



Enterprise DB как альтернатива проприетарным СУБД



Аннотация

Взрывной рост объема информации приводит к необходимости создания все более мощных и производительных баз данных. При этом лидеры рынка Oracle, IBM DB2 и Microsoft SQL Server - дорогие и закрытые СУБД. В таких условиях подходящей альтернативой является открытая СУБД корпоративного уровня EnterpriseDB, о которой пойдет речь в статье.

Обратная связь

Данный документ подготовлен Бюро Соломатина, подразделением НЦПР.

Сайт: www.bureausolomatina.ru
Электронная почта: bureau@ncpr.ru
Телефон: +7 495 988 27 09
Факс: +7 495 745 40 81
Адрес: 125375, Россия, Москва, ул. Тверская, д. 7, под. 7, этаж 2

«EDB – альтернатива проприетарным СУБД», Редакция от 12.10.2010

© ООО Национальный центр поддержки и разработки, 2010 г.

Вы имеете право воспроизводить, распространять, перерабатывать документ или использовать его иным образом в соответствии с условиями лицензии Creative Commons Attribution Share Alike 3.0 (текст лицензии доступен по адресу <http://creativecommons.org/licenses/by-sa/3.0/legalcode>).

Авторы: Пустозеров Евгений, Бюро Соломатина
Дизайнер: Акимова Оксана, Бюро Соломатина
Ответственный редактор: Черний Надежда, Бюро Соломатина

Права на товарные знаки, упомянутые в настоящем документе, принадлежат их законным правообладателям.

Введение

Увеличение объема потоков информации, глобализация коммуникаций и быстрые перемены сформировали основные черты современности — мобильность и динамичность. Ни для кого не секрет, что объем обрабатываемой информации с каждым годом увеличивается. В соответствии с отчетами IDC рост объема хранимых и обрабатываемых данных составляет более 70% в год. В среднестатистической современной компании три тысячи сотрудников ежедневно передают по электронной почте терабайт данных.

Взрывной рост объема информации приводит к необходимости создания все более мощных и производительных баз данных. Разумеется, при этом нельзя забывать о таких ключевых характеристиках СУБД как отказоустойчивость, масштабируемость, возможность оперативного мониторинга и восстановления данных после сбоев. При этом ИТ-бюджеты из года в год растут достаточно медленно. В результате через некоторое время большинство компаний сталкиваются с тем, что доступных средств перестает хватать даже на масштабирование существующих решений, не говоря уже о внедрении новых систем.

