

Специализированный программно-аппаратный комплекс IBM Netezza

Успех любого предприятия зависит от своевременного наличия максимально точной информации, используемой для принятия решений. Постоянно растущие объемы информации предъявляют все более высокие требования к системам хранения и анализа информации. По прогнозу Gartner, до конца 2015 г. более 85% компаний, входящих в рейтинг Fortune 500, не смогут научиться эффективно пользоваться «большими данными» для того, чтобы повысить свою конкурентоспособность.

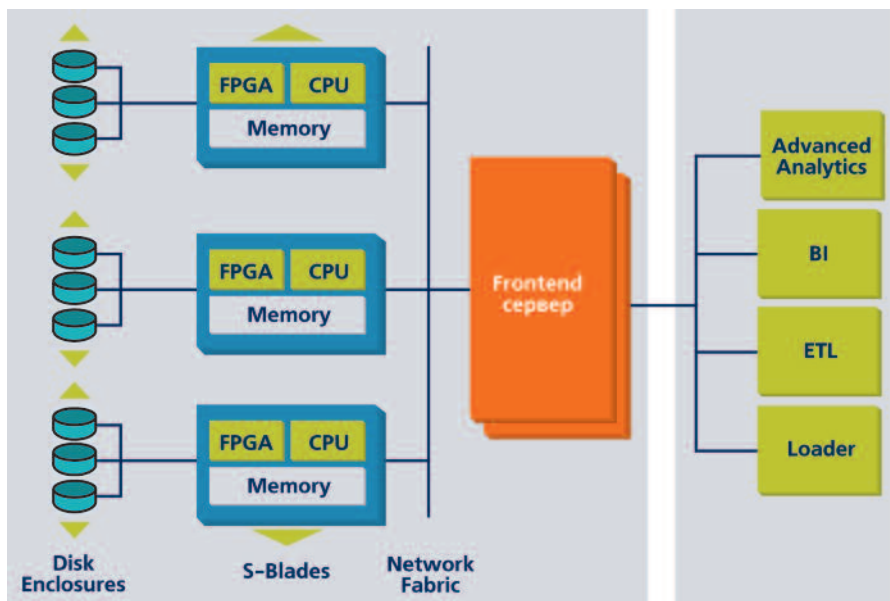


Рис. 1. Архитектурные компоненты СПАК IBM Netezza

Компания IBS предлагает воспользоваться высокотехнологичным решением от компании IBM — специализированным программно-аппаратным комплексом (СПАК) IBM Netezza Data Appliance, при создании хранилища данных для задач аналитики.

С точки зрения программного обеспечения, СПАК включает сервер БД, ОС с оптимизированными ядром и модулями, а также специализированное ПО для доступа и обработки данных и координации взаимодействия различных узлов.

Ключевым преимуществом СПАК является оптимизация стоимости как с точки зрения закупки, так и в дальнейшем обслуживании благодаря новейшим архитектурным решениям:

- вынесение емких вычислений на аппаратный уровень — программируемых логических матриц (FPGA);
- облачные интерфейсы обслуживания и управления;

- оптимизация производительности гарантируется на архитектурном уровне комплекса благодаря высокой интеллектуальности механизмов распределения нагрузки, аппаратного ускорения I/O и большинства операций.

В подобной сбалансированной системе ни один компонент не является узким местом, регулярно ограничивающим производительность. Вся работа по балансировке, отбору и настройке компонентов уже проделана производителем, помимо этого, производитель обеспечивает единую точку входа при поддержке как ПО, так и оборудования.

Узловым аппаратным компонентом IBM Netezza является управляющий Frontend сервер. Это буферная точка между пользователями и средней обработки запросов, представляющая собой набор высокопроизводительных блэйд-серверов и устройств хранения данных.



В силу простоты масштабируемости IBM Netezza предлагает очень гибкий подход к выбору конфигурации оборудования:

- Конфигурации на основе одного серверного «шкафа» (удовлетворяет потребностям небольших компаний):
 - Full Rack (12 серверов обработки запросов, хранилище данных общим объемом 32 Тб);
 - Half Rack (6 серверов обработки запросов, хранилище данных общим объемом 16 Тб);
 - Quarter Rack (3 сервера обработки запросов, хранилище данных общим объемом 8 Тб).
- Конфигурации на основе нескольких серверных «шкафов» (удовлетворяет потребностям крупных территориально распределенных компаний).

Таким образом, конфигурации имеют от 24 до 960 процессорных ядер, объем хранилища данных варьируется от 64 Тб до 10 ПТб.

