

Тяжелые времена – лучшая точка входа на рынок

В Москве состоялся Teradata Форум – одно из крупнейших в России мероприятий, посвященных эффективному ведению бизнеса, основанного на аналитике данных. Форум собрал представителей ведущих российских компаний: руководителей бизнес направлений, аналитиков, ИТ-специалистов, исследователей данных и маркетологов. Эксперты Teradata, а также представители компаний-клиентов рассказали о новейших технологиях в области аналитики и собственном опыте хранения и обработки больших данных. Речь шла о переходе от реактивной модели взаимодействия, когда задачи, поставленные бизнесом решаются в индивидуальном порядке, к проактивной, где бизнес и ИТ становятся полноценными партнерами и часто работают на опережение в разработке различных бизнес-сценариев. Следуя этой философии, мы побеседовали со Стивеном Бробстом, главным техническим директором корпорации Teradata.

– Компания Teradata известна на российском рынке приложениями для анализа данных и построения моделей, то есть офлайн-аналитикой. Какие новые решения Вы подготовили для российского финансового рынка, чтобы предложить их уже сейчас и в ближайшей перспективе?

Стивен Бробст: У Teradata есть продукты, которые могут быть интересны российскому финансовому рынку. Напомню, что под офлайн-аналитикой мы понимаем возможность строить модели и работать по ним. Сюда включаются «активные хранилища данных» – умение обслуживать процессы в режиме реального времени. В России это работает в банке ВТБ24. На Форуме в Москве о построении и использовании единого корпоративного хранилища данных на базе технологий Teradata рассказал Денис Гузовский. Обсуждалась унификация аналитических процессов в банке при помощи активных хранилищ данных, которые сегодня представляют собой не только платформу для отчетности, но и основу для проведения ключевых банковских бизнес-процессов по продажам и обслуживанию клиентов. Это наглядный пример реализации активного хранилища данных, когда у банка появляется возможность принимать операционные бизнес-решения в режиме реального времени и использовать для этого аналитику.

Сегодня мы предлагаем российскому финансовому рынку технологию, которая анализирует структуры данных в реальном времени, загружает их в систему хранилища данных Teradata. Нам удается обеспечивать обработку потоков событий – обработку в режиме реального времени – благодаря тому, что работа происходит с данными разных типов. Они загружаются как в систему хранилища данных Teradata, так и в другие системы.

Несколько лет назад мы приобрели компанию Aster Data Systems. Ее программное обеспечение обнаруживает всевозможные паттерны, закономерности, взаимосвязи и структуры данных. Сейчас Teradata выпускает новую систему – пакет приложений, объединивший различные паттерны, которые можно использовать в контексте финансовых услуг, и определять вероятность мошенничества клиентов или ухода их к другому поставщику услуг. Система может использоваться как отдельно, так и поверх открытых систем обработки данных.

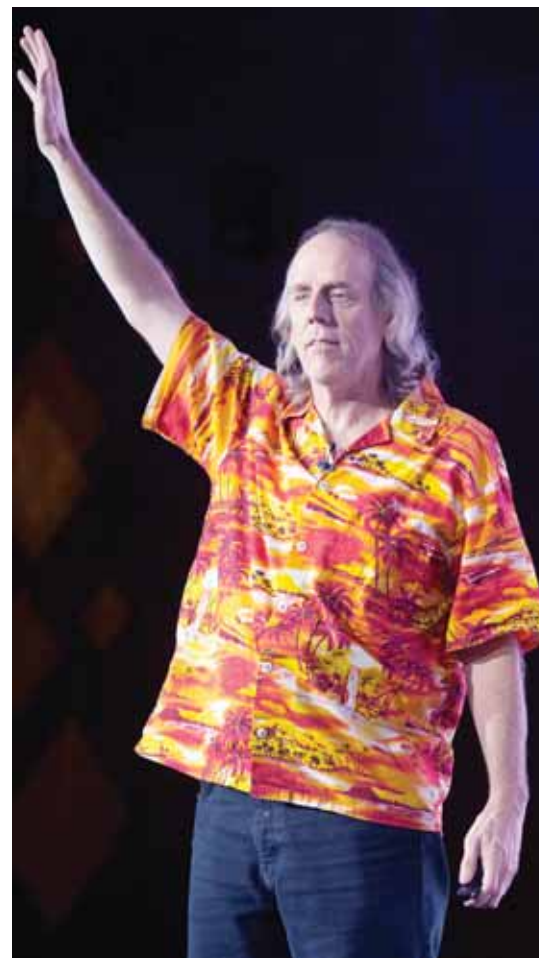
– Тема антифрода становится все более популярной у многих иностранных компаний, которые предлагают российскому финансовому рынку системы обработки данных. К примеру, компания SAS не так давно провела презентацию в Москве такого функционала. В чем специфика Вашего продукта?

