火龙果•整理
uml.org.cn

Teradata 数据仓库一体机暨技术特点概述

本文档及其所含信息为机密材料
并且 Teradata 所拥有。
本文档中的任何部分未经 Teradata 书面授权，
不得将材料泄露给第三方,也不得以任何手段、任何形式进行复制与传播
Copyright © 2012 Teradata 版权
保留所有的权利

1.1 一体机解决方案与传统解决方案对比

1.1.1 Teradata 与传统解决方案对比

1. 数据仓库平台的发展趋势

随着信息爆炸式增长、复杂结构数据的涌现以及云计算和虚拟化等关键技术的广泛应用，业界领先的数据库平台厂商纷纷推出数据仓库一体机产品以面对日益增长的海量信息分析的挑战。基于大规模并行处理架构（MPP）的数据仓库一体机适合为企业搭建一体化的集成数据分析环境(IDE)，能够有效地支撑海量数据分析和挖掘应用，从而为企业及时获取有价值的信息。

全球最具权威的 IT 研究与顾问咨询公司 Gartner 在 2010 年 3 月的研究分析报告中指出:“数据仓库专用平台是企业数据仓库的未来。在 2015 年前，全球 50%的企业会选择数据仓库专用平台做为自己的数据仓库解决方案”。

Gartner Research

Publication Date: 8 March 2010 ID Number: G00174689

Are Data Warehouse Appliances in Your Future? Plan On It!

Donald Feinberg

With their business value proven, and more choices available in the market, the use of data warehouse (DW) appliances is surging in the market. Because of their potential value, organizations must investigate DW appliances as an alternative or addition to a custom DW environment.

Key Findings

- DW appliances deliver easy to install, easy to manage DW platforms with predictable performance and single point-of-contact service.
- DW appliances are used to support all types of DW, from application-specific data marts to the enterprise DW, with complex mixed workloads.
- DW appliances are available from all the incumbent database management system (DBMS) vendors, as well as many vendors with appliance-specific offerings.
- For DW platforms, DW appliances have become a competing deployment standard to traditional self-configured hardware and storage deployments.

WHAT YOU NEED TO KNOW

DW appliances are available today from most of the vendors with a DW DBMS, as well as many vendors with only a specific DW appliance. They are being used for all types of data warehousing, from application-specific data marts to enterprise DW implementations with complex mixed workloads (see Note 1). Over the past two years, interest in and use of DW appliances have been growing fast. As the value proposition of DW appliances grows, more organizations will realize their benefits. By 2015, at least 50% of enterprises with DWs in production will include a DW appliance. This research should be of interest to CIOs, chief technology officers (CTOs), IT architecture teams and database administration teams, as well as anyone involved in the design and acquisition of a DW.

图 1 Gartner 对数据仓库平台发展趋势报告

2. 数据仓库一体机与传统解决方案的对比

数据仓库一体机与传统的基于小型机的解决方案相比，在如下表所示的数据仓库几方面关键领域存在明显差异。

关键领域	一体机方案	传统小型机方案
性能	高： 针对数据仓库应用优化定制	中： 典型配置很难满足数据仓库复杂应用环境
整体拥有成本	低： 一体机的优势体现在采购、运维、以及投资回报上	较高： 配置、运维复杂，投资回报相对较差
易用易管理性	好： 预先集成、统一管理，简单易用	一般： 配置复杂，管理依赖集成工具，对管理能力要求高
扩展能力	高： 提供搭积木式扩展	中： 内部扩展受限于单机(SMP)槽位，外部扩展受限于网络连接

3. Teradata 数据仓库一体机的特点和优势

Teradata 作为专注于数据仓库领域的厂商，在数据仓库一体机领域已经有 30 年以上的经验。Teradata 数据仓库一体机在性能、整体拥有成本、易用易管理性与扩展能力上表现更是优异，具体情况如下表所示：

数据仓库特性要求		Teradata 专业一体化平台的特点及优势
性能	数据规模与应用场景复杂	大数据量复杂运算的效率快，适合分析系统的数据规模和应用场景
	并行处理	Teradata 是为数据仓库处理设计的专业数据库，无条件的并行处理
	混合负载管理	提供成熟的混合负载管理和优先级调度能力，能够很好的支持各种复杂应用对资源的使用
	均衡配置	充分考虑计算能力、存储能力和 I/O 吞吐量的平衡，提供整体的最优计算能力
整体拥有成本	投资保护	多代产品共存，保护前期投资
	装机即可用	一体机，装机即可使用
	压缩比高	可以按列存储，提供卓越的压缩技术，更高的压缩比率
	无空间损耗	无需额外系统空间损耗，用户可用空间更高
	产品满足不同需求	多种平台产品家族，满足不同客户需求，保证投资回报
	运维管理方便	运维简单、管理自动化程度高，人工参与度低
易用性	系统易于管理	数据库完全自管理，空间、数据分布和管理无需人工干预
	系统管理成	统一的系统管理，系统自动化程度高，更少的 DBA

