



# Oracle 数据仓库参考架构与案例分享

Jason Guo  
Principle Solution Specialist  
Technology Strategy ,GC  
min.x.guo@oracle.com

*Copyright 2011 Oracle Corporation*

**姓名: 郭敏(Jason Guo)**

**职位: Oracle 技术战略部资深方案架构师**

• **简介:**

郭敏具有超过15年的IT专业背景和咨询服务经验，为国内主要的电信与金融行业客户提供过IT系统的咨询服务和项目实施，专长于数据仓库(DW)及商务智能(BI)解决方案架构设计、CRM应用系统规划设计、实施和项目管理，数据管控与数据质量管理。在加入甲骨文公司之前，郭敏分别供职于Teradata专业服务团队和亚信-联创的开发、实施、咨询团队。

• **主要项目经验:**

- 作为资深业务专家，为中国移动的不同省份的经营分析系统进行应用咨询服务，包括客户细分、客户价值、中高档客户保有、增值业务营销、反欺诈、客户体验改进等各类型专题
- 作为方案架构师，为中国移动某省设计实施营销管理平台与增值业务管理平台
- 作为数据架构方案专家，为中国移动某省进行数据管控咨询，提升数据质量，改进数据模型，优化数据处理流程
- 作为数据仓库方案专家，为中国移动、中国联通、中国电信各总部设计数据仓库架构与应用架构
- 作为方案架构师，为某大型商业银行提供ACRM(分析型CRM)提供整体架构的咨询与设计。
- 电信业核心业务系统(BOSS/BSS)的开发与实施
- 作为软件架构师，开发针对各类型IP业务的计费支撑软件的开发与实施

**服务的客户**

中国移动总部  
中国电信总部  
北京移动  
内蒙移动  
山西移动  
辽宁移动  
江西移动  
贵州移动  
吉林移动  
福建联通  
江苏联通  
山东网通  
北京网通  
香港HKnet  
中国银行

...

