

## 管理报告

---

2014 年 5 月

# IBM PureData System for Analytics 的 成本/效益案例

*Teradata Data Warehouse Appliance 的成本和时间  
与价值比较*

---

### International Technology Group

609 Pacific Avenue, Suite 102

Santa Cruz, California 95060-4406

电话：831-427-9260

电子邮箱：Contact@ITGforInfo.com

网站：ITGforInfo.com

# 目录

<b>执行摘要</b>	1
挑战 1	1
拥有成本	1
丢失的机会成本	3
结论	5
<b>用户视图</b>	6
引言	6
信息统计数据	6
FTE 人员配置	6
<i>Teradata Data Warehouse Appliance</i>	6
<i>IBM PureData System for Analytics</i>	7
部署时间	8
<b>技术视图</b>	10
Teradata	10
概述	10
最新发展	10
平台产品组合	11
<i>Data Warehouse Appliance</i>	12
IBM PureData System for Analytics	13
概述	13
架构和技术	14
<b>计算基础</b>	16
竞争对手档案	16
成本计算	16
成本分析	17

## 插图目录

1.	IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 的三年拥有成本 - 所有设施的平均值	2
2.	IBM PureData Systems for Analytics 和 Teradata Data Warehouse Appliances 的报告部署时间	3
3.	使用 IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 丢失的机会成本 - 所有设施	4
4.	IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 的三年总体使用成本 - 所有设施的平均值	4
5.	调查总体的行业分布	6
6.	Teradata Data Warehouse Appliance 的 FTE 人员配置示例	7
7.	Teradata Accelerate 产品	9
8.	Teradata 的特定工作负载平台	12
9.	Teradata Data Warehouse Appliance 2750 的用户数据容量	13
10.	主要的 PureData System for Analytics N200X 过滤引擎	14
11.	PureData System for Analytics S-Blade Structure	15
12.	当前的 PureData System for Analytics N200X 生产线	15
13.	竞争对手档案	16
14.	拥有成本分析	17

# 执行摘要

## 挑战

进入大数据和实时处理时代，数据仓库必须符合越来越多的要求。总体趋势是朝着设备使用迈进。比如，针对高性能分析查询工作负载优化集成硬件和软件包。

然而，设备已不再是唯一的解决方案。该领域被划分为多个产品，以适应不同类型和规模的工作负载。本报告的重点是中高端设备段的主要竞争对手，即基于 Netezza 技术的 IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750。

两大平台均实施大规模并行处理 (MPP)。但在其他架构和技术领域，二者具有显著区别，因而会影响相对性能、服务质量和经济效益，而后者也是本报告关注的焦点。

从最基本的级别上而言，可按照每太字节用户数据的美元数额 (\$/TB) 来比较平台成本。然而，该度量值仅涉及购置价格。通常，两个平台在数据结构、压缩等级和性能方面存在巨大差异，这些都将影响有效性。

另一个更能反映实际情况的度量指标便是拥有成本。计算不仅应包括购置成本，还应包含数据库的维护和支持、人员成本及系统管理、部署成本（将系统投入生产可能需要数周到数月），此外还有多年时间内的设施成本。

此外，另一种类型的成本比较也开始发挥作用。由于分析市场发展迅速，因此价值生成时间可极大地影响成本。部署延迟可能导致收入缩水和/或丢失利润机会，同时还可能产生其他底线效应（如部署延迟带来的安全应用、由于盗窃和欺诈造成的损失上升）。总体影响可能不可低估。

两组度量值均被应用于比较 IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 安装在数字媒体、金融服务、零售和电信公司四种代表性设施中时的成本。结果基于使用 IBM PureData System for Analytics 设备的 21 家组织和使用 Teradata Data Warehouse Appliances 的 17 家组织的输入信息，各组织在两个组别中的角色相当。

## 拥有成本

Teradata 2750 系统的三年拥有成本比 PureData System for Analytics 平均高 1.5 倍。针对可比较的应用和工作负载进行比较。结果可综述如下：

- **系统成本。** 尽管初始购置成本无甚差别（如图 1 所示），但 Teradata Data Warehouse Appliances 在三年中的维护成本更高。IBM 为 PureData System for Analytics N200X 提供一年的支持，费用包含在购置价格之内，并且根据百分比来计算，因此年度成本也更低。

根据采用最新一代的 Teradata 2750 和 PureData System for Analytics N200X 模型的情况来计算系统和设施成本。然而，由于这两种模型引入的时间相对较晚（分别为 2013 年 10 月和 2013 年 1 月），因此基于两个平台较早模型中的用户体验计算人员和部署成本。

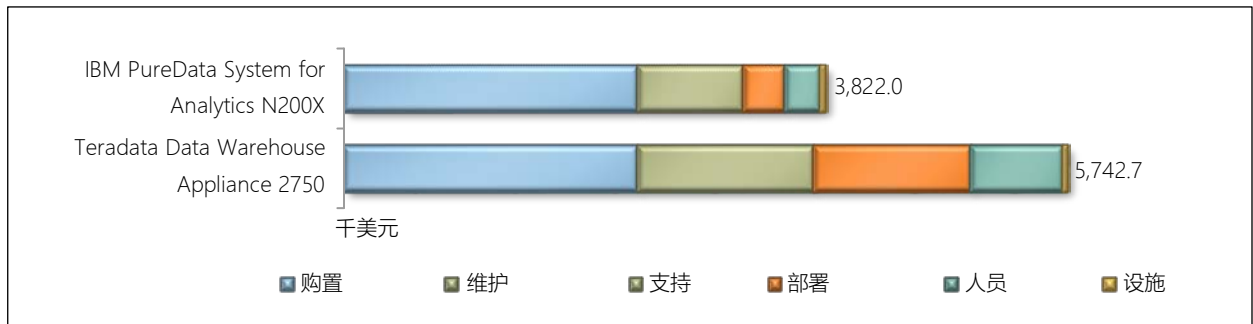


图 1: 使用 IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 的三年拥有成本 - 所有设施的平均值

- **人员成本。** Teradata 2750 系统的数据库管理员全时人工 (FTE) 人员配置明显较高。Teradata 2750 的人员使用成本比 PureData System for Analytics 平均高 2.6 倍。

