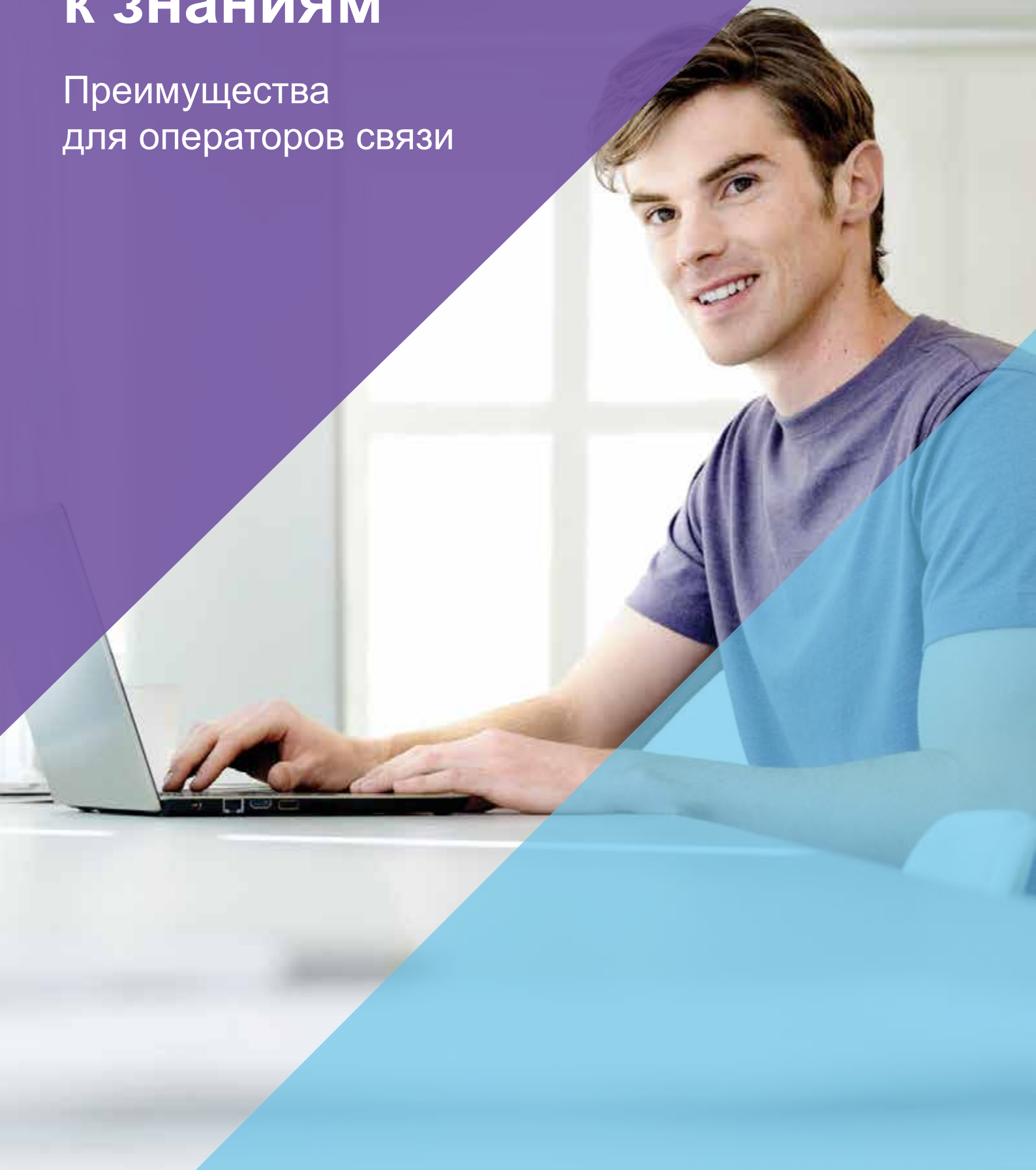


VERTICA

# От Больших данных к знаниям

Преимущества  
для операторов связи



## Краткое описание

Для операторов связи Большие данные — это возможность создать систему аналитики для более эффективной эксплуатации своих сетей, оценить выгоду от предоставления тех или иных услуг и обеспечить индивидуальный подход к клиентам. И директора по маркетингу, и вице-президенты по эксплуатации сетей, и руководители бизнес-подразделений в равной степени стремятся принимать решения, основываясь на результатах обработки значительных объемов информации. Такие компании ожидают, что их маркетологи смогут предложить им комплексные аналитические выводы, подготовленные с учетом данных, имеющихся в их ИТ- и сетевой инфраструктурах. В данном документе рассматриваются все преимущества преобразования данных операторов связи в знания. Этот процесс охватывает источники информации, инструменты сбора данных, аналитические СУБД с быстрым доступом и, наконец, сценарии использования бизнес-аналитики с представлением и визуализацией результатов и прогнозов.

## Введение

Выстраивая стратегию бизнеса, компании хотят взаимодействовать с клиентами с полным пониманием их потребностей, определять тенденции и имеющиеся возможности, своевременно выявлять случаи мошенничества и обеспечивать соответствие нормативным требованиям. Операторы связи обладают огромным количеством персональной информации о своих абонентах: об их списках контактов, адресах (домашнем, рабочем, электронной почты), об использовании Интернета и приложений, о загрузках, истории путешествий и даже о том, как долго они добираются до работы. Смартфоны становятся отражением повседневной жизни людей, источниками данных, которыми хотела бы обладать любая компания, занимающаяся маркетингом в социальных сетях.

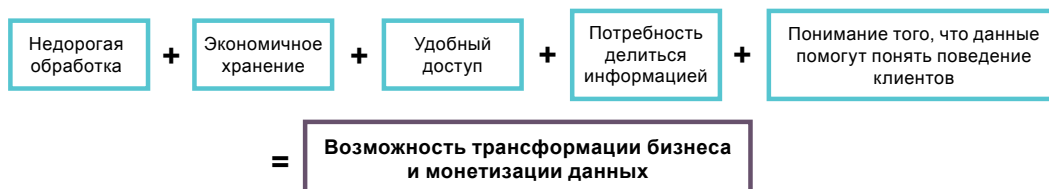
