

ORACLE DATABASE 11G
ДЛЯ ХРАНИЛИЩ ДАННЫХ
И БИЗНЕС-АНАЛИТИКИ






ORACLE IS THE INFORMATION COMPANY

Oracle Database 11g для хранилищ данных и бизнес-аналитики

Введение	4
Интеграция	4
Oracle Warehouse Builder	5
Основные возможности интеграции базы данных	7
Выполнение	7
Масштабирование	9
Секционирование	9
Сжатие	10
Real Application Clusters	11
Параллелизм	11
Анализ	12
Data Mining	13
OLAP	14
Заключение	15



Решение Oracle для хранилищ данных охватывает широкий спектр компонентов и функций баз данных, в том числе:

- Oracle Warehouse Builder
- Oracle Partitioning
- Oracle Real Application Clusters
- Oracle OLAP
- Oracle Data Mining
- Пакеты управления Oracle

Подробные сведения обо всех этих продуктах можно найти в дополнительных официальных публикациях и спецификациях по адресу: www.oracle.com/database

ВВЕДЕНИЕ

Oracle Database 11g представляет собой единую и функционально полную платформу для построения хранилищ данных и бизнес-аналитических систем, в которой сочетаются лучшие в отрасли масштабируемость и производительность, тесно интегрированная аналитика и встроенная интеграция и качество данных; единая платформа функционирует на основе надежной, недорогой Grid инфраструктуры. Oracle Database 11g обеспечивает непревзойденную функциональность для хранилищ данных и витрин данных с проверенной возможностью масштабируемости до сотен ТБ и лучшей на рынке производительностью. Oracle Database 11g также обеспечивает уникальную интегрированную платформу для аналитики; в результате встраивания OLAP, Data Mining и статистики непосредственно в базу данных корпорация Oracle объединяет всю функциональность автономных аналитических механизмов с масштабируемостью, безопасностью и надежностью Oracle Database. Интеграция данных — основное требование любого хранилища данных, поэтому в состав Oracle Database 11g входит лучший инструмент ETL, Oracle Warehouse Builder, в котором используются возможности Oracle по масштабируемому преобразованию данных и разнородному доступу к данным.

В этом информационном бюллетене приводится обзор возможностей Oracle Database 11g для хранилищ данных и рассматриваются важнейшие компоненты и технологии, посредством которых можно интегрировать информацию, выполнять быстрые запросы, проводить масштабирование до очень больших объемов данных и анализировать любые данные в бизнес-аналитике и банке данных на основе Oracle.

ИНТЕГРАЦИЯ

Современная информационная архитектура гораздо более динамична, чем несколько лет назад. В настоящее время компаниям требуется больше информации, они стремятся получить ее быстрее и передают более широкий набор аналитических сведений постоянно расширяющемуся кругу пользователей и приложений.

Oracle Warehouse Builder

Для удовлетворения потребностей бизнеса в состав Oracle Database 11g включен лучший инструмент интеграции данных, Oracle Warehouse Builder (OWB). Базовые функции OWB предоставляются как бесплатный компонент базы данных; дополнительно существует три варианта OWB, отвечающие различным требованиям интеграции.

- Базовый продукт. С помощью базового продукта каждый клиент Oracle сможет эффективно построить витрину данных или хранилище данных любого размера и сложности. Он предоставляет готовый к использованию на предприятии многопользовательский репозиторий метаданных, функции моделирования данных и широкое разнообразие методов преобразования и извлечения данных, а также производительность и масштабирование архитектуры «ELT».
- Опция Enterprise ETL. Эта опция специально предназначен для повышения производительности и продуктивности, она включает следующие возможности:
 - Расширенные функции загрузки данных
 - Повышение производительности труда разработчиков благодаря повторно используемым компонентам
 - Встроенное управление медленно меняющимися размерностями
 - Сквозной анализ происхождения и влияния данных
 - Поддержка расширенного управления конфигурацией
- Опция Data Quality. Полностью интегрированная часть продукта, которая обеспечивает четкое преимущество над другими имеющимися на рынке инструментами. Эта опция предоставляет поддержку для профилирования данных, правил данных (в сущности, бизнес-правил) и для проверки соответствия информации нормативным актам.
- Connectors. Обеспечивают оптимизированный доступ к ведущим рабочим приложениям. Корпорация Oracle предоставляет интерфейсы для Oracle e-Business Suite, Peoplesoft Enterprise, Siebel (CRM) и SAP R/3.

