

Открытые системы

№10
2011

ISSN 1028-7493

ИТ для бизнеса —
архитекторам
информационных систем

www.osmag.ru

СУБД



БОЛЬШИЕ ДАННЫЕ

СУБД для Больших Данных

- Бертран Мейер о надежном программировании •
- «Скорые» методы в ФБР • Данные по требованию
- «Большая семерка» ОС. Прогноз – 2012

Цикл конференций для руководителей и ведущих специалистов современных предприятий



**ОТКРЫТЫЕ
СИСТЕМЫ**
Open Systems Publications



Регистрация:
Тел.: (495) 956-3306, e-mail: kon@osp.ru
<http://www.ospcon.ru>

Издательство «Открытые системы» выпускает 14 различных журналов и газет, которые выходят ежегодным тиражом более 10 миллионов экземпляров. Лидирующие позиции на рынке СМИ обеспечиваются широким спектром изданий для специалистов и руководителей из таких отраслей, как информационные технологии, нефтегазовая отрасль, телекоммуникации, полиграфия и медицина. Активное сотрудничество с мировыми издателями, такими как IDG, PenWell, позволяет нам поддерживать высочайший уровень и обеспечивать актуальность публикаций. Наличие более 30 региональных представительств дает возможность издательству распространять свои журналы по всей территории России и СНГ.

Агентство корпоративных коммуникаций OSP-Con – подразделение издательства «Открытые системы», в которое в 2004 году был выделен весь бизнес, связанный с организацией и проведением мероприятий. Сегодня OSP-Con является организатором многочисленных конференций для отечественного ИКТ и бизнес-сообщества. Основной акцент в своей деятельности OSP-Con делает на детальную проработку контентной составляющей проводимых мероприятий, начиная с определения тематики конференций и кончая предоставлением ее участникам собственной аналитической экспертизы. Именно это обстоятельство сделало конференции, проводимые OSP-Con, заметным явлением на российском рынке и позволило агентству занять на нем ведущие позиции.

ГРАФИК ТЕМАТИЧЕСКИХ КОНФЕРЕНЦИЙ

Агентства корпоративных коммуникаций OSP-Con на бизнес-сезон 2012 г.

22	марта	Форум Big Data 2012, Международный информационно-выставочный центр «ИнфоПространство»
12	апреля	Форум «Бизнес-Видео 2012», гостиница «Radisson-Славянская»
18	апреля	СIO-Форум, гостиница «Radisson-Славянская»
24	мая	Форум Strategic IT Governance, гостиница «Radisson-Славянская»
30-31	мая	Форум «Мир ЦОД 2012» (DC World-2012), гостиница «Radisson-Славянская»

Письмо Деду Морозу

Многочисленные исследования аналитиков и игроков ИТ-рынка свидетельствуют, что одна из главных нынешних проблем компаний и организаций – переизбыток информации, причем высказывается даже мнение, что человечество производит сейчас данных больше, чем может хранить. Что значит для общества проблема Больших Данных и какие пути ее решения предлагают сегодня субъекты индустрии ИТ?

Большие Данные обычно определяют как сумму структурированных и неструктурированных данных постоянно растущих объемов, а также методологий, инструментов и методов обработки данных в распределенной вычислительной сети с учетом многообразия форматов и источников этих данных. Целью обработки является получение понятных человеку результатов, обобщающих поведение изучаемых объектов и позволяющих его прогнозировать. Ключевой момент здесь, как отмечают авторы этого номера журнала, посвященного практике Больших Данных, – непрерывный прирост данных. Однако аналитики не любят, когда данные меняются во время выполнения запроса, поскольку изменения могут исказить результат, следовательно, надо уметь анализировать их «мгновенно», за время, сравнимое со временем их поступления.