	本低	
可扩展性	线性扩展	节点和存储与性能线性扩展
	可扩展能力	最大可支持到 4096 个节点，186PB 数据处理规模
未来技术发展	开放的云平台	提供云计算解决方案：分析云方案已应用于 Apple iCloud,Amazon EC2
	虚拟存储	领先的虚拟存储技术，实现自动的多温度数据管理
	引领技术发展	融合了对未来大数据(非结构化数据)分析的解决方案，将继续专注于数据仓库与分析领域，持续提供领先的数据仓库产品与技术

4. Teradata 数据仓库一体机系列及配置

Teradata 超过三十年在数据仓库和企业数据分析的专注研究及技术创新，积累了丰富的数据仓库系统建设经验，同时，形成了丰富的数据仓库一体机产品家族，分别能够适应各行业客户不同业务应用场景。

Teradata 数据仓库一体机产品家族如下表所示：

	数据集市专用平台	巨量数据专用平台	数据仓库专用平台	超强性能专用平台	动态企业级数据仓库
适用范围	测试/开发或小型数据集市	巨量数据存储及巨量数据分析	数据仓库或部门级数据集市	操作分析型系统高性能应用	企业级决策及操作智能数据仓库
容量	高达 11TB	高达 186PB	高达 343TB	高达 24TB	高达 92PB

范围					
主要特点	部门级分析，入门级企业数据仓库	存档分析 ,深度历史数据分析	战略智能决策支持快速扫描	操作性智能，低容量，高性能	动态负载，实时更新，战术战略时间响应

Teradata 建议在具体设计与配置一个数据仓库系统时 ,一个重要的考核指标是系统节点与存储设备之间的 I/O 带宽配合，要力求做到“均衡”以使系统性能最佳化。如下图所示：如果存储设备和磁盘提供足够的 I/O 带宽，可以充分利用系统节点的处理能力。如果一个系统节点的 CPU 使用率能够达到峰值的 90%，该系统配置就是“均衡”的；如果一个系统节点的 CPU 使用率在峰值的 75%以下，该系统配置就是“IO 受限的”的，反之则声”CPU”受限的。在设计一个数据仓库系统时，总是配置“均衡”方案，以取得最佳的整体性能。

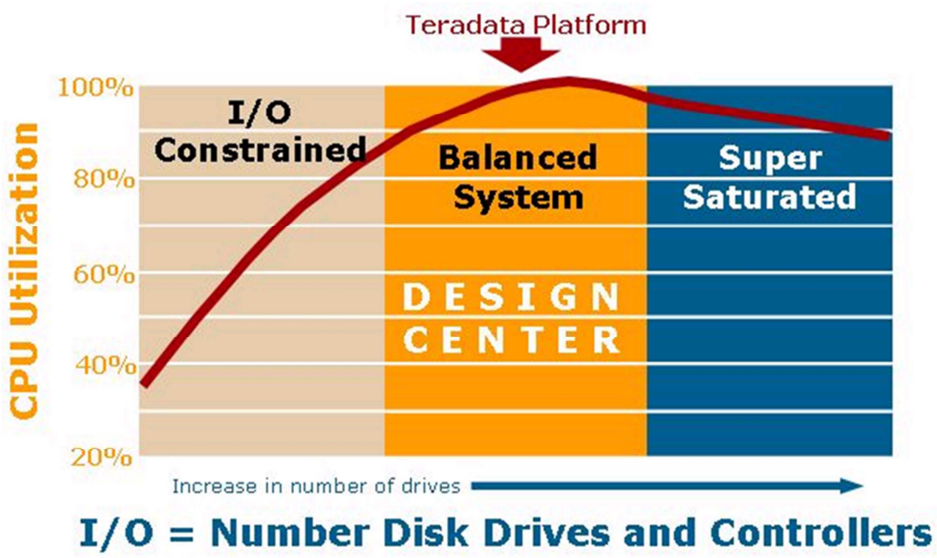


图 2 CUP 与 I/O 的均衡配置

因此，Teradata 数据仓库一体机产品的选型和配置要从客户的实际业务需求出发，分别从数据容量、系统性能、源数据获取、历史数据存储、系统业务支撑、动态负载管理、系统可用性、系统灾备要求、系统开发与测试要求等方面进行评估，并对每方面的评估结果进行综合分析与评估，从而得出所需要的系统能力指标，从而决定数据仓库一体机产品的选型与详细的产品配置规划。

5. 小结

作为 Gartner 评估的数据仓库平台十四连冠，Teradata 一体机解决方案面向海量数据分析应用，针对计算能力、存储能力和 I/O 吞吐能力进行了深度的优化，使软硬件平台达到完美融合。与传统解决方案相比，无论在性能、高可用性、整体拥有成本（TCO）、易用易管理性、扩展性、海量数据分析案例及最佳实践方面，都处于领先地位，符合技术发展趋势。

