

# PCWEEK RUSSIAN EDITION REVIEW

## БОЛЬШИЕ ДАННЫЕ

МАРТ • 2012 • МОСКВА

<http://www.pcweek.ru>



## Не пора ли вплотную заняться “большими данными”?

АНДРЕЙ КОЛЕСОВ

Разговоры о постоянном геометрическом росте объемов хранимых данных ведутся с середины прошлого века. По данным аналитиков, в 2010 г. суммарные цифровые активы в мире перевалили за отметку в 1 зеттабайт (зетта — 1000 в седьмой степени, единица, после которой идет 21 ноль). На темпах этого роста даже почти не сказались мировой экономический кризис: четыре года назад IDC предсказывала, что в 2011 г. объемы данных вырастут до 1,8 зеттабайт (в 10 раз по сравнению с 2006-м), и, кажется, этот прогноз оказался даже меньше реальных показателей.

Объемы растут постоянно, но только в последние один-два года на ИТ-рынке стала широко обсуждаться концепция Big Data (большие данные), появление которой связано с пониманием необходимости некоторых качественных изменений в подходах к хранению и использованию растущих объемов информации. Когда традиционный вариант прямолинейного увеличения ресурсов и мощностей уже не работает.

Проблема заключается еще и в том, что, несмотря на постоянное снижение удельной стоимости хранения данных, заказчики отмечают непрерывный рост затрат на хранение. Аналитики отмечают, что взрывной рост объема информации не является результатом увеличения числа деловых операций и, вполне вероятно, объясняется неуправляемыми процессами репликации данных. Даже поставщики устройств хранения все чаще говорят о том, что круг задач управления информацией сейчас быстро смещается от вопросов физического хранения данных к их использованию, что хранение данных — это не самоцель, а средство для того, чтобы ими можно было воспользоваться в нужный момент. При этом тема Big Data напрямую связана с другой, уже давно обсуждаемой глобальной ИТ-тенденцией — с переходом к широкому внедрению облачных вычислений.

Чтобы разобраться с кругом вопросов темы Big Data и понять актуальность этого направления, в том числе для российского ИТ-рынка, мы обратились к ряду экспертов из отечественных и зарубежных ИТ-компаний.

### Big Data — новые идеи или новое название известных вещей?

Очень часто действительно новые, важные и интереснейшие прорывы и открытия в технике лежат где-то на стыке различных методик, дисциплин, технологий и представлений о, казалось бы, довольно хорошо изученных вещах. Высказав этот тезис, менеджер по маркетингу продукции представительства Fujitsu в России и СНГ Александр Яковлев отметил, что нечто похожее происходит сегодня и с большими данными: “Новая концепция, а скорее новые возможности для работы появились на стыке облачных технологий хранения, новых функций

аналитических приложений и собственно самих больших данных”. При этом он подчеркнул, что работа с большими данными невозможна без облачных хранилищ и облачных вычислений — появление облачных технологий не только в виде идеи, а уже в виде законченных и реализованных проектов стало отправной точкой для нового витка интереса к концепции обработки больших данных, о которой впервые заговорили около трех лет назад. Отсюда делается вывод: “Сегодня это скорее не маркетинговая переменная названия давно существующей технологии, а выход технологий на новую ступень, которая позволяет совершить уже качественные изменения”.

Руководитель направления платформенных решений “SAS Россия/СНГ” Алексей Мещеряков сказал о том, что хотя Big Data — это уже популярный се-

**Решая задачи, связанные с Big Data, компании получают весомые конкурентные преимущества в виде доступности обоснованных результатов аналитических исследований, проведенных на громадных объемах структурированных и неструктурированных данных.**

годня термин, но все же он еще не очень устоялся по смыслу. Сейчас он обычно используется для обозначения потребности организации в загрузке, хранении и оперативном анализе сверхбольших объемов экспоненциально растущей структурированной и неструктурированной информации, лежащей за пределами традиционных вычислительных возможностей. Представитель SAS выделил в понятии Big Data три основные характеристики:

- объем — стремительно растущие размеры массивов данных;
- многообразие — все виды хранения информации: текст, структурированные таблицы, документы, электронная почта, SMS, данные измерительных приборов, видео, аудио, данные биржевых торгов и многое другое;
- быстродействие — время выдачи информации по запросу, способность оперативно реагировать на задачи и проблемы компании.

В целом данная проблематика напрямую связана с направлением высокопроизводительных вычислений (High Performance Computing, HPC) при выполнении аналитических исследований, что, в свою очередь, включает такие вопросы, как grid computing (распределенные вычисления на нескольких серверах), in-database analytics (частичный перевод нагрузки при аналитических вычислениях в СУБД,

а также регламентное применение готовых аналитических моделей к новым данным полностью на стороне СУБД) и in-memory analytics (применение аналитики прямо в оперативной памяти сервера СУБД).

По мнению Алексея Мещерякова, решая задачи, связанные с Big Data, компании получают весомые конкурентные преимущества в виде доступности обоснованных результатов аналитических исследований, проведенных на громадных объемах структурированных и неструктурированных данных, причем в условиях жестких ограничений по времени, вплоть до режима реального времени. Такой подход меняют саму парадигму ведения бизнеса, позволяя руководителям применять ранее недоступные методы и возможности для управления компанией. Развитие технологий для решения задачи Big Data может стать основой для развития конкурентоспособности, инноваций и роста в бизнесе.

А вот менеджер по маркетингу продуктов SQL Server и BizTalk Server российского представительства Microsoft Вячеслав Архаров считает, что в настоящее время четкого и однозначного определения понятия Big Data нет: “Под Big Data, как правило, понимается набор задач, связанных с обработкой большого количества неструктурированных данных за нерегламентированное время, но с гарантированным результатом. В этом заключается принципиальное отличие Big Data от традиционных СУБД и хранилищ данных, так как последние предполагают хранение информации в четко структурированном виде и гарантируют время выполнения операций. Под неструктурированными данными имеется в виду такая информация, как логи файлов, изображения, видео, информация, полученная с устройств, и т. п. Большинство таких данных никогда не будут храниться в традиционных СУБД, и в прошлом они достаточно часто удалялись за ненадобностью. Однако сейчас стало экономически выгодно не только хранить подобную информацию, но также обрабатывать и анализировать ее”.

С этой трактовкой в целом согласен ведущий технический эксперт Sybase CIS Андрей Хромов: “Под большими данными подразумеваются растущие объемы структурированных и неструктурированных данных, поступающих из разных источников. Причем работа с этими данными, сложными по структуре и многообразными по типам, не может выполняться традиционными системами”. В то же время он отмечает, что большие данные — не такое уж и новое явление. Еще в 50-х годах прошлого века объем данных в организациях уже становился проблемой — рост данных опережал развитие технологий хранения информации. Тогда IBM представила первый жесткий магнитный диск для хранения информации вне компьютера. После этого возникали новые проблемы, но тут же предлагались и но-

ПРОДОЛЖЕНИЕ НА С. 18 ►

### Наши эксперты



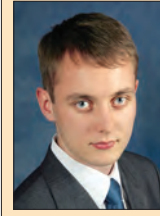
**СЕРГЕЙ АРТЁМОВ**, эксперт группы перспективных технологий, “Инфосистемы Джет”



**ВЯЧЕСЛАВ АРХАРОВ**, менеджер по маркетингу продуктов SQL Server и BizTalk Server, “Microsoft в России”



**АНДРЕЙ ВАКОВ**, заместитель директора департамента “Инжиниринговый центр” компании “Техносерв”



**МИХАИЛ ГОРБАТОВСКИЙ**, руководитель по развитию бизнеса сектора финансовых услуг, “IBM в России и СНГ”



**МАКСИМ ИСАЕВ**, менеджер по развитию бизнеса департамента системных решений, IBS



**АЛЕКСЕЙ МЕЩЕРЯКОВ**, руководитель направления платформенных решений, “SAS Россия/СНГ”



**АНДРЕЙ ХРОМОВ**, ведущий технический эксперт, Sybase CIS.



**ВАЛЕРИЙ ЮРИНСКИЙ**, директор отделения технологического консалтинга, ФОРС



**АЛЕКСАНДР ЯКОВЛЕВ**, менеджер по маркетингу продукции, представительство Fujitsu в России и СНГ



## Не пора ли...

◀ ПРОДОЛЖЕНИЕ СО С. 17

вые решения. Например, целую революцию за последние 10 лет произвели платформы массивно-параллельной обработки и БД с хранением данных по столбцам — эти технологии позволили организациям проводить качественно новый анализ больших данных.

Что касается высказываемых в последнее время прогнозов, что большие данные могут подорвать производительность аналитического хранилища данных и вызывать проблемы с масштабируемостью, поддержкой большого числа пользователей, то такие опасения Андрей Хромов считает беспочвенными: “Все крупнейшие ИТ-компании предлагают решения, которые призваны помочь организациям справиться с нарастающими объемами, сложностью данных и их анализом. Конечно, тут есть свои трудности, но все они преодолимы с помощью уже предлагающихся на рынке решений. А в будущем будут появляться все более усовершенствованные технологии”.

