

Интегрированные платформы для бизнес-аналитики

Современные условия ведения бизнеса как формируют новые задачи для бизнес-аналитики (Business Analytics – BA), так и выдвигают новые требования к уже решаемым задачам BA. Среди новых задач это, прежде всего, “аналитика больших данных”, среди новых требований – максимальное приближение решения задач бизнес-анализа к реальному времени. Появившиеся около 5 лет назад и активно продвигаемые в последний год специализированные аппаратные платформы, жестко интегрируемые с приложениями бизнес-анализа, способны в полной мере ответить новым вызовам времени. Своим видением развития нового направления подобных решений с SN делятся Сергей Золоторев – директор по продажам, Россия и СНГ Greenplum (подразделение EMC) и Андрей Свирщевский – руководитель направлений Аналитики и Гарантирования Доходов компании SAS Россия/СНГ – представители компаний, заключивших в 2011 г. соглашение о совместном продвижении на рынке приложений бизнес-анализа SAS на базе пакета SAS® High-Performance Analytics и платформы EMC® Greenplum® Data Computing Appliance.

Введение

Направления “аналитика больших данных” и “специализированные хранилища данных (Data Warehouse) для BA” тесно связаны. Причины их развития в последнее время в значительной степени пересекаются, среди них можно отметить следующие. Во-первых, в последние годы резко возросли технологические возможности микроэлектроники, что позволило повысить многоядерность микропроцессоров, емкость оперативной и твердотельной памяти, скорость интерфейсов и внутренних шин компьютеров. Все это произошло при значительном удешевлении производства и самих чипов.

Во-вторых, существенно увеличились объемы корпоративной информации – от десятков терабайт до сотен и более петабайт. При под определение “значимая информация” стала попадать не только структурированная информация (хранящаяся в БД) при работе продуктивных приложений (CRM-, ERP-систем и др.), но и разнообразная файловая информация (документы, видео-, аудиоконтент, веб-контент, сообщения различного рода и др.).

В-третьих, возрос спрос на более интеллектуальные системы аналитики, позволяющие: анализировать колоссальные объемы информации – как структурированной, так и неструктурированной; учитывать множество параметров и факторов, повышающих точность предсказаний; приближать решение задач бизнес-анализа к реальному времени с целью максимально эффективного управления бизнесом.

Следствием этих тенденций (где-то с 2005 г.) – стало развитие специализированных хранилищ для BA, и в 2008 г. к этому процессу уже подключились все крупные мировые игроки. В 2011 г. волна инсталляций таких решений докатилась и до России. Эти решения позволили резко поднять долю обработки на уровне хранения данных, а также повысить параллелизм обработки запросов на десятках и сотнях процессоров на уровне хранения данных. В ре-

зультате, скорость обработки запросов выросла в 10–100 раз, а также появилась возможность работы с DW объемом до многих петабайт без снижения производительности.

Аналитику больших данных во многом связывают с проектом Hadoop с открытым кодом (проект Apache Software Foundation – 2006 г.) для хранения (на базе распределенной файловой системы HDFS) и пакетной обработки в массивно-параллельном режиме больших данных (неструктурированных или плохо структурированных данных).

SAS стала продвигать тему больших данных с конца 2011 г. и к данному моменту на базе платформы Greenplum имеет несколько реализованных проектов в России и СНГ.

В настоящее время платформа EMC Greenplum поддерживает все основные технологии SAS, связанные с обработкой больших данных:

- аналитика in-memory, включая сервер SAS LASR Analytic Server;
- аналитика in-database;
- аналитика на базе распределенных вычислений (SAS Grid Manager).

С 2012 г. платформа EMC Greenplum может одновременно поддерживать как работу продуктивных приложений, так и приложений бизнес-аналитики в отдельных пулах хранения данных. При этом интеграция данных между продуктивными БД и DW в значительной степени упрощается и ускоряется и также поддерживается в решениях SAS.

Высокопроизводительная аналитика больших данных

SN. Какие технологии развивает SAS для решения задач бизнес-аналитики на больших данных?