**工作经验**

- Teradata 中国有限公司
- 亚信-联创

**学历**

清华大学，计算机科学技术，工学学士  
信产部15所，计算机应用，工学硕士

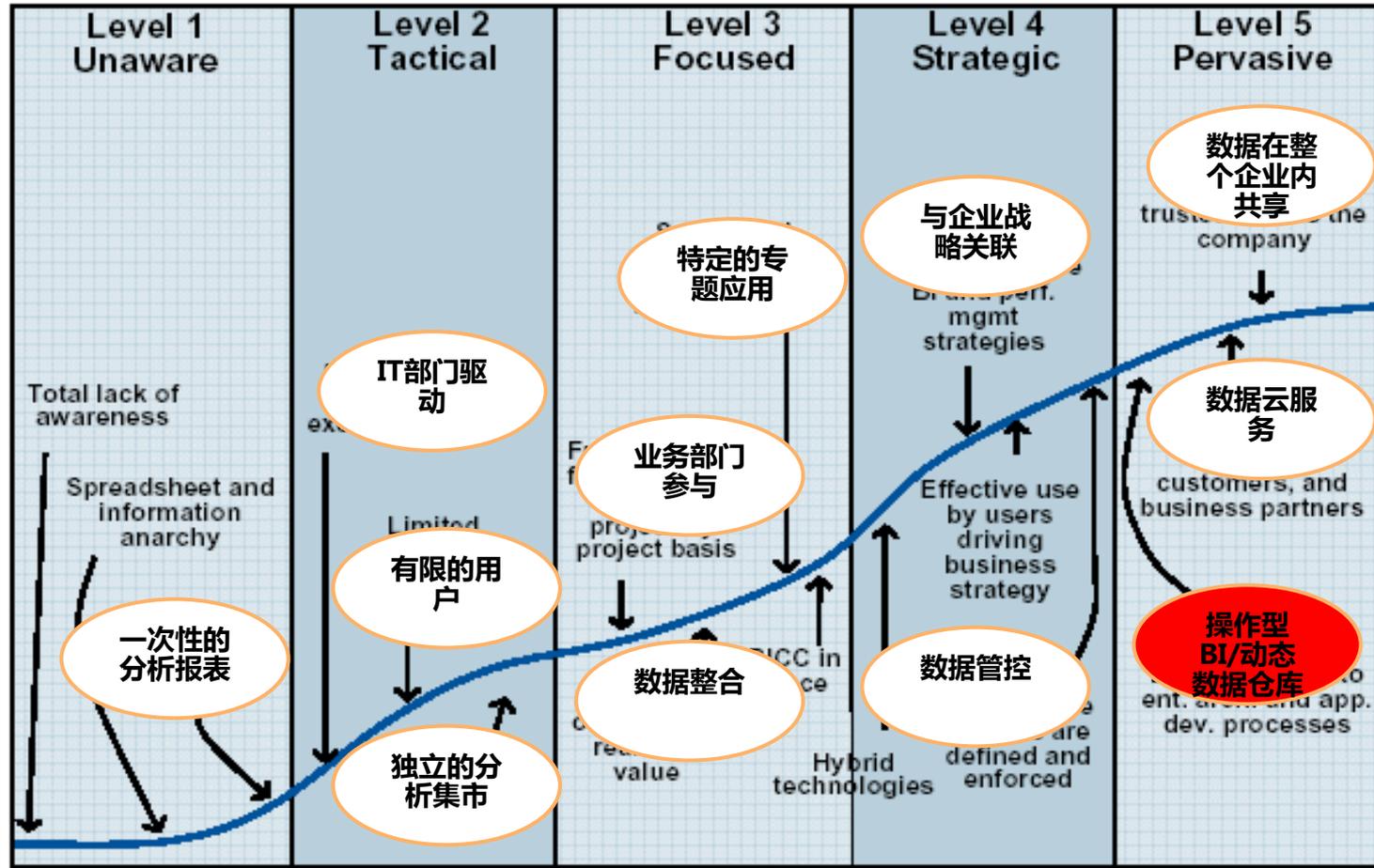
# 目录

**1. 数据仓库参考架构**

**2. 案例分享**

**3. 数据集成产品ODI介绍**

# 发展趋势：动态数据仓库 (Active DW)



\*资料来源：Gartner Business Intelligence Summit

Gartner

ORACLE

# 后视镜和仪表盘

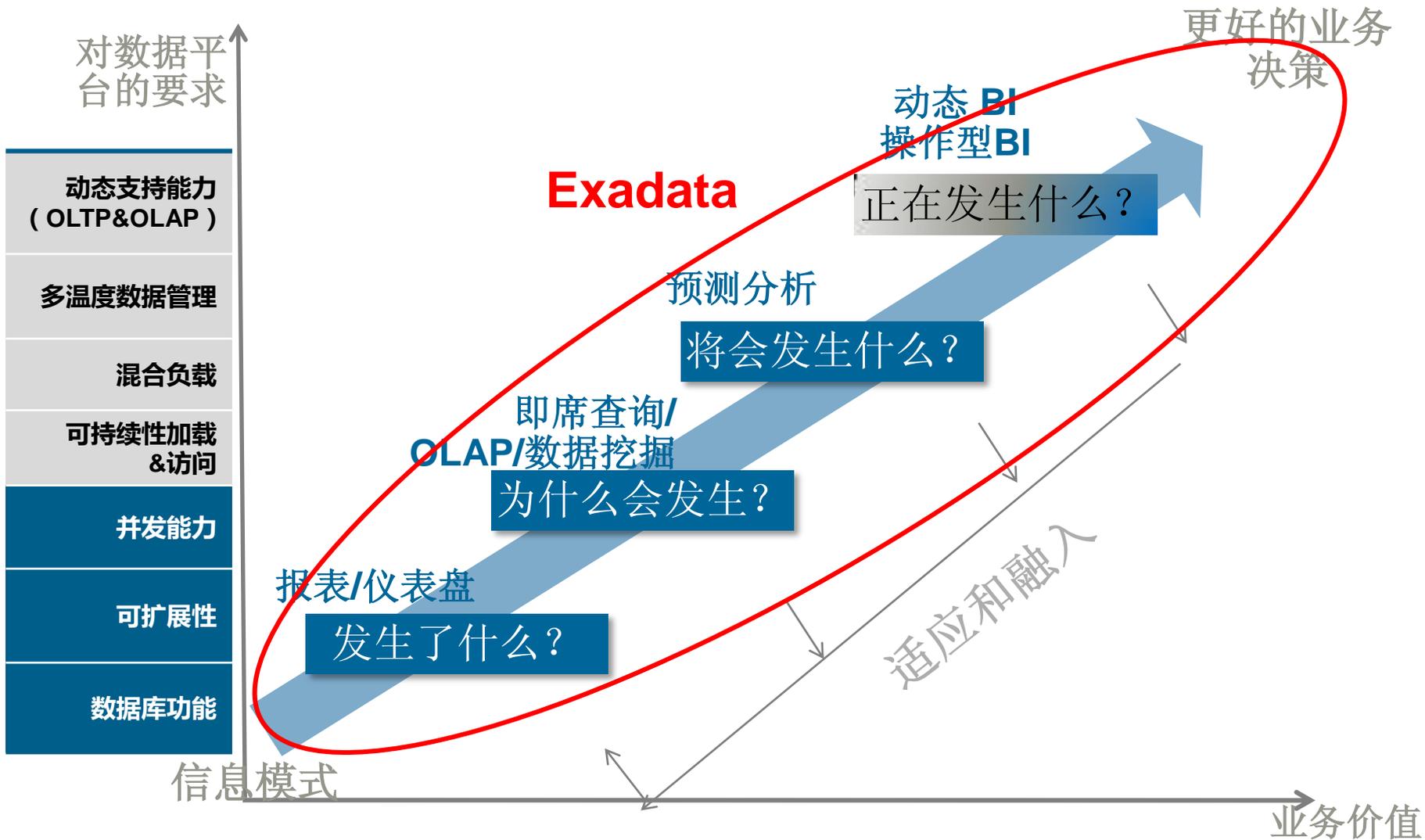
——从传统商务智能（BI）到动态商务智能（Active BI）



传统的商务智能更侧重历史分析，  
实现**战略决策**

今天的BI更侧重与业务流程实时整合，  
实现**运营决策**

# 商业智能发展趋势



# 构建Active BI系统



基于‘昨天’的数据  
做‘今天’的决策

- ✓ 主要面向分析员、决策者
- ✓ 数据的周期为日、周、月
- ✓ 定时的批量导出导入
- ✓ 历史数据查询为主
- ✓ 面向功能为主
- ✓ 被动服务

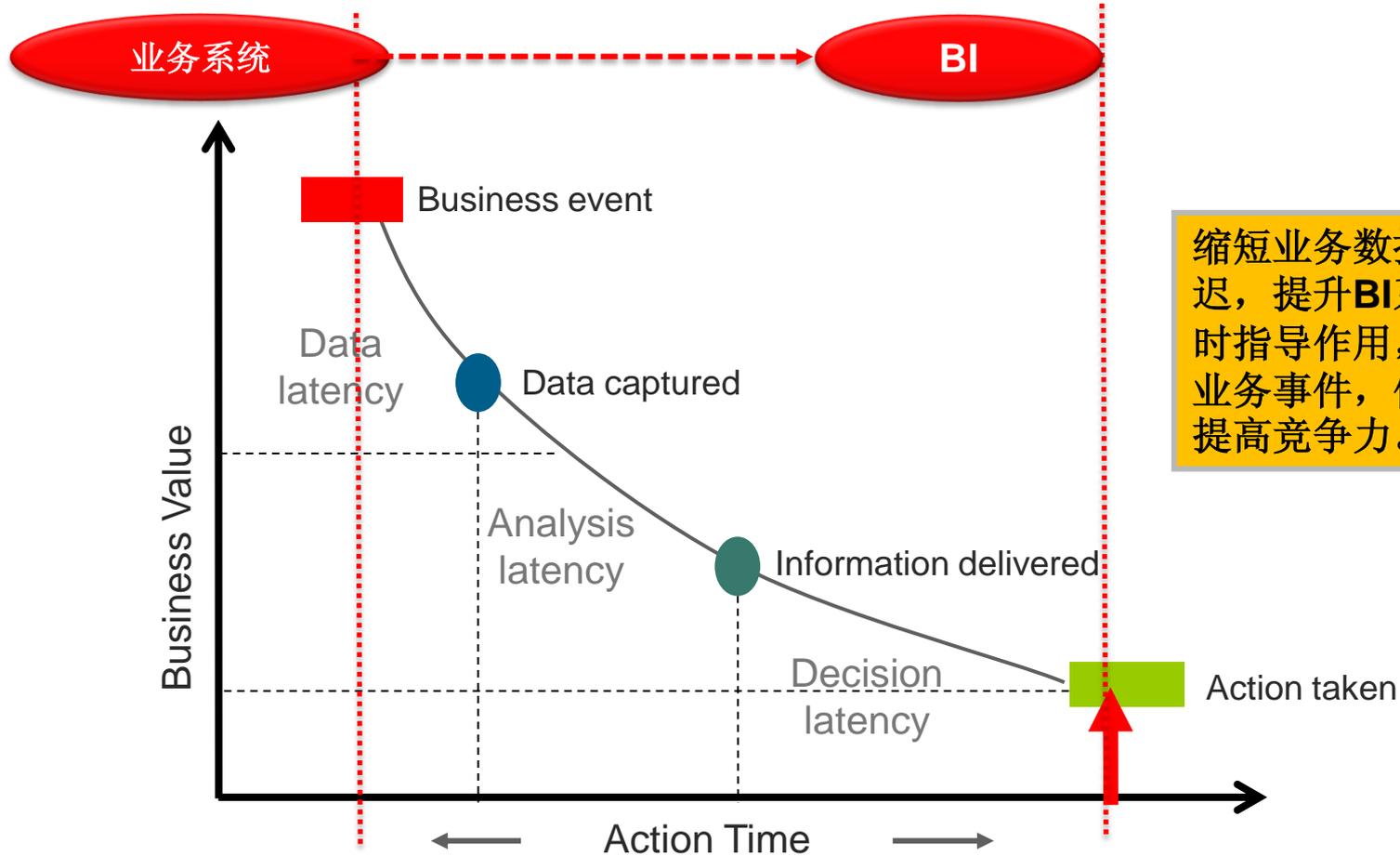


基于实时的数据  
做实时的运营优化

- ✓ 可以面向一线的客户经理、销售代表
- ✓ 数据的周期可以达到分钟
- ✓ 批量处理结合变化数据捕获
- ✓ 支持实时&历史查询
- ✓ 面向流程为主
- ✓ 支持主动的事件驱动的服务