По мнению руководителя по развитию бизнеса сектора финансовых услуг отделения IBM в России и СНГ Михаила Горбатовского, понятие Big Data обозначает новый феномен в аналитической обработке данных. Оно связано с четырьмя видами потребностей, которые ранее присутствовали только в отдельных уникальных проектах, а сегодня нужны все большему числу клиентов:

- скорость анализа: потребность в принятии решений все чаще смещается в сторону работы в режиме реального времени, что требует от аналитической системы выполнения расчетов с минимальными задержками. Если в традиционном хранилище отчет может строиться несколько часов, то в ряде задач произвести сложный анализ требуется за минуты и секунды на основе больших объемов данных;
- разнородность данных: для более глубокого анализа и более качественной поддержки решений в бизнес-среде востребованы возможности поиска закономерностей и ценной информации из всех доступных источников данных, внешних и внутренних, структурированных и неструктурированных. Примером внешних данных являются социальные сети и блоги, внутренних — геолокация, системные логи, поток информации с датчиков. Совместная обработка этих данных ранее была затруднена из-за ограничений технологий хранения и анализа данных;
- объемы данных: речь идет о двух типах задач, связанных с большим объемом данных на единицу времени (гигабайты в секунду) и большими объемами хранимых данных (петабайты). В первом случае алгоритмы анализа требуются применять для больших объемов данных и получать аналитику с низким откликом (например, надо рассчитывать риски по банковским операциям на рынках ценных бумаг для 10 млн. сообщений в секунду с откликом 80 мкс). Во втором случае ретроспективный анализ выполняется на всем объеме петабайтов транзакций, фото и видео, а также текстовых записей;

- глубина исследования данных: анализ нужно выполнять детально, на предмет зависимостей, аномалий, соответствия сохраненным шаблонам и т. п. Делать это нужно зачастую в режиме диалога, коррекции схемы исследования в зависимости от получаемых результатов анализа. Поэтому модели, которые раньше использовались в десктопных системах исследования данных, сегодня должны анализироваться на больших объемах. Например, в онлайн-рекламе приходится индивидуально реагировать на тысячи, миллионы моделей поведения клиентов.

Заместитель директора департамента “Инжиниринговый центр” компании “Техносерв” Андрей Вакатов считает, что понятие Big Data объединило в себе мас-

су подходов и методов обработки больших массивов данных. Сегодня все чаще бизнес-подразделения ставят перед ИТ задачи по обработке, анализу и сопоставлению уже накопленных или непрерывно поступающих данных. Это приводит к созданию разнородных инструментов и подходов, которые могли бы обеспечить достаточно быструю обработку и анализ информации. Взрывной интерес последнего времени к этой концепции вызван в первую очередь ростом потенциально полезных, но неструктурированных данных в глобальной сети.

Директор отделения технологического консалтинга компании ФОРС Валерий Юринский уверен, что Big Data — это действительно качественно новое явление в ИТ: “Уже более трех лет много пишут и говорят о больших данных в сочетании со словом “проблема”. Определений этому термину существует множество, но суть их можно свести к описанию ситуации, когда обработка и хранение больших объемов данных становится настоящей проблемой. Сейчас мы переживаем очередной этап технологической революции, характеризующийся тем, что имеем дело с данными, объем которых превышает возможности привычных нам технологических средств и потому требуется создание новых специальных инструментов для их адекватной обработки. Счет идет на сотни гигабайтов и далее до петабайтов, а в дальнейшем — зеттабайтов”. Он считает, что необходимо также учитывать такие факторы, как большое разнообразие данных, усложнение их структуры, возрастание числа отдельных элементов, скорости поступления и изменения данных, требование получения ответа системы в реальном времени и многое другое. И отмечает, что именно поэтому компания Gartner предложила формулировку “Big Data and Extreme Information Processing and Management” как более полную и лучше отражающую суть явления.

Свое понимание, что такое большие данные, эксперт группы перспективных технологий компании “Инфосистемы Джет” Сергей Артёмов сформулировал так: “Это маркетинговый термин. Как правило, под ним подразумевается большой объем плохо структурированных данных, обработка которых привычными методами невозможна или экономически нецелесообразна. Типичный пример — это записи о транзакциях (например, данные Call Data Records у сотовых операторах или данные платежей в процессинговых центрах), данные с телеметрических датчиков, журналы активности пользователей в крупных интернет-проектах или социальных сетях”.

### Как это используется в мире

Валерий Юринский считает, что говорить о решении проблемы больших данных имеет смысл только в контексте наличия единой цепочки “данные — информация — знание”. Хранимые сведения обрабатываются не просто так, а для получения информации, которой должно быть ровно столько, чтобы человек мог превратить её в знание. Впрочем, нужно иметь в виду, что сегодня машины уже стали и основными производителями информации, и её потребителями. Поэтому сейчас прежде всего требуется обеспечить не интерактивную работу людей с данными, а их эффективный доступ к ним, а машины могли бы работать с машинами. Это необходимо для автоматизации рутинной деятельности по обработке наблюдений, поиску информации и т. п. Причем задача заключается еще и в том, чтобы по возможности устранить большинство неточностей данных, когда их значения известны с той или иной степенью погрешности. Для её решения требуются особые методы анализа и очистки, особенно для петабайтных объемов.

Большие данные меняют концепцию аналитики — с этим тезисом согласен и

Андрей Хромов, который отметил, что в лавине информации главной проблемой становится поиск и выделение, причем часто в реальном времени, существенных для бизнеса сведений, а также их четкое и понятное представление: “В этой ситуации некоторые компании ищут решение, которое поможет им быстро устранить возникшие проблемы с традиционной архитектурой хранилища данных. Другие, более продвинутые, уходят от традиционных инфраструктур и осваивают новые технологии, например обработку и вычисление в оперативной памяти. Или даже рассматривают более федеративные архитектуры, которые позволяют абстрагировать и виртуализировать доступ к данным и анализ по всем распределенным хранилищам и архивам. Целью многих организаций в самом ближайшем будущем станет минимизация влияния больших объемов на производительность, материализации данных на стороне клиента перед проведением анализа, а также их зависимости от предварительно агрегированных данных”.

Александр Яковлев также уверен, что Big Data — это прежде всего новые возможности в современной аналитике: “Если огромные объемы данных хранятся консолидированно, то открываются качественно новые возможности для их совокупного и многостороннего анализа. С одной стороны, сбор и сохранение гигантских массивов информации стали возможны благодаря появлению облачных технологий, а с другой — облака в это же время представляют и практически неограниченные процессорные мощности для обработки этих самых объемов данных, причем по очень эффективной и гибкой схеме. Поэтому сегодня ведущие поставщики наряду с построением и вводом в эксплуатацию глобальных облачных ЦОДов уделяют особое внимание и различным системам сбора данных — различного рода сенсорам, считывающим устройствам и другим относительно простым приспособлениям, обеспечивающим получение и передачу исходных данных. Сегодня они встраиваются в огромное количество устройств и приспособлений в медицине, ритейле, транспортных системах, системах управления движением и т. д. А уже после собранные данные анализируются на базе мощных облачных ЦОДов”.

Понятие Big Data появилось как отражение объективной трансформации подходов к работе с данными. Поясняя этот тезис, Максим Исаев отметил, что сегодня многие компании должны управляться в режиме онлайн, в том числе на основании потоков неструктурированных данных. Текущий инструментальный — программный и аппаратный — был заточен и разработан под совершенно другую практику. В качестве примера он приводит опыт Google, которой для обработки статистики поисковых запросов понадобились новые инструменты. Компания стала придумывать новые алгоритмы, а после того как она поделилась ими с общественностью, тема была подхвачена ИТ-производителями. При этом нужно понимать, что на рынке имеется большая маркетинговая составляющая, сейчас множество продуктов пытается примерить на себя эту новую “одежку”. Нужно иметь в виду, что изначально концепция Big Data развивалась как концепция работы с неструктурированными данными, однако сегодня это понятие стало шире, теперь под ним понимают и работу просто с большими данными.

Алексей Мещеряков уверен, что для компаний, ориентированных на повышение качества предоставляемых услуг, совершенствование работы с огромной клиентской базой и улучшение внутренних процессов управления, уровень использования технологий Big Data будет возрастать. Понимание возможностей и выгод, которые скрыты в больших массивах данных, и практическое применение методов

углубленной аналитики обеспечат организациям дополнительные конкурентные преимущества. У поставщиков есть решения, которые позволяют обрабатывать данные в десятки раз быстрее, чем это было еще два-три года назад.