Ключевое преимущество OWB заключается в объединении в одном инструменте широких функциональных возможностей. Моделирование данных, соответствие

данных нормативным актам и качество данных — основные функции, которыми должен располагать каждый инструмент интеграции данных предприятия. Но ключевое архитектурное преимущество Oracle Warehouse Builder — интеграция компонентов. Все возможности Oracle Warehouse Builder реализованы в общем репозитории и интерфейсе пользователя.



Рисунок 1.

Благодаря всем этим возможностям, реализованным в одном инструменте на основе единого репозитория, OWB разрешает давнюю задачу интеграции данных. Многие решения интеграции предлагаются отдельными инструментами. Однако крайне неэффективно моделировать данные в одном инструменте, затем выполнять ETL отображение в другом, после чего проводить профилирование данных в третьем инструменте. OWB предоставляет один репозиторий метаданных и один интерфейс пользователя для всего процесса интеграции.

Oracle Warehouse Builder использует архитектуру «извлечение-загрузка-преобразование» (ELT). OWB не предоставляет механизм внешнего преобразования данных, а выполняет все преобразования в базе данных Oracle, используя масштабируемость и производительность платформы базы данных.

Основные возможности интеграции базы данных

База данных Oracle располагает широким набором функций извлечения, загрузки и преобразования данных. Эти функции задействованы OWB, но могут применяться в любой среде, в которой данные интегрированы в хранилище данных Oracle. К этим функциям относятся:

- Database Gateways для доступа к системам, отличным от Oracle
- Служебная программа Loader для быстрой загрузки данных из плоских файлов
- Расширения SQL преобразования данных: инструкция MERGE
- Табличные функции: эффективные параллельные пользовательские преобразования
- Сбор измененных данных для извлечения на основе журнала сведений из баз данных Oracle с малыми задержками.

ВЫПОЛНЕНИЕ

Корпорация Oracle обеспечивает оптимизацию производительности для хранилища данных любого типа. Рабочая нагрузка хранилища данных часто сложна, различные пользователи выполняют очень разные операции, столь же разнообразны ожидания и требования к производительности запросов.

Корпорация Oracle удовлетворяет требования к производительности хранилища данных, обеспечивая следующее:

- Широкий набор методов оптимизации производительности для каждого типа запросов и рабочей нагрузки:

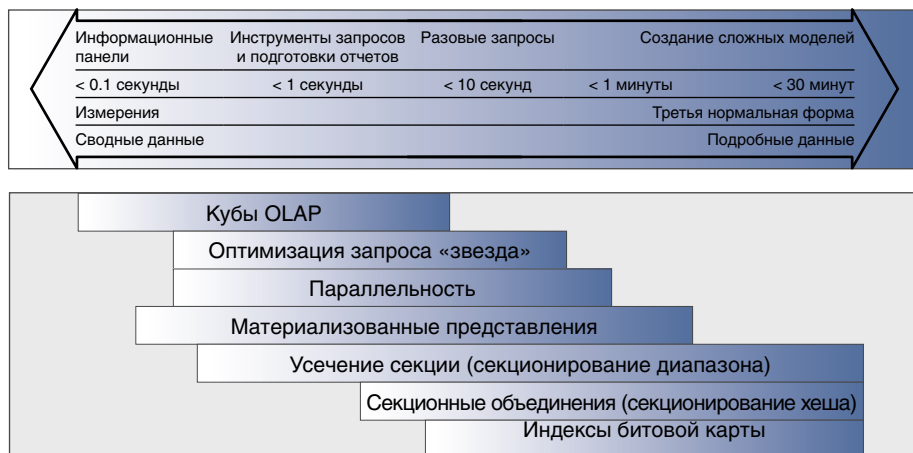




Рисунок 2.