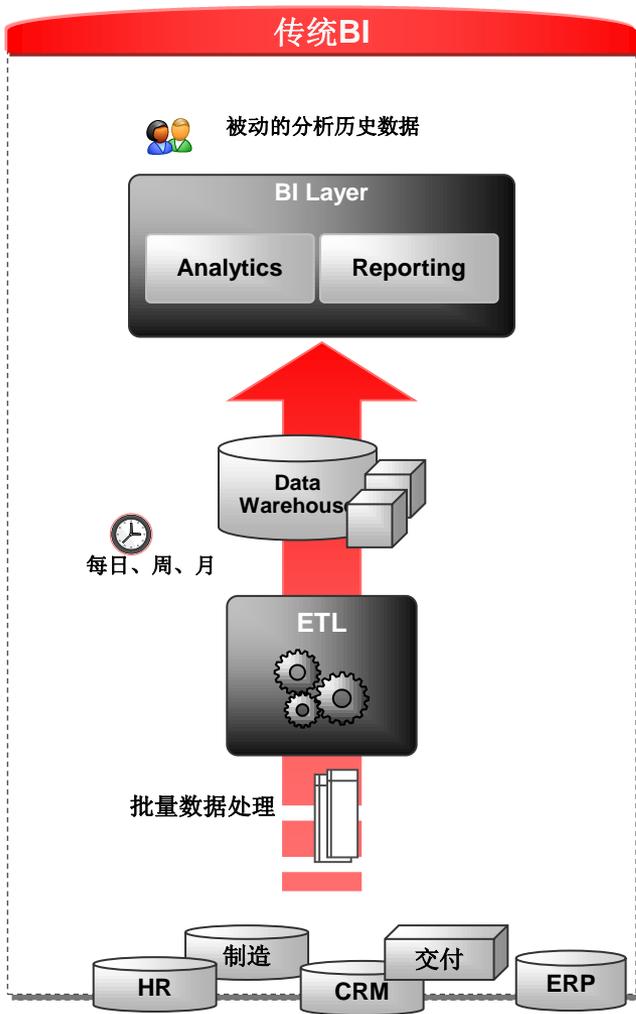
VS

# Active DW的意义 *提高企业信息的价值*

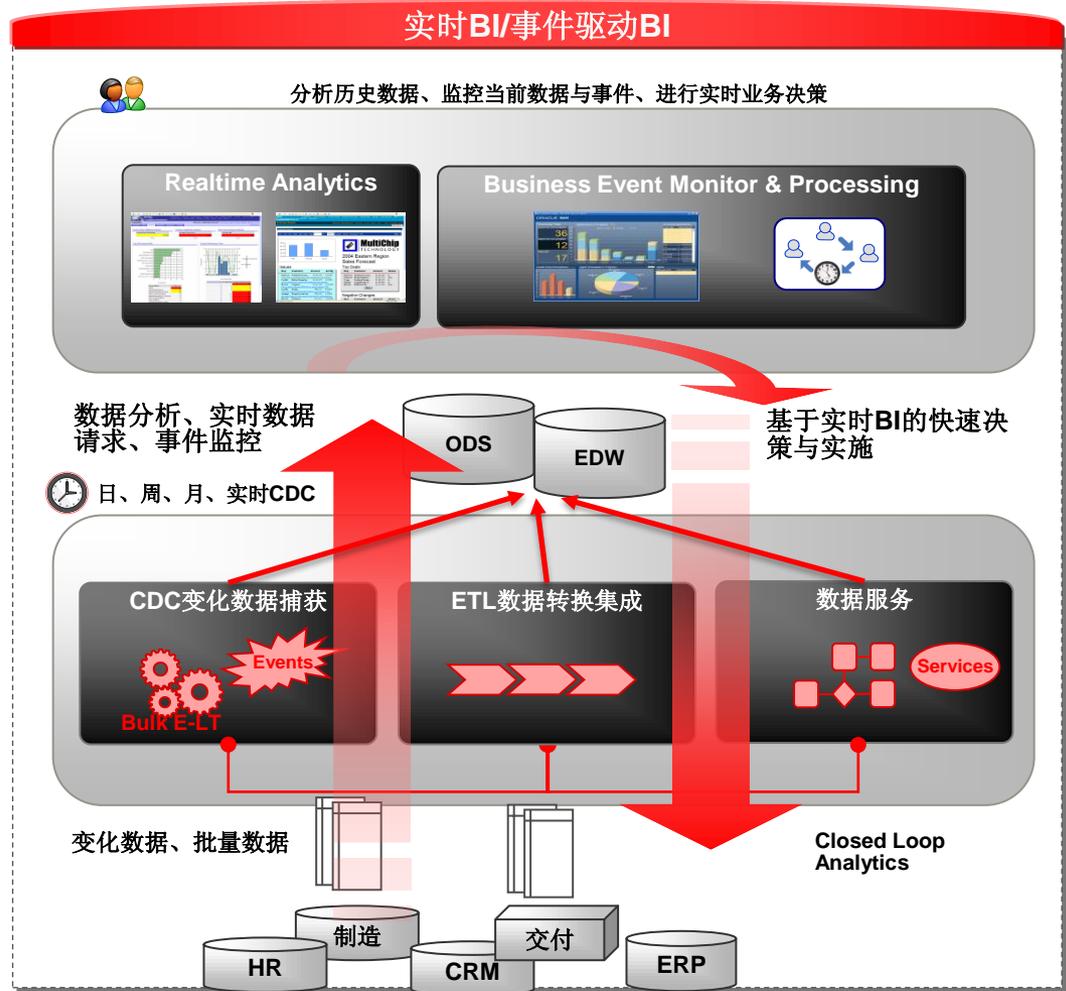


信息的价值取决于时效性与质量，而非数量！

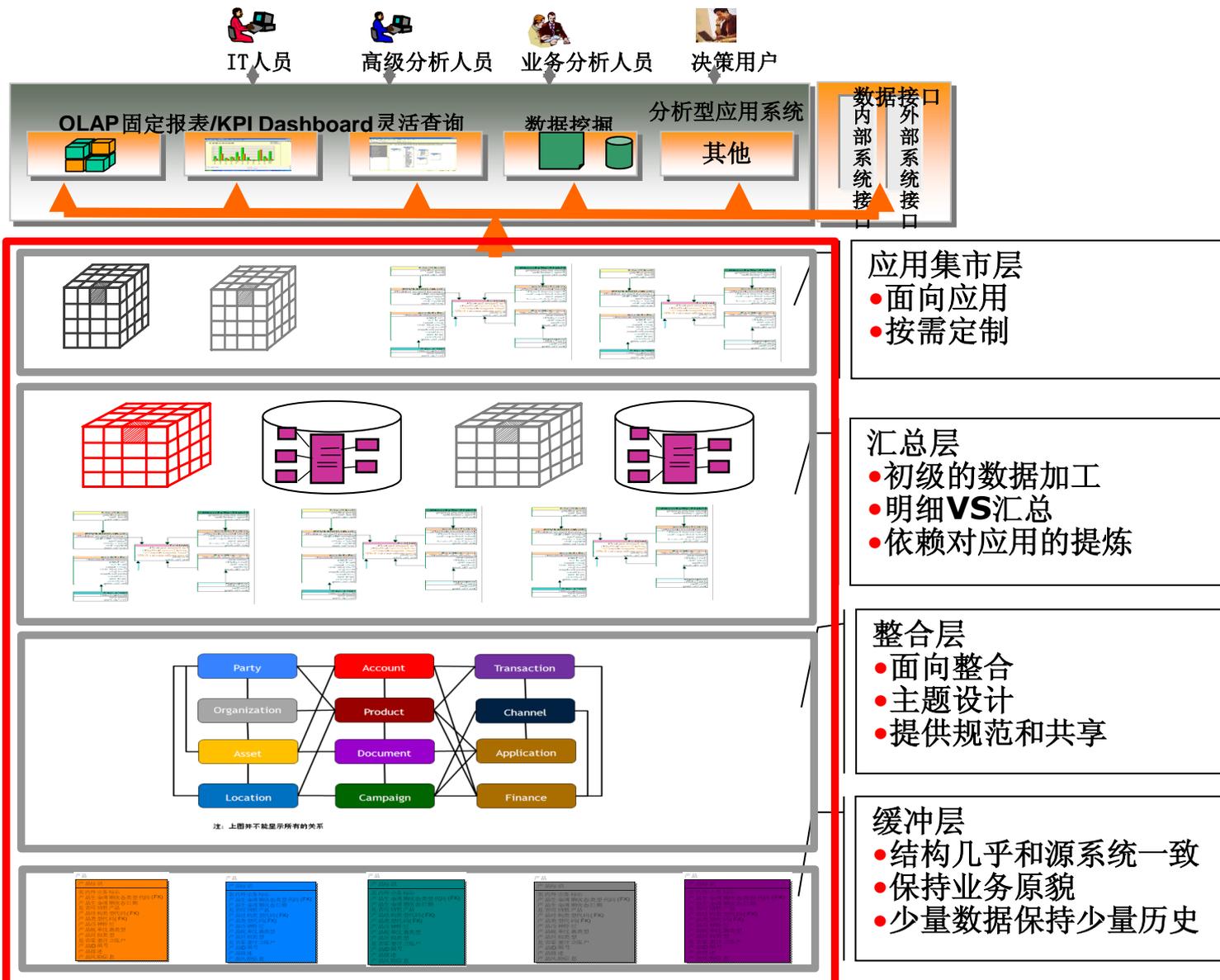
# 传统 DW/BI vs Active DW/BI



VS



# 传统数据仓库架构



# 传统数据仓库发展的思考——关于模型



数据模型是DW的核心，两大主要优点：

- 3NF，减少数据冗余，节约存储开销
- 面向主题整理，方便分析型应用

• 缺点：时效性！

- 应用时效性无法满足：缓冲层→整合层→汇总层→集市层（或应用）

传统数据仓库架构是用**时间换空间**的选择！

# 引入动态数据仓库的时机已经成熟

数据架构整体  
时效性提升

核心系统、数据抽取、数据传输、数据处理平台  
可实现（准）实时数据获取和分析

存储能力发展

02年主流DW设备使用36GB 硬盘; 06年使用  
73GB 硬盘; 目前主流使用600GB 硬盘——为”  
空间换时间”的思路调整提供了成本优势

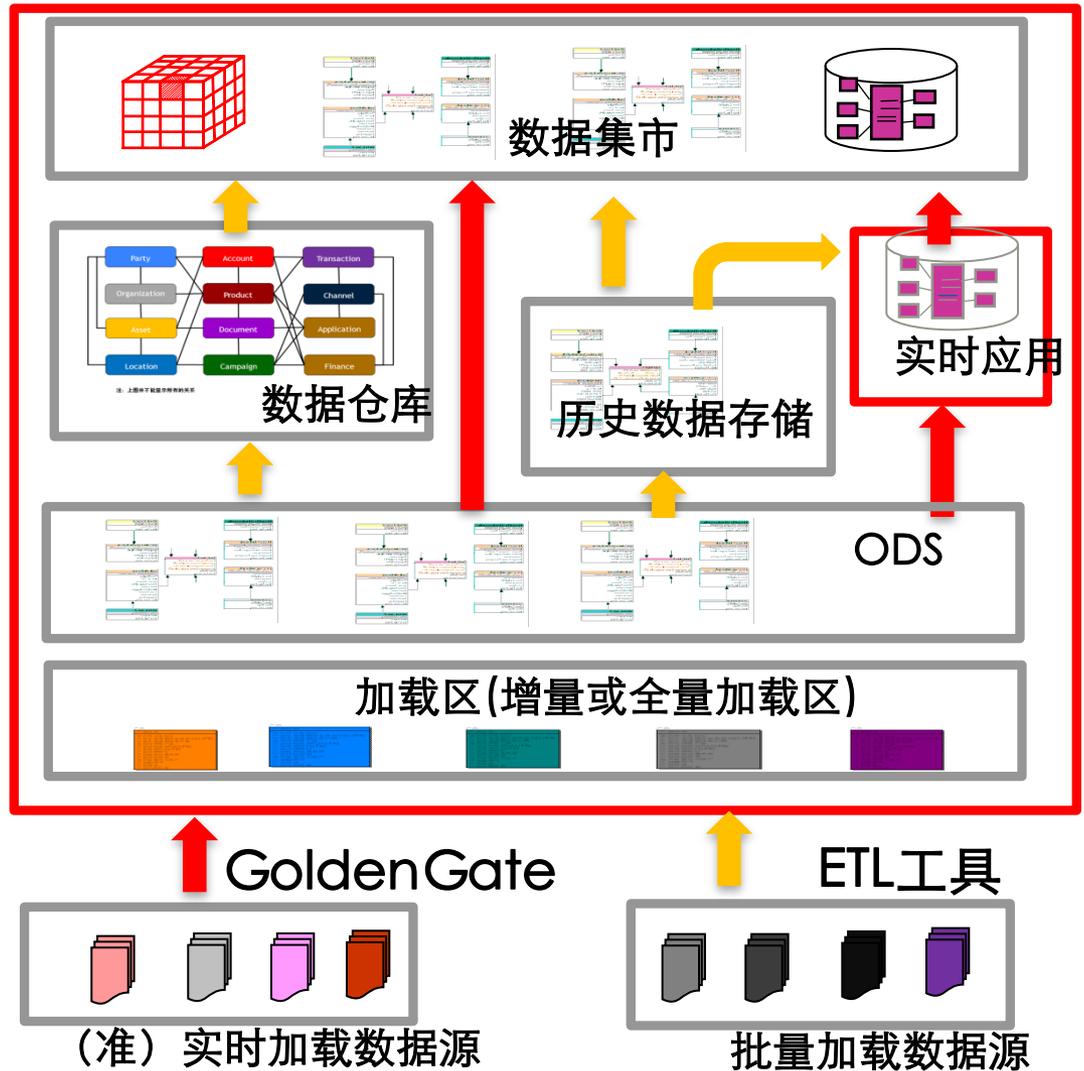
海量数据分析  
平台能力创新

Exadata的出现使同时支持OLTP&OLAP海量混  
合数据分析成为可能

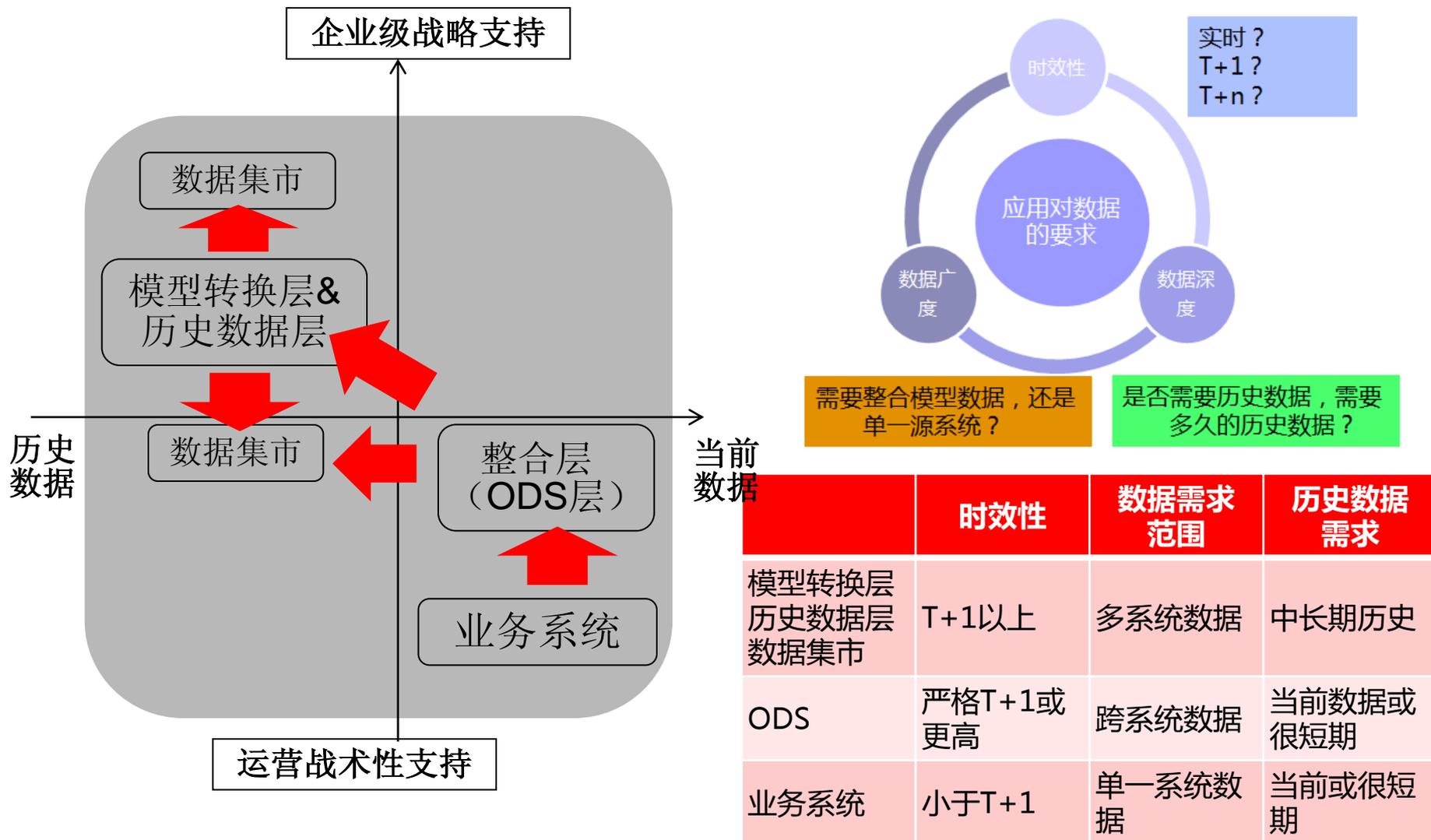
# 动态数据仓库整体架构

## Exadata在动态数据仓库架构的独特优势:

- 突出的混合负载管理能力
- OLAP & OLTP的均衡优异表现
- 同一物理平台实现多温度数据管理



# 动态数据仓库架构 —— 系统定位



# 动态智能数据中心架构——应用场景

	设计特点	应用场景