### Почему именно сейчас?

В результате эволюции технологий, происходящей на наших глазах в течение последних пяти лет, взрывной рост объемов данных стал отправной точкой для их трансформации и монетизации. Формула аналитики, представленная на рис. 1, позволила применять решения и подходы, которые еще совсем недавно были невероятно дорогими и сложными.

- Затраты на обработку данных и их анализ в режиме реального времени снижаются. Получаемые результаты теперь можно использовать для запуска необходимых в данной ситуации действий систем, а не только для создания статических отчетов, требующих дальнейшего реагирования сотрудников.
- Удобный доступ к информации, поступающей из множества источников, в том числе из Интернета, различных внутренних и внешних систем, от сетевого оборудования и датчиков, приводит к тому, что объемы данных для анализа стремительно увеличиваются.
- Стоимость хранения данных существенно снизилась. В сочетании с новыми методами обработки информации это дает возможность собирать, хранить и анализировать более крупные наборы данных.
- Новые источники данных о пользователях, а также потребность делиться информацией со смартфонов и в социальных сетях позволяют сформировать многомерное представление о клиенте.
- Данные могут пролить свет на мотивы поведения клиента, поэтому многие операторы применяют статистические методы, сложные связи между данными и графиками, анализ социальных сетей и собственных проблем бизнеса.

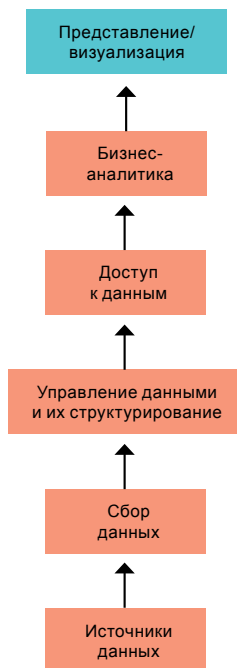
Крупная телекоммуникационная сеть и связанные с ней коммутаторы, биллинговые системы и отделы обслуживания могут многие годы ежедневно генерировать сотни миллионов записей с подробными данными о вызовах (CDR). По мере внедрения новых сервисов и увеличения объемов IP-трафика поступление данных, исчисляемых терабайтами, будет нарастать в геометрической прогрессии.

