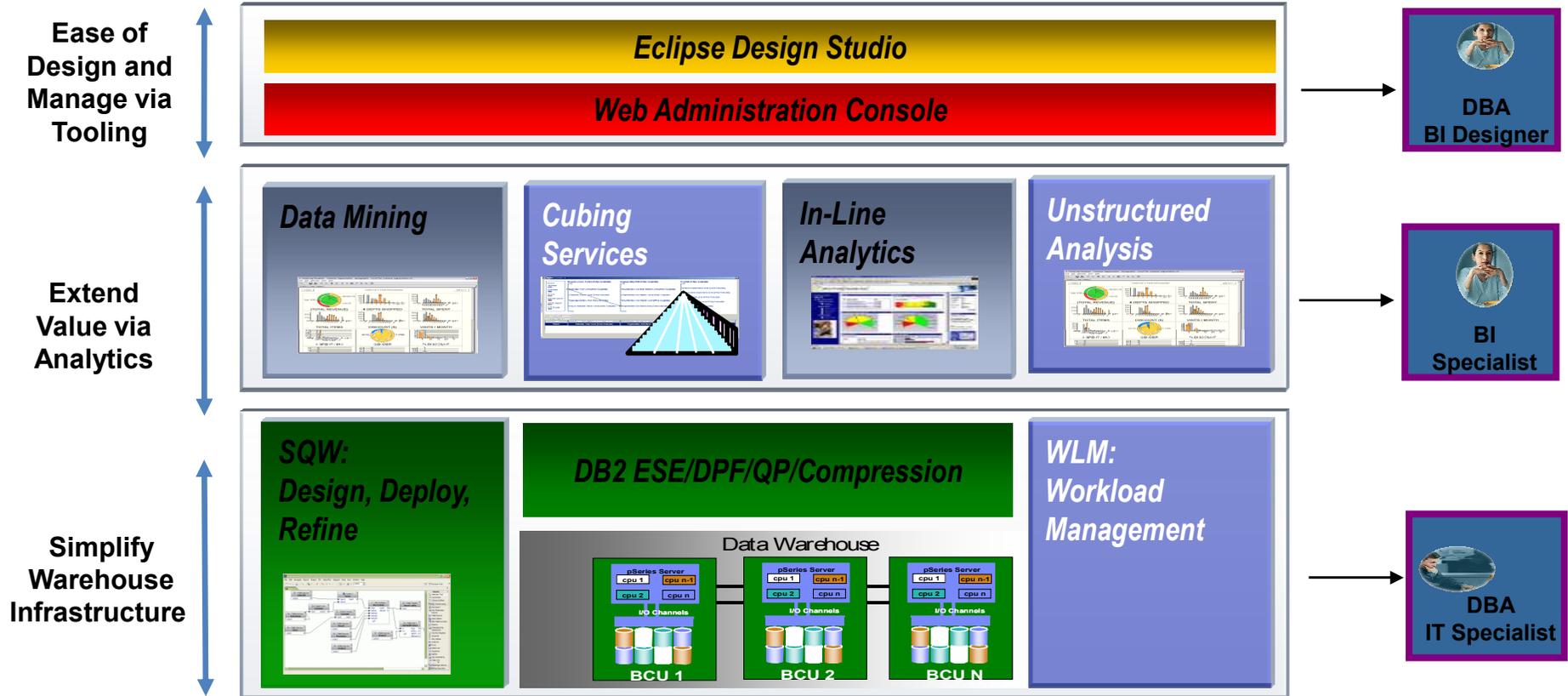


数据仓库系统软件架构





InfoSphere Warehouse

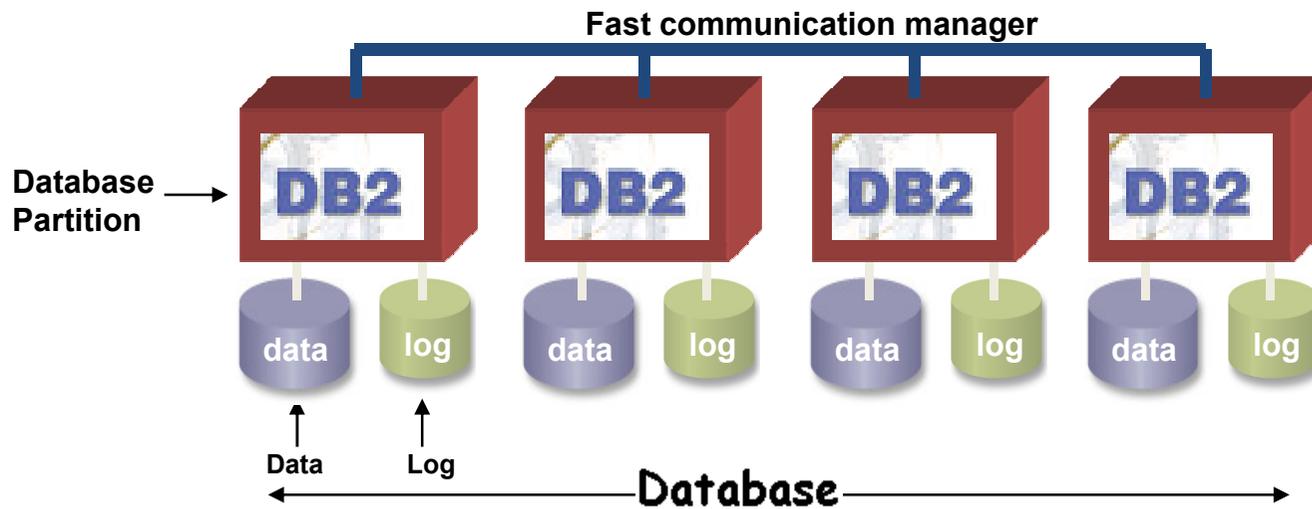




DB2 – 非共享体系结构

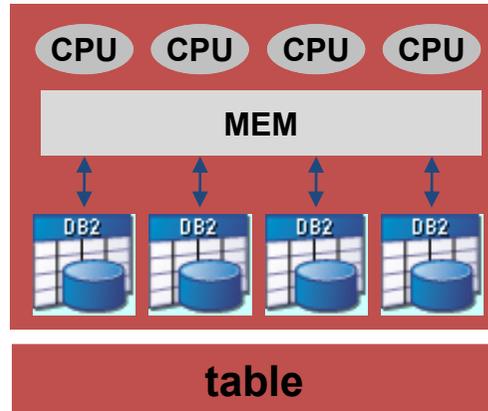
分区数据库模型

- 数据库被分成多个分区
- 数据库分区运行在各个节点上
- 每个数据库分区具有自己的资源 (Engine, LogMg., LockMg., Caches, etc.)
- 数据库协调所有分区进行并行处理
- 对用户和应用来看,是一个单独的系统