整合层 (ODS 层)	<ul style="list-style-type: none"> <li>• 贴源设计, 与源业务系统的数据模型基本相同</li> <li>• 经过简单的数据清洗和变型</li> <li>• 存储详细数据</li> </ul>	<ol style="list-style-type: none"> <li>1、追加至模型转换层</li> <li>2、支持实效性要求高的应用</li> <li>3、贴源的数据批量交换</li> <li>4、贴源系统的业务查询和灵活查询</li> </ol>
模型转换层	<ul style="list-style-type: none"> <li>• 共享的、面向主题的3NF设计</li> <li>• 不为支持某个特定的、预先定义的应用而设计</li> <li>• 存储详细、历史的数据</li> </ul>	<ol style="list-style-type: none"> <li>1、支持后台加工复杂性高的管理分析类应用(往往实效性要求不高)</li> <li>2、跨业务数据的灵活查询</li> </ol>
历史数据层	<ul style="list-style-type: none"> <li>• 贴源设计, 与源业务系统的数据模型基本相同</li> <li>• 从整合层导入数据, 基于数据生命周期管理要求, 采用不同的存储方式</li> <li>• 存储详细数据</li> </ul>	<ol style="list-style-type: none"> <li>1、基于源数据结构的应用;</li> <li>2、审计;</li> <li>3、有独立数据模型的应用</li> <li>4、追加至模型转换层</li> </ol>
数据集市	<ul style="list-style-type: none"> <li>• 针对特定主题域、部门或用户组的数据集合</li> <li>• 星形或雪花模型设计</li> <li>• 轻度或高度汇总数据</li> </ul>	<ol style="list-style-type: none"> <li>1、直接支持大量并发性数据多维分析查询</li> <li>2、通常直接支持套装软件包应用</li> </ol>

# Exadata : EDW->ADW->DBaaS

## EDW

- 企业级数据整合
- 分析型BI
- 案例：软银,BOA(美洲银行)