А.С. Большие данные создают ситуации, когда организация сталкивается с необходимостью загружать, хранить и анали-



Андрей Свирщевский – руководитель направлений Аналитики и Гарантирования Доходов компании SAS Россия/СНГ.

зировать разнородные (как структурированные, так и неструктурированные) данные, которые по своему экспоненциально растущему объему и требуемой скорости обработки не вписываются в рамки предельно допустимой нагрузки имеющихся у этой организации вычислительных мощностей. Для работы с этими данными, конечно, нужны новые инструменты и технологии. Компания SAS уделяет работе с Big Data огромное внимание. Так, разработаны и выпущены на рынок “прорывные” инструменты High Performance Analytics (или HPA), которые включают в себя: технологию 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, наконец, самую передовую технологию – in-memory analytics (применение аналитики непосредственно в оперативной памяти сервера СУБД).

Аналитические инструменты SAS HPA в интеграции со специализированными DW позволяют решать бизнес-задачи принципиально нового уровня сложно-

сти, давая возможность обрабатывать и анализировать большие объемы данных с гораздо большей точностью. Использование таких решений открывает для бизнеса невиданные ранее возможности. Более того, благодаря такому инструменту, как SAS Visual Analytics, работа с большими данными становится возможной даже для бизнес-пользователей, не имеющих специального технического образования и подготовки.

В качестве результатов использования этих инструментов можно привести следующие:

- расчет логистической регрессии на основе одного миллиарда измерений занимает порядка 60 секунд;
- один из лидирующих мировых банков уменьшил время расчетов по дефолтам с 96-ти часов всего до 4-х минут;
- время на решение задач оптимизации маркетинговых кампаний для 25 млн клиентов и 1000 различных предложений сократилось с 5,5 часов до менее чем 6-ти минут.

SN. Насколько актуально для российских предприятий решение задач, связанных с продвинутой аналитикой и большими данными?

А.С. Технологии для быстрой работы с большими данными наиболее актуальны и полезны в крупных компаниях, у которых сотни тысяч и миллионы клиентов или у которых накоплено много различной информации – о клиентах, партнерах, транзакциях, взаимодействиях, рынке, собственной операционной деятельности и т.д. В России такие компании и госструктуры также представлены. Это:

- **банки и операторы связи**, которые могут принимать решения о том, какое целевое маркетинговое предложение и по какому каналу коммуникаций сделать тому или иному клиенту. Пользуясь средствами высокопроизводительной аналитики SAS, они более точно сегментируют свою аудиторию и таргетируют предложения, получая в результате значительно более высокий отклик. Причем высокая скорость анализа дает возможность не только эффективно анализировать предпочтения и склонности клиентов, но и несколько раз изменять настройки итоговой маркетинговой кампании и просчитывать ее эффективность с тем, чтобы выбрать наиболее выгодный вариант в условиях имеющихся ограничений на бюджет и пропускную способность каналов коммуникаций;
- **розничные сети**, которые благодаря скоростному анализу больших данных, более точно устанавливают цену на каждый конкретный товар с учетом сезонных колебаний, уровня спроса, складских запасов и др. В результате вместо постепенного снижения и интуитивного "нащупывания" приемлемой цены, они сразу определяют, по какой цене оптимально продавать товар;
- **страховые компании и банки**, применяя инструменты высокопроизводительной аналитики SAS, могут намного эффективнее рассчитывать риски инвестиционного портфеля на рынках

капитала. Это достигается за счет более точного построения моделей рисков. Регулярный анализ кредитного портфеля помогает вовремя принять меры, повысить его качество, снизить долю "плохих" активов.

Аналогичная ситуация – и с оценкой рисков при кредитовании (расчет скоринг-балла для каждого заемщика). При этом за счет высокопроизводительной аналитики появляется возможность при посещении клиентом банка оценивать не только всю его кредитную историю, но и его психофизиологическое состояние, мотивацию поведения и др.

