

大数据大挑战 整合是关键

但彬

Informatica中国

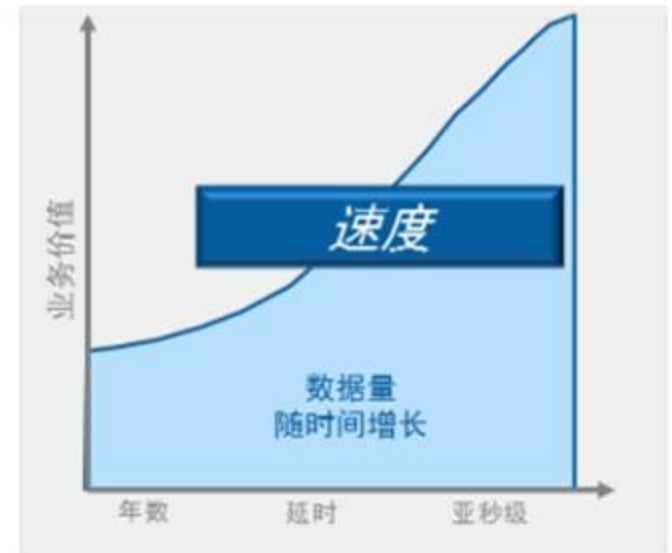
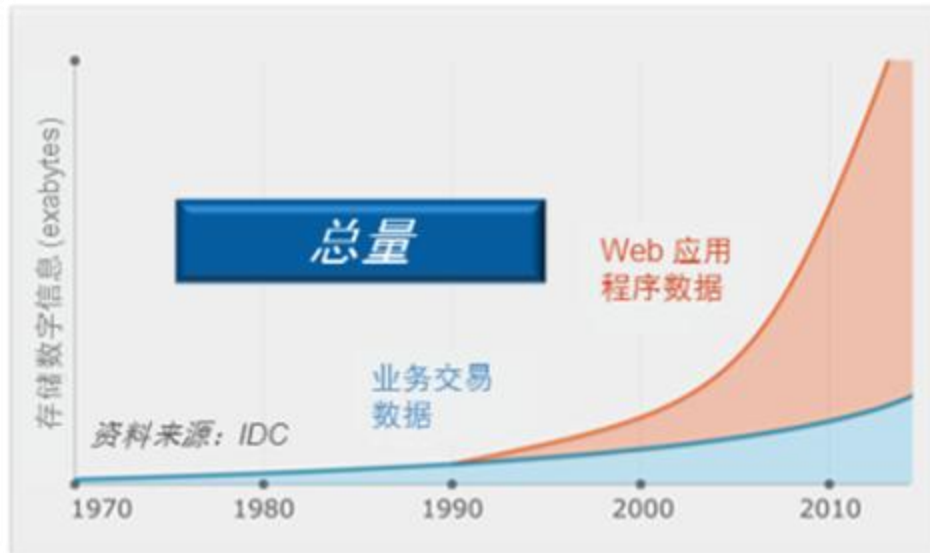
bdan@informatica.com

技术突破带来互联时代

- 24小时在线(Connected 24 hours Era)
- 全方位网络覆盖

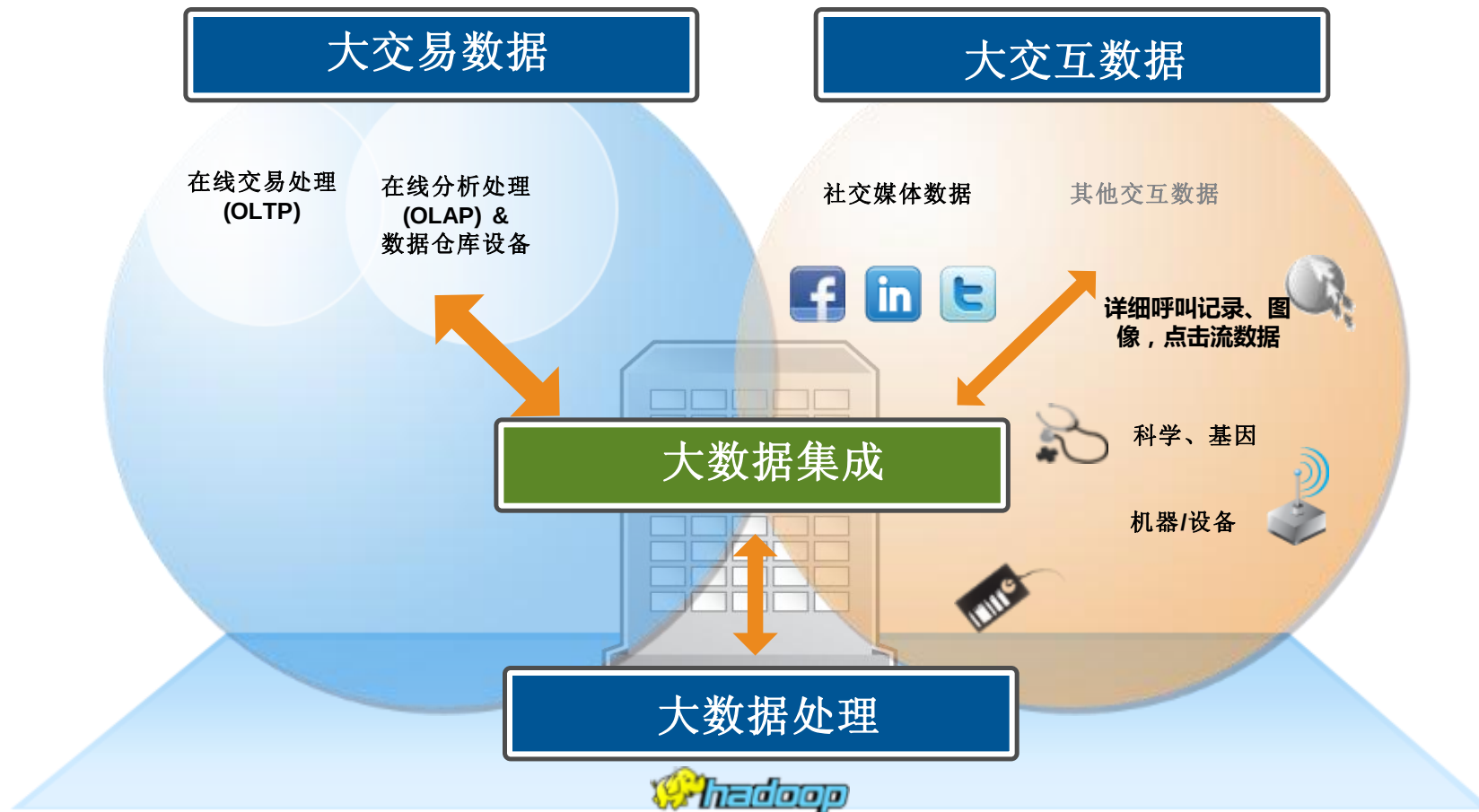


随之而来的数据爆炸



Big Data成为流行语

大数据(Big Data)是大交易数据, 大交互数据和大数据处理的统称。



大数据是关于...

创新

“到2015年，那些将高价值、多样和最新的信息类型及来源集成到统一连贯的信息管理基础设施的组织，其财务表现将较业内同行优越20%以上。”，

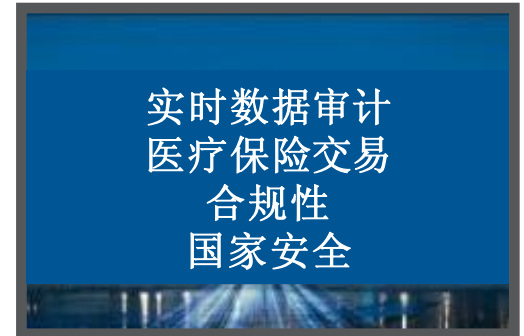
Neil Chandler, Gartner

成本

“当前部署的85%数据仓库项目，都不能快速扩展规模以满足新的信息数量和复杂性要求”，“

Mark Beyer, Gartner

您打算如何利用大数据来开发创新产品和服务？



实施久经验证
的创新之路

**数据
回报率**

=

数据价值

数据成本

随着数据呈指数级增长，
降低大数据成本

大家都说大数据是座金山…



Hadoop的特点

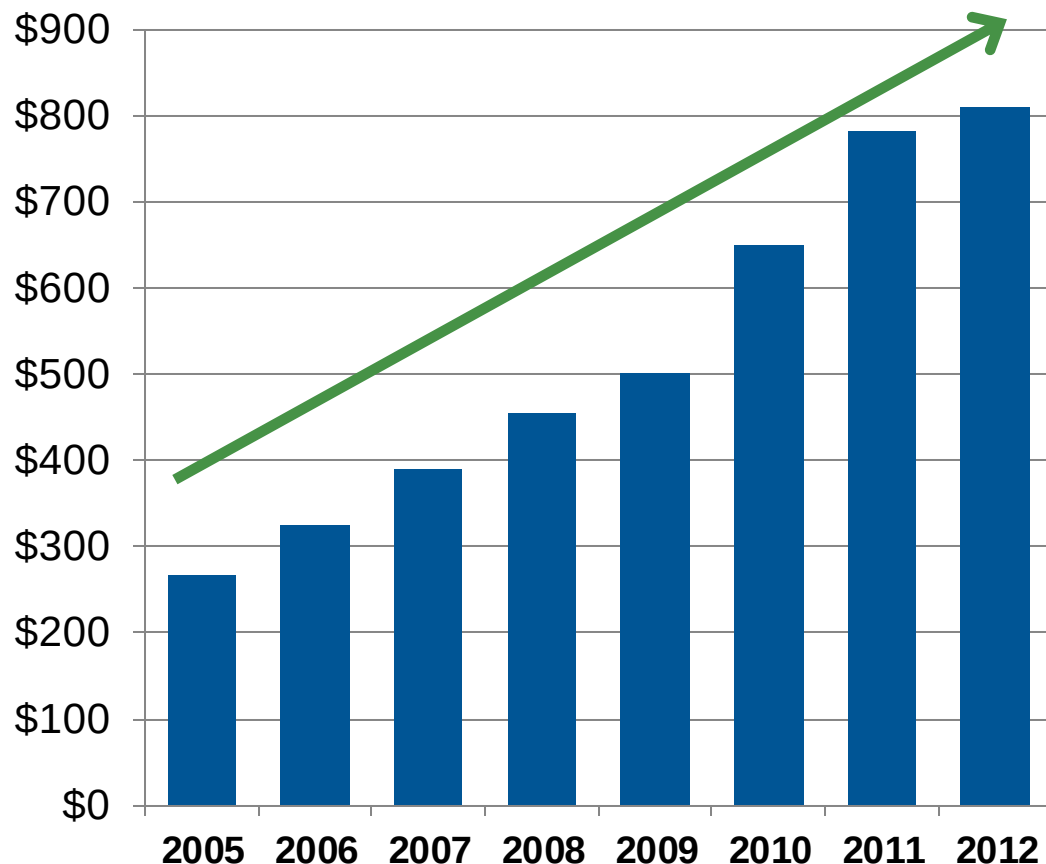
- 扩容能力（**Scalable**）
 - 能可靠地（reliably）存储和处理千兆字节（PB）数据。
- 成本低（**Economical**）
 - 可以通过普通机器组成的服务器群来分发以及处理数据。这些服务器群总计可达数千个节点。
- 高效率（**Efficient**）
 - 通过分发数据，hadoop可以在数据所在的节点上并行地（parallel）处理它们，这使得处理非常的快速。
- 可靠性（**Reliable**）
 - hadoop能自动地维护数据的多份复制，并且在任务失败后能自动地重新部署（redeploy）计算任务。

