HP Oracle Exadata Storage Server – оптимизированная платформа для Oracle BI-хранилищ данных

В конце сентября 2008 г. HP и Oracle анонсировали совместную разработку – HP Oracle Database Machine – специализированную платформу для хранилищ данных (data warehouse – DW) на базе Oracle 11.0.7, позволяющую от 10 до 70 раз и более повысить скорость обработки запросов в сравнении с реализациями DW на самых мощных традиционных компонентах серверов и систем хранения.

Введение

Суть DW-проблемы в том, что при определенном размере DW в диапазоне от 10 до 100 Тбайт (порог зависит от производительности системы хранения, которая поддерживает DW – традиционная NAS, массив среднего класса, High-End массив) начинает резко возрастать время реакции (в разы и даже на порядки) на запросы. Анонсированное решение позволяет поддерживать время реакции на минимальном (приемлемом) уровне в очень широких пределах масштабирования DW (сотни терабайт и более).

Тенденции в бизнесе

Почему возрастает интерес к BI-решениям – и не только у крупных компаний? Прежде всего, из-за постоянно возрастающих требований к современному бизнесу и изменения его "характера": в частности, все бóльшая его онлайновость, географическая распределенность, персонификация, зависимость от многочисленных быстро меняющихся факторов требуют одновременно и все бóльшей его интеллектуализации с точки зрения обеспечения понимания общих бизнес-процессов.

Растущая сложность бизнеса связана, во-первых, с резко возросшим потоком информации/данных, который необходимо обрабатывать/пропускать для принятия правильных и своевременных решений на всех уровнях ведения бизнеса и госуправления. Так, например, согласно WinterCorp TopTen программе, самые большие в мире хранилища данных утраиваются в размере каждые два года, начиная с 1999 г. (рис. 1). Эволюция поддержки сверхбольших хранилищ данных на базе Oracle представлена на рис. 2. Во-вторых, – со значительно возросшими требованиями к быстроте (или буквально "на лету") принимаемых решений, для чего прежде использовалась многочасовая или многодневная пакетная обработка.

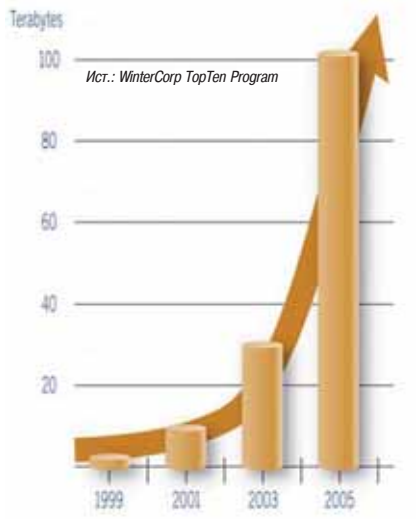

Рис. 1. Самые большие в мире хранилища данных утраиваются в размере каждые два года, начиная с 1999 г.

Другие исследования показывают, что при реализации хранилищ данных на основе традиционных систем хранения экспоненциальный рост времени ответа в зависимости от типа системы хранения может наблюдаться уже при размере DW от нескольких терабайт (рис. 3).

Эта проблема связана с тем, что существующие реализации хранилищ данных часто имеют проблему "бутылочного горлышка", ограничивающего передачу данных с дисковой системы, на которой физически хранится БД, на серверы БД. Оценки показывают, что каналы между системами хранения и серверами БД от 10 до 100 раз медленнее того, что требуется при передаче данных огромных объемов.

При реализации HP Oracle Database Machine эта проблема решалась тремя способами:

– передачей по каналам меньшего объема данных;

Рис. 2. Эволюция поддержки сверхбольших хранилищ данных.

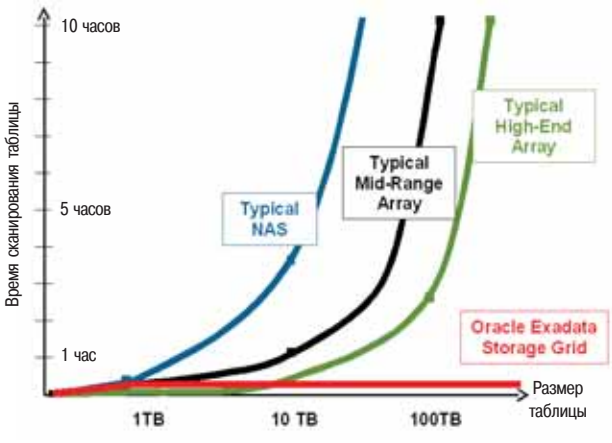

Рис. 3. Время доступа к таблице хранилища данных в зависимости от типа системы хранения, на которой оно реализовано, и размера таблицы.