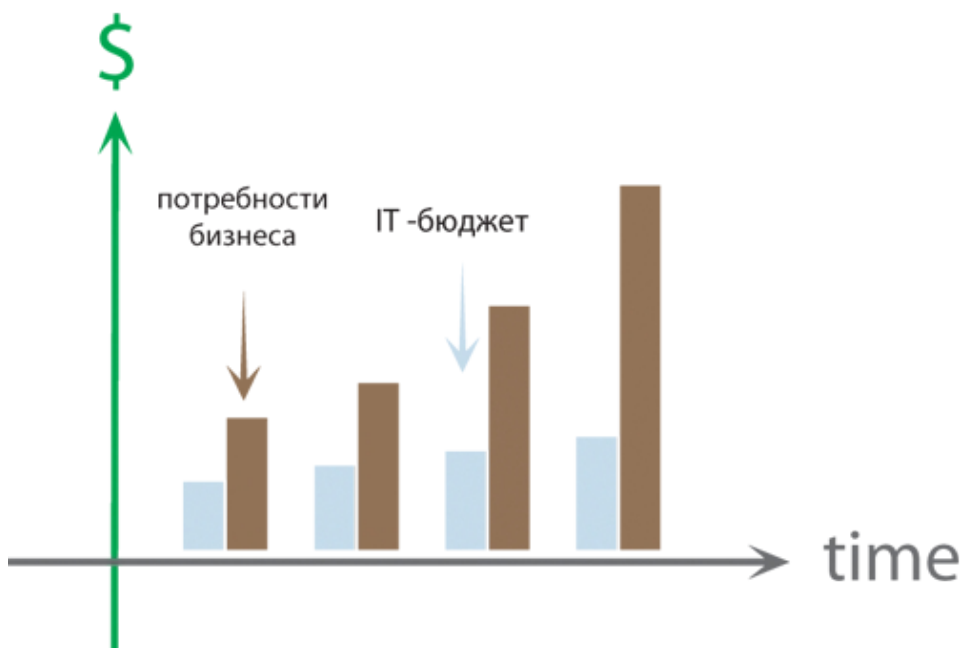


Рис. 1 Потребности бизнеса и ИТ-бюджет

Особую роль в современных условиях занимает безопасность хранения и обработки конфиденциальной информации и персональных данных, ведь защита информации очень важна для многих организаций и регламентируется законодательством.

Сегодня существует множество методов защиты хранимых данных и еще больше различных взглядов на этот вопрос. Отдельный вопрос — реализация самих механизмов



безопасности. Достаточно ли надежен конкретный способ защиты, можно ли убедиться в отсутствии вредоносного кода? Доступны ли исходные тексты системы для проверки? А если не доступны, то насколько в каждом конкретном случае можно доверять «черному ящику»? Для тех случаев, когда защита информации является критически важной, например, при защите конфиденциальной информации и персональных данных, законодательство предусматривает сертификацию механизмов защиты, в том числе проверку отсутствия недеklarированных возможностей («программных закладок»).

Одна из важнейших целей современного ИТ-подразделения компании — построение информационных систем, охватывающих все бизнес-процессы организации. Эффективное решение этой задачи требует создания высокопроизводительных СУБД, которые лежат в основе любой информационной системы. Такая СУБД должна обеспечивать защиту информации, надежность, отказоустойчивость и возможность простого масштабирования, а также обладать эффективными средствами администрирования, планирования и мониторинга.

В настоящее время, большинство компаний используют популярные СУБД — Oracle, IBM DB2 и Microsoft SQL Server. Все они имеют широкие возможности для реализации различных приложений на основе БД, в том числе очень больших масштабируемых систем высокой доступности. Имеющиеся технологии позволяют организовать распределенную отказоустойчивую систему любого типа. Основным минусом данных СУБД является их высокая стоимость. Кроме того они являются закрытыми, что не всегда подходит компаниям с высокими требованиями к безопасности и соответствию систем российским законам о защите персональных данных. Таким образом, СУБД из «большой тройки» можно использовать для решения многих задач, но при этом необходимо быть готовым к большим денежным затратам на реализацию систем и возможным проблемам, связанным с закрытостью ПО.

Альтернативой в данном случае являются открытые системы управления базами данных, из которых наиболее популярными являются EnterpriseDB и MySQL. Они имеют схожий функционал и рассчитаны на выполнение тех же задач. Технологически СУБД с открытым исходным кодом в большинстве случаев могут успешно заменить проприетарные аналоги. Кроме того, существенными преимуществами открытых СУБД являются более низкая стоимость владения, возможность гибкой адаптации для решения специфических задач, а также возможность проверки их исходных текстов на отсутствие недокументированных возможностей.

Однако стоит отметить, что СУБД MySQL является не самым подходящим вариантом для решения масштабных корпоративных задач. Это связано с наличием в MySQL ряда ограничений, связанных, например, с использованием системы хранения данных MyISAM, которая не поддерживает транзакции и внешние ключи, и тем самым затрудняет выполнение стандартных операций с БД. Кроме того MySQL с недавнего времени принадлежит компании Oracle и поэтому будущее этой СУБД весьма туманно.