2.数据仓库一体机平台技术分析

2.1 产品关注点

一体机相关产品的主要关注点如下表所示：

整体性能	系统对于海量数据的处理性能、并行处理能力与整体性能表现
扩展能力	扩展后相同查询量、数据量与系统性能应成正比并能线性扩展
高可用性	单点故障不影响系统持续对外服务能力，并支持双系统高可用性解决方案
并发处理能力	支持高并发访问同时运行，包括加载、库内转换、查询分析等
系统稳定性	处理大数据量、复杂查询、多类型多并发作业任务时，系统稳定可靠
库内分析	支持端到端的库内数据挖掘分析，避免数据搬迁开销；支持所有主流数据挖掘厂家工具及算法
与第三方工具融合性	支持所有主流的第三方 ETL 工具、报表及 OLAP 工具、数据挖掘工具、应用访问等工具等
易用性/易管理性	提供统一易用的界面综合管理系统硬件、软件、数据库、负载等管理；DBA 不介入数据分布
混合/动态负载管理	动态调整系统资源，同时支持各类负载需求。支持同时进行复杂的战略深度分析、大量的数据加载及更新
多应用支持能力	单一平台支持多类对系统资源使用各异的应用，具备平衡优化应用的能力
数据实时度	支持与业务系统同步的数据加载、更新速度，支持时延要求苛刻的业务活动及分析
产品技术架构	数据仓库一体机应具备并行处理、消除单一节点等性能瓶颈的架构优势

专业服务	售后服务能力、专业服务队伍构成及咨询服务能力、实施经验和实施方法论的完备性
案例经验	国内国际大数据量数据仓库案例、解决方案完整性、先进性
整体拥有成本	投资保护，系统运维简便，工作量少，回报率高
第三方评测	在 Gartner、 Forrester 等第三方权威评测机构的数据仓库综合评测中排名

2.2 Teradata

2.2.1 简介

Teradata 作为全球领先且专注于数据仓库技术与服务的提供商，从领先的技术、专业的服务，到众多的成功案例、丰富的实施经验，造就了 Teradata 数据仓库专用平台与众不同的成熟优势。

一、 全球领先的数据仓库和企业智能技术

1、 Teradata 数据仓库采用非共享 MPP 核心架构，以无可匹敌的高性能在主流数据仓库产品中独领风骚；

2、 “多温度数据存储调度”、“企业级固态硬盘”等领先技术的应用，使 Teradata 数据仓库平台始终在竞争中技高一筹；

3、 “线性可扩展”、“稳定的数据架构”、“成熟的行业逻辑数据模型”等特点，保障了 Teradata 企业数据仓库及应用的可持续发展；

4、 “7X24 高可用”、“混合负载管理与优先调度”，为 Teradata 数据仓库应用提供了高可靠的保证；

二、 专注于数据仓库建设三十年以上

1、 Teradata 被国际权威机构 Gartner、Celent、Ventana Research 和 AMR Research 评为业界一流的专业服务提供商；

2、 30 多年来形成了一支遍布全球范围的、经验丰富的咨询服务专业团队，在中国就有 800 多人。

三、 全球 500 强超过 80%都是 Teradata 成功解决方案的受益者

1、 Teradata 成熟的数据仓库解决方案，遍及能源、电力、金融服务、零售、保险、通信、媒体和娱乐、政府与公共服务、医疗与制药、制造、博彩、旅游与运输、公用事业等所有行业；

2、 目前，全球 3000 强中的 1000 多家和全球 500 强企业中的超过 80%都是 Teradata 成功解决方案的受益者；

2.2.2 技术架构

Teradata 的技术架构如下图：

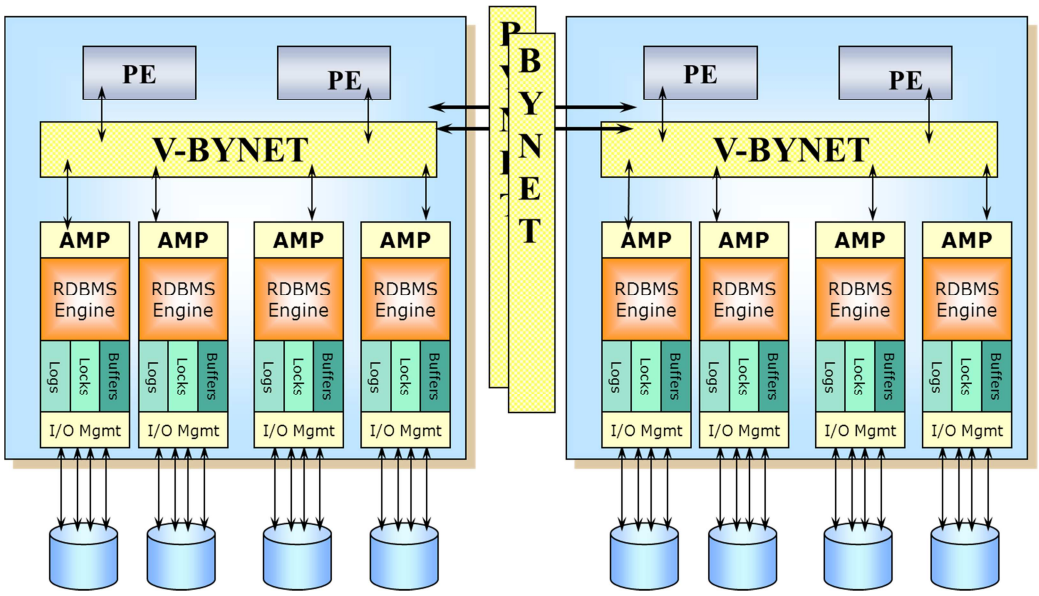


图 2 Teradata 技术架构

Teradata 采用非共享(Share-Nothing)的大规模并行处理(MPP)技术体系架构。在物理布局上，Teradata 系统主要包括三个部分：处理节点（Node）、用于节点间通信的内部高速互联网络（InterConnection）和数据存储介质（通常是磁盘阵列）。其中每个节点都是 SMP 结构的单机，多个节点一起构成一个 MPP 系统，MPP 系统内服务器节点中各个处理器单元管理自己的内存、IO 与存储，而实现并行处理的关键技术是如何将这些处理单元联成一个系统，并能在各处理单元之间进行负载动态平衡。Teradata 将 CPU、内存、与存储虚拟化，并采用其专利的 BYNET 互联技术连接各处理单元。