Сергей Артёмов считает, что проблема Big Data в принципе не нова и на рынке существует немало вариантов ее решения. Есть ряд коммерческих продуктов, имеющих множество инсталляций в самых разнообразных организациях и позволяющих создавать огромные хранилища данных и аналитические системы на их базе. Речь идет о системах, подобных Oracle Exadata, IBM Netezza, или продуктах компании Teradata. Но параллельно с промышленными решениями сейчас становится все более популярна модель работы с Big Data, реализованная в проекте Apache Hadoop. Уже сейчас множество вендоров стали использовать эту систему в составе своих решений, поэтому термин Big Data сегодня прочно ассоциируется именно с Hadoop.

С этим тезисом согласен Вячеслав Архаров: “Пожалуй, самой известной платформой Big Data сегодня является Hadoop, и наша компания также обеспечивает поддержку приложений Hadoop на своих платформах”. Он выделил в развитии направления Big Data несколько трендов.

Во-первых, это растущая сложность данных: исторически данные широко структурировались по типам. Успех поисковых систем и магазинов электронной торговли, которые открыли ценность данных о посещении веб-сайтов, развеял миф о том, что неструктурированные данные не имеют ценности. Для многих предприятий все более актуальным сегодня становится требование анализировать и использовать как структурированные, так и неструктурированные данные.

Во-вторых, растущая сложность анализа: чем сложнее становятся данные, тем труднее их анализировать, например при обработке изображений для распознавания лиц, классификации поисковыми движками видео или использовании данных о посещении веб-сайтов для анализа поведения потребителей. Модели анализа транзакционных данных давно известны и являются зрелыми. Именно они в течение двух последних десятилетий способствовали развитию хранилищ данных и BI. Модели, с помощью которых можно собирать и анализировать сложные данные, только-только развиваются.

В-третьих, растущие требования к BI: в среднем в компаниях количество пользователей, которые имеют доступ к инструментам бизнес-аналитики, составляет менее 20%, тогда как организациям все чаще необходимо буквально в режиме реального времени анализировать данные. Сочетание этих факторов дает толчок развитию аналитики и подготовке прогнозов в режиме реального времени.

В-четвертых, меняющаяся экономика вычислений: облачные вычисления и разнообразие используемых устройств радикальным образом уменьшают стоимость вычислительных мощностей и средств для хранения информации. Уменьшающаяся стоимость распределенных вычислений, памяти и систем хранения фундаментально меняют экономику обработки данных. Стоимость 1 Тб в современных хранилищах данных более чем в половину меньше, чем три года назад, а облачные хранилища данных предлагают десятикратное преимущество в ТСО по сравнению с традиционными системами on-premise.

И наконец, новые технологии: легко масштабируемое оборудование дополняется новыми фреймворками и инструментами для распределенной параллельной обработки данных, благодаря чему появляются недорогие платформы, предлагающие широкие возможности поддержки задач по обработке массивов данных.



# Sybase IQ 15.4 — СУБД для анализа "больших данных"

В наше время предприятиям уже недостаточно просто понимать текущую ситуацию на рынке. Нужно уметь заглянуть в будущее и, задав вопрос "что, если...", узнать вероятные риски и благоприятные возможности. Однако реализовать эту идею на практике не просто. Для этого надо научиться оперативно анализировать огромные объемы информации, которые постоянно поступают на предприятие. Проблему усложняет не только объем данных, но и то, что они записаны в разных форматах, являются неструктурированными и хранятся в различных источниках.

Компания Sybase предлагает единую платформу для аналитической обработки огромных разрозненных массивов данных — аналитическую СУБД Sybase IQ, в которой реализованы новаторские технологии для преодоления ограничений традиционных баз данных. При этом Sybase IQ является универсальной СУБД, предназначенной для анализа любых типов данных — структурированных, частично структурированных, неструктурированных — с помощью самых разных алгоритмов и аналитических систем.

Главными преимуществами Sybase IQ являются механизм поcolumn-based хранения таблиц, масштабируемая grid-архитектура, а также высокая производительность при небольшом потреблении аппаратных ресурсов. В версии Sybase IQ 15.3 реализована технология массово-параллельной обработки PlexQ с разделением всех ресурсов, которая обеспечивает обслуживание в реальном времени тысяч пользователей, поддержку множества смешанных нагрузок и анализ больших массивов данных.

Динамически распределяя рабочую нагрузку между узлами кластера, PlexQ позволяет решать сложные аналитические задачи

с ускорением от 10 до 100 крат по сравнению с традиционными хранилищами данных, а кластеризация с разделением всех ресурсов в сочетании с поколонным хранением и механизмом логических серверов обеспечивает высокое быстродействие для всех видов запросов и типов нагрузок применительно к любым наборам данных.

В недавно выпущенной версии Sybase IQ 15.4 возможности продукта расширены за счет встроенного интерфейса прикладного программирования (API) для применения модели распределенных вычислений MapReduce, гибкого механизма интеграции с технологией Hadoop и поддержки языка PMML (Predictive Model Markup Language). В состав пакета включены расширенные библиотеки алгоритмов статистической обработки и углубленного анализа данных (data mining), в полной мере использующие возможности распределенной обработки запросов в кластерной конфигурации PlexQ. Новые API позволяют разработчикам приложений быстро и безопасно реализовывать свои собственные (proprietary) алгоритмы, которые исполняются внутри базы данных, обеспечивая прирост быстродействия в десятки раз по сравнению с традиционными подходами. Кроме того, значительно усовершенствованы алгоритмы сжатия текста и интерфейсы массовой загрузки данных.

## Основные новые возможности Sybase IQ 15.4

**Прикладные сервисы.** Встроенный API MapReduce представляет собой систему для распределения аналитической обработки больших наборов данных по кластеру PlexQ. Библиотеки функций этого интерфейса можно писать на языке C++ и обращаться к ним из SQL.

Табличные параметризованные пользовательские функции (UDF) — это встроенный API, позволяющий разработчикам создавать и развертывать библиотеки функций на C++, которые могут исполняться параллельно на кластере PlexQ и в самом сервере базы данных Sybase IQ. С помощью этого интерфейса программисты могут реализовывать свои собственные алгоритмы и выполнять их внутри Sybase IQ, повышая скорость обработки.

Эмулятор для моделирования внутрибазовой аналитической обработки облегчает разработку и тестирование UDF, предназначенных для выполнения внутри Sybase IQ.

**Библиотеки функций внутрибазовой аналитической обработки.** Интеграция с технологией Hadoop и средства объединения (federation) позволяют интегрировать результаты анализа, выполняемого в Hadoop, с запросами для Sybase IQ, что дает возможность находить нужную информацию в больших массивах структурированных и неструктурированных данных, а затем интегрировать ее в Sybase IQ для совместного анализа с транзакционными данными и результатами, полученными из других источников.

Поддержку языка PMML обеспечивает сертифицированный подключаемый модуль компании Zementis, который позволяет автоматизировать выполнение аналитических моделей, описанных на этом языке с помощью таких инструментов, как SAS, SPSS Clementine и других популярных средств прогнозного моделирования.

Появилась возможность обращения посредством SQL-запросов к базам IQ из сервера Sybase IQ с помощью R — статистического инструмента с открытым кодом (поддерживается только интерфейс R JDBC).

Расширен функционал библиотеки анали-

тических и статистических алгоритмов DB Lytix компании Fuzzy Logix. Теперь библиотека может использовать API MapReduce в некоторых алгоритмах углубленного анализа данных для массово-параллельной обработки.

**Управление данными.** Ускорена массовая загрузка данных в Sybase IQ с использованием интерфейсов ODBC и JDBC. При этом значительно увеличивается масштабируемость и быстродействие приложений.

Улучшенное сжатие данных типа VARCHAR, VARBINARY, CHAR и BINARY позволяет более эффективно и экономично развертывать быстродействующие приложения текстового анализа.

Генератор эталонной архитектуры PowerDesigner 16.1 дает возможность быстро создавать хранилища или витрины данных на базе Sybase IQ.

Улучшен ряд аспектов администрирования в инструменте управления базой данных Sybase Control Center (SCC). Упрощено управление часто используемыми функциями SCC и развертывание встроенной, внешней и текстовой внутрибазовой аналитики. Введено также новое окно исполнения команд SQL, облегчающее тестирование SQL-программ для вычислений ad hoc.

## Заключение

СУБД Sybase IQ 15.4 преобразует большие массивы данных в практически ценную информацию, ориентированную на все категории сотрудников. При этом возможности, которые дает аналитическая обработка крупных информационных массивов, становятся доступны всем категориям пользователей и для всех бизнес-процессов в масштабе организации.

► "В своей жизни мы практически постоянно сталкиваемся с применением тех или иных методов, объединенных определением "большие данные", — сказал Андрей Вакатов. — Множество крупных компаний осознало, что кроме данных, накопленных в собственных корпоративных сегментах, появилась не менее насыщенная информацией глобальная сеть. И эта информация может быть очень полезна компаниям как с точки зрения текущей деятельности, так и стратегических направлений бизнеса".

Характеризуя ситуацию в целом, Михаил Горбатовский прогнозирует, что поскольку спрос со стороны бизнес-компаний и госсектора на решение таких задач уже назрел, можно ожидать высоких показателей роста рынка этих решений.