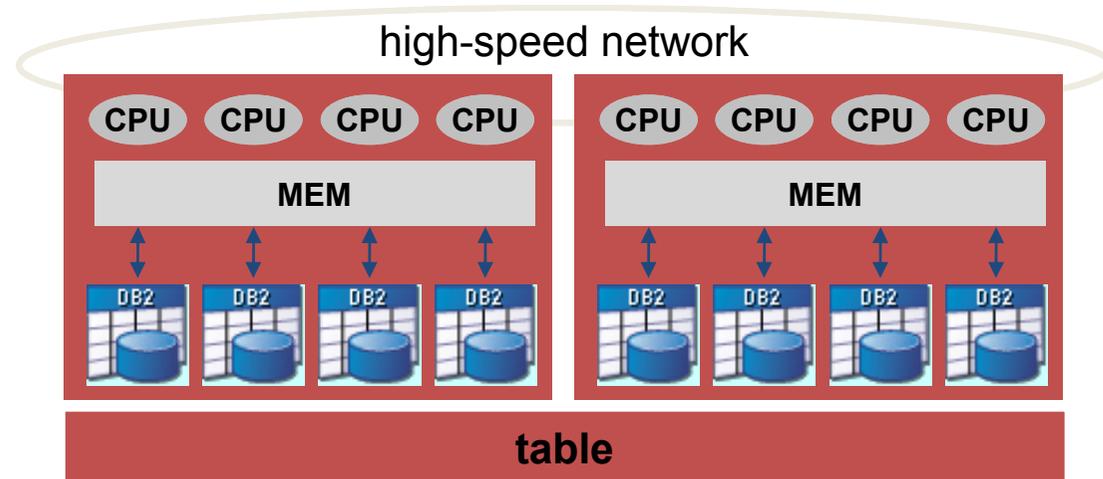


DB2 并行机制



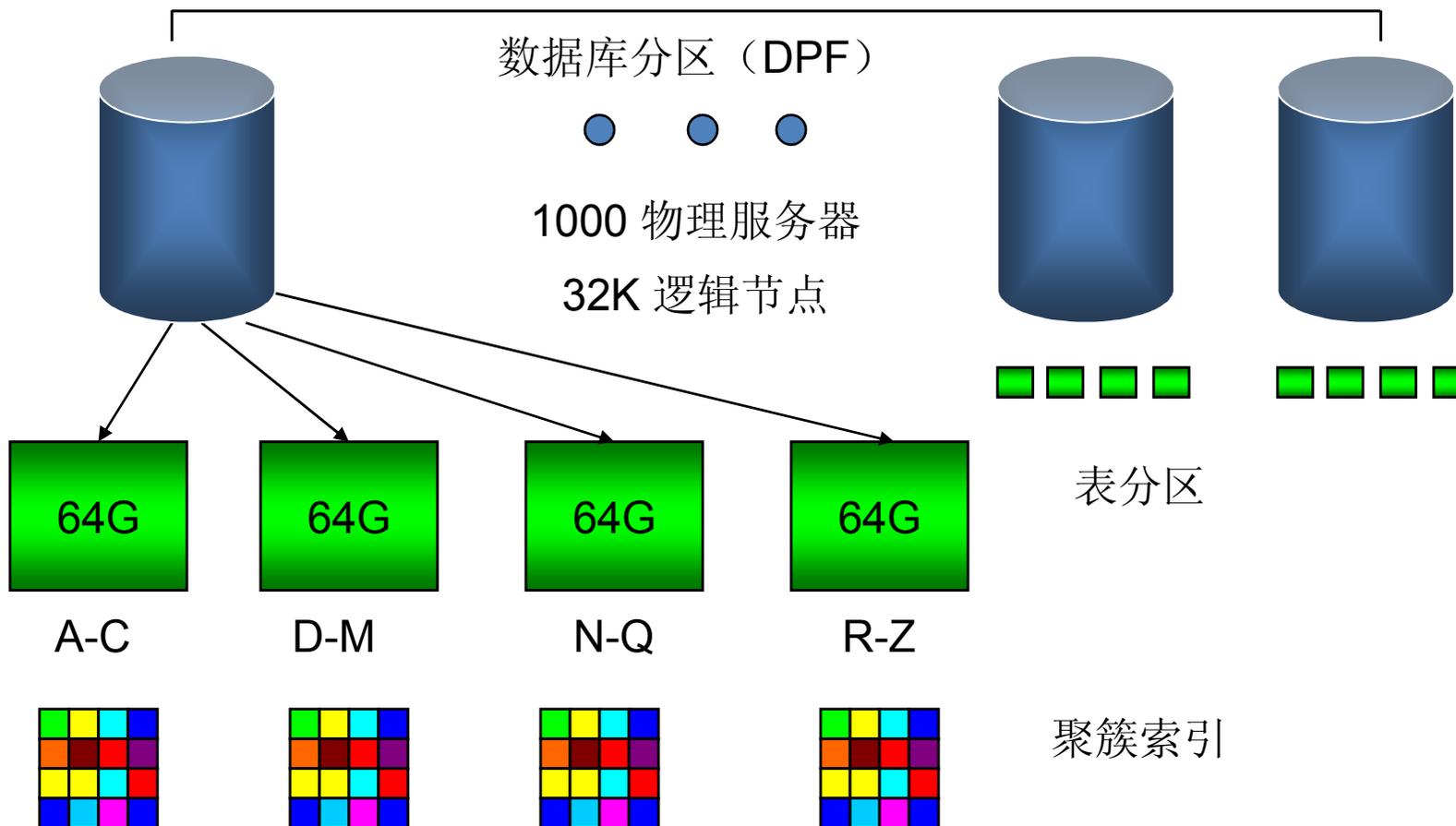
SMP服务器

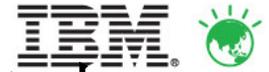
MPP集群





DB2混合分区极大提升数据仓库性能



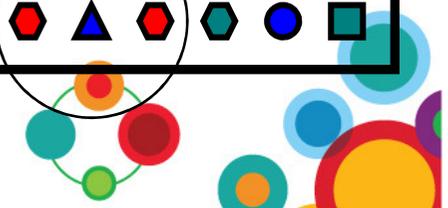
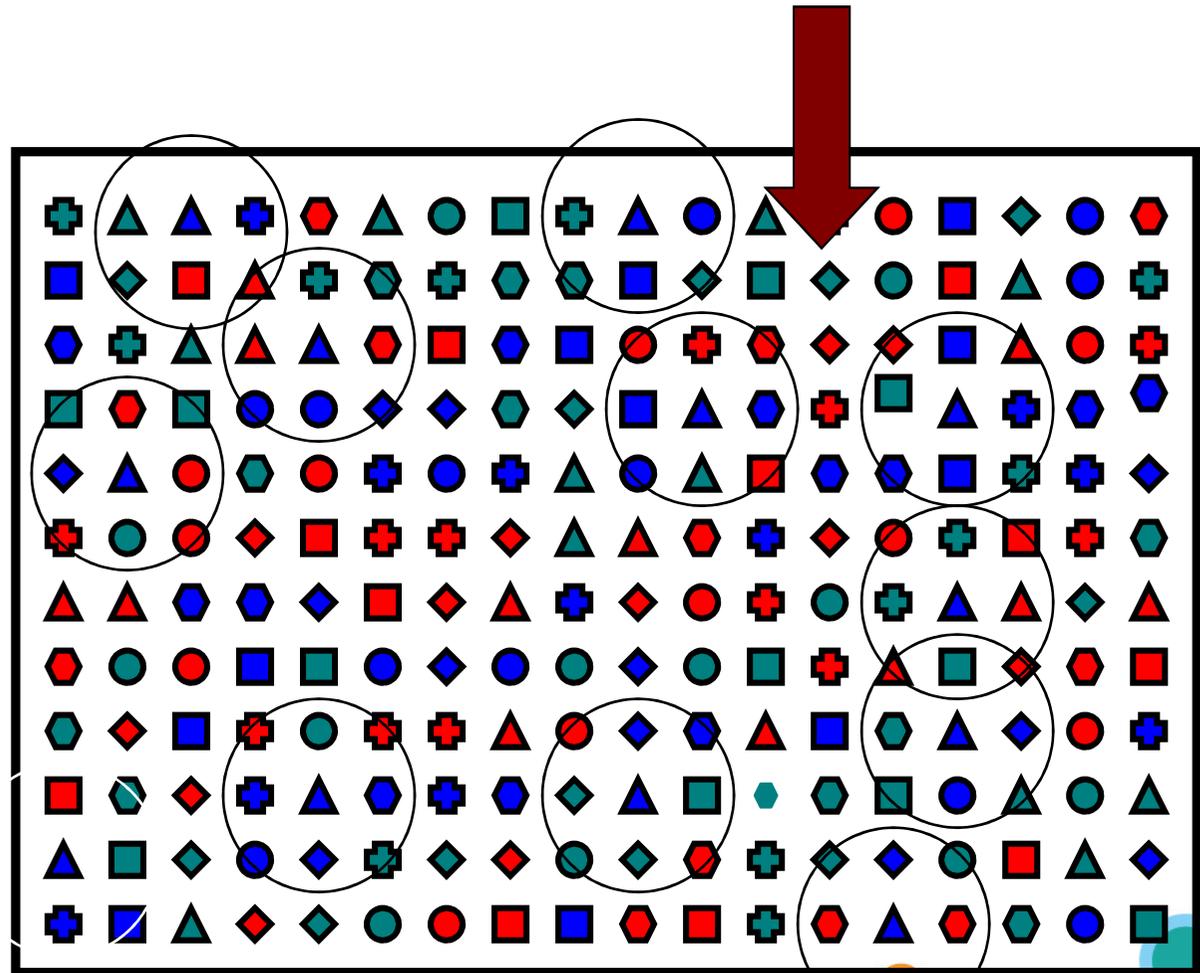


如果数据在单个数据库一张大表中

- 当数据存储在一个大表中, 许多的BI查询需要访问表中大部分数据

— SELECT ▲NAME,
TOTAL_SPEND,
LOYALTY_TIER
from CUSTOMERS
where REGION=

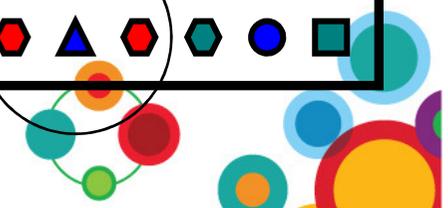
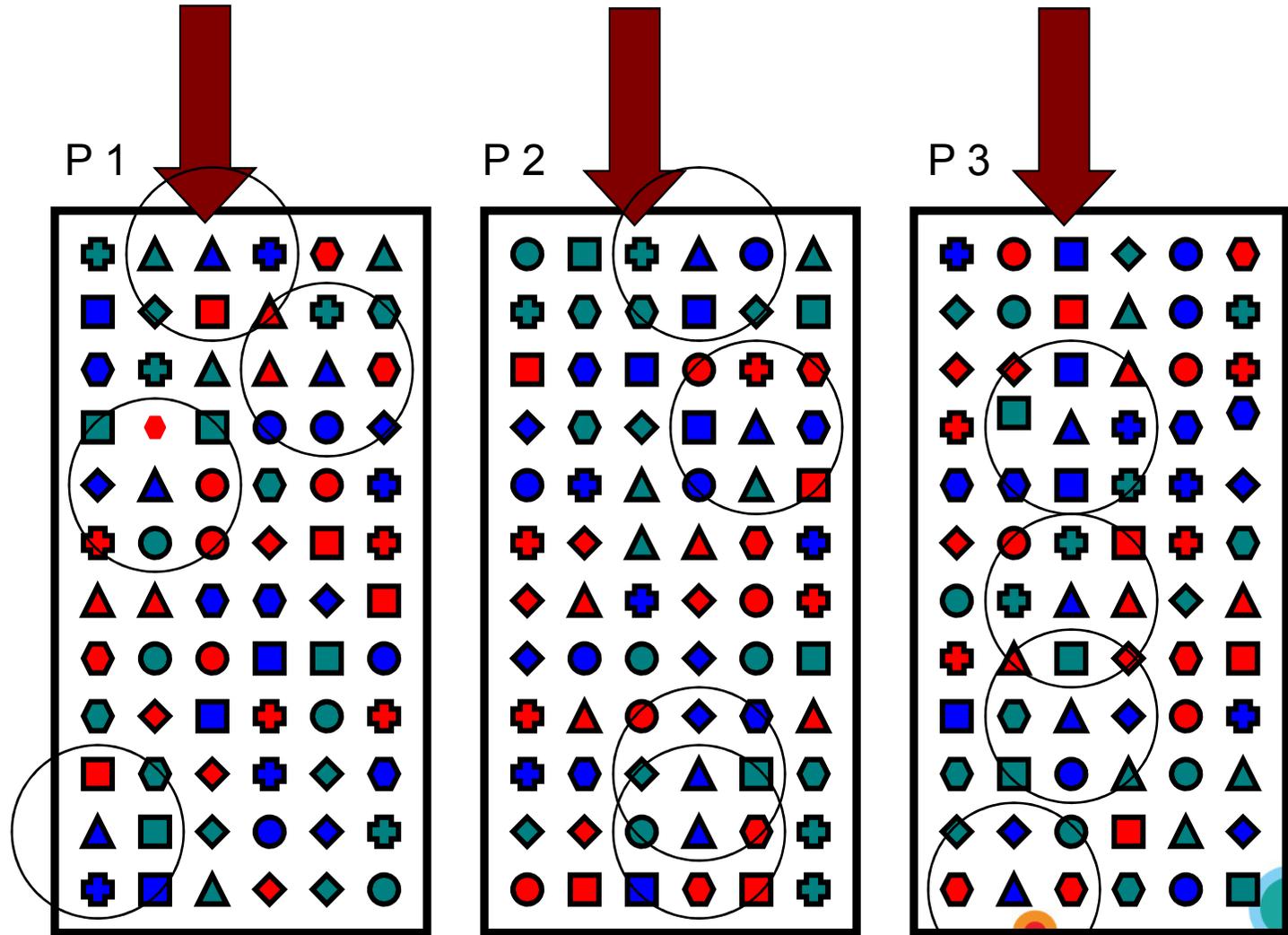
- 如果没有使用DPF, 查询时大部分的处理都只能利用1个CPU