Teradata 2750 平台定位于简化的低成本产品，以替代公司的旗舰产品 Active Enterprise Data Warehouse (Active EDW)。然而，EDW 架构的主要特征仍被保留。仍需要大量的 DBA 时间和工作来构建和维持指标、调优性能、管理数据模型及相关任务。

当系统的应用、数据量和/或工作负载时常发生变化时，影响尤为明显。出现这种情况时，可能需要在多年时间内持续调优，方可维持一致的性能。

相比之下，PureData System for Analytics 的数据结构更为简单，需要的干预更少。DBA 的 FTE 人员配置比 Teradata 低，且除了非常大的设施外，其他设施均由同样的个人处理数据库和系统管理任务。仅需要很少或根本不需要性能调优。

受访的 Teradata 用户中，FTE DBA 数量为一个到多于五个，平均值为 1.7。PureData System for Analytics 用户中，管理开销为每年 20 个小时到两个 FTE (当公司运营 30 个以上系统时)。平均值低于 0.5 FTE；也就是说，Teradata 系统的平均 FTE 人员配置比 PureData System for Analytics 高 3.4 倍。

人员成本基于美国平均 DBA 工资水平进行计算，并为福利、奖金和相关项目预留了相应的空间。培训成本已涵盖在人员总成本中。

- **部署成本。** Teradata 环境的复杂度更高，这意味着，其部署时间和工作量大幅高于 PureData System for Analytics N200X。Teradata 2750 系统的部署成本 (主要用于提供外部专业服务) 比 PureData System for Analytics N200X 平均高 3.8 倍。

部署成本基于适当专业服务人员已发布的供应商价格进行计算，并为适用的折扣预留了空间。成本包括差旅和招待 (T&E) 费用。

就这点而言，部署时间指的是决定部署特定平台到开始为大部分或整个用户群体提供生产使用所经过的时间（也就是说，概念验证和先导试验并不视为部署流程的终点）。

在用于比较的四种部署设施中，Teradata Data Warehouse Appliance 的部署时间为四周到九个月，而支持相等应用、用户数据量和工作负载的 Teradata Data Warehouse Appliance 设备的部署时间为四天到三个月。

整体受访人群的部署时间如图 2 所示。

部署时间	IBM PureData System for Analytics	Teradata Data Warehouse Appliance
1-2 天	4	–
3-10 天	5	–
10-20 天	7	–
20-50 天	3	2
50-100 天	1	3
100 天到 6 个月	1	5
7-12 月	–	4
12 月以上	–	3
总计	21	17

图 2：IBM PureData Systems for Analytics 和 Teradata Data Warehouse Appliances 的报告部署时间

引人注目的是，当四分之三以上的 PureData for Analytics 客户都已完成部署时间，尚无任何 Teradata 客户完成部署。

由于 Teradata 的架构复杂度更高，因此不仅需要更长时间建造数据结构，同时也延长了测试、调优、系统集成和相关任务的时间。

参与部署的 FTE 内部人员数量不易量化。然而有一点毋庸置疑，那就是 Teradata Data Warehouse Appliance 的内部人员和成本更高。

- **设施成本。** 这些成本主要用于能源消耗，PureData System for Analytics N200X 比 Teradata 2750 略高。

有关安装、配置和方法的更多信息以及精细的成本分析请参见本报告的“计算基础 (Basis of Calculations)”部分。

## 丢失的机会成本

经验表明，数据仓储应用通常可在数周到数月内产生巨大的底线收益。因此，延迟将这些应用投产可能意味着因丢失收入和/或利润机会导致巨额成本。此外，如果在此期间这些机会为他人所用，那么竞争地位也将受到影响。

随着时间的流逝，这种延迟的影响越来越大。众所周知，近年来分析循环时间在逐渐缩短。各行各业中，预测和规划循环的时间已从数月缩短至数周，甚至数小时。越来越多的组织正移至实的模型。

对于在不稳定的因特网和社交媒体中运营的数字媒体公司而言，这种转变尤为明显。但也有少数企业未受到电子商务和蒸蒸日上的移动商务增长的影响。对机会即时响应已成为一种必不可少的竞争力。

这些影响在进行拥有成本比较的四种设施中同样显而易见。在这些案例中，部署 Teradata 2750 系统丢失的机会成本明显高于 PureData System for Analytics N200X。图 3 阐明了差距。