– добавлением числа каналов;

– расширением полосы пропускания каналов.

Архитектура HP Oracle Exadata Storage Server в составе HP Oracle Database Machine

В конце сентября 2008 г. Oracle анонсировала 2 программно-аппаратных решения: HP Oracle Exadata Storage Server, позиционируемый как интеллектуальное устройство хранения для реляционных баз данных, и HP Oracle Database Machine – оптимизированное преинсталлированное преконфигурированное DW на базе кластера Oracle Database с Real Application Clusters (RAC) и HP Oracle Exadata Storage Server. Хотя HP Oracle Exadata Storage Server и представляется как самостоятельное решение, оно не продается отдельно и поставляется только в составе HP Oracle Database Machine. Эти решения полностью реализованы на продуктах HP, изготавливаются и собираются на заводах HP по заказу Oracle. В настоящее время эти решения продвигаются и сопровождаются Oracle или ее сертифицированными партнерами.

HP Oracle Exadata Storage Server работает только в составе HP Oracle Database Machine, начиная с версии Oracle Database 11g, Release 11.1.0.7. HP Oracle Exadata Storage Server призван заменить внешние дисковые массивы (SAN/NAS) и, самое главное, значительно повысить производительность обработки запросов в многотерабайтных BI-хранилищах.

Поставляется две версии HP Oracle Database Machine – полная и "половинчатая", соответственно, стандартная 42U стойка или ее половина.

Общая архитектура HP Oracle Database Machine представлена на рис. 4. В ней один или множество (для RAC) серверов БД соединяются через Infiniband-коммутаторы с HP Oracle Exadata Storage Server.

В полной комплектации HP Oracle Database Machine содержит:

– 8 DL360 Oracle Database серверов (2 quad-core Intel Xeon, 32GB RAM) с установленными Oracle Enterprise Linux и Oracle RAC;

– 14 Exadata Storage Cells с дисками SAS или SATA, соответственно, допуская масштабирование до 21 Тбайт и 46 Тбайт некомпрессионных пользовательских данных;

– 4 InfiniBand-коммутатора по 24 порта;

– оборудование для управления (1 Gigabit Ethernet-коммутатор, Keyboard, Video, Mouse (KVM) hardware).

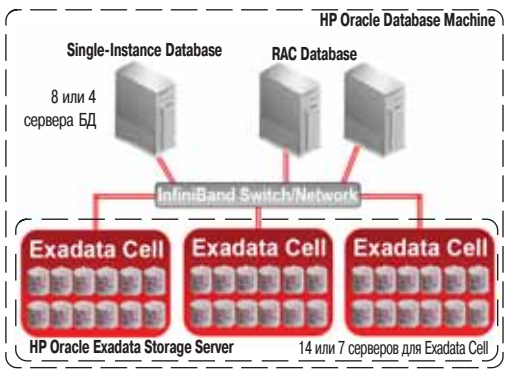

Рис. 4. Архитектура HP Oracle Database Machine.

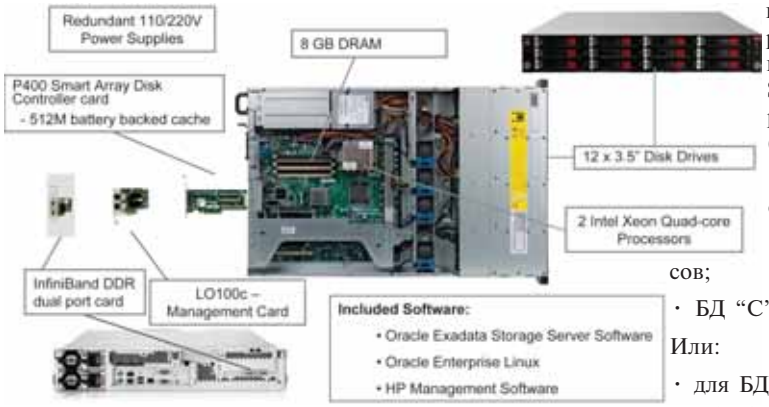

Рис. 5. Конструктивная реализация отдельной ячейки HP Exadata Storage Server.