HDFS和MapReduce是核心

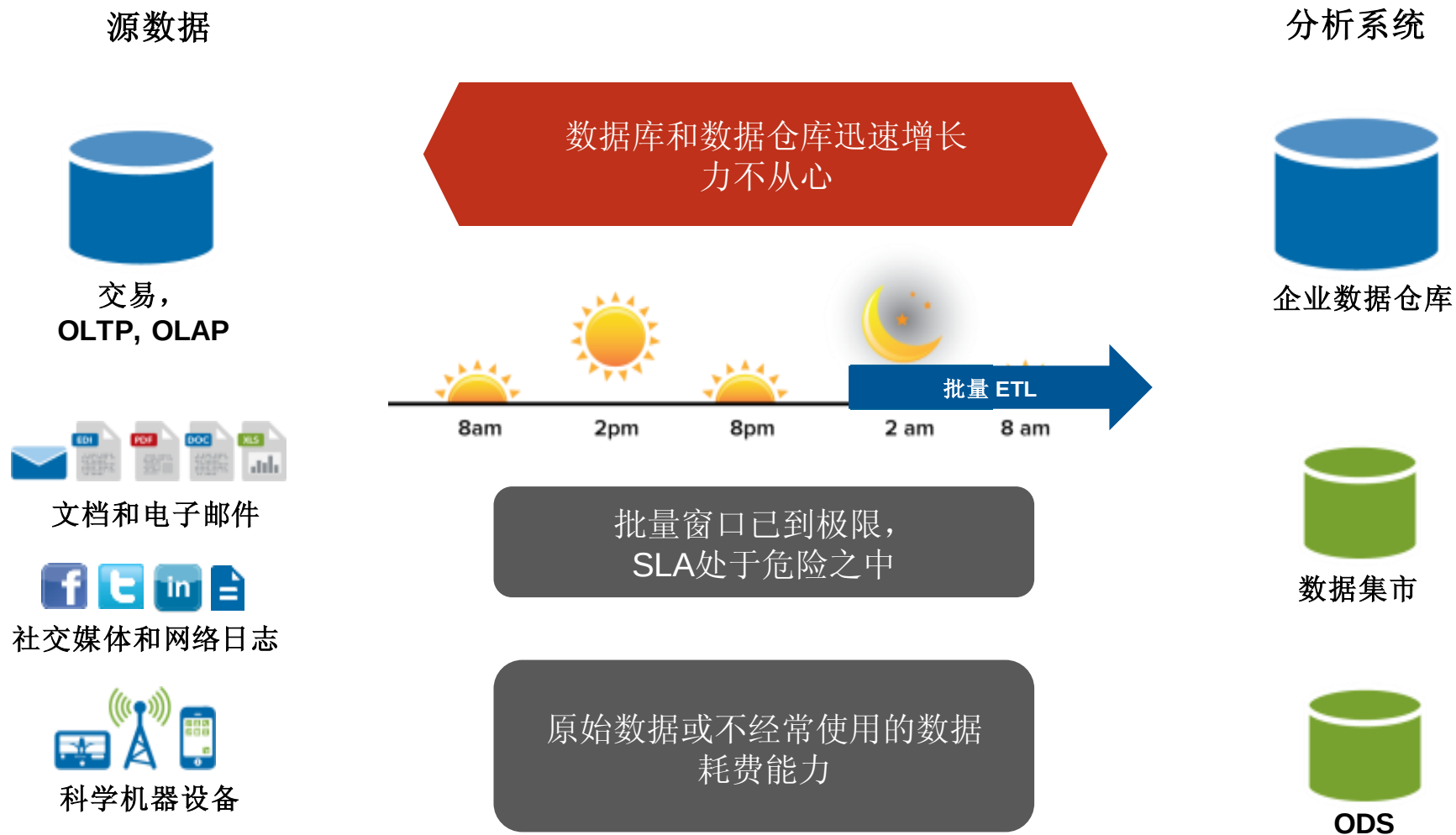
Informatica

数据集成行业翘楚

- 成立时间：**1993**
- **2012年收入为：8.116亿美元**
- **7年平均增长率：每年17%**
- 员工人数：**2,810多人**
- 合作伙伴：**450多家**
 - 主要的系统集成商、独立软件供应商、原始设备制造商和按需生产的先驱
- 客户：**超过5,000家**
 - 客户遍布全球82个国家
 - 直接客户遍布全球28个国家
 - 客户忠诚度排名位居榜首（已连续7年获此殊荣）