Набирают популярность в этом секторе рынка решения Big Data для борьбы с мошенничеством – SAS Fraud Framework. По нашим оценкам, тема мошенничества вызывает экспоненциальный рост интереса со стороны банков и страховых компаний. И не даром: например, по собственным оценкам страховщиков, минимум 8% обращений за возмещением убытков сфабрикованы мошенниками. Сегодня по мере усиления конкуренции и роста давления на мировых финансовых рынках, по мере разрастания международных криминальных группировок, включая и Россию, по мере роста интернетизации населения и все более активного участия в социальных сетях у финансовых институтов появляется понимание необходимости проводить детальный анализ своих данных с помощью современных аналитических решений. Это позволит им избежать ненужной утечки капитала и потери репутации.

Например, крупнейший в Европе по размеру рыночной капитализации банк HSBC, внедрив систему SAS по противодействию кредитному мошенничеству, повысил эффективность службы по выявлению случаев мошенничества в 3 раза, точность выявления мошенничества – в 10 раз. При этом в первые же две недели эксплуатации решения 7 специалистов службы безопасности HSBC выявили новые криминальные группы и схемы с общим потенциальным ущербом более \$10 млн;

- **предприятия ТЭК (нефтегаз, энергетика)**, которые могут использовать скоростной анализ многочисленных пока-

зателей работы оборудования для выявления факторов, влияющих на его работу, для предотвращения сбоев и отказов, анализа наступления нежелательных событий, выработки оптимального расписания технического обслуживания;

- **госструктуры**, которые с помощью SAS Fraud Framework могут бороться с мошенничеством. Например, для каждой декларации, каждого таможенно-го груза рассчитать вероятность мошенничества. Это то, что в мире сейчас активно внедряется и применяется.

Для бизнес- и госструктур в последнее время особую актуальность приобретает анализ жалоб и обращений, включая и гарантийное обслуживание. Это – как раз тема больших данных: информации тут много, она и структурированная, и неструктурированная, и к ней нужно применять комплексные методы, то есть и text mining, и data mining, и статистический анализ и др. Высокоскоростной анализ жалоб и предложений помогает своевременно понять, на что стоит обратить внимание, принять решение об изменении стандартов обслуживания или замене субподрядчиков и поставщиков сырья для производства, оборудования, запчастей.

Построенная в Гонконге система интеллектуального анализа обращений граждан позволила:

- быстро выявлять основные причины недовольства граждан;
- повысить качество и скорость реагирования на обращения;
- планировать стратегию развития города;
- повысить удовлетворенность горожан качеством жизни.

Необходимо также отметить, что с появлением различных технологий высокопроизводительной аналитики для больших данных становится доступным анализ неструктурированной информации о клиентах, например, получаемой из интернета или в процессе работы с ними (обслуживания их). Такая информация может быть использована как для решения задач целевого маркетинга, так и для управления рисками и предотвращения мошенничества. Это происходит за счет возможности добавить к Интеллектуальному Анализу Данных (Data Mining) Ин-



Рис. 1. Высокопроизводительная аналитика: от простых отчетов – к прогнозированию будущего и оптимизации результата.

теллектуальный Анализ Текста (Text Mining).

SN. Как меняется характер работы сотрудников в компании при внедрении решений высокопроизводительной аналитики?

А.С. Во-первых, с появлением высокопроизводительной аналитики меняется сам характер задач. Появляется возможность перестать воспринимать слово "аналитика" как аналитическую отчетность с ответами на вопросы "что было и происходило" и перейти к проактивной – прогнозной аналитике – и отвечать на вопросы "что будет происходить", "что будет происходить, если я поступлю так или иначе" (рис. 1). Далее, когда компания накапливает достаточно информации о принятых решениях и их последствиях, то уже актуально переходить от прогнозирования и прогнозного моделирования к оптимизации – "что конкретно и как нужно сделать, чтобы добиться оптимального результата". И тут – очень много классов задач. Это может быть оптимизация расписаний, например, поездов, технического обслуживания или работы сотрудников в отделениях банка. Или оптимизация распределения ресурсов, например, электроэнергии или бюджета.

Во-вторых, возрастает скорость и точность принимаемых решений. Аналитик, благодаря высокой скорости обработки данных, может изменять настройки и запускать выявление закономерностей и построение модели несколько раз, применять больше алгоритмов и в более сложной конфигурации. Также появляется возможность анализировать данные не на выборках, как раньше, а на всей генеральной совокупности.