- 
- Диспетчер ресурсов с передовыми возможностями для обеспечения производительности даже в базах данных со сложной, разнородной рабочей нагрузкой. Диспетчер ресурсов базы данных позволяет объединять конечных пользователей в «группы потребителей ресурсов», а администратор БД может точно контролировать выделение ресурсов каждой группе. Для каждой группы администратор БД может управлять показателем использования ЦП, а также назначать политики для упреждающего управления запросами и организации очереди запросов.
 - Уникальная, патентованная модель согласованности чтения, которая исключает влияние загрузки данных на производительность выполнения запроса. Корпорация Oracle решает проблемы одновременного доступа с помощью технологии, именуемой многовариантной согласованностью чтения; эта уникальная технология была основой параллельного доступа Oracle более 15 лет. Многовариантная согласованность чтения гарантирует, что пользователь всегда получает согласованное представление запрошенных данных. Если другой пользователь изменяет базовые данные во время выполнения запроса (например, непрерывное обновление большой таблицы хранилища данных), то сохраняется экземпляр данных, существовавших во время начала выполнения запроса. Данные, возвращенные запросу, всегда отражают состояние базы данных (в том числе все завершённые транзакции) на момент времени, когда поступил запрос, независимо от других обновлений, которые могли происходить во время выполнения запроса. Благодаря этой технологии корпорация Oracle располагает уникальной возможностью выполнять загрузки данных в режиме, близком к реальному времени, в среде хранилища данных.
 - Пакеты управления Management Packs для автоматизации текущей настройки производительности хранилища данных.

Одной из самых интересных новых возможностей повышения производительности в Oracle Database 11g стали материализованные представления на основе OLAP. Обычные методы оптимизации производительности для схем «звезда» в хранилищах данных издавна были связаны с использованием сводных таблиц (или материализованных представлений). При данном подходе производительность запросов повышается благодаря превари-тельному вычислению одного или нескольких суммарных



подитогов и сохранению этих подитогов в хранилище данных. Эти подитоги совершенно прозрачны для приложения, поскольку база данных обеспечивает возможность «повторной записи запроса», при которой запрос к базовому набору таблиц со схемой «звезда» повторно записывается для доступа к сводным данным. Одна из трудностей при работе с материализованными представлениями заключается в том, что одна схема «звезда» может иметь тысячи и даже миллионы сводных показателей, хранящихся в сводных таблицах. Практически невозможно подготовить материализованное представление для каждого возможного сводного показателя; затраты на использование места на диске и обслуживание могут быть чрезмерно высокими. Однако эта проблема может быть решена с использованием материализованных представлений на основе OLAP. В материализованном представлении на основе OLAP задействованы структуры данных OLAP, из которых можно быстро извлечь все возможные сводные показатели. Таким образом, схема «звезда» с материализованным представлением на основе OLAP обеспечивает производительность схемы «звезда» с тысячами реляционных материализованных представлений... но одним простым в управлении кубом OLAP с большим уровнем сжатия и возможностями эффективного управления. Такое применение технологии OLAP к рабочим нагрузкам общего назначения хранилища данных уникально для корпорации Oracle. Доступ к предшествующим решениям OLAP осуществлялся через API-интерфейс на основе OLAP, а Oracle OLAP обеспечивает выигрыш в производительности для всех пользователей хранилища данных, работающих с любым инструментом на основе SQL.


МАСШТАБИРОВАНИЕ

Размер крупнейших хранилищ данных растет экспоненциально. В настоящее время все больше бизнес-процессов полностью автоматизируются и больше данных собираются с углубленным уровнем детализации, поэтому не наблюдается признаков замедления темпов роста объемов данных.

В Oracle Database существует четыре основных способа повышения масштабируемости: секционирование, сжатие, кластеризация и параллелизм.

Секционирование

Секционирование чрезвычайно важно для управления крупными базами данных. Благодаря ему обеспечивается



метод «разделяй и властвуй» при управлении большими таблицами в базе данных, особенно если размер этих таблиц увеличивается.

В базе данных через год может храниться вдвое больше данных, чем в настоящее время, но для конечных пользователей неприемлемо, чтобы быстродействие их приложения снизилось вдвое, поэтому нельзя вдвое увеличить время, отводимое на обслуживание базы данных и пакетную обработку, а ИТ-менеджеры не удвоят бюджет на закупку оборудования для хранилища данных. Секционирование позволяет базе данных выполнять масштабирование для очень больших наборов данных без потерь производительности и неоправданного увеличения административных и аппаратных ресурсов. В результате секционирования большие таблицы разбиваются на мелкие фрагменты, таким образом окна управления данными и обработка многих запросов конечного пользователя сохраняются на прежнем уровне, несмотря на увеличение количества данных.

