

"Radio" is monthly publication on audio, video, computers, home electronics and telecommunication

УЧРЕДИТЕЛЬ И ИЗДАТЕЛЬ: ЗАО «ЖУРНАЛ «РАДИО»

Зарегистрирован Министерством печати и информации РФ 01 июля 1992 г.
Регистрационный ПИ № ФС77-50754

Главный редактор В. К. ЧУДНОВ

Редакционная коллегия:

А. В. ГОЛЫШКО, А. С. ЖУРАВЛЁВ, Б. С. ИВАНОВ,
С. Н. КОМАРОВ, А. Н. КОРОТКОНОШКО, К. В. МУСАТОВ,
И. А. НЕЧАЕВ (зам. гл. редактора), Л. В. МИХАЛЕВСКИЙ,
С. Л. МИШЕНКОВ, О. А. РАЗИН, Б. Г. СТЕПАНОВ
(первый зам. гл. редактора), В. В. ФРОЛОВ

Выпускающие редакторы: С. Н. ГЛИБИН, А. С. ДОЛГИЙ

Обложка: В. М. МУСИЯКА

Вёрстка: Е. А. ГЕРАСИМОВА

Корректор: Т. А. ВАСИЛЬЕВА

Адрес редакции: 107045, Москва, Селиверстов пер., 10

Тел.: (495) 607-31-18. Факс: (495) 608-77-13

E-mail: ref@radio.ru

Группа работы с письмами — (495) 607-08-48

Отдел рекламы — (495) 608-99-45, e-mail: advert@radio.ru

Распространение — (495) 608-81-79; e-mail: sale@radio.ru

Подписка и продажа — (495) 607-77-28

Бухгалтерия — (495) 607-87-39

Наши платёжные реквизиты:

получатель — ЗАО "Журнал "Радио", ИНН 7708023424,
р/сч. 40702810438090103159

Банк получателя — ОАО "Сбербанк России" г. Москва
корр. счет 3010181040000000225 БИК 044525225

Подписано к печати 18.09.2013 г. Формат 60×84 1/8. Печать офсетная.

Объём 8 физ. печ. л., 4 бум. л., 10,5 уч.-изд. л.

В розницу — цена договорная

Подписной индекс:

по каталогу «Роспечати» — 70772;

по каталогу Управления федеральной почтовой связи — 89032;

по каталогу Российской прессы ПОЧТА РОССИИ — 61972.

За содержание рекламного объявления ответственность несёт рекламодатель.

За оригинальность и содержание статьи ответственность несёт автор.

Редакция не несёт ответственности за возможные негативные последствия использования опубликованных материалов, но принимает меры по исключению ошибок и опечаток.

В случае приёма рукописи к публикации редакция ставит об этом в известность автора. При этом редакция получает исключительное право на распространение принятого произведения, включая его публикации в журнале «Радио», на интернет-страницах журнала, CD или иным образом.

Авторское вознаграждение (гонорар) выплачивается в течение одного месяца после первой публикации в размере, определяемом внутренним справочником тарифов.

По истечении одного года с момента первой публикации автор имеет право опубликовать авторский вариант своего произведения в другом месте без предварительного письменного согласия редакции.

В переписку редакция не вступает. Рукописи не рецензируются и не возвращаются.

© Радио®, 1924—2013. Воспроизведение материалов журнала «Радио», их коммерческое использование в любом виде, полностью или частично, допускается только с письменного разрешения редакции.

Отпечатано в ЗАО «ПОЛИГРАФИЧЕСКИЙ КОМПЛЕКС «ЭКСТРА М»,
143400, Московская обл., Красногорский р-н, а/м «Балтия», 23 км.
Зак. 13-09-00314.



Компьютерная сеть редакции журнала «Радио» находится под защитой Dr.Web — антивирусных продуктов российского разработчика средств информационной безопасности — компании «Доктор Веб».

www.drweb.com

Бесплатный номер
службы поддержки
в России:
8-800-333-79-32

Big Data, или как справиться с цунами информации

А. ГОЛЫШКО, канд. техн. наук, г. Москва

*"Упрощать — сложно,
а усложнять — просто".*

Откуда это?