В-третьих, расширяется круг лиц, связанных с аналитикой. Например, система SAS Visual Analytics, сочетая в себе высокопроизводительные средства анализа и графический интерфейс визуализации данных, позволяет пользователям разной квалификации, не прибегая к помощи ИТ-специа-

листов, анализировать большие объемы данных и получать точные, наглядные и удобные для интерпретации отчеты. В процессе исследования пользователи могут создавать произвольные иерархии данных для многомерного анализа и применять фильтры на любом уровне иерархии, агрегировать и детализировать данные, проводить вычисления по новым показателям и добавлять их к любому представлению. Полученные отчеты можно публиковать в интернете, а благодаря приложениям для мобильных устройств с ними можно работать, например, на iPad.

Эти же возможности SAS Visual Analytics позволяют каждому руководителю самостоятельно, работая с большими данными, посмотреть и понять, что происходило с бизнесом (в том числе и необычного) и требует внимания, по каким закономерностям развивается бизнес и к чему приведет текущие тенденции.

В-четвертых, поскольку пользователи работают с данными напрямую, самостоятельно, не делая для этого никаких запросов в ИТ-службу, то меняется весь бизнес-процесс проведения анализа и принятия решений. В свою очередь, ИТ-специалисты освобождаются от рутинных операций по обработке пользовательских запросов на предоставление доступа к информации и создание новых представлений данных и разовых отчетов.

EMC DCA: платформа для высокопроизводительной аналитики больших данных

SN. Какие технологии SAS, ориентированные на высокопроизводительную аналитику, поддерживает платформа EMC DCA?

С.З. В настоящее время платформа Greenplum DCA поддерживает все основные технологии SAS для высокопроизводительной аналитики – SAS In-Database, SAS In-Memory, SAS Grid Computing", позволяя при этом полностью отказаться



Сергей Золоторёв – директор по продажам, Россия и СНГ Greenplum (подразделение EMC).

от использования каких-либо выделенных серверов аналитики вне платформы Greenplum DCA и оставляя на стороне клиента только простейшие средства (iPad, iPhone, ноутбук и др.) для визуализации результатов.

В развитие поддержки SAS-технологий EMC предлагает также несколько решений, интегрирующих вышеназванные технологии SAS для различных задач/ применений:

- SAS/Access – интерфейс/язык для управления компонентами инфраструктуры GP DCA и вычислениями над данными, который:
 - поддерживает работу на всех платформах SAS;
 - поддерживает функции SQL в БД;
 - обеспечивает прямой доступ к таблицам GP;
 - поддерживает высокопроизводительную пакетную загрузку-выгрузку;
 - обработку данных в БД (например, команды FREQ, SORT, SUMMARY/MEANS, REPORT, RANK, TABULATE);