## ADW

- 操作型BI
- 动态BI
- 案例：BPN银行,美林证券

## DBaaS

- DBaaS
- 数据库整合
- 案例：澳洲联邦银行

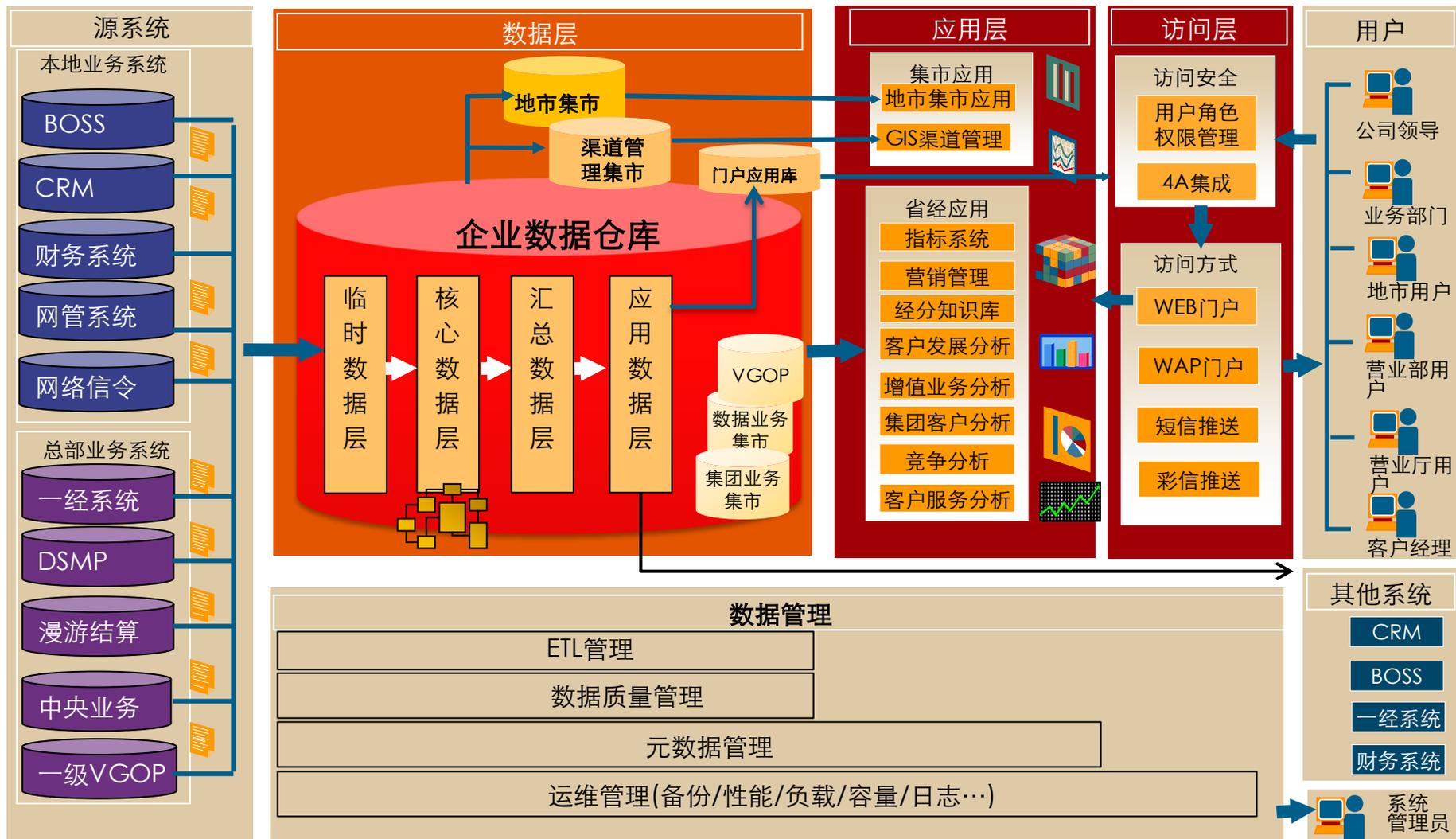
# 目录

1. 数据仓库参考架构

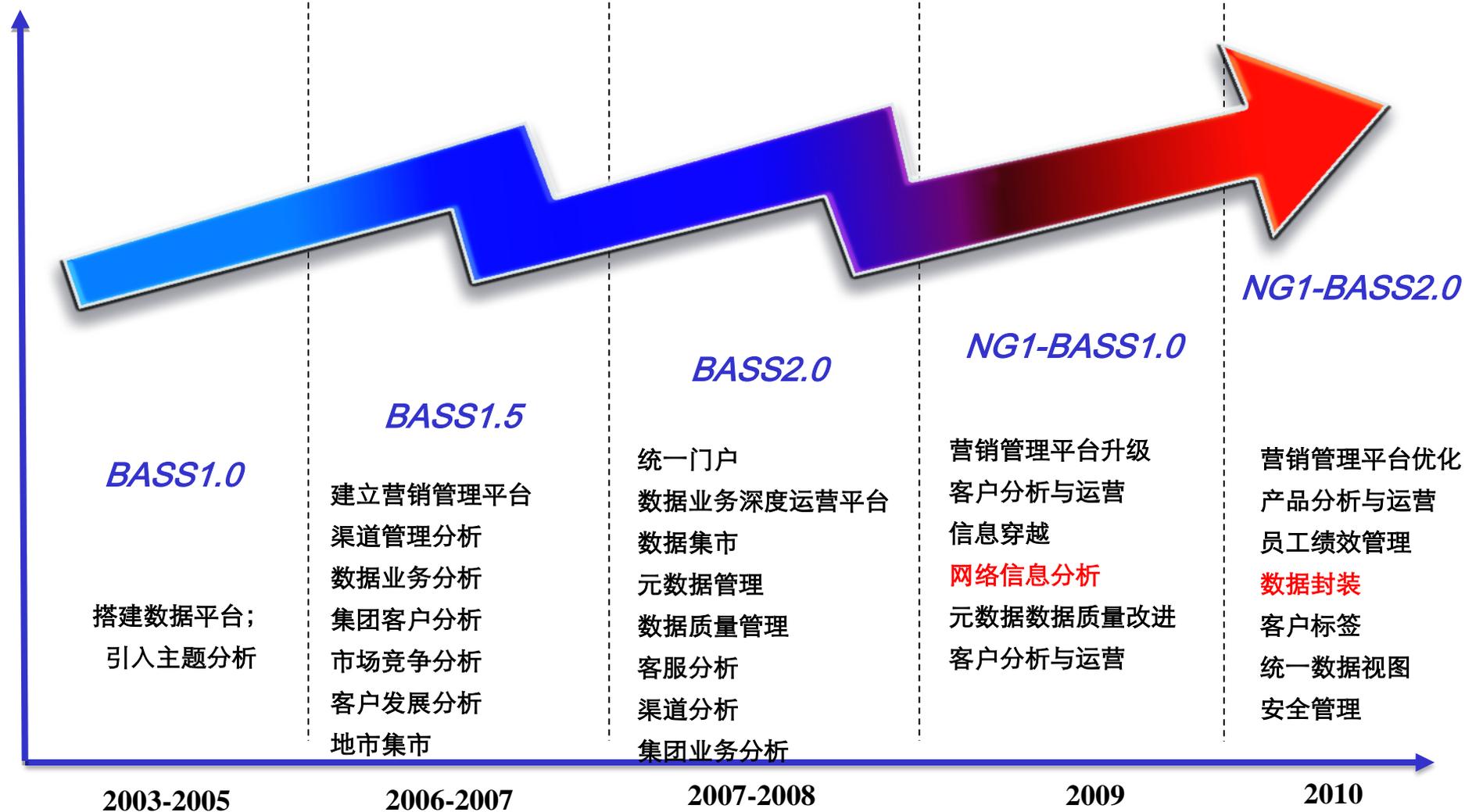
2. 案例分享

3. 数据集成产品ODI介绍

# 某电信运营商省公司数据仓库架构



# 某省电信运营商省公司系统建设历程



2003-2005

2006-2007

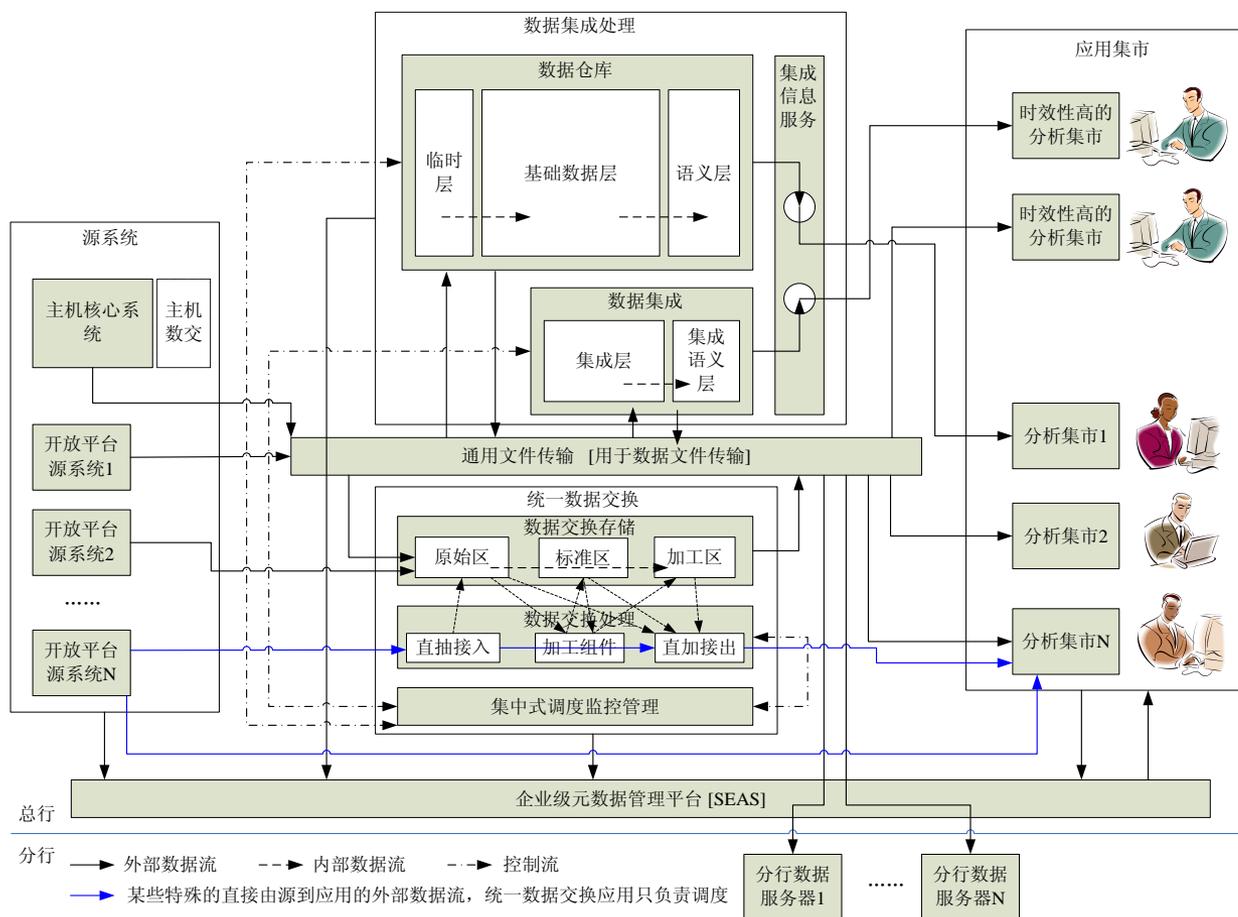
2007-2008

2009

2010

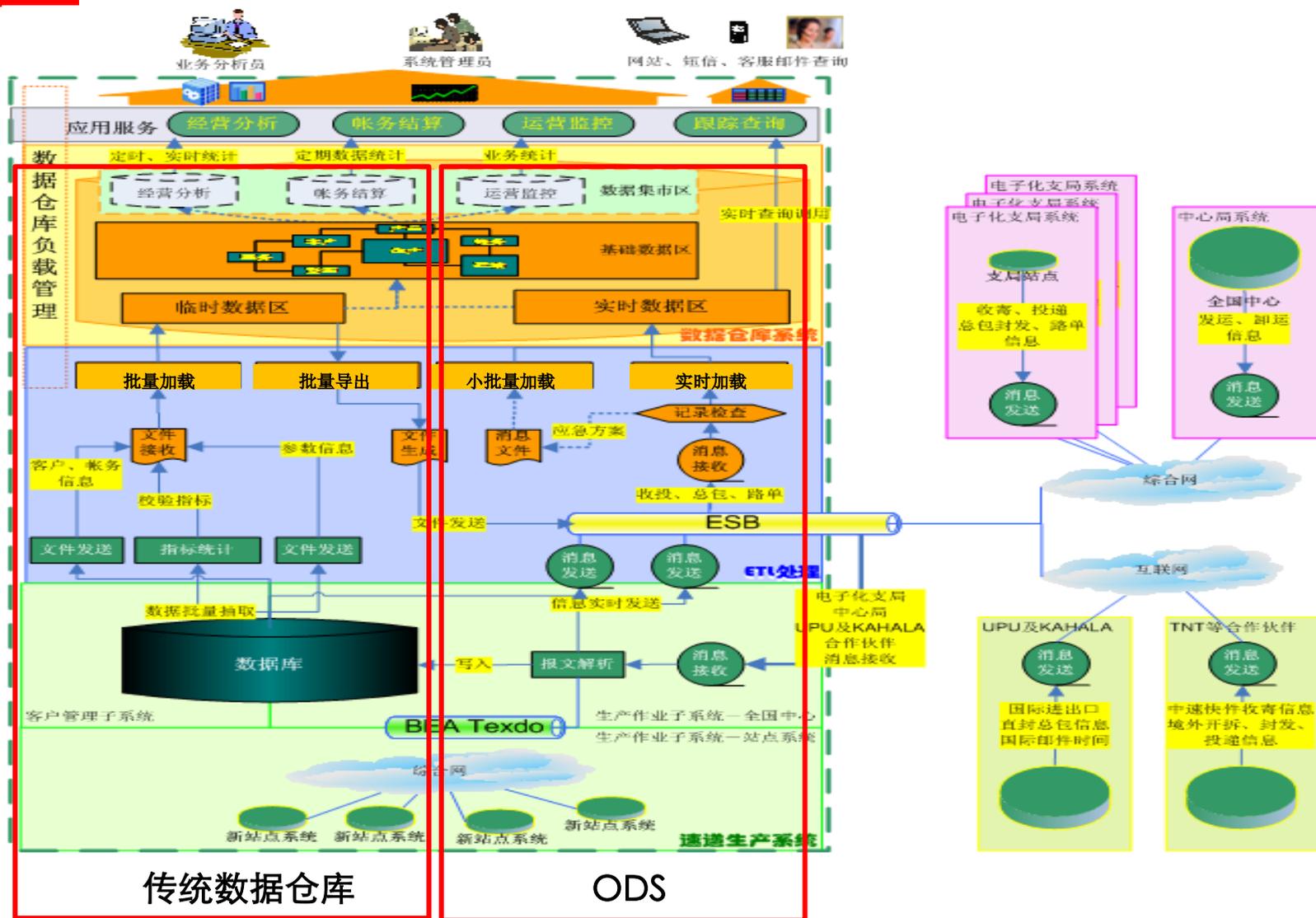
# 某大型商业银行数据仓库架构

## 应用集市区



- PCRM集市(
- 业务营运风险集市(
- 银税接口集市(
- 法人不良贷款欠款扣收集市
- 非居民个人业务统计集市
- 信用风险视图
- 理财自动化集市
- 数据质量管理集市
- 1104报表集市(
- 指标库集市(
- 非零售信用风险内部评级集市(
- 个人客户内部评级集市(
- 两化绩效评价集市
- 个人目标客户集市
- 资本金管理集市
- 私人银行集市
- ATM流程优化集市
- 信用风险集市

# 某物流企业数据仓库架构



# 基于EXADATA, 实现动态数据仓库应用

## 法国BNP银行投行部

客户需求：35个交易大厅**交易实时监控**系统；（OLTP）

合规和风控系统；（OLAP）

历史交易查询和报表系统（5年）（OLAP）

最近数据量达20TB；客户查询性能远远跟不上

客户买2台半配X2-2 Exadata（上周交易数据放在闪存中）

结果：历史交易查询提高了20倍以上；从30-50分钟到1分钟左右（5年数据）

数据库运维和变更从数个小时缩短到10-20分钟

业务部门新的查询需求可以实现而没有性能瓶颈的担忧

## 美林证券

美林证券大规模应用Oracle的Exadata和Goldengate，在IBM/TD/Netezza中全面胜出  
Bank of America 把多个业务系统的数据抽取整合到Exadata支持的ODS系统中；实现  
业务部门要求**准实时**的SLA

多个Web应用系统数据装载速度大大提高；

**支持典型的OLTP应用（连接到ODS系统的应用系统）和DW应用；语句查询速度快了14-572倍不等。**

风险控制系统、历史查询等应用是典型海量数据情况下支持OLTP和OLAP的混合应用

# 迁移至Exadata, 实现云数据整合平台

CommonwealthBank



## ——澳洲联邦银行案例分享

### ➤ 客户概况

- 澳洲最主要的房地产信贷银行、最大的信用卡发行银行、最大基金经理、最大零售退休金管理公司和最大人寿保险公司
- 1991年实施股份化, 成为上市公司
- 1994年初, 在澳大利亚政府的支持下, 澳洲联邦银行制定了把中国作为重点对象的亚洲发展计划, 长期致力于推动中澳贸易发展