Рис. 1. Формула аналитики



## От данных к знаниям: принятие более правильных решений

Рис. 2. Преимущества Больших данных



Чтобы повысить эффективность бизнеса, операторы связи должны объединить в одно комплексное решение все компоненты, помогающие преобразовать данные в знания. К этим компонентам относятся (см. рис. 2):

- **источники данных:** от телекоммуникационных сетей, биллинговых систем и профилей абонентов до клиентских устройств и социальных сетей;
- **сбор данных:** в том числе с помощью различных технологий, таких как сетевые зонды;
- **управление данными и их структурирование:** аналитические СУБД, являющиеся, по сути, «сердцем» бизнес-знаний, предоставляют быстрый доступ к данным;
- **доступ к данным:** поддержка сессий запросов делает аналитику по-настоящему интерактивной и доступной для самых разных подразделений и сотрудников компании;
- **бизнес-аналитика:** процесс аналитики, применяемый операторами связи в конкретных сценариях использования.
- **представление и визуализация:** получение в удобном формате прогнозов и результатов исследования данных.

Рассмотрим каждый из них более подробно.

### Источники данных

Успех стратегии в области Больших данных заключается в определении различных типов их источников, использовании технологий интеллектуального анализа для поиска в каждом из них ценной информации, в последующей интеграции и представлении этих новых сведений в соответствии с уникальными целями компании, что позволит принимать эффективные управленческие решения. У операторов связи информация стекается из различных источников.

- **Использование сети:** CDR, записи с подробными данными об IP-трафике (IPDR), информация из систем поддержки бизнеса и операционной деятельности.
- **Датчики:** объем мирового рынка беспроводных датчиков, используемых в конечных вертикальных приложениях, был равен 532 млн долл. в 2010 году и 790 млн долл. в 2011-м. Ожидается, что совокупный среднегодовой темп роста (CAGR) этого рынка составит 43,1 %, и к 2016 году его объем достигнет 4,7 млрд долл<sup>1</sup>.
- **Подключенные устройства:** на сегодняшний день к Интернету подключено около 9 млрд устройств, а к 2020 году это число увеличится до 24 млрд (согласно новым статистическим данным, опубликованным GSMA)<sup>2</sup>.
- **Мобильные устройства:** общее количество мобильных устройств, подключенных к Интернету, удвоится — с 6 млрд в настоящее время до 12 млрд к 2020 году<sup>2</sup>.
- **Приложения:** Согласно прогнозам, количество загруженных по всему миру приложений в 2016 году достигнет 44 млрд, что приведет к увеличению как объемов данных, так и количества маркетинговых анализов, осуществляемых на их основе. Информация о профилях абонентов поступает из разных систем, в числе которых — домашний регистр местоположения/домашний абонентский сервер (HLR/HSS), решения для автоматической настройки телефонов (provisioning), CRM-системы.
- **Сервисы:** взаимодействие клиентов с сервисами по продаже, покупке и оказанию различных услуг позволяет анализировать их поведение, создавать целевые модели и многое другое.
- **Профили в социальных сетях:** анализ профилей, зарегистрированных в Facebook, LinkedIn, Yahoo, Google, тематических соцсетях или на сайтах о путешествиях, дает возможность собирать демографическую информацию, сведения о пользователях и потенциально интересных им ресурсах.

<sup>1</sup>«Глобальные рынки и технологии беспроводных датчиков (Global markets and technologies for wireless sensors)», код отчета: IAS042A, ZCCResearch, февраль 2012 г.

