

Специализированный программно-аппаратный комплекс Teradata Data Warehouse Appliance

В 2009 году каждая крупная компания хранила свыше 200 терабайт данных. Этот показатель вдвое больше, чем содержало хранилище Walmart в 1999 году. На 2010 год, мировые корпорации хранили уже более 7 экзбайт (миллионов терабайт) данных, а потребители – еще 6 экзбайт. По оценкам консультантов McKinsey Global Institute, эти показатели в 52 000 раз больше, чем в библиотеке конгресса США, причем эти данные переполняют практически все секторы глобальной экономики.

При создании передовых корпоративных аналитических систем, компания IBS предлагает воспользоваться решением от компании Teradata, общепризнанного мирового лидера по разработке и внедрению высокотехнологичных корпоративных хранилищ данных и аналитики на их основе, – специализированным программно-аппаратным комплексом (СПАК) Teradata Data Warehouse Appliance (далее Teradata), ядром которого является СУБД Teradata.

СУБД Teradata 13.10 полностью параллельная СУБД, в которой все данные и вся нагрузка распределяются равномерно по узлам системы. В результате, Teradata обладает практически линейной масштабируемостью – коэффициент роста производительности к росту числа обрабатывающих узлов близок к единице. Необходимо отметить, что распределение данных и параллельная обработка не требуют настройки или сопровождения и выполняются полностью автоматически. Оптимизатор Teradata расщепляет SQL-запрос на небольшое число операций базы данных, называемых шагами, и направляет эти шаги на конвейерное выполнение. Фактически, в СУБД Teradata конвейерная обработка применяется везде, где это необходимо, кроме того, оптимизатор использует сложные правила, обеспечивающие непрерывность конвейера.

В зависимости от целей и задач бизнеса, Teradata предлагает 5 классов систем для построения хранилищ данных (ХД) предприятия (см. рис. 1).

Функциональные возможности и преимущества Teradata

■ Гибкость при изменении бизнес-задач

Массивно-параллельная архитектура «shared nothing», линейно масштабируемая по объему данных, количеству пользователей и приложений, обеспечивает стабильные и предсказуемые параметры производительности и развитие системы. Линейность объясняется тем,

что в случае двукратного увеличения нагрузки на систему (количество обрабатываемых данных/количество пользователей) достаточно вдвое увеличить мощность системы для сохранения производительности на начальном уровне.

■ Высокая надежность и доступность системы

Несколько уровней защиты от сбоя позволяют сохранить данные и продолжить бесперебойное незаметное для пользователя формирование отчетов даже при выходе из строя части узлов, дисков.

■ Расширенное управление нагрузкой (workload management)

Система позволяет гибко распределять и приоритезировать нагрузку процессов использования и обработки данных в зависимости от времени суток, ролей пользователей, количества сессий, сложности запроса, загруженности ресурсов (CPU/memory).

■ Мощные встроенные аналитические возможности

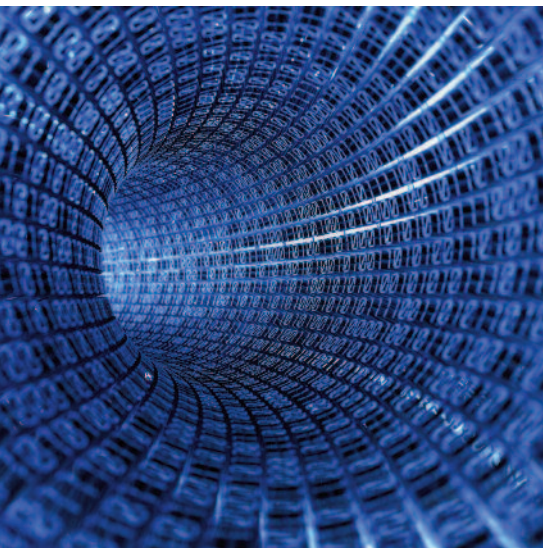
Встроенные функции статистического и периодического анализа данных, data mining, виртуальные OLAP-кубы, возможность анализа геопространственной информации и временной анализ, интегрированные пользовательские сервисы на расширяемой открытой платформе с параллельной архитектурой улучшают эффективность и дифференциацию бизнес-анализа.

■ Улучшенные возможности сжатия данных

Появление в СУБД Teradata 13.10 улучшенных алгоритмов компрессии (сжатие с целью экономии дискового пространства до 75%).