Неверно характеризовать Большие Данные лишь их объемом – любой фиксированный объем хранения, вчера казавшийся огромным, сегодня помещается в мобильном устройстве, а послезавтра, как отмечает Кирилл Вахрамеев, его не хватит и на один фрагмент семейной хроники. Большие Данные – типичный пример перехода количества в качество, это не определенное количество байтов, а, по аналогии с Большой Водой или Большой Нефтью, такой объем, который трудно хранить традиционными способами, и поступает он таким потоком, что его трудно принимать, формализовывать и обрабатывать. Чтобы проводить анализ, нужно уметь извлекать данные из хранилища быстрее, чем они поступают, а это возможно, только если изменить подходы к хранению и

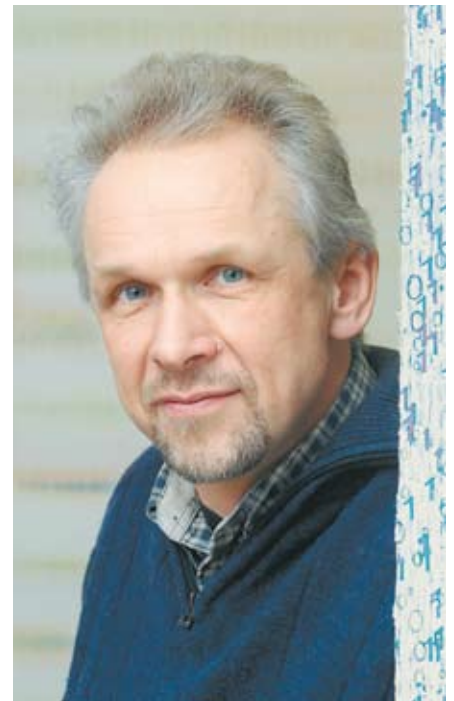
обработке, однако всего этого сегодня не рынке еще нет.

Как следствие, ценная информация либо теряется, либо устаревает еще до того, как оказалась предоставленной руководителям. Согласно наблюдениям, проведенным ведущими ИТ-производителями, треть менеджеров считает, что более половины информации внутри компаний и организаций вообще никак не используется – недоступно и не связано с другими данными. Каждый третий директор уверен, что проведению эффективной информационной стратегии в его компании мешает неразбериха в корпоративных данных. Мало того, только 2% руководителей уверены, что по их запросам всегда доставляется достоверная и актуальная информация, однако более половины топ-менеджеров считает, что залог успеха их деятельности в эффективной работе с информацией. Такой процент, как отмечает в своей статье Леонид Черняк, свидетельствует о том, что работающих технологий поддержки принятия решений и приложений, критически важных для эффективного управления предприятиями, на самом деле нет, а все предлагаемые ИТ-продукты, по сути, предназначены лишь для автоматизации рутинных операций.

Вывод неутешительный, но вместе с тем уже намечились пути создания платформ для разработки приложений, материализующих предприятие, управляемое в реальном времени. Похоже, что решение проблемы Больших Данных будет возможно уже теми средствами, о которых индустрия узнает в 2012-м и последующих годах.

Итак, какие просьбы Деду Морозу от ИТ-индустрии могли бы в канун нового, 2012 года сформулировать пользователи информационных систем, требующие актуализации информации в соответствии с запросами бизнеса, обеспечения защиты данных и разработки сбалансированной политики работы компаний с разными проявлениями публичной активности, в частности с социальными сетями?

Сделай так, чтобы можно было эффективно работать с неструктурированными данными (аудио, видео, сообщения из социальных сетей и блогов, электрон-



ная почта, телевизионные передачи, презентации в различных форматах и т. п.), составляющими сегодня 85% всего объема генерируемых данных и ежегодно растущими на 62% (примечательно, что количество структурированных данных увеличивается лишь на 22% в год).

Сделай так, чтобы бизнес-аналитика вообще и прогнозная аналитика в частности обогатились инструментами работы с Большими Данными.

Сделай так, чтобы облака действительно превратились в коммунальную услугу, интегрировав в себя все необходимые для этого ресурсы.