## Какова актуальность темы больших данных для России

Максим Исаев уверен, что в России спрос пока в стадии формирования: "Решения есть, а готовность заказчиков их внедрять на невысоком уровне. Сейчас вендоры всячески пытаются разогреть этот рынок. Думаю, в ближайшие год-два будут преобладать пилотные проекты. В нашей стране первыми технологии Big Data адаптируют те компании, которые тратят свой ИТ-бюджет на инновационные составляющие. Это банковская сфера и телеком, также к ним примкнет госсектор". Он пояснил свой прогноз. В телекоме это может быть все, что связано с анализом абонентской базы. Например, социальный анализ: выявление связи между абонентами, групп, влияющих абонентов; анализ базы вкпе с данными социальных сетей и веб-приложений, геоинформационными данными. В российских банках использование Big Data будет, но с оговорками. На Западе наиболее активные пользователи — инвестиционные банки, работающие с большими потоками информации, по-

ступающей с большого числа торговых площадок. Нашим банкам такой режим работы не нужен, поэтому их интересы лежат в плоскости снижения затрат на обработку данных и в организации более качественной работы с клиентскими базами. Госсектор, пожалуй, наиболее емкий с точки зрения применения Big Data, особенно в связи с переводом госуслуг в электронный вид и созданием электронного правительства объем задействованных данных растет на порядки. Одним из наиболее активных потребителей технологии больших данных становится медицина. Помимо названных отраслей хорошие перспективы у Big Data есть в логистике, ритейле, автомобилестроении. Правда, уточнил Максим Исаев, технологическая концепция в нынешнем ее виде ориентирована главным образом на крупных потребителей, применение Big Data целесообразно, когда объемы данных составляют сотни терабайтов.

По мнению Александра Яковлева, уровень актуальности больших данных сильно коррелирует с актуальностью облаков: "Хотя все же аспект облаков в России кажется намного важнее, так как именно здесь у нас есть отставание от остального мира как минимум на несколько лет. Уже сегодня можно создать облачное хранилище огромной емкости, высокой надежности, гибкое, недорогое и с использованием большого количества стандартных компонентов".

Алексей Мещеряков считает, что тема больших данных с неизбежностью становится актуальной и для России, особенно для тех компаний, которые уже успели оценить преимущества аналитических процедур для своего бизнеса. Речь идет о банковских организациях, телекоммуникационных и страховых компаниях, розничных сферах торговли. Там накоплены огромные массивы данных по клиентам и операционной деятельности. А конку-

ренция заставляет использовать весь спектр доступной информации, включая и открытые источники. В числе новых аналитических методов, связанных с темой Big Data, Алексей Мещеряков назвал анализ социального окружения клиентов. В целом у финансовых институтов появляется понимание необходимости проводить детальный анализ своих данных с помощью современных аналитических решений, что позволит им избежать ненужной утечки капитала и потери репутации.

С тем, что большие данные нужны российским заказчикам, согласен Андрей Хромов: "Прежде всего это касается компаний, у которых к настоящему дню накопились базы данных больших и очень больших объемов. Я имею в виду объемы в 1 Тб и выше. Во-первых, это компании, за много лет накопившие многотерабайтные базы, например крупные банки или федеральные ведомства. Ко второму типу обладателей больших данных я бы отнес тех, чья деятельность связана с необходимостью постоянного учета и переработки огромных объемов данных, в частности крупные телекоммуникационные компании, ежедневно обрабатывающие гигабайты CDR-данных. Еще одна категория компаний — те, чей бизнес связан с большим числом клиентов и большим и очень большим числом операций (сделок). Это крупные торговые сети. Сюда же можно отнести и те банки, которые активно развивают розничный бизнес, потребительское кредитование".

Михаил Горбатовский перечисляет примерно такой же список: "К тем, кто в России интересуется решениями для больших данных, относятся инвестиционные и розничные банки, телекоммуникационные операторы, биржи, диспетчерские и ситуационные центры, транспортные компании, государственные учреждения. Но речь идет, конечно, только о крупных организациях".

"Особенно остро в технологиях, способных обрабатывать неструктурированные массивы данных, нуждаются компании, работающие на высококонкурентных рынках: телекомы и финансовые структуры, — считает Андрей Вакатов. — Трудно сказать, для каких предприятий или отраслей тематика больших данных не является интересной. Виртуализация и облачные вычисления все больше вовлекают в эту гонку сегмент малого и среднего бизнеса. И даже если они не готовы приобрести собственные решения для анализа Big Data, они будут генерировать спрос на анализ и обработку больших данных как на услуги".

По мнению Валерия Юринского, наша страна находится только в самом начале освоения темы больших данных, но интерес к ней растет: "Область, где с беспрецедентными объемами данных традиционно сталкиваются чаще всего — это сложные научные эксперименты в разных областях знания, а финансирование научных экспериментов в России сейчас не на самом высоком уровне. У нас пока нет крупных компаний — держателей больших данных, подобных Amazon или Google, однако их появление уже не за горами — в первую очередь это "Яндекс" и Mail.ru".

Он также отметил, что на интерес к теме существенно влияет укрупнение бизнеса. До определенного времени крупные финансовые и телекоммуникационные компании в России имели децентрализованную распределенную структуру, сейчас же штаб-квартиры требуют от своих филиалов все более детализированные данные, которые затем нужно где-то хранить и как-то обрабатывать. К тому же все шире стали использоваться исторические и мультимедийные данные. Ссылаясь на мнения аналитиков, Валерий Юринский сказал о том, что под влиянием больших данных существенной трансформации под-

ПРОДОЛЖЕНИЕ НА С. 23 ►



# Big Data требует новых инфраструктурных решений

**В** последнее время в области ИТ произошел качественный скачок. Развитие технологий позволило людям оцифровывать весь окружающий мир, причем степень детализации очень быстро растет. Этот процесс порождает гигантские массивы данных, которые поступают от всевозможных цифровых датчиков, сенсоров, фото- и видеокамер, оборудования для научных исследований, медицинских приборов и т. д. Как справиться с этой информационной лавиной и извлечь из нее пользу? О новых инфраструктурных решениях, обеспечивающих решение этой проблемы, рассказывает Николай Зезюлинский, директор по развитию бизнеса компании ФОРС.



Николай Зезюлинский

**Проблема накопления огромного количества данных существует уже давно, но термин Big Data появился относительно недавно. Что изменилось? И что подразумевает это понятие?**

Действительно, уже давно объемы данных растут быстрее, чем развиваются средства для работы с ними. И хотя технологии отвечают на возрастающие запросы, это происходит с отставанием. В эпоху дискета было не просто скопировать файл размером в 1 Гб, а сейчас это не проблема. Зато возникают другие задачи.

Например, чтобы осуществить перенос хранилища данных объемом 11 Тб, требуется больше суток для одной лишь передачи этой массы информации по стандартным каналам связи.

Проблема Big Data заключается не в способности хранить эти массивы, это так или иначе делается, а в том, способны ли мы извлечь пользу из этих данных. Ведь если мы ничего не можем с ними сделать, значит, это просто цифровой мусор. А между тем их анализ может дать совершенно невероятные результаты.

К тому же принципиально изменился и подход к хранению информации. Сейчас просто невозможно собрать все данные в одном месте, а потом их анализировать. Поэтому они остаются на своих местах, а система анализирует их, выбирая из разных мест и подключая для обработки большие вычислительные мощности. Нарастают технические проблемы вокруг надежного хранения больших объемов данных, выборки из больших массивов, поиска нужных данных, их структуризации и очистки. Поэтому мы говорим о том, что требуются новые технологии, особый класс сверхпроизводительных инфраструктурных решений.

**Для каких заказчиков задача обработки больших объемов данных стоит наиболее остро и с чем это связано? Какие новые требования это накладывает на функционал аналитических систем?**

Сейчас информационные технологии используют все предприятия, даже малый бизнес. Но есть ряд отраслей, для которых автоматизация не просто обеспечивает удобство работы, а является основой бизнеса — например, это телеком и банки. Государственный сектор тоже сталкивается с проблемой Big Data. Ведь с появлением задачи предоставления населению услуг в электронном виде тут же появилась необходимость обработки огромных массивов информации.

Что касается функционала аналитических систем, то изменения в них связаны не с Big Data, а с переменами в способах использования таких систем. Речь идет о всепроникающей аналитике.

Раньше аналитическими системами пользовались лишь специалисты-аналитики, которые строили диаграммы и графики для руководства. Но сейчас функциональность практически всех аналитических систем стала такой, что с ними может работать любой человек, а не только профессионал.

К сожалению, сейчас далеко не все заказчики знают, что такие системы доступны. Требуется просветительская работа. Кроме того, в нашей стране многие предприятия все еще решают вопросы первичной автоматизации. Они с интересом смотрят на передовые аналитические системы, но необходимой для этого ИТ-инфраструктуры у них еще нет. Тем не менее движение в сторону всепроникающей аналитики наблюдается.