1. 单个节点的硬件结构

Teradata 系统中的每个节点在物理上都是一个 SMP 处理单元，事实上就是

一台多 CPU 或多核的计算机。节点硬件包括 CPU、内存、用于安装操作系统和应用软件的本地磁盘、与外界交互的网卡及 BYNET 端口。节点的网卡根据具体的网络环境而不同，通常包括两种，一种是与大型主机连接的 Channel Adapter，另一种是局域网网卡。通常情况下，一个节点上只会使用一种网卡，但会有多块网卡，分别用于不同的连接和冗余。

2. 单个节点的软件结构

在软件结构上，每个节点自下向上包括操作系统软件（OS）、Teradata 并行数据库扩展（PDE）和相关应用程序，其中 PDE 的主要职责是管理和运行虚拟处理器，其中主要包括 PE 和 AMPs。

- Teradata 并行数据库扩展（PDE，Parallel Database Extensions），是直接架构在操作系统之上的一个接口层，用于为 Teradata 提供并行环境，并保证这个并行环境的可运行性和健壮性。PDE 的主要功能是执行虚拟处理器、进行 Teradata 并行任务调度、进行操作系统内核和 Teradata 数据库的运行故障处理。
- 虚拟处理器（VPROC，Virtual Processor），是一系列软件进程，这些进程驻留在一个节点上，依赖 PDE 环境运行，并接受 PDE 调度。VPROC 为 Teradata 的底层服务进程，虚拟处理器完成 Teradata 数据处理的主要工作，按照工作性质的不同，虚拟处理器主要包括两大类——解析引擎和存取模块处理器。
- 解析引擎（PE，Parsing Engine），用于进行客户系统（通常是使用 Teradata 数据库的应用程序的 SQL 请求）和存取模块处理器之间的通讯和交互，主要的功能包括任务控制（Session Control），SQL 语句的解析、优化、查询步骤的生成和分发，并行化预处理和返回查询结果。当多个用户同时访问系统时，Teradata 能够通过 PE 在各节点间自动平衡负载，不需要数据库管理员 DBA 或者系统管理员的干预。每个节点中 PE 和 AMP 的数目可以由 DBA 进行配置。
- 存取模块处理器（AMP，Access Module Processor），这是 Teradata 数据库的关键进程，它控制所有与数据库有关的操作。每个节点会有多个 AMP，每个 AMP 只负责磁盘阵列的一部分磁盘，是 Teradata 数据库 Share Nothing 架构的核心表现。
- 虚拟磁盘(VDisk,Virtual Disk),这是一个纯粹的逻辑概念。典型的 Teradata

MPP 系统的数据存储都是以磁盘阵列（Disk Arrays）的形式实现的，在物理上是一个个存放于标准磁盘阵列柜中的磁盘阵列模块。Teradata 系统中的每个 AMP 在处理数据存储时，会根据哈希算法把不同的数据自动地均匀地分散存储到磁盘阵列中的不同的磁盘上，这样，在逻辑上，我们就把磁盘阵列中不同磁盘上存储着的那些由同一个 AMP 负责存储和维护的数据合并在一起，就像它们在一个磁盘上一样，这就是 VDisk 的作用。

3. BYNET

在 Teradata MPP 系统中，实现其海量并行处理的关键技术就是如何将各 SMP 节点联成一个系统，并能在各节点之间进行负载动态平衡。Teradata 用来实现节点互联的这种关键技术就是 BYNET。BYNET 是一组硬件和运行在这组硬件上的一些处理通讯任务的软件进程组合体，用于节点之间的双向广播（bidirectional broadcast）、多路传递（multicast）和点对点通信（point-to-point communication），同时，BYNET 还实现 SQL 查询过程中的合并功能。

在数据仓库中，经常有各种复杂的数据库查询。Teradata 的基本思想是，把一个复杂查询分解成多个简单子查询，并均衡地分配到各个结点上执行，这样，总的查询处理时间就能缩短很多。BYNET 是实现这种设想的重要保障。

另一方面，通过 BYNET 连接，MPP 系统中任何两个 SMP 节点间都是互相连接的，这也是 Teradata 系列计算机具有线性可扩展能力的主要原因。由于两两互连，因此增加结点时，系统总带宽也成比例增加。

2.2.3 技术优势

1.线性扩展能力

Teradata 采用并行非共享(share-nothing)MPP 架构，采用“搭积木”方式建设数据仓库系统，以使数据仓库系统得以“按需”延展。提供线性度为 1 的可扩展性，例如，在数据量不变的情况下，增加一倍的系统节点平台，处理能力即能提升一倍。而 Teradata 独有的多代并存的扩展技术，也充分地保护了用户的原有投资。

