

# Аналитика “больших данных”

## — НОВЫЕ ПЕРСПЕКТИВЫ

*Аналитика “больших данных” — приоритетный вектор развития корпорации EMC с середины 2011 г. Это направление, имеющее в основе решение EMC Greenplum, за последние полгода было существенно расширено новыми компонентами и функциональностью, благодаря чему EMC удалось вплотную приблизить эту группу решений к промышленным стандартам. Данная публикация является продолжением статьи в SN № 2/46 (2011) — “EMC Greenplum — новое поколение BI-хранилищ”.*



Денис Серов — руководитель направления технического консультирования “EMC Россия и СНГ”.

### Введение

В последние годы в мире ИТ произошла “тихая” революция, обусловленная устойчивым ростом вычислительных мощностей и лавинообразным ростом объемов данных. Сочетание этих двух факторов дало начало новому направлению — аналитике “больших данных” — мощной силой. Компьютерные технологии, проникшие в нашу жизнь, создают бесчисленное множество новых источников информации. Этому способствует и огромное количество мобильных цифровых устройств: сотовые телефоны, смартфоны, лэптопы, персональные датчики и др.

Все большее число предприятий и правительственных агентств открывают для себя стратегические возможности использования больших объемов данных. По мере того, как все эти системы начинают связываться друг с другом, появляются новые программные инструменты и аппаратные технологии для анализа этих данных на предмет их взаимосвязи и извлечения полезной информации. Во всем мире воплощается радикально новый вид инфраструктуры знаний. Развивается новая эра “больших данных”, оказывающая значительное влияние на бизнес, правительство, демократию и культуру. Данные теперь — не просто офисные файлы и компьютерные базы данных, которые существуют в течение многих десятков лет, они все чаще, используются как инструмент принятия решений в реальном времени.

Исторически в наибольшей степени в аналитике с “большими данными” продвинулись интернет-компании, предоставляющие услуги поиска, социальных сетей, онлайн-магазинов. Но, по мере того, как хранение данных становилось дешевле, а пропускная способность публичных сетей возрастала, другие сектора бизнеса, а также правительственные агентства, учебные и научные организации стали осваивать новые техники анализа данных.

Современные кластерные системы предоставляют среду для хранения, средства сбора и интеграции метаданных из многочисленных источников, а также высокопроизводительные средства анализа. В соединении с новыми формами вычислений и статистического анализа теперь исследователи способны разрабатывать многомерные статистические модели на основе “больших данных” и определять, как система должна реагировать при изменении поведения этих моделей.

Так, например, компания Netflix, занимающаяся прокатом DVD-дисков, может использовать автоматизированные алгоритмы машинного обучения для определения корреляций предпочтений своих заказчиков и предлагать им свои рекомендации.

Различные гиганты, такие как Walmart, изучают массивные базы данных продаж с более чем 260 млн транзакций в день для того, чтобы определить наилучшие стратегии ценообразования и обеспечить контроль за складами.

Разведывательные агентства теперь должны уметь обрабатывать данные не только со своих собственных спутников и телефонных перехватов, но и из интернета и других публикаций. Многие научные дисциплины становятся все более компьютерными и зависящими от этих данных: физика, астрономия, океанография и биология.

При петабайтных масштабах информация не имеет ничего общего с простыми тремя или четырехмерными моделями ее представления. Необходим совершенно иной подход для выявления/построения зависимостей данных, который, в свою очередь, приводит к необходимости отказаться от попыток визуализировать данные в целом и “смотреть” на них математически, и только потом устанавливать для них контекст.

Например, Google захватил мир рекламы не больше и не меньше, чем с помощью при-

кладной математики. Ее основатели не претендовали, что много знают о культуре или/i основах рекламы. Они просто предположили, что лучшие данные с лучшими аналитическими инструментами являются ключом к победе. И были правы.

### Аналитика в эру “больших данных”

Аналитика “больших данных” неразрывно связана с темой специализированных хранилищ данных для BI (Business Intelligence), позволивших вывести бизнес-анализ на качественно новый уровень. И, если в начале своего появления, они, в основном, были полностью ориентированы на работу с традиционными хранилищами (DW — Data Warehouse) и структурированными данными и позволяли многократно повышать производительность BI-приложений при работе с многотерабайтными DW, то в последнее они стали специализироваться и для работы с большими объемами неструктурированных данных, например, за счет глубокой интеграции с Hadoop-кластерами.