■ Простота управления

Администраторам базы данных не нужно задавать тонкие настройки, управлять табличным пространством или реорганизовывать данные, т.к. существенная часть процесса администрирования автоматизирована.



	Data Mart Appliance 5XX	Extreme Data Appliance 1XXX	Data Warehouse Appliance 2XXX	Extreme Performance Appliance 4XXX	Active Enterprise Data Warehouse 6XXX
Цель	Тестирование, разработка, небольшие витрины данных	Аналитика на огромных объемах данных, для новых типов данных	ХД или витрины данных уровня департамента	Экстремальная производительность для операционного анализа	ХД уровня предприятия, активное ХД уровня предприятия
Масштабируемость	До 12 TB	До 200 PB	До 343 TB	До 18 TB	До 114 PB
Сегмент применения	Аналитика уровня отдела, сервер начального уровня	Аналитический архив, Deep Dive Analytics	Strategic Intelligence, поддержка принятия решений, быстрое сканирование	Operational Intelligence, низкий объем, высокая производительность	Активная нагрузка, обновления в реальном времени, тактические и стратегические запросы

Рис. 1. Семейство целевых платформ Teradata для построения хранилищ данных

Инновационные особенности технологий Teradata

- **Быстрое выполнение запросов**
Архитектура «parallel everything» и средства оптимизации Teradata Optimizer обеспечивают быструю обработку запросов.
- **Параллельная обработка данных**
За счет равномерного распределения данных каждой таблицы, нагрузка распределяется между всеми виртуальными процессорами (AMP) в системе, обеспечивая максимально возможную производительность.
- **Teradata Temporal**
Учет временного изменения данных для автоматизированного управления данными и интеллектуальной обработки временных запросов. Сокращение количества ошибок данных, упрощение разработки приложений и возможность анализа данных во временном интервале, доступная для бизнес-пользователей.
- **Teradata Virtual Storage**
Представляет дисковое пространство, как единый пул по хранению данных, контролирует выделение дискового пространства для конкретных AMP, что упрощает управление дисками. Данная технология позволяет классифицировать данные по частоте доступа к ним на «холодные» и «горячие», система автоматически собирает статистику использования данных и перераспределяет данные внутри хранилища, размещая «горячие» данные на самые быстрые носители.

- **Fallback резервирование данных**
Данная опция доступна при создании таблиц, обеспечивая дублирование строк таблицы между двумя AMP, таким образом, в случае недоступности одного из AMP, данные продолжают быть доступными на втором AMP, все изменения строк таблицы автоматически сохраняются в журнал для дальнейшего восстановления произведенных изменений при восстановлении первого AMP.
- **Standby Node**
Позволяет в случае выхода из строя одного из серверных узлов (нодов) переключить обработку данных на резервный нод без падения производительности, при восстановлении нода нет необходимости перезапускать систему – восстановленный нод может далее использоваться как standby нод.

В марте 2011 Teradata приобрела компанию Aster Data Systems, также предоставляющую решения ХД, основным преимуществом которых является инновационный подход к анализу больших объемов неструктурированных и слабоструктурированных данных (Big Data). Анализ данных выполняется не на транзакционном, а на событийном уровне, который является намного более детальным. Данная модель позволяет анализировать такие данные, как клики пользователя на сайте при веб-серфинге и при наборе корзины покупок в интернет-магазине, оценивать наиболее часто используемые функции мобильного устройства (телефона/планшета), поведение клиентов в социальных сетях, анализировать машинные логи и логи приложений.

Услуги компании IBS при переходе на Teradata

Специалисты компании IBS обладают всей необходимой компетенцией и опытом для предоставления следующих услуг при разработке нового решения или переводе существующих решений на использование Teradata Data Warehouse Appliance:

- расчет экономически целесообразной, масштабируемой программной платформы, способной соответствовать текущим и будущим потребностям организации;
- проектирование модели хранилища данных по заданным отраслям, формирование благоприятной среды для загрузки и анализа данных с учетом их роста;
- своевременная разработка эффективного хранилища данных с учетом всех требований клиента;
- определение готовности к миграции, разработка методик и проведение соответствующих работ по миграции;
- доработка баз данных (в случае необходимости);
- предоставление консалтинга на любом этапе;
- предоставление технической поддержки.