2.易用易管理

Teradata 数据库完全自管理，空间、数据分布自动管理，对 DBA 的人工干预要求最低，一体机装机即可使用。

另外，Teradata 提供图形化的方便易用的整合界面，对数据仓库环境进行自助、快速又功能强大的管理工具。

3.混合工作负载管理

Teradata 可以对同一环境中的各种负载进行管理（报表、即席查询、动态数据加载、动态事件侦测、动态访问等），通过完善的优先级调度机制，支持动态的负载管理，并自动进行负载均衡。

4.高效的数据库 SQL 优化器

最优的基于成本的数据库优化器可以处理任意复杂的 SQL 语句，通过智能的查询执行分解，既结合数据的分布情况、又结合系统的可用资源情况，并充分利用产品的并行处理机制，从而可以产生极其高效的执行步骤与，最终以最快的方式返回 SQL 语句结果，支持多种方式的数据仓库应用类型。

5.优异的并行处理能力

Teradata 采用非共享(share-nothing 架构),其并行处理机制的实现方式称为多维并行处理机制，包括查询并行(Query 并行)、步内并行(Within-a-Step 并行)、多步并行(Multi-Step 并行)等无条件的并行处理，正是由于其不受限的并行处理机制，以及对混合负载的高度管理能力结合最优的基于成本的 SQL 解析优化器，Teradata 可以提供最好的并发查询性能，具有在业界公认的海量数据处理性能。

6.实时数据处理能力

Teradata 以其“动态数据仓库”的技术与理念引领数据仓库向实时数据处理的方向发展，其能力通过以下 6 个方面来反映：

1）实时访问：对实时或者近实时数据的、对第一线操作类型的决策支持或者服务支撑，SLA 通常在 5 秒以下;

2) 实时加载：每日多频次的小批量或者近实时的数据加载，SLA 通常在秒级到分钟级；

3）动态事件响应：主动监控业务活动，基于一定的规则和环境主动触发业务流程；对于系统和用户，支持操作层面的业务流程；

4）动态负载管理：在各类混合负载下，动态管理系统的各类资源以保证系统性能的最优化和系统资源被充分利用；

5）动态企业集成：决策支持服务能够动态集成到企业的 IT 架构中；

6）高可靠性:7X24 的高可靠性手段保证实时数据的处理机制。

7.高可用性

Teradata 数据仓库采用多个层面的技术和架构保证系统的高可用性，在数据仓库中采用的热备节点(Hot Standby)、Clique 内节点互备、磁盘 Raid 技术等多种选项，以及在数据仓库系统间的 Dual-Active 技术等，从软件、硬件、技术架构三个方面都实现了数据仓库的高可用性。

8.系统稳定可靠

Teradata 数据库提供一系列的数据保护机制，从系统级到数据库级，全面保证数据在意外事件中能够得到保护并提供数据恢复能力。另外 Teradata 针对对海量数据、并发处理、复杂负载的数据仓库环境所构建的成熟与强大的软件体系也是系统整体健壮性的保证，例如其高效的并行处理优化器，保证复杂的查询语句都能保证其顺利的运行，不象其他系统会造成死机的现象。

9.开放性

Teradata 是一个符合工业标准的开放式的系统，其一体机平台都采用业界标准的服务器、存储、芯片等产品设计。服务器平台采用 LINUX 操作系统，数据库为标准关系型数据库，遵从 ANSI 92/99 等标准，并提供 ODBC/JDBC/OLEDB 以及 C/Java 等多种编程语言 API 接口。

正是由于 Teradata 的这种开放性，使得其可以支持所有主流的数据仓库/BI 工具提供商，包括主流的 ETL 工具、报表及 OLAP 工具、数据挖掘工具、应用访问等工具。而 Teradata 更是与行业中全球领先的软件供应商和系统整合者一起通力合作，提供经优化的方式访问 Teradata。

10.数据库内含丰富的分析功能

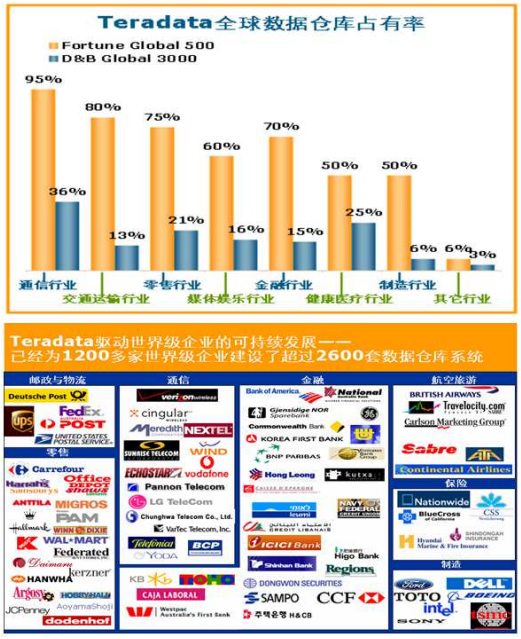
Teradata 是专为数据仓库设计的，主要用来进行数据的综合分析和处理，因此在开发时嵌入了丰富的分析功能，其中既 OLAP 函数，也有数据挖掘函数，包括：移动平均 MAVG、移动和 MSUM、移动差分 MDIFF、采样 SAMPLE、分位 QUANTILE、限定 QUALIFY、线性回归、因素分析、决策树、分群等等。这些函数可以和标准的 SQL 语句一起使用。