Аналитику “больших данных” можно классифицировать по трем признакам: 1) возможность анализа данных петабайтного объема; 2) возможность анализа очень большого трафика данных в режиме максимально приближенному к реальному времени; 3) возможность анализа данных разного типа — структурированных и неструктурированных одновременно.

В качестве источников “больших данных” могут выступать: новые типы датчиков, порождающие потоки цифровой информации от телескопов, видеокамер, мониторов дорожного движения, томографов, а также биологические и химические сенсоры окружающей среды; интернет-магазины с большим объемом логов и высоким онлайн-трафиком посетителей сайта. Миллионы людей со своих мобильных устройств/телефонов расширяют эти потоки информации собственными персонифицированными данными.

Потребителями решений для аналитики “больших данных” может быть как бизнес, так и госсектор.

Приведем три примера применения решений “с человеческим лицом” для аналитики “больших данных”, продвигаемых

EMC и развиваемых на основе массово устанавливаемых датчиков разного типа (<http://humanfaceofbigdata.com/examples>).

Япония с 2005 г. с целью раннего оповещения о землетрясениях, потратила порядка \$500 млн на установку датчиков по всей стране, для детектирования т.н. М-волн, которые происходят приблизительно за 80 секунд до возникновения Р-волн, которые и наносят основные разрушения. Это позволяет вовремя остановить сверхскоростные пассажирские экспрессы и сборочные линии. В будущем эту диагностику планируется делать на основе ноутбуков от IBM (Lenovo), Acer и HP, в которые с 2005 г. интегрированы акселерометры. Они обнаруживают внезапное ускорение, например, когда ноутбук падает со стола. Специальное ПО (Quake Catcher Network) анализирует толчки, ощущаемые акселерометром компьютера, и сообщает только о существенных движениях центральному серверу, игнорируя колебания от проходящего грузовика, удара о стол или даже незначительных землетрясений. Образец сигналов, полученных центральным сервером, позволяет распознать существенное землетрясение.

Перепись населения в США производится раз в 10 лет при огромных затратах. В условиях современного быстро меняющегося мира результаты переписи могут уже устареть еще до того, как официально опубликованы. Это связано как со случайными многочисленными ошибками при заполнении опросных форм, так и внесением в них умышленно неверных данных (из-за нарушения иммиграционного законодательства, нежелания делиться своими персонализированными данными и пр.). Но, если вам необходимо, например, знать только число людей, проживающих в доме, районе, регионе, то в качестве показателя, коррелирующего с числом жителей, может служить, например, объем потребляемой воды. И это, фактически, намного более точный показатель, чем данные, добровольно сообщаемые при переписи.

Intel и GE работают над серией продуктов, используя сенсоры и собираемые ими данные для помощи самостоятельному живущим пожилым людям, например, для предсказания инсульта. Intel и GE создали устройство, которое постоянно измеряет и передает данные о походке человека. Нет никаких стандартных профилей хорошей или плохой походки, система самостоятельно после анализа создает персонализированный профиль нормальной походки (с учетом возраста и состояния) для каждого человека. Оказывается, что за 2–3 дня до того, как происходит инсульт, в ней происходят небольшие изменения.

Благодаря постоянному мониторингу и анализу состояния человека через централизованную систему удается своевременно распознавать предвестники приступов и своевременно принимать меры по их предотвращению, оповещая родственников или/и медицинских работников. При этом удается не прибегать к помощи персонала, осуществляющего постоянный контроль за таким человеком.

Таким образом, на основе хитроумных техник корреляции, появляется возможность извлекать полезные корреляции

информации, которые до недавнего времени оставались спрятанными в петабайтах данных. Например, компания Google изучает пестроту и локализацию поисковых запросов, чтобы предсказывать эпидемии гриппа, а также тенденции безработицы еще до того, как выйдет официальная государственная статистика.

