

# Большие Данные. Требования к системам хранения

Умение работать с Большими Данными сегодня становится важной задачей для все большего числа предприятий из самых разных сфер деятельности. При этом результат эффективного анализа больших объемов данных может выражаться и в росте оборота предприятия, и в увеличении эффективности и в оптимизации работы различных служб и бизнес-процессов. Сегодня можно уверенно сказать, что применение технологий обработки Больших Данных перешло от стадии некой модной маркетинговой идеи к вполне зрелой стадии. В то же время постепенно сформировался определенный набор требований к аппаратному обеспечению, необходимому для подобного анализа.

Аналитические задачи изначально предполагают отдельные, и зачастую довольно специфические требования для современных систем хранения, архивирования и резервного копирования. Современные задачи по обработке Больших Данных зачастую подразумевают одновременную работу как продуктивных OLTP-приложений и транзакционных задач, так и собственно аналитических задач.

Конечно, это очень эффективно и удобно с точки зрения как постановки, так и выполнения аналитической задачи – есть возможность анализировать самые последние изменения продуктивных баз данных, сопоставлять с предыдущими результатами и т.д. При этом необходимо одновременно гарантировать необходимое количество операций в секунду (IOPS) и гарантированную полосу пропускания для критически важных транзакционных задач, массив должен параллельно иметь воз-

можность обрабатывать огромные объемы данных для аналитических заданий, при этом выделяя под бизнес-аналитику необходимое количество ресурсов. А это подразумевает высокий уровень производительности массива в целом (рис. 1).

Все эти требования бизнеса начинают суммарно предъявлять довольно жесткие требования к современному хранилищу, зачастую заставляя разработчиков пересматривать всю его внутреннюю архитектуру. Например, новое поколение дисковых систем Fujitsu ETERNUS DX S3 (рис. 2) было изначально нацелено на задачи бизнес-ориентированного хранения и одновремен-

ной работы с Большими Данными. Для увеличения общей производительности пришлось использовать новейший интерфейс подключения дисковых полок SAS3, что позволило вдвое увеличить пропускную способность дискового массива. Было существенно увеличено количество кэш-памяти, добавлена возможность разбивать кэш-память дискового массива на выделенные разделы, введена возможность увеличивать кэш-память за счет различных SSD-носителей.

Использование SSD в современных хранилищах на сегодня является не просто модным трендом, но и оправданной новой техноло-

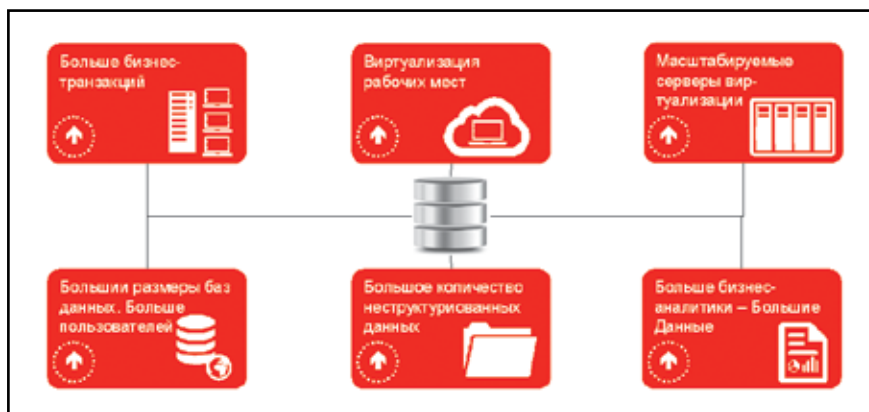


Рис. 1. Современные приложения существенно увеличивают трафик данных на хранилищах



Рис. 2. Гибкая архитектура семейства для поддержки растущих объемов данных

гией, которая зачастую позволяет существенно сократить издержки хранения огромных объемов данных. Причем наряду с созданием сильно нишевого направления массивов All-Flash-Array, изначально предназначенных для работы исключительно с SSD, все большую популярность получают гибридные схемы. Если All-Flash-Array характеризуется высокой скоростью работы и очень высокой ценой, то использование Flash-накопителей совместно с обычными SAS- и NL-SAS-дисками в рамках единой системы позволяет говорить о высокоэффективном решении для обработки больших объемов данных с необходимыми параметрами. Для оптимального решения таких задач наряду с разными типами носителей необходимо также использование технологии QoS (гарантированное обеспечение необходимого уровня сервиса) и многоуровневого хранения с возможностью автоматического переноса блоков данных между уровнями (рис. 3). Причем указанные технологии должны быть не только указаны в характеристиках того или иного хранилища, но и быть действительно эффективными, при этом оставаясь простыми в настройке и удобными в ежедневной работе.