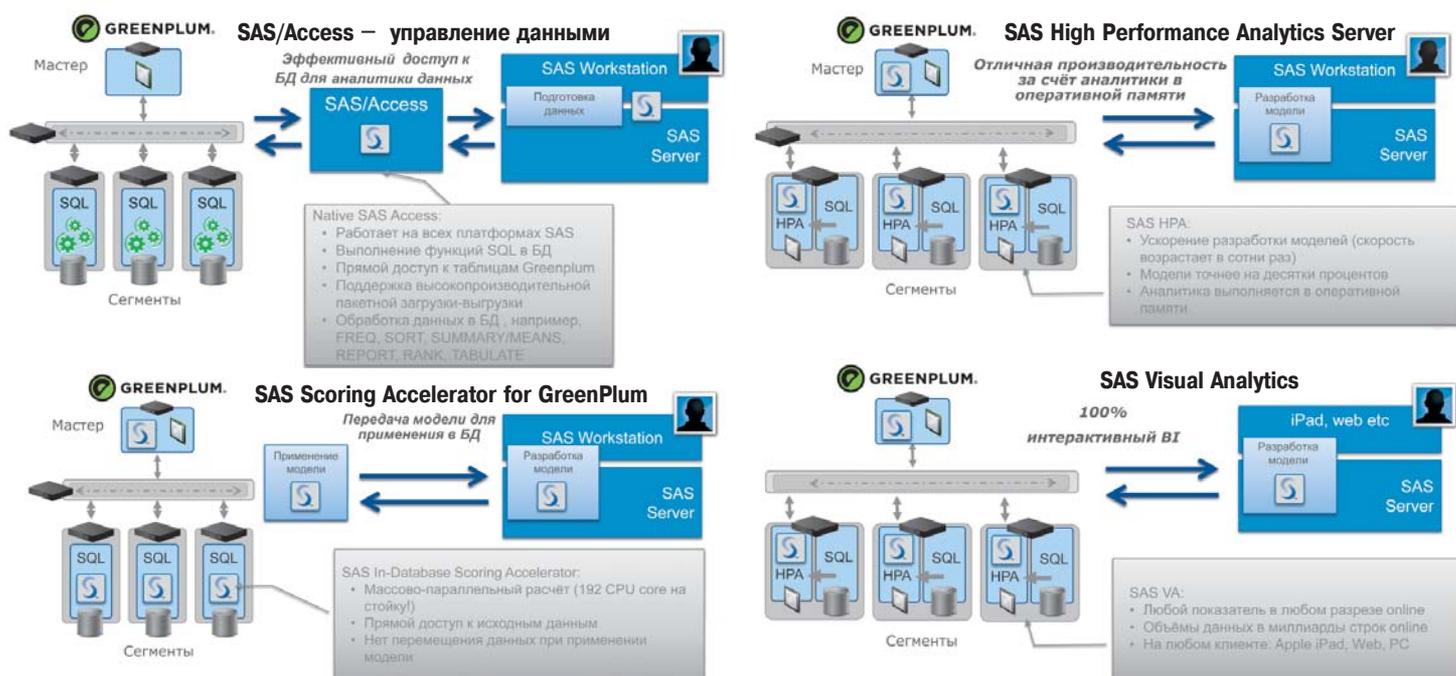


Рис. 2. Четыре технологии EMC на базе EMC DCA, позволяющие интегрировать особенности трех базовых технологий SAS и расширить возможности SAS-приложений: SAS/Access – управление данными; SAS Scoring Accelerator for GreenPlum; SAS High Performance Analytics Server; SAS Visual Analytics.

– *SAS Scoring Accelerator for GreenPlum* – аналитическая модель, например, для оценки кредитного рейтинга заемщика, которая устанавливается как хранящая процедура непосредственно на уровне хранилища GP. При этом рейтинги сотен тысяч и даже миллионов заемщиков могут пересчитываться многократно по нескольку раз в день при различных параметрах. Это достигается за счет:

- поддержки массово-параллельных расчетов (до 192 ядер на стойку);
- прямого доступа к исходным данным;
- отсутствия перемещения данных при применении модели.

При увеличении числа клиентов БД может линейно масштабироваться без снижения производительности;

– *SAS High Performance Analytics Server (SAS HPA Server)*. На GP устанавливается SAS in-темогу компонент, который может быстро “поднимать” данные из БД на каждом узле в ОП, обрабатывать их и возвращать результат клиенту. Например, необходимо ускорить разработку какой-либо модели или повысить ее точность. Если это касается линейной регрессии, то в ней очень сложно подобрать коэффициенты/переменные, которые в наибольшей степени влияют на поведение модели. Или, если ухудшилась финансовая ситуация или поменялось законодательство, то эффективность модели может резко снизиться и потребуются ее тюнинг к изменившимся условиям. Благодаря технологии SAS HPA Server можно генерировать огромное количество (сотни) идей, моделей и моментально проверять их. В дальнейшем выбранная модель может быть реализована в качестве, например, процедуры SAS Scoring Accelerator. За счет SAS HPA Server достигается:

- ускорение разработки моделей в десятки и даже сотни раз;
- увеличение точности моделей на десятки процентов;

– *SAS Visual Analytics* – простая в использовании система для быстрого изучения и визуализации данных. Поставляется SAS и EMC Greenplum, как совместно-разработанный программно-аппаратный комплекс. За счет высокопроизводительной платформы GP появляется возможность:

- оценки любого показателя в любом разрезе в режиме online;
- анализа данных объемом в миллиарды строк в режиме online;
- получения результатов на любом клиенте: Apple iPad, Web, PC и др.