Компании, выпускающие кредитные карты, просеивают колоссальные объемы информации, связанные с переписью населения, финансовыми операциями, персональными данными для того, чтобы обнаружить мошенничество или определить тренды потребительского спроса.

Медицинские организации обрабатывают и осмысливают карты тысяч пациентов с целью определения корреляций между принятыми мерами лечения и их результатами. Компании, предоставляющие web-сайты для социальных сетей, проводят анализ огромных данных персональной информации для того, чтобы выявить потребительские предпочтения и разработать лучшие маркетинговые стратегии.

Появляется совершенно новый класс геолокационных данных, который дает компаниям возможность анализировать данные мобильных устройств, что, в свою очередь, позволяет получить самые неожиданные выводы о жизни людей и об экономике. Оказывается, например, то расстояние, которое люди готовы проехать до супермаркета (эти данные собираются при помощи отслеживания мобильных телефонов людей и их звонков), является прекрасным показателем для измерения потребительского спроса в экономике. Уникальные техники, которые используются при анализе “больших данных”, могут дать взгляд изнутри на множество сложных проблем — в большинстве случаев с высокой точностью и своевременностью.

### EMC Greenplum UAP — унифицированная платформа для аналитики “больших данных”

Учитывая активно развивающийся сегмент рынка информационных хранилищ нового поколения, компания EMC объявила аналитику “больших данных” своим приоритетом в 2012 г. Впервые EMC анонсировала свою программу развития аналитики “больших данных” в мае 2011 г. В частности, EMC представила

комплексную стратегию дистрибуции, интеграции и поддержки ПО с открытым исходным кодом Apache Hadoop.

Платформа Apache Hadoop возникла как инновационный модуль анализа “больших данных”, предназначенный для поддержки распределенных приложений, интенсивно работающих с данными, и ускорила анализ корпоративных данных, особенно неструктурированных. Hadoop — это программная платформа с открытым исходным кодом, которая предназначена для выполнения больших пакетных заданий анализа информации в кластерах типовых серверов.

Hadoop состоит из двух основных компонентов: MapReduce и распределенной файловой системы Hadoop (HDFS). MapReduce — это механизм параллельной обработки заданий, который принимает запросы и выполняет их параллельно на нескольких узлах, чтобы эффективнее обрабатывать большие наборы данных. HDFS — это распределенная файловая система, которая используется в обычных кластерах Hadoop для хранения данных, которые должны анализироваться с помощью MapReduce. Hadoop часто используется с рядом своих компонентов, например HBase, Pig и Hive, которые расширяют возможности работы с базами данных.

Этот проект получил наибольшее распространение на рынке свободного ПО для организации хранения “больших данных” в распределенных приложениях с интенсивной обработкой данных. EMC также анонсировала первое в мире специализированное высокопроизводительное устройство Hadoop для обработки структурированных и неструктурированных данных — Greenplum HD Data Computing Appliance.

Модуль Greenplum HD объединяет Hadoop и Greenplum Database в одно специализированное устройство Data Computing Appliance.

Greenplum HD — это полностью открытая, сертифицированная и поддерживаемая версия стека Apache Hadoop, которая включает HDFS, MapReduce, Hive, Pig, Hbase и ZooKeeper:

- Hive — язык запросов к данным в распределенном хранилище данных на базе HDFS;
- Pig — язык параллельной обработки данных в среде MapReduce;