Информация, собираемая из источников данных: вызывающий абонент, вызываемый абонент, джиттер, потеря пакетов, задержка, длительность звонка, время установки соединения, кодек, пропускная способность, идентификаторы мобильного абонента и мобильного оборудования (IMSI, IMEI), номер телефона, имя пользователя, IP-адрес, MAC-адрес, дата и время входа в систему/выхода из системы, тема сообщения электронной почты/чата/веб-почты, отправитель, получатель, вложенные документы, время отклика, сеансы передачи данных (тип, содержимое, время), посещенный сайт, содержимое страницы, проведенное на сайте время, корзина, объект ссылки и т. д.

Аналитика сетевого зонда предназначена для встраивания в приложение, поэтому сведения, представляемые зондом в режиме реального времени, можно использовать для разработки приложений, обработки информации о трафике либо ее хранения для последующей подготовки отчетности или формирования трафика.

## Сбор данных

Сетевой зонд декодирует протоколы и извлекает информацию, встроенную в трафик или передаваемую поверх трафика, которая затем предоставляется (в виде метаданных и потоков контента) приложению, разработанному пользователем, для последующего применения.

Требуемая информация собирается в соответствии с заданными параметрами. Сетевой зонд выводит данные в табличном формате — точно так же, как это происходит в базе данных, оснащенной механизмом хранения. Кроме того, эта технология может доставлять пакеты и содержимое пакетов. Процесс извлечения и доставки информации из сети осуществляется в режиме реального времени со скоростью до 10 Гбит/с. Поддерживаются различные протоколы — как сетевые, так и уровня приложений (веб-почты, баз данных электронной почты или любого сетевого приложения). Для каждого протокола поставляются десятки метаданных, создающие в приложении тысячи метаданных. Эти протоколы регулярно обновляются, и в библиотеку плагинов добавляются новые протоколы.

NPE Internet Usage Manager (IUM) — широко распространенное в отрасли конвергентное решение для учета используемых сетевых ресурсов, которое поддерживает огромное количество разновидностей систем. NPE IUM обрабатывает данные для конвергентного посредничества, онлайн-учета стоимости, а также для IMS-сервисов передачи голоса и данных по проводным, беспроводным, кабельным и широкополосным сетям. В NPE IUM реализовано множество методов сбора информации в реальном времени и в режиме пакетов событий. Предлагаются различные механизмы поиска данных: FTP, SCP, HTTP, GTP, FTAM, Radius, Diameter, SIP, CSG, Cisco SCE (P-Cube), а также методы сбора локальных файлов. Поскольку все эти компоненты можно настраивать в приложении, заказчики получают весь спектр возможностей в полностью готовом к использованию продукте.

## Управление данными и их структурирование

Аналитические СУБД обеспечивают быстрый доступ к данным, дают более глубокое понимание первопричин и позволяют выводить в окне анализа больше информации, чем строчные СУБД. Многие аналитические СУБД различаются по способу хранения данных на диске. В колоночных аналитических СУБД в файлы, находящиеся на диске, записываются значения отдельных колонок, а не полных строк. Благодаря такому физическому разделению часто используемые данные можно разместить на уровне хранилища с более быстрым доступом. Кроме того, не относящиеся к запросу колонки исключаются из доступа, что ведет к повышению (иногда весьма существенному) производительности при выполнении важного для предприятия класса запросов. Кластеризация значений, повторяющихся в колонке, тоже расширяет возможности сжатия колоночных баз данных. Исключение колонок и сжатие помогают решить проблемы ввода-вывода, которые в течение многих лет возникали при аналитических запросах

Используются следующие методы сжатия:

- кодирование длины последовательностей, в результате чего значения колонок, повторяющиеся в последовательных строках, сохраняются один раз;
- алгоритм сжатия с использованием словаря, абстрагирующий реальные значения и сохраняющий в записи только маркеры (ключи);
- алгоритм дельта-сжатия, сохраняющий только отклонения от заданного значения. Некоторые аналитические СУБД (в частности, платформа Vertica) с точки зрения хранения являются «гибридными»: несколько колонок можно сохранить в одном файле на диске. Тем самым оптимизируется процесс в конце запроса, когда результирующий набор колонок сводится вместе для представления. Таблицы с большим числом колонок (как в случае с CDR) эффективнее всего преобразовать в многоколоночные дисковые файлы.