数据量持续增长，您如何降低 & 控制成本？



降低数据管理成本

- 将原始数据临时存储在低成本的商用硬件上
- 将 ETL/ELT 处理转移到低成本的商用硬件上



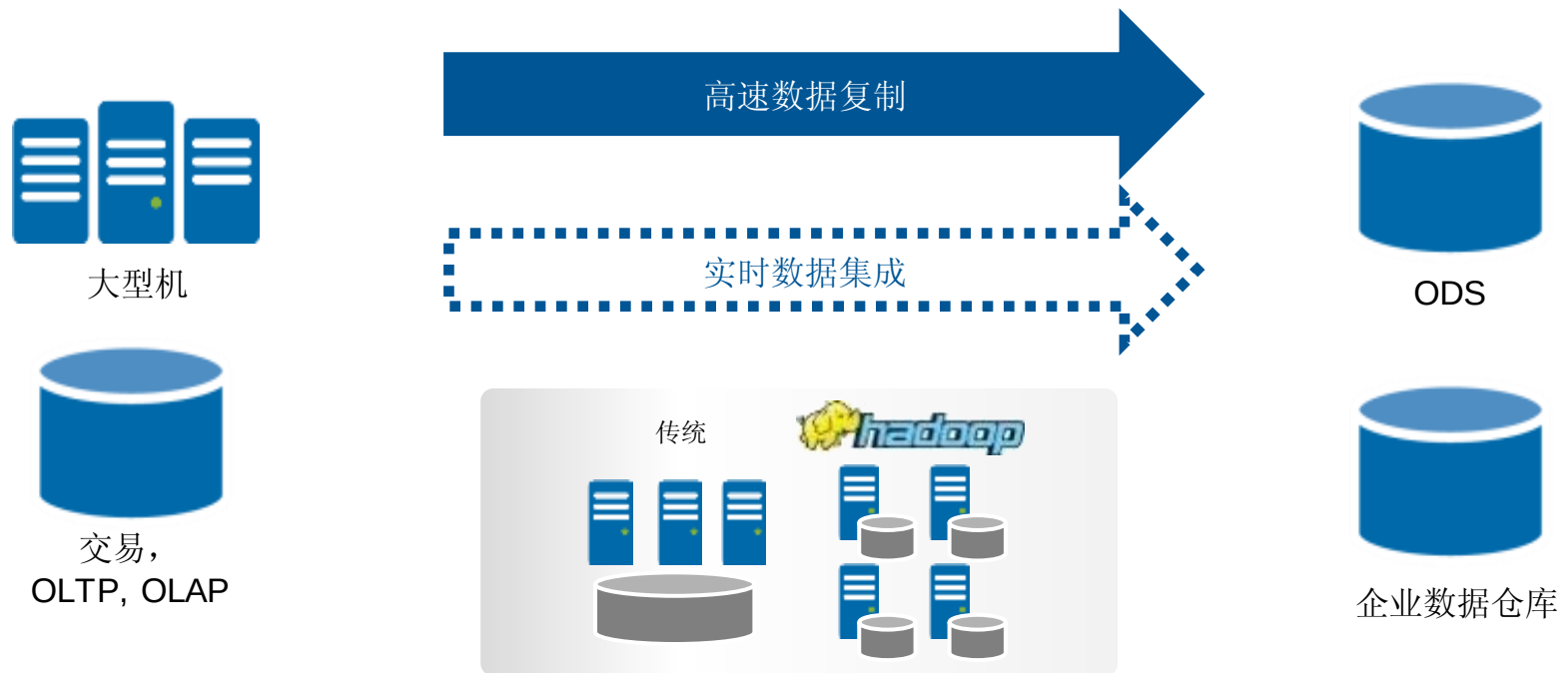
交易，
OLTP, OLAP



企业数据仓库

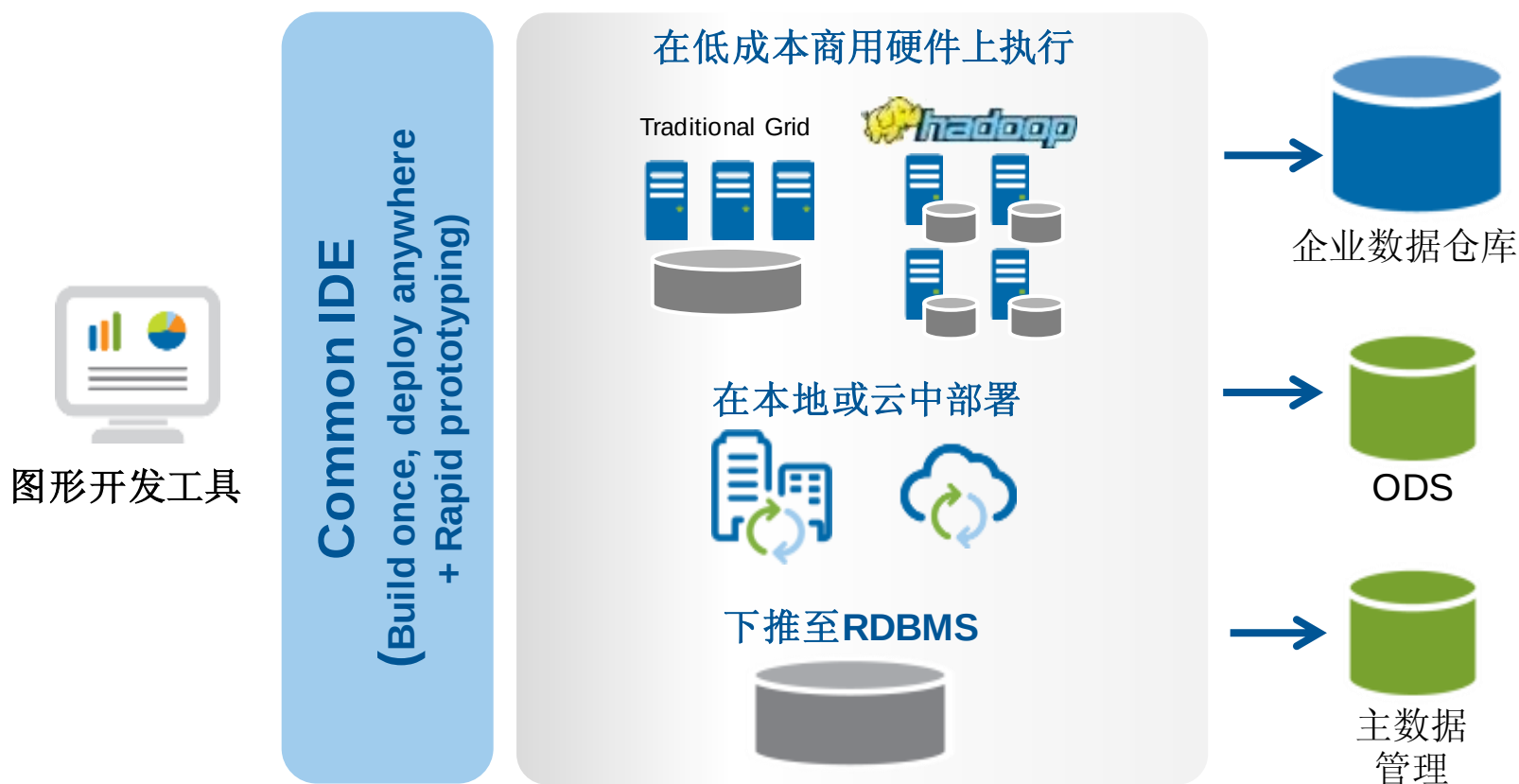
降低数据管理成本

- 借助实时数据集成，平滑实现ETL处理
- 借助高速数据复制，从源系统中卸载处理



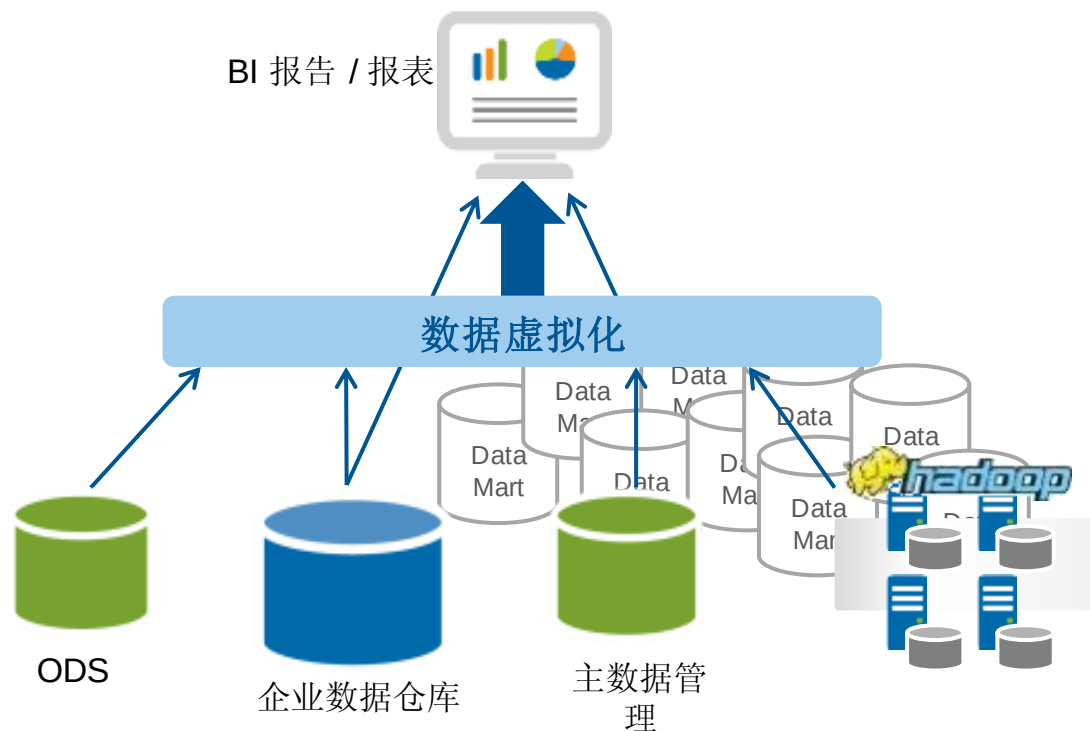
降低数据管理成本

- 借助通用的**IDE**，将生产效率提升两倍。开发人员通过一次开发，即可实现随地部署。



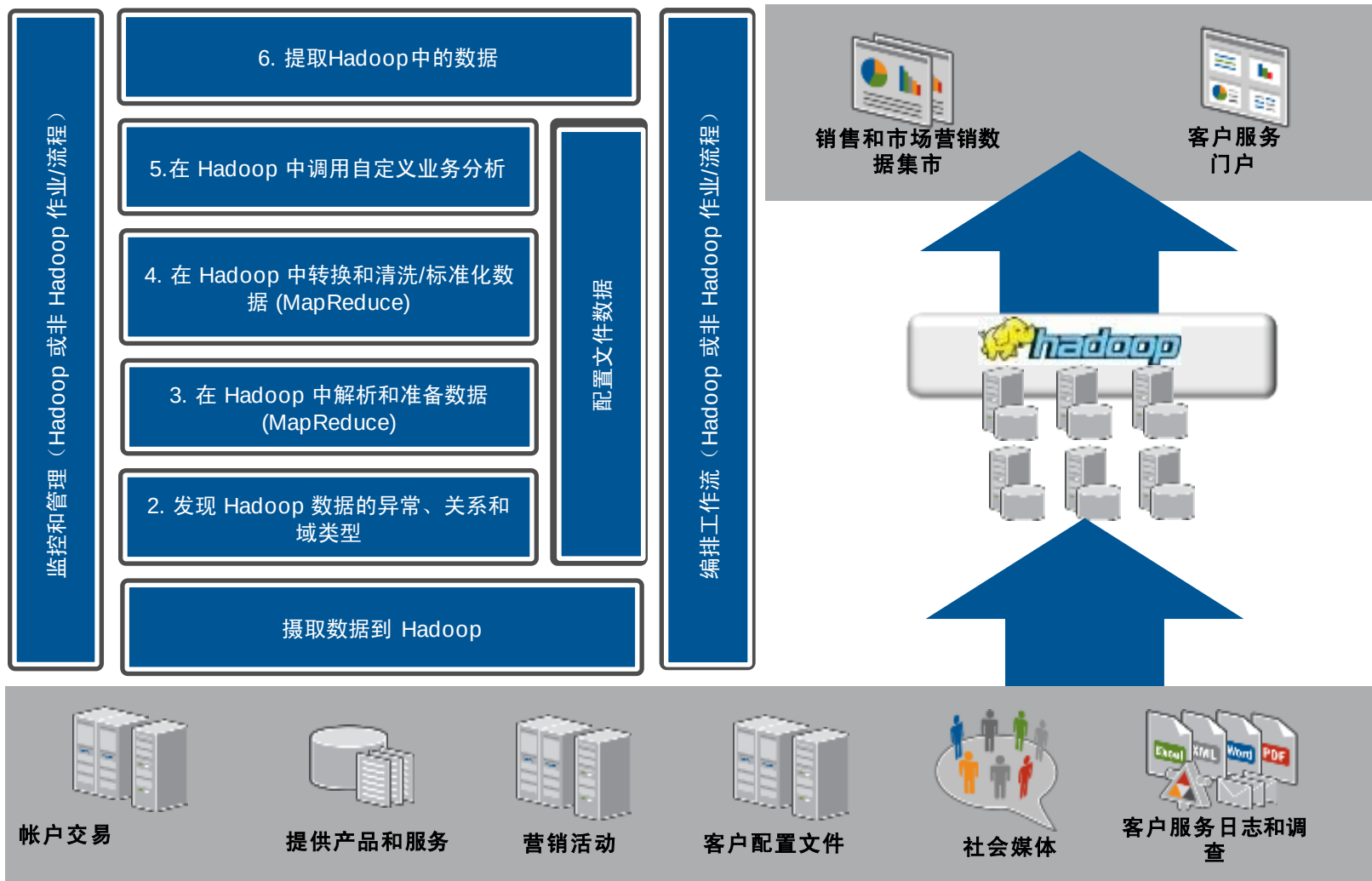
降低数据管理成本

- 消除数据副本，通过数据虚拟化提升数据仓库能力



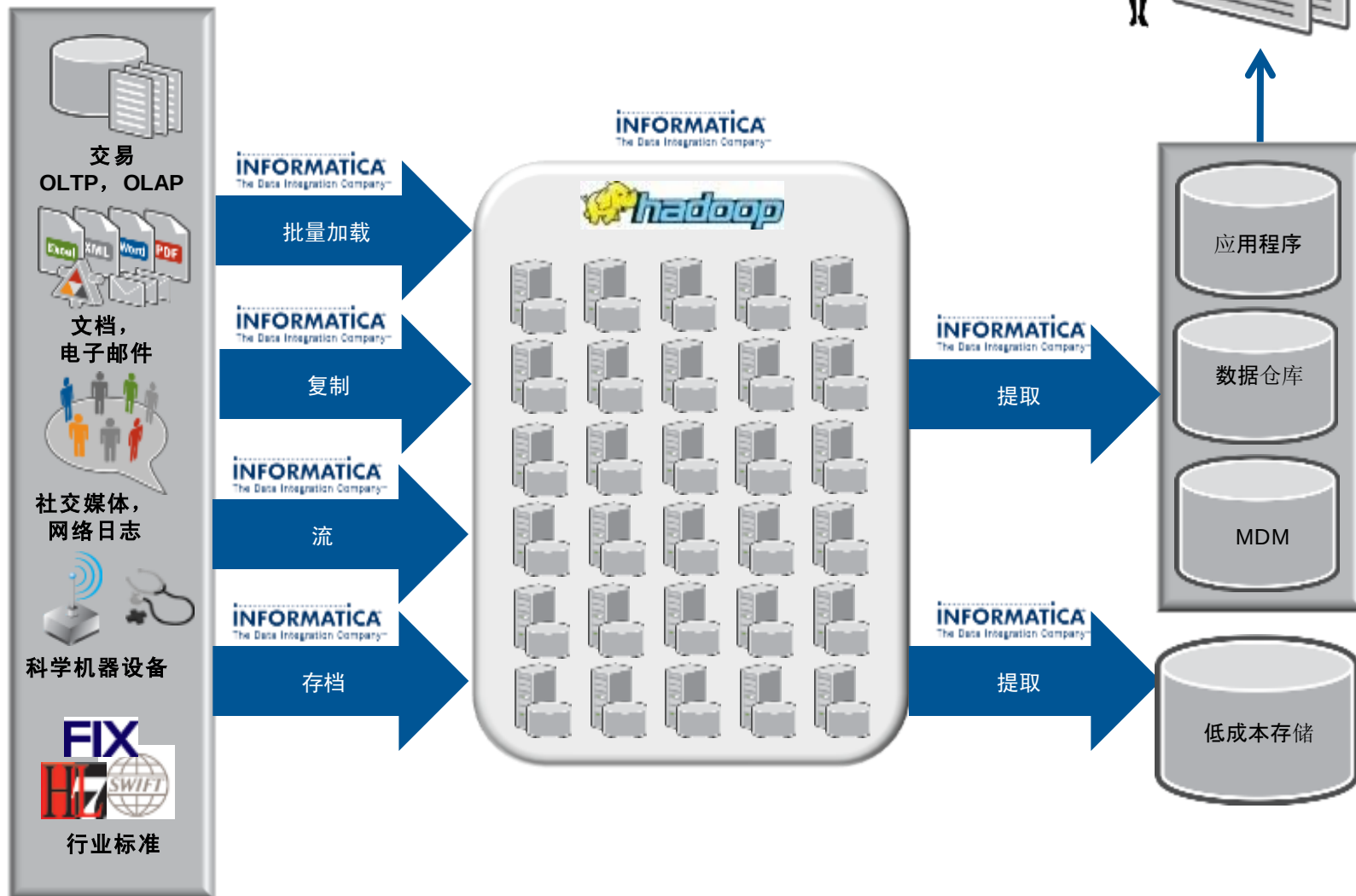
通过Informatica

释放Hadoop 的强大潜能



数据摄入和提取

每小时交易、交互和数据流移动的数据量达到数十TB



释放 Hadoop 能力

通过高性能的通用数据访问



传讯
和 Web 服务



关系文件和
展平文件



大型机
和中型机



非结构化数据
和文件



MPP 设备



WebSphere MQ
JMS
MSMQ
SAP NetWeaver XI

Web 服务
TIBCO
webMethods

Oracle
DB2 UDB
DB2/400
SQL Server
Sybase

Informix
Teradata
Nettezza
ODBC
JDBC

ADABAS
Datacom
DB2
IDMS
IMS

VSAM
C-ISAM
二进制展平文件
磁带格式...

Word, Excel
PDF
StarOffice
WordPerfect
电子邮件 (POP, IMPA)
HTTP

展平文件
ASCII 报告
HTML
RPG
ANSI
LDAP

EMC/Greenplum
Vertica

AsterData

JD Edwards
Lotus Notes
Oracle E-
Business
PeopleSoft

SAP NetWeaver
SAP NetWeaver BI
SAS
Siebel

Salesforce CRM
Force.com
RightNow
NetSuite
SAP By Design
Oracle
OnDemand

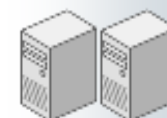
EDI-X12 AST
EDI-Fact FIX
RosettaNet
HL7
HIPAA