Корпорация Oracle работала над технологией секционирования более 10 лет, со времени выпуска Oracle8. В выпуске 11g предоставляется девять методов секционирования таблиц наряду с возможностью для администратора БД определять специальные схемы секционирования, мощный набор административных команд для секционированных таблиц, консультанта по секционированию, чтобы помочь администратору найти лучший способ секционирования.

Секционирование также позволяет применять методы ИЛМ (управления жизненным циклом информации) в базе данных Oracle. После секционирования единственную таблицу можно распределить по нескольким уровням хранения. Старые, редко используемые данные, соответствующие старым секциям, можно хранить на менее дорогостоящих устройствах памяти. В больших базах данных такой подход обеспечивает существенную экономию.

Сжатие

Со времени выпуска Oracle9i Release 2 сжатие стало одним из методов, наиболее широко применяемых клиентами хранилищ данных. Клиентам требуется сохранять все более крупные массивы данных, поэтому сжатие — естественное решение. Алгоритмы сжатия Oracle обеспечивают уникальный механизм для сжатия данных, сохраненных в реляционных таблицах практически без отрицательного влияния на производительность запросов. Учитывая, что



типичные уровни сжатия составляют от 2:1 до 5:1, широкое распространение этого метода неудивительно.

В Oracle Database 11g методы сжатия были усовершенствованы и поддерживают частые малые обновления и вставки. В предыдущих релизах данные сжимались при вставке в таблицу с использованием «прямых» операций («массовые» операции, в которых тысячи строк вставляются с использованием одной инструкции SQL или с помощью служебной программы загрузки). Новые данные не сжимались при меньших обновлениях и вставках в сжатую таблицу. Новые улучшения в релизе 11g обеспечивают преимущества сжатия для таблиц хранилища данных, которые часто обновляются непрерывным методом или иным способом.


Real Application Clusters

Компонент Real Application Clusters (RAC) обеспечивает масштабирование одной базы данных на нескольких серверах. RAC предоставляет решение масштабирования на сотни ЦП для хранилищ данных. Архитектура RAC обеспечивает уникальные преимущества для хранилищ данных Oracle. В дополнение к возможности увеличения чистой вычислительной мощности системы хранилища данных RAC обеспечивает высокую доступность при отказе узла.

RAC также обеспечивает чрезвычайную гибкость в управлении несколькими рабочими нагрузками в одной базе данных. С помощью «RAC-служб» различные приложения могут быть назначены отдельным службам, которые в свою очередь могут выполняться на отдельных узлах кластера. Например, пользователям витрины данных можно назначить один набор узлов, а операциям ETL — другой набор узлов. Ключевое преимущество RAC — возможность динамически изменять и регулировать эти рабочие нагрузки. Можно без труда динамически назначать дополнительные узлы для ETL-обработки, например, в ночное время и переназначать эти узлы для обработки запросов в течение дня.

Параллелизм

Параллелизм — возможность использовать несколько процессоров для выполнения одной операции базы данных. Масштабируемая архитектура параллельного выполнения Oracle динамически перенастраивается для




выполнения всех потребностей пользователей. Крупные операции базы данных параллельно распределяются на все доступные узлы и ЦП. Для меньших операций параллельность динамически регулируется с целью повышения общей пропускной способности системы, а диспетчер ресурсов базы данных обеспечивает выделение достаточных ресурсов для приоритетных операций.

АНАЛИЗ

Такие технологии, как OLAP, статистическая обработка и извлечения знаний, не новшество для банков данных и бизнес-аналитики (в основном все эти технологии существовали до появления банков данных). Практики, работавшие с хранилищами данных, в течение многих лет могли без труда приобрести продукты с любой из этих возможностей. Однако продукты OLAP обычно располагают собственным вычислительным механизмом, статистические продукты имеют собственный механизм данных, и продукты извлечения знаний обладают своими механизмами сбора данных. Одним словом, в среде бизнес-аналитики предприятия существует по крайней мере четыре различных типа «механизмов данных», для каждого из которых требуются собственные серверы, собственная архитектура управления, собственное администрирование безопасности, собственная инфраструктура высокой доступности. Каждый механизм располагает собственными API-интерфейсами и собственным набором инструментов разработчика и конечного пользователя. Значительны сложность и стоимость репликации всех наборов технологий бизнес-аналитики.