所有这些函数都是在 Teradata 内部以并行方式来工作，相较于外部工具部

署的方式，这种库内运行的方式可以充分利用 Teradata 的并行处理能力，处理效率与速度更高更快。

2.2.4 经典案例

自 Teradata 公司成立以来的 30 多年时间里，先后在全球实施了 2600 多个大型的数据仓库项目，其中，数据量超过 PB 级的 Teradata 客户有 20 家。在大型数据仓库平台领域，Teradata 的市场占有率远远超过其它竞争对手。



Teradata PB级客户名单（20家）：
— 截止到2011年6月

客户名称	数据量（PB）
eBay Inc. *	47.74
Wal-Mart Stores, Inc.	6.99
Bank of America Corporation	4.70
Apple Inc.	4.69
AT&T Inc.	4.41
Verizon Wireless	2.22
Lloyds TSB Bank Plc	1.90
Wells Fargo & Company	1.89
WellPoint	1.86
Dell Inc.	1.71
Morgan Stanley	1.66
Boeing	1.28
T-Mobile USA	1.16
Barclays Bank Plc *	1.16
Intel Corporation	1.16
Capital One Financial Corporation	1.14
Best Buy Co., Inc.	1.12
Pfizer, Inc.	1.10
Vodafone Group Services GmbH	1.03
The Home Depot, Inc.	1.02

图 3 Teradata 全球行业占有率和超 PB 级的客户名单

在国内，Teradata 在金融、保险、证券、电信、邮政、航空、制造等行业拥有丰富的数据仓库建设和实施经验。基于国家电网公司企业规模和信息化应用现状与未来规划，我们选择 Teradata 在类似规模企业—中国工商银行和中国移动集团的数据仓库平台实施案例进行介绍。

一、中国工商银行企业级数据仓库实施案例

1、项目背景与发展历程

中国工商银行于 2000 年底起着手启动全行数据仓库建设。由于中国工商银行数据仓库建设初期业务需求迫切，为了加速建设步伐，确保周期短、投资小、

见效快，中国工商银行经过分析论证，决定采用先建数据集市、后建数据仓库的方案进行数据仓库建设。

2004 年 10 月，中国工商银行启动了数据仓库逻辑数据模型客户化建设项目，整个项目建设历时 5 个月，基本形成了中国工商银行自有知识产权的数据仓库逻辑数据模型。

2005 年 3 月，在顺利完成数据仓库逻辑数据模型建设项目后，中国工商银行启动了数据仓库物理平台选型测试的任务。该任务历时 8 个月，于 05 年 11 月顺利完成。物理平台选型测试工作作为中国工商银行数据仓库物理平台产品选型的参考、考察数据仓库产品厂商的技术水平、服务能力和实施经验的任务，为中国工商银行数据仓库物理平台的选型奠定了坚实的基础。

在进行数据仓库商务的同时，为保证后续数据仓库项目的顺利实施，2006 年 12 月，中国工商银行启动了数据仓库体系建设规划研究项目，从架构、功能、规模等方面对数据仓库和数据集市进行规划，以指导后续数据仓库建设工作。

2007 年 4 月，中国工商银行数据仓库实施项目启动，到 2008 年 1 月正式上线试运行，开发测试工作历时 10 个月，已经顺利完成投产工作和系统上线。

截至到目前，中国工商银行数据仓库建设已经进入第三期，入仓源系统 62 个，源系统覆盖范围达到 90%以上。基础数据平台支持下游 17 个应用，提供 160 个对外数据接口。

2、企业级数据仓库平台架构与应用

中国工商银行采用了集中式的建设模式，其用户数据量已经超过 1PB，目前是国内最大的金融业企业级数据仓库应用。

中国工商银行企业级数据仓库平台的逻辑架构如下图所示：

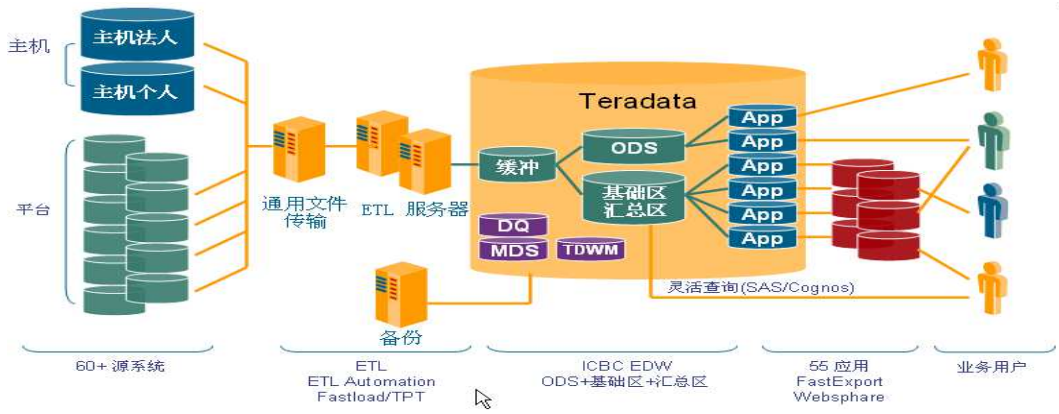


图 4 工商银行数据仓库平台逻辑架构

3、项目实施效果

经过近多年的持续开发和建设，中国工商银行数据仓库系统在不断的发展壮大，随着整合源系统数目的增加，仓库已经积累了大量的数据，成为全行整合度最高、信息量最大的系统之一。支持的应用的数目也从一期的 3 个扩充到现在的 17 个主题，并建立起比较完善的元数据管理、数据质量管理体系。