Cargo IMP
MVR

XML ebXML
LegalXML HL7 v3.0
IFXCORD (AL3, XML)
cXML

Facebook
Twitter

LinkedIn

打包的应用程序



SaaS/BPO

salesforce.com
experience success



行业标准



XML 标准



社交媒体

PowerExchange for Hadoop

HDFS 和 Hive 适配器

Informatica Developer

File Edit Mapping Layout Navigate Search Run Window Help

Mapping

支持源和目标连接的下推，确保获得最高性能和最大范围

在吸收前执行任何所需要的预处理

本地 HDFS 和 Hive 来源/目标支持

与元数据的集成开发环境和预览支持

Customer_Mas...
Customer_Dev...
Join_Master_D...
Join_Master_D...
Lkp_Plan_Features
HDFS_Customer

Name	Native Type
CUSTOMER_ID1	string
PLAN_ID1	number
AGE	number
MARITAL_STAT...	string
INCOME	number
HOUSEHOLD	number
LIFETIME_SPEND	number
CUSTOMER_ID2	string
CUSTOMER_NA...	string
COMPANY_NA...	string
ADDRESS1	string

Customer ID	Plan ID	Age	Marital Status	Income	Household	Lifetime Spend	Customer ID2	Customer Name	Company Name	Address1	Address2	Address3
15954283	2	27	MARRIED	145138	5	11232	15954283	CLEAVLAND H	CITCO	5900 NORTH A	FORT LAUDER	FL
19132715	1	31	MARRIED	85073	5	10680	19132715	SHIFRAH ROTT	DAVID A NOYE	6 DOUGLAS AV	ELGIN	IL
10111257	4	35	SINGLE	128335	6	3480	10111257	GERMAYNE M	GENEVA TRADI	980 N. MICHIG	CHICAGO	IL
15951670	2	47	MARRIED	39315	1	10800	15951670	KRYSTAL SAX	HUNT OIL COM	1445 ROSS AT	DALLAS	TX