В этом году на страницах журнала мы рассказывали о виртуальных сетях, дополненной реальности, 3D-моделировании, растущем взаимодействии по линиям "человек—машина" и "машина—машина" и "Всеохватывающем Интернете" (IoE), который имеет гораздо больший охват, чем "Интернет вещей" (IoT). Каждое из этих явлений — настоящий технологический и сервисный прорыв прогрессивного человечества, следствием которого является генерация всё новых и новых информационных потоков или просто потоков данных. Ежесекундно в мире производится и потребляется огромный объём информации. Рост данных уже сейчас серьёзным образом влияет на деятельность бизнеса — банков, промышленных и производственных компаний, торговых предприятий (и не только), которые ищут решения, способные подготовить бизнес к постоянно растущим потребностям. И всё это, собственно, последствия всеобщей цифровизации. Казалось бы, ну чего ещё надо? Дело за малым — нужно суметь обработать все эти вновь сгенерированные данные, чтобы как минимум как-то воспользоваться полученной информацией. А иначе, зачем всё это?

Следует сразу заметить, что разговор идёт отнюдь не о хранении данных, а об их обработке. Электронные хранилища данных разработаны и продолжают развиваться в огромном разнообразии, масштабах и объёмах, но как достать нужную вам информацию, перелопатив колоссальные массивы данных? Вот тут-то и возникло понятие Больших Данных или Big Data, под которыми понимаются очень большие объёмы слабоструктурированных данных самого разнообразного вида.

А данные, зная себе, генерируются в невообразимых объёмах: в каждой компании их источниками может являться что угодно: персонал и его действия, сети, клиенты, компьютеры, всевозможные сенсоры и сенсорные сети, а их счёт идёт на многие миллиарды. Множатся различные мобильные приложения, от местоопределения до медицины. И в этих данных нетрудно и "утонуть". Вот тут-то и возникает необходимость обращения с накопленной (хотя бы корпоративной) информацией как с Большими Данными, которые требуют определённой технологии обработки, хранения, анализа и защиты. Наконец, работа с ними стала сегодня ещё активнее рассматриваться в качестве катализатора уже известных нам "облачных" инфраструктур. Большие Данные продолжают быстро расти — по оценкам IBM за последнюю пару лет было создано 90 % мировой информации.

Конечно, основным источником и даже потребителем Больших Данных являются преимущественно машины, поскольку рост населения планеты и его потребности растут отнюдь не в соответствии с законом Мура, которому пока подчиняются, к примеру, компьютерные системы. За последние десять лет население Земли увеличилось всего на 20 %, в то время как число транзисторов (а вместе с ними и ёмкость жёстких дисков) увеличилось более чем на 2000 %.

Существует важная разновидность данных, сулящих ощутимые преимущества, но пока почти не востребованных. Речь идёт о данных, получаемых в реальном времени из разных источников (от различных устройств, датчиков, видеосистем и т. п.) и требующих быстрой практической реакции. Они, собственно, и приносят наибольшую пользу, если используются в режиме реального времени. Поэтому одной из разновидностей Больших Данных стали так называемые быстрые данные. Во многих случаях обрабатывать базы данных надо "на лету", пока они не утратили актуальности.

ИНФОРМАЦИОННАЯ ПОДДЕРЖКА — КОМПАНИЯ «РИНЕТ»



Телефон: (495) 981-4571

Факс: (495) 783-9181

E-mail: info@rinet.ru

Internet Service Provider

Сайт: <http://www.rinet.net>

Типовые задачи — клиентская аналитика (точечный маркетинг), противодействие мошенничеству, оптимизация настроек оборудования.

В частности, компания Cisco называет эти данные Data in Motion (в буквальном переводе — данными в движении). Сеть связи может дополнить данные этого типа полезной контекстуальной информацией: сведениями о местоположении человека или устройства, идентификационной информацией, данными о доступности или недоступности. Приложения могут использовать эти данные для принятия решений и выполнения практических действий, необходимых в данный момент, а также для прогнозирования будущих событий. Технология Data in Motion может принести особую пользу в системах автоматизации производства для коммуникаций "машина—машина", где её можно использовать с целью оптимизации производственных процессов. Согласно опубликованному компанией Cisco прогнозу роста глобального мобильного трафика на ближайшие пять лет, к 2017 г. в мире будет работать более 1,7 млрд соединений "машина—машина".

Да что там машины, многие из нас сталкиваются с ситуацией, когда накопленные дома архивы с фото, видео, музыкой достигли таких объёмов, что стали неуправляемыми, и вы не можете быстро найти что-либо нужное (если вообще можете его найти). Горы DVD заполонили шкафы, масса информации записана на флешках, серверах Google и пр. И информационные технологии пока не могут предоставить пользователю удобные и понятные инструменты для работы с персональными архивами, хотя предложений с обещаниями моментально найти всё, что вам нужно, предостаточно. Но в очередной раз купившись на "выгодное" предложение, пользователь снять проблему не может.