二、中国移动集团总部数据仓库实施案例

1、项目背景与发展历程

中国移动经营分析系统从 2003 年开始建设，已经经历了 9 年的建设，Teradata 作为中国移动数据仓库平台的实施方，按照集中化、标准化、开放化、体系化的发展要求，逐步构建企业级数据仓库，支持应用快速、灵活部署以及与其他 IT 系统的高效互动，促进中国移动业务支撑网逐步向智能型支撑网演进。项目发展历程如下图所示：

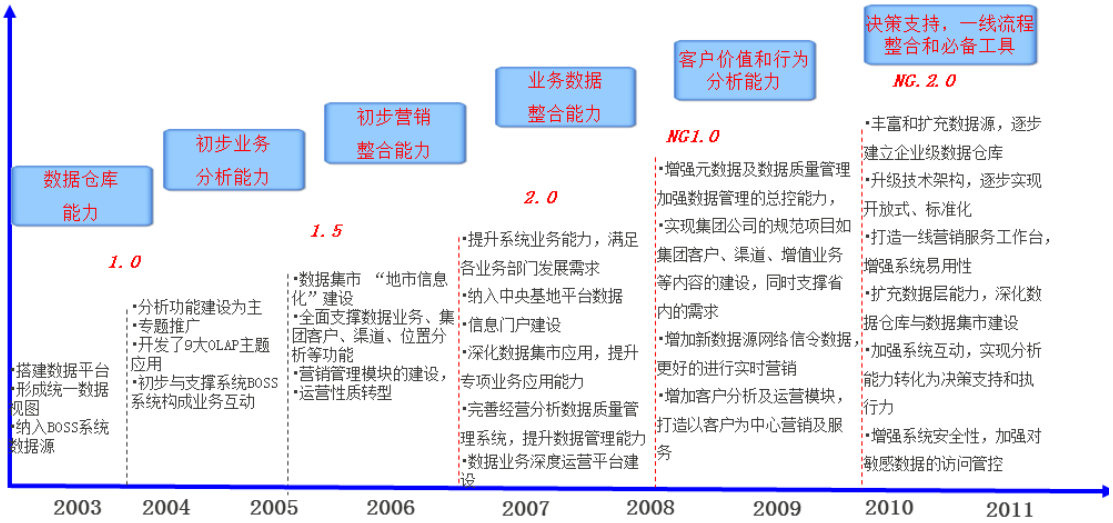


图 5 中国移动集团数据仓库平台项目发展历程

2、企业级数据仓库平台系统与应用架构

中国移动经营分析系统采用总部和省公司两级部署模式，省级经营分析系统部署面向省公司的相关应用，一级经营分析系统基于省级经营分析系统上传的数据，部署面对全网的应用。 中国移动经营分析系统物理架构如下图所示：