**Готовы ли вендоры сегодня предоставить заказчикам программно-аппаратные решения для работы с Big Data?**

Да, готовы. Так, наш стратегический партнер корпорация Oracle после слияния с Sun разработала целую линейку таких продуктов. Впервые были созданы готовые вычислительные комплексы на основе специализированного программного и типового аппаратного обеспечения, оптимизированные для решения задач по обработке и хранению данных.

Сначала появилась Exadata Database Machine — машина баз данных, предназначенная для кардинального повышения производительности работы баз данных. Следующей была выпущена Exalogic Elastic Cloud, оптимизированная для достижения максимальной скорости работы приложений и развертывания частных и публичных облаков. А недавно в продажу поступила машина «больших данных» — Oracle Big Data Appliance для обработки огромных массивов неструктурированных данных.

Важно, что во всех этих продуктах в единое целое объединены серверы, устройства хранения, сетевое оборудование и программное обеспечение. Они не требуют предвзвешенной настройки и поставляются как полностью готовый к эксплуатации и простой в управлении программно-аппаратный комплекс. По сути, это своего рода «строительные блоки» для создания ЦОДов нового поколения — инфраструктуры, необходимой для развития облачных услуг.

Уже больше года в нашей компании действует демонстрационный инженерный центр FORS ExaStack Studio, где эти системы можно не только увидеть, посмотреть, как они работают, но и протестировать на своих базах данных и приложениях.

**Но, как вы отметили, все предприятия уже используют ИТ-системы. Оправдан ли отказ от работающего оборудования? Как просчитать экономическую эффективность перехода на новый класс решений?**

Общий принцип обычно такой: если у вас есть исправно работающее оборудование, которое справляется с имеющимися задачами, то нет никакого смысла от него отказываться. Для перехода на что-то новое долж-

на быть необходимость. Но когда такая необходимость появляется, возникает вопрос: на какое оборудование переходить?

Судя по нашему опыту, есть целый ряд заказчиков, которые интересуются программно-аппаратными системами, потому что старые системы не справляются с накопившимися объемами данных. Можно, конечно, наращивать вычислительные мощности, но тогда нередко оказывается, что отдельные компоненты слабо связаны и передача данных становится «узким» местом. Система должна быть хорошо сбалансированной с точки зрения распределения нагрузки.

Идея Oracle как раз и заключается в обеспечении такой сбалансированности. Системы, построенные на новых машинах, можно наращивать по мере необходимости без нарушения общей производительности всего комплекса. К тому же это — машины, разработанные на основе общепринятых стандартов, поэтому они совместимы с оборудованием других производителей.

Что касается экономической эффективности, то это зависит от того, насколько организация заинтересована в извлечении пользы из огромных массивов данных. Она несомненно будет, если с помощью более совершенных ИТ-инструментов удастся привлечь новых клиентов и удержать существующих.

**Но Oracle — не единственный поставщик программно-аппаратных решений для хранения и обработки больших массивов данных. Как компания продвигает свои программно-аппаратные решения для Big Data в России?**

У Oracle большая доля на российском рынке СУБД, и сейчас компания стремится использовать эту клиентскую базу, с тем чтобы предложить программно-аппаратные решения.

При этом в России Oracle придерживается стратегии работы через партнеров. В первую очередь именно на них ориентирован наш инженерный центр ExaStack Studio, о котором я уже упоминал. Хотя заказчики тоже могут провести тестирование своих систем на совместно работающих комплексах Exadata и Exalogic. В скором времени к ним добавится и Exalytics. Таких инженерных центров всего несколько в мире, поэтому наш ориентирован не только на Россию, но и на всю Восточную Европу.

К нам уже обратилось немало партнеров — проекты, связанные с тестированием, оптимизацией приложений и миграцией баз данных на новые системы, идут все более активно.

**Но ведь другие партнеры Oracle — это ваши конкуренты. Они не опасаются приводить к вам своих клиентов?**

Пока к нам в основном обращаются разработчики софта, например банковских и биллинговых систем, мы также тестировали работу приложений SAP.

Другие партнеры Oracle в чем-то — наши конкуренты, а в чем-то — партнеры и для нас тоже. Мы часто совместно участвуем в крупных проектах как члены единой команды внедрения, кто-то у кого-то может работать на субподряде, есть и другие формы сотрудничества — например, обучение. И, наконец, у нас есть собственная партнерская сеть, которая осуществляет дистрибуцию продуктов Oracle. Если у такого партнера есть свои разработки, которые он хотел бы протестировать на новом оборудовании, или у него есть заказчик, который этим оборудованием интересуется, то мы готовы предложить такую услугу на нашей площадке. Результаты иногда бывают просто ошеломляющими — за счет применения специальных алгоритмов сжатия объемы данных уменьшаются в разы, а благодаря сбалансированности системы скорость

формирования отчетов возрастает многократно.

**ФОРС несет издержки на создание, содержание и обслуживание этого центра. Как же это окупается?**

Действительно, раньше мы не занимались поставками оборудования, и создание экспертизы по новым решениям потребовало значительных инвестиций в оборудование и специалистов. Но такие траты необходимы для продвижения принципиально новых на рынке продуктов. Сейчас мы проводим тестирование бесплатно, но какие-то другие услуги — такие, к примеру, как миграция баз данных, оптимизация приложений, последующее сопровождение систем, мы оказываем на коммерческой основе. Мы также рассчитываем на то, что партнеры Oracle будут размещать у нас заказы на поставку вычислительных комплексов. Сейчас у Oracle не предусмотрена привязка партнеров к дистрибьюторам, мы работаем на открытом рынке, и такие услуги позволяют нам дополнительно привлекать партнеров, предлагая им более выгодные условия сотрудничества.

Тем не менее главная задача инженерного центра остается прежней — ознакомление партнеров и заказчиков Oracle с возможностями и преимуществами этих систем. Мы регулярно проводим обучающие тренинги и семинары, осуществляем проекты, направленные на демонстрацию возможностей новинок для решения задач заказчиков.

**Есть ли спрос на такие комплексы — в мире и в России? Кто основные заказчики — коммерческие компании или госсектор? Есть ли варианты систем для малого и среднего бизнеса?**

Число заказов в России перевалило за десяток. Интерес проявляют государственные и коммерческие организации, которые занимаются обработкой больших массивов данных. И хотя вышеперечисленные системы ориентированы на крупные компании, у Oracle есть решение и для малого и среднего бизнеса — система Database Appliance. Мы уже привезли ее в Россию и установили в своем центре. Это — машина стоимостью от 50 тыс. долл., предназначенная для работы баз данных высокой готовности среднего размера. Она очень проста в установке и управлении и требует минимального обслуживания.

Общая черта этих машин заключается в том, что все они направлены, с одной стороны, на сокращение затрат на администрирование, а с другой — на повышение сбалансированности всей системы и максимально полное использование вычислительных ресурсов. Когда мощности будет не хватать, всегда можно добавить дополнительные стойки или другую машину класса Exa — на сбалансированность системы в целом это никак не повлияет.

Не следует забывать и том, что Oracle продолжает выпускать традиционные программные продукты, которые поставляются отдельно от оборудования. У заказчиков всегда есть выбор — приобрести нужные им компоненты ИТ-инфраструктуры по отдельности или сразу готовый преднастроенный программно-аппаратный комплекс. Ведь по разным причинам не все готовы перейти на принципиально новый класс решений. У кого-то в этом пока не появилось необходимости, а где-то просто преобладают консервативные настроения.

**Как вы планируете развивать новое направление, связанное с продвижением оборудования?**

Мы предлагаем услуги по оптимизации, настройке и миграции приложений заказчиков при переходе на новые системы. Это сложные технические задачи, но заказчикам нет смысла создавать для их решения свою экспертизу, поскольку эти задачи одноразовые. А у нас такие проекты поставлены на поток.

# PC WEEK RUSSIAN EDITION

## КОРПОРАТИВНАЯ ПОДПИСКА

### Уважаемые читатели!

Только полностью заполненная анкета, рассчитанная на руководителей, отвечающих за автоматизацию предприятий; специалистов по аппаратному и программному обеспечению, телекоммуникациям, сетевым и информационным технологиям из организаций, имеющих **более 10 компьютеров**, дает право на **бесплатную** подписку на газету PC Week/RE в течение года с момента получения анкеты. Вы также можете заполнить анкету на сайте: [www.pcweek.ru/subscribe\\_print/](http://www.pcweek.ru/subscribe_print/).

**Примечание.** На домашний адрес еженедельник по бесплатной корпоративной подписке не высылается. Данная форма подписки распространяется только на территорию РФ.