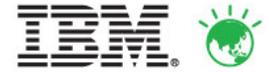




HASH数据库分区

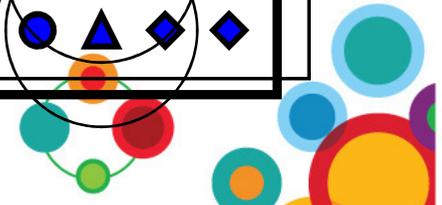
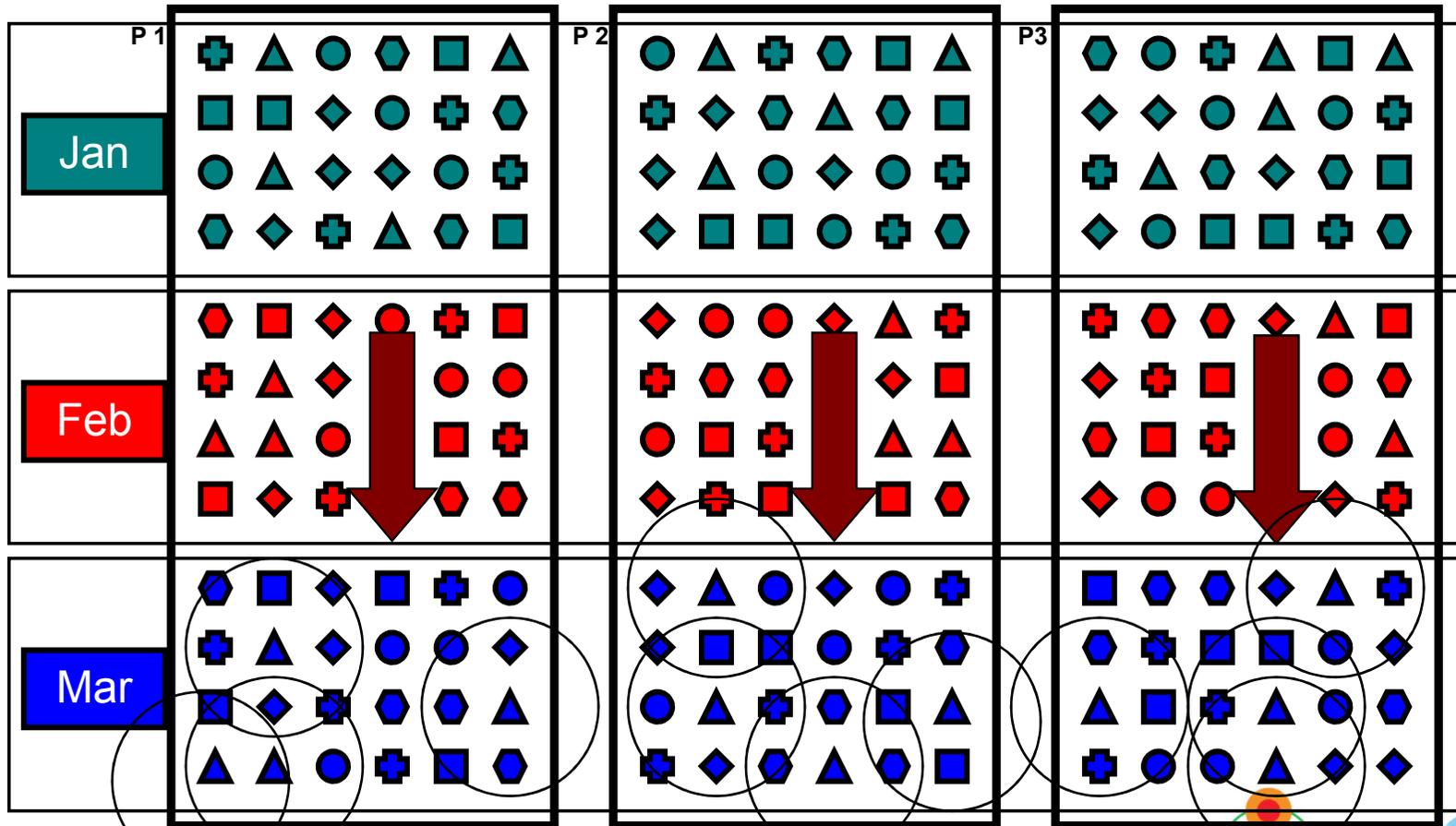
- 使用DPF, 查询也许仍然读取大部分的数据, 但查询却可以并行到所有的节点





HASH数据库分区 + Range 表分区

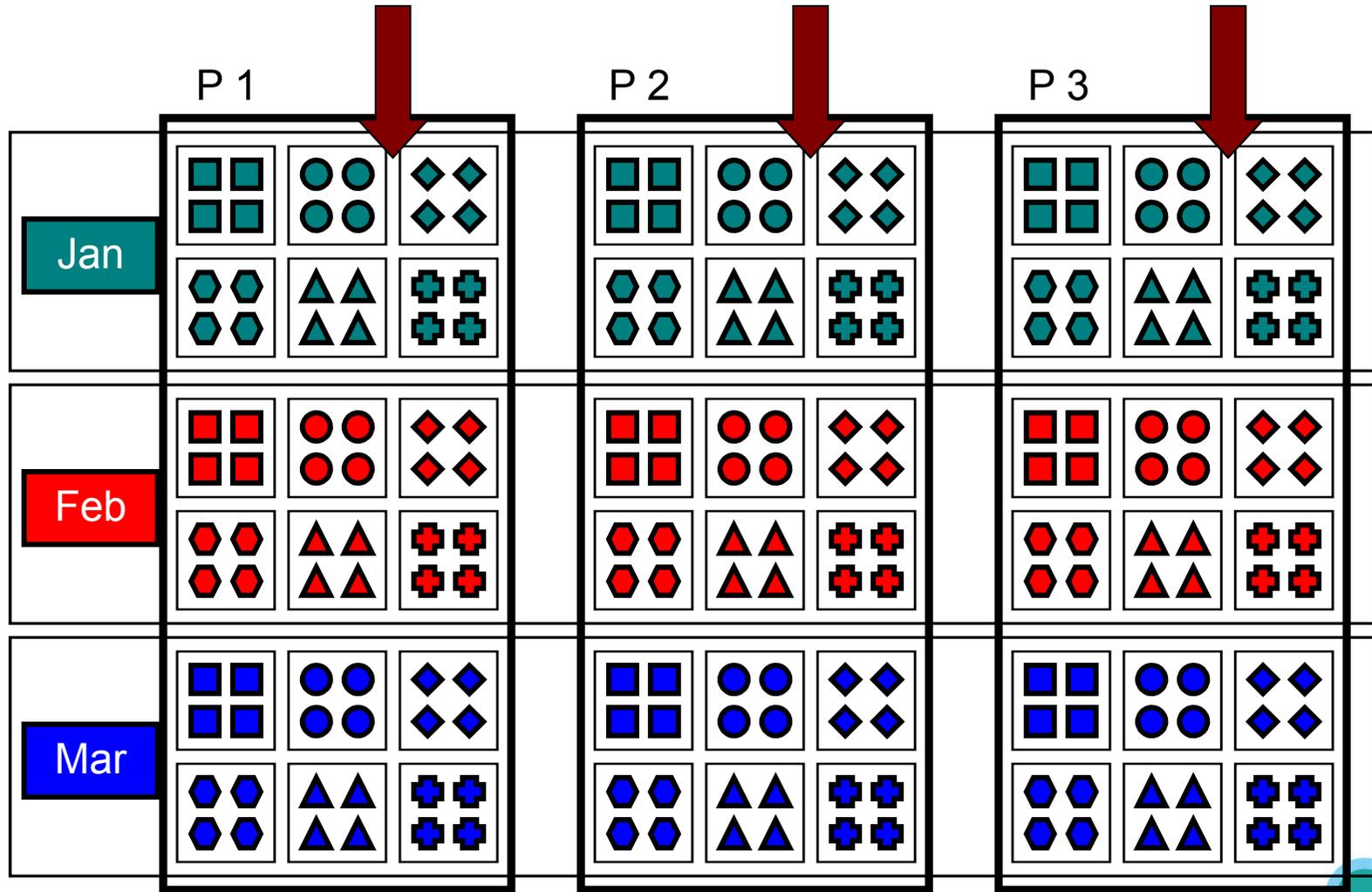
- 所有的相同范围的数据在一个分区中
- 数据库仅读取适合分区的数据
- 所以range分区节省I/O资源

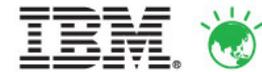




HASH数据库分区 + Range表分区 + MDC聚簇索引

- 有了MDC, 数据进一步根据属性簇组织
- 更少的I/O
- 而且, 表可以更容易的roll-in/roll-out of dat.