Уповать на Деда Мороза, конечно, можно, но для воплощения всех этих пожеланий в жизнь кроме технологий потребуются еще и люди новых специальностей, чья деятельность будет связана с анализом больших объемов данных с целью совершенствования систем управления бизнесом, компаниями и организациями: ученые по данным, архитекторы данных, проектировщики пользовательских запросов и специалисты, способные учесть опыт пользователей. Спрос на представителей этих специальностей, как отмечают в этом номере топ-менеджеры ведущих ИТ-компаний, будет расти экспоненциально. А пока, согласно еще одному исследованию, компании принимают сегодня решения на основе данных более чем годичной давности, точность которых не превышает 55%. ■

– Дмитрий Волков

COVER FEATURES

BIG DATA

18 Big Data: New Theory and Practice Leonid Chernyak

The symbols of the first Industrial Revolution were cast iron and steam, whereas the second has brought steel and production lines. With the third polymers, aluminum, and electronics have arrived, while the new age would come under the sign of composite materials and information. Is Big Data just a buzzword, or the future of the industry?

26 Big Data in Close-Up Natalia Dubova

During the recent years, IBM has systematically been building an information management platform based on its own and acquired technologies comprising all the tools enterprises need to work with data. However, the platform would not be complete without including Big Data processing tools into the supply chain.

30 A DBMS for Analyzing the Big Data Kirill Vakhrameev

Virtualization has significantly changed IT except for the area of high-load applications and large database servers where the virtualization approach would not be efficient. However, today revolutionary changes come to this area too, yet of completely different nature.

OS PREDICTION

14 The 3rd Platform of IT: The OS Big Seven '2012 Leonid Chernyak

In the distant 1950's, the train of IT industry has taken off from its departure platform, which later has been called the mainframe. In the 80's despite the significant development of minicomputers and UNIX systems the IT train had come to the PC platform, and with the arrival of the current decade it began approaching a third platform. This journey could be talked about not only in the railway, but also in the geology terms – today we are in the midst of a truly tectonic-scale paradigm shift, and there's a hope that in 2012 the features of the new platform would take their final distinct form.

OS GUESTROOM

34 Perspectives on the Big Data Leonid Chernyak

Hugh Yoshida, Stephen Brobst, Luke Lonergan and Joseph Reger – views on current trends in Big Data processing technologies.

SECURITY

40 Human Ear Recognition Arun Ross, Ayman Abaza

Ear recognition technology is a potentially valuable tool in the biometric arsenal.

SOFTWARE ENGINEERING

42 The FBI Gets Agile

Chad Fulgham, Jeffrey Johnson, Mark Crandall, Leo Jackson, Nathan Burrows

What can happen in a day at the Federal Bureau of Investigation? A SWAT team might rescue a hostage while an analyst pieces together intelligence linked to a terrorist threat and agents arrest a criminal. However, as these parts of the organization focus on high-profile issues such as counter intelligence, white-collar crime, and cybersecurity, a new face of the FBI is emerging in the basement of the Hoover building in Washington, DC. That's where FBI employees and contractors are working together to create Sentinel, a case-management software system designed to change how the FBI gathers, stores, and links data.

45 Optimizing Compilers Andrey Anufrienko

Optimizing compilers, which are considered among the most complicated software applications, are developing to become even more complex. That is why the ISVs have to dedicate increasingly more resources for support and development of optimizing compilers. What is interesting about today's compilers and in particular, Intel's optimizing compiler? What primary concepts and technologies for increasing applications' performance are used today?

EXTREME TECHNOLOGIES

47 DaaS: Data on Demand Leonid Chernyak

In addition to IaaS, SaaS, PaaS, and all the other *aaS concepts, there is a need for data provided as a service as well.

APPLICATIONS

50 Making Sense of Large Data Volumes by Association Andrey Revutskiy

Any company working with large data volumes needs to solve two key issues: how to store the data and how to use it efficiently. Business Discovery solutions widen capabilities of business intelligence systems using associative search and analysis.