С другой стороны, во многих крупных научных экспериментах возникает необходимость работы с беспрецедентными объёмами данных и на грани возможностей, которые предоставляют компьютерная индустрия и информационные технологии. Взять хотя бы Большой Адронный Коллайдер или Большой Обзорный Синоптический Телескоп (LSST) с петабайтами информации. К примеру, детектор ATLAS, предназначенный для поиска бозона Хиггса, выдаёт около петабайта "сырых" данных в секунду. Обработку таких объёмов и с такой скоростью современные информационно-коммуникационные технологии (ИКТ) обеспечить не могут, и в результате данные обрабатываются только выборочно, а "неинтересные события" фильтруются. В результате получается около 100 МБ в секунду или порядка петабайта в год.

Современный бизнес, особенно ИКТ-бизнес, также сталкивается с проблемами хранения, обработки и анализа гигантских объёмов данных. В частности, серверы компании Google обрабатывают примерно 1 ПБ информации в час. С тем же самым сталкиваются крупные торговые интернет-площадки и социальные сети. С проблемой Больших Данных работают и корпоративные аналитики, и создатели электронного государства.

В частности, планирующийся к запуску в 2015 г. гигантский радиотелескоп SKA с суммарной площадью антенн 1 км² будет передавать сигналы с одной антенны со скоростью 160 Гбит/с, а ожидаемый объём получаемых "сырых" данных в день оценивается в 1 ЭБ (эксабайт), которые сожгут до 10 ПБ для дальнейшего хранения. Для обработки этой информации в реальном масштабе времени необходимы будут вычислительные мощности около 100 Тфлопс (терафлопс). Просто анализ научных данных не требует учёта их изменчивости, когда нужно принять решение на основе только что поступившей информации. Зато, например, на рынке ценных бумаг, особенно в условиях высокой конкурентности запросов, нужно "шевелиться" гораздо активнее.

Когда говорят о Больших Данных, то зачастую имеют в виду даже не обеспечение какой-либо интерактивной работы с данными, а технологический программный доступ к этим данным, чтобы можно было автоматизировать рутинную обработку каких-либо наблюдений или поиска, чтобы машины могли работать с машинами. Автоматизация работы с данными сопряжена с рядом трудностей, в первую очередь связанных с необходимостью стандартизации описания данных (метаданными) и их форматов. Зачастую проблемы в организации межмашинного взаимодействия возникают из-за неточной и некорректной информации, когда приходится работать с данными, чьи значения известны неточно, например, с результатами измерений. К примеру, сенсоры и любые другие датчики всегда производят данные с некоторой погрешностью. Ну а гигантские сенсорные сети, вроде тех, что помогают следить за погодой, производят настоящее "цунами" данных, которые надо уметь не только хранить и обрабатывать, но и анализировать, а также принимать решения — и всё это с учётом "неточности" как оригинальных данных, так и процедур обработки.

Порой мы любим рассуждать о компьютерах или даже искусственном интеллекте, который чуть ли не завтра будет принимать различные решения от прогноза погоды и готовки еды до управления атомными электростанциями и медицинскими процедурами, не говоря уже о проведении хирургических операций. Однако, оказывается, все эти решения приходится принимать на основе изначально не совсем точных данных. С этим довольно легко справляется человек, но попробуйте добиться верного ответа от компьютера. И главное, сумеете ли поверить, что ответ правильный и довериться машине. В связи с этим хорошо известно, что такие популярные у специалистов задачи, как сведение данных, полученных из разных источников (Data Cleaning, Data Merging, Deduplication), требуют особых методов анализа в случае неточных данных, особенно для данных петабайтных объёмов.

Вот, к примеру, недавно компания IBM объявила о создании хранилища данных размером 120 ПБ для задач предсказания ураганов. Теперь, как говорится, дело за малым — надо научиться более-менее точно предсказы-

вать ураганы. И, как мы регулярно видим по телевизору или в Интернете, тут отнюдь не всё ещё получается.

