

# Большие данные — это не проблема, а набор задач и решений, доступных уже сегодня

*Создание технологий работы с большими данными (Big Data), включая большие геопространственные данные (Big Geo Data), — важнейшая задача современной ИТ-отрасли из-за постоянного колоссального увеличения потока информации, требующей обработки и анализа. Одним из лидеров рынка информационных и коммуникационных технологий, активно работающих в области разработки систем хранения данных, является компания Fujitsu, наработки и продукция которой, на наш взгляд, представляют большой интерес для геоинформационной отрасли и сферы космического мониторинга. Мы попросили рассказать о компании и ее решениях менеджера по развитию бизнеса СХД Александра Яковлева.*



## А. Яковлев. Блицпортрет

А. Яковлев работает в сфере информационных технологий более 12 лет. В Fujitsu он пришел в 2005 г. на должность менеджера по маркетингу продукции, отвечал за продвижение систем хранения и резервного копирования, критически важных серверов и программно-аппаратных комплексов. До прихода в Fujitsu работал в компании HP, где занимался технической поддержкой, инсталляцией и настройкой UNIX-серверов и систем хранения данных. А. Яковлев — выпускник МИФИ, который он окончил в 1999 г. по специальности «Автоматика и электроника физических установок». В 1999–2002 гг. обучался в аспирантуре МИФИ.

**Редакция:** Расскажите, пожалуйста, о компании Fujitsu, основных направлениях ее деятельности. Каковы особенности ее работы в России?

**А. Яковлев:** Fujitsu — это японская компания, лидер рынка информационных и коммуникационных технологий (ИКТ), предлагающая полный спектр технологических продуктов, решений и услуг. В активе компании большой коллектив

экспертов мирового уровня в области систем и сервисов, высоконадежная ИТ и телекоммуникационная продукция, а также передовая микроэлектроника, обеспечивающие заказчикам реальную выгоду. Fujitsu поставляет ИТ-решения более чем в 70 стран мира, является четвертым по величине поставщиком ИТ-услуг в мире, а также одним из пяти крупнейших мировых поставщиков серверов.

Подразделение компании в России существует с 2000 г. Сегодня представители Fujitsu работают не только в Москве, но и в других крупнейших городах России, а также на территории стран СНГ. Мы способны удовлетворить запросы различных заказчиков: государственных структур, крупных корпораций, малого и среднего бизнеса. Стратегическим направлением деятельности компании является разработка динамических инфраструктур и их внедрение в масштабе всего предприятия. Fujitsu предлагает широкий спектр ИТ-продуктов, решений и сервисов — от рабочих мест до решений для Центров обработки данных (ЦОД) и услуг по управлению ИТ-инфраструктурами, а также инфраструктуру как услугу (IaaS).

Благодаря высокой квалификации своих специалистов и многолетнему опыту в области разработки комплексных ИТ-решений компания Fujitsu выполнила ряд крупных проектов на территории России. Так, ИТ-инфраструктура известной торговой сети «Юлмарт» практически полностью построена на основе серверных систем Fujitsu, включая оборудование для функционирования ERP-системы «Юлмарт», работающей на базе продуктов SAP. А для Северного Арктического федерального университета имени М. В. Ломоносова (САФУ) специалисты Fujitsu создали суперкомпьютер с пиковой производительностью вычислительных узлов в 11,68 ТФлопс.

**Р.:** Fujitsu в России — один из ведущих поставщиков инфраструктурных решений. Что Вы можете сказать о проблеме больших данных (Big Data), возникающей при реализации проектов? Какие технологии Вы применяете при создании центров обработки данных?

**А. Я.:** Компании уже достигают конкретных бизнес-результатов на основании анализа больших данных, принимая оптимальные

решения, которые позволяют увеличивать обороты и прибыльность. Поэтому большие данные — это, скорее, не проблема, а набор задач и решений, доступных уже сегодня. Требования к Системам хранения данных (СХД) предъявляются повышенные, причем не только и не столько к их размеру, сколько к эффективности, возможности динамически перераспределять ресурсы, к их пропускной способности, которая должна быть, без преувеличения, огромной, в частности при анализе данных продуктивных систем. С одной стороны, есть жесткие требования по скорости доступа к ним со стороны, например, ERP-системы, в плане гарантированного времени отклика и достаточности полосы пропускания продуктивных систем. С другой стороны, к тем же данным (или смежным с ними, лежащим в том же хранилище) должны иметь доступ и аналитические системы, которые также могут создавать серьезную нагрузку на базу данных. То есть возникает задача разграничения доступа к одним и тем же данным. И здесь у компании Fujitsu имеются очень серьезные наработки как в сегменте специализированных хранилищ, так и для обычных дисковых СХД.