### ➤ 现状与挑战

- 业务系统独立; 数据分散; 数据库紧跟着每个业务系统 (共300个左右)
- 由于数据量的不断增加, 业务出报表速度相比原先大幅下降
- 客户级别的实时查询达不到要求; TB级历史数据查询缓慢
- 数据中心的利用率和运维管理急需提高

### ➤ 解决方案

- 把**250+**个数据库和数据仓库、**30+**个左右的应用系统类 (几百个不同业务部门使用的应用) **实施整合**
  - 配置Exadata 数据库云服务器
  - Oracle EBS、Siebel CRM

### ➤ 为什么选择Exadata

- 面向数据仓库、OLTP应用程序、混合负载的超高性能
  - 卸载可伸缩的密集型查询处理和数据挖掘评分来提高查询性能管理的自动化和管控的集中化
  - 智能闪存缓存透明地将“热”数据缓存到快速固体存储上来加快查询响应速度和提高吞吐量
  - 大规模并行网格是整合数据仓库和OLTP应用程序的理想之选, 并且数据库云服务器的服务质量资源管理功能确保快速响应所有用户

### ➤ 成果

- 运营成本减少 50%
- 第 1 年即实现盈亏平衡
- 历史交易查询提高了20倍以上;
- 灵活性 — 可以根据需要从资源池中获取 CPU

# 目录

1. 数据仓库参考架构

2. 案例分享

3. 数据集成产品ODI介绍

# ODI是Oracle 数据集成整体解决方案的主要部分

IT data Metadata, data analyses      Real-time data BAM, Real-time BI      Corporate data Traditional BI, CPM

高性能，智能  
数据库存储一  
体机系统

数据评估与质  
量管理

Data Profiling      Data Quality

高效的数据整  
合平台

Oracle Data Integrator

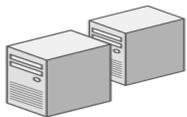
变化数据捕获  
CDC

Oracle GoldenGate

元数据管理平  
台

Metadata Manager

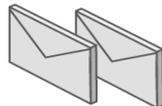
ExaData



ERP



Databases



Messages



Flat files



XML

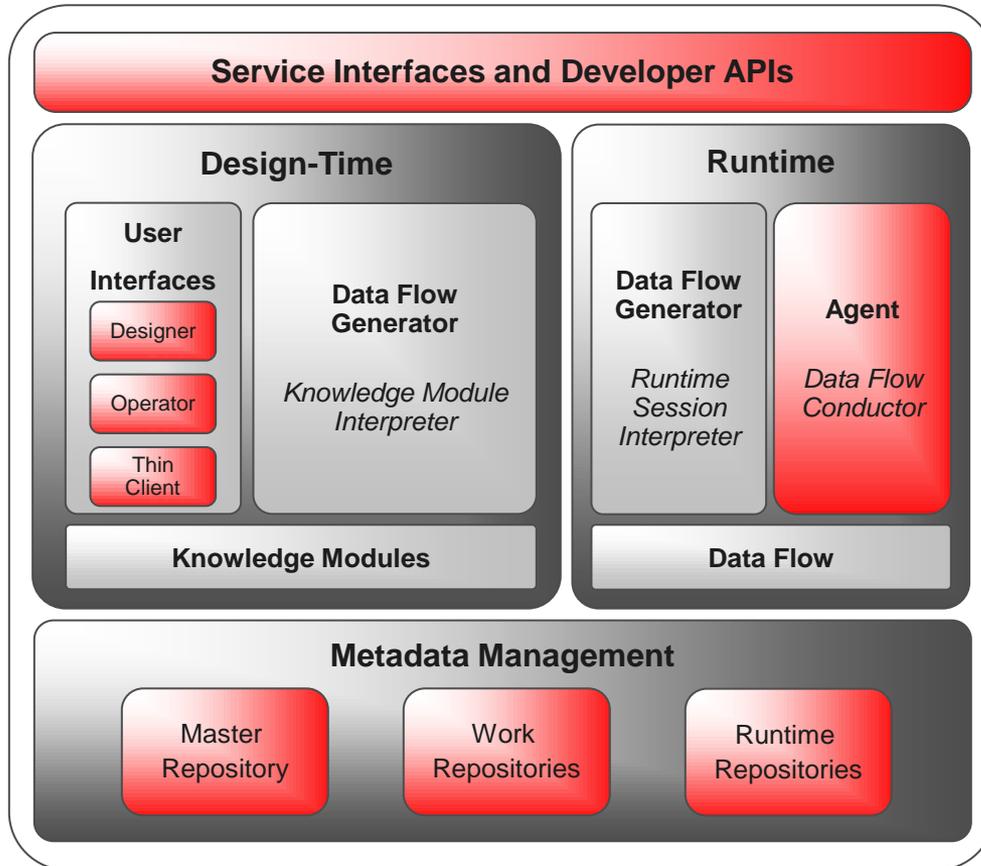


Unstructured Data



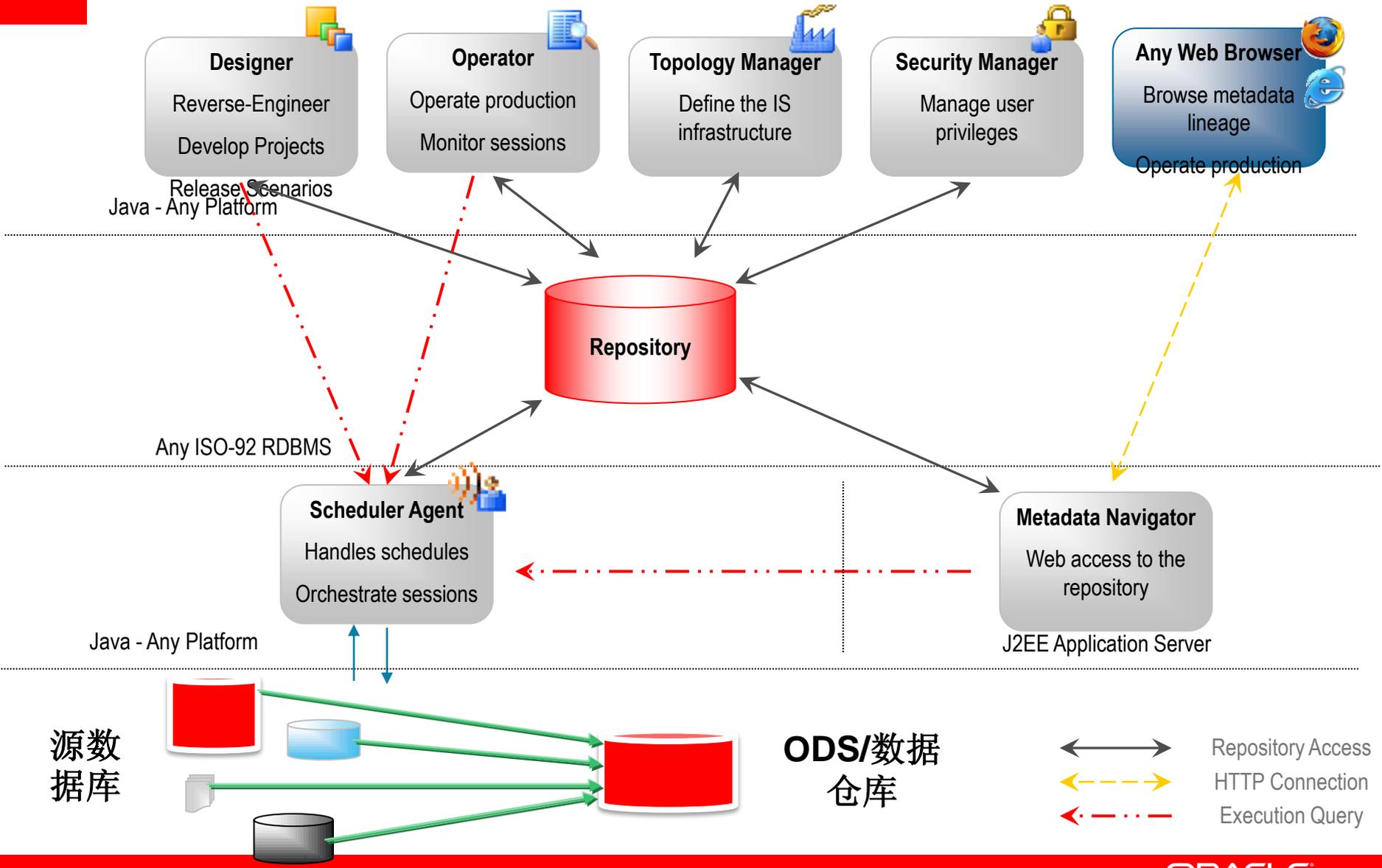
Mainframe

# 架构: 概念视图



- 纯Java的设计环境
  - 可以在任何平台上运行
  - 通过瘦客户端浏览元数据
- 纯Java的运行环境
  - 可以在任何平台上运行
  - 协调执行数据流
- 元数据库
  - 可以装在任何关系数据库上
  - 便于开发
  - 组件化的、可扩展的元数据

