

Неоклассика в сфере BI-технологий:

ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ



Людмила ОСТРОВСКАЯ,
руководитель проектов Отделения
комплексного проектирования
информационных систем, компания
«ФОРС – Центр разработки» (ГК ФОРС)

Аналитика для всех

Рынок продуктов для анализа данных не перестает активно развиваться уже долгое время. Появляются не только новые платформы для создания аналитических систем, но и новые подходы к их реализации, организации работы с данными. Появились и стали стандартом новые технологии, позволяющие обрабатывать все большие объемы информации.

Сегодня каждая компания может найти на рынке удобное аналитическое решение, соответствующее именно ее потребностям. Навсегда ушли в прошлое времена, когда использование продуктов класса Business Intelligence (BI) могли позволить себе только крупные компании, обладающие значительным бюджетом на ИТ.

Аналитические технологии постоянно развиваются, а в сочетании с машинным обучением и облачными технологиями они способны решать совершенно иной класс задач. В статье, предлагаемой вниманию читателей журнала Connect, мы расскажем о том, что собой представляет аналитический инструментарий сегодня, почему наша «добрая старая классика» будет востребована всегда и как эти технологии будут развиваться в будущем.

Основной движущей силой, сделавшей BI-системы (впрочем, и не только их) доступными для среднего и малого бизнеса, стали облачные технологии. Это благодаря им появились сервисы, позволяющие приобретать аналитическое решение как услугу. На сегодняшний день такой способ стал настолько распространенным, что существуют и универсальные платформы, и готовые решения, предназначенные для формирования определенных видов стандартной отчетности компаний. При этом массовое использование обеспечило вполне приемлемую стоимость подобного ИТ-решения для небольших организаций и для частных предпринимателей.

В нашей стране заметной тенденцией в мире Business Intelligence стали укрепление позиций российских разработчиков, а также появление все большего числа новых отечественных продуктов. Среди них можно назвать Visiology, «Форсайт» и ряд других. Отчасти это произошло благодаря санкционной войне, в результате которой не только государственные органы, но и многие коммерческие компании вынуждены отказываться от закупки лицензий на продукты западных вендоров.

Такие заказчики оказались перед выбором: переходить на свободное программное обеспечение или использовать российские разработки; и во многих случаях выбор делается в пользу российских продуктов.

Основные причины этого существенны для всякого заказчика, планирующего использовать созданную систему долго: присутствие постоянной команды разработки, занимающейся развитием продукта, наличие технической поддержки и программ обучения. Кроме того, российские разработчики обеспечивают сертификацию своих платформ в точном соответствии с требованиями российских регуляторов в части информационной безопасности и защиты от несанкционированного доступа.

Новые возможности классического BI

Но каковы бы ни были масштабы бизнеса потребителя, какого бы он ни выбрал для себя поставщика, современную аналитическую платформу довольно трудно представить без инструментов, расширяющих возможности классического Business Intelligence.

С развитием Интернета вещей и появлением разнообразных

датчиков весьма ценным дополнением к обычным, прежде всего статическим визуализациям стала потоковая аналитика. В основе потоковой аналитики – подключение к источнику непрерывно поступающих данных и использование алгоритмов обработки непосредственно этого потока. В результате пользователь может в онлайн-режиме наблюдать за процессами, а если это предусмотрено применяемыми алгоритмами, то и получать различные сигналы и предупреждения.

Не только новые источники данных и визуальные элементы, но и дополнительные возможности, которые пользователь стремится получить, используя свои данные, присутствуют в современных аналитических системах. Пользователь хочет не только видеть информативную картину настоящего, но и прогнозировать, что будет дальше, оценивать различные варианты развития событий.

Традиционно промышленные платформы Business Intelligence содержат инструментарий и для выполнения анализа «что если», и алгоритмы построения трендов. Это необходимый минимум, но современные возможности технологий машинного обучения существенно шире. Во многих отраслях, таких, например, как финансы, ритейл, телекоммуникации, машиностроение, использование машинного обучения (ML) стало уже неотъемлемой частью бизнеса.

Многими организациями накоплен значительный объем корпоративных и внешних данных. Предиктивная аналитика, построенная на обработке этих массивов с помощью машинного обучения, позволяет: руководителям – принимать более взвешенные решения, специалистам по управлению финансами – минимизировать риски, инженерам – искать потенциальные дефекты оборудования.

