* 1. 成功案例1-汤姆森路透(Thomson Reuters)利用Oracle大数据解决方案实现互联网资讯和社交媒体分析



* **Oracle Customer:**  Thomson Reuters
* **Location:**  USA
* **Industry:**  Media and Entertainment/Newspapers and Periodicals

汤姆森路透(Thomson Reuters)成立于2008年4月17日，是由加拿大汤姆森公司(The Thomson Corporation)与英国路透集团(Reuters Group PLC)合并组成的商务和专业智能信息提供商，总部位于纽约，全球拥有6万多名员工，分布在超过100个国家和地区。

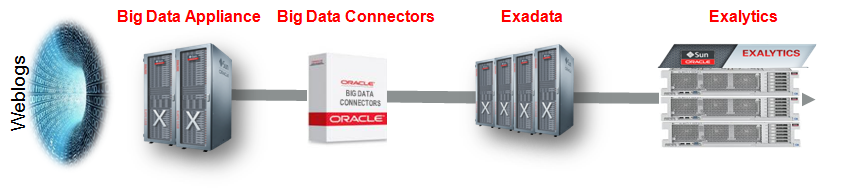
汤姆森路透是世界一流的企业及专业情报信息提供商，其将行业专门知识与创新技术相结合，在全世界最可靠的新闻机构支持下，为专业企业、金融机构和消费者提供专业财经信息服务，以及为金融、法律、税务、会计、科技和媒体市场的领先决策者提供智能信息及解决方案。

在金融市场中，投资者的心理活动和认知偏差会影响其对未来市场的观念和情绪，并由情绪最终影响市场表现。随着互联网和社交媒体的迅速发展，人们可以方便快捷的获知政治、经济和社会资讯，通过社交媒体表达自己的观点和感受，并通过网络传播形成对市场情绪的强大影响。汤姆森路透原有市场心理指数和新闻分析产品仅对路透社新闻和全球专业资讯进行处理分析，已不能涵盖市场情绪的构成因素，时效性也不能满足专业金融机构日趋实时和高频交易的需求。

因此汤姆森路透采用Oracle的大数据解决方案，使用Big Data Appliance大数据机、Exadata数据库云服务器和Exalytics商业智能云服务器搭建了互联网资讯和社交媒体大数据分析平台，实时采集5万个新闻网站和400万社交媒体渠道的资讯，汇总路透社新闻和其他专业新闻，进行自然语义处理，通过基于行为金融学模型多维度的度量标准，全面评估分析市场情绪，形成可操作的分析结论，支持其专业金融机构客户的交易、投资和风险管理。

Oracle大数据解决方案为汤姆森路透构筑了全面涵盖互联网资讯分析和专业资讯分析的市场心理指数和新闻分析产品，帮助汤姆森路透的专业金融机构客户扩展投资视野，针对市场情绪及时修正投资战略，准确把握交易时机，以及更好的管理风险和监控交易流程，从而进一步提升汤姆森路透在金融资讯和分析服务领域的竞争力。同时Oracle大数据解决方案中所应用的一体机产品的预优化配置和一体化集成能力，使得汤姆森路透的大数据分析平台不再需要复杂的配置、排错、调优，大大缩短了系统部署和业务扩展的时间，并能在线扩容和滚动升级，有效满足了其未来的业务发展需求。

* + 1. 挑战
* 实现海量非结构化互联网资讯和社交媒体信息的实时捕获、组织和分析，捕捉市场情绪的变化趋势。
* 关联非结构化的互联网资讯/社交媒体信息和结构化的金融行情/交易数据，形成可度量的分析体系，指导金融机构的交易、投资和风险管理。
* 提供高性能、高可用性和安全性的IT基础架构，并实现更高的能源利用率和更低的空间要求，降低大数据分析平台的部署成本和维护成本
* 提供可快速部署并能在线扩容和滚动升级的大数据分析平台，以适应高速增长和不断变化的业务需求
  + 1. 解决方案