Стивен Бробст: Наш конек – высокая производительность и масштабируемость обработки данных. Поскольку вы упомянули SAS, приведу пример работы крупного банка в США. Замечу, что Teradata и SAS – партнеры в части реализации параллельных вычислений и их интеграции в систему аналитики данных. Американский банк использовал систему противодействия отъему денег SAS. Банк выделил для этого один сервер. Затем этот банк приобрел много других банков. Система, соответственно, разрослась, появились и другие серверы, но все равно они не успевали обрабатывать данные в соответствии с требованиями действующего законодательства (напомню, что суммы сделок, превышающие \$10 тыс., необходимо оценить по методологии AML в течение двух часов). Тогда мы перенесли код SAS

в контекст вычислений внутри базы данных, масштабировали систему, параллелизовали вычисления внутри базы данных. То есть, мы не стали терять время на то, чтобы переносить данные из базы данных на сервер, где бы происходили вычисления, а сэкономили ресурсы, и то, что раньше SAS не успевал сделать за два часа, теперь база данных делает за несколько минут. Teradata не заменяет решение SAS, а помогает ему быстрее выполнять свои задачи (мы можем делать это с открытым исходным кодом).

– В настоящее время технологии стремительно развиваются, развиваются и методы анализа данных, появляются новые, которые, возможно, не были доступны ранее. Что привнесла Teradata за последнее время в методы аналитики?

Стивен Бробст: Раньше финансовые институты, в основном, работали с транзакционными данными, которые хранились в базах данных, к примеру, банков. Чтобы обращаться к этим данным использовался язык структурированных запросов. Сейчас во многих банках происходит процесс цифровизации. Действительно, если раньше чтобы что-то обсудить требовалось прийти в банк и обратиться к менеджеру. Сейчас, если, к примеру, требуется принять решение об инвестициях или еще о чем-то, все чаще используют онлайн-чаты. Клиент заходит на сайт банка и взаимодействует с менеджером. Данные чата записываются, анализируются и на основании полученных результатов делаются выводы о том, каковы предпочтения или особенности инвестирования конкретного клиента. Но для этого требуется более сложная система анализа, потому что данные чата на сайте не структурированы, к примеру, там могут быть опечатки. Банкам часто требуется сегментация клиентов по географическому признаку – региону их размещения или проживания. Замечено, что у людей, проживающих на разных территориях не одинаковые предпочтения. То же можно сказать и о предпочтениях в выборе формата коммуникации с банком. Так, в одних регионах люди определенной возрастной категории (к примеру, старше 30 лет) могут выбирать голосовые или видео-коммуникации. Текстовая, голосовая, видеoinформация – все это «грязные» данные, и для того, чтобы во всем этом разобраться, нужно использовать алгоритмы с применением нечеткой логики. Сравнительно недавно появилась аналитика больших данных. Как раз это –



Стивен Бробст (Stephen Brobst), главный технический директор корпорации Teradata, окончил Массачусетский Технологический Институт (Massachusetts Institute of Technology), где его исследования в рамках дипломной работы и кандидатской диссертации сосредоточились на высокопроизводительной параллельной обработке данных. Получил степень MBA в рамках совместной программы обучения и защиты диссертации Гарвардской бизнес-школы и Школы менеджмента Слоана при МТИ. Недавно он занял четвертое место в рейтинге ExecRank среди 10 тыс. технических директоров США, уступив коллегам из таких компаний, как: Amazon, Tesla Motors и Intel.