Отдельно следует остановиться на использовании SSD-носителей для расширения количества кэша. В новом поколении дисковых систем Fujitsu добавлена возможность использования многоуровневого кэша. Наряду с большим количеством "обычного" кэша система может наращиваться отдельными блоками PCIeSSD, которые также будут использоваться для расширения кэш-памяти. Кроме того, для организации вторичного кэш-буфера можно будет использовать и обычные флеш-диски, установленные в дисковых полках. Такая система распределения кэш-ресурсов позволяет с высокой степенью granularity выделять необходимые ресурсы системы под различные задачи, что особенно важно при консолидированном хранилище и различных видах предоставления облачных ресурсов для продуктив-



Рис. 3. Пример бизнес-ориентированного использования ресурсов



Рис. 4. Возможность биллинга по факту использования ресурсов

ных баз данных и аналитики Больших Данных.

При этом незаменимой становится опция формирования отчетов о предоставленных и использованных ресурсах хранения за установленный период с максимальной детализацией – сколько ресурсов было использовано и с какой пропускной способностью. На основании этих данных провайдер облачных ресурсов хранения, будь то глобальное или локальное облако, может и выставить детализированные счета конечным пользователям (рис. 4), и точно понимать, какие ресурсы используются наибольшим спросом, и представлять, что, возможно, придется поменять через какое-то время в конфигурации хранилища. Для удовлетворения этим требованиям в новых системах Fujitsu ETERNUS были реализованы автоматизированные возможности QoS, которые обеспечивают, например, определенное администратором время отклика для различных групп приложений и пользователей. Это обеспечивает работу приложений бизнес-анализа с большими объ-

емами данных, причем анализ может происходить в онлайн-режиме. А сегодня скорость принятия решения на основании проведенного анализа зачастую является важной составляющей успеха работы всего предприятия.

Отдельная задача, о которой не стоит забывать при разговоре о Больших Данных, – это необходимость создания эффективного архивного хранилища, на которое также накладывается ряд серьезных технических требований. С одной стороны, требуется хранить действительно огромные объемы данных, с другой – необходимо делать это экономически эффективно. При этом архивное хранилище должно иметь возможность предоставить для анализа большие объемы хранимых данных в требуемый промежуток времени (с точки зрения отдела бизнес-аналитики лучше всего это осуществлять мгновенно).

Кроме всего прочего, и архив, и основные хранилища нуждаются также в современных и эффективных процедурах резервного копирования. Причем важным моментом является не только скорость и



Рис. 5. Система Fujitsu ETERNUS CS8000

время, затрачиваемое на создание резервной копии, но и гарантированная скорость восстановления данных в случае сбоя. В компании Fujitsu более 10 лет назад начато создание подобной системы, которая является сверхмасштабируемым решением как для целей резервного копирования, так и для архивирования. Работа с Большими Данными не всегда предполагает онлайн-обработку, иногда требуется глубокий ретроспективный анализ в различных срезах.

На сегодня очевидно, что для создания эффективного хранилища нужно использовать всю сумму технологий – и хранение на дисках и на лентах, и применение современных технологий дедупликации данных. При этом следует предусмотреть масштабируемость решения и возможность создания катастрофоустойчивых конфигураций и множественное хранилище отдельных объемов данных в соответствии с заданным уровнем обслуживания.