### Я хочу, чтобы моя организация получала PC Week/RE !

Название организации: \_\_\_\_\_

Почтовый адрес организации: \_\_\_\_\_

Индекс: \_\_\_\_\_ Область: \_\_\_\_\_

Город: \_\_\_\_\_

Улица: \_\_\_\_\_ Дом: \_\_\_\_\_

Фамилия, имя, отчество: \_\_\_\_\_

Подразделение / отдел: \_\_\_\_\_

Должность: \_\_\_\_\_

Телефон: \_\_\_\_\_ Факс: \_\_\_\_\_

E-mail: \_\_\_\_\_ WWW: \_\_\_\_\_

(Заполните анкету печатными буквами!)

#### 1. К какой отрасли относится Ваше предприятие?

1. Энергетика
2. Связь и телекоммуникации
3. Производство, не связанное с вычислительной техникой (добывающие и перерабатывающие отрасли, машиностроение и т. п.)
4. Финансовый сектор (кроме банков)
5. Банковский сектор
6. Архитектура и строительство
7. Торговля товарами, не связанными с информационными технологиями
8. Транспорт
9. Информационные технологии (см. также вопрос 2)
10. Реклама и маркетинг
11. Научно-исследовательская деятельность (НИИ и вузы)
12. Государственно-административные структуры
13. Военные организации
14. Образование
15. Медицина
16. Издательская деятельность и полиграфия
17. Иное (что именно) \_\_\_\_\_

#### 2. Если основной профиль Вашего предприятия – информационные технологии, то уточните, пожалуйста, сегмент, в котором предприятие работает:

1. Системная интеграция
2. Дистрибуция
3. Телекоммуникации
4. Производство средств ВТ
5. Продажа компьютеров
6. Ремонт компьютерного оборудования
7. Разработка и продажа ПО
8. Консалтинг
9. Иное (что именно) \_\_\_\_\_

#### 3. Форма собственности Вашей организации (отметьте только один пункт)

1. Госпредприятие
2. ОАО (открытое акционерное общество)
3. ЗАО (закрытое акционерное общество)
4. Зарубежная фирма
5. СП (совместное предприятие)
6. ТОО (товарищество с ограниченной ответственностью) или ООО (Общество с ограниченной ответственностью)

#### 4. К какой категории относится подразделение, в котором Вы работаете? (отметьте только один пункт)

1. Дирекция
2. Информационно-аналитический отдел
3. Техническая поддержка
4. Служба АСУ/ИТ
5. ВЦ
6. Инженерно-конструкторский отдел (САПР)
7. Отдел рекламы и маркетинга
8. Бухгалтерия/Финансы
9. Производственное подразделение
10. Научно-исследовательское подразделение
11. Учебное подразделение
12. Отдел продаж
13. Отдел закупок/логистики
14. Иное (что именно) \_\_\_\_\_

#### 5. Ваш должностной статус (отметьте только один пункт)

1. Директор / президент / владелец
2. Зам. директора / вице-президент
3. Руководитель подразделения
4. Сотрудник / менеджер
5. Консультант
6. Иное (что именно) \_\_\_\_\_

#### 6. Ваш возраст

1. До 20 лет
2. 21–25 лет
3. 26–30 лет
4. 31–35 лет
5. 36–40 лет
6. 41–50 лет
7. 51–60 лет
8. Более 60 лет

#### 7. Численность сотрудников в Вашей организации

1. Менее 10 человек
2. 10–100 человек
3. 101–500 человек
4. 501–1000 человек
5. 1001–5000 человек
6. Более 5000 человек

#### 8. Численность компьютерного парка Вашей организации

1. 10–20 компьютеров
2. 21–50 компьютеров

3. 51–100 компьютеров
4. 101–500 компьютеров
5. 501–1000 компьютеров
6. 1001–3000 компьютеров
7. 3001–5000 компьютеров
8. Более 5000 компьютеров

#### 9. Какие ОС используются в Вашей организации ?

1. DOS
2. Windows 3.xx
3. Windows 9x/ME
4. Windows NT/2K/XP/2003
5. OS/2
6. Mac OS
7. Linux
8. AIX
9. Solaris/SunOS
10. Free BSD
11. HP/UX
12. Novell NetWare
13. OS/400
14. Другие варианты UNIX
15. Иное (что именно) \_\_\_\_\_

#### 10. Коммуникационные возможности компьютеров Вашей организации

1. Имеют выход в Интернет по выделенной линии
2. Объединены в intranet
3. Объединены в extranet
4. Подключены к ЛВС
5. Не объединены в сеть
6. Dial Up доступ в Интернет

#### 11. Имеет ли сеть Вашей организации территориально распределенную структуру (охватывает более одного здания)?

- Да  Нет

#### 12. Собирается ли Ваше предприятие устанавливать интрасети (intranet) в ближайший год ?

- Да  Нет

#### 13. Сколько серверов в сети Вашей организации ?

1. ЕС ЭВМ
2. IBM
3. Unisys
4. VAX
5. Иное (что именно) \_\_\_\_\_

#### 14. Если в Вашей организации используются мэйнфреймы, то какие именно?

1. ЕС ЭВМ
2. IBM
3. Unisys
4. VAX
5. Иное (что именно) \_\_\_\_\_
6. Не используются

#### 15. Компьютеры каких фирм-изготовителей используются на Вашем предприятии?

- |                   |                          |                          |
|-------------------|--------------------------|--------------------------|
| “Аквариус”        | Настольные ПК            | Серверы                  |
| ВИСТ              | <input type="checkbox"/> | <input type="checkbox"/> |
| “Формоза”         | <input type="checkbox"/> | <input type="checkbox"/> |
| Acer              | <input type="checkbox"/> | <input type="checkbox"/> |
| Apple             | <input type="checkbox"/> | <input type="checkbox"/> |
| CLR               | <input type="checkbox"/> | <input type="checkbox"/> |
| Compaq            | <input type="checkbox"/> | <input type="checkbox"/> |
| Dell              | <input type="checkbox"/> | <input type="checkbox"/> |
| Fujitsu Siemens   | <input type="checkbox"/> | <input type="checkbox"/> |
| Gateway           | <input type="checkbox"/> | <input type="checkbox"/> |
| Hewlett-Packard   | <input type="checkbox"/> | <input type="checkbox"/> |
| IBM               | <input type="checkbox"/> | <input type="checkbox"/> |
| Kraftway          | <input type="checkbox"/> | <input type="checkbox"/> |
| R.&K.             | <input type="checkbox"/> | <input type="checkbox"/> |
| R-Style           | <input type="checkbox"/> | <input type="checkbox"/> |
| Rover Computers   | <input type="checkbox"/> | <input type="checkbox"/> |
| Sun               | <input type="checkbox"/> | <input type="checkbox"/> |
| Siemens Nixdorf   | <input type="checkbox"/> | <input type="checkbox"/> |
| Toshiba           | <input type="checkbox"/> | <input type="checkbox"/> |
| Иное (что именно) | _____                    | _____                    |

#### 16. Какое прикладное ПО используется в Вашей организации?

1. Средства разработки ПО
2. Офисные приложения
3. СУБД
4. Бухгалтерские и складские программы
5. Издательские системы
6. Графические системы
7. Статистические пакеты
8. ПО для управления производственными процессами
9. Программы электронной почты
10. САПР
11. Браузеры Internet
12. Web-серверы
13. Иное (что именно) \_\_\_\_\_

#### 17. Если в Вашей организации установлено ПО масштаба предприятия, то каких фирм-разработчиков?

1. “1С”
2. “Айти”
3. “Галактика”
4. “Парус”
5. BAAN
6. Navision
7. Oracle
8. SAP
9. Epicor Scala
10. ПО собственной разработки
11. Иное (что именно) \_\_\_\_\_

#### 18. Существует ли на Вашем предприятии единая корпоративная информационная система?

- Да  Нет

#### 19. Если Ваша организация не имеет своего Web-узла, то собирается ли она в ближайший год завести его?

- Да  Нет

#### 20. Если Вы используете СУБД в своей деятельности, то какие именно?

1. Adabas
2. Cache
3. DB2
4. dBase
5. FoxPro
6. Informix
7. Ingress
8. MS Access
9. MS SQL Server
10. Oracle
11. Progress
12. Sybase
13. Иное (что именно) \_\_\_\_\_

#### 21. Как Вы оцениваете свое влияние на решение о покупке средств информационных технологий для своей организации? (отметьте только один пункт)

1. Принимаю решение о покупке (подписываю документ)
2. Составляю спецификацию (выбираю средства) и рекомендую приобрести
3. Не участвую в этом процессе
4. Иное (что именно) \_\_\_\_\_

#### 22. На приобретение каких из перечисленных групп продуктов или услуг Вы оказываете влияние (покупаете, рекомендуете, составляете спецификацию)?