LoadTags

hStream with MapR – 持续摄取

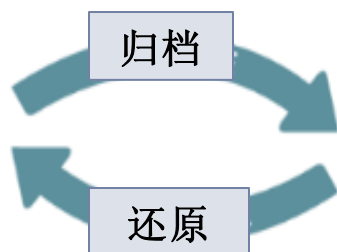


Informatica 数据存档

归档至 Hadoop



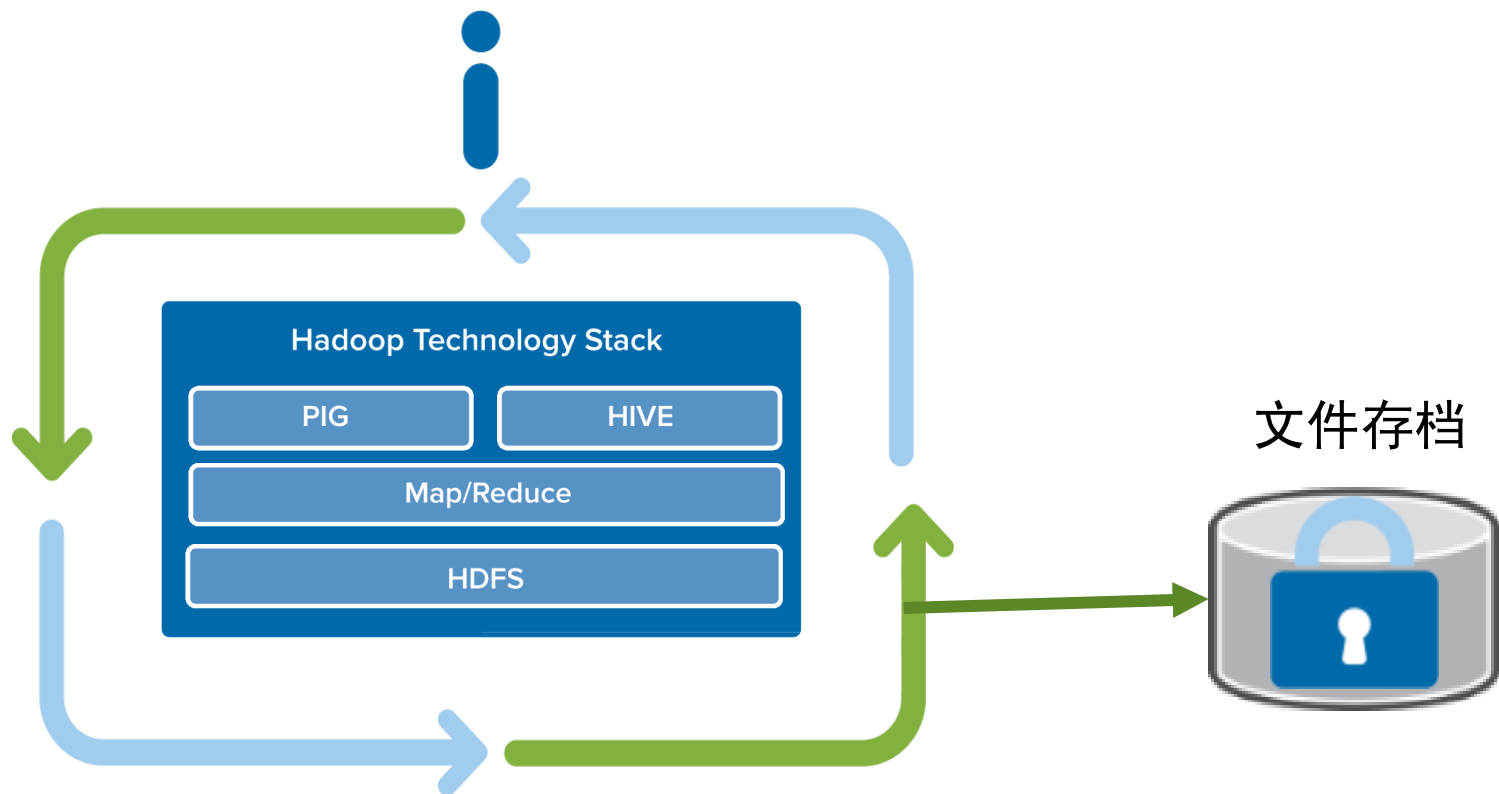
产品数据



- 将数据存档为优化的文件格式以减少存储
 - 压缩率（高达 90%）
 - 不可修改
- 可访问 (SQL, ODBC, JDBC)

Informatica 数据存档

从 Hadoop 归档



Informatica Hparser

处理各种各样的大数据

最广范围的大数据

平面文件和文档

名称 = 值

^/>限定</^



XML



行业标准

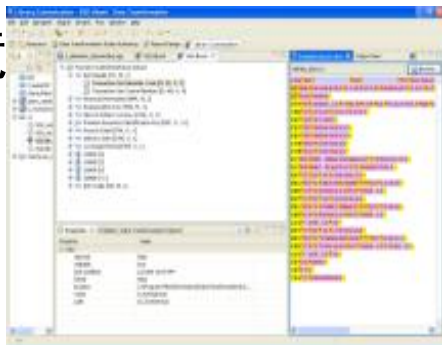


交互数据



生产力

- 直观解析环境
- 预定义转换



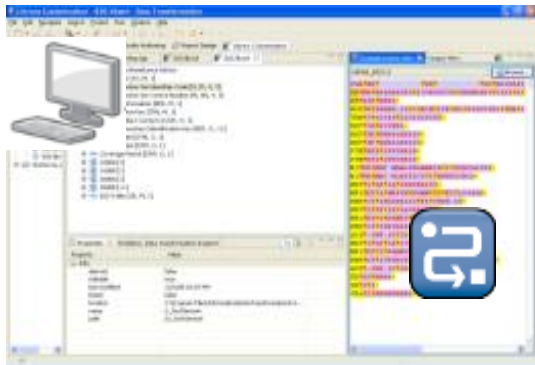
Any DI/BI architecture



EDW
MDM

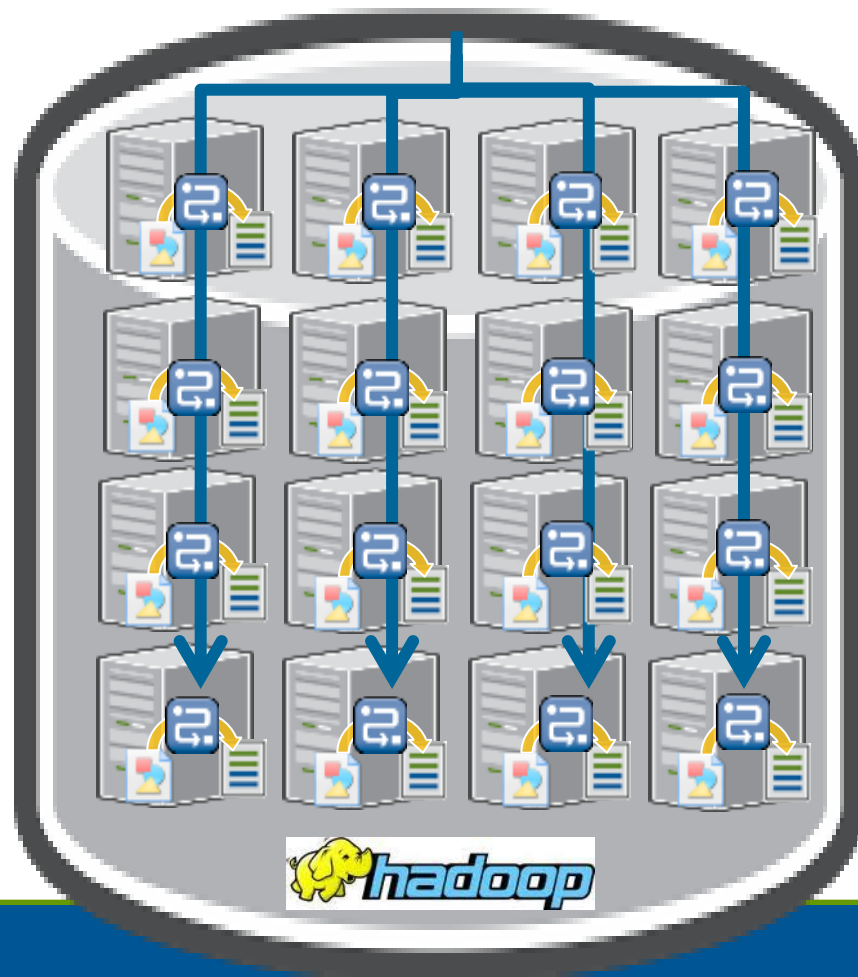
在 Hadoop 上解析和准备数据

工作原理是怎样的？



```
hadoop ... dt-hadoop.jar  
... My_Parser /input/*/input*.txt
```

1. 在 HParser 可视化工作室中定义解析器
2. 在 Hadoop 分布式文件系统 (HDFS) 上部署解析器
3. 运行 HParser 提取数据，并在 Hadoop 产生表格格式



工作效率：数据转换工作台



发现 Hadoop 问题/异常

通过浏览器或 Eclipse 客户检查
和共享结果

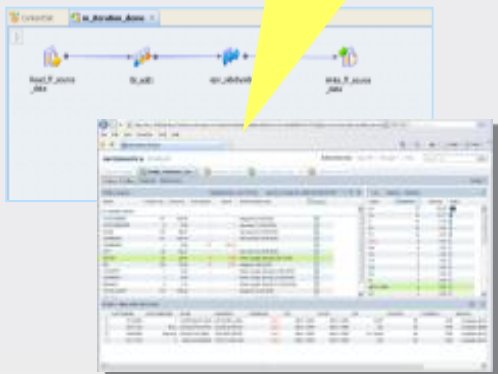
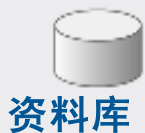
- 单一表格/数据对象
- 交叉表格/数据对象
- 数据域发现

1

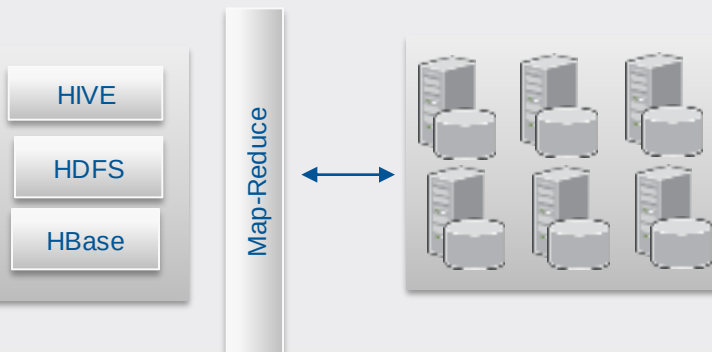
通过到 Hadoop 的本地
连接导入元数据（Hive、
HDFS、Hbase 等）

3

Informatica



Hadoop



2

创建/运行配置文件发现 Hadoop 数据属性
配置文件自动转换成 Hadoop 查询/代码（Hive，
MapReduce 等）

在 Hadoop 上本地执行

Hadoop 数据探查结果



Hadoop 数据探查结果
- 通过浏览器接触企业中的任何人员

值和模式频率与不一致的/脏数据或意外模式隔离

客户标识示例

Statistic	Value
Maximum Length	8
Minimum Length	6
Bottom (5)	10110090 10110091 10110092 10110122 10110124
Top (5)	A5B334 A44563 A23456 19134136 19134134

标识数据中的异常和反常现象的统计数据

INFORMATICA Analyst

2. 值和模式分析 Hadoop 数据

1. 探查统计数据: 最小值/最大值, 空值推导的数据类型等

3. 追溯分析 (Hadoop 数据)

CUSTOMER_ID	CUSTOMER_NAME	COMPANY_NAME	ADDRESS1	ADDRESS2	ADDRESS3	ZIP_OR_POSTAL	ISO_CTRY_CD
10110239	GORDIE SPARROW	MCGREGOR GROUP	1740 BROADWAY	NEW YORK	NY	10019	US
10116657	GORDON SPARROW	MCGREGOR GRP	1740 BRDWDY	NY	NY	10019	US
10178890	GORDY SPAROW	UNKNOWN	BROADWAY	NEW YORK	NY	10019	USA