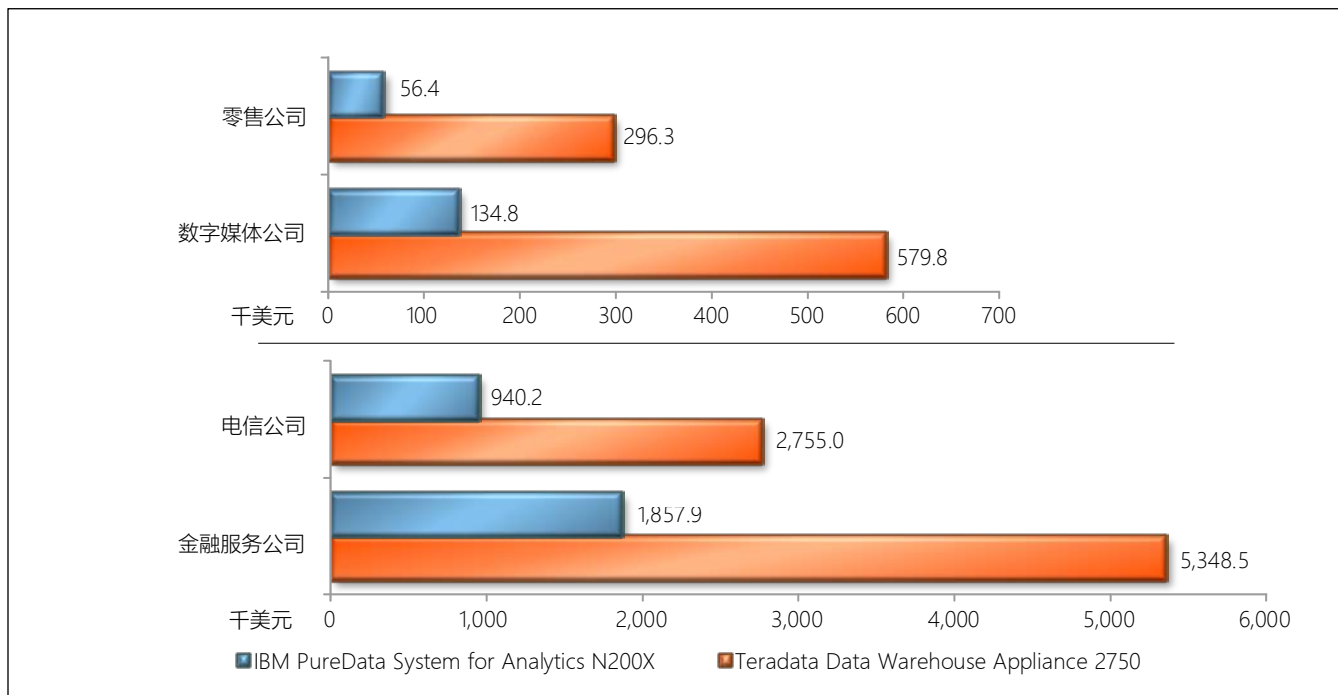


图 3 : 使用 IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 丢失的机会成本 - 所有设施

使用 Teradata 2750 丢失的机会成本比 PureData System for Analytics N200X 高 2.9 到 5.3 倍，平均高 3 倍。如果将丢失的机会成本加入到拥有成本，那么这些平台的总体成本可能还要增加，如图 4 所示。Teradata 2750 系统的总体使用成本比 PureData System for Analytics N200X 平均高 1.7 倍。

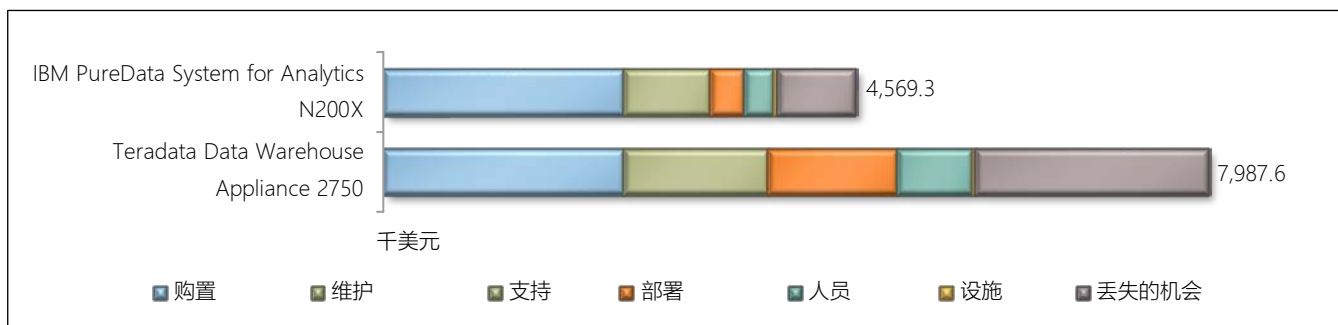


图 4 : 使用 IBM PureData System for Analytics N200X 和 Teradata Data Warehouse Appliance 2750 的三年总体使用成本 - 所有设施的平均值

## 结论

20 多年来，基于 SQL 的传统数据仓库为用户提供了良好的服务。然而，它们已变得日益复杂，并且随着供应商试图采用新的硬件和软件技术，其复杂程度仍将加深。与此同时，业务用户迫切需要更快的应用交付，因此产生了越来越强烈的挫败感。

简而言之，*时间就是金钱*。部署时间影响成本，也影响着长期的业务绩效。从这个角度看，上文所提及的丢失机会成本比较可能低估了 PureData System for Analytics N200X 的优势。

图 3 和图 4 中所示的丢失机会成本仅针对于初步应用。实际上，组织仍将继续部署新应用。快速部署在多年时间内的累积效应可能大得多。丢失机会成本的差距将大幅增加。

此外，PureData for Analytics 系统还具有另一个更大的优势。最终用户可借其轻松开发和部署自身应用，因此，不仅减少了延迟，还可比传统的数据仓库方法提供更高的业务一致性。

因此，PureData System for Analytics N200X 的成本/效益案例并非简单在于比 Teradata 2750 更具成本效益，或者能够更快地完成部署。而在于：该平台的独特功能可以比任何竞争对手都更紧密地反映至高性能数据仓库的长期需求。

# 用户视图

## 引言

该部分提供报告中受访用户组织的更多相关细节及其在 FTE 人员配置和部署时间方面的输入信息。下一部分（技术视图）提供本报告涵盖平台的技术和供应商定位的更多相关信息。

最后一部分（计算基础）概括了设施和详细配置及 FTE DBA 人员级别。此外，还阐述了成本计算所采用的方法和值。

## 信息统计数据

PureData System for Analytics 和 Teradata 的用户涵盖范围从新近成立、员工不到 200 人的数字媒体创业公司到财富 100 强和全球 100 强企业。

行业分布如图 5 所示。Teradata Data Warehouse Appliances 在数字媒体公司中的采用率较低，这似乎反映了总体的市场信息统计数据。

	电信	数字媒体	金融服务	零售	其他	总计
IBM PureData System for Analytics	6	5	3	3	4	21
Teradata Data Warehouse Appliances	4	2	4	3	4	17

图 5：调查总体的行业分布

七个案例（41%）中，Teradata 用户已采用 Active EDW 系统。多数情况下，Data Warehouse Appliances 用于从这些系统卸载专业应用，或部署于互补角色中。一些组织已将数据仓库从 Active EDW 系统迁移至 Data Warehouse Appliances。

这些组织可利用 Teradata 系统架构的现有技能和经验。因此，Data Warehouse Appliances 的平均 FTE 人员配置级别和部署时间可能会被低估。

## FTE 人员配置

### *Teradata Data Warehouse Appliance*

17 位 Teradata 用户中，有三位用户表示使用不到一名 FTE DBA 来管理系统。其余受访者报告的 FTE DBA 为一到五名。总平均值为 1.7 名 FTE。人员通常具有多年使用 Teradata 数据库的经验。