<sup>2</sup>«Интернет вещей к 2020 году охватит 24 млрд устройств (Internet of things will have 24 billion devices by 2020)», GigaOM Pro, октябрь 2011.

### Платформа аналитики Vertica

Эта аналитическая платформа позволяет операторам связи анализировать данные и принимать решения почти в режиме реального времени с непревзойденной эффективностью, производительностью и масштабируемостью.

Нашим заказчикам и партнерам из числа операторов связи приходится регулярно решать следующие задачи:

- выявление моделей поведения, которые могут привести к оттоку клиентов;
- оценка клиентского опыта на уровне транзакций и определение критериев инвестирования;
- оптимизация предложений и портфеля услуг на основе эмпирических данных, что позволяет добиться высокой рентабельности инфраструктуры и приложений в условиях жесткой конкуренции;
- анализ больших объемов данных в режиме реального времени (чего не могут обеспечить устаревшие инфраструктуры).

С помощью Vertica наши заказчики и партнеры из числа операторов связи улучшают управление мощностями, повышают производительность, масштабируемость и доступность своих аналитических систем. Вот несколько примеров улучшений:

- значительный рост показателей удовлетворенности клиентов, коэффициента их удержания и рентабельности;
- оптимизация портфеля с постепенным вытеснением высокочрезвычайных и низкоприбыльных продуктов и услуг;
- динамическое управление и масштабирование портфеля продуктов без потери подробной информации об отдельных клиентах, транзакциях или продуктах;
- хранение, доступ, анализ и монетизация огромных объемов данных о клиентах и сети без ущерба для времени обработки, масштабирования или детализации.

Благодаря инновационной архитектуре, существенно уменьшающей количество дисковых операций ввода-вывода, Vertica гарантирует на порядок более высокую производительность и значительное снижение стоимости владения. На рис. 3 приведены результаты тестов производительности, проведенных одной из телекоммуникационных компаний (на основе реального набора данных) в марте 2011 года, которые иллюстрируют преимущества Vertica.

В течение последних 30 лет в области СУБД было не так уж много инноваций. Существующие строчные СУБД оптимизированы для обработки запросов с интенсивными операциями записи, а не для аналитики с интенсивным чтением, поэтому произвольные запросы к большим объемам данных выполняются неэффективно. Пытаясь повысить производительность, пользователи строчных СУБД ежегодно тратят миллионы долларов на добавление ресурсов DBA, создание и поддержание OLAP-кубов или замену своих СУБД на дорогое, проприетарное оборудование для развертывания хранилищ данных.

**Рис. 3.** Платформа HPE Vertica обеспечила сокращение времени выполнения запросов с нескольких часов до нескольких минут при снижении стоимости оборудования более чем на 500 000 долл.

Хранилище данных CDR	Платформа аналитики Vertica	СУБД со строчным хранением данных	Преимущества Vertica
Набор данных для сравнительного тестирования	1,2 ТБайт записей с подробными данными о вызове		
Набор данных для использования в продуктивной среде	50 ТБайт записей с подробными данными о вызове		
Оборудование для сравнительного тестирования	3 сервера Dell — 2 двухъядерных процессора Opteron 2220, ОЗУ 16 ГБайт, локальный диск 1 ТБайт	24-процессорный SMP сервер, крупная SAN	Экономия на оборудовании более 500 000 долл.
Среднее время выполнения запроса (4 запроса)	5,4 минуты	2 часа	В 22 раза быстрее
Время загрузки	5,5 часа	82 часа	В 15 раз быстрее
Размер базы данных (1,2 ТБайт необработанной информации)	220 ГБайт	4 ТБайт	В 18 раз меньше емкости хранилища (коэффициент сжатия 82 %)

