

# Технологии Big Data

## как новый катализатор телеком-индустрии

### Больше, чем данные

Примерно с 2012 года все больше российских компаний, работающих в сфере ИТ, называют сегмент Больших данных (Big Data) в числе наиболее вероятных драйверов роста отрасли в среднесрочной перспективе. По мере того как пилотные проекты на основе этих технологий показывают ожидаемый, а иногда и неожиданный возврат на инвестиции, растет уверенность в том, что одним из основных активов компании в XXI веке станет информация. Как результат, предоставление инструментария для ее сбора, накопления и анализа должно стать первоочередной задачей ИТ-игроков.

Своего рода парадоксом является то, что уверенность в перспективах Больших данных у представителей ИТ сформировалась несколько позже, чем у специалистов других областей. К примеру, известный в сфере менеджмента цикл Шухарта-Деминга Plan-Do-Check-Act (Планирование-Реализация-Проверка-Действие) как минимум в двух своих составляющих базируется на работе с данными. Без качественного исследования предметной области нельзя в должной степени проработать шаг Plan. Что касается стадии Check, то она всецело опирается на анализ данных, полученных в результате Реализации. Таким образом, менеджеры, и в первую очередь, менеджеры качества, легко могут обосновать, почему возможность оперировать большим объемом данных позволяет получить лучший результат.

Цикл Plan-Do-Check-Act присутствует и в области маркетинга. По итогам запланированной акции, например маркетинговой кампании, на некоторой выборке собирается информация о ее эффективности, и по итогам анализа этой информации формируются новые предложения. Возможность в разумные сроки проанализировать результат акции на полном наборе данных (по всей аудитории) позволяет составить

целостную картину ее эффективности без типичного риска ошибки по причине нерепрезентативной выборки. Технология, дающая возможность быстрого получения достоверных ответов относительно эффективности предпринимаемых шагов, может стать решающим преимуществом маркетолога перед его коллегами, такой технологией не располагающими.

В качестве еще одного интересного примера приверженцев Big Data из не-ИТ-сферы можно привести биржевых аналитиков. В своей книге «Большие данные», опубликованной в 2014 году и уже ставшей бестселлером, Виктор Майер-Шенбергер и Кеннет Кукьер затрагивают вопрос целесообразности признания накопленной компанией информации активом в бухгалтерском смысле этого слова, так как балансовая стоимость предприятия, оперирующего данными (например, Google, Amazon, MasterCard), и его рыночная стоимость сегодня могут отличаться в десятки раз именно из-за этого неучтенного актива. Таким образом, на рынке акций ценность больших объемов данных признается буквально и имеет денежный эквивалент, который закладывается участниками сделок в стоимость бумаги.

Сегодня, все ближе к середине 2015 года, тематика Big Data находится в топе рубрикаторов отраслевых мероприятий не только в индустрии ИТ, но и в телекоммуникациях, банковской сфере, ритейле, страховании. Представители различных секторов экономики на практике убеждаются в том, что игнорирование возможностей Больших данных влечет потерю доли рынка и дает возможность конкурирующим компаниям уйти в отрыв.

### Бутылочное горлышко (теперь в хорошем смысле)

Известно, что компания Twitter на коммерческой основе предоставляет (вернее, предоставляла до апреля



**Юрий СЕНЧЕНКО, К.Т.Н.,**  
руководитель направления  
«Управление трафиком  
и широкополосные сети» ООО «НТЦ ПРОТЕЙ»

2015 года) доступ к полной ленте твитов компаниям-дистрибуторам информации, и подобные оптовые продажи данных обеспечили ей около 10% выручки в 2014 году. Продажа информации по модели Twitter — один из интересных бизнес-кейсов для операторов связи. Будучи той самой трубой, и в данном случае не столь важно, smart или dumb, оператор является транспортом для всей Интернет-активности абонента на протяжении периода подписки. То есть на практике оператор сотовой связи располагает несравнимо большим набором данных об абоненте, чем тот же Twitter, так как через его инфраструктуру проходит весь информационный обмен. Оператор владеет полной историей посещений Интернет-ресурсов, историей перемещений абонента, данными о динамике движения средств на счете, SMS перепиской, историей вызовов (т.е. всей адресной книгой), в ограниченной форме — историей платежей через банковские карты, если у абонента подключена услуга 3D-secure, списком установленных приложений, типом устройства, информацией о наличии других устройств, наличием у абонента объектов M2M и многими другими. Неприятный для любого связиста термин bottleneck, т.е. «бутылочное горлышко», неожиданно приобретает позитивный оттенок: оператор связи — это идеальная точка для получения и накопления информации об абоненте с целью построения абонентского профиля, так как это самое узкое место, через которое информационные потоки абонента попадают во внешний мир.