Осознание указанных изменений в науке и бизнесе привело к пониманию того, какой должна стать новая архитектура информационных систем, чтобы справиться с этими "цунами" данных. В первую очередь, это должен быть кластер независимых многоядерных компьютеров, имеющих собственные диски, соединённые быстродействующей сетью, на котором должна быть установлена система управления базами данных (СУБД), ориентированная на работу с распределёнными данными очень больших объёмов и их параллельную обработку.

Терминология

В общем виде термин Big Data появился для описания той ситуации, когда данные становятся большой проблемой. Разумеется, всё относительно, и ваш персональный архив — это тоже Большие Данные, но для вас лично. Зато для какой-нибудь научной организации или корпорации это будет уже несколько десятков терабайт. Принято также использовать этот термин для описания данных, объём которых превышает возможности доступных на данный момент технологических средств и когда требуется разработка специальных инструментов. Например, для данных астрономического проекта Pan-STARRS используется специально разработанная для него масштабируемая кластерная архитектура GrayWulf, а крупные интернет-компании используют "домашние" (разработанные внутри компании) базы данных.

Несмотря на то что, как утверждает Wikipedia, термин Big Data появился достаточно давно, заметный информационный шум по этому поводу начался где-то в 2008 г., когда несколько уважаемых изданий, таких как "Nature" (2008 г.), "CACM" (2008 г.), "The Economist" (2010 г.), "Science" (2011 г.), посвятили специальные номера проблеме Больших Данных.

Ну а "полное признание" проблемы Big Data настало после упоминания в вышедшем в 2011 г. традиционном исследовании компании Gartner "Hype Cycle for Emerging Technologies, 2011" нового тренда — Big Data and Extreme Information Processing and Management. Там говорилось, что процесс оформления нового ИКТ-явления находится ещё на участке первоначального роста, а рынок даже ещё не до конца понял, что можно ожидать от этой технологии, и что многие, возможно, возлагают на неё слишком большие надежды.

Термин Big Data, конечно же, не совсем точен, поскольку непосредственно указывает только на объём данных, в то время как необходимо учитывать такие факторы, как большее разнообразие данных, скорость их поступления и изменения, требование быстрого (в реальном времени) ответа. Именно поэтому Gartner использовала более длинное название Big Data and Extreme Information Processing and Management. Но народ уже привык к короткому названию.



Государственные интересы

Правительства многих развитых стран сегодня уже хорошо понимают роль и значение национальной информационно-коммуникационной инфраструктуры в экономическом и социальном развитии государства. В связи с этим в ИКТ-отрасли появилась примечательная тенденция, которая предоставляет госучреждениям отличную возможность для развития целевой доставки услуг компаниям и физическим лицам. Вот, в частности, три особенности Больших Данных, которые уже сегодня оказывают влияние на государственные организации во всех регионах мира.

Объём. Каждый случай взаимодействия с государственной организацией порождает сетевой трафик и цифровые записи, занимающие определённое место в системах хранения. По прогнозам, за период 2010—2015 гг. годовые темпы роста глобального пользовательского и корпоративного трафика вырастут соответственно на 36 и 22 %.

Скорость сбора данных непрерывно растёт. Вот лишь один пример: британское правительство решило принимать налоговые декларации от работодателей в реальное время раз в месяц, а не раз в год, как раньше.

Разнообразие. Сегодня, помимо традиционных документов и форм, госучреждениям приходится иметь дело с огромными потоками неструктурированной информации (видеоматериалы от систем безопасности, материалы социальных сетей и т. д.). Ещё одну проблему создаёт разнообразие каналов, по которым граждане обращаются в государственные учреждения.

Инновационные возможности для правительства создают не сами данные, а анализ огромных объёмов данных разного типа и получение на этой основе полезной информации и новых идей. Благодаря технологии Больших Данных с помощью анализа потоков городского транспорта и управления сигналами светофоров в реальное время государство может, например, сократить для своих граждан время, затрачиваемое на поездки на работу. Эта технология поможет государственным учреждениям быстро выявлять и подавлять эпидемические очаги распространения гриппа и других серьёзных болезней, а также предотвращать распространение импортных продуктов питания и пр.

Кое-кому стратегия Больших Данных пока кажется малореальной, но реальность такова, что некоторые правительства уже подают примеры использования данных, поступающих из разнообразных источников, для принятия даже политических решений. Вместе с тем изолированный характер этих попыток увеличивает стоимость их реализации и поддержки. Поэтому настёт удобный момент для того, чтобы сделать технологию Больших Данных главным вопросом при обсуждении путей строительства государственного "облака".