- Системы**
1. Мэйнфреймы
  2. Миникомпьютеры
  3. Серверы
  4. Рабочие станции
  5. ПК
  6. Тонкие клиенты
  7. Ноутбуки
  8. Карманные ПК

#### Сети

9. Концентраторы
10. Коммутаторы
11. Мосты
12. Шлюзы
13. Маршрутизаторы
14. Сетевые адаптеры
15. Беспроводные сети
16. Глобальные сети
17. Локальные сети
18. Телекоммуникации

#### Периферийное оборудование

19. Лазерные принтеры
20. Струйные принтеры
21. Мониторы

22. Сканеры
23. Модемы
24. ИБП (UPS)

#### Память

25. Жесткие диски
26. CD-ROM
27. Системы архивирования
28. RAID
29. Системы хранения данных

#### Программное обеспечение

30. Электронная почта
31. Групповое ПО
32. СУБД
33. Сетевое ПО
34. Хранилища данных
35. Электронная коммерция
36. ПО для Web-дизайна
37. ПО для Интернета
38. Java
39. Операционные системы
40. Мультимедийные приложения

#### Средства разработки программ

41. Средства разработки программ
42. CASE-системы
43. САПР (CAD/CAM)
44. Системы управления проектами
45. ПО для архивирования

#### Внешние сервисы

46. \_\_\_\_\_
47. \_\_\_\_\_

#### 23. Каков наивысший уровень, для которого Вы оказываете влияние на покупку компьютерных изделий или услуг (служб)?

1. Более чем для одной компании
2. Для всего предприятия
3. Для подразделения, располагающегося в нескольких местах
4. Для нескольких подразделений в одном здании
5. Для одного подразделения
6. Для рабочей группы
7. Только для себя
8. Не влияю
9. Иное (что именно) \_\_\_\_\_

#### 24. Через каких провайдеров в настоящее время Ваша фирма получает доступ в интернет и другие интернет-услуги?

1. “Демос”
2. МТУ-Интел
3. “Релком”
4. Combellga
5. Comstar
6. Golden Telecom
7. Equant
8. ORC
9. Telmos
10. Zebra Telecom
11. Через других (каких именно) \_\_\_\_\_

Дата заполнения \_\_\_\_\_

Отдайте заполненную анкету представителям PC Week/RE либо пришлите ее по адресу: **109147, Москва, ул. Марксистская, д. 34, корп. 10, PC Week/RE.**

Анкету можно отправить на e-mail: [info@pcweek.ru](mailto:info@pcweek.ru)



# ВЫБЕРИ

# ЧЕВИДНОЕ!



# ПОДПИШИСЬ

# НА 2012 ГОД

Я подписываюсь

на 6 месяцев и плачу за 17 журналов 1020 рублей (в т. ч. НДС 10%)  
 на 12 месяцев и плачу за 35 журналов 2100 рублей (в т. ч. НДС 10%)

Ф.И.О. \_\_\_\_\_ дата рождения \_\_\_\_\_ индекс \_\_\_\_\_  
 обл./край \_\_\_\_\_ р-н \_\_\_\_\_  
 город \_\_\_\_\_ улица \_\_\_\_\_  
 дом \_\_\_\_\_ корп. \_\_\_\_\_ этаж \_\_\_\_\_ кв. \_\_\_\_\_ домофон \_\_\_\_\_  
 код \_\_\_\_\_ тел. \_\_\_\_\_

Копия квитанции об оплате от \_\_\_\_\_ с отметкой банка прилагается



ИЗВЕЩЕНИЕ	ИНН 7707010704 КПП 770701001 ЗАО «СК Пресс»	получатель платежа	
	Учреждение банка Сбербанк России, ОАО Вернадское ОСБ г. Москвы № 7970		
Кассир	Расчетный счет № 40702810938100100746	БИК	044525225
	Кор. счет: 30101810400000000225	фамилия, и. о., адрес	
КВИТАНЦИЯ	ИНН 7707010704 КПП 770701001 ЗАО «СК Пресс»	получатель платежа	
	Учреждение банка Сбербанк России, ОАО Вернадское ОСБ г. Москвы № 7970		
Кассир	Расчетный счет № 40702810938100100746	БИК	044525225
	Кор. счет: 30101810400000000225	фамилия, и. о., адрес	
	Назначение платежа	Дата	Сумма
	Подписка на журнал «PC WEEK»		
	Плательщик:	Всего:	

**Стоимость подписки:**

На 6 месяцев (17 журналов) — 1020 рублей (в т. ч. НДС 10%)

На 12 месяцев (35 журналов) — 2100 рублей (в т. ч. НДС 10%)

Данное предложение на подписку и указанные цены действительны до 30.06.2012

**Чтобы оформить подписку Вам необходимо:**

- Заполнить прилагаемый купон-заявку и платежное поручение.
- Перевести деньги (стоимость подписного комплекта) на указанный р/с в любом отделении Сбербанка.
- Отправить заполненный купон-заявку и копию квитанции о переводе денег по адресу:

109147, г. Москва, ул. Марксистская, 34, корп.10,  
 3 этаж, оф. 328 (отдел распространения, подписка),  
 или по факсу: (495) 974-2263. Тел. (495) 974-2260,  
 отдел распространения, менеджеру по подписке.

Журнал высылается заказной бандеролью.

Цена подписки включает в себя стоимость доставки в пределах РФ.

Если мы получили Вашу заявку до 10-го числа текущего месяца и деньги поступили на р/с ООО «СК Пресс», подписка начинается со следующего месяца. Не забудьте, пожалуйста, указать в квитанции Ваши фамилию и инициалы, а также Ваш точный адрес с почтовым индексом.

Внимание! Отдел подписки не несет ответственность, если подписка оформлена через другие фирмы.

Редакционная подписка осуществляется только в пределах РФ.

Деньги за принятую подписку не возвращаются.

**Условия подписки:**

\* Минимальный период подписки — 3 месяца.

\*\* Начало доставки — следующий месяц за месяцем, в котором оплачена подписка.

\*\*\* Оформляя подписку, подписчик соглашается, что его персональные данные могут быть предоставлены третьим лицам для выполнения доставки издания.

Справки по телефону: +7 (495) 974-2260, доб. 1736; e-mail: podpiska@skpress.ru.

В случае если Вам не доставляют издания по подписке, сообщите об этом по e-mail: pretenzii@skpress.ru.

## Не пора ли...

◀ ПРОДОЛЖЕНИЕ СО С. 19

вернутся сферы производства, здравоохранения, торговли, административного управления и наблюдения за индивидуальными перемещениями. Уже сейчас быстрый рост данных наблюдается в сфере телевидения и киноиндустрии, что связано с бурным развитием специализированных цифровых технологий, а также в социальных сетях, фармации, генной инженерии и нефтедобыче.

Сергей Артёмов считает, что Big Data так же актуальны для России, как и для всего остального мира, но только к этой теме нужно относиться не как к неожиданно возникшей досадной проблеме, а как к новой возможности сделать свой бизнес более эффективным — использовать на полную катушку информацию, которая раньше не анализировалась или анализировалась частично. Это уже делают сотовые операторы, которые сейчас хранят и анализируют данные о звонках абонентов. На основе этой информации работает ряд информационных систем сотовых операторов: системы противодействия мошенничеству (Fraud Control), бизнес-аналитики и биллинговые системы.

В подобных решениях могут быть заинтересованы крупные онлайн-магазины, подобные Ozon.ru. По словам Сергея Артёмова, с помощью аналитики систем на базе Hadoop можно быстро формировать персональные рекомендации на базе совершенных ранее покупок для каждого клиента, а не просто предлагать сопутствующие товары. Также возможно рекомендовать другие варианты, которые почти наверняка заинтересуют покупателя, например новые книги авторов, которых ранее читал клиент, новые диски музыкальных групп, фильмы и т. д. Государственные заказчики также не исключение: на основе данных системы наблюдения за дорожным трафиком можно оперативно получать информацию о том, где камеры в последний раз зафиксировали требуемую машину, ее маршрут, характеристики машинного потока, среднюю скорость движения, прогноз плотности трафика.

### Проблемы на пути реализации идей Big Data

По мнению Александра Яковлева, не очень понятным остается вопрос форматов данных: “С одной стороны, мы имеем огромное количество различных типов хранимых данных, с другой — даже однотипные данные можно хранить в несовместимых форматах. Здесь мы уже приходим к постановке задачи о некоем свободном формате данных, когда на момент сбора данных еще неизвестно, в каких приложениях и для каких именно целей они будут впоследствии использоваться. Но при этом нельзя допустить низкую эффективность или сильную избыточность хранимых данных”.

Пока российские пионеры освоения новых технологий не озвучат экономических эффектов от их использования, все остальные будут выжидать, присматриваться. Говоря об этом, Максим Исаев проводит параллель с RFID:

“Всем понятны преимущества, много пилотных зон, но широко распространения технология не получила, потому что при нынешней цене на оборудование сфера ее применения ограничена. Но все опросы на Западе, да и у нас, показывают, что множество компаний считают неструктурированные данные и те возможности, которые дает работа с ними, очень важными для бизнеса. Поскольку сейчас быстрыми темпами развиваются подходы к работе с большими данными (MapReduce, Hadoop), к моменту, когда российский рынок будет готов к проектам, мы будем пользоваться уже апробированными, зрелыми технологиями”.