Для работы “чистых” SAS-технологий (In-Database, In-Memory, Grid Computing) в Greenplum DCA предоставляется возможность формирования выделенных пулов ресурсов на базе модулей EMC DIA SAS Grid и SAS HPA для GP DB (каждый модуль в своем составе имеет: емкость 9 Тбайт – до сжатия; 2 CPU – 12 ядер; 48 Гбайт ОП; 2,25 Тбайт полезной емкости). При этом SAS Grid Computing полностью сертифицирован для выполнения в составе Greenplum DCA.

SN. Какие особенности платформы Greenplum DCA отличают ее от аналогичных, представленных на рынке?

С.З. Таких особенностей достаточно много. Из наиболее интересных хотелось бы отметить следующие три:

– *полиморфное хранение (Polymorphic Table Storage™)*. С версии 3.3.4 в рамках одной таблицы EMC Greenplum стало возможным хранить данные в поколоночном и построчном форматах. Те данные, которые постоянно изменяются, эффективнее хранить в построчном формате, с точки зрения скорости обращения к ним. Те данные, которые становятся историческими (архивными, т.е. остаются неизменными), эффективнее хранить в поколоночном формате, благодаря чему снижаются объемы БД до 10 раз и одновременно увеличивается скорость обработки OLAP-запросов;

– *большие возможности по управлению производительностью*, которые обеспечиваются высокоэффективным управлением пулом подключенных пользователей и приоритизацией запросов за счет:

- управления числом подключенных пользователей;
- освобождения и использования ресурсов простаивающих сессий;
- управления приоритетами пользовательских очередей;
- выполнения динамической балансировки ресурсов между выполняемыми запросами (например, от времени суток, важности и др.);

– *высокая скорость выполнения процедур ETL (Extract, Transform, Load)*. В настоящее время стойка GP DCA поддерживает скорость загрузки данных до 10 Тбайт/час. Соответственно, если добавляются дополнительные стойки, скорость линейно масштабируется – это один из важных дифференциаторов GP на рынке. Как правило, узким местом в ETL-процедурах является трансформация данных. И очень немногие производители подобных решений могут ее сделать по системе grid. На GP процедуры ETL переводятся ELT, что обеспечивает выполнение трансформации данных на множестве параллельно работающих серверов с линейной масштабируемостью, и это поддерживается всеми партнерами GP.

SN. Одно из самых быстро развивающихся направлений – аналитика неструктурированных данных. Что в этом плане привносит Greenplum DCA на рынок и как DCA интегрируется с решениями SAS для этих задач?

С.З. Вы совершенно правы – интерес к этой теме стремительно возрастает. Аналитику неструктурированных данных, в основном, связывают с проектом Hadoop. Родившись в Google, и с первой реализацией в Yahoo, Hadoop первоначально использовался в различных web 2.0 компаниях. Сейчас он активно используется и развивается в научных, образовательных учреждениях, а также большой интерес к нему проявляют банки и телеком (рис. 3).

Основное, что привносит GP в работу бизнес-аналитиков при работе с неструктурированными данными на базе Hadoop – это упрощение их использования. Благодаря этому, пользователи SAS для анализа файловых данных обращаются к Hadoop так же, как к любому другому источнику (например, реляционной СУБД). Это достигается за счет:

– возможности установки модулей EMC Hadoop или модулей EMC Hadoop HDFS на базе Isilon непосредственно в Greenplum DCA (с полной их интеграцией в архитектуру GP);