DB2 的数据压缩技术

- 基于字典的压缩——为压缩/解压数据建立符号表
- 使用基于静态字典表的Lempel-Ziv (LZ) 算法
- 数据在内存/磁盘中均以方式驻留
 - 节省大量I/O
 - 显著节省内存占用
 - 占用CPU负载: 行数据处理前, 需要进行解压缩操作
- 如果有些行数据的压缩并不能显著节省空间, 那么DB2将自动选择不进行压缩

Name	Dept	Salary	City	State	ZipCode
Fred	500	10000	Plano	TX	24355
John	500	20000	Plano	TX	24355

原始行 →

Fred	500	10000	Plano	TX	24355	John	500	20000	Plano	TX	24355	...
------	-----	-------	-------	----	-------	------	-----	-------	-------	----	-------	-----

压缩行 →

Fred	(01)	10000	(02)	John	(01)	20000	(02)	...
------	------	-------	------	------	------	-------	------	-----

Dictionary

01	Dept 500
02	Plano, TX, 24355
...	...





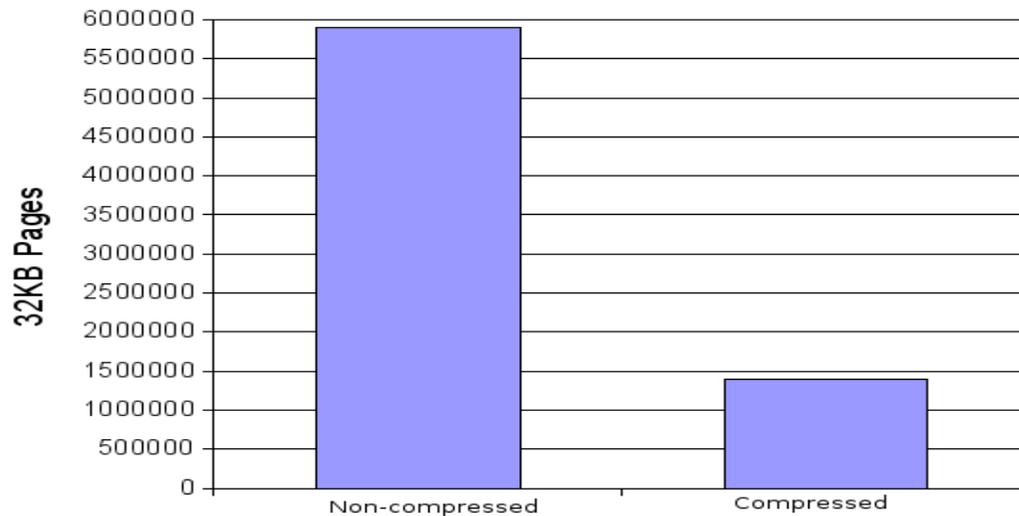
DB2 提供强大的数据压缩能力

大大减少存储成本，提高I/O效率

压缩类型	32KB 页数	磁盘空间需求
无压缩	5893888	179.9GB
记录压缩	1392446	42.5GB

% 节省存储: 76.4%

T1 Compression - 179.9GB Initial Size



国内测试结果:

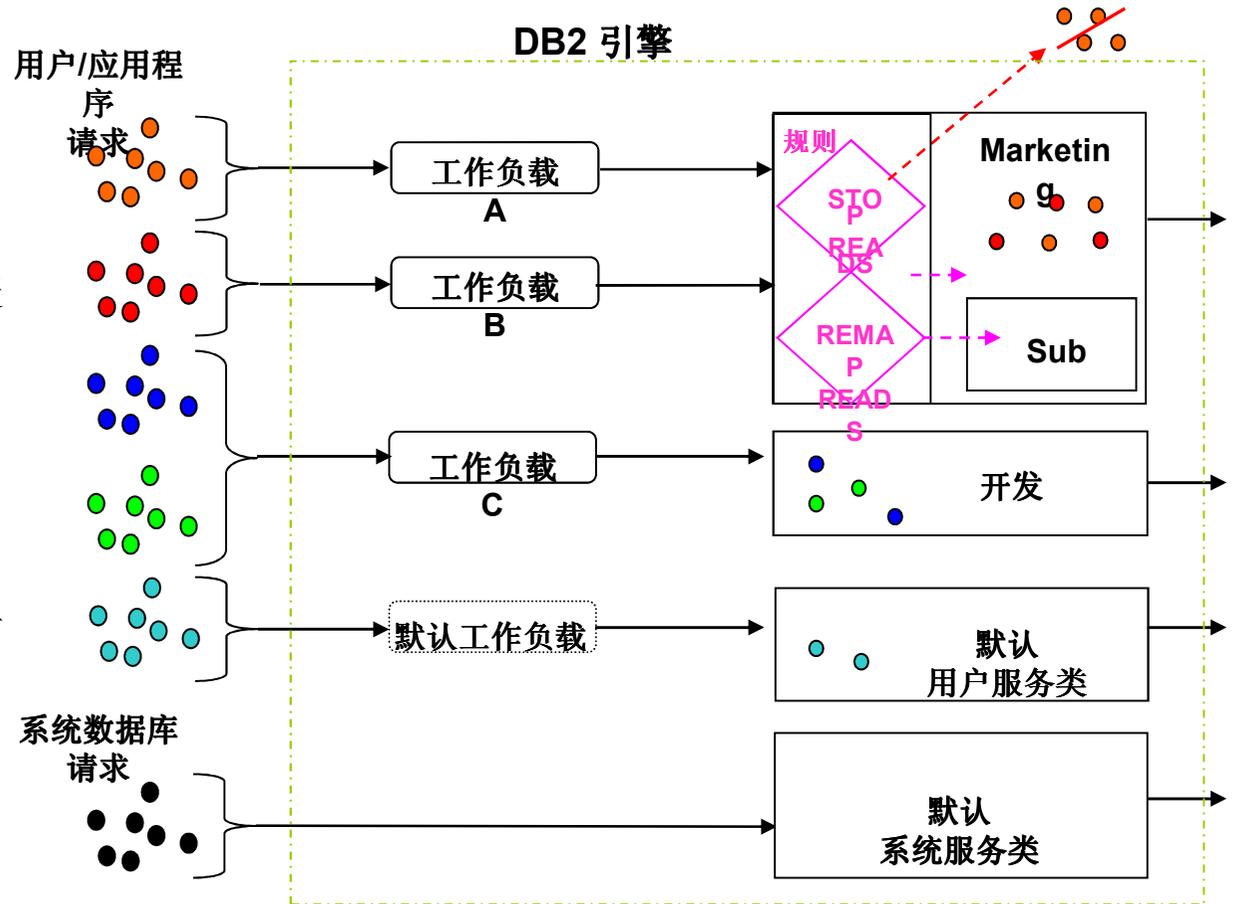
- XX软件: 数据存储减少60%以上, 同时交易性能提升10%
- XX移动: 数据存储减少50%
- XX银行: 数据存储减少50%





灵活的工作负载管理能力

- 防止系统资源过度消耗
 - 并发作业数等控制
 - 对作业运行的总时间进行限制
 - 阻止”垃圾“SQL、”流氓“查询
- 实现不同服务级别
 - 对作业进行优先级管理
- 资源公平使用
 - 分组资源总量控制
- ...

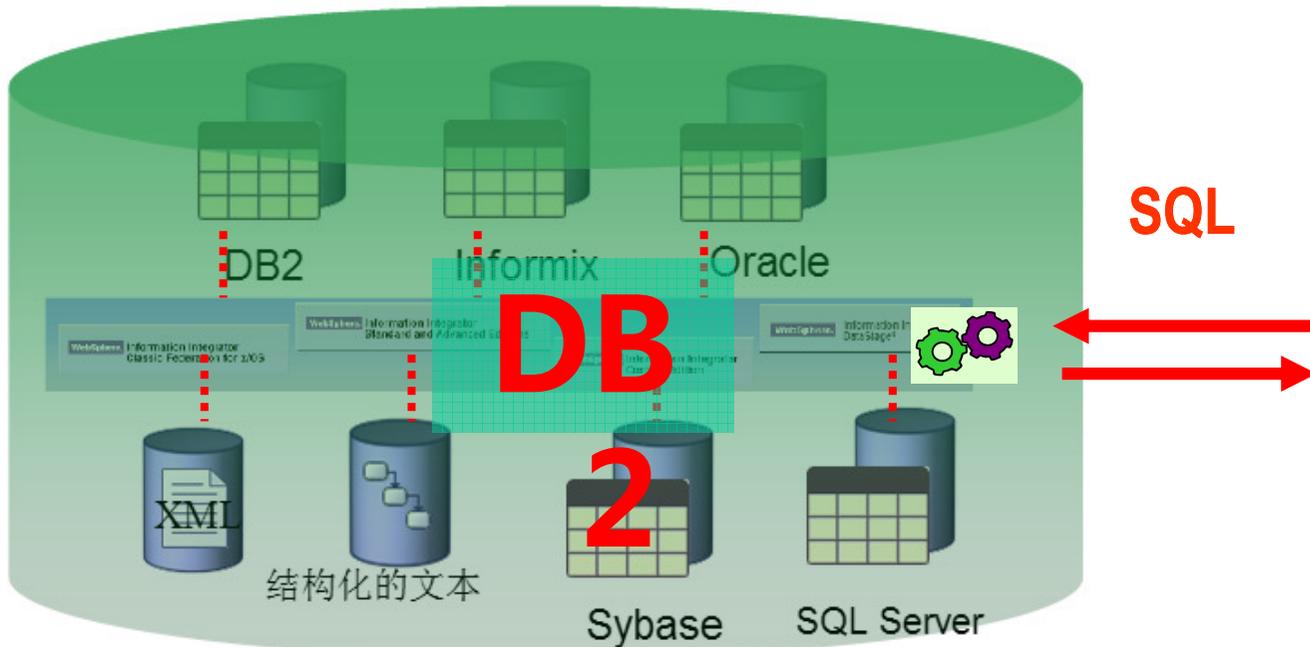




联邦访问

联邦技术可以组合多种不同的数据源，实现虚拟整合，对于最终用户、数据管理者、开发人员来看，所有数据如同在同一个DB2数据库中，可以方便的进行查询和处理。

- 独立性，几乎对原有系统应用没有影响，无需在原有系统上安装程序、软件, 无需特有的接口
- 高性能，优化分布式访问，SQL转换下推
- 安全性，与被访问数据源的用户建立映射，双层用户认证/授权机制