此外，通常还雇佣了系统管理员/工程师、技术支持和其他人员。图 6 提供了示例。

由于非 DBA 管理员的相关数据通常不完整，因此未将其包括在人员成本的计算中。



金融服务	制造业	分布	数字媒体
2 名 DBA 2+ 系统管理与支持	1.5 名 DBA 1-2 名 FTE 系统管理员	1 名 DBA 1-1.5 名系统管理员	1 名 DBA 1 名系统管理员

图 6：Teradata Data Warehouse Appliance 的 FTE 人员配置示例

根据组织是否已雇佣具有 Teradata 系统经验的人员，Teradata DBA 所需的培训量各异。当组织已雇佣经验丰富的人员时，通常仅需针对 Teradata Database 13.10 或 14 开展技能更新方面的培训，同时让人员熟悉系统软件和硬件更改。

当组织不具备富有经验的 DBA 时，通常需要开展为期数周的培训，即便如此，组织也不愿雇佣无经验的人员来负责关键业务系统的支持事宜。内部转移无法满足人员配置需求时，外部雇佣十分普遍。

尚不明确 Teradata Database 15 的 DBA 人员和培训需求是否有所区别。它在 2014 年 4 月新近上市，因此在准备本报告时仍缺乏充足的用户体验。

## IBM PureData System for Analytics

在 21 位 PureData System for Analytics 用户中，有 18 位表示雇佣了不到一名 FTE 管理员。只有一家组织拒绝公布采用的系统数量，但将设施描述为超过一百万兆字节（雇佣了一名 FTE）；其他组织都表示采用了 20 多个或 30 多个系统（雇佣了两名 FTE）。

在表示 FTE 不到一名的组织中，12 个组织（67%）估计实际数量少于 0.5。管理开销据说代表了一个人的时间比例，可能为每周一次.....每周两个小时.....每周数个小时.....每月数小时.....每天不到半小时（管理五个系统）.....也可能为每三个月 6 个小时.....每年 20 个小时。

PureData System for Analytics 的管理员通常处理 DBA 任务，负责系统管理、系统工程、开发和其他功能。三家组织表示，无需为 PureData System for Analytics 配备传统意义上的 DBA。

多数情况下，PureData System for Analytics 管理员都具有其他商业智能系统和/或数据库的过往经验，或曾任职系统管理员和工程师。仅有四家组织表示雇佣了外部专家。

对于人员配置水平较低的原因，这里提及了一些原因。最常见的原因是（最终）用户直接与系统连接。两家已从 Teradata 迁移至 PureData System for Analytics 的组织 and 一家同时采用这两大平台的组织给出了更为详细的解释。

一家组织表示，我们无需构建索引.....由用户直接写入系统，不必求助于 DBA.....我们处理的是完整的数据集，而非首先将一切汇总和总结.....我们无需使用数据模型。与 Teradata 系统不同，据说几乎不存在性能调优开销。

PureData System for Analytics 管理员所需的培训量从零（五个案例）到两天不等。一个组织表示，未经任何培训，其系统便正常运行了六个月，而其后开展的培训是由于（系统）升级。此外，PureData System for Analytics 管理员的学习曲线据说也比 Teradata 平缓。

还有迹象表明，PureData System for Analytics 开发人员的 FTE 人员配置通常也低于 Teradata 系统。一位受访者表示，大多数工作由（最终）用户完成。除非出现异常情况，否则他们不必求助于开发人员。

## 部署时间

任何设备的启动时间都取决于多种因素。保持业务一致性、识别数据源和构建提取、转换和加载 (ETL) 机制等任务所需的时间和工作负载通常与平台无关。此外，启动时间还受到应用及必须加载和处理的数据量的影响。

然而，调查反馈却显示，PureData System for Analytics 设备的投产时间比 Teradata 短。Teradata 需要更多的时间来完成架构设计、数据模型和索引建造、配置、测试及其他任务。即使小范围部署，通常也需要进行大量的性能调优。

部署时间的差异极大。已报告最快的 PureData System for Analytics 部署可在四天内让 500 多位最终用户使用报告应用和数据，不到三周内为 3,000 多位用户提供完全的生产操作支持。

最快的 Teradata Data Warehouse Appliance 部署可在四周内初步投入使用，且可在两个月内实现面向 600 位报告用户的完全生产运营。

另一方面，PureData System for Analytics 和 Teradata 系统的最长报告部署周期分别为 3 个月和 12 个月。这些周期涉及近实时处理数百太字节数据的项目。

对于较小的 Teradata 项目，其报告部署时间与该公司针对 Accelerate 计划所报告的数据一致，该计划提供了单一价格的援助包，由公司的专业服务组织提供特定类型的部署。

图 7 总结了主要的 Teradata 产品。前两种 (Accelerate for Do It Yourself 和 Accelerate for Load and Go) 不需要进行生产部署。其他产品的部署时间均为 60 到 120 天。实际部署时间通常更长，还可能发生其他的专业服务费用。

此外，该公司还与软件公司 Kalido 合作，Kalido 将其现有的数据模型加速转换至 Teradata Data Warehouse Appliance。Teradata 宣称，其客户可在 90 天或更短的时间内构建或扩展数据仓库。然而，标准部署时间似乎再次超过了可比较的 PureData Systems for Analytics 设施。

产品	描述	时间 (天)	起价
Teradata Accelerate for Do it Yourself	初始数据加载和样本查询。为富有经验的数据仓库用户设计	10	\$350,000
Teradata Accelerate for Load & Go	数据加载	30	\$440,000
Teradata Accelerate for Application Re-Host	数据集市与运营报告系统集成到数据仓库	120	\$550,000
Teradata Accelerate for Data Warehousing	针对首次用户的完全数据仓库部署	120	\$830,000
Teradata Accelerate for Finance	预置财务报告和分析解决方案, 包含财务导向的数据模型。	70-80	\$700,000
Teradata Accelerate for Marketing	针对客户细分、营销活动管理、一对一目标定位和其他市场营销应用的预置解决方案	70	\$670,000
Teradata Accelerate for Demand Signal Repository	为消费品制造商实施需求信号库的预置解决方案	60	\$400,000
Teradata Accelerate for Gaming	为博彩公司实施分析和报告应用的预置解决方案	90	\$600,000
Teradata Accelerate for Retail	用于销售与库存分析、种类规划、市场篮子与存储性能分析及相关应用的预置解决方案	90	\$640,000