## Доступ к данным

Обычно поставщики аналитических платформ заявляют, что производительность обработки запросов «улучшается, невзирая на сложности», или что заказчики не в полной мере используют имеющиеся у них технологии. Если вы привязаны к стеку оборудования, единственным реальным решением этой проблемы остается наращивание аппаратных средств. Однако большинство программных систем уже полностью оптимизированы для имеющегося в компаниях аппаратного обеспечения. Если бы удалось повысить производительность выполнения запросов в бизнес-аналитике, не прибегая к аппаратным средствам, это произвело бы революцию в распространении информации, обеспечив следующие возможности:

- по-настоящему интерактивные сессии запросов (не ограниченные низкой производительностью, которая позволяет сделать только три интерактивных запроса вместо 10, 20 или 100), позволяющие получать ценные для бизнеса сведения;
- упрощение развертывания аналитической среды для всех сотрудников, клиентов, партнеров в цепи поставок и других потенциальных пользователей данных;
- хранение больших объемов ранее накопленных данных, которые могут быть запрошены;
- подробный анализ CDR, данных о посещении сайтов и прочей ресурсоемкой информации;
- расширенный анализ сложных типов данных, таких как плоские файлы, XML, графики и таблицы.

Благодаря одному только сокращению времени выполнения запросов компании смогут обеспечить соответствие нормативным требованиям в отношении сроков выполнения работ, повысить качество анализа жизненного цикла клиентов, предотвращать мошенничество, предоставлять клиентам привлекательные предложения в режиме реального времени, подробно анализировать данные и выявлять закономерности для улучшения клиентского опыта.

## Бизнес-аналитика

Решение бизнес-задач на основе анализа данных, находящихся в информационных системах, опирается на знания из области статистики и на результаты исследований операций и вычислений. Обнаружив закономерности в данных, можно использовать статистические методы для создания моделей и алгоритмов прогнозирования будущих событий. Аналитический процесс состоит из двух частей: моделирования и прогнозирования.

Моделирование начинается с процесса интеллектуального анализа данных с целью выявления закономерностей и взаимосвязей, которые позволят объяснить совокупность действий или найти аномалии в рамках сектора или кластера. После определения закономерностей и взаимосвязей создается модель, описывающая поведение: например, вероятность оттока клиентов, влияние видео на пропускную способность сети, вероятность принятия пользователями сервисов, включенных в новые пакеты услуг.

Общий потенциал аналитических инструментов можно реализовать в различных сценариях использования (см. рис. 4). Некоторые уже широко известны, но в большинстве случаев из-за уникального характера данных, требований и процессов приходится применять индивидуальный подход и осуществлять внедрение решения, разработанного с учетом особенностей бизнеса конкретного заказчика.

Чтобы подробнее узнать о примерах использования аналитических систем и о том, как они помогают принимать правильные бизнес-решения, ознакомьтесь с информационным документом «От Больших данных к знаниям: сценарии использования аналитики для операторов связи».

Платформа Vertica создана для поддержки аналитических запросов, в том числе произвольных запросов к накопленным данным и добавления новых данных (массовая загрузка или непрерывный поток для аналитики в режиме реального времени).

Частота прогона модели зависит от приложения, вычислительной мощности, объема данных и сложности самой модели. Кроме того, могут быть ограничения, связанные со скоростью поступления новых данных из их источника. С появлением более мощных технологий баз данных, таких как Vertica, время, необходимое для запуска аналитической модели, сокращается.