В свою очередь EnterpriseDB (или Postgres Plus Advanced Server) — это СУБД корпоративного уровня на базе PostgreSQL с гарантированной поддержкой производителя, лидер среди СУБД с открытым исходным кодом. На ее основе можно строить крупные приложения, обеспечивая эффективное масштабирование и кластеризацию, а также создавать различного рода системы высокой доступности, о которых пойдет речь ниже. Также EnterpriseDB хорошо совместима с Oracle, MSSQL, MySQL, Sybase (поддержка процедур, триггеров, пакетов, типов данных, функций и т.д.),



что дает возможность достаточно быстро мигрировать существующую БД на EnterpriseDB. Кроме того в состав стандартного дистрибутива EnterpriseDB входят мощные инструменты разработки, мониторинга и управления базой данных.

Эффективность применения EnterpriseDB подтверждена многими успешными внедрениями. Вот некоторые из них:

- **Sony Online Entertainment®**

Sony Online Entertainment LLC (SOE) является признанным мировым лидером в области многопользовательских онлайн-игр с миллионами преданных геймеров по всему миру. Анализируя свой бизнес, руководители SOE пришли к выводу, что используемая ими СУБД Oracle 9i в конфигурации Real Application Cluster обходится компании слишком дорого, и что правила использования, описанные в лицензии Oracle, не допускают изменений в коде БД. В то же время, компания испытывала реальную потребность в создании и поддержке дополнительных БД. После тщательного отбора выбор руководства компании SOE пришелся на EnterpriseDB. Важным фактором при выборе послужила возможность запуска в Postgres Plus Advanced Server приложений, написанных под Oracle. В SOE отметили, что 80% приложений, отобранных для миграции с Oracle, либо заработали сразу, либо потребовали минимальных изменений для функционирования в Postgres Plus Advanced Server. К тому же персонал компании, имеющий опыт в работе с Oracle, смог сразу же приступить к своим непосредственным обязанностям в новой БД. Другим существенным фактором в пользу продукта компании EnterpriseDB стала БД PostgreSQL, лежащая в основе всех продуктов серии Postgres Plus Advanced Server. PostgreSQL гарантировала доступность и стабильность, в которых очень нуждались в компании SOE. И наконец, что тоже немаловажно, Postgres Plus Advanced Server был гораздо дешевле Oracle.

- **Florists' Transworld Delivery®**

Florists' Transworld Delivery (FTD) представляет собой одну из крупнейших в мире фирм по составлению и доставке букетов. В начале 2007 FTD начала использование новой системы администрирования транспортировки цветов, которая поддерживалась стратегически важным приложением по внешней отчетности, основанном на Oracle. В том же году, в день Св. Валентина, когда нагрузка на систему возросла в десятки раз, производственная инфраструктура на базе Oracle была не в состоянии справиться с количеством поступавших сообщений, которые занимали в среднем 90% 12-процессорной системы и ее производительность ухудшилась настолько, что отчеты пришлось готовить вручную. Первоначальным вариантом решения проблемы с отчетностью в пиковые сезоны являлось внедрение новой системы от Oracle, но это оказалось слишком дорого. Тогда FTD успешно решил эту проблему с помощью репликации (копирования) данных с Oracle на EnterpriseDB и построения отчетности на базе EnterpriseDB. Сама репликация была осуществлена посредством Postgres Plus Replication Server.

- **NTT Group**

Nippon Telegraph and Telephone, также известная как NTT — это крупнейшая телекоммуникационная компания Японии. Несколько лет назад в NTT осознали насколько выгодным и полезным является открытое ПО с точки зрения экономии расходов, повышенной надежности кода, а также гибкости в эксплуатации и внедрении. Являясь одним из крупнейших операторов, осуществляющим действительно широкомасштабные операции, компания NTT предъявляет чрезвычайно высокие требования к ПО. В апреле 2006 года компанией был основан центр Open Source Software Center (OSSC), предназначенный для удовлетворения этих требований с помощью открытого ПО. Сегодня OSSC продолжает оказывать поддержку и предоставлять техническую информацию по открытому ПО группе компаний NTT, одновременно внося свой вклад в сообщества разработчиков открытого кода. По расчетам центра OSSC, за последние два года использования EnterpriseDB, компания NTT снизила общую стоимость владения базами



данных на несколько сотен миллионов Иен (миллионов долларов США). В дальнейшем NTT собирается увеличить экономию до нескольких миллиардов Иен, удвоив масштабы использования EnterpriseDB. NTT выбирает EnterpriseDB для критически важных систем, обосновывая свой выбор первоклассными рабочими характеристиками, масштабируемостью и надежностью этой БД.