OS ACADEMY. IT UNIVERSITIES

52 Builder of Reliable Software Natalya Dubova

ETH Zurich Professor of Software Engineering Bertrand Meyer is one of the experts that form the face of today's software engineering. His scientific research, which covers among others the area of software verification show that software engineering can be a full-fledged scientific discipline, while his practical efforts has contributed greatly to applying the results of research into everyday work of software programmers.

OS ACADEMY. LIBRARY

56 Software Engineering and Related Disciplines Sergey Kuznetsov

The October and November 2011 issues of Computer Magazine (IEEE Computer Society, Vol. 44, No. 10-11, 2011) is dedicated to interoperation of software engineering with other fields of computer science and exaflops systems in accordance..

<http://infozone.osp.ru>

- Около 400 000 посетителей сайта ежемесячно из России, Украины, Белоруссии, Казахстана и других стран ближнего и дальнего зарубежья
- Свыше 9000 зарегистрированных подписчиков

Открытые
системы

computerworld

Windows IT PRO

Директор

Мир ПК

ЖУРНАЛ СЕТЕВЫХ РЕШЕНИЙ

LAN

Сети

ОТКРЫТЫЕ СИСТЕМЫ

Open Systems Publications

Продукты
и услуги OSP.RU

События
Статьи
Вебкасты
Конференции
Аналитика
OSP TV

Читатели
и посетители сайта



Темы

Обеспечение
непрерывности
бизнеса
Слияния и поглощения
Управление
бизнес-процессами
ИТ в здравоохранении
...
Десятки тем и зон
по продуктам и
решениям

Заказчики
ИТ



ОТКРЫТЫЕ СИСТЕМЫ
Open Systems Publications

<http://infozone.osp.ru>

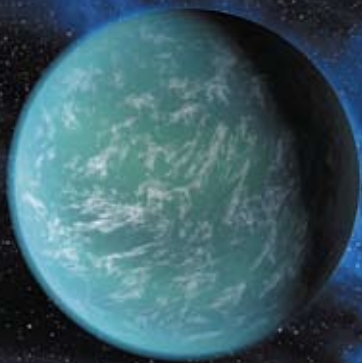


Иллюстрация: NASA/Ames/JPL-Caltech

Найдена первая планета, предположительно пригодная для жизни

Агентство NASA сообщило, что космический телескоп Kepler впервые обнаружил планету в зоне обитаемости звезды, похожей на наше Солнце. Зоны обитаемости – условные области, в которых на поверхности планет может присутствовать вода в жидком состоянии. Kepler был выведен на орбиту специально с целью поиска экзопланет, похожих на Землю.

В зоне видимости телескопа – около 150 тыс. звезд. Космическая обсерватория идентифицирует планеты по «транзитам» – изменениям в яркости звезд, вызванных прохождением планет между звездой и телескопом. Чтобы убедиться, что обнаружена действительно планета, необходимо не меньше трех «транзитов». Радиус открытой планеты, Kepler-22b, примерно в 2,4 раза больше, чем у Земли. От нас она находится на удалении 600 световых лет. Ее период обращения вокруг своей звезды — 290 суток. Звезда принадлежит к той же категории, что и Солнце, – G-классу, но несколько меньше и холоднее.

NASA также сообщает, что Kepler открыл уже более 1000 новых кандидатов в планеты, в том числе десять размером примерно с Землю, обращающихся в зоне обитаемости своих звезд.

«Ментальные протезы» помогают сдерживать пагубные порывы

Сотрудник MIT Media Lab Саджид Сади предложил идею «рефлектонов» (ReflectOns) – предметов, которые помогают людям задумываться о последствиях своих действий. Он называет их «ментальными протезами». Например, рефлектон SpendTrend представляет собой систему в кредитной карте, помогающую сдерживать «импульсные» покупки – карта прогнозирует долгосрочные последствия таких покупок и выстраивает график, показывающий, в какие суммы в итоге они выливаются, если совершать их постоянно.