Он отметил еще один момент: работа с большими данными невозможна без соответствующей заточки и программной части, и аппаратной части (“одна рука не хлопает”). Поэтому те компании, кто занимается и “железом”, и ПО, быстро подхватили идею предоставления единого продукта, хотя для заказчика в таком подходе есть как минусы, так и плюсы.

Алексей Мещеряков сказал о том, что самым важным фактором для успеха больших данных является создание гибкой инфраструктуры, которая обеспечит правильное сочетание различных аспектов, затрагивающих данную технологию. “Прежде всего она должна опираться на бизнес-цели и бизнес-требования организации, — подчеркнул эксперт. — Необходимо учитывать такие важные составляющие, как доступ ко всем источникам данных в режиме, близком к реальному времени, и даже в реальном времени, поддержку различных типов устройств, управление данными, интегрированную аналитику и т. д.”.

Андрей Хромов указал на наличие пяти факторов, которые определяют корпоративную аналитику: 1) рост объема данных, необходимость в больших объемах памяти; 2) рост числа пользователей; 3) отсутствие границ для неструктурированных данных — многообразие данных становится стандартом; 4) скорость попадания данных в систему; 5) качество данных. Далее он пояснил: “Расширение традиционной инфраструктуры в направлениях обработки данных в оперативной памяти и хранения данных по столбцам позволит компаниям анализировать как структурированные, так и неструктурированные данные в единой консолидированной среде, а также обрабатывать их в реальном времени и реагировать с наименьшей задержкой на события. Особенно это касается сред с произвольными запросами и различными профилями пользователей и запросов. Кроме того, инвестиции в мобильную аналитику позволят бизнес-сообществу получить максимальную ценность от данных и повысить производительность своих работников. Внедрение новых программных моделей, например фреймворка MapReduce, а также поддержка аналитическим сервером парадигм, обеспечивающих массивно-параллельные и распределенные вычисления, таких как Hadoop, позволит создать более управляемую, интегрированную и доступную аналитическую среду”.

Ценность систем, основанных на технологиях Big Data, состоит в качестве и скорости предоставляемой аналитики. Отметив это, Михаил Горбатовский отметил: “Так как подобные аналитические задачи, выходящие за рамки хранилищ данных, ранее для заказчиков решались только в отдельных уникальных проектах, то для запуска проектов на новом наборе технологий нужна разработка новых промышленных моделей, показателей для конкретных типов клиентов. Кроме промышленной направленности требуется также креативный подход к исследованию данных — возможности по исследованию Big Data колоссальны”.

Андрей Вакалов уверен, что основная проблема — это дефицит квалифицированных ИТ-кадров, которых сейчас не так много как на российском рынке, так и на мировом.

“В компаниях, столкнувшихся с необходимостью обработки больших объемов данных, важно обеспечить возможность прозрачного масштабирования без прерывания их работы, — сказал Валерий Юринский. — В такой среде необходимо обеспечить возможность платить за ресурсы по мере роста при едином и простом управлении всеми элементами инфраструктуры. Поэтому мы считаем, что гибкая работа с большими данными невозможна без облачных хранилищ и облачных вычислений, причем в виде законченных решений. Однако сами по себе облачные технологии недостаточно эффективно работают с большими блоками информации. Очевидно, что требования к масштабированию систем хранения, аналитическим приложениям и соответствующим программно-аппаратным комплексам становятся все выше. Задача состоит в том, чтобы обеспечить максимальную производительность без кардинального наращивания серверов или дисковых массивов. Последние технологические достижения ведущих вендоров позволяют это сделать”.

Сергей Артёмов обратил внимание на то, что основной задачей клиента при решении проблем Big Data является выбор подходящей ему технологии: “Если организация не готова ждать адаптации своего ПО для работы с Hadoop или не имеет желания работать с СПО, то, скорее всего, взгляд будет направлен в сторону готовых систем. Если же речь идет о решениях на базе Hadoop, то на данном этапе я вижу три основные проблемы. Первая состоит в том, что эти системы не являются самодостаточным продуктом, подобно новому серверу или массиву. Таким образом, вариант “поставил новый, более мощный сервер и решил проблему” не работает. Второе — это правильное позиционирование систем. Их, разумеется, нельзя продвигать как универсальную замену больших баз данных, они конкурентоспособны в своей области приложения (анализ и обработка больших объемов разнородных данных). Наконец, подобные решения вышли на рынок относительно недавно и имеют ряд особенностей, которые надо учитывать при внедрении в промышленную эксплуатацию”.

## РАСПРОСТРАНЕНИЕ PC WEEK/RUSSIAN EDITION

**Подписку можно оформить в любом почтовом отделении по каталогу:**

• “Пресса России. Объединенный каталог” (индекс 44098) ОАО “АРЗИ”

**Альтернативная подписка в агентствах:**

• ООО “Интер-Почта-2003” — осуществляет подписку во всех регионах РФ и странах СНГ. Тел./факс (495) 580-9-580; 500-00-60; e-mail: interpochta@interpochta.ru; www.interpochta.ru

• ООО “Агентство Артос-ГАЛ” — осуществляет подписку всех государственных библиотек, юридических лиц в Москве, Московской области и крупных регионах РФ. Тел./факс (495) 788-39-88; e-mail: shop@setbook.ru; www.setbook.ru

• ООО “Урал-Пресс” г. Екатеринбург — осуществляет подписку крупнейших российских предприятий в более чем 60 своих филиалах и представительствах. Тел./факс (343) 26-26-543

**ВНИМАНИЕ!**  
Для оформления бесплатной корпоративной подписки на PC Week/RE можно обращаться в отдел распространения по тел. (495) 974-2260 или E-mail: [podpiska@skpress.ru](mailto:podpiska@skpress.ru), [pretenzii@skpress.ru](mailto:pretenzii@skpress.ru)  
Если у Вас возникли проблемы с получением номеров PC Week/RE по корпоративной подписке, пожалуйста, сообщите об этом в редакцию PC Week/RE по адресу: [editorial@pcweek.ru](mailto:editorial@pcweek.ru) или по телефону: (495) 974-2260. Редакция

(многоканальный); (343) 26-26-135; e-mail: info@ural-press.ru; www.ural-press.ru

**ПРЕДСТАВИТЕЛЬСТВО В МОСКВЕ**  
**ООО “УРАЛ-ПРЕСС”**

Тел. (495) 789-86-36; факс (495) 789-86-37; e-mail: moskva@ural-press.ru

**ПРЕДСТАВИТЕЛЬСТВО В САНКТ-ПЕТЕРБУРГЕ**  
**ООО “УРАЛ-ПРЕСС”**

Тел./факс (812) 962-91-89

**ПРЕДСТАВИТЕЛЬСТВО В КАЗАХСТАНЕ**  
**ООО “УРАЛ-ПРЕСС”**

тел./факс 8(3152) 47-42-41; e-mail: kazakhstan@ural-press.ru

• ЗАО “МК-Периодика” — осуществляет подписку физических и юридических лиц в РФ, ближнем и дальнем зарубежье.

Факс (495) 306-37-57; тел. (495) 672-71-93, 672-70-89; e-mail: catalog@periodicals.ru; info@periodicals.ru; www.periodicals.ru

• Подписное Агентство KSS —

осуществляет подписку в Украине. Тел./факс: 8-1038- (044)585-8080 www.kss.kiev.ua, e-mail: kss@kss.kiev.ua

**PCWEEK**  
RUSSIAN EDITION

№ 5  
(790)

БЕСПЛАТНАЯ  
ИНФОРМАЦИЯ  
ОТ ФИРМ!

ПОЖАЛУЙСТА, ЗАПОЛНИТЕ ПЕЧАТНЫМИ БУКВАМИ:

Ф.И.О. \_\_\_\_\_  
ФИРМА \_\_\_\_\_  
ДОЛЖНОСТЬ \_\_\_\_\_  
АДРЕС \_\_\_\_\_  
ТЕЛЕФОН \_\_\_\_\_  
ФАКС \_\_\_\_\_  
E-MAIL \_\_\_\_\_

1С .....1     IBM .....5  
 РОСКО .....3  
 ELKO GROUP ....24     PANASONIC .....9  
 HEWLETT-  
PACKARD .....7     VELES-DATA .....11

ОТМЕТЬТЕ ФИРМЫ, ПО КОТОРЫМ ВЫ ХОТИТЕ ПОЛУЧИТЬ ДОПОЛНИТЕЛЬНУЮ ИНФОРМАЦИЮ, И ВЫШЛИТЕ ЗАПОЛНЕННУЮ КАРТОЧКУ В АДРЕС РЕДАКЦИИ: 109147, РОССИЯ, МОСКВА, УЛ. МАРКСИСТСКАЯ, Д. 34, КОРП. 10, PC WEEK/RUSSIAN EDITION; или по факсу: +7 (495) 974-2260, 974-2263.