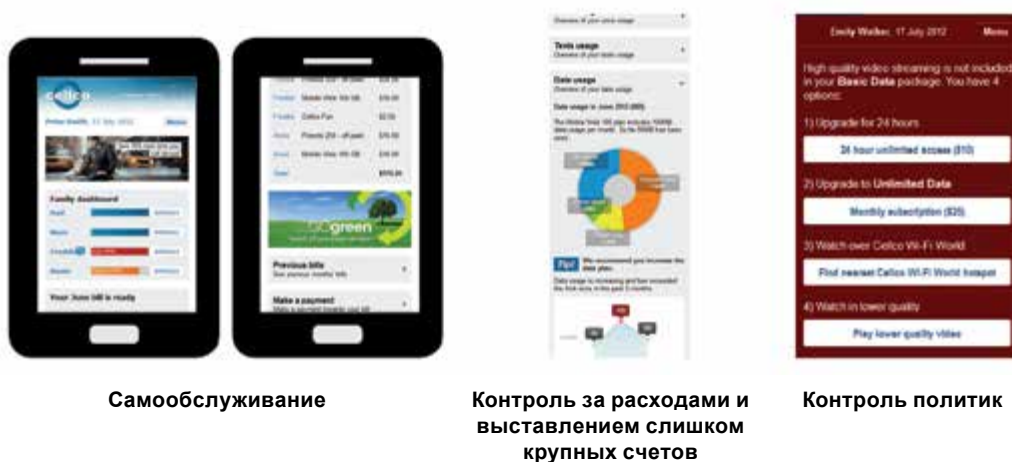
Рис. 4. Сценарии использования аналитики для операторов связи<sup>3</sup>

Продажи и маркетинг	Сеть	Продукты и услуги	Поставщик	Управление клиентами	Управление текущей деятельностью и ресурсами	Предприятие
<ul style="list-style-type: none"> <li>Анализ партнеров</li> <li>Анализ каналов сбыта</li> <li>Анализ маркетинговых и прочих кампаний</li> <li>Изучение потребностей клиентов</li> <li>Анализ продаж</li> <li>Изучение профилей в социальных сетях</li> <li>Сегментация клиентов</li> <li>Анализ абонентской сети</li> <li>Анализ оттока клиентов</li> <li>Перекрестные продажи и допродажи клиентам</li> <li>Клиент приносит выгоду в течение всего срока работы с ним</li> </ul>	<ul style="list-style-type: none"> <li>Управление мощностями</li> <li>Управление производительностью</li> </ul>	<ul style="list-style-type: none"> <li>Анализ производительности</li> <li>Маржинальный анализ</li> <li>Влияние ценообразования и стимулирование продаж</li> <li>Эффект «каннибализации» при внедрении новых сервисов</li> <li>Анализ альтернатив для запуска нового продукта</li> <li>Влияние цепочек поставок</li> </ul>	<ul style="list-style-type: none"> <li>Анализ стоимости и вклада</li> <li>Контроль качества</li> <li>Нормативные требования</li> <li>Анализ выполнения обязательств</li> <li>Анализ качества</li> <li>Аналитика роуминга</li> <li>Расчеты</li> </ul>	<ul style="list-style-type: none"> <li>Отток (удержание) клиентов</li> <li>Удобство для пользователя</li> <li>Соглашения об уровне обслуживания</li> <li>Оценка платежеспособности</li> <li>Анализ удержания клиентов</li> <li>Анализ проблемных ситуаций при работе с клиентами</li> </ul>	<ul style="list-style-type: none"> <li>Оперативный анализ основных процессов: – среднее время от заказа до оплаты</li> <li>– обработка заявок о неисправностях</li> <li>– управление производительностью</li> <li>Аналитика рентабельности каналов связи</li> </ul>	<ul style="list-style-type: none"> <li>Обнаружение и предотвращение мошенничества</li> <li>Гарантия получения прибыли</li> </ul>

## Визуализация аналитики

Благодаря функциям представления и визуализации, прогнозы и результаты передаются в формате, удобном и для персонала, и для автоматизированных систем. Эта задача всегда выполнялась вручную, но теперь она в значительной степени автоматизирована, поскольку компаниям необходимо реагировать на изменения очень быстро, в том числе использовать данные о клиентах для персонализированного взаимодействия с ними через их мобильные устройства (см. рис. 5). Это помогает повысить качество обслуживания клиентов и открывает доступ к новым источникам получения прибыли.