Для получения нужного результата используются статистические вычисления, сложные алгоритмы углубленного анализа данных. Современный data scientist отдает

Системы бизнес-анализа

- Бизнес-анализ данных
 - визуализация, интерактивный анализ, информационные панели
- Углубленная аналитика
 - Прогнозирование, кластеризация, статистическая обработка
- Анализ показателей эффективности предприятия, аналитика результатов маркетинговой кампании, мониторинг бизнес-процессов

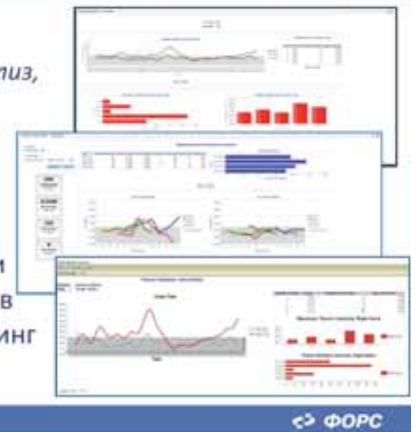


Рис. 1. Системы бизнес-анализа

предпочтение специализированным инструментам и языкам программирования, в которых доступны необходимые библиотеки обработки данных. Ведь ему требуется не только применить алгоритм к данным, но и провести значительную подготовительную работу.

Даже в самом простом случае специалисту класса data scientist необходимо преобразовать дан-

а может и вовсе прийти к выводу, что для этой задачи ему следует применить другой алгоритм или последовательность из нескольких алгоритмов.

Это серьезная творческая работа, для которой недостаточно лишь знаний методов и инструментов. Такой специалист должен обладать аналитическим складом ума и развитым критическим мышлением, ему важно уметь

Промышленные платформы Business Intelligence содержат инструментарий и для выполнения анализа «что если», и алгоритмы построения трендов.

ные для обработки тем или иным методом машинного обучения, подобрать параметры для запуска алгоритма и оценить качество результата после завершения работы. Часто это итеративный процесс, и по результатам предыдущей оценки аналитик старается усовершенствовать каждый его элемент. Например, обогащает сформированный набор данных новыми признаками, которые могут оказать влияние на результат, выполняет дополнительную нормализацию, изменяет параметры работы алгоритма,

глубоко погружаться в прикладную область, а иногда и в бизнес определенной организации. Именно потому компании ценят и берегут таких специалистов, и они сегодня крайне востребованы на рынке.

А инструменты – уже дело вкуса и привычки. Многие используют удобные визуальные средства, что значительно ускоряет процесс, особенно если инструмент уже содержит шаблоны сценариев для той или иной прикладной области. Однако сегодня наиболее широкое распространение получило

Анализ чрезвычайных ситуаций (для МЧС)



Рис. 2. Анализ чрезвычайных ситуаций (для МЧС)

использование языка Python – кстати, не в последнюю очередь благодаря активному, опережающему многие другие инструменты развитию библиотек обработки данных и машинного обучения для этого языка.

Машинное обучение (ML) и BI

Интеграция инструментов машинного обучения с системами Business Intelligence всегда дает положительный результат, многократно повышающий эффективность анализа данных пользователем. Обращение к результатам работы алгоритма ML может быть реализовано с помощью сервиса, но чаще для этого используется база данных, в которой сохраняется результат работы алгоритма для последующего его анализа с помощью BI-инструмента.

Некоторые СУБД и сами содержат встроенный инструментарий статистической обработки данных и прогнозирования. Так, Oracle Database давно включает полезную опцию Data Mining, в дополнение к которой предоставляется визуальный инструмент конструирования алгоритмов и подготовки данных для них. Позже этот инструментарий был дополнен R – языком статистического моделирования с открытым исходным кодом. Сейчас встроенные возможности машинного обучения появились и в других современных СУБД.

Почти любой человек сегодня практически не расстается с мобильным телефоном – это устройство зачастую становится рабочим инструментом в поездках и других ситуациях, когда пользователь не может оставаться в офисе. Поэтому для решений класса Business Intelligence полезным спутником будет не только инструмент ML, но и приложение мобильной аналитики, а интеграция с мессенджерами сделает информирование более эффективным, чем с помощью электронной почты.

Благодаря развитию технологий искусственного интеллекта (AI) важным дополнением к традиционным методам управления аналитическим процессом стал голосовой помощник – умный робот, который может распознавать человеческую речь и формировать аналитические представления по голосовым командам пользователя.

BI и облачные технологии

Создание комплексов решений, в которых ИТ-инструменты взаимно дополняют и обогащают возможности друг друга, – одна из наиболее актуальных современных тенденций. В особенности это относится к упомянутым облачным технологиям.

С одной стороны, пользователь должен доставлять свои данные в облако, потому появляются

дополнительные сервисы в виде хранилища данных и ETL-инструмента. С другой стороны, компания, предоставляющая облачный ресурс, может интегрировать в него и другие полезные сервисы. И пользователь уже не стоит перед необходимостью самостоятельной настройки и внедрения дополнительных инструментов, а в зависимости от своих потребностей и компетенции решает, покупать ли тот или иной дополнительный сервис.