* 汤姆森路透大数据分析平台所应用的Oracle大数据解决方案使用Big Data Appliance大数据机对海量低价值密度的非结构化互联网信息进行采集和实时分析处理，通过Big Data Connectors将分析结果加载到基于Exadata数据库云服务器的企业级数据仓库，与路透社新闻和其他专业新闻分析结果进行关联聚合，利用Exalytics商业智能云服务器进行数据挖掘和可视化分析。
* Oracle Big Data Appliance大数据机为汤姆森路透大数据分析平台提供了一个低成本、可伸缩并享有全面支持的大数据基础架构，Big Data Appliance提供了应对大数据挑战的所有需要，它通过集成优化的硬件和最全面的软件体系，提供一个全面的、易于部署的解决方案，用于获取、组织、分析大数据以及将其加载到Oracle数据库中，
* 通过Oracle Big Data Appliance大数据机为汤姆森路透大数据分析平台的所有数据类型提供极致性能的全面分析，并通过Big Data Connectors，与 Oracle Exadata数据库云服务器紧密集成，无缝实现对汤姆森路透大数据分析平台中所有数据（结构化和非结构化）的分析。
* 利用Oracle Exadata的智能扫描（Smart Scan）和闪存技术（Smart Flash Cache），极大地提升了汤姆森路透大数据分析平台中数据仓库的处理效率，并且能够将数据分析和联机事务处理混合部署，实现了资源整合和动态共享。
* 利用Oracle Exadata的高级混合列压缩技术（EHCC），汤姆森路透大数据分析平台中数据仓库的在线数据使用查询压缩比率以及历史数据使用归档压缩比率提高数倍，大大减少使系统所需的设备和数据备份工作量，也节约了机房空间和能耗费用
* 汤姆森路透大数据分析平台利用Exalytics提供的一系列内存优化分析技术，在主内存中保存数据，避免了网络延迟或磁盘I/O，大幅提升数据分析的响应速度。
* 汤姆森路透大数据分析平台通过Exalytics提供的针对硬件专门优化的Oracle商务智能基础以及同类最佳的报表、仪表盘、即席查询、OLAP 和记分卡等商业智能分析功能，实现高度互动和高密度可视化的数据挖掘分析，让用户以思考的速度进行分析成为现实。
* Oracle一体机产品的内外部节点都通过40Gb/秒的InfiniBand端口进行互联，提供了从互联网资讯及社交媒体信息采集到在线商业智能分析，端到端的数据交互高速通道
* 借助Oracle一体机产品预优化配置的功能以及开箱即用的平衡系统，汤姆森路透大数据分析平台不再需要复杂的配置、排错、调优，系统部署时间比原有数据仓库平台缩短了数倍
* 利用Oracle一体机产品的可在线扩容和滚动升级特性，汤姆森路透大数据分析平台实现了容量和处理性能的弹性扩充，使其获得更大的存储空间以存储更长周期的数据，同时保证了系统性能的可持续增长，从而有效提高了系统的商务智能分析能力
  + 1. 为什么选择Oracle

汤姆森路透首席技术官Rick King表示：我们一直致力于利用一流的技术来提升汤姆森路透的产品，更好的为客户服务。Oracle的大数据解决方案和其一体机产品为我们提供了高效的大数据处理体系，完美契合了汤姆森路透的业务发展诉求。

* 1. 成功案例2 – 旅游和航空服务提供商大数据案例
     1. 客户背景

客户介绍

客户一家是总部位于美国德克萨斯州的高科技企业，它在超过60个国家和地区拥有约10,000名员工。客户旗下的四家公司分别提供高端旅游服务，酒店解决方案，旅游网络服务和航空解决方案，为全球旅游业和航空业提供直销、分销和技术解决方案。业务和技术服务涉及全球各地的旅客，旅行社，航空公司和旅游供应商。其中：

* 旅游网络服务
  + 运营世界最大的旅游网络系统
  + 系统连接35万从业者，400家航空公司，93,000家酒店，25家租车行，50家铁路公司，13家游轮公司
  + 系统每年出售3亿张机票
* 航空公司解决方案
  + 运营行业最大的SaaS业务，为全球航空公司提供航空预订、商业计划及系统运营服务
  + 解决方案被全球300家航空公司，100家机场采用
  + 年服务空运乘客3亿人次
* 酒店解决方案
  + SaaS架构系统提供预订、营销供应及电子商务等业务
  + 为全球12,000家酒店提供服务，年营业额达120亿元美金
* 高端旅游服务
  + 直接提供旅行相关产品及服务的在线营销系统
  + 提供多个品牌网站、呼叫中心，以及供应商及合作伙伴平台