Стивен увлечен данными. Он ведет курс по визуализации продвинутой технологии уже более десяти лет в Институте хранения данных (Data Warehousing Institute) и в других организациях. В курс включено углубленное ознакомление с паттернами супер-графа геномов. Стивен является фанатом туризма и часто выбирается в походы по Новой Зеландии и США.



Текст:
Ирина
Измайлова

Отдельного внимания заслуживает прошедшая в рамках форума выставка проекта «Искусство аналитики», арт-проекта компании Teradata в области аналитики больших данных. Были представлены 18 уникальных кейсов от ведущих исследователей данных Teradata со всего мира, переданных в виде ярких аналитических картин, в которых сложнейшие математические связи становятся понятными не только продвинутому аналитикам. Примеры работ и подробное описание проекта можно найти по ссылке: www.teradata.ru/artofanalytics.

умение работать с большими объемами информации – и есть сфера компетенции Teradata.

– Какие методы позволяют из «грязных» данных извлечь верные выводы? Появились ли какие-то новые подходы и методологии?

Стивен Бробст: Из «грязных» данных невозможно сделать на все 100% верные выводы. Но можно делать вероятностные утверждения. Например, исходя из текущей информации о предпочтениях, настроениях, о риске, связанном с конкретным клиентом, можно сказать с вероятностью, например, 90%, что этот клиент уйдет в течение ближайших 30 дней, и поэтому нужно сделать какое-то вмешательство, например, как-то на него повлиять, сформулировать ему предложения и так далее. Но при этом нужно иметь в виду, что возможны ошибки первого и второго уровней: согласно предсказанию клиент может уйти в течение ближайших 30 дней, но по факту он не уходит; или, наоборот, согласно предсказанию клиент не должен был бы уйти, но по факту он уходит.

Здесь традиционные методы анализа данных, такие, к примеру, как логистическая регрессия (1), не работают, поэтому нужно использовать другие методы, например, методы нейронных сетей или машинного обучения. Это итерационные методы, которые в данном случае оказываются более эффективными.

Такие технологии известны давно, но именно сейчас, когда требуется быстро обрабатывать большие объемы данных и появились мощные компьютеры, которые способны справиться с этими данными, эти методы используются, в частности, в продуктах Teradata.

– Объемы данных стремительно нарастают. Возможно, существует угроза того, что бизнесу приходится обрабатывать слишком много данных, а результат не приносит ему ценности.

Или же такой опасности нет, и верным является утверждение: чем больше данных, тем больше ценности?

Стивен Бробст: Конечно, такое утверждение не верно: «плотность» ценности больших данных, обычно, гораздо меньше, чем у данных транзакционных, хранимых в базе банка.

К примеру, клиент звонит в службу поддержки банка. Разговор с ним записывают в звуковой файл. Затем распознают голос, конвертируя его в текст. Объем текстовой информации в сто раз меньше, чем голосового файла. Потом текстовая информация анализируется и в результате получаются факты измерения, например, настроения (насколько клиент удовлетворен) в зависимости от того, какой продукт банка используется, на какой сегмент он ориентирован и так далее. Эти данные банк занесет в базу данных Teradata, и по объему они будут в 10 тысяч раз меньше, чем исходные голосовые файлы. Поскольку в этом примере «плотность» ценности голосовых файлов гораздо ниже, то для их хранения потребуется не дорогое хранилище, например, открытое хранилище Hadoop, а результирующие данные имеет смысл поместить в хранилище Teradata. Таким образом, можно сказать, что Teradata извлекает ценности из данных, повышая их плотность.

