

Открытые системы

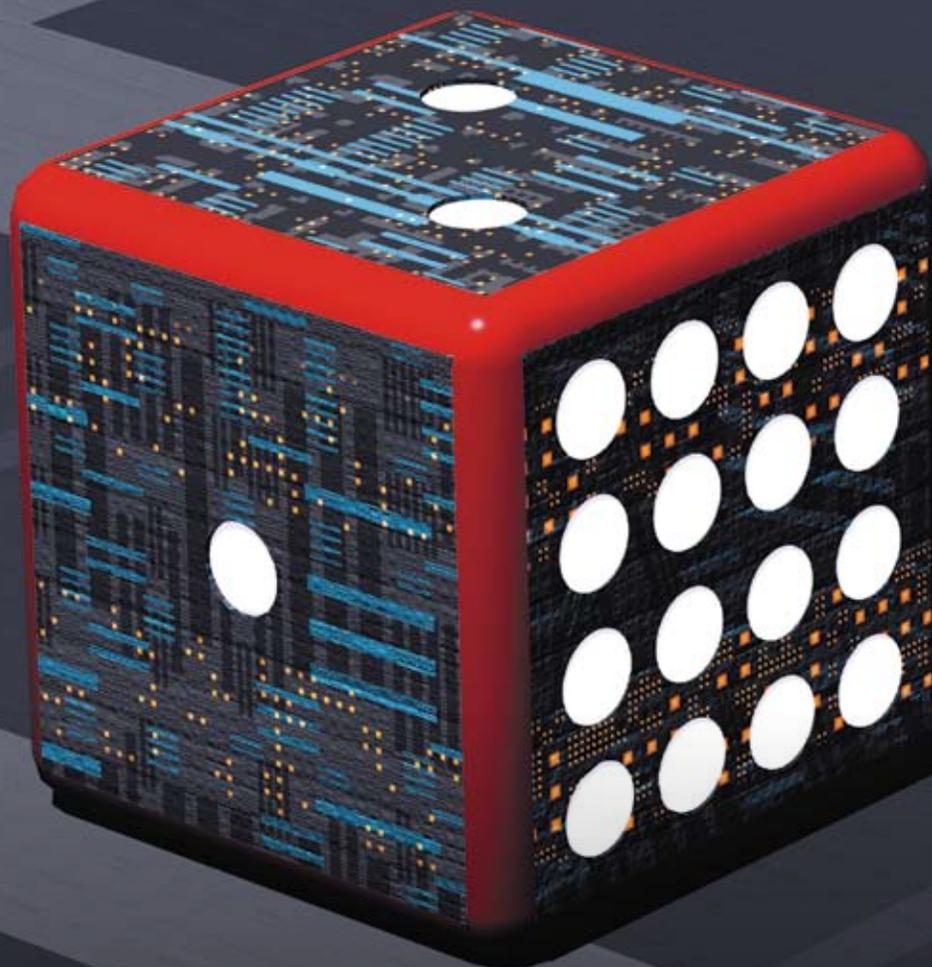
Открыты для вас. 20 лет

№08
2013

ISSN 1028-7493

ИТ для бизнеса —
архитекторам
информационных систем

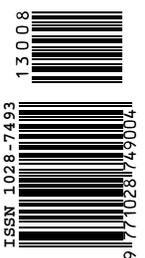
www.osmag.ru



НРС: НА ГРАНИ ЭКЗАФЛОПСА

Задачи для экзамасштабных компьютеров

- Секреты машинного обучения в «Яндексе»
- Аналитика Больших Данных и социальные сети
- Новые возможности разработки в облаках
- Как программировать в параллельных мирах



ПОДПИСЧИКАМ БОНУСЫ



1650 руб.

2100 руб.*

*Стоимость подписки указана за 12 месяцев, с учетом доставки «Почта России» на территории РФ

Вы - участник бонусной программы «Спасибо от Сбербанка»?
Оплатите любую подписку на сайте www.osp.ru, картой Сбербанка
и получите 5% от ее стоимости бонусами СПАСИБО! Узнайте
больше о «Спасибо от Сбербанка»: www.spasibosberbank.ru



ПОДПИСАТЬСЯ

www.osmag.ru

+7 495 725-47-85

e-mail: xpress@osp.ru

Организатором Акции является ЗАО «Издательство "Открытые системы"» ОГРН 1037700120886. Акция проводится на территории Российской Федерации в период с «30» августа 2013 г. и не ограничена сроком действия. На Бонусный счет Участника в виде Бонусов начисляется 5% от суммы покупки в дополнение к Бонусам, начисляемым в рамках Программы «Спасибо от Сбербанка». С Правилами Акции можно ознакомиться на Сайте Организатора Акции www.osp.ru. Акция проводится на базе Программы «Спасибо от Сбербанка», в рамках которой на Бонусный счет Участника в виде Бонусов начисляется 0,5% от суммы покупки (новые участники программы в течение первых 3-х месяцев получают 1,5% от суммы покупки) Организатором Программы является ОАО «Сбербанк России» (Генеральная лицензия Банка России № 1481 от 08.08.2012). С Правилами Программы, порядком начисления и списания Бонусов можно ознакомиться на сайте www.spasibosberbank.ru