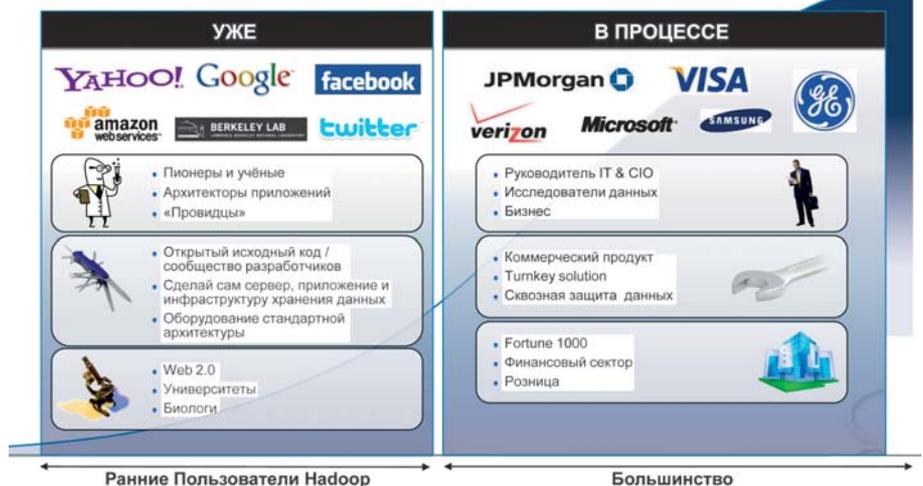


Рис. 3. Эволюция рынка Hadoop.

*) Greenplum DCA поддерживает все типы In-Database Analytics: встроенную (Greenplum DB, Greenplum Spatial), партнерскую (SAS Scoring Accelerator, SAS/HPA и др.), Open-Source (MADlib), настроенную, пользовательскую, прим. ред.

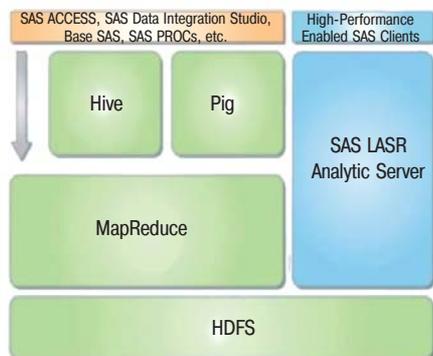


Рис. 4. Архитектура интеграции решений EMC и SAS при работе с Hadoop-кластерами.

— развертывания программного дистрибутива EMC Hadoop HDFS на отдельном выделенном пуле серверов с расширенными возможностями в части:

- скорости обработки запросов (благодаря модулям, написанным на C, вместо стандартных Java-скриптов, поддерживаемых интерпретатором);
- большей управляемости кластера серверов, поддерживающих Hadoop HDFS;
- большей устойчивости от потери данных;
- высокой доступности, без единой точки отказа и поддержкой 24/7.

Hadoop — это прежде всего распределенная файловая система — HDFS, которая обеспечивает хранение данных на тысячах узлах кластера, и MapReduce — специальный программный фреймворк, который позволяет организовать распределенную обработку данных в кластере, например, вычленивать какие-то ключевые слова из логов и др. (рис. 4). MapReduce позволяет писать логику запросов и гарантирует ее правильное применение на тысячах узлов и возвращение результатов, обеспечивая при этом толерантность к потере узлов, данных и т.п. Все, что представлено на рис. 4 зеленым цветом, может быть реализовано на уровне GP.

SAS LASR Analytic Server не требуется использование MapReduce, он может напрямую загружать данные из HDFS и работать с ними. При этом сам SAS LASR Analytic Server может полностью разворачиваться на серверах GP.

Также к hadoop-данным можно обращаться через SAS ACCESS, используя стандартные компоненты Hadoop — Hive, Pig и MapReduce. При этом все интегрировано в рамках одной платформы GP.

SN. Были ли уже внедрения решений EMC и SAS на базе Greenplum и какие результаты были достигнуты?

С.З. В 2012 году первое в России решение такого рода было внедрено в банке «Тинькофф Кредитные Системы». Это один из ведущих банков России, обслуживающих частных лиц. Банк «Тинькофф Кредитные Системы» на протяжении последних нескольких лет демонстрирует высочайшие темпы роста портфе-

ля кредитных карт, которые стали возможными за счет уникального сочетания характеристик продукта, уровня клиентского обслуживания и персонализированного маркетинга, основанных на глубокой аналитике потребностей потенциальных и существующих клиентов.

Основными критериями, определившими выбор Банка в пользу EMC Greenplum, стали: высочайшая скорость загрузки и обработки данных, масштабируемость решения, возможность полиморфного хранения данных, наличие технологии сжатия, возможность интеграции с используемыми Банком аналитическими продуктами компании SAS.