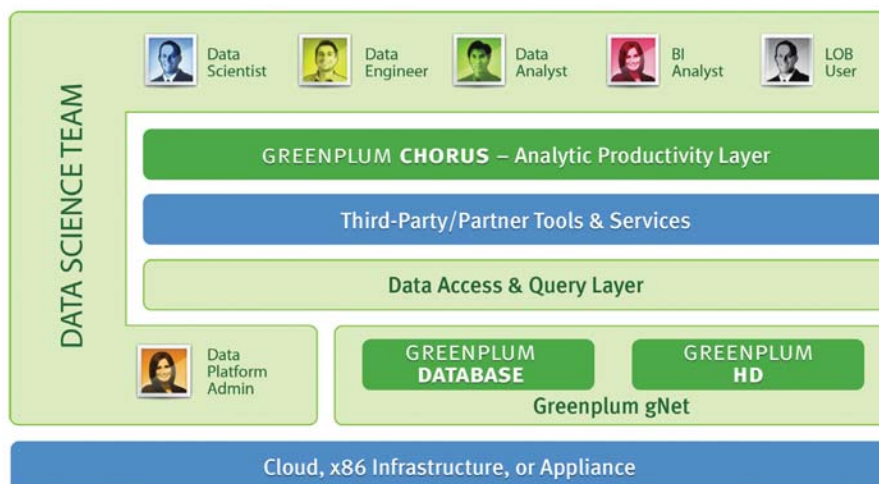


Рис. 1. Архитектура EMC Greenplum Unified Analytics Platform.

- Hbase — база данных Hadoop с линейной и распределенной масштабируемостью;
- Zookeeper — высоконадежный сервис централизованной координации распределенного выполнения заданий

MADlib — это надстройка верхнего уровня над вышеописанной конструкцией, которая является библиотекой с открытым кодом для масштабируемой аналитики БД. Этот набор обеспечивает разработчиков заготовками для параллельной обработки данных с математическими, статистическими и машинными методами обучения для структурированных и неструктурированных данных. ML специально спроектирован для того, чтобы способствовать повсеместному распространению навыков масштабируемой аналитики, совмещая усилия коммерческой практики, академических исследований и разработок сообщества СПО.

В результате объединения Hadoop с базой данных EMC Greenplum появляется возможность в рамках одного хранилища обрабатывать как структурированные, так и неструктурированные данные. Кроме того, EMC объявила о доступности основанного на технологиях Hadoop программного обеспечения EMC Greenplum HD Community Edition и EMC Greenplum HD Enterprise Edition. В сочетании с сертификацией продукта двенадцатью ведущими партнерами это позволит применять такие технологические инновации, как взаимодействие с данными в реальном времени, улучшит надежность и упростит развертывание и использование Hadoop.

Уникальный дополнительный функционал EMC для Hadoop:

- *EMC Greenplum HD Data Computing Appliance* — Apache Hadoop "бесшовно" интегрируется с базой данных Greenplum в Greenplum HD Data Computing Appliance. Решение поддерживает внешние таблицы Hadoop, что позволяет пользователям обращаться к данным, хранящимся в Hadoop Distributed File System (HDFS), без извлечения их из файловой системы. Администраторы могут читать и записывать файлы параллельно из Greenplum на HDFS, что обеспечивает быстрое и простое совместное использование информации. Кросс-платформенный анализ можно выполнять, используя мощь Greenplum SQL и расширенные функции аналитики для доступа к данным в HDFS. Комбинированное решение реализует единственную в индустрии законченную платформу для аналитики больших данных;
- *EMC Greenplum HD Enterprise Edition* — Enterprise Edition на 100% совместима по интерфейсам со стеком Apache Hadoop. Будучи совместимой с интерфейсами Hadoop, Enterprise Edition обеспечивает "бесшовную" переносимость приложений и в то же время реализует расширенные функции, востребованные в больших организациях, в том числе:
  - функции управления данными, включая мгновенные снимки и репликацию на большие расстояния;
  - простую загрузку и доступ к данным с помощью "родного" интерфейса network file system (NFS);

– полную управляемость, в том числе простое развертывание кластера, автоматическое распознавание сбоев и оповещение о них, управление несколькими площадками и функцию rolling upgrades.

Наряду с реализацией этих функций производительность улучшается в 2-5 раз по сравнению со стандартной версией Apache Hadoop;

- *EMC Greenplum HD Community Edition* — Community Edition это бесплатная версия, полностью сертифицированная на совместимость с open source, поддерживающая стек Apache Hadoop, который состоит из HDFS, MapReduce, Zookeeper, Hive и Hbase. EMC Greenplum обеспечивает отказоустойчивость для Name Node и Job Tracker. Оба этих компонента являются единичными точками отказа стандартных внедрений Hadoop.

Также EMC построила для Hadoop мощную и динамичную экосистему с 12-ю компаниями: Concurrent, CSC, Datameer, Informatica, Jaspersoft, Karmasphere, Microstrategy, Pentaho, SAS, SnapLogic, Talend и VMware, специализирующимися на business intelligence, data transfer и других технологиях.