国家/地区代码示例

Value	Fre...	Per...
NULL	15	3.20
United States	2	0.40
USA	8	1.60
US	464	92.80
U.S.A.	6	1.20
U.S.	3	0.60
*	1	0.20

Value	Fre...	Per...
NULL	33	6.60
Unknown	4	0.80
N/A	2	0.40
999999999	1	0.20
999999	2	0.40
98101	1	0.20
98006	1	0.20
98005	1	0.20
97210-3676	1	0.20
95821	1	0.20
95135	1	0.20
94903	1	0.20

邮政编码示例

追溯实际数据值来检验整个数据集中的结果, 包括可能的重复

IN(GORDY SPAROW G)

CUSTOMER_ID	CUSTOMER_NAME	COMPANY_NAME	ADDRESS1	ADDRESS2	ADDRESS3	ZIP_OR_POSTAL	ISO_CTRY_CD
10110239	GORDIE SPARROW	MCGREGOR GROUP	1740 BROADWAY	NEW YORK	NY	10019	US
10116657	GORDON SPARROW	MCGREGOR GRP	1740 BRDWDY	NY	NY	10019	US
10178890	GORDY SPAROW	UNKNOWN	BROADWAY	NEW YORK	NY	10019	USA

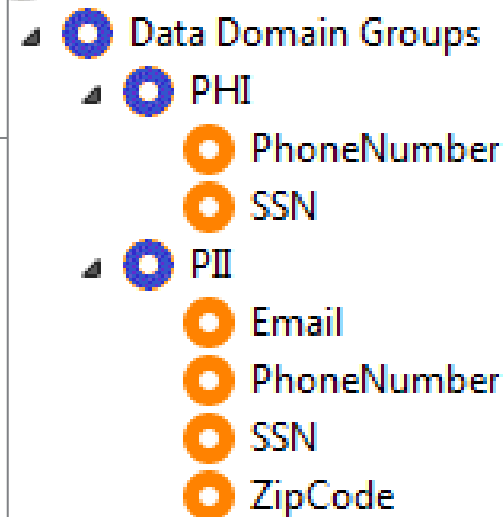
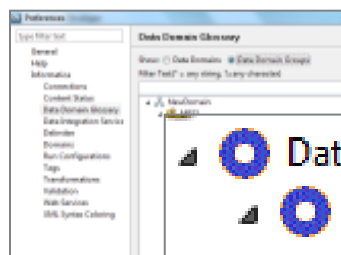
Hadoop 数据域发现

寻找 Hadoop 数据的功能意义

1. 利用 INFA 规则/mapplet 识别 Hadoop 数据的功能意义

- 敏感数据（例如 SSN、信用卡号等）
- 债务和合规性风险？

自动向下发现逻辑，并在 Hadoop 上运行



PHI: 受保护的健康信息

PII: 个人识别信息

可扩展到查找/发现任何域类型

Columns and Domains				
Show: ☐ Data Domain ☒ Columns				
Name	% Data Conforming	Column Name Match	Data Domain Groups	Documented Data Type
zip5				
ZipCode	97.00%	Yes	PII	decimal(5)
phone				
PhoneNumber	76.00%	Yes	PHI, PII	string(14)
email				
Email	100.00%	Yes	PII	string(17)
credit_card_num				
CreditCardNumber	100.00%	Yes	PCI	string(19)
cust_no				
SSN	100.00%	No	PHI, PII	string(11)

2. 查看/共享 Hadoop 中包含的数据域/敏感数据的报告。追溯了解可疑数据值的能力。

保护敏感数据——数据脱密



生产数据隔离



敏感信息屏蔽



- 通过数据变换的方法去除数据中的敏感信息。
- 可以在数据抽取中批量完成，也可以在应用和数据库之间做动态隔离。
- 使企业信息资产得到保护，同时满足合规要求。

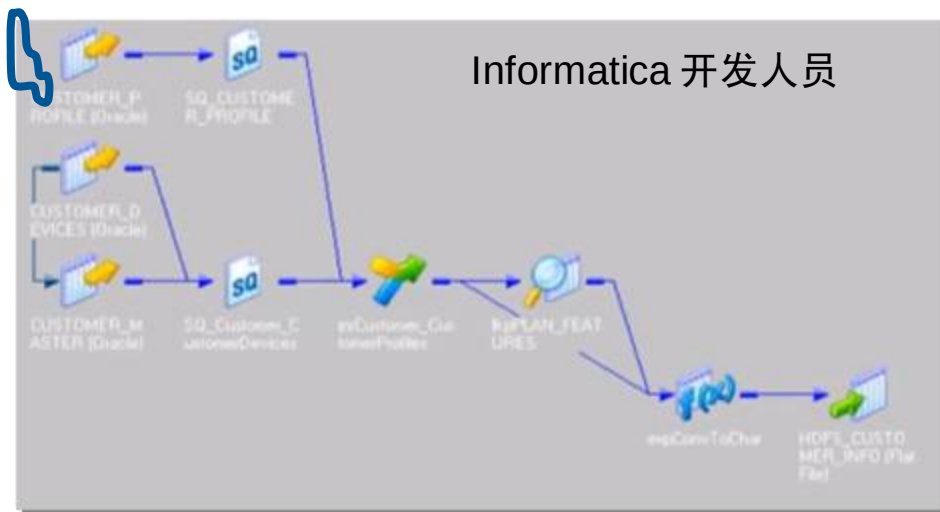
John Smith
654-65-8945
4739-1146-8075-5716
100 Cardinal way
Redwood city

Glen Carter
654-45-2643
4739-1102-3517-8842
342 54th Street
New York



数据集成和数据质量

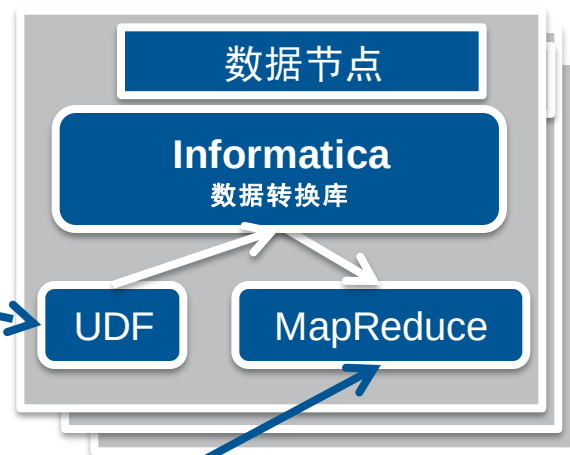
Hadoop MapReduce 处理



1. Informatica 映射转换成优化的 Hive HQL
2. HQL 调用 Informatica DTM 中的自定义 UDF 进行某些特殊的数据转换
3. 优化的 HQL 转换为 MapReduce
4. 在 Hadoop 上执行的 MapReduce 和 UDF

```
SELECT
T1.ORDERKEY1 AS ORDERKEY2, T1.li_count, orders.O_CUSTKEY AS CUSTKEY,
customer.C_NAME, customer.C_NATIONKEY, nation.N_NAME,
nation.N_REGIONKEY
FROM
(
SELECT TRANSFORM (L_Orderkey.id) USING CustomInfra
FROM Informatica
GROUP BY L_ORDERKEY
) T1
JOIN orders ON (customer.C_ORDERKEY = orders.O_ORDERKEY)
JOIN customer ON (orders.O_CUSTKEY = customer.C_CUSTKEY)
JOIN nation ON (customer.C_NATIONKEY = nation.N_NATIONKEY)
WHERE nation.N_NAME = 'UNITED STATES'
) T2
INSERT OVERWRITE TABLE TARGET1 SELECT *
INSERT OVERWRITE TABLE TARGET2 SELECT CUSTKEY, count(ORDERKEY2) GROUP BY
CUSTKEY;
```

Hive HQL



重复使用和导入 Hadoop 的 PC 元数据

Project Import

Select the project content to import.

Source: type filter text

Target: text

Repositories

- REP95
 - Test
 - Sources
 - Flat_File
 - TestSRC
 - WebServices_Provider
 - countryInfoByIana_input
 - Targets
 - countryInfoByIana_output
 - TestTGT
 - Transformations
 - Maplets
 - Mappings
 - m_TestMAP
 - Source Instances
 - Target Instances
 - Transformation Instances
 - m_TestWSP
 - Source Instances
 - Target Instances
 - Transformation Instances
 - User-Defined Functions

Buttons: Add to Target >, << Remove All, < Remove Selected

Buttons: Validate, < Back, Next >, Finish, Cancel

Location: Resolution

Oracle

	Type
MER_ID	string
	decimal
	decimal
	decimal
	string
	decimal
_DURATI...	decimal
UMBER	decimal
KBYTES	decimal
_CALLS	decimal
TIME_CAL...	decimal
PEAK_TI...	decimal

与往常一样设计映射...

The screenshot displays the Informatica Developer interface with a mapping workflow. The workflow starts with a source 'HDFS_CallDetailRecords', followed by a 'Parser' task. The data then flows into three target tasks: 'Filter_by_Region', 'Lkp_Device', and 'Aggregate_Duration_Day...'. Each task has a table of columns and their types. The 'Filter_by_Region' table has columns: Name, Type, YEAR (string), Region (string), DATA_KBYTES (string), SMS_NUMBER (string), and CALL_DURATION. The 'Lkp_Device' table has columns: Name, Native Type, Source Ports, CUSTOMER_ID1, DAY, ISDN, and IMSI1. The 'Aggregate_Duration_Day...' table has columns: Name, Type, CUSTOMER_ID1 (string), DAY (string), ISDN (string), IMSI1 (string), and MONTH. The final target is 'Write_HDFS_Target'. A yellow callout box points to the mapping logic with the text: '图形和元数据驱动环境中的 Hadoop 设计集成和质量逻辑'.