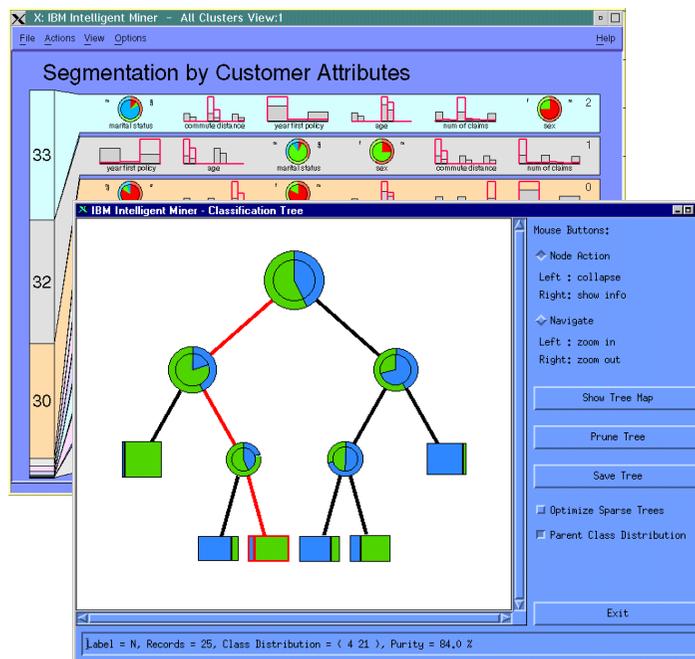




内嵌数据挖掘 - DB2 Intelligent Miner



- 关联挖掘
- 聚类分析
- 预测分析
- 决策树分析

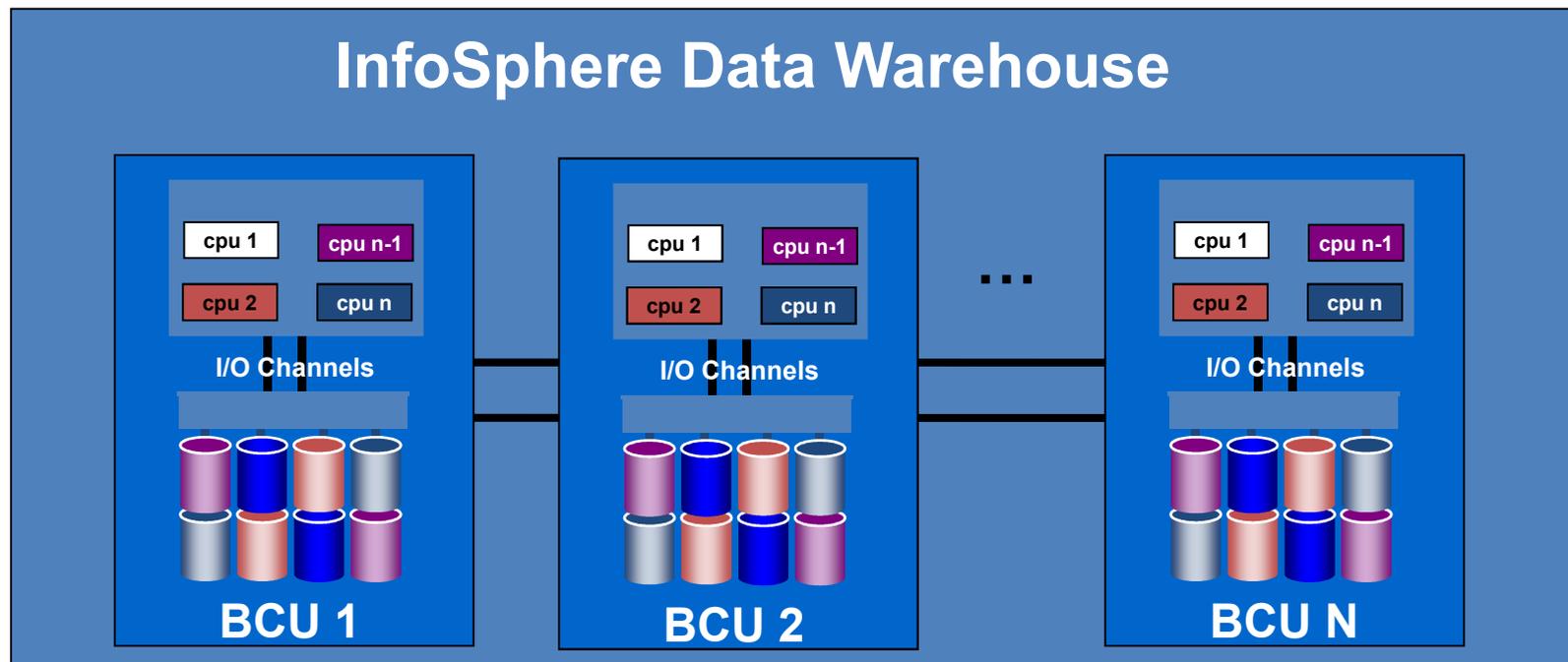




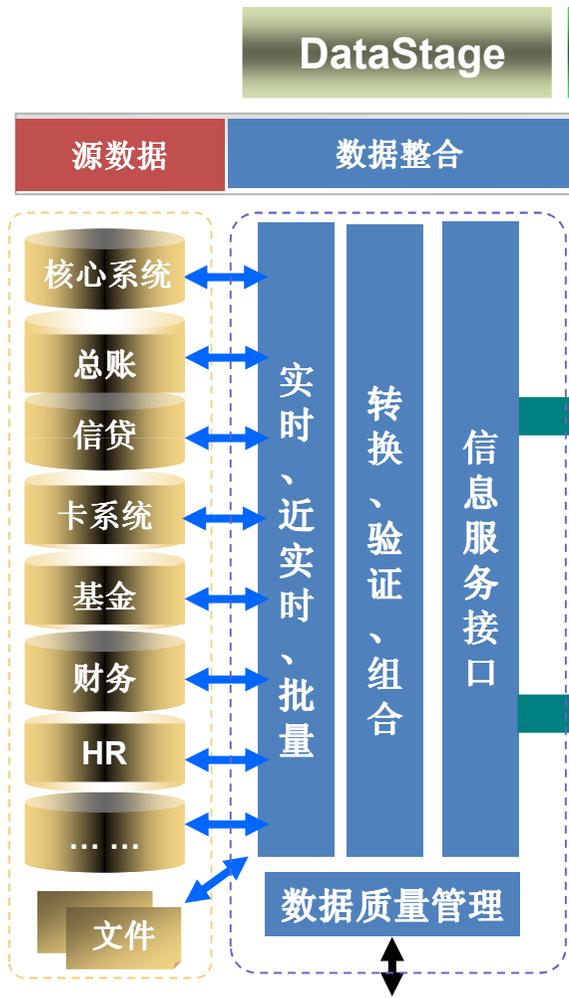
基于均衡配置方法论的标准单元模块

灵活增长

- 增加 BCUs 以满足需求增长
- 多种配置方案满足不同需求（Power6 570， Power7 740）
- 可靠，开放平台，可重用硬件



数据仓库系统软件架构





数据整合工具 - DataStage

- 通过数百种内置转换功能实现数据流的无代码可视化设计
- 优化数据整合对象的重用
- 利用并行处理，而不要求更改设计
- 能够支持批处理和实时操作
- 完备的ETL能力
- 支持团队协作开发
- 支持最广泛的数据源整合



开发人员



架构师

转换

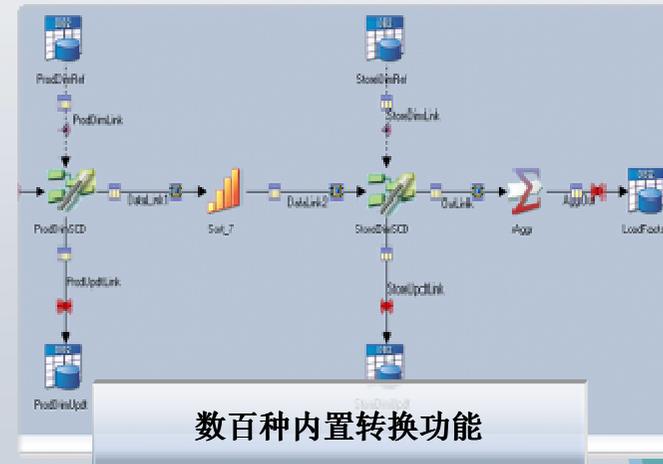


交付



WebSphere DataStage®

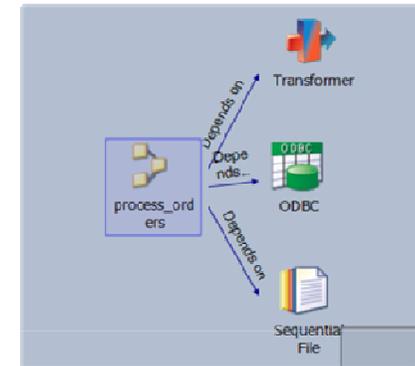
通过可视化设计的逻辑，以批处理或实时方式转换和集合任意数量的信息



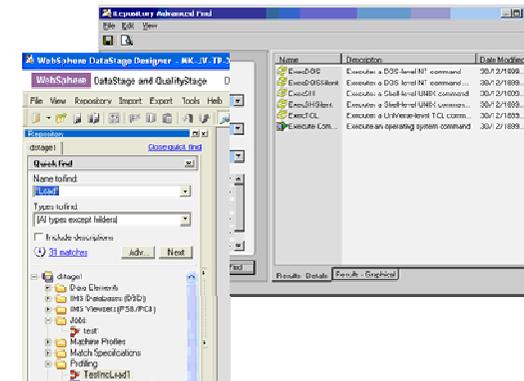


元数据驱动的综合

- 在不同的产品模块中共享元数据
 - 团队成员之间更好更快的沟通
 - 直接访问所有对象的定义和注解
 - 更好的理解，更佳的数据质量
- 强大的元数据设计工具
 - 快速查找和高级查找数据对象
 - 数据对象关联影响分析
 - 数据血缘报告
 - 更好的工作效率，更简单的维护和重用