Услуги компании IBS при переходе на IBM Netezza

Специалисты компании IBS обладают всей необходимой компетенцией и опытом для предоставления следующих услуг при переходе организаций на использование IBM Netezza:

- определение готовности к миграции;
- разработка методики миграции (миграция на платформу Linux, перенос БД и приложений, и т.д.);
- доработка баз данных (в случае необходимости);
- работы по миграции;
- консалтинг на любом этапе;
- техническая поддержка.

Основные характеристики IBM Netezza

- служит универсальной платформой как для OLTP, так и для аналитической обработки больших массивов данных;
- простое в развертывании решение для размещения баз данных, время ввода в эксплуатацию 1-2 дня;
- является модульным компонентом с универсальными интерфейсами, что позволяет подключать дополнительные средства сторонних производителей как в области бизнес-аналитики, так и в части загрузки данных;
- при расширении бизнеса, за счет линейной масштабируемости обес-

печена возможность без существенных затрат увеличивать как объемы хранимых данных в комплексе, так и производительную мощность;

- имеет простой встроенный механизм разработки приложений;
- обеспечивает непрерывность и надежность работы при аппаратных и программных сбоях;
- встроенные средства обеспечивают наивысшую безопасность хранимых данных.

В IBM Netezza обработка части SQL-команд производится на аппаратном уровне в FPGA матрицах, что значительно снижает нагрузку на процессоры серверов обработки и сокращает объем данных, передаваемых между серверами обработки данных. Помимо этого, использование технологии FPGA предоставляет возможность увеличения производительности комплекса только за счет обновления программного обеспечения.

Технические характеристики IBM Netezza

Архитектура IBM Netezza сочетает в себе лучшие черты технологий симметричного мультипроцессорирования (Symmetric Multiprocessing, SMP) и массово-параллельной обработки (Massive Parallel Processing, MPP) для создания комплекса, позволяющего оперировать петабайтами информации.

В IBM Netezza реализовано частичное выполнение SQL-запросов на аппаратном уровне, что уменьшает затраты на операции ввода вывода в несколько десятков раз и, тем самым, позволяет преодолеть основной потолок известных систем хранилищ данных – низкую скорость операций взаимодействия с дисками.

Компоненты системы тщательно подбираются для достижения скорости обработки информации сравнимой со скоростью считывания данных с диска. IBM Netezza показывает производительность, более чем в 10 раз, превышающую скорость работы традиционных СУБД. За секунду IBM Netezza считывает от 40 Гб (конфигурация на основе одного Full Rack «шкафа») до 400 Гб несжатых данных и выполняет от 1 млн. до 15 млн. операций ввода/вывода.

Все данные в IBM Netezza хранятся в сжатом виде для достижения максимальной эффективности использования дисковых хранилищ. Коэффициент сжатия варьируется от 4-х до 32-х. Таким образом, IBM Netezza позволяет хранить от 130 Тб (конфигурация

на основе одного Full Rack «шкафа») до 10 ПТб пользовательских данных.

Для взаимодействия внутренних компонентов и коммуникаций с внешними системами используются новейшие технологии передачи данных – шины SAS (скорость передачи: 4x3 Гбит/сек сжатых данных), протоколы Fiber Channel и 2xEthernet с поддерживаемой скоростью передачи данных от 10 Гбит/сек.

Дополнительные преимущества IBM Netezza

В IBM Netezza используется модульный подход, который позволяет легко и быстро масштабировать данное решение с нескольких гигабайт до десятков петабайт запрашиваемых пользователями данных.

Использование взаимозаменяемых компонентов позволяет модифицировать цепочку диск-процессор-память для получения конфигураций, ориентированных на производительность и/или хранение данных.

Простота настройки и масштабирования комплекса IBM Netezza позволяет запустить решение в промышленную эксплуатацию в течение 2-х рабочих дней.

Большинство мероприятий по оптимизации работы и эффективному распределению нагрузки на все имеющиеся ресурсы производится в автоматическом режиме, с помощью специально разработанных алгоритмов, поэтому, для поддержания и администрирования системы, достаточно одного специалиста. Это существенно снижает совокупную стоимость владения комплексом. Стоимость хранения данных в IBM Netezza в 5-10 раз дешевле, чем у других аналогичных решений на рынке.

Многочисленные передовые технологии, реализуемые как на аппаратном, так и на программном уровне позволяют повысить производительность и надежность комплекса. Так, с помощью технологии Zone Maps системе предоставляется информация о том, в каких частях физических дисков присутствует нужная информация, тем самым уменьшая объем просматриваемого дискового пространства, а технология Clustered Base Tables предоставляет информацию о нахождении требуемых данных в таблицах.