Понятия и основные признаки системы высокой доступности

При построении системы высокой доступности на базе той или иной СУБД, необходимо однозначно определиться с требованиями, которые будут предъявляться к системе. Для того, чтобы их сформулировать необходимо ответить на ряд вопросов по назначению системы.

1. Насколько доступность данных критична для работы компании? Как долго компьютерные данные могут оставаться недоступными без каких-либо последствий для компании?
2. Какой объем денежных средств компания готова выделить на приобретение и поддержку такой системы?

Ответив на эти вопросы, вы определите тип создаваемой системы:

Тип системы	Уровень доступности, в долях	Максимальное время простоя	Среднее время простоя за раз
Обычная	0,99	3,5 суток в год	1-2 часа
Высокая доступность	0,999	8,5 часов в год	Менее 1 часа
Отказоустойчивая	0,9999	1 час в год	Несколько минут
Безотказная	0,99999	5 минут в год	Несколько секунд

Следующая таблица приведена в качестве примера соответствия необходимого уровня доступности и назначения СУБД:

Уровень доступности, в долях	Пример
0,99	База данных web-сайта/портала общего характера
0,999	База данных интернет-магазина/крупного портала
0,9999	Почтовый сервер крупного предприятия
0,99999	База данных банка/мобильного оператора

После того как уровень доступности СУБД определен, можно переходить к рассмотрению конкретного вида архитектуры и соответствующему способу ее реализации.

Способы построения систем высокой доступности

EnterpriseDB предлагает различные способы организации систем высокой доступности, но любой из них основан на использовании дополнительных серверных ресурсов. При этом все варианты организации системы можно разделить на две группы:

- конфигурация с пассивным резервным сервером (active/passive)
- конфигурация с активным резервным сервером (active/active)

В свою очередь конфигурация с использованием активного резервного сервера возможна в нескольких вариантах:

- использование резервного сервера для работы сторонних приложений
- репликация сервера на резервный
- кластер с функцией балансировки нагрузки.

Пассивный резервный сервер

Основные компоненты конфигурации, представленной на Рис. 2: два сервера, общая (разделяемая) СХД и кластерное ПО. Приложения выполняются на основном сервере, но в случае его отказа автоматически запускаются на резервном. Кластерное ПО (например,

RedHat Cluster Suite) отвечает за мониторинг кластера, а также принимает меры в случае сбоя одного из серверов.



Рис. 2 Пассивный резервный сервер

Такая конфигурация обеспечивает достаточно высокую доступность системы, которая может быть увеличена за счет добавления в кластер дополнительных серверов. А развертывание и настройка такой системы довольно просты. Данная конфигурация оправдана в случаях, когда необходима простая в настройке и по содержанию система высокой доступности и при этом нет жестких требований ко времени восстановления системы в случае сбоя (от нескольких до десятков минут). Еще одним недостатком такой системы является необходимость содержать 1 (или более) незадействованных в повседневной работе серверов, что является экономически невыгодным.

Активный резервный сервер

Использование резервного сервера для работы других приложений

Основное отличие конфигурации с использованием активного резервного сервера от предыдущей в том, что ресурсы резервного сервера используются в повседневной работе, например, для работы разных экземпляров БД (Рис. 3).