С точки зрения оператора, организация кластера по сбору, хранению и анализу данных в первом приближении представляет собой относительно несложный проект. На каналы передачи данных устанавливаются ответители трафика, которые копируют передаваемые пакеты на модули анализа. Так как ответители включаются пассивно, т. е. не являются терминирующей точкой на каком-либо уровне OSI, а также не вносят задержки при передаче пакетов, подключение кластера к СПД не оказывает воздействия на вероятностно-временные характеристики канала. При этом оптический ответитель не требует подведения питания и не является активным устройством, т. е. не образует точку отказа, и его не нужно сопровождать.

Следующее (и заключительное) действие, требующее вовлечения оператора, — предоставление емкости в стойках дата-центра для размещения модулей анализа и модулей хранения данных. Оба типа модулей исполнены на основе обычных commodity-серверов, т. е. от технического блока оператора не требуется специальной экспертизы по монтажу, подключению и диагностике. Работы по сопровождению кластера, как правило, исчерпываются периодической заменой выходящих из строя жестких дисков. Следует отметить, что подобная ситуация не является аварийной, так как в кластерах большого объема отказ диска считается штатной ситуацией и не приводит к отказу в обслуживании за счет резервирования данных. Замена диска выполняется в течение нескольких секунд на работающем сервере сотрудником с любым уровнем квалификации.

### Ad usum internum

Процесс сбора и хранения информации осложняется технической проблемой диалектического характера: при просчитывании кейсов большой объем информации открывает большие возможности, так как позволяет искать ответы на вопросы с учетом более широкого перечня переменных. В то же время, не все данные окажутся полезными для анализа, и итогом их извлечения и хранения будет удорожание комплекса, а также увеличение времени получения конечного результата. Таким

образом, при планировании состава анализируемых данных оператор взаимодействует с производителем кластера для «калибровки» своих интересов и поиска баланса между стоимостными параметрами комплекса, глубиной поиска, набором параметров и скоростью получения информации.

На собранном наборе данных сотрудники оператора или сопровождающие проект специалисты производителя формируют аналитические кейсы и получают ответы (выборки, корреляции) на интересующие вопросы, исходящие, как правило, из отдела маркетинга или эксплуатации. В качестве простейших примеров выборки можно указать TOP-абонентов по какому-либо признаку, TOP-локации по определенному виду трафика, корреляцию между параметром профиля абонента и структурой потребления услуги, выборку абонентов в определенной локации по какому-либо признаку в режиме реального времени и др.

После того, как параметры кластера согласованы и определен перечень собираемых данных, дальнейшие возможности применения комплекса ограничиваются только потребностями оператора и законодательством в области защиты персональных данных. Так как передача данных абонентов третьей стороне является юридически сложным предприятием, подавляющая часть кейсов операторов связи находится в категории «для внутреннего пользования». В числе наиболее популярных сфер применения технологий больших данных — мониторинг состояния сети, удержание абонентов, повышение продаж и кросс-продажи, сегментирование абонентской базы для точечных рекламных кампаний.

Говоря о конкретных примерах кейсов Big Data, можно упомянуть повышение ARPU за счет продажи пакетов трафика абонентам, отправляющимся за рубеж. В связи с непредсказуемостью итогового счета, а также высокой стоимостью услуги передачи данных в роуминге многие абоненты на все время путешествия отключают услугу GPRS, а для доступа в Интернет покупают SIM-карту местного оператора. Для увеличения продаж выезжающим абонентам отправляется SMS с информацией о пакетном предложении на



выгодных условиях, например X Гбайт трафика в сутки на скорости Y Мбит/с за Z рублей. Система аналитики Big Data позволяет отправить SMS только тем абонентам, которые действительно покидают страну, и исключить из списка получателей, которые только что прилетели, а также сотрудников аэропортов и авиакомпаний.

Открывающиеся с появлением технологий Больших данных перспективы сложно переоценить. Технология Big Data дает подразделениям информацию, которая ранее не была доступна из-за отсутствия технической возможности обработки обширных информационных массивов. В эпоху Big Data специалисты эксплуатации имеют возможность получить точный ответ на вопрос «что происходит с сетью», так как могут проверять свои предположения и искать причины, анализируя весь доступный объем информации (статистику с устройств, индикаторы аварий, динамику параметров во времени, CDR и т. п.). Специалисты маркетинга могут точно выделить сегмент пользователей, удовлетворяющий их запросу на основе полной картины поведения абонентов, а не только остатка средств на счете и города регистрации SIM-карты. Отдел предотвращения фрода получает возможность выявления всех аномалий без исключения, так как им доступен полный набор данных о поведении всех абонентов на интересующем временном промежутке. Положительный опыт применения технологий Big Data из эпизодического становится повсеместным. Таким образом, сегодня можно уверенно говорить о том, что ящик Пандоры закрыть не удастся, и альтернативы интеграции технологий Больших данных в бизнес-процессы у операторов связи нет. ■