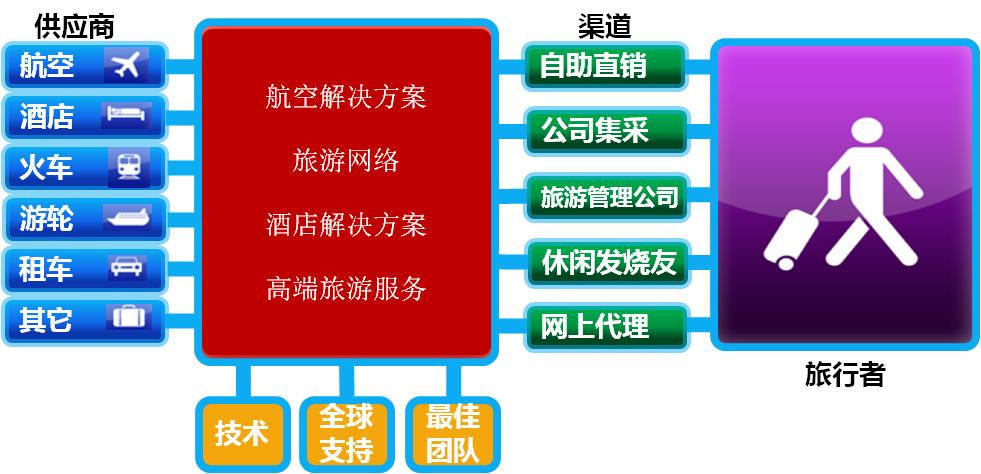
客户的技术能力

客户作为行业领先的技术提供者，在2012年的Information Week TOP 500中排名第26位，它的直销/分销平台和解决方案所能提供的技术指标都是行业内首屈一指的：

* 每秒80,000交易数
* 全球100多个国家的63,000家代理商使用客户的分销系统
* 每日1.5亿次航班状态请求
* 每日10亿次应用访问请求
* 每日2.45亿次网络服务请求
  + 1. 客户遇到的挑战

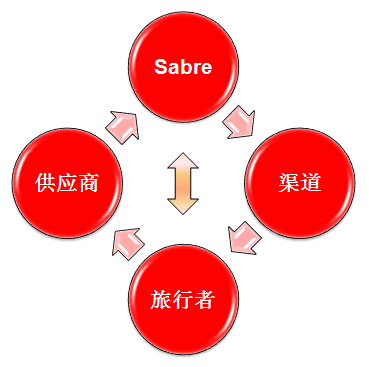
客户作为行业直销/分销交易平台，从航空公司、旅行社、租车公司等获得各种旅游资源，一方面通过自己的直销渠道（Travelocity），另一方面通过代理商分销为公众消费者提供这些资源。

虽然客户已经成为行业的业务和技术领跑者，但是在业务发展过程中发现其目前的业务模型和系统存在一些问题。



图表 1 客户目前的业务模型

上图为客户的业务模型图，对应的支撑系统也依照此模式设计。从中可以看出客户目前的业务模型是开环的：供应商、代理商和消费者是分离的，之间只是通过流程联系在一起，没有形成完整的闭环管理，如：客户可以了解各个供应商提供的旅游资源；直销/分销渠道了解客户的实际需求，而客户无法将它们有机的联系在一起。



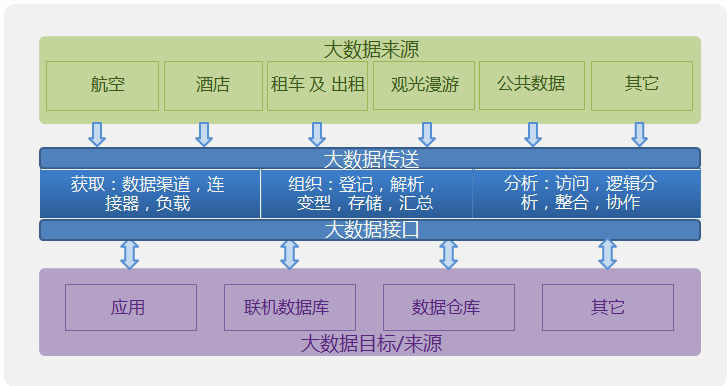
图表 2 闭环管理的业务模型

在当今的市场环境下，仅仅向消费者提供可供的资源信息已不能够全面满足消费者的需求，追踪消费者的消费领域、消费过程、消费习惯、消费周期，才有可能探索、挖掘潜在的客户需求，并为其提供更加贴心的全方位解决方案，形成完整的服务链。