В конце 2011 г. EMC сделала ряд важных анонсов, существенно продвинувших ее решения для аналитики "больших данных".

Все решения, объявленные в мае 2011 г., появились на рынке в конце сентября 2011 г. В частности, стала доступна EMC Greenplum Modular Data Computing Appliance (DCA) — первая в индустрии платформа для аналитики корпоративного класса, которая на основе модульной архитектуры позволяет комбинировать MPP-обработку (Massively Parallel Processing) реляционных баз данных и неструктурированных данных на базе Apache Hadoop-кластеров.

При этом все BI-приложения и ELT/ETL-инструменты могут развертываться в одном и том же DCA с помощью новых Greenplum Data Integration Accelerator модулей. Как результат — унифицированная Big Data платформа для структурированных и неструктурированных данных и приложений в составе одной инфраструктуры с общим мониторингом, управлением и поддержкой от одного вендора — EMC.

В итоге, в составе DCA могут одновременно сочетаться четыре модуля:

- *Greenplum Database Module* — высоко-масштабируемый DW-модуль, который архитектурно интегрирует базу данных, вычисления, хранение и сеть в одной системе. Оптимизирован с точки зрения цены и производительности;
- *Greenplum Database High Capacity Module* — предназначен для петабайтных размеров хранилищ данных. Оптимизирован с точки зрения цены за единицу хранения при высокой производительности;
- *Greenplum HD Module* — высокопроизводительный Hadoop-appliance. Обеспечивает бесшовную обработку структурированных и неструктурированных данных в пределах одного решения;
- *Greenplum Data Integration Accelerator (DIA) Module* — предназначен для хостирования партнерских аналитических

приложений и подключения их на одном и том же высокопроизводительном интерконнекте с низкими задержками.

DCA можно развертывать, начиная с одного стандартного модуля (или высокоемкого Greenplum Database) и двух мастер-серверов, которые ответственны за аутентификацию, оптимизацию очереди, балансировку нагрузки среди различных сегментных серверов, управление отказоустойчивостью данных и других механизмов в кластере. В дальнейшем решение может масштабироваться, используя Greenplum Database, Greenplum HD или Greenplum DIA модули.

В декабре 2011 г. EMC анонсировала дальнейшее развитие платформы DCA — EMC Greenplum™ Unified Analytics Platform (UAP). Новая платформа соединяет в себе три решения для поддержки аналитики "больших данных" (рис. 1): 1) EMC Greenplum database — для структурированных данных; 2) EMC Greenplum HD — для аналитики и обработки неструктурированных данных; 3) EMC Greenplum Chorus™, представляющее собой производительную машину и социальную сеть для коллективов аналитиков. Основная цель создания UAP — максимальная интеграция результатов аналитики "больших данных" с бизнес-процессами.

В конце января 2012 г. EMC анонсировала интеграцию файловой горизонтально-масштабируемой СХД EMC Isilon с распределенной файловой системой Hadoop (Hadoop Distributed File System — HDFS), что в значительной степени расширило возможности Hadoop-кластеров по хранению данных с точки зрения надежности, доступности/скорости доступа, а также простоты управления и масштабирования.

По сути, интеграция EMC Isilon с протоколом HDFS положила начало переходу использования для хранения данных в Hadoop-кластерах — от кустарных ненадежных СХД к промышленным.

Горизонтально-масштабируемая система NAS EMC Isilon является первой и единственной корпоративной системой NAS со встроенной поддержкой слоя распределенной файловой системы Hadoop Distributed File System. Используя HDFS как протокол передачи данных, можно быстро развернуть комплексное решение для анализа "больших данных", которое состоит из системы Greenplum HD и горизонтально-масштабируемой системы NAS Isilon, и в результате можно получить мощную, высокоэффективную и гибкую экосистему хранения и анализа данных с возможностью масштабирования до 15 Пбайт при общей производительности 85 Гбайт/с в единой файловой системе.