Impact Analysis



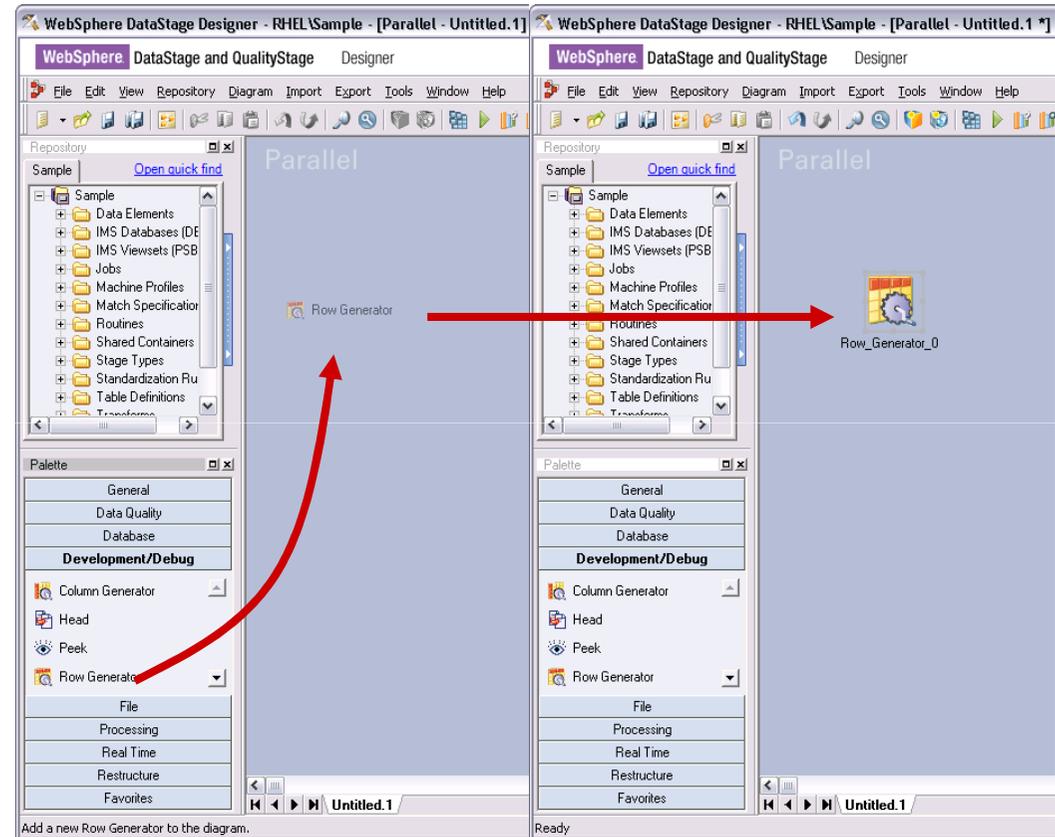
Find Capability





图形化的设计开发面板

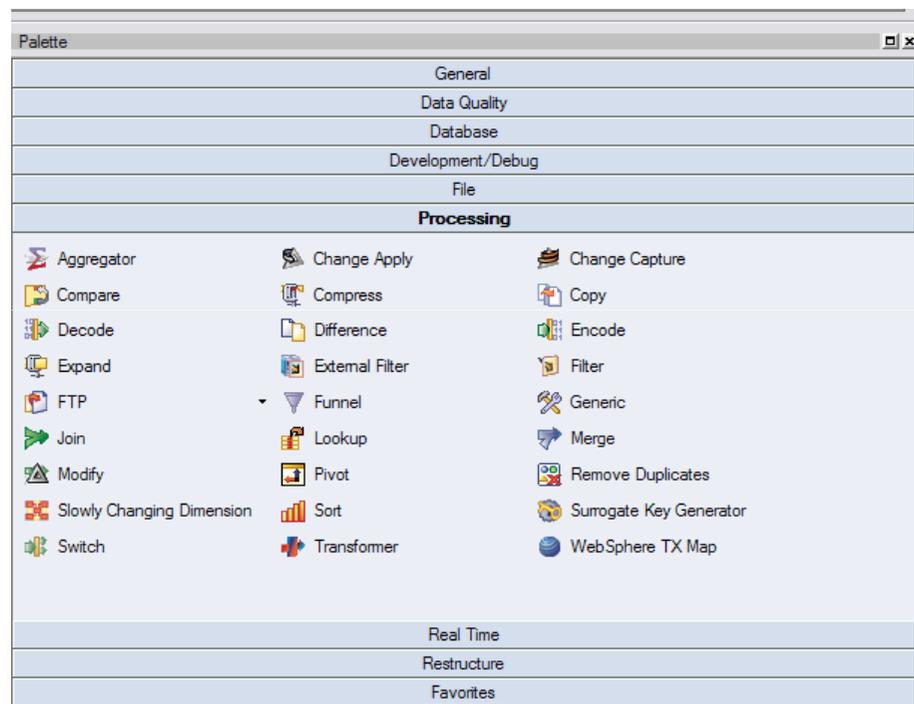
- 拖拖拽拽的工作方式进行ETL逻辑设计
 - 鼠标左键选择预定义的计算组件，拖拽到设计面板中释放
 - 或者鼠标左键点击计算组件，然后在设计面板中再次点击鼠标左键放置计算组件。





内置大量成熟的计算组件

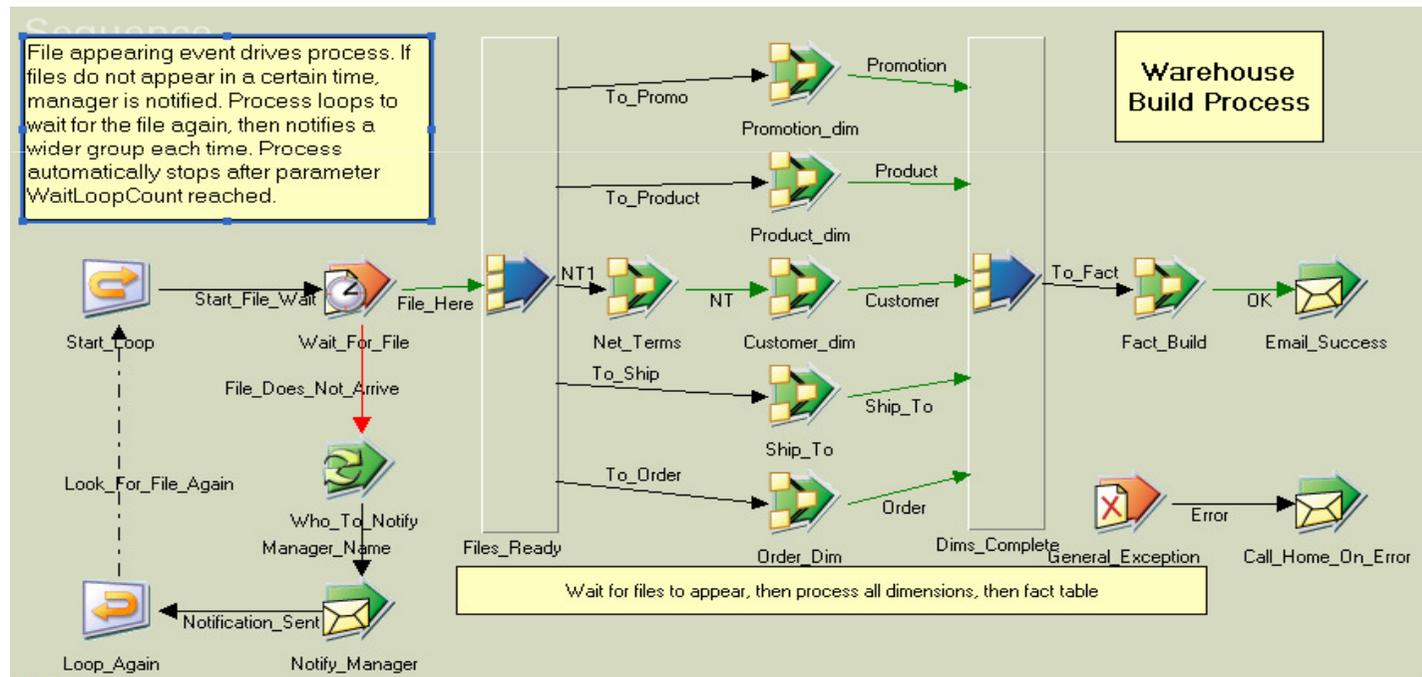
- 超过50种预定义的转换计算组件，快速实现ETL任务开发，大大提高开发效率
- 大量的内建函数使得复杂的转换逻辑开发变得简单
- 简单易用的计算组件编辑器轻松进行参数配置