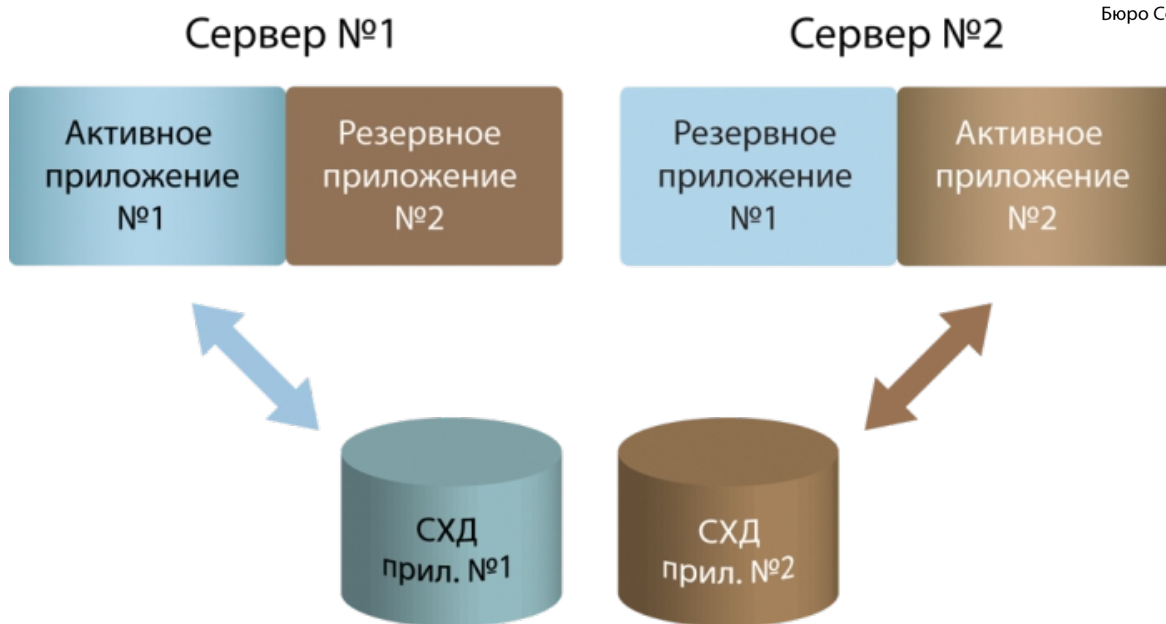


Рис. 3 Активный резервный сервер

В этом случае повышается эффективность использования серверных ресурсов. Однако, в случае сбоя, нагрузка на резервный сервер заметно возрастет, что не может не сказаться на производительности (Рис. 4).