Oracle Database предоставляет совершенно иной подход. Корпорация Oracle реализует технологии OLAP, Data Mining и статистической обработки внутри ядра СУБД. Вместо того чтобы перемещать данные из хранилища данных в другие аналитические механизмы для дальнейшего анализа, корпорация Oracle реализовала передовые аналитические механизмы в своей базе данных, в которой хранятся данные. Кроме того, предоставляется SQL-доступ ко всем аналитическим функциям, таким образом, их можно реализовать в любом инструменте или прикладной среде на основе SQL.

Наряду с серьезными преимуществами более рациональной серверной архитектуры данных бизнес-аналитики предприятия, интеграция аналитических механизмов с Oracle Database обеспечивает множество преимуществ, недоступных в автономной среде. Например, можно ли масштабировать автономный OLAP-сервер на сотни ЦП или кластер серверов? Насколько просто интегрировать




механизм извлечения знаний с серверов проверки подлинности пользователей? И можно ли прозрачно применить все политики безопасности данных? Насколько легко можно интегрировать результаты статистического анализа с данными хранилища данных? В Oracle Database все эти проблемы решены просто благодаря глубокой интеграции OLAP, Data Mining и статистической обработки.

Data Mining

Oracle Data Mining — мощное программное обеспечение, встроенное в Oracle Database, с помощью которого можно обнаружить новые полезные сведения, скрытые в данных. С помощью Oracle Data Mining компании могут ориентироваться на своих лучших потребителей, обнаруживать и предотвращать попытки мошенничества, отыскивать наиболее важные факторы, влияющие на ключевые показатели деятельности и находить новую ценную информацию, скрытую в данных. Oracle Data Mining помогает техническим специалистам обнаруживать закономерности в данных, идентифицировать ключевые атрибуты, отыскивать новые объединения и связи, получать ценные знания.

Для решения этих бизнес-задач Oracle Data Mining позволяет компаниям отыскивать новую информацию в данных с использованием разнообразных новейших алгоритмов. Алгоритмы извлечения знаний представляют собой методы машинного обучения для анализа данных с целью обнаружения закономерностей и взаимосвязей. Корпорация Oracle предоставляет многочисленные алгоритмы, так как разные алгоритмы эффективны для различных типов анализа и разных задач бизнеса.

Большинство алгоритмов извлечения знаний можно приблизительно разделить на методы «контролируемого обучения» и «неконтролируемого обучения». Для контролируемого обучения аналитик данных должен указать целевой атрибут или зависимую переменную (например, потребителей, купивших определенный продукт). Затем алгоритм контролируемого обучения просматривает данные в поисках закономерностей и связей между другими атрибутами и целевым атрибутом (например, характеристиками, показывающими, насколько велика вероятность приобретения определенного продукта потенциальным потребителем). В Oracle Data Mining используются следующие алгоритмы контролируемого обучения: упрощенный алгоритм Байеса, дерево решений, общие линейные модели и метод опорных векторов.



Другая широкая категория алгоритмов извлечения знаний предназначена для «неконтролируемого обучения». В этом случае не существует «целевого атрибута», вместо него алгоритмы извлечения знаний применяются для поиска в данных связей и кластеров, независимых от любой, заранее определенной бизнес-цели. К этим алгоритмам относятся расширенная кластеризация по k-средним, кластеризация с ортогональным секционированием, ассоциативные правила (анализ рыночной корзины) и факторизация неотрицательной матрицы.

В программу Oracle Data Mining входит графический интерфейс пользователя, Oracle Data Miner, с помощью которого аналитики данных могут строить, оценивать и применять модели анализа данных. Oracle Data Miner проводит аналитику данных по процессу извлечения знаний, обеспечивая полную гибкость и представление результатов в графическом и табличном форматах. Oracle Data Miner может формировать код PL/SQL, связанный с действиями по анализу данных.


OLAP

Oracle OLAP — полнофункциональный механизм интерактивного анализа (OLAP), встроенный в Oracle Database. Благодаря Oracle OLAP расширяются возможности хранилища данных как вследствие повышения производительности запросов (это отмечалось в разделе производительности), так и добавления расширенного аналитического арсенала.