Объединенное решение на базе Greenplum HD и Isilon — это готовая комплексная система Hadoop, которая позволит избежать сложного и ресурсоемкого развертывания. Этот подход позволяет сконцентрировать внимание на анализе своего бизнеса вместо того, чтобы тратить ценные ресурсы на преодоление технических трудностей, которые возникают при настройке кластера Hadoop и управлении им.

Дело в том, что хранимые в Hadoop данные обычно реплицируются несколько раз и распространяются по кластеру

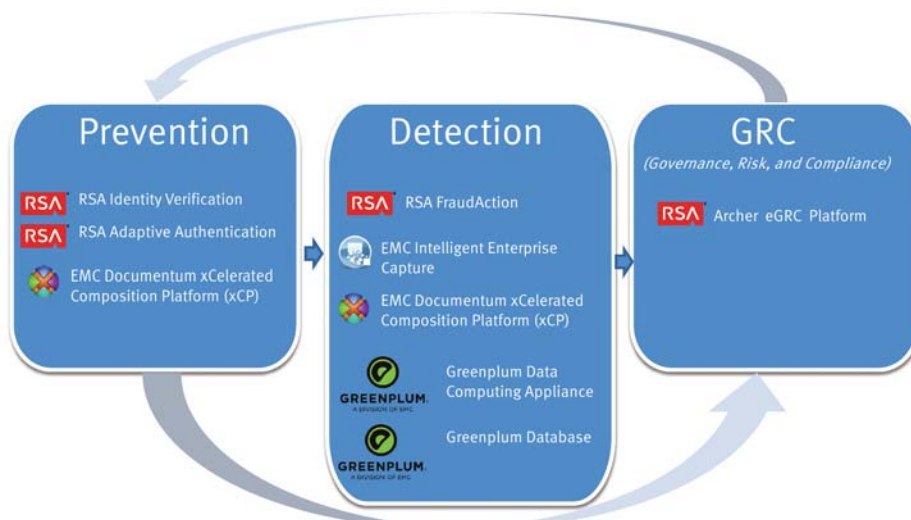


Рис. 2. Семейство решений EMC для борьбы с мошенничеством в трех областях: предотвращение (i), обнаружение (ii) и управление рисками и согласием (GRC) (iii).

Nadoop, чтобы оптимизировать производительность и надежность. При традиционном развертывании Nadoop вычисления (MapReduce) выполняются в кластере стандартных серверов, а для хранения данных используется система хранения с прямым подключением (HDFS) по сети. Метаданные для файлов, которые хранятся в HDFS, сохраняются на едином сервере, называемом NameNode. Этот традиционный подход к средам Nadoop вызывает ряд проблем:

- внедрение Nadoop и управление этим решением – технически сложная задача;
- NameNode является критической точкой отказа;
- ресурсы системы хранения с прямым подключением используются неэффективно из-за репликации HDFS;
- управление большими пулами локально подключенных дисков является сложным и дорогим;
- перенос и загрузка данных неэффективны;
- невозможно независимое повышение производительности и емкости;
- интеграция нескольких проектов с открытым исходным кодом – сложная задача;

Решение этих проблем – комплексный продукт, который сочетает в себе мощност Greenplum HD и горизонтально масштабируемую систему NAS Isilon. В результате – получилось мощное, надежное и универсальное хранилище “больших данных” для аналитиков.

## Направления внедрений

### Пример 1. Увеличение скорости аналитики для социальной сети

Tagged – социальная сеть знакомств – базируется в Сан-Франциско (Калифорния, США). Tagged ранжируется среди самых крупных социальных сетей с числом членов 100 млн и ежедневным числом просмотров страниц – 200 млн.

По мере роста числа членов сети и интенсивности ее посещения, традиционное хранилище данных для бизнес-анализа перестало удовлетворять текущим бизнес-требованиям. На обработку теку-

щих BI-запросов могло уходить несколько часов и даже целый день, а об использовании более предиктивной и расширенной аналитики не могло быть и речи. В современных условиях и быстро меняющемся бизнесе требовалась реакция в течение минут и меньше.

В связи с этим, было принято решение о переходе на специализированное хранилище данных для BI. После анализа предложений на рынке выбор был сделан в пользу EMC Greenplum Data Computing Appliance (DCA) благодаря тому, что это решение в наилучшей степени удовлетворяло ряду ключевых требований: по доказанной масштабируемости, производительности обработки данных, цене и др.