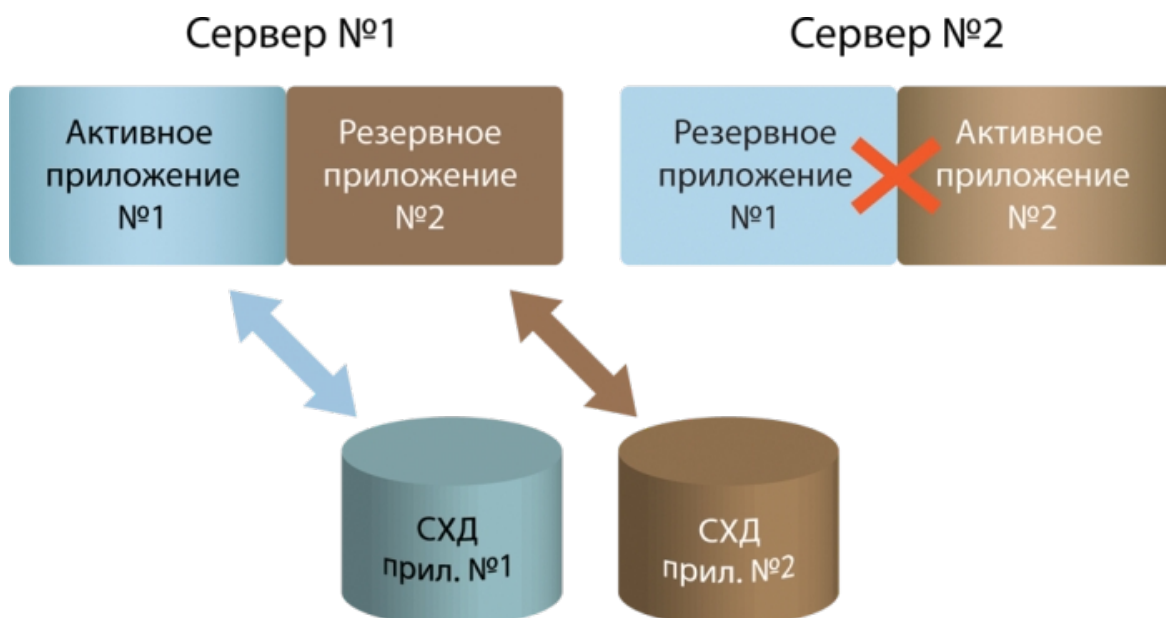


Рис. 4 Активный резервный сервер в случае сбоя

Данную конфигурацию целесообразно использовать, когда необходимо построить систему высокой доступности в условиях ограниченности серверных ресурсов и при отсутствии жестких требований к производительности системы в случае сбоя.

Репликация

Репликация — это механизм синхронизации содержимого нескольких копий БД, расположенных на разных серверах, возможно географически распределенных. Репликация является важным компонентом систем высокой доступности, поскольку она обеспечивает необходимую избыточность данных. На Рис. 5 представлены варианты создания репликаций на базе EnterpriseDB.

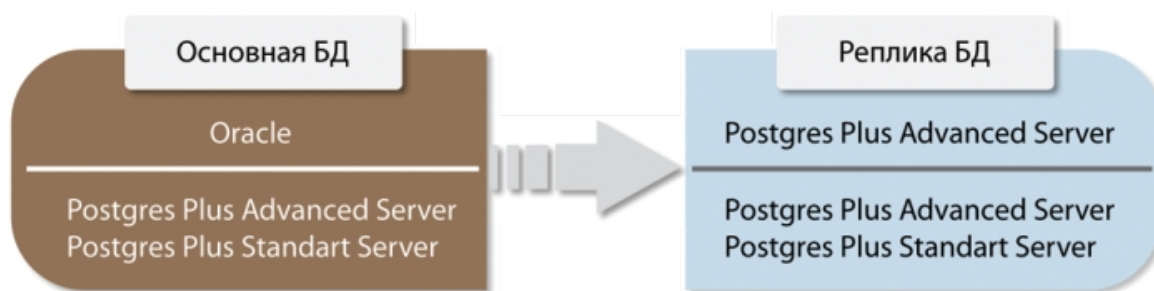


Рис. 5. Варианты репликаций

Репликация «Postgres Plus to Postgres Plus» с помощью технологии Slony

Для репликации экземпляров СУБД EnterpriseDB используется Slony —гибкая асинхронная технология репликации с открытым исходным текстом. С ее помощью можно создавать многоуровневую каскадную репликацию (Рис. 6).