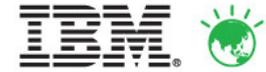
灵活 workflow 管理

- 在 workflow 内部嵌套定义动态子 workflow
- 可以调用外部流程支持复杂计算
- 支持高级 workflow 概念例如分支、循环

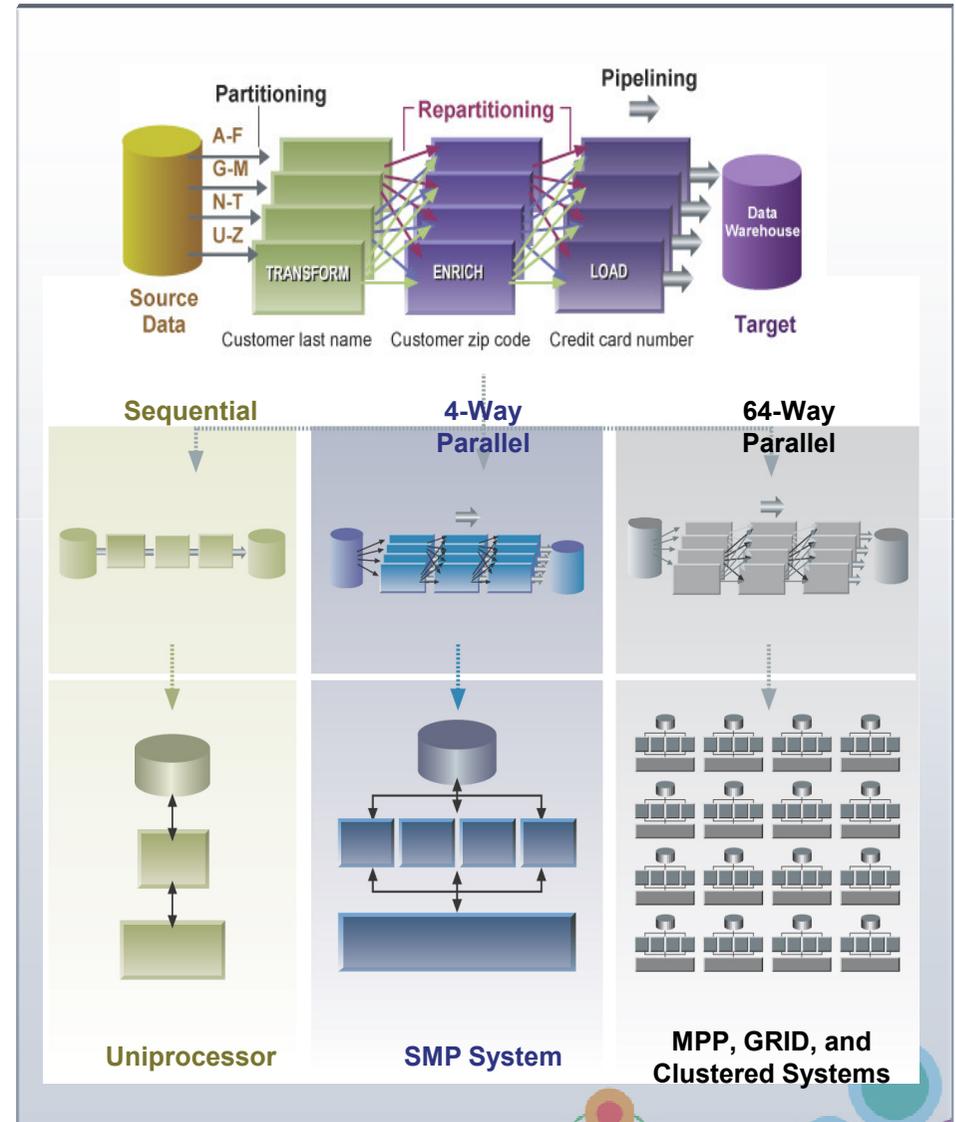




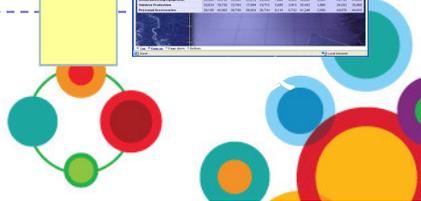
并行处理的数据整合平台



- 无需考虑数据规模和时间约束
- 利用数据库的分区能力加速数据装载
- 简化必要时的并行处理定义流程
- 简单方便的添加CPU等资源提高处理能力
- 无需编程就可以启用多路CPU的处理能力
- 支持多路CPU、多机集群等大规模计算平台

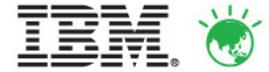


数据仓库系统软件架构



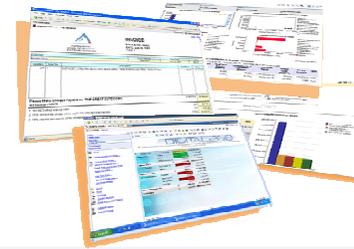


IBM Cognos提供的广泛能力



报表

- 支持各种报表格式
- 通过各种不同的格式报表显示一致的信息
- 支持用户与用户之间以及用户与IT之间的协作
- 提供各种展现途径：email, portal, MS-Office, Mobile, 搜索引擎



分析

- 为多维分析提供维度信息向导展示
- 可以简单快捷的进行复杂的分析和场景建模
- 揭示趋势后面的征兆和原因
- 轻松实现在不同层次粒度之间钻取



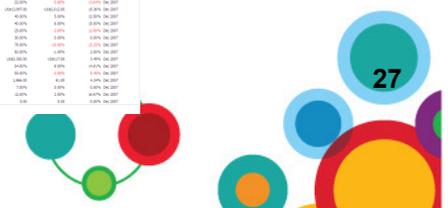
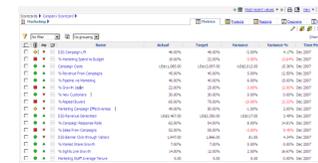
仪表盘

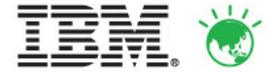
- 一目了然地纵览全局
- 有助于迅速定位有问题需要采取行动的地方
- 形成直观的、富有冲击的视觉效果
- 组合显示不同来源的信息



记分卡

- 提供每个目标相应的度量指标
- 将战略与具体战术相结合
- 支持平衡计分卡方法论
- 确保责任明晰





IBM Cognos 8

完备

全面的BI能力

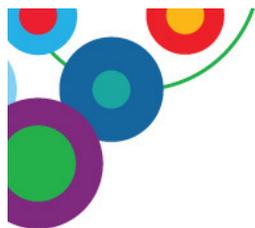
简便

统一架构

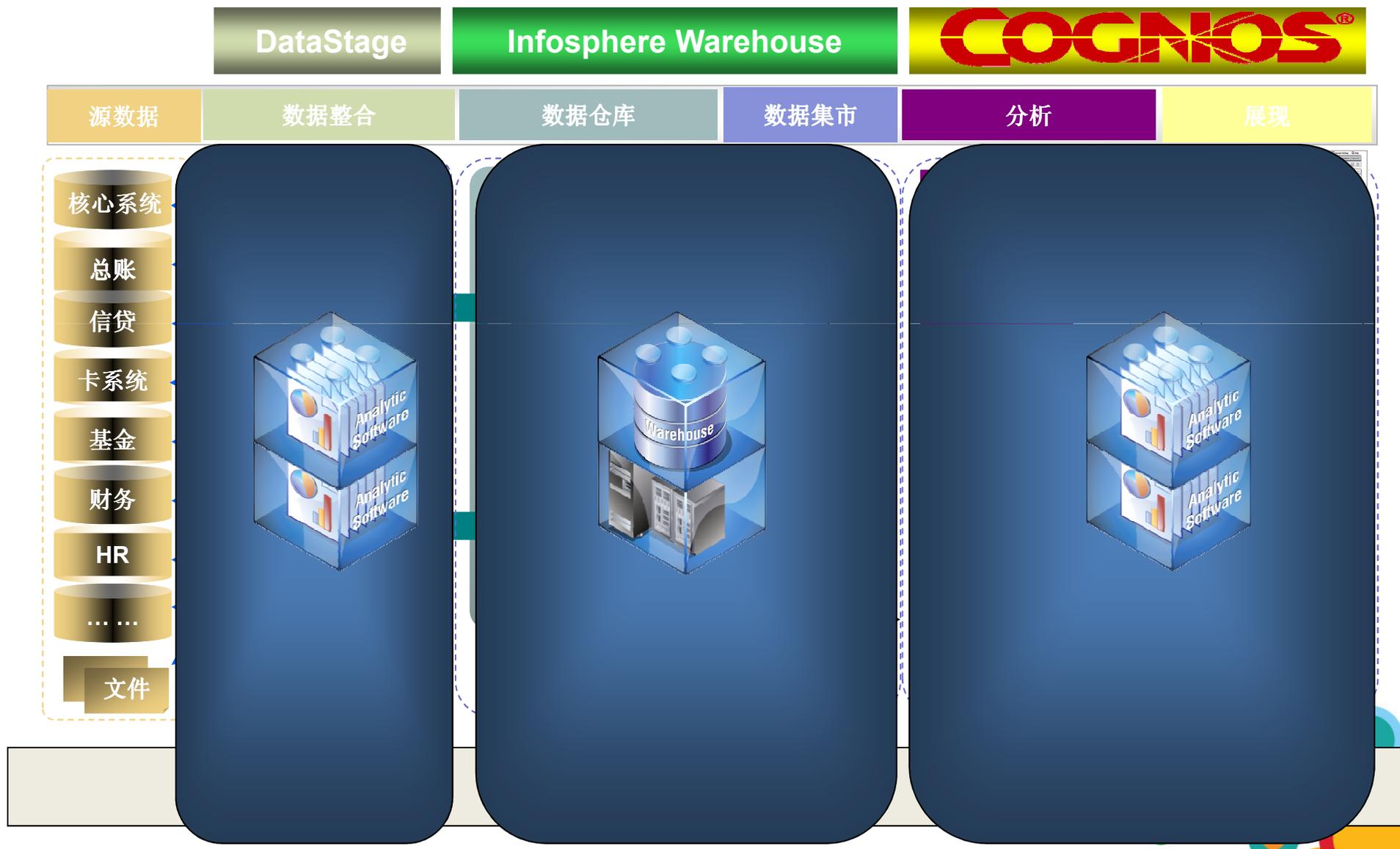
可靠

可靠的领先技术体系



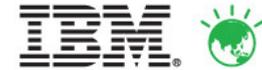


IBM 智能分析系统(ISAS) — 软硬件一体的解决方案

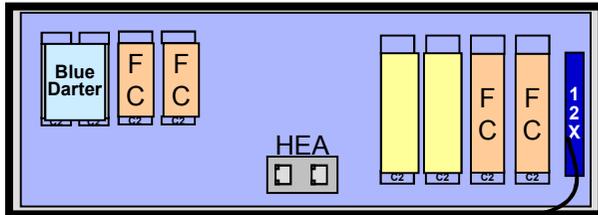




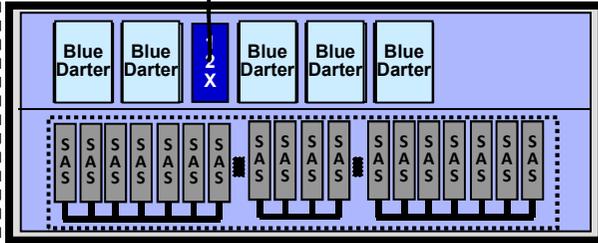
ISAS 7700 硬件方案



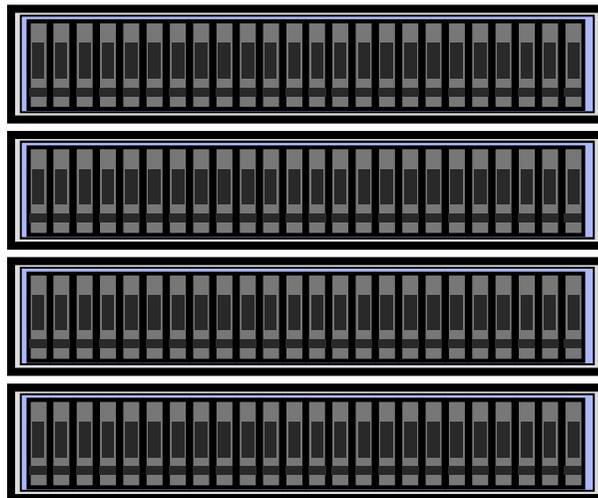
740



Exp12X (Optional)