Основой Oracle OLAP являются кубы. Эта структура данных, управляемая внутри базы данных Oracle, хранит данные в тщательно оптимизированном многомерном формате. Кубы обеспечивают масштабируемое и сжатое хранение многомерных данных, их быстрое поэтапное обновление, высокую производительность обработки запросов и возможность вычислять или сохранять сложные аналитические расчеты.

Цель корпорации Oracle в отношении Oracle OLAP — реализовать базовые преимущества OLAP в хранилище данных. Это достигается благодаря доступу к ключевым возможностям Oracle OLAP через стандартный SQL, благодаря чему любые инструменты бизнес-аналитики или другие приложения на основе SQL могут использовать OLAP.

Главные функции интеграции SQL включают: возмож-



ность доступа к кубам OLAP через представления SQL, что раскрывает многомерный куб как набор таблиц размерностей и таблиц фактических данных, материализованные представления на основе кубов, которые позволяют использовать кубы OLAP как материализованные представления, интеграцию словаря данных, благодаря которой инструменты запросов на основе SQL и разработчики SQL могут быстро идентифицировать все кубы и их атрибуты и интеграцию путей доступа к кубу в оптимизаторе Oracle.

ЗАКЛЮЧЕНИЕ

Oracle Database — лидер на рынке хранилищ данных, предоставляемый надежность, масштабируемость и производительность и дополненный такими новаторскими компонентами, как уникальная модель Oracle согласованности чтения для хранилищ данных, функционирующих в близком к реальному времени режиме, гибкий и мощный набор функций секционирования таблиц, использование технологии OLAP для развития реляционной среды и непревзойденную поддержку grid архитектур. Oracle Database 11g также предлагает наиболее функционально полную платформу для построения хранилищ данных и бизнес-аналитических приложений, благодаря встроенной интеграции данных, OLAP и функциям извлечения знаний. Кроме того, Oracle Database 11g обеспечивает доступность, безопасность и управляемость — ключевые требования к хранилищам данных и любым системам предприятия.



Oracle Россия

119435, г. Москва,
Краснопресненская наб., 18,
Москва-Сити, бизнес-центр
«Башня на набережной» блок С
Тел.: +7 (495) 641 1400
Факс: +7 (495) 641 1414
E-mail: oracle_ru@oracle.com
Internet: www.oracle.com/ru/

191186, г. Санкт-Петербург,
Невский пр., 25
Тел.: +7 (812) 363 3257
Факс: +7 (812) 363 3258

Oracle Украина

01601, г. Киев,
Бизнес-центр «Парус»
ул. Мечникова, 2-а, 16 эт.
Тел.: + 380 (44) 490-90-50/51
Факс: + 380 (44) 490-90-52

Oracle Казахстан

480099, г. Алматы,
микрорайон Самал-2,
Самал Тауэрс, оф. 97, блок А-2,
6-й этаж
Тел.: +7 (3272) 58 47 48/40
Факс: +7 (3272) 58 47 44

Oracle Республика Беларусь

220004, г. Минск,
ул. Немига, д. 40
БЦ «Немига-Сити», к. 511
Тел.: +375 172007810/11
Факс: +375 172007817

Oracle Азербайджан

AZ1065, г. Баку,
ул. Джафар Джаббарли, 14,
БЦ «Каспиан Плаза III», 11 этаж
Тел.: +994 50 204 80 72

Copyright © 2008 Oracle Corporation. Все права защищены.

Данный документ представлен исключительно в информационных целях и его содержание может быть изменено без уведомления. Этот документ не гарантирует отсутствие ошибок и не подразумевает никаких гарантий или условий, выраженных явно или подразумеваемых законом, включая косвенные гарантии и условия окупаемости или пригодности для решения конкретной задачи. Мы отказываемся от любой ответственности, связанной с этим документом, и никакие договорные обязательства не могут быть оформлены, прямо или косвенно, на основании данного документа. Этот документ не может быть воспроизведен или передан в любой форме и любыми средствами, электронными или механическими, для любых целей, без нашего письменного разрешения.

Oracle является зарегистрированным товарным знаком корпорации Oracle и/или входящих в нее компаний. Другие наименования могут быть товарными знаками соответствующих владельцев.