Каждый HP Exadata Storage Server имеет потоковую производительность до 1 Гбайт/с и реализован на базе сервера HP DL180 G5 (2 Intel quad-core processors, 8 Гбайт RAM, Dual-port 4X DDR InfiniBand card, 12 SAS- или SATA-дисков). Он поставляется со следующим установленным ПО: Oracle Exadata Storage Server Software, Oracle Enterprise Linux, HP Management Software (рис. 5).

Доступность данных на ячейках Exadata поддерживается за счет зеркалирования данных с помощью ПО Automatic Storage Management (ASM) и возможности горячей замены отдельных дисков. Зеркалирование данных отдельной ячейки на множестве других гарантирует, что отказ ячейки не будет вызывать потерю данных или снижать их доступность.

Каких-либо ограничений по масштабированию (по заявлениям Oracle) HP Oracle Exadata Storage Server не существует, поэтому приобретая дополнительные стойки Exadata, можно в онлайновом режиме объединять их в общее консолидированное хранилище.

В настоящее время поставки осуществляются только в виде двух вышеуказанных базовых конфигураций.

Где целесообразно использовать HP Oracle Exadata Storage Server?

HP Exadata Storage Server ориентировано, прежде всего, на работу с приложениями, которым приходится обрабатывать таблицы БД размером от сотен мегабайт до нескольких терабайт, и где часто необходимо выполнять полное сканирование таблиц. В качестве классических примеров можно назвать BI-системы, отчетные системы и им подобные. Транзакционные системы, имеющие высокую интенсивность по чтению/записи большого количества файлов размером от нескольких десятков килобайт до нескольких десятков мегабайт, преимуществ при работе с Exadata не получат или это будет не настолько эффективно, как в первом случае.

Однако это не означает, что для разных классов приложений или групп пользователей требуется создание разных физических хранилищ данных. Exadata позволяет смешивать и приоритезировать различные нагрузки как между различными группами/классами пользователей/приложений внутри одной базы, так и между базами данных, гарантируя при этом заданный (в соответствии с SLA) уровень выделения ресурсов ввода/вывода (рис. 6). Например:

· БД "A" – 30% IO-ресурсов;

· БД "B" – 20% IO-ресурсов;

· БД "C" – 50% IO-ресурсов.

Или:

· для БД "A" – 60% IO-ресурсов для отчетов и 40% – для ETL-задач;

· для БД "B" – 30% IO-ресурсов для интерактивных задач и 70% – для пакетных задач.

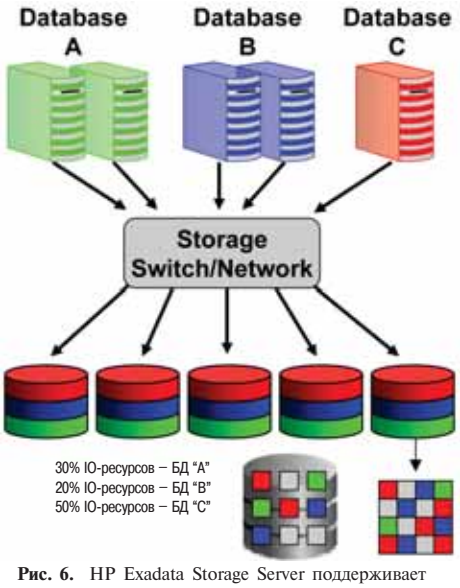

Рис. 6. HP Exadata Storage Server поддерживает гарантированное распределение IO-ресурсов как между БД, так и группами пользователей/приложений.

За счет чего достигается основное преимущество при использовании HP Exadata Storage Server?

Необходимо сразу отметить, что за счет устанавливаемого специализированного ПО каждая ячейка "понимает" структуру таблицы. Само повышение производительности при работе с таблицами большого размера происходит за счет двух факторов. *Во-первых*, при записи таблицы на HP Exadata Storage Server она равномерно "размазывается" на все ячейки – все серверы HP Exadata Storage Server. Поэтому при обработке запроса происходит его распараллеливание между всеми ячейками и, соответственно, чем и их больше, тем больше коэффициент параллелизма. Однако даже при наличии только одной ячейки эффект повышения производительности может быть многократным – это *второй* фактор. Это связано с тем, что за счет того, что сервер (ячейка) Exadata понимает структуру таблицы, при обработке запроса она не просто отсылает серверу БД часть таблицы (которая хранится на данной ячейки), а производит его обработку. Т.е. при использовании Oracle Exadata, обработка SQL перемещается с сервера базы данных к Oracle Exadata Storage Server. Oracle Exadata самостоятельно выполняет функцию отгрузки данных в дополнение к обеспечению традиционных блоковых сервисов к базе данных. Это одна из уникальных вещей, которые Exadata-хранение делает по сравнению с традиционным хранением – возвращение только строк и столбцов, которые удовлетворяют запрос к базе данных, а не полную таблицу, как обычно делалось.