来源：Teradata

图 7：Teradata Accelerate 产品

# 技术视图

## Teradata

### 概述

Teradata 是数据仓库设备市场上规模最大、历史最长的参与者。自 1983 年首次发行系统以来，该公司已逐步增强其核心 MPP 系统和数据库架构。

公司早年在基于 MPP 的数据仓储方面取得了成功，这为它赢得了极为广泛的固定客户群。通常，用户看重的是定制化系统和 Teradata 特有的应用。此外，Teradata DBA 技能很难向其他平台转移。

鉴于以上因素，Teradata 比小型和/或新的竞争对手收费更高。尽管该公司可能会在招标中报出较高的系统价格，但其软件预订和维护价格仍然不菲。

Teradata 平台围绕 Teradata Database 构建，其最新版本 13.10、14 和 15 分别于 2010 年 8 月、2011 年 10 月和 2014 年 4 月面世。

Teradata 13.10 包含了新的时间和空间数据管理功能。Teradata 14 实施了列数据结构，被该公司描述为行列混合数据库。新的 Teradata Database 15 可支持 Java Script Object Notation (JSON) 和 NoSQL 数据库及其他开放式源代码环境通常采用的标准。

硬件方面，Teradata 平台采用基于 OEM Intel 处理器的系统部件，配备了 SUSE Linux Enterprise Server (SLES) 和第三方存储、I/O 及其他硬件组件。所有系统均采用专有的 Teradata BYNET 互联技术。

### 最新发展

尽管 Teradata 在高端市场上的地位仍不可撼动，但自从 2000 年代中期以来，该公司面临来自创业公司和知名供应商日益加剧的竞争。正是由于这些竞争压力，Teradata 决心推出更小、价格更低廉的 2000 系列 Data Warehouse Appliances (其首款产品于 2008 年面世)，并增加其他专业的设备产品。

此外，该公司还力图整合列数据结构、Hadoop 和 NoSQL、固态硬盘 (SSD) 及最近的内存计算等新技术。如前所述，Teradata Database 的连续版本为上述技术增加了支持。

此外，Teradata 于 2011 年收购了基于 Hadoop MPP 分析软件的专业开发公司 Aster Data Systems。同年，它引入了使用 Aster 技术的一系列分析设备。

在最近几代系统中，Teradata 还非常重视公司所谓的*系统范围的温度管理*，即针对热（最常访问）、冷（很少访问）和中间型数据，使用不同类型的存储设备。这种方法与存储阵列供应商采用的分层机制类似。

关键功能包括启用 SSD 和传统磁盘的 Virtual Storage；允许特别热门的数据在 RAM 中处理的 Intelligent Memory。这些功能分别被引入到了 Database 13 和 14 中。

采用多种基于软件的压缩方法（算法、块级别和多值）的数据压缩被引入到了更早的 Teradata 版本中。此外，Database 14 和 15 还支持高性能的列压缩。

## 平台产品组合

Teradata 提供广泛的产品，该公司将其称为 *特定工作负载平台*。这些产品目前包括：

- *Active Enterprise Data Warehouse 6000* 系列是 Teradata 的高端旗舰平台。2014 年 4 月面世的最新一代产品使用 Intel Ivy Bridge 处理器。标准配置可扩展至 2,048 节点，较大系统可在自定义的基础上实施。

Teradata 宣称，6750 旨在开发 Teradata 14 和 15 中的新功能，并使用 SAS HDD 和 SSD。

- *Data Warehouse Appliance* 实施与 Active EDW 同样的核心架构，但不支持 SSD 的使用。它被 Teradata 定位为低成本产品，从而为计算需求更为简单的组织提供替代 Active EDW 系统的产品；作为开发平台；支持特定主题的数据集市。

此外，Data Warehouse Appliances 通常还作为支持 EDW 的故障转移系统。最新一代的 2750 于 2013 年 10 月面世，我们将在下文中对其进行更为详细的讨论。

Teradata 宣称，2013 年的 Data Warehouse Appliance 销售额占公司产品收入的 14%，超过 1.7 亿美元。这还不包含咨询和维护服务收入。

- *Integrated Big Data Platform 1700* 专为涉及海量（数百太字节到百万兆字节）数据的分析应用而设计。该平台于 2013 年 10 月面世。它是原 Teradata Extreme Data Appliance 的升级版。

1700 支持 Teradata MPP 架构和 Database 14.10 和 15，原则上可扩展至 2,048 节点。它使用速度较慢、容量更大的 SAS 2 TB 和 3 TB 7.2K 驱动器。

- *Data Mart Appliance* 提供单一节点的入门级平台，专为小规模生产及测试和开发应用设计。它支持 Teradata Database 13.10 和更高版本中的大多数功能，包括支持混合 HDD 和 SSD 配置，但使用对称多处理 (SMP)，而非 MPP 架构。

最新一代的 670 于 2013 年 4 月面世，使用 Intel Xeon 处理器和 NetApp E2600 磁盘阵列，可支持 8 TB 的未压缩用户数据。Teradata 称，可实现 50% 到 70% 的数据压缩级别。

图 8 对这些平台进行了总结。

	Active Enterprise Data Warehouse 6750	Data Warehouse Appliance 2750	Integrated Big Data System 1700	Data Mart Appliance 670
引言	04/14	10/13	10/13	04/13
数目	高达 2,048	高达 2,048	高达 2,048	单个
机箱数量	2 到 32 个	四分之一到 32 个	四分之一到 32 个	半个到一个
处理器节点	2/24 x Ivy Bridge 2.7 GHz	2/24 x Ivy Bridge 2.7 GHz	2/16 x Sandy Bridge 2.6 GHz	2/16 x 2.6 GHz 或 2/12 x 2.0 GHz Xeon
磁盘类型	300, 450 或 600 GB 10K SAS ; 400 GB SSD	300, 600 或 900 GB 10K SAS	2 TB 或 3.0 TB 7.2K SAS	300, 450 或 600 GB 10K SAS ; 400 GB SSD
最大用户数据 (未压缩)	61 PB	21 PB	234 PB	8 TB