* + 1. 客户采用的Oracle大数据解决方案

只依靠传统关系型架构很难解决上述问题，客户转而从大数据方面寻求突破点。由于航空公司数据、旅客数据和历史订座数据目前都是以关系型数据的形式存储，所以客户采用了将大数据与传统关系型数据相结合的架构。

客户采用的逻辑模型



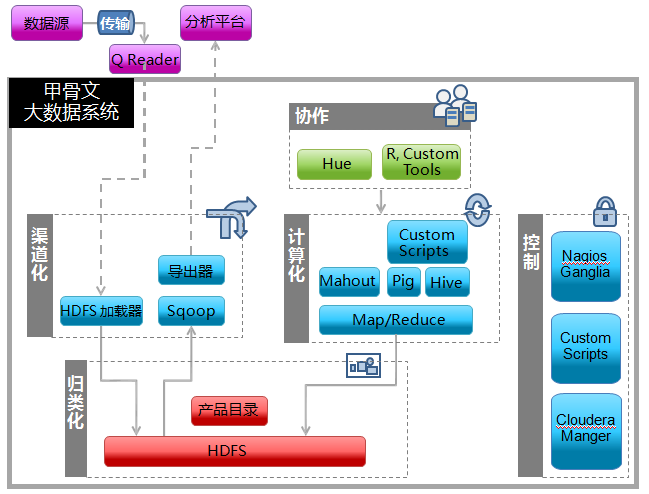
图表 3 客户大数据逻辑模型

客户采用的数据流程

图表 4 客户大数据架构的数据流

客户采用的软件架构

对应上述数据流程，客户采用了完整的Oracle大数据软件架构：依靠MapReduce解析和分解原始数据，并存储在HDFS分布式文件系统中，最后将大数据的处理结果通过Sqoop导出至Oracle 11g分析平台中。

****

图表 5 客户大数据的软件架构

客户采用的硬件架构



图表 6 客户大数据的硬件架构

客户的大数据解决方案的核心是3台Oracle Big Data Appliance（BDA），其中每台BDA的配置如下：

18个计算节点通过40Gb/s的Infininband互联，每个节点包含：

* + 2 个6核Intel® Xeon® X5675 处理器 (3.06 GHz)
  + 48MB内存
  + 12 x 3TB 3.5” 7200 RPM SAS 磁盘
    1. 为什么选择Oracle大数据解决方案

客户在众多的大数据解决方案提供商中选择了Oracle，是因为Oracle的大数据解决方案存在以下优势：

* 完整的大数据解决方案，并经过众多的客户实践
  + 由于目前客户绝大部分数据都以关系型数据库的方式存储，所以大数据需要与传统关系型数据结合使用。而Oracle的解决方案既是独立的大数据平台，也可以与传统Oracle数据库一起构建完整的数据处理平台
  + 软硬一体，全集成系统，工程化调优，插电即用
  + 全Infiniband互联，解决了大数据采用分布式计算时可能存在的网络瓶颈
* 全面大数据产品线，清晰的产品定位，多结构化数据体系结构，降低可能的风险
  + 开源软件商业支持 Hadoop
  + 操作系统商业支持 Linux
  + 硬件商业支持 Sun x86
* 品牌优势
  + Oracle大数据解决方案是与大数据领域的领先厂商Cloudera “强强联手”推出的完整解决方案，充分发挥了各自在领域内的优势，并完美的集成在一起。
    1. 客户采用Oracle大数据方案获得的收益

图表 7 客户实施大数据后的收益

通过使用Oracle的大数据解决方案，客户将供应商、代理商和消费者成功的联系在一起。通过从社交媒体上获得的大数据，与自身的订座系统、直销网站内存储的结构化数据相结合：

一方面更好的使客户在客户自己的直销渠道Travelocity上获得真正需要的旅游产品；

另一方面客户作为航空业的解决方案提供商，将Oracle的大数据解决方案与航空业务模型相结合，作为整体解决方案帮助航空公司和旅游供应商获得了以下业务提升：

* 更精确的了解旅客的需求，做出更精确的市场预测
* 通过分析应用系统日志的方法，预测各个系统可能出现的峰值，以便做好预案
  + 1. Oracle大数据方案在客户的后续利用