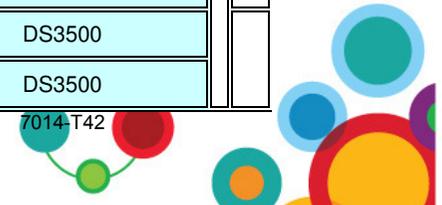
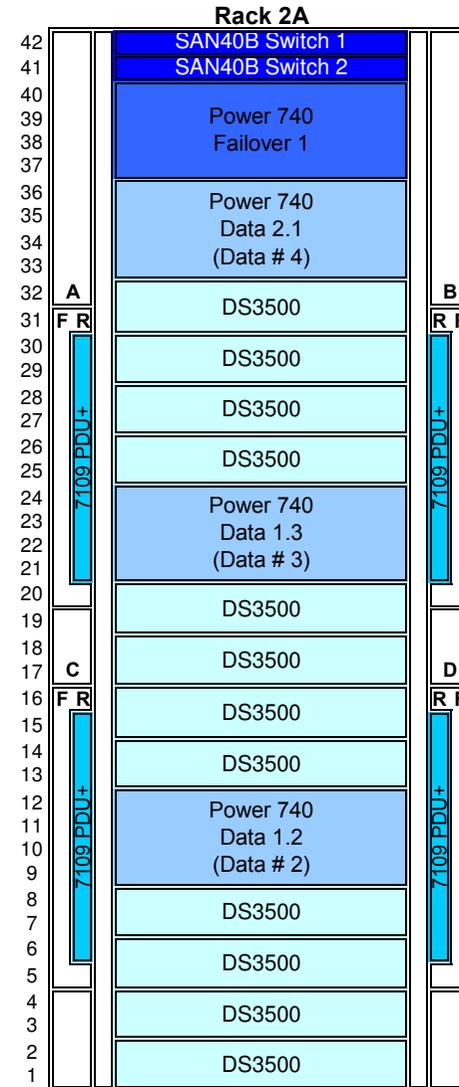


4 DS3500



- Power 740 服务器
 - 16 CPU Cores
 - 128 GB 内存
 - 8 逻辑分区
 - 7.2 GB/s 带宽
 - 万兆以太网
- 4 DS3500 存储
 - 双控制器
 - 8GB 镜像Cache
 - 96 块SAS高性能磁盘
 - 28.8TB 磁盘空间

机柜配置





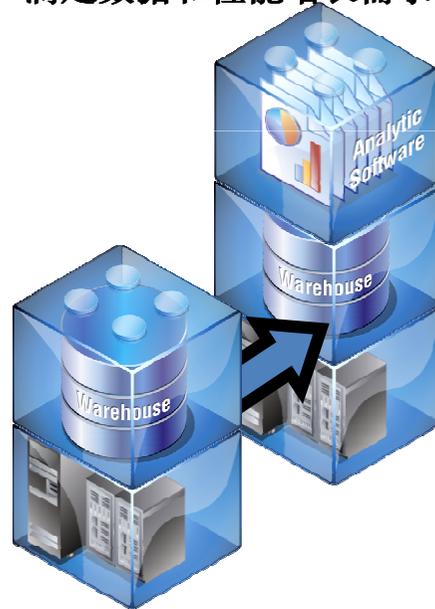
IBM Smart Analytics System

高度灵活性和扩展性

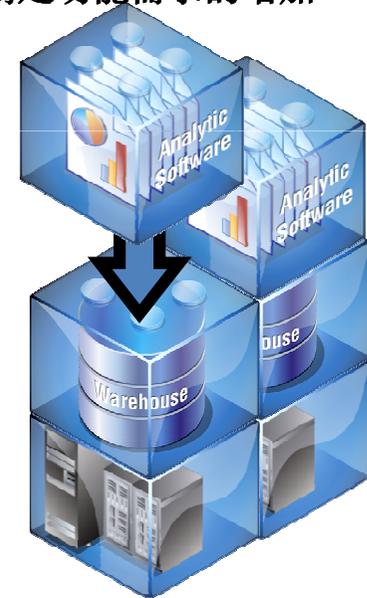
从最小的需求出发，
选择最小模块



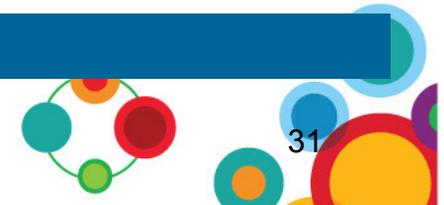
模块的横向扩展
满足数据和性能增长需求



模块的纵向扩展
满足功能需求的增加



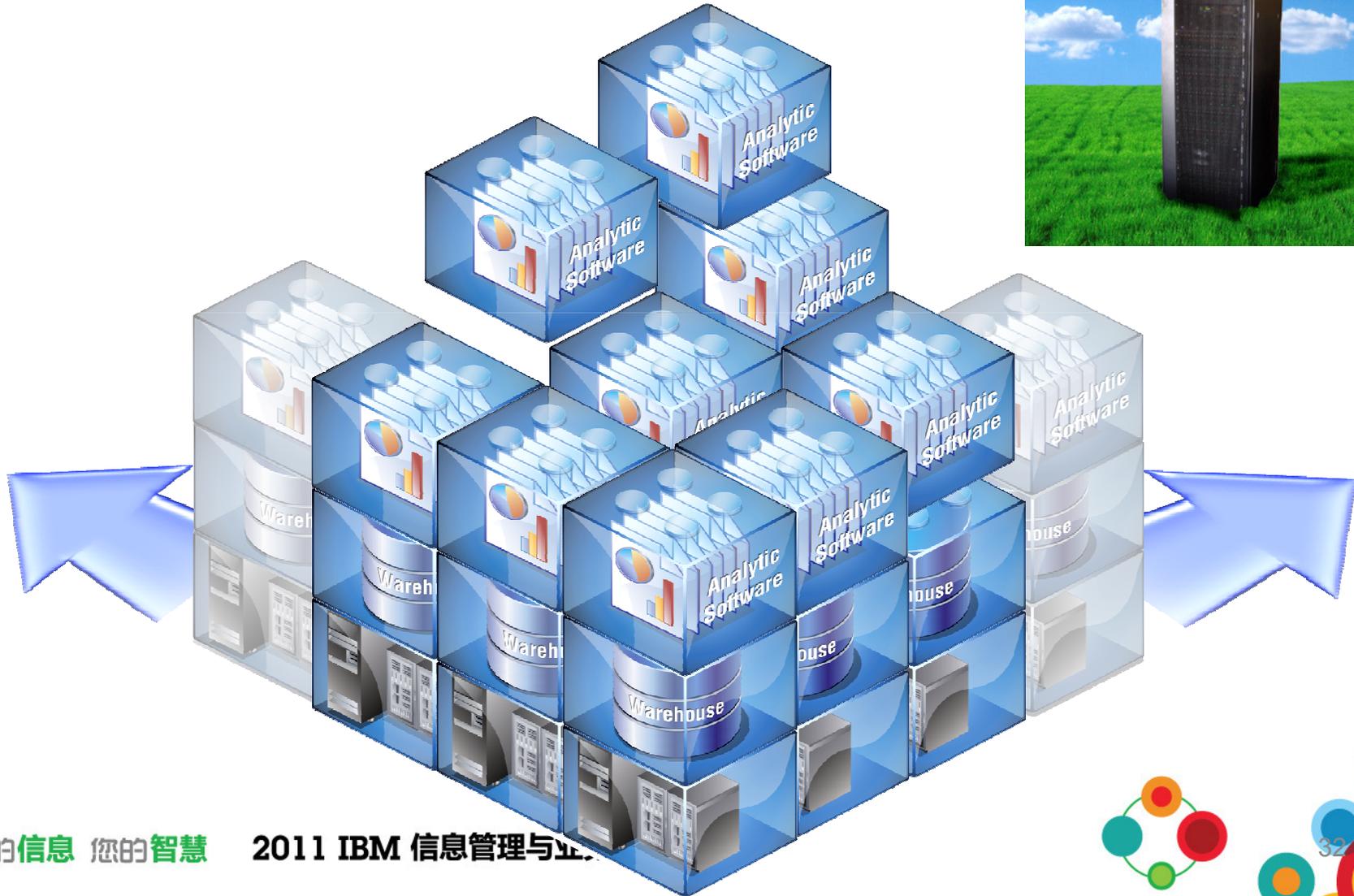
IBM SMART ANALYTICS SYSTEM





IBM 智能分析系统

—— 根据客户的需要灵活增长



您的信息 您的智慧 2011 IBM 信息管理与企业



I1

are the warehouses federated?
IBM_USER, 2009-7-16



THANK
YOU

Good enough never is.

© Copyright IBM Corporation 2008 All rights reserved. The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, these materials. Nothing contained in these materials is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software. References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. IBM, the IBM logo, Cognos, the Cognos logo, and other IBM products and services are trademarks of the International Business Machines Corporation, in the United States, other countries or both. Other company, product, or service names may be trademarks or service marks of others.