图 8 : Teradata 的特定工作负载平台

该表包含了已针对特定模型发布的 Teradata 用户数据容量，以及其他产品按比例分配的估计值。

此外，Teradata 还提供 *Teradata Aster Big Data Analytics Appliance*。它结合了 Aster Hadoop 和 MapReduce 技术，配备的工具可让用户借助 SQL 方法访问 Hadoop 数据，并使用基于 SQL 的分析和提取、转换和加载 (ETL) 工具来开发 Hadoop 数据库。该设备部署于配备 SLES 11 的 Intel Sandy Bridge 处理器上。

Teradata 已定义了统一数据架构 (Unified Data Architecture)，它旨在整合 Teradata Databases 和 Aster 环境。Teradata 宣称，借助该架构的扩展 QueryGrid，用户将能够对 Teradata 和 Hadoop 数据库中的数据开展多阶段分析。QueryGrid 将于 2014 年第三季度面世。

其他两种设备目前未着重强调：Teradata Appliance for SAS High Performance Analytics ( 尽管很多 Teradata 客户使用 EDW 上的 SAS，但事实证明，专用设备的吸引力有限 ) 和 Teradata Extreme Performance Appliance 4600，一种针对实时分析应用的全 SSD 设计。

Teradata 认为，使用混合 HDD 和 SSD 的配置和内存技术可更有效地提供数据仓库应用的高级别性能。

## *Data Warehouse Appliance*

Data Warehouse Appliance 2750 是原 2700 的升级版。Teradata 2750 系统配备了 12 核的 Ivy Bridge 2.7 GHz 处理器和高达 512 GB 支持 Intelligent Memory 的 RAM，而非 2700 中使用的配备 128 GB RAM 的 8 核 Sandy Bridge 2.66 GHz 处理器。

其他方面，Teradata 2750 与 2700 类似。与 2700 一样，它也可使用 300 GB、600 GB 或 900 GB 10K SAS 驱动器进行配置，并使用 Database 14 中用以开发列数据结构、基于应用特定集成电路 (ASIC) 的压缩引擎。Teradata 发布的值表明典型数据仓库工作负载 65% 到 70% 的压缩级别。

2750 平台可在四分之一到 32 个机箱之间的增量中配置，这些机箱支持的用户数据容量为 6.8 TB 到 21 PB 未压缩数据，或 22.4 TB 到 70 PB 的 70% 压缩数据。图 9 总结了一到六个机箱配置的容量。

驱动器类型 用户数据	300 GB		600 GB		900 GB	
	未压缩 (TBs)	70% 压缩 (TBs)	未压缩 (TBs)	70% 压缩 (TBs)	未压缩 (TBs)	70% 压缩 (TBs)
四分之一机箱	6.8	22.4	13.7	45.7	20.5	67.7
半个机箱	13.7	45.7	27.4	91.3	41.4	135.3
四分之三个机箱	20.5	68.3	41.4	138	61.7	203.6
整个机箱	27.4	91.3	54.8	182.7	82.3	271.6
2 个机箱	54.8	182.7	109.6	365.3	164.6	543.2
3 个机箱	82.2	274	164.4	548	264.9	874.2
4 个机箱	109.6	365.3	219.2	730.7	329.2	1,294.3
5 个机箱	137	456.7	274	913.3	411.5	1,358
6 个机箱	164.4	548	328.7	1,095.7	493.8	1,629.5

图 9 : Teradata Data Warehouse Appliance 2750 User Data Capacities

该表包含了已针对特定模型发布的 Teradata 用户数据容量，以及其他产品按比例分配的估计值。

## IBM PureData System for Analytics

### 概述

IBM PureData System for Analytics 基于 Netezza Performance Server (NPS) 架构。Netezza 在 2002 年发布了首款 NPS 产品，是 2000 年代数据仓库设备市场的先驱。IBM 于 2010 年收购了该公司。

核心 NPS 的设计采用独特的 MPP 组合、过滤、流处理和压缩技术，可提供行业领先的查询性能。借助成本相对较低的现场可编程逻辑阵列 (FPGA) 处理器和商品磁盘驱动器，可提供极具竞争力的价格/性能级别。

此外，NPS 架构的简单易用也是其广受欢迎的原因之一。一直以来，NPS 系统以极快的部署速度和极低的管理开销而著称。随着时间的推移，这成为了它吸引消费者的一个越来越重要的因素。

第四代 NPS 系统 TwinFin 系列产品于 2009 年面世，后更名为 IBM PureData System for Analytics N1001。第五代产品 N200X 于 2013 年 1 月面世。

N200X 系统采用更强大的 Intel 和 FPGA 处理器，配备容量更大的 SAS 磁盘。IBM 宣称，其性能大约比 N1001 高三倍。当前可提供半机架到四个机架的模型。

改良版 DB2 Analytics Accelerator for z/OS 能够从运行 IBM DB2 数据库 z/OS 版本的 IBM System z 主机中卸载分析处理。

## 架构和技术

NPS 系统架构的关键组件概述如下：

- **MPP** 采用专有的编译器，可将查询工作负载划分为名为 Snippet 的片段，随后这些 Snippet 由基于刀片的 *Snippet Processors* (S 刀片) 并行执行。
- **流处理**可缩短数据从磁盘的往返进程，相比传统 MPP 架构，时间可大幅缩短。数据以异步方式在磁盘和 S 刀片间移动（即传输延迟和协议开销为最低）。

在 PureData System for Analytics N200X 系统中，每个 S 刀片最多可从同样数目的磁盘、以高达 130 MB/秒的速度处理 40 个同步数据流。在配备 240 个活动磁盘驱动器的全机架系统中，聚集带宽为 240 x 130 MB/秒 = 31.2 GB/秒，或数据压缩后的接近 128 GB/秒。IBM 引用整体吞吐量级别为每机架 450 TB/小时。

- **过滤**确保在转到处理器内存之前，筛选出特定查询所不需要的数据。IBM 表示，通常 95% 到 98% 的用户数据都以这种方式排除。过滤由多种基于软件的引擎完成，大部分都是图 10 中所示的引擎。

<b>项目引擎</b>	根据处理中 SQL 语句的 SELECT 子句中规定的参数，过滤出无用的列数据。
<b>可视性引擎</b>	过滤出不应为执行查询所见的数据行，原因可能是该记录已被标记为被更早的查询删除，或它们在当前查询开始之后已被添加至数据库。以流处理的速度维持 ACID（原子性、一致性、独立性和耐久性）的合规性。
<b>限制引擎</b>	根据处理中 SQL 语句的 WHERE 谓词子句，过滤出无用的行级别数据。