在使用Oracle大数据解决方案提升客户感受和销售额的基础上，客户成功的将大数据的应用范围扩大到航班运行控制领域：客户的研发实验室使用Oracle大数据解决方案，通过从Twitter 和Facebook 上获得的飓风信息（提到飓风的评论），并结合传统的航空气象服务，更好的预测航空公司的航班在通过加勒比地区是否会遭遇飓风。

* 1. 成功案例3-大型银行大数据案例
     1. 客户背景

客户是一家总部位于西班牙巴塞罗那的金融服务公司，该公司负责总集团旗下的综合银行和保险业务。截至2013年1月10日，旗下有超过6000家分行，是西班牙分行最多的银行。曾荣获2013年世界最有创造力银行的奖项。

* + 1. 遇到的挑战

客户面临着同行业银行的挑战，同时也面临着互联网金融服务公司的挑战，这时候不仅仅要考虑日常的战术性运维，信息化更要从战略的角度来为公司的决策和运维做出支撑。

* 数据困扰于在线交易系统，系统孤立和复杂性
* 过去的十几年，信息系统架构已经演化为一种方式，数据从交易型和运维系统转换为信息型和分析型数据集市，通过复杂和昂贵的流程，导致：
  + 昂贵的固定资产和运维资产（基础架构，维护，开发等等）
  + 花费大量的时间，无论是批处理（日或者周），还是获得不可预测的KPI（周或者月）
    - 对最新数据的有限访问能力
  + 需要预测哪些要转换和加载
    - 缺少未预料的需求的风险，新的KPI不能按计划加载，往往需要几个月的时间
* 当前IT基础架构有着诸多的限制和依赖性
  + 难以访问非结构化格式，有限的扩展性，负载性难以提供SLA

“At the Internet age, we can not confront the challenges of the future by doing the things as was thought 20 years ago”

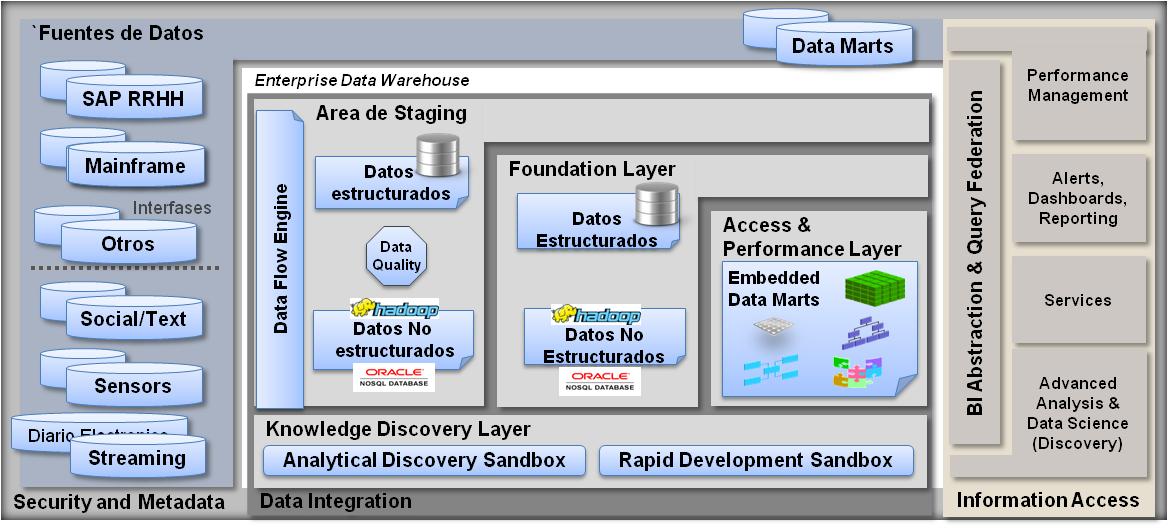
Customer executive statement for Inspiring the “data pool” and User Experience Initiatives