图形和元数据驱动环境中的 Hadoop 设计集成和质量逻辑

Properties - General

Specify execution environment:

- ☒ Native
- ☒ Hadoop

Hadoop connection: HadoopConnection

配置应在何处运行集成逻辑 – Hadoop 或本地

查看生成的 HiveQL

The screenshot shows the Informatica Developer interface. In the background, a mapping is visible with sources like CustomerProfile, Customer_Master, and Customer_Devices, and targets Lkp_Plan_Features and HDFS_Customer. A yellow callout bubble points to the 'View Hive Query' window, containing the text: 从 Hadoop 映射查看完整生成和向下推动的 Hive 或 MR 代码 (View complete generated and pushed-down Hive or MR code from Hadoop mapping).

View Hive Query

Hive Query:

```
FROM  
(  
  SELECT  
    T1.ORDERKEY1 AS ORDERKEY2, T1.li_count, orders.O_CUSTKEY AS CUSTKEY,  
    customer.C_NAME , customer.C_NATIONKEY , nation.N_NAME ,  
    nation.N_REGIONKEY  
  FROM  
    (  
      SELECT L_ORDERKEY AS ORDERKEY1 , count(*) AS li_count  
      FROM lineitem  
      GROUP BY L_ORDERKEY  
    ) T1  
  JOIN orders ON (T1.ORDERKEY1 = orders.O_ORDERKEY)  
  JOIN customer ON (orders.O_CUSTKEY = customer.C_CUSTKEY)  
  JOIN nation ON (customer.C_NATIONKEY = nation.N_NATIONKEY)  
  WHERE nation.N_NAME = 'UNITED STATES'  
) T2  
INSERT OVERWRITE TABLE TARGET1 SELECT *  
INSERT OVERWRITE TABLE TARGET2 SELECT CUSTKEY , count(ORDERKEY2) GROUP BY  
CUSTKEY ;
```

Output

CUSTOMER_ID1	PLAN_ID1	AGE
15954283	2	23
19132715	1	31
10111257	4	35

Configuration: (Default Settings)

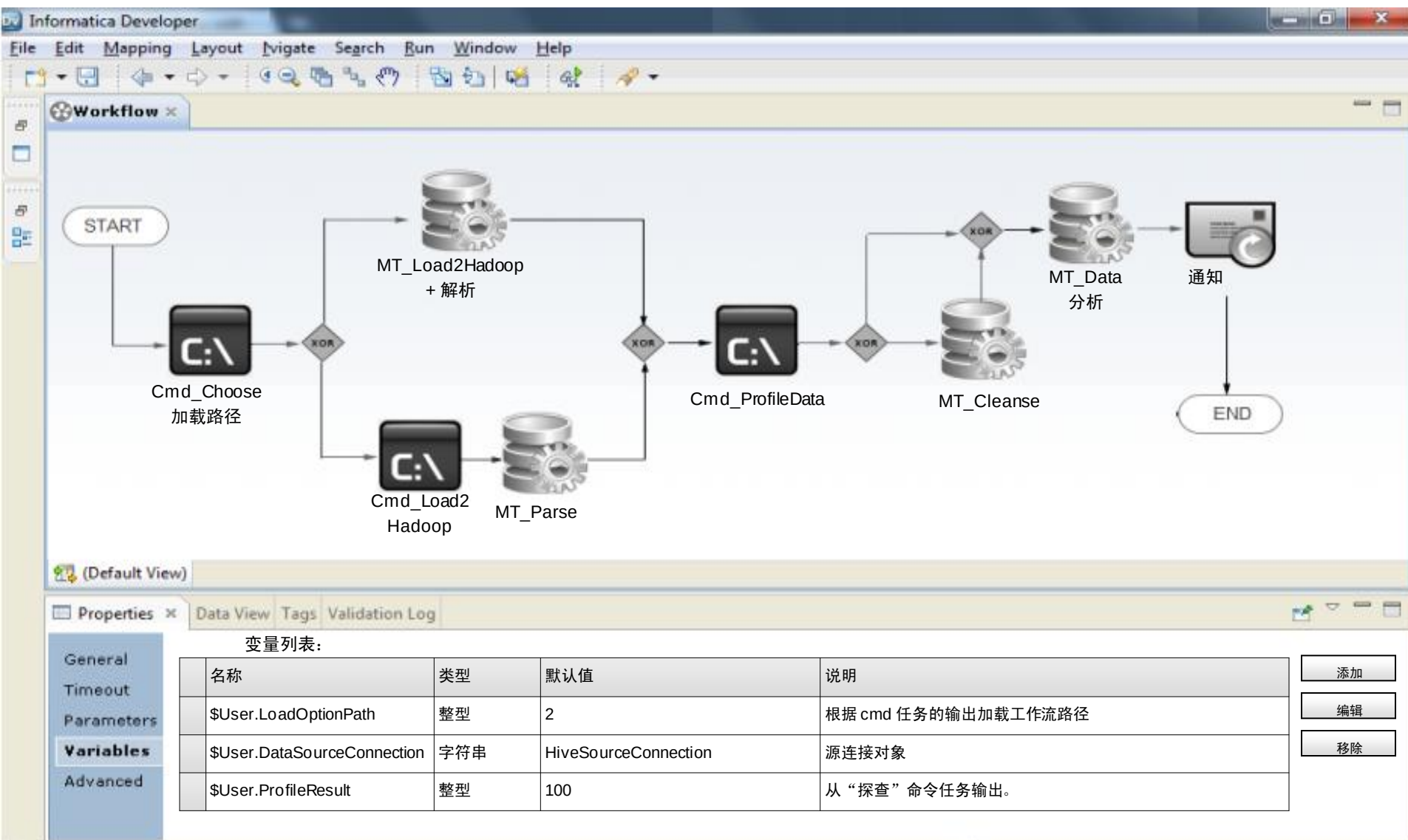
Show: (All Outputs) Choose...

COMPANY_N...	ADDRESS1	ADDRESS2	ADDRESS3
CITCO	5900 NORTH A	FORT LAUDER	FL
T DAVID A NOYE	6 DOUGLAS AV	ELGIN	IL
GENEVA TRADI	980 N. MICHIG	CHICAGO	IL

LoadTags

混合工作流编排

Hadoop 和本地环境的单一工作流运行任务



监控 – Hive 查询计划详情

INFORMATICA Administrator

Administrator Log Out | Manage ▼ Help ▼

Domain Logs Monitoring Reports Security

Monitoring

Navigator Actions ▼

- myDOMAIN
 - DIS
 - Jobs
 - Application
 - Deployed Mapping Jobs
 - SQL Data Services
 - Logical Data Objects
 - Web Services
 - Workflows

Workflows

Instance Id	Name	Type	State	Started by	Start time	Elapsed time	End time	Updated at
00							20:30 31 Oc...	
00					11:22 30 O...	20:30 31 Oc...		
00					14:25 30 O...	20:30 31 Oc...		
00					14:50 30 O...	20:30 31 Oc...		
00					18:45 30 O...	20:30 31 Oc...		
00						20:30 31 Oc...		
00						20:30 31 Oc...		
00					19:10 30 O...	20:30 31 Oc...		
00						20:30 31 Oc...		
00						20:30 31 Oc...		

Hive Query Plan

```
FROM
(SELECT T1.ORDERKEY1 AS ORDERKEY2, T1.li_count, orders.O_CUSTKEY AS CUSTKEY, customer.C_NAME ,
customer.C_NATIONKEY, nation.N_NAME , nation.N_REGIONKEY
FROM
(SELECT L_ORDERKEY AS ORDERKEY1 , count(*) AS li_count
FROM lineitem
GROUP BY L_ORDERKEY
) T1
JOIN orders ON (T1.ORDERKEY1 = orders.O_ORDERKEY)
JOIN customer ON (orders.O_CUSTKEY = customer.C_CUSTKEY)
JOIN nation ON (customer.C_NATIONKEY = nation.N_NATIONKEY)
WHERE nation.N_NAME = 'UNITED STATES'
) T2
INSERT OVERWRITE TABLE TARGET1 SELECT *
INSERT OVERWRITE TABLE TARGET2 SELECT CUSTKEY , count(ORDERKEY2) GROUP BY CUSTKEY ;
```

开发人工具中同
样可用的 hive 查询。

OK Cancel

Number of Hive Queries: 3 (View Hive Query Plan)

监控 – Hive 查询追溯 M/R

INFORMATICA Administrator

Administrator Log Out | Manage | Help

Domain Logs Monitoring Reports Security

Monitoring

Navigator

- myDOMAIN
- DTS
- Jobs
- Application
 - Deployed Mapping Jobs
 - SQL Data Services
 - Logical Data Objects
 - Web Services
 - Workflows

Workflows

Instance Id	Name	Type	State	Started by	Start time	Elapsed time	End time	Updated at
OiksEhrn8346	workflow_customerAnalysis	Workflow	In Progress	Administrator	10:15 30 O...	10:05:00		20:30 31 Oc...
OiksEhrn8346	Cnd_ChoseLoadPath	Command Task	Completed	Administrator	10:15 30 O...	01:07:00	11:22 30 O...	20:30 31 Oc...
18i239k2nn	MT_Load2Hadoop+Parse	Mapping Task	Completed	Administrator	11:22 30 O...	03:03:00	14:25 30 O...	20:30 31 Oc...
0os93k9GHJH	Cmd_ProfileData	Command Task	Completed	Administrator	14:25 30 O...	00:20:00	14:50 30 O...	20:30 31 Oc...
Pu723939793	MT_Cleanse	Mapping Task	Completed	Administrator	14:50 30 O...	03:55:00	18:45 30 O...	20:30 31 Oc...
K8923889ki2	MT_DataAnalysis	Mapping Task	In Progress	Administrator	18:50 30 O...	01:20:00		20:30 31 Oc...
K8923889ki2	MT_DataAnalysis_WI	Mapping	In Progress	Administrator	18:50 30 O...	01:07:00		20:30 31 Oc...
K8923889ki2	M_DataAnalysis-Hive1	Hive Query	Completed	Administrator	18:50 30 O...	00:30:00	19:10 30 O...	20:30 31 Oc...
K8923889ki2	M_DataAnalysis-Hive2	Hive Query	In Progress	Administrator	19:15 30 O...	00:30:00		20:30 31 Oc...
K8923889ki2	M_DataAnalysis-Hive3	Hive Query	In Progress	Administrator	19:25 30 O...	00:07:00		20:30 31 Oc...