Еще один рефлектон помогает не переедать. Устройство смонтировано на вилке. Если пользователь во время еды начинает часто подносить вилку ко рту, рефлектон начинает менять центр тяжести вилки, «намекая» хозяину, чтобы тот не торопился. Как объясняет изобретатель, если есть высококалорийную пищу медленно, то чувство насыщения, зависящее от уровня глюкозы в крови, еще до наполнения желудка успеет дать знать о том, что калорий потреблено уже достаточно.

Фото: Sajid Sadi/MIT Media Lab



IBM избавит астрономов от скуки

IBM завершила работу над прототипом программной системы Information Intensive Framework, которая поможет астрономам работать с огромными потоками данных, поступающими с радиотелескопа Square Kilometre Array – распределенной системы параболических антенн общей площадью в 1 кв. км. По замыслу разработчиков телескопа, он сможет беспрецедентно тщательно следить за источниками радиоволн в космосе. Антенны будут установлены в Австралии, Новой Зеландии или ЮАР, в зависимости от того, какая страна победит в конкурсе. ИИФ до некоторой степени автоматизирует классификацию астрономических объектов с использованием стандартной таксономии. Последняя позволяет астрономам в повседневных наблюдениях вместо цифр оперировать именованными объектами, например «галактика», «сверхмассивная черная дыра» и т. д. Как поясняет Дугал Ватт, директор по технологиям новозеландского офиса IBM и председатель отраслевого консорциума Новой Зеландии по проекту Square Kilometre Array, задача системы в том, чтобы повысить результативность и эффективность работы астрономов, так чтобы они могли уделять больше времени творческим аспектам своей деятельности. Ватт надеется, что, когда астрономов освободят от скучной работы по каталогизации, у них будет больше времени на то, чтобы обнаруживать что-то необычное в наблюдениях.



Иллюстрация: SPDO

Нарисовано наукой

С картинками, победившими в ежегодном конкурсе Принстонского университета Art of Science, можно ознакомиться на его официальном сайте. Тема конкурса – Intelligent Design (разумный замысел), отсылка к направлению креационизма, утверждающему, что жизнь была сотворена с определенной целью. Как объясняют организаторы, этот термин употреблен в более широком смысле: в чертах красивых объектов – и природных, и рукотворных, – всегда видна элегантность их строения. Жюри отдавало предпочтение научным изображениям, имеющим эстетическую ценность, а не специально созданным произведениям, навеянными наукой.

Первый приз достался Кристофу Гиссинжеру за компьютерную модель изменения полярности магнитного поля Земли со временем. Второе место у Чжена Джеймса Сяна за картину, полученную в результате разбиения изображения дерева алгоритмом распознавания составных частей, а третье у Сюэнин Бай и Джеймса Стоуна за изображение одного из этапов формирования планет. Денежные призы – 250, 154,51 и 95,49 долл. – распределены по правилу золотого сечения, соотношения, широко встречающегося в эстетически приятных формах, от морских раковин до древнегреческих храмов.



Аэроп: Christophe Gissinger/Princeton Plasma Physics Laboratory

ДИСК В ПОДАРОК!



**ОФОРМИТЕ ГОДОВУЮ
ПОДПИСКУ НА ЖУРНАЛ
"ОТКРЫТЫЕ СИСТЕМЫ"
ДО 31 ДЕКАБРЯ**

**ПОЛУЧИТЕ ДИСК С
АРХИВОМ 1993-2010 В ПОДАРОК.**

Журнал «Открытые системы» издается с 1993 года и сегодня является одним из самых авторитетных в России периодических изданий, посвященных построению гетерогенных информационных систем корпоративного уровня; архитектурам современных компьютерных систем и микропроцессоров; операционным системам; СУБД и хранилищам информации, а также Web-технологиям. Большое внимание на страницах журнала уделяется освещению практического опыта разработки и внедрения конкретных решений на промышленных предприятиях, в научных и государственных учреждениях, на транспорте, в медицине и в телекоммуникационных компаниях.