* + 1. 解决方案

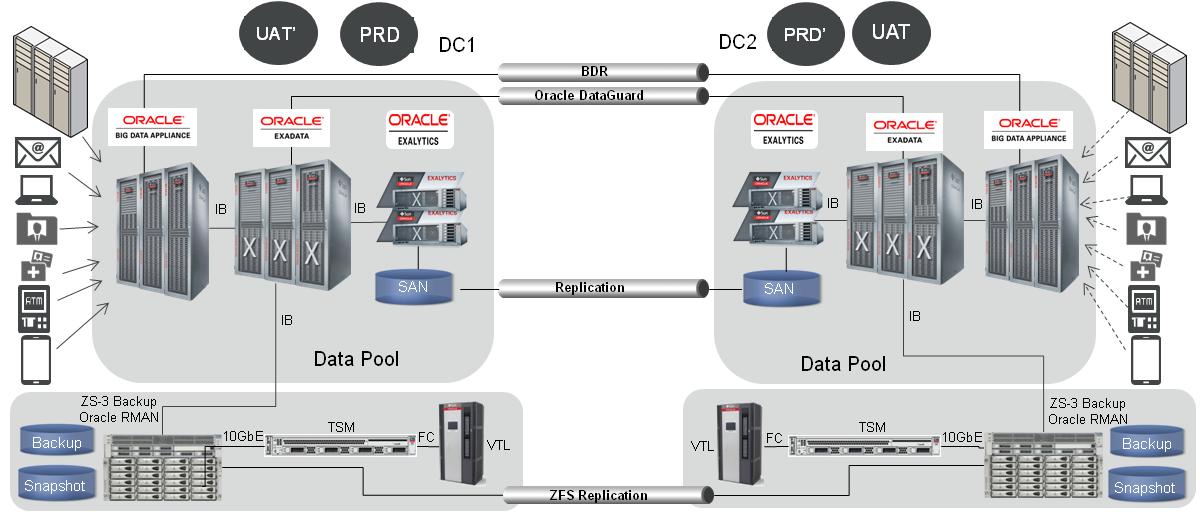
客户启动了大数据时代下提供企业全数据分析能力的“数据资源池”的计划，期望从数据资产入手最大化业务价值。

客户采用了Oracle大数据整体解决方案，构建企业级的统一的数据仓库，并且基于此提供灵活的数据分析和数据服务能力。通过Hadoop和NoSQL来保存，处理非结构化数据，通过传统的关系型数据库来存储和处理结构化数据，在统一的企业全数据基础上，构建数据集市提供给各个业务部门使用。

通过数据集成工具，能够把客户企业内部各种业务系统的数据与企业级数据仓库集成，例如SAP，基于主机开发的系统等，也可以把新兴的数据集成，例如社交网络，传感器等，构建统一的知识发现层，以数据沙箱为各个部门提供分析发现和快速开发的能力。在交互层提供统一的信息访问能力，支持商务智能抽象和查询交互，实现绩效管理，告警、控制面板和报告，数据服务，以及高级分析和数据科学（发现）。



逻辑架构图



系统部署示意图

在具体实现方面，客户全面采用了Oracle提供的一体化设备，大大加快了整体项目的进程，并且通过预先配置和集成的设备，保证了系统的性能和可靠性。整个系统提供了双中心互备，分别承担生产和用户接受测试功能。

基于企业全数据的“数据资源池”实现以下功能:

* 完整的，统一的视图，展现内部和外部与业务有关的有价值的数据，包括整个数据的生命周期/阶段（在线/产品，暂存，企业级，消费者等等）
* 增加使用数据的敏捷性，透明性和安全性，特别是解决新兴业务需求和满足来自于业务线的新需求的灵活性（加快上市时间）
* 数据发现和高级分析的能力，能够发现模式和关联，新使用和转化，提取与格式无关的新数据，灵活增加新的属性以创造附加值。
* 提高服务质量，提供数据保护，可用性，恢复和处理偶发事件，能够处理各种数据并且不影响运维。
  + 1. 带来的价值
* 实现了供应模式的现代化
  + 灵活性和敏捷性

验证了 “数据优先” vs “模式/模型优先”

* + 节省在大机上投入的成本(预计30%)

大机批处理窗口时间减少

ETL to ELT

* + 信息系统演进到接近实时时代

减少了上市时间，提高了转化价值的时间和与业务需求保持一致性（预计70%）

* 实现了当前数据集市的整合

减少了 20-40% OPEX (\*)