Реклама 16+

Явление Больших Идей

Любому современному предприятию или организации нужны высокопроизводительные системы, обеспечивающие масштабируемость и эффективность в реальном времени для работы систем бизнес-аналитики, управления корпоративными данными, моделирования или поддержки гибких облачных сред. Вместе с тем, какими бы изощренными ни были нынешние суперкомпьютеры, они пока остаются инструментами прошлого века — для решения перечисленных задач, да и любой проблемы из портфеля Больших Данных, требуются компьютеры постмуровской эпохи, и ожидается, что первые такие компьютеры экзафлопсной производительности появятся уже через несколько лет. А что произойдет, когда исследователи или промышленники не будут в своей работе ограничены вычислительной мощностью или объемами доступных для хранения данных? Как изменится работа инженеров, которые благодаря комбинации экзафлопсных систем, облаков и средств работы с Большими Данными получат возможность масштабировать свои ИТ-инфраструктуры вплоть до глобальных гетерогенных конфигураций, доступных сегодня лишь компаниям из Fortune 500? Откроется ли тогда перед нами пространство Больших Идей?

Вполне вероятно, но пока неумолимый прогресс в процессорных технологиях привел к тому, что производительность приложений уже не ограничена возможностями процессоров, а все больше зависит от других компьютерных компонентов, считавшихся ранее вспомогательными. Вместе с тем архитектуры экзамасштабных систем со сложной иерархией памяти и параллелизмом совсем не обязательно приведут к ускорению работы программ: невысокая скорость доставки и передачи данных, неспособность унаследованных приложений автоматически распределяться на множество одновременно выполняемых потоков уже сегодня резко контрастируют с высокой производительностью процессоров.

Как отмечают авторы этого номера, посвященного обсуждению технологий создания экзамасштабных систем, сегодня перед разработчиками стоит задача не столько построения компьютерной конфигурации, при запуске на которой производительность специально созданной тестовой программы превысила бы рубеж

в один экзафлопс, сколько обеспечения максимально эффективного режима выполнения реальных приложений, задействующих все выделенные им ресурсы суперкомпьютера.

Данных, которые сегодня приходится обрабатывать в различных областях, настолько много, они настолько сложны и прибывают так быстро, что обычные технологии оказываются бессильны, поэтому в ближайшие годы ожидается появление новых алгоритмов, позволяющих в огромных массивах разнообразных наборов обнаруживать скрытую информацию. Сегодня, как уже всем известно, Агентство национальной безопасности США собирает и хранит данные обо всем и обо всех, однако у экспертов все-таки нет уверенности, что они действительно обрабатываются и анализируются. Как отмечает Леонид Черняк, архитектуры нынешних универсальных компьютеров, пусть и отлично приспособленные для выполнения расчетов, плохо адаптированы к задачам с интенсивной работой в условиях нерегулярного доступа к памяти, в частности информационного анализа и поиска, — предполагается, что собранные данные должны быть кем-то и как-то рассортированы. Однако, например, сознание человека оперирует не индексами, а ассоциациями, опирается на паутину ячеек мозга, объединенных сложными путями, а эта паутина характеризуется высокой динамикой связей, непостоянством наполнения и огромным быстродействием. Сегодня нужны большие идеи по организации эффективных механизмов доступа к данным, а одно лишь стремление любой ценой победить в гонке за экзафлопсами неконструктивно — не лучше ли сначала тщательно разобраться, какие на самом деле нужны архитектуры для решения каждой конкретной задачи?

Эту идею продолжает и Леонид Эйсьмонт, отмечающий, что рост популярности облаков и обострение проблемы Больших Данных заставляют пересмотреть традиционные взгляды на аппаратно-программные платформы, хорошо выдающие флопсы, но деградирующие на задачах с интенсивной нерегулярной работой с памятью, что сильно снижает уровень развиваемой реальной производительности по отношению к пиковой. В этой связи неудивительно, что разработчики из развитых стран, работающие над созданием новых экзамасштабных технологий в



рамках таких программ, как DARPA UHPC и STARnet, уже обратили свои взоры на гибридность и гетерогенность. В результате уже появились суперкомпьютеры с архитектурой, включающей неоднородные сегменты, ориентированные на решение задач конкретного типа и реализуемые на базе специализированных микропроцессоров, сетей и устройств.

Неспособность существующих систем быстро обрабатывать большие объемы разнообразных данных, поступающих из различных источников, отмечают и отраслевые эксперты, в качестве причин называя слабость имеющихся ИТ-конфигураций — по разным оценкам, 40% компаний из США, 33% из Европы и 49% из Азии работают на системах с производительностью много ниже необходимой. Одна из причин этого, конечно, стоимость и, как следствие, увлечение так называемыми серверами стандартной архитектуры, однако тот, кто выбирает только по цене, заслуживает быть обманутым.

Гражданские и компьютерные СМИ наперебой сегодня обсуждают проблемы Больших Данных, проводятся конференции, на которых каждый производитель и спикер считает своим долгом коснуться этой темы. Попутно оказалось, что существующие аппаратно-программные инфраструктуры неспособны эффективно решать возникающие в этой сфере задачи. В полной мере это касается и традиционных высокопроизводительных систем, предназначенных «тяжелых» расчетов, в частности в рамках вычислительного моделирования, однако сегодня этого уже недостаточно — чтобы наблюдать явление Больших Идей, нужны новые архитектуры. ■

Дмитрий Волков