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

Система Fujitsu ETERNUS CS8000 позволяет организовать эффективное хранение больших объемов данных, работать как мощный архив и при этом 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 не только увеличить 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 соглашением об уровне предоставления услуги (Service Level Agreement — SLA). При этом уникальность решения состоит в том, что при его использовании можно не только гарантировать скорость записи в архив или скорость создания резервной копии, но и гарантировать скорость восстановления или извлечения данных из архива.

Потребности бизнеса — главное, но не единственное условие успешного внедрения и использования 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 наших новых системах были 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.

**Р.: Какие сервисы на основе облачных вычислений предлагает ваша компания**

**для поддержки развития бизнес-проектов в России?**

**А. Я.:** Мы предлагаем облачные хранилища и реализацию резервного копирования и/или архивирования из облака. Причем можно получать эти сервисы из нашего облака, из облака наших партнеров либо создать частное облако на своей площадке. Все необходимые механизмы уже заложены в наших решениях

**Р.: Можете ли вы назвать главные отрасли, в наибольшей степени заинтересованные во внедрении технологий больших данных?**

**А. Я.:** Самый простой и наиболее частый пример — онлайн-ритейлер. С одной стороны, налицо очень жесткие требования по времени отклика на запрос покупателя через сайт. С другой стороны известно, что средства аналитики в области розничной торговли способны увеличить на десятки процентов посещаемость сайта, обороты и прибыль компании. Возникает задача разграничения уровней доступа и возможностей, предоставляемых различным группам пользователей. Решение данной задачи осложняется нелинейностью продуктивной нагрузки — могут возникать как прогнозируемые (перед праздниками), так и непрогнозируемые пики продаж, возможно снижение покупательской активности. Соответственно в одной ситуации реализуется осознанное уменьшение полосы пропускания, количества операций в секунду для аналитики с целью гарантировать минимальное время доступа для продуктивной системы, в другой — полоса пропускания для получения аналитики может быть увеличена. У ритейла уже есть понимание, что маркетинговые программы необходимо реализовать на основе анализа больших данных, только тогда они приносят ощутимый для бизнеса позитивный результат. Для банковской сферы использование СХД и для

продуктивных систем, и для аналитики также актуально. Только там все сложнее, поскольку добавляются серьезные требования по обеспечению уровня надежности СХД выше 0.99999, гарантирующей непрерывность ИТ-поддержки бизнеса. Любые сбои в оказании банковского онлайн-сервиса, даже если они чисто технического характера, могут вызвать панические настроения среди клиентов и иметь крайне неприятные последствия для банка.

**Р.:** Приходится ли Вам работать с пространственными данными? Какие технологии хранения и обработки таких данных Вы предлагаете?

**А. Я.:** Да, у нас есть проекты по хранению пространственных данных. С точки зрения аппаратного обеспечения используются те же хранилища, разница лишь в требуемом уровне SLA, производительности и, собственно, ПО для обработки.

**Р.:** В последнее время значительно увеличился поток данных, получаемых в результате космического мониторинга. Зачастую большой объем информации требует оперативной обработки и анализа (так называемые Fast Data). Есть ли в вашем арсенале предложения по решению аналогичных проблем?

**А. Я.:** В наших системах хранения существует целый комплекс возможностей и механизмов, позволяющих гарантировать заданное время отклика для тех или иных наборов данных и логических томов. Причем средства управления хранилищами позволяют динамически изменять эти требования, что позволяет менять уровень приоритета, скорость работы и время отклика для тех или иных наборов данных без остановки функционирования приложений. Для удобства администрирования разработаны интуитивно понятные средства управления.

**Р.:** Назовите, пожалуйста, наиболее значимые проекты, связанные с хранением и обработкой больших данных, реализованные Вашей компанией.

**А. Я.:** Это большое количество крупных проектов в ритейле и торговых сетях, ряд проектов в государственном секторе и финансовых организациях, сотрудничество с рядом компаний, работающих на рынке геоинформационных и космических услуг. В качестве примера успешно реализованного проекта, связанного с хранением и обработкой пространственных данных можно привести сотрудничество с ОАО «Сибнефтегеофизика», многопрофильным геофизическим предприятием, выполняющее все виды полевых сейсморазведочных работ 2D, 3D, промыслово-геофизические исследования скважин, их обработку и геолого-геофизическую интерпретацию, мониторинг изученности территорий сейсморазведкой и бурением. Этой компании был необходим надежный вычислительный кластер для обработки информации, получаемой в ходе геофизических исследований. Проблема состояла в том, что при проведении обработки и интерпретации (моделировании структуры земли) требуются вычислительные ресурсы большой мощности. Компания смогла предложить серверные решения, сочетающие в себе приемлемую стоимость закупки и сервисного обслуживания с очень высоким уровнем производительности, надежности и масштабируемости. Для модернизации вычислительной системы была выбрана интегрированная система Fujitsu PRIMEFLEX, состоящая из системы хранения данных и серверов.

**Р.:** Большое спасибо, Александр, за подробный рассказ на актуальную тему, представляющую большой интерес для наших читателей.