满足高可用性，业务连续性和数据保护的服务水平协定

提高了性能，同时减低了整体拥有成本

* 数据联邦

简化和加强了数据访问的控制，不需要数据复制多套就实现关联

提高了敏捷性，实现了数据层面业务理念的统一和一致性愿景

更好地上线时间和对业务需求的响应

* 统一的数据模型

通过统一的，增量的数据模型，避免数据复制，从而减少了成本

更好地上线时间，通过数据资源池能够支持数据集市

* 高级分析

实现了基于海量数据的高级分析，能够支持快速决策和提高数据模式和关系的发现，基于已有的传统的方法论

通过实现了内存处理的能力减少了整体拥有成本

* 支持各种类型数据
  + 通过对“数据资源池“渐增的和逐步进步的部署方式减少成本，包括新的数据源和非结构化数据

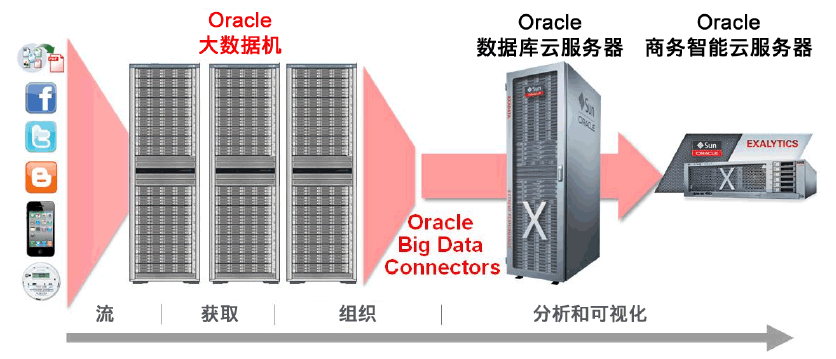
同时减少了运维的复杂性

* + 数据增长

丰富了信息以提高了知识和业务价值

“Business innovation is the key for success in today’s highly competitive banking environment,” said CEO of Customer. “The implementation of this Big Data solution will help *us* remain at the forefront of innovation in the financial sector, delivering the best and most competitive services to our customers”.

* + 1. 为什么选择Oracle



Oracle提供了预先配置集成的一体化设备。无论是在大数据预处理，数据仓库和数据分析阶段，都提供了卓越的性能和可用性。

Oracle提供了大数据整体解决方案，支持各种数据，能够实现数据的整个生命周期的管理，整体最优，大大减少了项目实施风险。

Oracle不仅仅提供了技术和产品，而且提供了大数据和数据架构的方法论，能够基于全球众多的成功案例来为项目成功保驾护航。

* 1. dunnhumby

dunnhumby Accelerates Complex Segmentation Queries from Weeks to Minutes—Gains Competitive Advantage

* http://www.oracle.com/ocom/groups/public/@ocom/documents/digitalasset/2138811.gif
* **Oracle Customer:**  [dunnhumby Ltd.](http://www.dunnhumby.com/us" \t "_blank)  
  **Location:**  London, England  
  **Industry:**  [Professional Services](http://www.oracle.com/us/industries/professional-services/index.html)  
  **Employees:**  2,000

dunnhumby is the world’s leading customer-science company. It analyzes customer data and applies insights from more than 400 million customers across the globe to create better customer experiences and build loyalty. With its unique analytical capabilities, dunnhumby helps retailers better serve customers, create a competitive advantage, and enjoy sustained growth.

* + 1. Challenges
* 扩展服务领域到维护在客户科学行业具有竞争力的优势
* 为客户--包括英国和北美的主要零售企业--提供扩展的历史和实时洞察，针对客户行为，购买趋势和促销活动和产品供给的影响。
* 在公司客户分析服务方面保证具有竞争力的价格，基于增长的客户基础交付增值服务
* 快速和全面地分析快速增长的大量的数据
* 保证敏感数据的安全性，包括保护个人信息，降低风险和符合法规要求
* 保护数据丢失和减少备份和恢复窗口，因为数据对于竞争力优势和业务成功来说非常关键
* 优化IT投资和跨技术密集型业务的性能
* 减少以前分析型和数据仓库软件的许可和维护成本
  + 1. Solutions