Кое-что из продуктов

Корпорация IBM анонсировала две новые технологии, которые помогут

предприятиям в обработке больших объёмов данных. Они позволят сделать их анализ проще, быстрее и экономичнее. В частности, BLU Acceleration позволяет пользователям гораздо быстрее получить доступ к важной информации, повышая качество принимаемых решений. Это ПО расширяет возможности традиционных систем, работающих в оперативной памяти (которые позволяют загружать данные в ОЗУ вместо записи на жёсткий диск, что увеличивает производительность обработки). Характерная для таких систем производительность обеспечивается даже в том случае, когда объём наборов данных превышает размер ОЗУ. Во время тестирования некоторые запросы типовых аналитических задач выполнялись более чем в 1000 раз быстрее при использовании всех инноваций BLU Acceleration, сообщает IBM.

IBM PureData System for Hadoop призван упростить и ускорить развёртывание Hadoop на предприятиях. Hadoop — это программная среда с открытым исходным кодом для организации и анализа больших объёмов структурированных и неструктурированных данных, таких как сообщения в социальных сетях, цифровые изображения и видео, онлайн-новые транзакции и данные о местонахождении сотовых телефонов.

Новая система способна сократить с недель до минут срок освоения организациями технологии корпоративного класса Hadoop с её простыми в использовании инструментами анализа и визуализации, которые могут понадобиться как бизнес-аналитикам, так и специалистам по данным. Помимо этого, она предоставляет расширенный инструментарий для работы с Большими Данными, позволяющий осуществлять мониторинг, разработку и интеграцию со многими другими корпоративными системами.

Корпорация Oracle планирует выпустить набор приложений, который будет использовать преимущества концепции in-memory, позволяющей очень быстро проводить операции с данными и работать с очень большими объёмами информации практически в режиме реального времени. Параллельно с Oracle концепция in-memory в своих продуктах также развивает и компания SAP. Компания разрабатывает более десятка приложений, использующих этот подход. Все они будут работать на аппаратном обеспечении Oracle Exadata и Exalogic, новые версии которых были презентованы в прошлом году. За счёт того что в продуктах in-memory значительная часть данных и транзакций хранится непосредственно в оперативной памяти, Oracle обещает почти 20-кратный прирост скорости работы программного обеспечения. Также в компании говорят, что новинки будут адаптированы для использования на современных многоядерных чипах и сверхскоростных сетевых интерфейсах, таких как Infiniband или Gigabit Ethernet.

Корпорация Hewlett-Packard представила HAVen — инновационную платформу Big Data, которая использует программное, аппаратное обеспечение и услуги этой компании, позволяя за-

казкику выстроить современную платформу анализа Больших Данных. Первой интегрированной платформой, созданной на базе технологии HAVen, стало решение HP Operations Analytics, которое позволяет пользователю получить всестороннее представление о своих ИТ-операциях, чтобы гарантировать соответствие заявленному качеству. HP Operations Analytics помогает эффективно анализировать огромные массивы данных, поступающих из самых разных решений Hewlett-Packard, включая HP ArcSight Logger и продукты линейки HP Business Service Management, а также из решений других производителей.

Подразделение технологических услуг этой корпорации расширяет пакет услуг Big Data Consulting Practice, призванный помочь клиентам обеспечить необходимую производительность ИТ-инфраструктуры для поддержки проектов Больших Данных. Новые услуги по формированию ИТ-стратегии, построению архитектуры и инфраструктуры систем, а также защите данных позволяют синхронизировать бизнес и ИТ, гарантировать соответствие стандартам и увеличить прибыль.

Продуктами для Больших Данных занимаются также VMware, Teradata, Dell, EMC, Cisco, Microsoft, Accenture и ещё десятки компаний.

Проблемы Big Data

Одной из проблем Больших Данных является разное понимание самого понятия, которое чаще ассоциируется с модной тенденцией, а не с реальной методологией. К тому же сегодня нужно "продавать" не саму технологию, а её ценность для бизнеса, и с осознанием данного факта пока не у всех хорошо, что является лишь отражением незрелости рынка.