图 10：主要的 PureData System for Analytics N200X 过滤引擎

该方法极大地降低了内部延迟。在传统 MPP 架构（如 Teradata Active EDW 和 2750 系统所采用的架构）中，磁盘和处理器内存之间的多个交互发生的方式可减缓整个过程周期。

- **压缩**使用一组算法来压缩数字、整数，以及在加载、插入和更新操作期间写入磁盘的时序（日期和时间）数据。系统基于数据特征自动选择最好用的压缩算法。

更早的 PureData System for Analytics 模型通常可实现两到四倍的压缩率。IBM 宣称，在 PureData System for Analytics N200X 中，借助算法改进和更强大的 FPGA，标准速度已比 N1001 系统提高两到三倍。

在 PureData System for Analytics N200X 中（如同更早的系统），S 刀片将执行核心 NPS 逻辑的 Intel 处理器与基于 FPGA 的过滤和控制引擎相结合。将其转移至 Intel 处理器前，由单独的 Compression Engine 解压数据。该结构如图 11 所示。

当前的 S 刀片将 IBM HX5 刀片服务器与双 Intel E7-2830 八核 2.13 GHz Linux 处理器和 128 GB RAM 以及双八核 Xilinx FPGA 相结合。在全机架中，支持多达七个活动的 S 刀片。

磁盘存储由每机架 12 机柜中的标准 600 GB 10K SAS 驱动器提供。完整的单架系统包含 288 个驱动器，其中 240 个为活动驱动器，14 个提供交换/记录空间，其余 34 个充当备件。数据在主磁盘中被条块化，在二级磁盘中被复制。



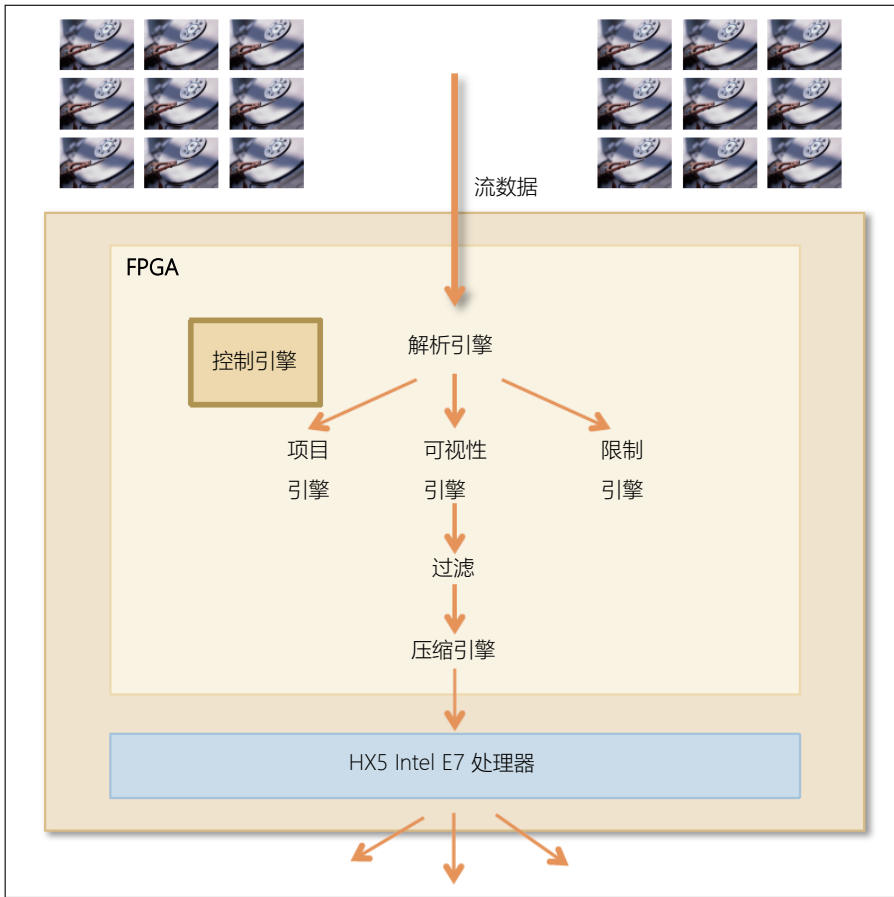


图 11 : PureData System for Analytics S-Blade Structure

当前的 PureData System for Analytics N200X 生产线如图 12 所示。

模型	002	005	010	020	040
用户数据	32 TB	96 TB	192 TB	384 TB	768 TB
机架	四分之一	二分之一	1	2	4
活动的 S 刀片	2	4	7	14	28
Intel 处理器核心	32	64	112	224	448
FPGA 核心	32	64	112	224	448

图 12 : 当前的 PureData System for Analytics N200X 生产线

此外，双冗余对称多处理 (SMP) 主机实施系统范围的 SQL 编译器、查询计划、管理和优化以及其他功能。在 PureData System for Analytics N200X 设备中，IBM HX5 刀片服务器和 Red Hat Linux 6 被用于担当这一角色。

# 计算基础

## 竞争对手档案

本报告得出的计算基于图 14 中所示的四个组合概要。

零售公司	数字媒体公司	电信公司	金融服务公司
<b>业务概况</b>			
多渠道零售商 80 TB + 用户数据	在线内容提供方 150 TB+ 用户数据	固网电信、移动与因特网服务提供商 350 TB+ 用户数据	多样化的跨国交易服务提供商 600 TB+ 用户数据
<b>应用</b>			
销售与库存分析、客户行为建模、规划、销售规划、预测、竞销活动管理、一对一营销、多种	客户定位；在线广告与内容传递及其他金融服务应用、在线服务、远程通信、差旅与其他业务	呼叫详细记录 (CDR) 分析、社交媒体与应用范围的账单数据如成本与盈利率分析、使用情况预测、降低流失率	合规与监管应用的实时交易分析；识别非正常模式，以检测欺诈、反洗钱等。
<b>IBM PUREDATA SYSTEM FOR ANALYTICS N200X</b>			
半机架	全机架	2 个机架	4 个机架
96 TB 用户数据*	192 TB 用户数据*	384 TB 用户数据*	768 TB 用户数据*
0.5 名 FTE DBA	0.4 名 FTE DBA	0.5 名 FTE DBA	1.0 名 FTE DBA
部署时间：4 天	部署时间：2 周	部署时间：2 个月	部署时间：3 个月
<b>TERADATA DATA WAREHOUSE APPLIANCE 2750</b>			
半机架	全机架	2 个机架	4 个机架
91.3 TB 用户数据*	182.7 TB 用户数据*	365.3 TB 用户数据*	730.7 TB 用户数据
1.2 名 FTE DBA	0.75 名 FTE DBA	1.25 名 FTE DBA	* 2 名 FTE DBA
部署时间：4 周	部署时间：2 个月	部署时间：6 个月	部署时间：9 个月