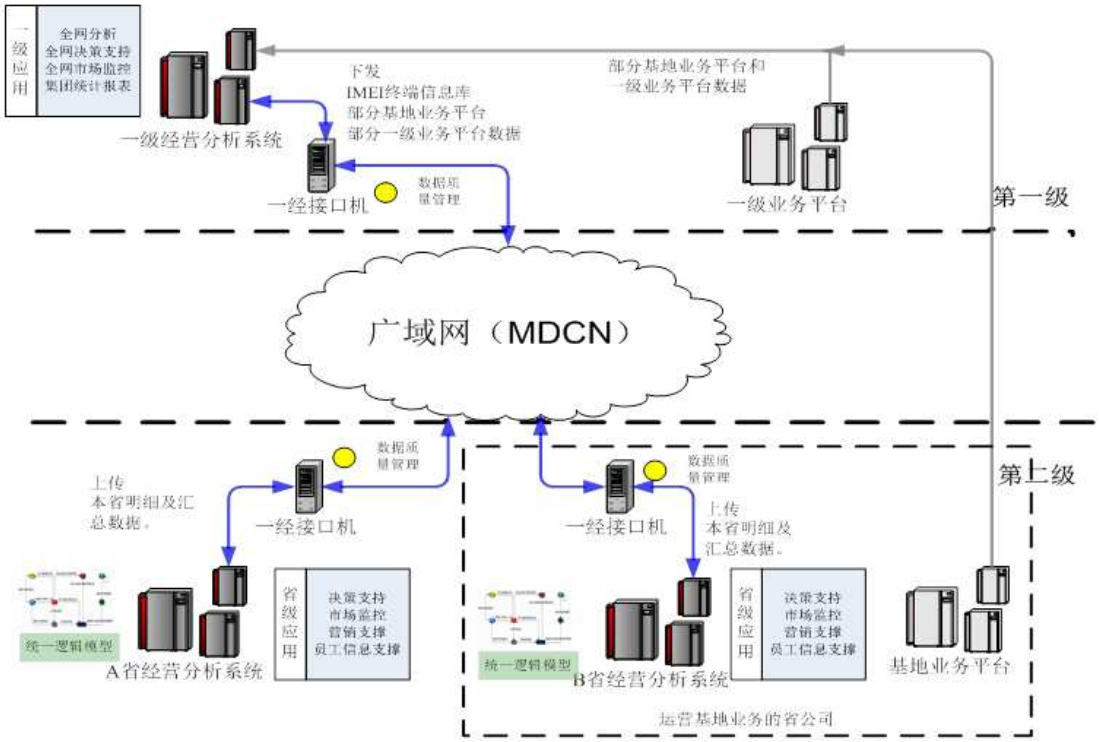


图 6 中国移动集团数据仓库平台物理架构

中国移动经营分析系统从应用架构方面，分为三个区域：

经营服务域：面对各类使用角色提供多样化和多渠道的经营分析服务，提高企业决策、运营管理和市场一线营销工作的能力；

信息服务域：通过商业智能技术手段深化企业数据的精加工服务，改善企业价值信息结构，并由互动功能为企业提供各类信息服务，促进信息的开放、重用，提升信息的价值；

数据管理域：依托元数据管理模块，完善数据质量监控应用和数据运维管理，逐步实现企业全程数据的监管、运维管理；

3、项目实施效果

经过近多年的数据仓库平台持续建设和深化应用，中国移动经营分析系统形成了涵盖九大主题域的企业逻辑数据模型，其中包括 3000 多个数据实体。在业务应用效果方面，中国移动某省公司的经营分析系统涵盖多达 6000 多个报表应用，日运营指标约 1500 个，日作业约 2000 多个，总部日考核指标 28 个，考核指标随总部市场的关注点每月进行调整。中国移动通过使用经营分析系统对海量的业务数据进行分析和挖掘，从而有效地支撑了公司管理层的企业战略决策制定、业务管理人员运营分析和业务一线人员日常运营操作，进一步提升了企业管理与运营的智能化水平。

2.3 综合评价

Teradata 作为专注于数据仓库行业 30 多年的领导厂商，推出全世界第一台数据仓库一体机，自 Gartner 开始评估数据仓库平台以来连续 14 年位居第一，Teradata 数据仓库一体机具有业内普遍认同的海量数据处理与分析能力，因此，Teradata 数据仓库一体机更适合于面向于数据分析的应用。

下表为 Teradata 数据仓库一体机各项评价指标方面的综合说明：

整体性能	整体性能优异：凭借其不受限的并行处理机制、对混合负载的高度管理能力、最优的基于成本的 SQL 解析优化器，Teradata 可以提供最好的并发查询性能。
扩展能力	Teradata 可以提供经验证的斜率为 1 的线性扩展能力
高可用性	7X24 高可靠性保证，并具备双系统保护方案
并发处理能力	并发处理能力优异：Teradata 以其非共享(share-nothing 架构),其并行处理机制的实现方式称为多维并行处理机制，包括查询并行(Query 并行)、步内并行(Within-a-Step 并行)、多步并行(Multi-Step 并行)等无条件的并行处理
系统稳定性	Teradata 提供多层次的数据保护，并通过其成熟与强大的软件体系保证了系统整体的健壮性
库内分析	数据库内分析与挖掘计算的先行者，提供功能丰富的分析与挖掘函数。
与第三方工具融合性	Teradata 支持所有主流的数据仓库/BI 工具，包括主流的 ETL 工具、报表及 OLAP 工具、数据挖掘工具、应用访问等，并提供多种语言的开发接口
易用性/易管理性	简单易用、自动化程度高、人工参与少
混合/动态负载管理	Teradata 对数据仓库的各种负载进行高效管理（复杂查询、数据加载与更新、动态事件侦测等等），通过完善的优先级调度机制，支持动态的资源调配管理，并自动进行负载均衡

多应用支持能力	通过对资源的优化调度结合强大的并发响应与处理能力，提供满足多种应用场景的数据仓库环境。
数据实时度	Teradata 是以多种措施满足实时数据处理的动态数据仓库理念的提出者和实践者。
产品技术架构	非共享 MPP 架构，保证了无条件的并行处理能力与线性的扩展能力
专业服务	Teradata 提供一整套的数据仓库专业服务，从战略、设计到实施与支持，其专业服务咨询顾问结合专利的实施方法论与经验，提供快速规划、建立有价值并使风险最小化的项目服务，并涵盖了系统的整个生命周期。
案例经验	Teradata 的客户多居于各行各业的前列。其成功建设的大型数据仓库客户包括了 80%的全球顶级商业和储蓄银行、全球十大电信运营商、70%的全球顶级零售商、100%的全球顶级航空公司和 83%的世界顶级运输/物流公司
整体拥有成本	整体拥有成本低：运维方便支持费用低、线性扩展与多代并存保护投资、实施能力与经验确保投资回报
第三方评测	连续 14 年被 Gartner Group 评为第一位的数据仓库 DBMS 平台