Exadata выполняет SQL-запрос, максимально оптимизируя его для аппаратных средств, добиваясь максимального параллелизма работы дисков. В целом, это уменьшает загрузку центрального процессора на сервере базы данных, уменьшает требуемую полосу пропускания при перемещении данных между серверами базы данных и серверами хранения, обеспечивает балансировку нагрузки на ячейках Exadata.

Процесс фильтрации данных (в английской терминологии этот процесс еще называют Predicate Offload или Smart Scan), т.е. как именно Exadata возвращает меньше данных, чем обычный дисковый массив. Вот как описывает его более подробно в своем блоге Дмитрий Волков *(менеджер по развитию бизнеса, Oracle)*.

"Оптимизатор может использовать режим обращения Predicate Offload только, если запрос использует Direct Read Full table scan и таблица расположена на дисковой группе, которая состоит из дисков Exadata. Обычно Direct Read производится, используя Parallel Query. В Parallel Query один процесс выполняет роль координатора, другие (PQ slaves) – выполняют собственно чтение. PQ slave определяет фильтр и набор записей, который ему нужно прочитать. Если это возможно, вместо кода чтения direct path вызывается код взаимодействия с Exadata (речь идет о kernel code path). Этот код, с помощью ASM, переводит имена сегментов в диски смещения. Дальше открывается специальный поток, в котором эти данные передаются на Exadata (если несколько cell, это также легко определить с помощью метаданных ASM). Таким образом, правильный набор команд посылается нужному cell.

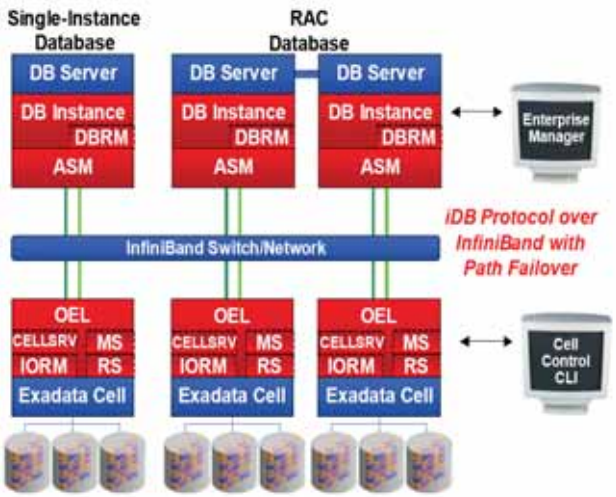

Рис. 7. Покомпонентное представление архитектуры HP Oracle Database Machine.

Процесс CELLSRV на стороне Exadata получает поток команд на чтение и необходимый фильтр. Далее с помощью библиотеки из состава ядра Oracle он выполняет необходимую фильтрацию и возвращает результат – только необходимые записи и колонки.

Ячейкам Exadata (cell) нет необходимости общаться между собой в момент выполнения запроса. Каждая ячейка получает нужную ей команду.

Поскольку речь идет о Direct Path Read, нет проблемы с read consistency. Перед началом Direct Path Read всегда выполняется tablespace checkpoint object-checkpoint, т.е. сброс грязных блоков этого объекта".

DW на основе Exadata полностью прозрачно для приложений и не требует никакой модификации SQL-инструкций. При поставках HP Oracle Database Machine каких-либо модификаций приложения в целом также не требуется. Exadata одинаково хорошо работает как с единственным образцом (single-instance) Oracle БД, так и с Oracle БД, развернутой на Real Application Clusters. Функциональные возможности и управление такими инструментами БД, как: Data Guard, Recovery Manager (RMAN), Streams и др. те же самые, как с Exadata, так и без нее.

Развернутое покомпонентное представление архитектуры HP Oracle Database Machine дано на рис. 7.