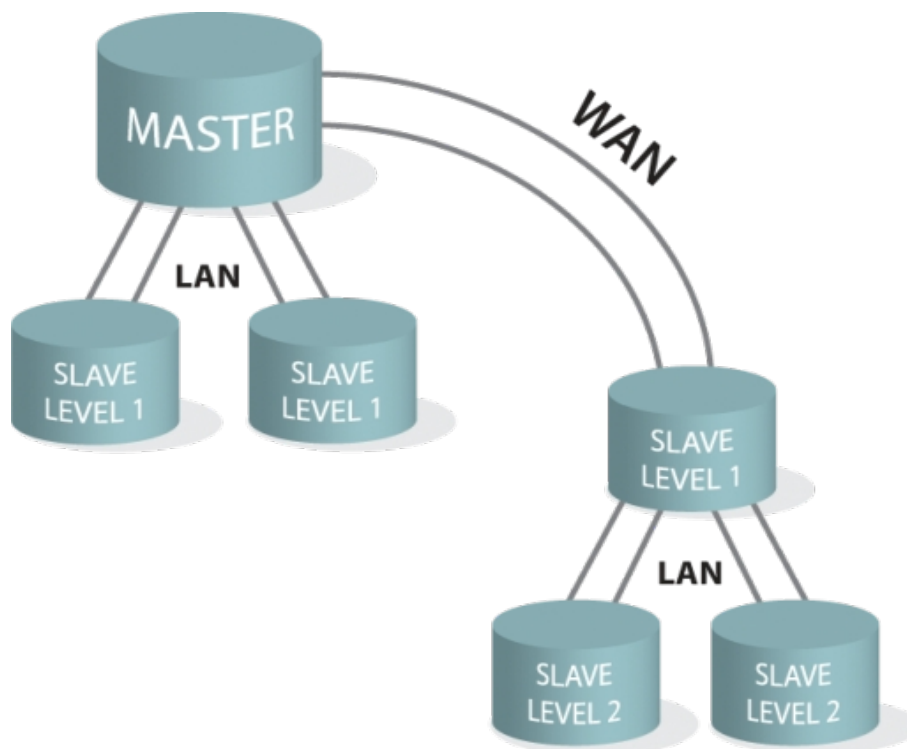


Рис. 6. Технология Slony

Благодаря гибкой архитектуре Slony умеет реплицировать как целую БД, так и отдельные наборы таблиц. А удобное переключение между активным и пассивным серверами позволяет эффективно обновлять ОС и СУБД без простоев системы.

Репликация «Oracle to Postgres Plus» с помощью Postgres Plus Replication Server

Postgres Plus Replication Server обеспечивает двухстороннюю репликацию между EnterpriseDB и Oracle по технологии схожей с Slony. Postgres Plus Replication Server можно использовать не только для создания отказоустойчивых систем, но и для эффективного перераспределения нагрузки транзакционных и отчетных приложений на разные сервера. Многие организации (например, Florists' Transworld Delivery, о которой упоминалось ранее) используют репликацию «Oracle to Postgres Plus» для значительного сокращения затрат на поддержание БД. Ее суть состоит в следующем: транзакционные системы продолжают работать на базе Oracle, в то время как системы отчетности переносятся на EnterpriseDB, тем самым уменьшая нагрузку на исходную систему и заметно сокращая затраты на содержание всей системы в целом.

Кластер с функцией балансировки нагрузки

Помимо упомянутых выше, у EnterpriseDB имеется ряд других функциональных возможностей построения систем высокой доступности. Наиболее интересная из них — это PGCluster.

PGCluster — кластерное решение с открытым исходным кодом, позволяющее организовывать репликацию БД, а также балансировку нагрузки между ними.



Рис. 7 Кластер с функцией балансировки нагрузки

Кластер может состоять из нескольких серверов, которые одновременно участвуют в работе системы и между которыми распределяется (балансируется) общая нагрузка. Данная конфигурация хороша тем, что мы получаем систему высокой доступности без простоя в случае отказа одного из серверов. Еще одним преимуществом данной схемы является возможность эффективного масштабирования системы, основанного на добавлении новых серверов в кластер.

Производительность и масштабирование EnterpriseDB

В условиях быстрого роста объемов хранимых данных требование масштабируемости для СУБД не менее важно, чем высокая доступность. EnterpriseDB включает большое количество утилит для повышения производительности СУБД и ее масштабируемости. Вот некоторые из них:

- **Infinite Cache**

Уникальная технология, позволяющая увеличивать кэш СУБД за счет подключения дополнительных серверных ресурсов. Это обеспечит хорошую масштабируемость, а также высокую производительность, особенно при выполнении запросов на чтение.