Система Fujitsu ETERNUS CS8000 (рис. 5) позволяет организовать эффективное хранение больших объемов данных, может работать как мощный архив и при этом позволяет реализовать эффективную процедуру резервного копирования. Система обладает высокой производительностью благодаря grid-архитектуре и может масштабироваться до 20 процессорных узлов. При этом без прерывания работы хранилища может увеличиваться как емкость хранения, так и производительность хранилища. Если заказчику необходимо увеличить не только емкость и производительность хранилища, но и повысить его надежность, в системе можно организовать зеркалирование данных и создание необходимого количества копий хранимых данных в соответствии с заданным SLA. При этом уникальность решения состоит в том, что при его использовании можно гарантировать не только

скорость записи в архив или скорость создания резервной копии, но и скорость восстановления или извлечения данных из архива.

Потребности бизнеса – главное, но не единственное условие успешного внедрения и использование технологий Больших Данных. Важно, чтобы современная ИТ-инфраструктура могла гибко подстраиваться под особенности бизнес-процессов. Конечно, производители стараются максимально учесть эти требования при разработке нового оборудования. Например, во всех новых системах компании Fujitsu реализованы возможности временного увеличения производительности систем в моменты возрастания нагрузки на базу данных, создаваемой аналитическими приложениями. Благодаря ряду новшеств стало возможным эффективно балансировать характеристики работы системы в зависимости от внешних условий и текущих требований бизнеса.

Сегодня опыт работы заказчиков компании Fujitsu показывает, что популярность применения технологий Больших Данных будет только расти, так как эти технологии позволяют эффективно сокращать издержки и генерировать новые возможности для бизнеса. При этом немаловажным фактором успеха является готовность хранилищ заказчика к дополнительной нагрузке и возможность интеллектуально адаптировать технические характеристики всего хранилища или его части для проведения аналитических задач.

По материалам компании Fujitsu

## НОВОСТИ

### Power BI для пользователей Office 365

Компания Microsoft Россия анонсировала доступность сервиса Power BI для бизнес-аналитики. С Power BI стало проще получать необходимую информацию из разных сред и управлять данными любого объема и структуры с помощью знакомых инструментов Office 365, в частности таблиц

Excel, которыми сотрудники пользуются ежедневно.

Сегодня многие компании сталкиваются со стихийным ростом объема данных. Но данные сами по себе не имеют ценности. Конкурентные преимущества на рынках могут обеспечить только новые знания. Их можно получить с помощью инструментов бизнес-анализа, которые легко справляются с

разными типами данных из разных источников.

Инструменты Power BI позволяют аккумулировать любые типы данных из гетерогенных сред, будь то информация, хранящаяся в частном, публичном облаке или в облаке провайдера. Сотрудники могут делиться отчетами со своими коллегами, совместно работать с данными и более эффективно управлять рабо-

чими процессами в режиме реального времени на любых мобильных устройствах. Power BI объединяет необходимый для усовершенствования бизнес-аналитики функционал: публичный и корпоративный каталог данных, новейшие инструменты поиска информации, работу с приложениями, богатую интерактивную визуализацию, возможности для совместной работы.



# INNOPROM

## ИНТЕЛЛЕКТУАЛЬНАЯ ПРОМЫШЛЕННОСТЬ

9 —12 Июля 2014

- Более **570 компаний-экспонентов**
- Бизнес-делегации из **70 стран** мира
- Более **53 000 посетителей**, в том числе владельцы и топ менеджеры крупнейших российских и международных компаний
- **50 000 кв м** высокотехнологичного выставочного пространства МВЦ «Екатеринбург ЭКСПО»

Деловая программа **INNOPROM**:

- Более **150 мероприятий**
- **Актуальные** темы
- **Глобальные** лидеры
- **Деловая элита** России, Европы, Азии, США, Латинской Америки и Африки

ТЕЛЕФОН ГОРЯЧЕЙ ЛИНИИ

**8-800-700-82-31**

(звонок по России бесплатно)

[WWW.INNOPROM.COM](http://WWW.INNOPROM.COM)

Организаторы:



ПРАВИТЕЛЬСТВО  
СВЕРДЛОВСКОЙ  
ОБЛАСТИ

Оператор:

**FORMIKA**

При поддержке:



**СОСПП**

Свердловский областной  
Союз промышленников  
и предпринимателей