Сервер БД и Exadata Storage Server Software взаимодействуют, используя протокол iDB – Intelligent Database protocol. iDB реализован в ядре базы данных и

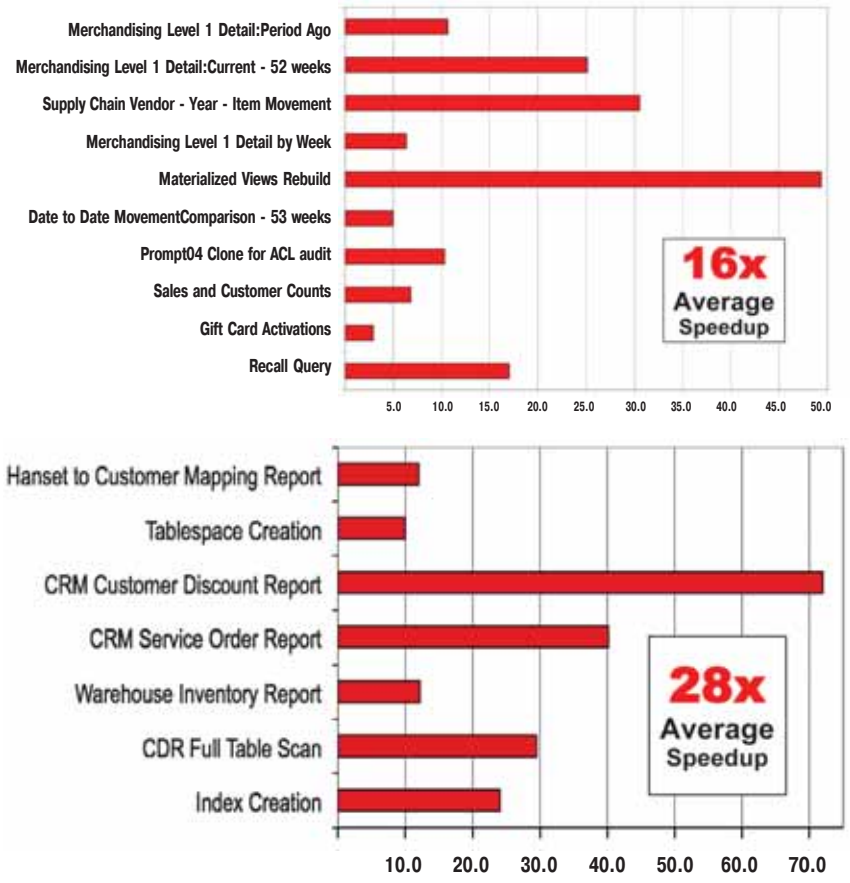

Рис. 8. Ускорение операций при обращении к DW в зависимости от типа DW (компания розничной торговли – вверху, телеком-компания – внизу) может достигать от 3–50 до 10–72 раз соответственно.

прозрачно отображает операции базы данных к расширенным операциям Exadata. iDB используется, чтобы отправить SQL-операции к Exadata-ячейкам для выполнения и возвратить результат запроса к базе данных.

iDB построен на промышленном стандарте – Reliable Datagram Sockets (RDSv3) протоколе и работает на транспорте InfiniBand. ZDP (Zero-loss Zero-copy Datagram Protocol) и zero-copy имплементации RDS используются, чтобы устранить ненужное копирование блоков.

Тестирование производительности HP Oracle Database Machine

Тестирование HP Oracle Database Machine на улучшение скорости обработки запросов в сравнении с реализациями на традиционных компонентах корпоративного класса показало улучшение времени обработки от 10 до 100 раз. Это подтверждают как ряд западных компаний, проводивших тестирование, так и независимые агентства, например, Winter Corp.

Тестирование (рис. 8) HP Oracle Database Machine в телекоммуникационной компании – M-Tel, по словам руководителя администрирования БД *(Plamen Zyumbyulev)*, показало, что "каждый запрос выполняется быстрее на Exadata по сравнению с существующими системами. Минимальное улучшение производительности было в 10 раз, а самое большое – в немыслимые 72 раза."

Другой пример – компания розничной торговли CME Group (USA, Chicago). Как заявил директор ее подразделения Enterprise Database Systems, "Oracle Exadata на сегодняшний день превосходит по результатам все, что мы тестировали ранее, от 10 до 15 раз." (см. рис. 8).

Вместо заключения

Интерес к рынку специализированных аппаратно-программных решений для DW в последние годы резко вырос. Это связано не только с возрастающими значимостью BI-систем и, соответственно, требованиями к ним по скорости обработки запросов, но и с их имплементацией и дальнейшим масштабированием.

В соответствии с проведенными несколько лет назад исследованиями, в среднем, более 50% проектов по созданию хранилищ данных терпят неудачу из-за того, что от 60% до 80% средств тратится на "чистку" и объединение данных от наследуемых систем, а управление данными и интеграция являются главными технологическими причинами, из-за которых CRM-проекты не оправдывают ожидания.