Важнейшей задачей в эпоху Больших Данных является обучение специалистов, способных эффективно использовать достижения ИТ для работы с такими данными. Однако таких специалистов не хватает, что показывает постоянно увеличивающийся спрос на них. Для науки, в частности, это означает, что учёный должен не только уметь ставить научную задачу, анализировать результаты, но и разбираться в таких ненаучных вопросах, как форматы данных, язык запросов, алгоритмы, используемые для анализа. В реальности любой научный коллектив на сегодня обязан иметь специалистов по данным, которые планируют такие работы, как организация хранения и безопасности информации, описание и обработка сырых данных, доступ, визуализация данных и результатов. Даже если вдруг на некотором этапе окажется, что, наконец-то, реализован набор инструментов, позволяющий учёным эффективно работать с результатами экспериментов, то растущие запросы учёных бросят новые вызовы технологии и потребуются следующий этап разработки новых инструментов, модификация старых и т. д., а это обычно приводит к построению уникальной, нетиражируемой технологии, что мало кому интересно.

Применение

Способность захватывать, хранить и анализировать данные, а затем применять добытую информацию к нужным процессам в реальном времени требует тонко отлаженных платформ Больших Данных и аналитики. Остановимся лишь на двух примерах.

BNSF Railway Company — один из крупнейших в Северной Америке железнодорожных грузоперевозчиков, применяет **IBM BLU Acceleration**, чтобы быстрее разобратся с огромным объемом данных, поступающих с 1700 серверов, отслеживающих выполнение программ технического обслуживания, погоду, составление графиков, состояние складских запасов, безопасность, поставки и многое другое. **BNSF** ежедневно использует свыше 1400 поездов, которые проходят 32,5 тыс. миль в 28 штатах США и двух канадских провинциях.

А вот компании **Ford Motor** использование аналитики с Большими Данными обусловило прибавку в 100 млн долл. США к годовой прибыли. Команда приглашенных ИТ-специалистов задействовала комбинацию средств работы с Большими Данными и бизнес-аналитическими приложениями, формируя наборы данных из исходно разносортной информации, связывая бизнес-процессы с данными газетной обложки, маркетинга и т. п. Масса важной информации получалась путём переработки сообщений пользователей внутренних и внешних социальных сетей и определения, как они связаны с бизнес-процессами и как их можно использовать. Сбор гигантских объёмов данных, генерируемых транспортными средствами, можно использовать не только для улучшения дизайна автомашин, но и для того, чтобы понять, какие типы данных можно собирать дополнительно. Последнее имеет особенно большое значение, так как производители автомобилей устанавливают в машины всё больше разных датчиков, чтобы отслеживать их эксплуатационные характеристики, повышать уровень техобслуживания и совершенствовать проекты будущих моделей. Например, датчики гибридного автомобиля **Ford Fusion Energi** с электрической подзарядкой от внешней сети генерируют и записывают 25 Гб разнообразных данных в час — от температуры двигателя, скорости и нагрузки на автомобиль до дорожных условий и общей рабочей эффективности. Этот поток данных может возрастать до 4 Тб в час при тестировании с использованием специальных инструментов, хотя, по словам специалистов, через несколько лет эти инструменты могут превратиться в стандартное оборудование.

И ещё один пример

Программа сбора информации от спецслужб США, известная как **PRISM**, стала всемирной сенсацией. Однако за шумихой в СМИ и истерикой в гражданском обществе необходимо разглядеть главное — реализация этой программы стала как минимум подтверждением базовых концепций и технологий, лежащих в основе обработки Больших Данных.

Не секрет, что некоторые из самых уважаемых имён в мире технологий имеют давние глубокие связи с военно-промышленным комплексом США. Ещё в 70-е годы прошлого века программа ЦРУ по сбору и обработке информации выросла за рамки ведомства и изменила мир технологий. Эта программа привела к появлению реляционных (от англ. relation — отношение) баз данных, к созданию компании и одноимённого продукта под названием **Oracle**. Это название со временем даже превратилось в синоним базы данных. Предложенная концепция радикально отличалась от существовавших ранее подходов и стала гигантским шагом вперёд в управлении, обработке и представлении данных. Даже намёк на оракула (**oracle**) подразумевал способность предугадывать будущее.

Реляционные базы данных хорошо подходят для хранения ассоциируемых элементов с высокой степенью структуризации, но благодаря повсеместной распространённости стали местом, где хранится всё, независимо от своих структурных или реляционных свойств. Эти базы данных стали основной средой совместно используемого хранения данных даже там, где они, очевидно, малопригодны, например, в сфере динамики изменения рыночных цен. И активные пользователи сегодня всё больше чувствуют ограниченность такого подхода, ибо для него нужны затраты на оборудование, ПО, сеть, обслуживающий персонал. Кроме того, эта модель ограничена в масштабировании.