БЛАНК ЗАКАЗА

Ф.И.О. (полностью) _____

Индекс _____ Область _____

Город _____

Улица _____

Дом _____ Корпус _____ Квартира _____

Место работы _____

Должность _____

Телефон _____

E-mail _____

Оплатите прилагаемую квитанцию в любом отделении Сбербанка и пришлите копию об оплате по адресу dinara@osp.ru до 31 декабря 2011 г.

Также на сайте <http://www.osp.ru> доступны электронные способы оплаты.
По всем вопросам, связанным с подпиской, звоните по тел.: (495) 725-47-85

Извещение

Кассир

Квитанция

Кассир



ЗАО "Издательство "Открытые системы"

ИНН 7706128372/770601001 № р/с 40702810438170101424

в Краснопресненском ОСБ 1569, Сбербанк России г. Москвы

кор./с 30101810400000000225 БИК 044525225

ФИО _____

Индекс _____ Адрес доставки _____

Подписка на «Открытые системы», в т.ч. НДС 10%

(назначение платежа)

Сумма платежа: _____ руб. _____ коп.

Сумма платы за услуги: _____ руб. _____ коп.

Итого: _____ руб. _____ коп.

007

ЗАО "Издательство "Открытые системы"

ИНН 7706128372/770601001 № р/с 40702810438170101424

в Краснопресненском ОСБ 1569, Сбербанк России г. Москвы

кор./с 30101810400000000225 БИК 044525225

ФИО _____

Индекс _____ Адрес доставки _____

Подписка на «Открытые системы», в т.ч. НДС 10%

(назначение платежа)

Сумма платежа: _____ руб. _____ коп.

Сумма платы за услуги: _____ руб. _____ коп.

Итого: _____ руб. _____ коп.

007

Форма № ПД-4

Уважаемый клиент!
Пожалуйста, введите в поле "назначение платежа" всю информацию, выделенную рамкой

Уважаемый подписчик!
Пожалуйста, аккуратно и разборчиво заполните все данные о вашей подписке в "назначении платежа".

УПРОЩАЙТЕ

построение сети хранения данных
(SAN) с помощью масштабируемой
и отказоустойчивой системы хранения.

HP P4000 G2 SAN (LeftHand)

Данная система имеет уникальную кластерную архитектуру, при которой каждый узел является полноценной СХД. Благодаря тому, что все узлы работают как единое целое, HP P4000 G2 имеет следующие ключевые преимущества перед классическими системами хранения:

- **Двойная отказоустойчивость.** Так как данные защищены не только дисковым рейдом на уровне отдельного узла, но и сетевым рейдом на уровне всего кластера, то даже выход из строя целой ноды лишь незначительно влияет на общую работоспособность системы.
- **Линейная масштабируемость.** Так как у HP P4000 каждый узел – это отдельная СХД, то при апгрейде добавляется не только емкость, но так же линейно растет производительность и надежность всего кластера.

Кроме того, в начальной конфигурации HP P4000 G2 Вы **БЕСПЛАТНО** получаете целый набор полезных утилит:

- синхронная и асинхронная репликация;
- Snapshot;
- Thin Provisioning;
- SmartClone и т.д.

HP P4000 G2 применяется для виртуализации, баз данных, почтовых и веб-приложений, и т.д. А благодаря отличной масштабируемости и широкому модельному ряду, эти системы хранения используют как малые и средние организации, так и международные корпорации.

Более подробно о продукте
Вы можете узнать на сайте:
www.hp.ru/promo/P4000



© Copyright 2011 Hewlett-Packard Development Company, L.P.
Содержащаяся здесь информация может быть изменена без предварительного уведомления. Все гарантии в отношении продуктов и услуг компании HP выражены в явном виде в гарантийных обязательствах, прилагаемых к соответствующим продуктам и услугам. Никакие содержащиеся здесь сведения не могут рассматриваться как дополнение к этим условиям гарантии. HP не несет ответственности за содержащиеся здесь технические или редакторские ошибки или упущения.