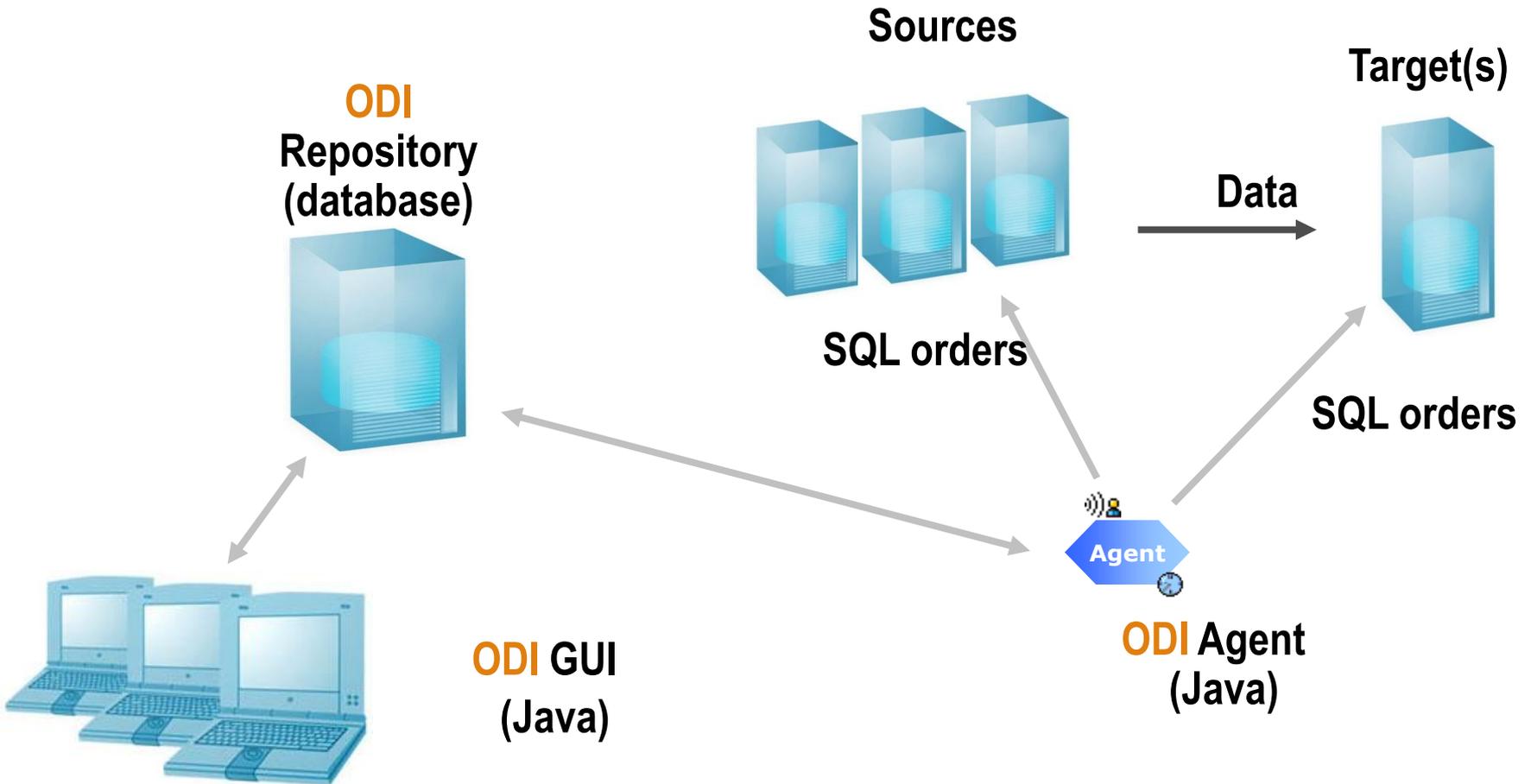
# ODI整体架构



源数据库

ODS/数据仓库

# Oracle Data Integrator 架构示例



# Oracle Data Integrator的4大技术优势

- 将多个数据源中的数据迁移、转换到不同种类的目标中

## 优势

1. 高性能:
2. 高效性:
3. 开放体系架构:
4. 热插拔:

## 关键的特性

高效的“E-LT”系统架构  
声明式设计（所见即所得）  
**100%JAVA和SOA无缝集成**  
通过知识模块动态支持各种系统

# 高性能: 采用E-LT体系架构

## 传统的ETL: 单独的ETL 服务器

- 私有的 ETL 服务器
- 性能差
- 需要购买单独的硬件服务器
- IBM & Informatica采用这种方式

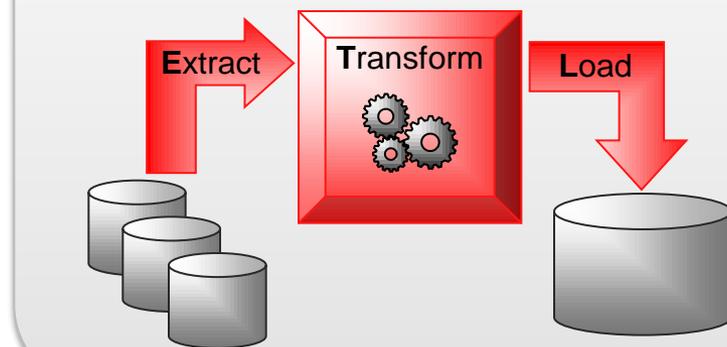
## Oracle: 不需要单独的硬件服务器

- 费用低: 充分利用计算资源 & 有效分发工作负载
- 高效: 有效利用数据库的优化技术
- 速度快: 调用数据库的内部接口, 如批量加载
- 高扩展: 可以在源端或者目的端加强计算能力, 有效分发负载

## 优势

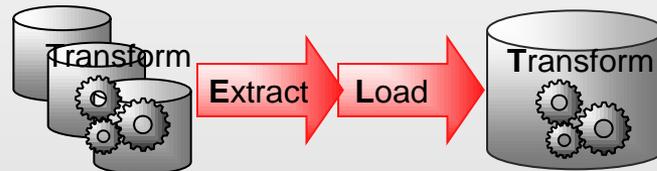
- 优化的效率和可扩展性
- 充分利用现有硬件资源
- 管理简单, 费用低

## 传统的 ETL 架构



## 下一代的ELT架构

# “E-LT”



# 特性: 声明式设计

提高开发人员生产力

## 需要指定每个ETL的数据流

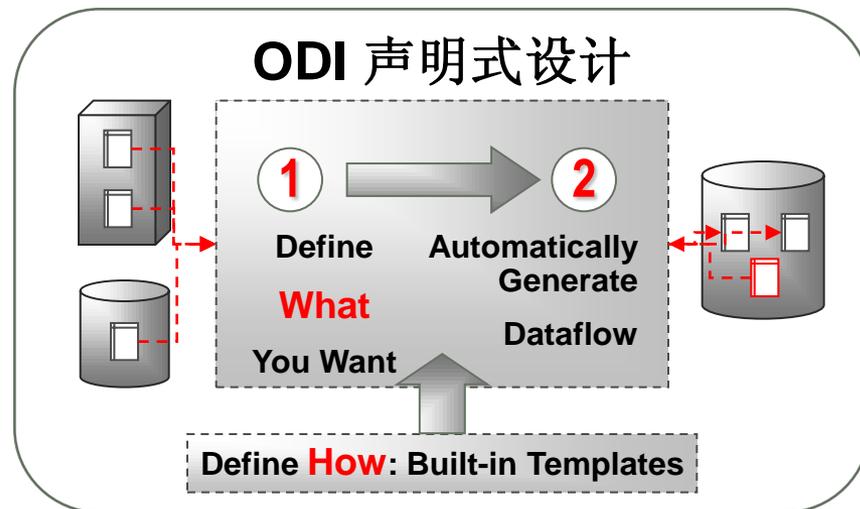
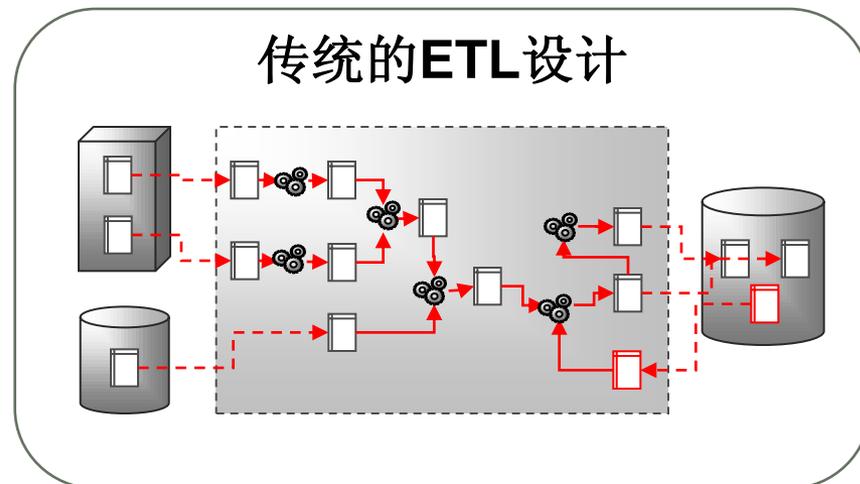
- 开发人员必须定义每个步骤
- 传统的开发方法, 需要专门的ETL技术
- 开发时间长, 维护困难