Соответствие перечисленным выше требованиям было подтверждено результатами полномасштабного тестирования производительности, проведенного специалистами Банка в лаборатории EMC в условиях максимально приближенных к реальной ситуации (Proof of Concept).

Во время внедрения проекта существовавшая в Банке инфраструктура в виде стандартной СУБД, работающая на тяжелых серверах SPARC-архитектуры, была заменена системой EMC Greenplum. В ходе проекта был осуществлен комплексный реинжиниринг более 350 процессов загрузки и преобразования данных. В результате, время решения аналитических задач сократилось минимум в 10 раз, а для некоторых — более чем в 100 раз.

Еще интересный пример проекта внедрения Greenplum — компания «Тройка Диалог» (теперь входящая в состав Сбербанка). Цель проекта — создание хранилища данных для отчетности. Внедрение позволило:

- ускорить предоставление отчетности. Теперь запросы стали обрабатываться в 10–100 раз быстрее, а результаты стали доступны в течение минут (ранее это требовало дней);
- разгрузить БД бэк-офисной системы;
- создать линейно-масштабируемую платформу для развития нового функционала, например, алгоритмической торговли, системы мониторинга нор-

мативных требований, количественного анализа и др.;

- оптимизировать затраты на ИТ-инфраструктуру.

Еще несколько совместных проектов сейчас в работе.

SN. Развернут ли в России демо-центр для тестирования интегрированных решений EMC DCA и приложений SAS для бизнес-анализа на задачах пользователя?

С.З. Да такой центр располагается в Москве. Он начал работать с конца 2012 г., и в настоящий момент у наших заказчиков есть возможность лично убедиться в преимуществах данного решения, оттестировав свои реальные задачи на интегрированной платформе Greenplum DCA в совместном демо-центре EMC и SAS.

Заключение

Интегрированные платформы на базе Greenplum DCA и приложений для бизнес-аналитики от SAS позволяют решать весь комплекс задач, охватывающих: анализ закономерностей, Data Mining, прогнозирование временных рядов, оптимизацию, визуализацию данных, отчетность. При этом решения адаптируются практически для всех вертикальных рынков: разведка, добыча и переработка нефти и газа; поставка и реализация продукции; торговля тепловой и электроэнергией (включая тарифное моделирование); телеком; банки; страховые компании и др. (рис. 5).

В большинстве случаев за счет увеличения производительности обработки запросов в десятки и сотни раз открываются совершенно новые перспективы для управления бизнес-процессами, включая стратегическое видение их развития. Одновременно обеспечивается линейная масштабируемость по производительности и объему хранилища по требованию на многие годы вперед.

За счет анализа колоссальных объемов как структурированных, так и неструктурированных данных появляется возможность гораздо более глубокого понимания происходящих процессов, особенностей поведения и интересов клиентов или отдельных индивидуумов и др.

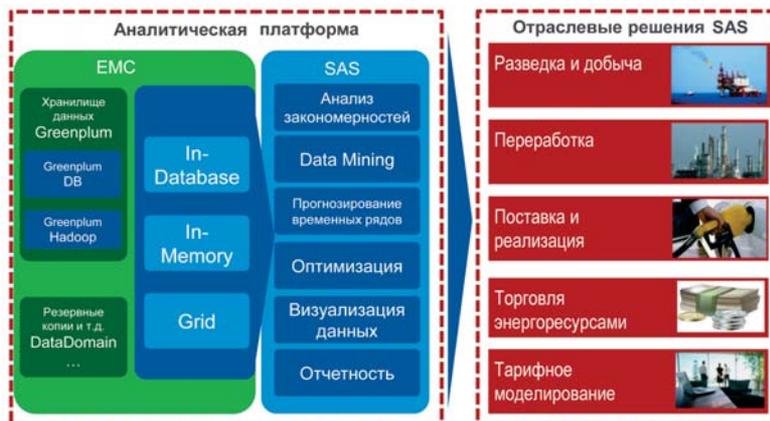


Рис. 5. Высокопроизводительная аналитическая платформа на базе решений SAS и EMC позволяет решать задачи бизнес-аналитики для большинства вертикальных рынков.