С введением DCA Tagged получила возможность анализировать более сложные наборы данных, которые включают: пользовательские профили, демографию, пользовательские логи, контакты в сети, технические данные браузеров и др. Также Tagged удалось снять все ограничения предыдущей ИТ-инфраструктуры. Теперь стало возможным выполнять непрерывную загрузку данных, аналитику в реальном времени и получать более детальные отчеты.

В результате более продуманной политики сервисов, время, в течение которого посетитель оставался на сайте, увеличилось на 50%, что, в свою очередь, увеличило доход от рекламы. Также повысилась производительность администраторов баз данных и сотрудников по развитию бизнеса.

### Пример 2. Борьба с кибермошенничеством в госструктурах

Борьба с кибермошенничеством – одно из ключевых направлений предотвращения неоправданных потерь в экономике для большинства стран современного мира. В соответствии с отчетом Gartner<sup>1)</sup> и данными Национальной Ассоциации Антимошенничества в Здравоохранении (National Healthcare Anti-Fraud Associa-

tion – NHCAA), из более чем \$2 трлн, выделяемых ежегодно американским правительством на здравоохранение, около 3%, или \$60 млрд, расхищаются.

В этой связи, системы, которые могут предотвращать, идентифицировать и сокращать мошенничество, начинают получать приоритетность в списке затрат американского правительства. Так, например, в сентябре 2011 г. Департамент труда правительства США выделил \$191 млн сорока штатам округов Колумбия и Пуэрто-Рико для модернизации системы социального страхования по безработице с целью предотвращения и снижения несоответствующих выплат<sup>2)</sup>. Также Министерство здравоохранения США (U.S. Department of Health and Human Services) в бюджет 2012 г. включило запрос на \$581 млн для улучшения контроля за выделением финансов в Центрах бесплатной медицинской помощи и на программы медицинской помощи<sup>3)</sup>.

В настоящее время нет ни одного решения, которое могло бы полностью предотвращать кибер-мошенничество, но снижение его уровня уже возможно. Так, EMC предоставляет основные ИТ-технологии (рис. 2), которые могут играть главную роль при проектировании решений, нацеленных на проблемы в трех ключевых областях: предотвращение (i), обнаружение (ii) и управление рисками и согласием (GRC) (iii), и которые обеспечивают:

- мониторинг и аутентификацию активности пользователей, основываясь на уровне рисков и организационных политиках;
- подтверждение пользовательской идентичности в реальном времени, снижая риски, связанные с ложной идентификацией;
- предотвращение фишинговой, phishing и троянские атак в онлайн-канале;
- обнаружение мошенничества посредством анализа в реальном масштабе времени многих источников данных, как структурированных, так и неструктурированных;
- генерирование предупреждающих сообщений, когда идентифицируется активность вне рамок стандартных правил или/и подозрительное поведение;
- консолидированное управление рисками и комплайном информации любого типа;
- унификация политик и контроля, а также оценки рисков;
- принятие соответствующих мер для блокирования попыток мошенничества.

## Заключение

Аналитика “больших данных” – стремительно развивающееся направление бизнес-анализа, которое, по оценкам многих экспертов, уже в ближайшем будущем может произвести целую революцию как в бизнесе, так и в обществе, благодаря возможности получения новых знаний, в реальном времени извлекаемых из информации, которая до недавнего времени была не подвластна какой-либо аналитике вообще.

Денис Серов,  
руководитель направления технического консультирования “EMC Россия и СНГ”

1) O’Neil, Maureen and Howard, Rick. “U.S. Healthcare Reform Advances Fraud Prevention.” Gartner Industry Research (2010): 2. 17 Sep 2011.

2) United States Department of Labor, “U.S. Department of Labor Awards \$191 Million of Unemployment Insurance Modernization,” Sep 14, 2011

3) O’Brien, Adelaide. “Business Strategy: U.S. Government 2011 IT Spending and Key Business Imperatives.” IDC Government Insights. #G I229587. (2011): p. 7. Web. 6 Sep. 2011.