\* 压缩

图 14：竞争对手档案

Teradata 2750 系统利用 Teradata Database 14 进行配置，系统规模和 FTE 人员配置计算反映了该版本的使用情况。Teradata 2750 和 PureData System for Analytics N200X 均使用 600 GB 10K SAS 驱动器进行配置。

## 成本计算

成本计算如下：

- **系统成本**基于供应商提供的绑定配置的已贴现购置和维护 (Teradata) 或支持 (IBM) 费用进行计算。
- **人员成本**用于上图所示 FTE 数量。成本根据年平均工资进行计算，Teradata 2750 和 PureData System for Analytics N200X DBAs 的年平均工资分别为 114,138 美元和 97,156 美元。为了给福利、奖金和相关项目预留空间，工资增加了 56.7%，并乘以三，以示三年。
- 零售、数字媒体和电信公司中的 Teradata 2750 设施**部署成本**分别根据该公司 Accelerate for Retail、Accelerate for Marketing 和 Accelerate for Data Warehousing 产品的定价进行计算。金融服务公司的成本基于适用的 Teradata Professional Services 技能级别和价格进行计算。

PureData System for Analytics N200X 在零售和数字媒体设施中的部署成本基于 IBM 业务合作伙伴产品进行计算。不存在与 Teradata Accelerate 程序直接对等的 IBM 产品。远程通信和金融服务业公司的成本基于适用的 IBM Global Services 技能级别和价格进行计算。

两个平台中，成本均包含外部专业服务人员为现场访问人员提供的差旅和培训 (T&E) 成本。

- **培训成本**包括针对 Teradata 2750 的 17 天 DBA 课程以及针对 PureData System for Analytics N200X 的 5 天 DBA 课程，外加其他在线教育课程。假定在每个案例中，现场派驻两名人员；即未预留 T&E 费用。
- **设施成本**为设备的能源消耗。计算基于供应商规格以及（适当情况下的）ITG 估算，并假定在三年时间内几乎 365 天 24 小时不间断运营。

所有值均针对美国。

## 成本分析

拥有成本分析如图 15 所示。

	零售公司	数字媒体公司	电信公司	金融服务公司
<b>IBM PUREDATA SYSTEM FOR ANALYTICS N200X</b>				
购置	701,250	1,229,500	2,475,000	4,922,500
支持	252,450	442,530	891,000	1,772,100
部署	98,530	165,985	388,900	663,458
人员	228,365	182,692	228,365	456,730
培训	6,000	6,000	11,000	11,000
设施	11,446	20,439	40,878	81,757
<b>总计 (美元)</b>	<b>1,298,041</b>	<b>2,047,146</b>	<b>4,035,143</b>	<b>7,907,545</b>
<b>TERADATA DATA WAREHOUSE APPLIANCE 2750</b>				
购置	625,600	1,244,400	2,488,800	4,970,800
维护	375,360	746,640	1,493,280	2,982,480
部署	279,502	753,888	1,362,550	2,598,694
人员	633,112	395,695	659,492	1,055,186
培训	32,200	32,200	52,800	52,800
设施	8,743	18,068	36,136	72,273
<b>总计 (美元)</b>	<b>1,954,517</b>	<b>3,190,891</b>	<b>6,093,058</b>	<b>11,732,233</b>

图 15 : 拥有成本分析

# International Technology Group

*ITG 帮助您认识当前形势，了解您的竞争优势  
... 这可能影响您未来的发展和获利前景*

International Technology Group (ITG) 成立于 1983 年，是一家独立的研究和管理咨询公司，专注于信息技术 (IT) 投资策略、成本/效益指标、基础架构研究、部署策略、业务一致性和财务分析。

ITG 是开发总拥有成本 (TCO) 和投资收益率 (ROI) 流程和方法的早期创新者和开拓者。2004 年，该公司被信息技术财务管理协会 (ITFMA) 授予“十年教育奖”。这个主导性专业协会专注于教育，致力于促进最终用户 IT 组织中的财务管理实践。

客户服务旨在提供真实的数据和可靠的文档记录，以协助决策制定流程。所提供的信息为发展战术和战略计划奠定了基础。使用最有效的方法分析重要发展并提供实践指导，以响应可能影响复杂的 IT 部署议程的更改。提供广泛的服务，为客户提供必要信息，以补足其内部能力和资源。

客户包括公共和私有领域不同层面的 IT 最终用户，他们分别代表跨国企业、工业公司、金融机构、服务型组织、教育机构、联邦和州政府机构以及 IT 系统供应商、软件供应商和服务企业。联邦政府客户已包括国防部、运输部和财政部中的机构（如 DISA、FAA 和 US Mint）。

WAL12377-CNZH-01

---

Copyright © 2014 International Technology Group. All rights reserved. 未经 International Technology Group (ITG) 事先书面同意，不得以任何方式或任何形式复制或分发本文档中的全部或部分材料，包括原件。所含信息均来自可靠来源，并反映了当时的结论。尽管本文可能会使用各种不同来源的公开资料，但就本文所述的各个问题而言，本文并不代表此类资料来源方的观点。本文档中包含的材料和提出的结论如有更改，恕不另行通知。不保证此类材料的准确性、完整性或充分性。对于本文档中所含材料或与其解释相关任何错误、遗漏或不足将不承担任何责任。本文档中包含的商标归其各自的所有者所有。

---

IBM PureData System for Analytics 的成本/效益案例：  
Teradata Data Warehouse Appliance 的成本和时间与价值比较