Рис. 5. Визуализация аналитики для персонализированного взаимодействия с клиентом на мобильном устройстве



<sup>3</sup> Источник: «Определение аналитики: оптимизация бизнес-процессов с помощью уже имеющихся у операторов связи данных», Analysis Mason, январь 2012 г. Визуализация аналитики.

## Услуги Software Services

Получите максимальную отдачу от инвестиций в программное обеспечение.

Micro Focus предоставляет услуги по технической поддержке программного обеспечения, затрагивающие все аспекты жизненного цикла ПО. Вы получаете возможность выбрать уровень поддержки, отвечающий конкретным потребностям в области ИТ и бизнеса. Не тратя средства на содержание в штате дорогостоящих специалистов, вы сможете высвободить ИТ-ресурсы, что позволит сосредоточиться на стратегических задачах и инновационном развитии компании.

Услуги Micro Focus обеспечивают:

- «единое окно» для обслуживания всего вашего программного и аппаратного обеспечения в круглосуточном режиме 365 дней в году;
- поддержку VMware, Microsoft®, Red Hat и SUSE Linux, а также HPE Insight Software;
- быстрые ответы на любые вопросы, удаленный доступ к технической документации, оперативное решение проблем и предупреждение их возникновения;
- опыт и знания специалистов Micro Focus, работающих по всему миру.

Дополнительная информация приведена по адресу:

[www.microfocus.com/support-and-services/](http://www.microfocus.com/support-and-services/)

## Vertica: платформа аналитики для операторов связи

С каждым днем контакты между людьми становятся все более тесными. SMS, MMS, видео по запросу, электронная почта, разговоры по мобильным телефонам — вот лишь несколько примеров электронного их взаимодействия. Все эти способы общения генерируют уникальные, подробные записи, и правильный анализ этих данных позволяет не только понять поведение и ожидания абонентов, но и разработать сервисы и приложения для улучшения качества их обслуживания.

Пришло время инноваций для СУБД, используемых в телекоммуникационных компаниях. Число абонентов мобильных и фиксированных сетей, а также пользователей кабельных линий связи увеличивается на несколько миллионов человек ежегодно. Согласно прогнозам, количество CDR, IPDR, информации о профилях абонентов, данных от сетевых зондов и M2M-данных, которые телекоммуникационные компании должны хранить и анализировать, будет расти на 12–13 % в год<sup>4</sup>.

### Архитектура СУБД, разработанная для аналитики в телекоммуникационных компаниях

Аналитическая СУБД Vertica позволяет выполнять высокоскоростной, экономически эффективный анализ больших объемов данных CDR, SNMP, IPDR, информации от сетевых зондов, информации о профилях абонентов и данных M2M. Партиционирование данных по колонкам, широкое использование сжатия и гибридная модель хранения — все эти функции Vertica способствуют уменьшению числа операций ввода-вывода при выполнении запросов. Записи телекоммуникационных компаний, такие как CDR, содержат множество колонок по каждому абоненту, но для отдельных отчетов извлекается лишь несколько. Типы аналитических приложений, использующих небольшое произвольное подмножество колонок при каждом запросе, идеально подходят для вертикального партиционирования, выполняемого аналитической СУБД Vertica. В то же время партиционирование данных в Vertica, распределяющее нагрузку по нескольким узлам в вычислительном кластере, поддерживает очень большие объемы данных, характерные для таких приложений. Эти аналитические приложения, как правило, дают возможность обеспечить поддержку работы сотрудников разных подразделений, которые обращаются к СУБД, чтобы получить нужную именно им информацию. Способность Vertica поддерживать несколько физических проекций данных незаменима в подобных сценариях использования. Таким образом, архитектура Vertica разработана специально с учетом общих характеристик аналитических приложений, применяемых в сфере телекоммуникаций.

## Дополнительная информация

Узнайте, как Vertica помогает улучшить анализ Больших данных, посетив наш сайт: [www.vertica.com](http://www.vertica.com).

<sup>4</sup> Источник: «Тенденции рынка: возможности Больших данных в вертикальных отраслях (Market trends: big data opportunities in vertical industries)», Gartner, июль 2012 г. Услуги Micro Focus