- **GridSQL**

Технология GridSQL позволяет создать кластерную СУБД, ориентированную на приложения связанные с хранилищами и витринами данных, с практически линейной масштабируемостью. Это позволяет серьезно увеличить производительность при использовании обычных недорогих серверов.

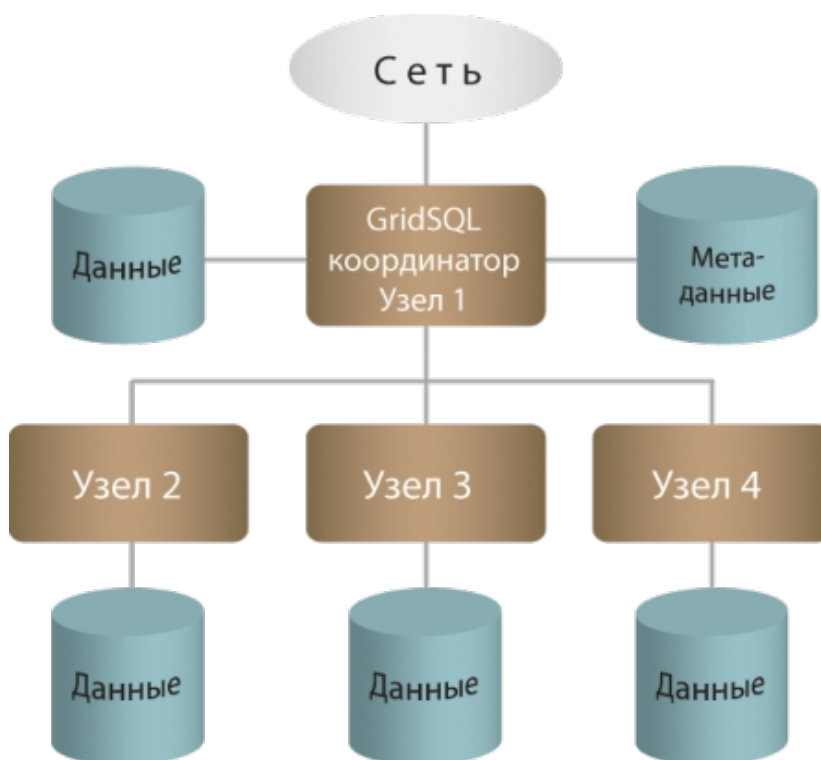


Рис. 8 GridSQL

- **Query Optimization Hints Postgres Plus**

Поддерживает более 15 хинтов (специальных ключевых слов) для оптимизации запросов при разработке приложений, что позволяет заметно улучшать производительность работы с БД.



Заключение

Во многих организациях рост объема обрабатываемой информации приводит к необходимости создания все более мощных и производительных баз данных. Разумеется, при этом нельзя забывать о таких ключевых характеристиках СУБД как отказоустойчивость, масштабируемость, возможность оперативного мониторинга и восстановления данных после сбоев. Кроме того, для многих организаций актуальной задачей является сочетание высокой безопасности и низкой стоимости СУБД.

Лидеры рынка, Oracle, IBM DB2 и Microsoft SQL Server, — довольно дорогие СУБД, и далеко не все компании могут их себе позволить. Кроме того, они являются закрытыми СУБД и поэтому не удовлетворяют требованиям российского законодательства в части защиты конфиденциальной информации и персональных данных.

Наиболее подходящей альтернативой может служить аналогичная по функционалу открытая СУБД корпоративного уровня EnterpriseDB — лидер среди СУБД с открытым исходным кодом. На ее основе можно строить крупные приложения, обеспечивая эффективное масштабирование и кластеризацию, а также создавать различного рода системы высокой доступности. Еще одним преимуществом EnterpriseDB является хорошая совместимость с Oracle, MSSQL, MySQL, Sybase (поддержка процедур, триггеров, пакетов, типов данных, функций и т.д.), что обеспечивает быструю миграцию с существующей СУБД на EnterpriseDB. Функциональность EnterpriseDB подтверждена множеством успешным внедрений по всему миру.