– Где Вы видите потенциал применения решений Teradata на российском финансовом рынке?

Стивен Бробст: В более чувствительных к результату бизнес-процессах, например, для управления рисками. Это одно из наиболее частых случаев применения решений Teradata во всем мире. В России банки также заинтересованы удержать существующих и привлечь новых клиентов. Те же самые данные, которые они используют для этого, можно применять и для целей рискованной аналитики.

– Как Вы оцениваете перспективы бизнеса компании в России?

Стивен Бробст: Teradata пришла на российский рынок надолго. Финансовый рынок развивается циклично. Периоды падения неизбежно сменяют периоды роста. А тяжелые времена – лучшая точка входа на рынок, потому что именно в это время банки начинают задумываться о том, как обслу-

живать своих клиентов еще лучше, и что для этого делать. У нашей компании уже есть такой опыт.

Поскольку финансовый рынок сейчас переживает не лучшие времена, банки в России не готовы много инвестировать в системы, но начинать с малого уже пора, пока есть время на построение и наладку системы. В дальнейшем, когда рынки вырастут, соответственно, будет расти и система, ее масштабы. И тогда мы надеемся расти вместе с нашими клиентами.

Россия отличается от других стран высокой культурой научных, математических знаний. Здесь много талантливых людей, которые знают, как использовать данные. В ряде стран нам было очень трудно найти людей, которые знают и умеют работать с данными, а здесь – очень хорошая, подготовленная интеллектуальная среда.

– Сегодня в России очень динамично развивается рынок микрофинансовых организаций. Очень часто они наследуют банковских специалистов, культуру и решения технологии Teradata могут быть доступны для среднего и малого бизнеса, в том числе малым финансовым организациям?

Teradata (NYSE: TDC) помогает компаниям извлекать максимальную пользу из накопленных данных. Передовое портфолио Teradata в сегменте аналитических решений по анализу больших данных и сервисов помогает организациям получить устойчивое конкурентное преимущество на основе работы с бизнес-данными.

Стивен Бробст: В микрофинансах, где очень много маленьких клиентов с небольшими балансами, очень важна способность системы к масштабированию и к принятию решений в автоматическом режиме. У Teradata большой опыт работы с микрофинансами, например, в Южной Африке и в Индии.

– Можно ли ожидать в России появление и развитие подобных проектов?

Стивен Бробст: Да, конечно. Мы уже работаем с крупными банками, такими как Сбербанк и ВТБ24, но думаем, что будем развиваться и дальше. У нас есть приложения и для небольших организаций на финансовом рынке, которые как раз позволяют начать работать с современными технологиями, но без отвлечения на это значительных средств.

Круглый стол
БУДУЩЕ ПЕНСИОННОГО РЫНКА
25 мая 2016 г., Москва

Обсуждаемые вопросы:

- Стратегические инвестиции НПФ: практика финансирования масштабных проектов.
- Сценарии обострения конкуренции на рынке застрахованных лиц и качественных эмитентов.
- Мораторий на текущие переводы по пенсионным накоплениям: потери для рынка и акционеров.
- Как изменятся планы инвестирования пенсионных денег в связи с новациями в законодательстве?

Аналитической базой мероприятия станет исследование, включающее рейтинги крупнейших НПФ, итоги 2015 года и прогноз развития рынка на 2016–2017 годы.

www.raexpert.ru

Руководитель проекта – Дина Джиенбекова
Тел.: (495) 225-34-44 (доб. 1824), dzhienbekova@raexpert.ru

По вопросам партнерства – Мария Индьева
Тел.: (495) 225-34-44 (доб. 1896), yandjeva@raexpert.ru

(1) Логистическая регрессия или логит-регрессия (англ. logit model) — это статистическая модель, используемая для предсказания вероятности возникновения некоторого события путём подгонки данных к логистической кривой.