New Notifications

M_DataAnalysis-Hive3

State: In Progress

General Properties

Workflow Name: workflow_customerAnalysis

Type: Hive Query (View Query)

Hadoop Cluster Details

Number of Nodes: 2 (View Nodes)

Cluster Heap Size: 613.52 MB

MR Job Details

Job ID	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduce Completed	Job Scheduling Info
job_352	High	Admin	Analysis	33	3	1	0	5	0	NA
job_343	High	Admin	Analysis	33	3	1	0	5	0	NA
job_252	High	Admin	Analysis	50	4	2	50	4	2	NA

INFORMATICA

Beta

查看 Hive 查询详情

单个 M/R 作业的可跟踪性。作业跟踪器链接 URL

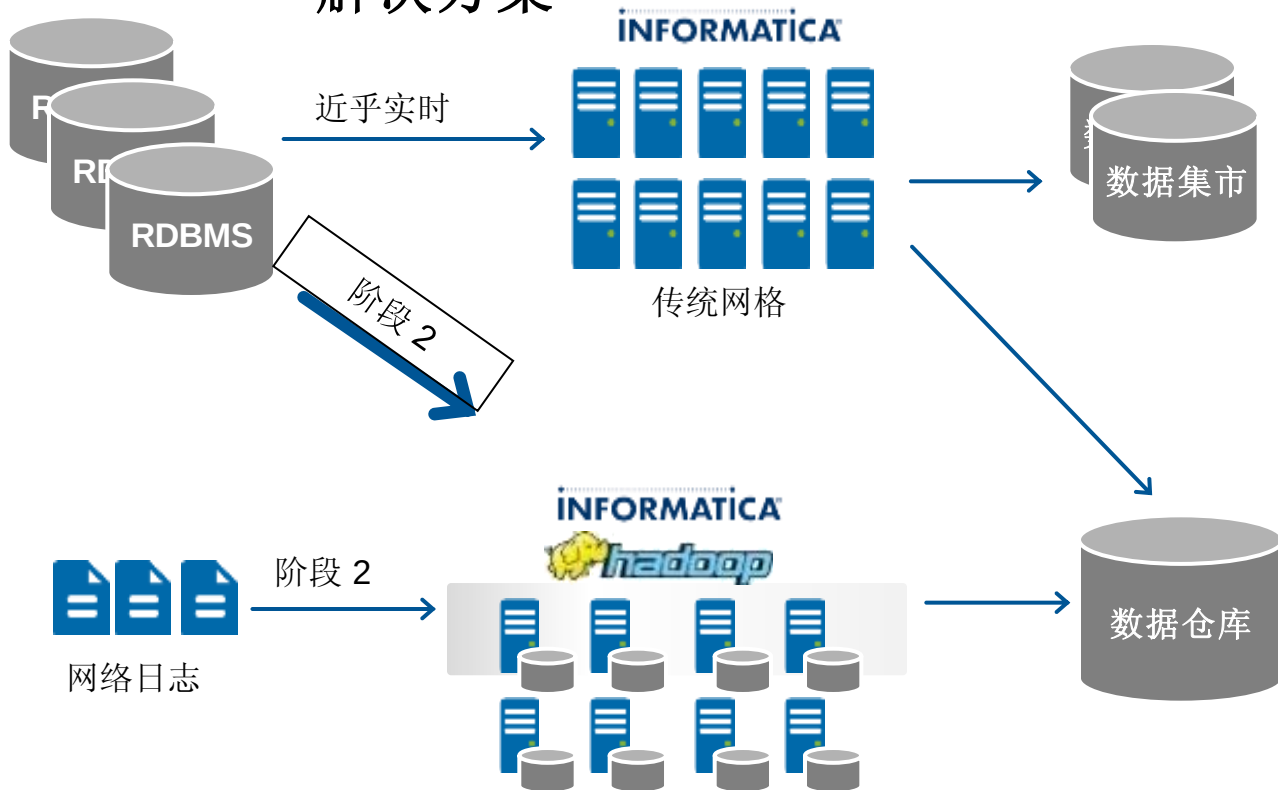
作业跟踪器状态摘要

扩展数据集成并控制成本

为大数据分析奠定基础

挑战：随着数据量和处理负荷的迅速增长，对更快的数据驱动型决策的需求不断增加

解决方案



结果

- 经济高效地拓展性能
 - 降低硬件成本
- 通过在统一数据集成平台上的标准化，增加了灵活性

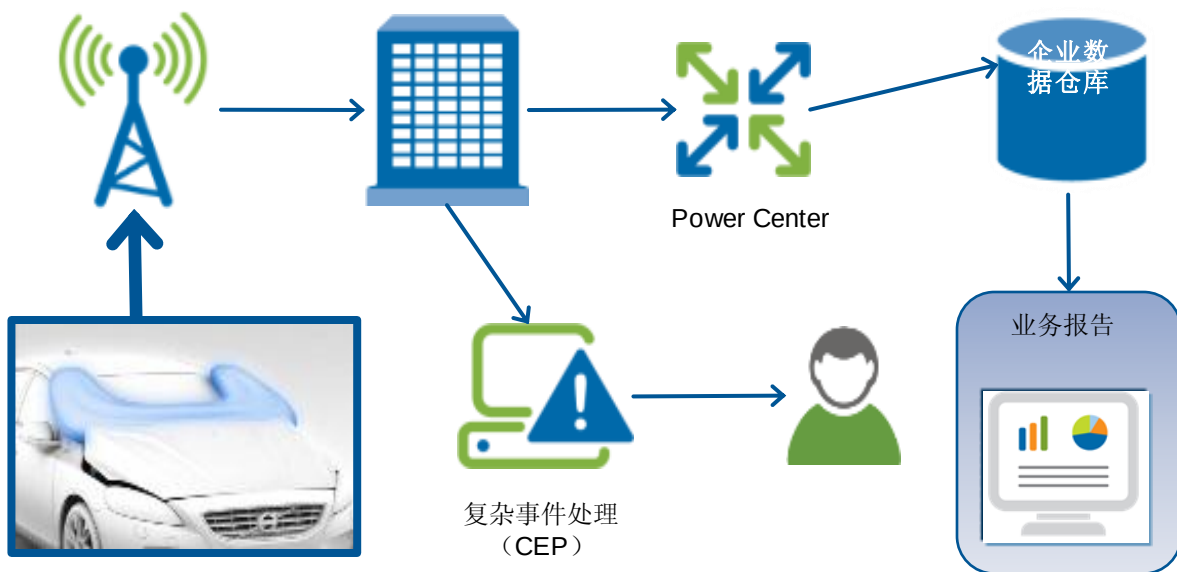
大型国际金融机构

互联车辆项目

开发创新产品和服务

挑战：为“互联车辆”计划，在年底前实现实时收集汽车数据

解决方案



- 持续收集所有车辆的所有信息
- 所有车辆在年底时，都将把数据传送到中央Teradata数据仓库
 - 利用PowerCenter, CDC和CEP 实现实时数据集成

结果

- 助力实现互联车辆的目标：
 - 嵌入移动技术提升客户体验
 - 预测维修维护和提高燃料效率
 - 电话道路救援和自动调度服务

大型国际汽车制造商

更多大数据解决方案

业务需求

提供
分析洞见

提高业务
灵活性

改善
业务流程

提高效率&
降低成本

兼并收购&
资产剥离

获取及挽留
客户

外包
非核心职能

治理风险
合规

提升合作伙伴网
络效率

HealthNow
Administrative Services

DirectEdge

U.S. EXPRESS
FAST FORWARD
U.S. EXPRESS FAST FORWARD
U.S. EXPRESS FAST FORWARD

EMC²
where information lives!

FOSTER'S
WORLDWIDE

SELECTION
CISCO

DB
Dun & Bradstreet



zyme

大数据仓库 &
运营商业智能

Ultra
messaging

大数据服务

大数据存档

大数据整合

实时客户视图

社交/大数据
同步

复杂事件处理

大数据收集及
汇总

通过数据源，将客户、风险和投诉等各种来源的数据(数据仓库、遗留数据、30000数据集和10M投诉)的访问迅速提高了5倍，而成本仅是之前的1/3

节省25%的数据中心空间(100万美元)，将延迟降低83%至340微秒，将每天10亿交易的吞吐量提高580%，且还在持续提高

通过传感器、移动和地理空间数据，改善货运业务并赋予业务Hadoop形式，节省了数百万美元

合理化应用程序组合，并在6个月内得到回报节省了100万美金。将用于服务监测和图案识别的大型数据年龄降低87%

将源自100多个国家的200多个品牌的业务数据从五个系统中转移到一个系统中

每月插槽收入提高4%，并在社交数据和机器数据中，实时从500个来源中将目标客户段从40扩大到160

为全球177多个业务部门及53,000,000个联系人提供云访问。保持360应用程序与LinkedIn和Twitter的更新

通过地理空间和视频追踪，将人工检查转换为以秒数计的自动提醒，保证了航海安全

通过On-Boarding New Data Sources Faster 将进入市场时间缩短90%，并可支持更多数据类型

从Informatica获取竞争优势

- 重复性
 - 可预测、可重复的部署和方法
- 与快速的 **Hadoop** 变化隔离
 - 经常推出新版本和项目
 - 避免对错误的技术下注
- 现有资产的重复使用
 - 应用现有集成逻辑向 Hadoop 加载数据
 - 重新使用现有数据质量规则验证 Hadoop 数据
- 现有技能的重复使用
 - 使 ETL 开发人员能够利用 Hadoop 的功能
- 治理
 - 执行并验证数据安全性、数据质量和法规遵从政策
 - 可管理



谢谢！