По некоторым данным, за **PRISM** стоят те же организации, которые выступали инициаторами ряда прошлых технологических импульсов, в том числе и реляционных баз данных. Если до истории с **PRISM** думалось, что Большие Данные являются сферой интересов только Интернет-гигантов или компаний, связанных с социальными сетями, то теперь технологии замахиваются на гораздо большее. И мир видит всё больше и больше примеров использования в реальном мире приложений для Больших Данных. Просто в случае с **PRISM** это использование очень конкретно и очень серьёзно.

Специалистам понятно, что **PRISM** в широком масштабе использует и "облачные" технологии (массово распределённая система, абстрагирование от аппаратуры, использование стандартных компонентов) и технологии Больших Данных (**Hadoop**, обучение машины и распознавание образов). Там должна присутствовать массивно распределённая файловая система, способная вмещать огромные объёмы неструктурированных данных и обеспечивать их быструю параллельную обработку. Подобная платформа должна быть самоисправляемой, горизонтально масштабируемой и построенной из серийно выпускаемых компонентов.

Всё это говорит о том, что Большие Данные вышли за рамки периода исследования или экспериментального внедрения. Понятно, что как используемая платформа, так и получаемая из неё информация обладают реальной ценностью для заинтересованных лиц, и поэтому (если отвлечься от сопутствующих юридических и этических проблем)

само физическое существование **PRISM** является жестоким, но весьма убедительным подтверждением жизнеспособности концепции Больших Данных.

До недавнего времени многие считали, что указанная выше программа технически нереализуема. И вот наступили другие времена. И **Big Data** теперь это не проблема, это — возможность.

По материалам **CNews**, **PCWeek**, **Суперкомпьютер**, **Cisco**, **IBM**, **Oracle**, **IBS**, **HP**, **Computerworld**.

МОДУЛЬНАЯ РЕКЛАМА

Модули и наборы от Ekits:

— Цифровой вольтметр + амперметр постоянного тока **EK-SVAL0013** до 100 В, до 50 А: белая подсветка, негатив — 820 руб., без подсветки, позитив — 640 руб., белая подсветка, позитив — 840 руб.

— Цифровой вольтметр + амперметр постоянного тока **EK-SVAL0013** до 100 В, до 10 А: без подсветки, позитив — 626 руб., белая подсветка, негатив — 810 руб., белая подсветка, позитив — 830 руб.

— Набор электролитических конденсаторов, 12 номиналов, всего 108 шт. **EK-C/ELECTR** — 560 руб.

— Набор выводных керамических конденсаторов, 40 номиналов (от 1 pF до 0,1 mF), каждого по 20 шт., всего 800 шт., **EK-C RADIAL** — 510 руб.

— Набор резисторов: 171 номинал, каждого по 20 резисторов, **EK-R20** — 1400 руб.

FCLG-meter — универсальный измеритель частоты, ёмкости, индуктивности и напряжения (по мотивам **sqham.ru**), собранная плата с индикатором и корпусом — 2550 руб.

Измеритель ёмкости и последовательного эквивалентного сопротивления электролитических конденсаторов **C/ESR-meter** — 1140 руб.

SC Analyzer 2005 — 890 руб.

Бесплатные макетные платы в широком ассортименте и перемычки к ним.

А также:

— **EK-R0603/170** — набор ЧИП резисторов (единицы Ом — единицы МОм), типоразмер 0603, 170 номиналов по 24/25 шт. — 950 руб.

— Набор ЧИП резисторов, типоразмер 1206, **EK-R1206/168** — 950 руб.

— Набор ЧИП резисторов, типоразмер 0805, **EK-R0805/169** — 820 руб.

— **ХИТ!** Набор деталей **ALX007** для сборки термостата на DS18B20 и ATmega8 — 640 руб.

— **Программатор** PIC-контроллеров и ¹C (IIC) EEPROM EXTRA-PIC — 850 руб.

— **ХИТ!** Набор "Частотомер 10 Гц — 250 МГц" — 650 руб.

— Цифровая шкала трансивера — 850 руб.

ЗВОНИТЕ! ЗАКАЗЫВАЙТЕ! По бесплатному междугородному номеру: 8-800-200-09-34 с 9-30 до 18-00 MSK, по e-mail: zakaz@dessy.ru или на сайте www.dessy.ru