**Oracle Product and Services**

* [Oracle Exadata Database Machine](http://www.oracle.com/us/products/database/exadata-database-machine/overview/index.html)
* [Oracle ZFS Storage Appliance](http://www.oracle.com/us/products/servers-storage/storage/nas/overview/index.html)
* [Oracle GoldenGate](http://www.oracle.com/us/products/middleware/data-integration/goldengate/overview/index.html)
* [Oracle Partitioning](http://www.oracle.com/us/products/database/options/partitioning/index.html)
* [Oracle Advanced Analytics](http://www.oracle.com/us/products/database/options/advanced-analytics/overview/index.html)
* [Oracle Advanced Compression](http://www.oracle.com/us/products/database/options/advanced-compression/overview/index.html)
* [Oracle Advanced Security](http://www.oracle.com/us/products/database/options/advanced-security/index.html)
* [Oracle Identity and Access Management Suite](http://www.oracle.com/us/products/middleware/identity-management/oiam/overview/index.html)
* [Oracle Directory Services Plus](http://www.oracle.com/us/products/middleware/identity-management/oracle-directory-services/overview/index.html)
* [Oracle Enterprise Single Sign-On Suite Plus](http://www.oracle.com/us/products/middleware/identity-management/oracle-enterprise-sso/overview/index.html)
* Deployed Oracle Exadata Database Machine and accelerated queries that previously took two-to-three weeks to just minutes, enabling the company to bid on more complex, custom analyses and gain a competitive advantage
* Achieved 4x to 30x more data compression using Hybrid Columnar Compression and Oracle Advanced Compression across sets—reducing storage requirements, increasing analysis and backup performance, and optimizing IT investment
* 安全地整合数据仓库和数据集市到Exadata上，实现了海量数据的极限速度预测
* Consolidated data marts securely with data warehouse schemas in Oracle Exadata, enabling extremely faster presummarizations of large volumes of data
* Accelerated analytic capabilities to near real time using Oracle Advanced Analytics and third-party tools, enabling analysis of unstructured big data from emerging sources, like smart phones
* Accelerated segmentation and customer-loyalty analysis from one week to just four hours—enabling the company to deliver more timely information as well as finer-grained analysis
* Improved analysts’ productivity and focus as they can now run queries and complete analysis without having to wait hours or days for a query to process
* Generated more accurate business insights and marketing recommendations with the ability to analyze 100% of data—including years of historical data—instead of just a small sample
* Improved accuracy of marketing recommendations by analyzing larger sample sizes and predicting the market’s reception to new product ideas and strategies
* 提升安全处理和管理60TB数据，增长率每周500万客户记录，包括用户的客户忠诚度活动信息
* Ensured data security and compliance with requirements for safeguarding protected personal information and reduced risk with Oracle Advanced Security, Oracle Directory Services Plus, and Oracle Enterprise Single Sign-On Suite Plus
* Gained high-performance identity virtualization, storage, and synchronization services that meet the needs of the company’s high-volume environment
* Ensured performance scalability even with concurrent queries with Oracle Exadata, which demonstrated higher throughput than competing solutions under such conditions
* Deployed integrated backup and recovery using Oracle’s Sun ZFS Backup Appliance—to support high performance and continuous availability and act as a staging area for both inbound and outbound extract, transform, and load processes
  + 1. A word from dunnhumby Ltd.

“Oracle Exadata Database Machine is helping us to transform our business and improve our competitive edge. We can now complete queries that took weeks in just minutes—driving new product offerings, more competitive bids, and more accurate analyses based on 100% of data instead of just a sampling.”

– Chris Wones, Director of Data Solutions, dunnhumby USA

* + 1. Why Oracle

dunnhumby considered Teradata, IBM Netezza, and other solutions, and selected Oracle Exadata for its ability to sustain high performance and throughput even during concurrent queries. “We needed to see how the system performed when scaled to multiple concurrent queries, and Oracle Exadata’s throughput was much higher than competitive offerings,” said Chris Wones, director of data solutions, dunnhumby, USA.

* + 1. Implementation Process

dunnhumby began its Oracle Exadata implementation in September 2012 and went live in April 2013. It has installed four Oracle Exadata machine units in the United States and four in the United Kingdom. The company is using three of the four machines in each country as production environments and one machine in each country for development and testing. dunnhumby runs an active-active environment across its Oracle Exadata clusters to ensure high availability.