Некоторые компании, добиваясь успеха на ранних стадиях имплементации DW, не смогли его сохранить при дальнейшем масштабировании BI-проекта. Это явилось следствием значительных трудностей, возникших из-за: 1) возрастания сложности хранилища данных или недостаточной его производительности при высокой интенсивности запросов, или при высоких размерностях таблиц баз данных

***Дмитрий Семынин** – заместитель директора департамента "Центр разработки инфраструктурных решений" компании "Ай-Теко"*

Oracle Database Machine является очень интересным решением для систем анализа данных. Это не прорыв технологий, но профессиональное, продуманное, архитектурно-сбалансированное программно-аппаратное решение, направленое на устранение проблемных мест, которое обеспечивает фантастическое увеличение производительности на специфичных запросах BI-систем, отчетных систем и хранилищ данных. Решение не только открывает новые горизонты объемов информации, которыми можно оперировать в аналитических задачах. Оно также позволяет повысить утилизацию вычислительной мощности непосредственно самих дисковых систем, используя эту мощность для предварительного агрегирования информации на уровне отдельной дисковой полки.

Как видно из графиков, разброс увеличения производительности запросов по сравнению с классической архитектурой аппаратных достаточно большой, что порождает некоторые сомнения пользователей по целесообразности применения данного решения. Трудности обоснования, связанные с невозможностью точно рассчитать среднюю стоимость транзакции существующей системы, преодолеваются достаточно легко двумя способами. Можно помочь просчитать плановую эффективность на специально подготовленном программном эмуляторе. С его помощью можно просчитать ожидаемое увеличение производительности системы, в зависимости от структуры данных, сложности и частоты выполняемых запросов. Более сложный и затратный с точки зрения времени и ресурсов путь – возможность протестировать собственную BI-систему на реальном оборудовании, которое для этих целей может быть предоставлено компанией Oracle. Уникальность технологии Exadata в том, что она абсолютно прозрачна для приложений. Нужно просто запустить приложение на Exadata, чтобы сразу увидеть его поведение. Отмечу, что компания "Ай-Теко" тоже обладает компетенциями по технологиям BI-систем и Grid & Consolidation, которые так гармонично увязаны в этом программно-аппаратном решении.

хранилищ; 2) невозможности поддерживать актуальную версию хранилища.

В этом контексте появление решения HP Oracle Database Machine, позволяющего прозрачно для пользователей Oracle DW мигрировать на новые технологии с возможностью дальнейшего беспроблемного масштабирования DW имеет большое значение.

Ряд storage-вендоров также проявили интерес к сектору DW специализированных хранилищ. В частности, в декабре прошлого года EMC заявила об открытии экспертного центра по аналитическим системам и хранилищам данных, основной задачей которого станет совместная разработка специализированных прикладных решений инженерами EMC и специалистами ведущих разработчиков систем и хранилищ данных, включая Greenplum, IBM, Microsoft, Netezza, Oracle, ParAccel, Sybase, Teradata и Vertica.

Microsoft приобрела Datallegro и, как ожидается, уже в течение года представит свое MPP DBMS (massively parallel processing database management system) развертывание SQL-сервера для больших DW.

Заметно вырос интерес к BI-решениям и в России. По данным IDC, этот сектор – один из самых быстрорастущих с ежегодным ростом 30%. Одна из главных причин столь высоких темпов роста – переход крупных компаний, внедривших интегрированные системы управления предприятием, в следующую фазу автоматизации бизнес-процессов. В условиях глобализации экономики бизнес-аналитика становится необходимым инструментом поддержания конкурентоспособности компаний. Немалую роль играет и растущий интерес к BI-технологиям со стороны среднего бизнеса.

BI становится востребованным по мере того, как все большее количество менеджеров убеждается в преимуществах наличия подручных инструментов, помогающих принятию решений. BI-системы поддерживают процессы управления эффективностью компании, предоставляя оперативный анализ соотношения плановых и текущих показателей и быстро выявляя "проблемные" зоны в бизнесе.

Интересные результаты опроса, проведенного среди участников прошлогоднего IDC CEMA BI Roadshow 2007, привел в своем выступлении Робер Фариш (Robert Farish Vice President, Regional Managing Director, CIS, IDC). На вопрос, "какие функции вы планируете добавить к вашим БА инструментам в течение следующих 12 месяцев", 54% опрашиваемых ответили: "в хранение данных" – самый высокий приоритет (за ним идут: "сбор данных" – 45%, "BPM" – 37%, "панели информации" – 33%, "ETL" – 32% и др.).