## 声明式设计

- 简化的步骤
- 根据源和目标数据库自动产生数据流
- 易于定制化

## 好处

- ✓ 便于知识传递
- ✓ 缩短开发时间
- ✓ 非IT人员也能使用



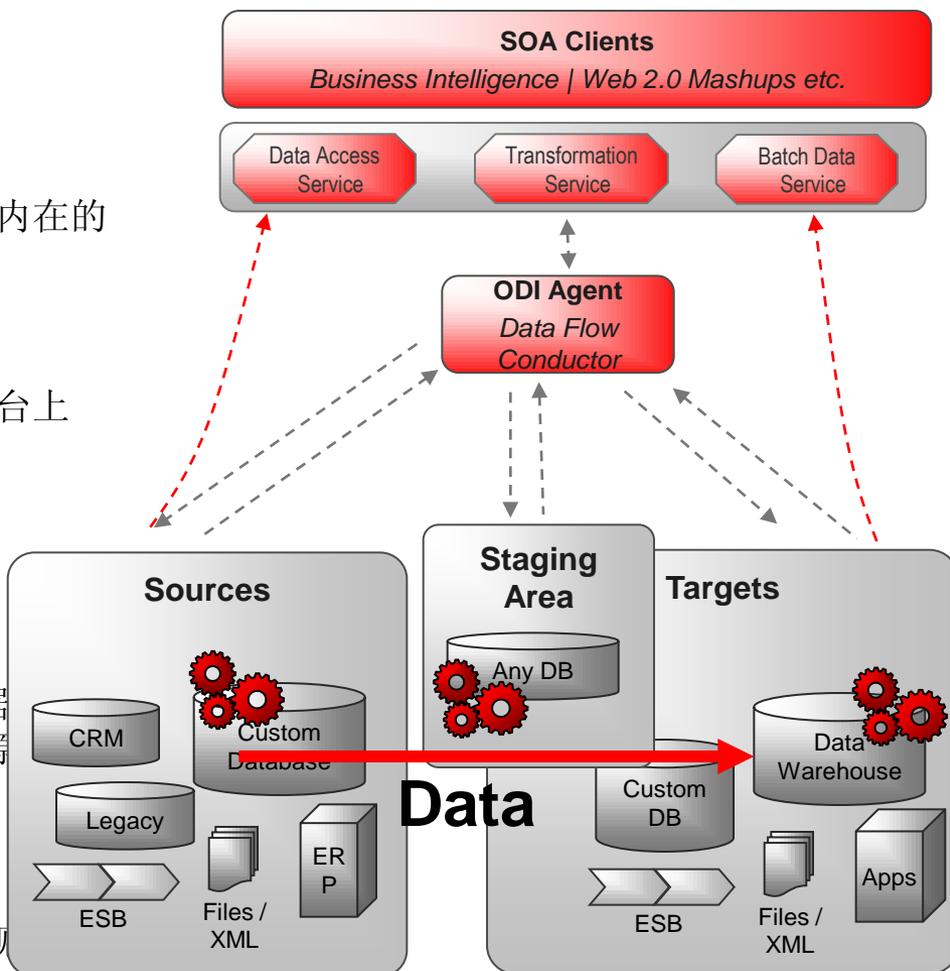
# SOA ready: 内在的SOA架构

传统ETL工具: 主要基于C++

- 过时的重量级体系架构
- SOA是系统的附加- 不是一体集成的
- 对Java的支持只是一个“wrapper” - 不是内在的

ODI: 轻量级的 & 100% Java实现

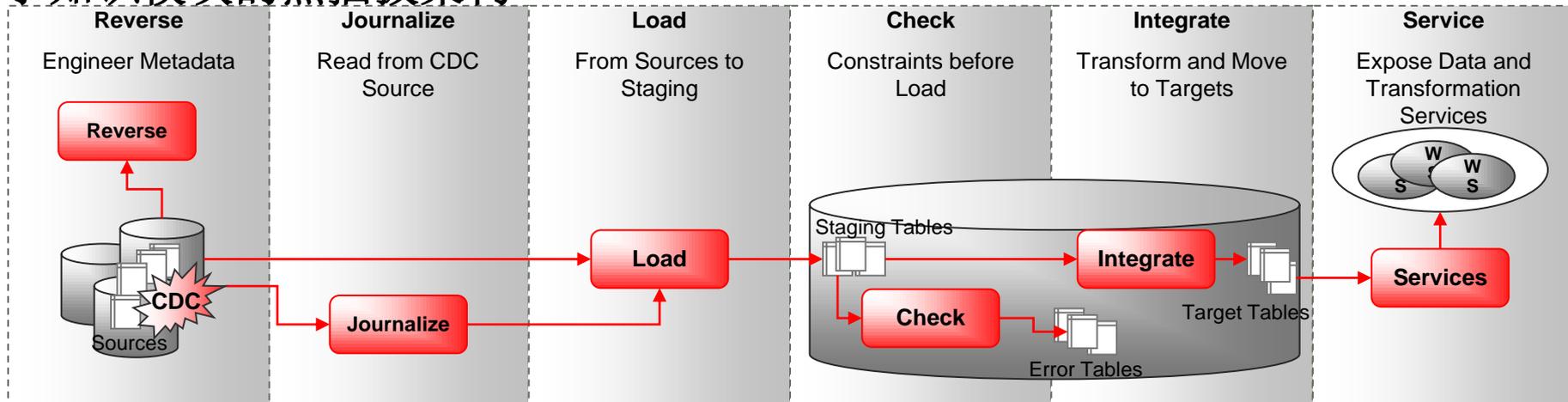
- 轻量级的JAVA运行引擎
  - <36MB memory, 在任何支持JAVA的平台上
- 可以通过Web Service访问数据流程
  - 开箱即用的数据访问服务
  - 将服务发布到UDDI (Service Registry)
  - 易于测试发布
- 数据清洗转化服务
  - 可以部署成基于SQL的转化(结构化的数据源)
  - 或者基于XSL-T的转化(非结构化的数据源)
- 非常容易的实现SOA集成
  - 在ESB中调用数据流
  - 通过SOA调用工作流程来进行错误处理
  - 作为服务嵌入到Oracle SOA套件中, 实现



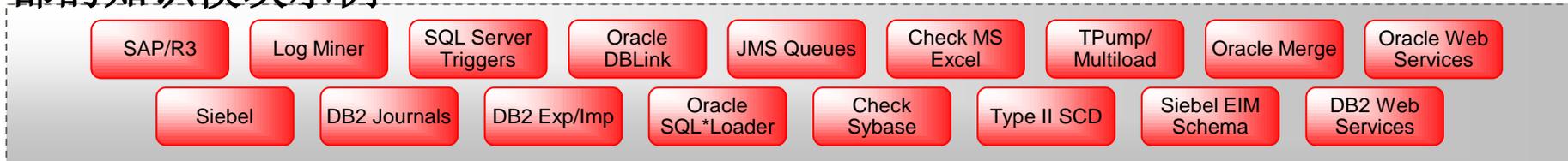
# 特性: 知识模块

热插拔: 组件化的, 灵活的, 可扩展的

## 基于知识模块的热插拔架构



## 外部的知识模块示例



## 好处

- 充分利用数据库的优化功能:  
Native SQL; Native Functions; Native Loads; Native Journaling / CDC
- 针对企业的最佳实践可以对其进行优化
- 易于管理
- 性价比好

# 广泛的数据源支持

结构化, 半结构化, 非结构化

## DB & Application KMs

- Generic SQL DB
- Oracle DB 9i
- Oracle DB 10g
- Oracle DB 10g XE
- IBM DB2/400
- IBM DB2/UDB
- IBM Informix SE
- IBM LDAP Server
- MS SQL Server 2000
- MS SQL Server 2005
- MS SQL Server 2005 SE
- MS Office Access 2000
- MS Office Excel 2000
- MS Active Directory
- Sybase ASA 8.x & 9.x
- Sybase IQ 12.x
- Sonic MQ v7.0
- Teradata V2R5.x
- Teradata V2R6.x
- Netezza PerformanceServer 2.2.1
- Hyperion Essbase
- PostgreSQL 8.1
- MySQL 4.0
- MySQL 5.0
- Oracle BI Suite 10g
- Oracle BAM 10g
- Oracle Internet Directory 9i
- OpenLDAP 2.3
- Siebel CRM 7.8
- JD Edwards
- PeopleSoft
- SAP R/3
- Oracle EBusiness Suite
- Oracle AQ 10g

## Application Adapters

- Oracle ESB 10g
- Salesforce.com App Exchange
- Any JMS Standard Implementation

- Ariba
- AXIOM mx/open
- Baan
- BroadVision
- Clarify
- Commerce One
- Hogan Financials
- i2 Technologies
- Lawson
- Livelink
- LotusNotes
- Manugistics
- Microsoft CRM
- Vantive
- Walker Interactive
- Remedy
- Sales Force

## Data Adapters

- AccountMate S
- ALLBASE/SQL
- Alpha Four
- Apache Common Logfile
- BizTalk XML
- Btrieve DDF
- Clarion
- Clipper
- Cloudscape
- Data Junction Logfile
- DataEase
- DataFlex (ODBC 3.x)
- IBM DB2
- DB2/400
- DBASE
- Dialog
- Dodge Bidders
- Enable
- DB2 (\*)
- ERWIN
- Essbase
- Excel
- eXcelon
- FOCUS Data Access
- Folio Flat File T
- Foxbase+
- FoxPro
- GoldMine
- Great Plains
- Hitachi HiRDB (ODBC)

- IDAPI
- IDS-II
- ImageSQL
- Informix(\*)
- Ingres
- MicroFocus Cobol
- Microsoft Access
- Microsoft SQL Server(\*)
- MUMPS
- Navision Financials
- Nucleus
- OpenIngres
- Oracle(\*)
- Paradox
- Pointbase
- PostgreSQL
- Progress
- Quattro Pro Windows v5
- Rbase (ODBC 3.x)
- Rdb
- Red Brick
- RMS
- Statistica
- SUPRA
- Sybase(\*)
- Teradata
- UniVerse
- Velocis (ODBC 3.x)
- Visual dBase 5.5
- Visual FoxPro
- XDB

## Legacy Adapters

- Adabas
- Bull TDS
- Bull TP8
- CA-Datcom
- CA-IDMS/DB
- CA-IDMS/SQL
- C-ISAM (Informix)
- C-ISAM (Microfocus)
- C-ISAM ACCUCORP
- CICS(\*)
- D-ISAM
- Hitachi HiRDB (ODBC)

- IMS/DB(\*)
- IMS/TM(\*)
- Software AG Natural
- Tuxedo(\*)
- Unisys DMS 1100/2200
- VSAM(\*)

## Tech Adapters

- ActiveX
- AQ(\*)
- COM / DCOM
- CORBA
- Email
- Flat Files(\*)
- FTP(\*)
- HTTP / HTTPS
- IBM MQSeries(native)
- JMS(\*)
- Microsoft DTS
- MS MQ (Native)
- NET
- Oracle Advanced Queues
- RPG
- Socket
- Telnet
- Tibco Rendezvous (Native)

## B2B Standards

- EDI
- UCCNet
- RosettaNet
- CIDX
- PIDX
- VICS
- ebXML
- UBL
- UN/EDIFACT
- X12
- X12, Property & Casualty
- X12, HIPPA
- X12N, Life & Annuity
- X12N, Healthcare
- NCPDP SCRIPT
- HL7
- OAG
- cXML
- xCBL

## Content Adapters

- DEC WPS Plus
- Display Write 2, 3, 4 & 5
- Enable (Writer & Spreadsheet)
- First Choice
- Framework
- IBM Writing Assistant
- Lotus Manuscript
- Lotus AMI/AMI Professional
- Microsoft Word & Works
- MultiMate
- Novell WordPerfect
- Office Writer
- PFS: Write
- Total Word
- Wang PC (IWP)
- WordMARC
- WordStar
- JustSystems Ichitaro
- JustWrite
- Adobe FrameMaker
- Lotus Word Pro
- Microsoft Write
- Microsoft Word
- Microsoft WordPad
- Microsoft Works
- MacWrite II
- Microsoft Word (Mac)
- Novell Perfect Works
- OpenOffice Writer
- Professional Write Plus
- StartOffice Writer
- WordStart
- Lotus 1-2-3
- Lotus Symphony
- Microsoft Excel (all versions)
- Mosaic Twin
- Novell Perfect Works (SS)
- OpenOffice Calc
- PFS: Professional Plan
- Quattro Pro (DOS)
- Quattro Pro (Win)
- SmartWare II
- StarOffice Calc
- SuperCalc 5
- VP Planner 3D

# Q&A