Так, в 2019 г. «Яндекс» запустил в своем облаке сервис DataLens – облачный сервис аналитики данных. Облачные гиганты, такие как Yandex.Cloud, предоставляют не просто Cloud-сервисы, а полноценную экосистему ведения бизнеса, позволяющую создавать и постепенно наращивать его возможности. Впрочем, это уже не только об аналитике.

Все же большинство компаний, особенно наш крупный бизнес, предпочитают создавать собственные аналитические системы с использованием собственных ресурсов на основе готовых решений от зарубежных или российских вендоров.

BI и люди

Внедрение большинства продуктов класса Business Intelligence требует сегодня участия квалифицированных специалистов, которые не только создают отчеты и информационные панели, но и настраивают для пользователей системы многомерную модель данных, которая служит основой созданной отчетности. Как правило, это реализуется с помощью слоя метаданных, которые описывают их трансформацию от физических структур к прикладным терминам, понятным пользователям системы.

Пользователи корпоративных BI-систем традиционно делятся на пользователей отчетов и продвинутых аналитиков. Последние не только используют в работе серьезные отчеты, настроенные специалистами, но и активно

выполняют свои запросы к данным, используя многомерную модель данных. И вот именно эти пользователи рано или поздно начинают сталкиваться с нехваткой того или иного элемента в настроенной специалистами модели. Однако для ее изменения нужно снова обращаться к специалисту, ожидать реализации и тестирования реализованных дополнений. В зависимости от процессов организации может потребоваться более или менее длительный период согласования. Пока все необходимые мероприятия будут проведены должным образом, сама потребность выполнить новый запрос уже может утратить свою актуальность.

Традиционный подход имеет свои плюсы. Это прежде всего единая и непротиворечивая модель расчета показателей и централизованное управление доступом к данным. Но и минусы у такого подхода очевидно присутствуют.

В результате, если у компании не было BI-системы или возникла необходимость в более простом инструментарии, в ход довольно часто шел MS Excel. Чудеса аналитики, которые хороший экономист может продемонстрировать с помощью этого инструмента, внушают уважение даже специалистам в области информационных технологий.

Но Excel, как и любое настольное решение, имеет очевидные недостатки. Это не только ограниченная производительность, но и возможные проблемы с защитой данных от несанкционированного доступа при копировании и пересылке файлов.

Между тем продвинутый аналитик часто обладает необходимыми знаниями, чтобы самостоятельно выполнить запрос и сконструировать относительно несложную модель данных. Поскольку возможности для этого в традиционном BI-инструменте были ограничены, назрела необходимость в инструментарии, позволяющем работать без жестко настроенной модели, предоставляющей более демократичный доступ к данным.

Аналитика для поддержки лечебного процесса



- Повышение точности диагностики
- Прогнозирование развития болезни
- Поддержка доказательной медицины
- Персонализация курса лечения
- Непрерывный мониторинг состояния пациента
- Предиктивная аналитика
- Обработка изображений
- Корреляционный анализ
- Анализ временных рядов
- Технологии больших данных

ФОРС

Рис. 3. Аналитика для поддержки лечебного процесса

Шаги в будущее

Поэтому уже больше десяти лет назад появились и к настоящему времени отвоевали свое место на рынке инструменты класса Data Discovery. Этот класс аналитических решений позволяет бизнес-пользователям подключать новые источники данных, а также создавать относительно несложные модели и строить отчетность на их основе уже без участия IT-специалистов. Конечно, работать с таким инструментом несколько сложнее, чем с готовой моделью, но в данном случае благодарный пользователь быстро находится, и сейчас инструменты Data Discovery используются наравне с традиционным Business Intelligence в самых разнообразных областях деятельности. В качестве примера можно привести сферу государственной безопасности и здравоохранения.

Меняются не только инструменты анализа данных – меняются их источники. Сегодня уже редко найдешь организацию, у которой не было бы хотя бы небольшого аналитического хранилища. Для крупных же организаций и холдингов, особенно из финансовой сферы, телекома и торговли, конфигурация источников данных для анализа давно вышла за рамки концепции единого централизованного хранилища.

Как правило, такие организации уже выстроили систему взаимосвязанных хранилищ данных,

использующих различные технологии. В них хранится структурированная и неструктурированная информация, используются технологии Big Data, компании формируют на их основе множество аналитических ресурсов. И это уже не только традиционные витрины, но и целые озера данных, и различные песочницы для пользовательских экспериментов. Для работы с последними, кстати, хорошо подходят инструменты Data Discovery, особенно такие, которые могут обрабатывать неструктурированную информацию и работать с большими данными.

В целом же здесь находится применение для самых разных аналитических инструментов. Востребованы и традиционный Business Intelligence, и старая добрая отчетность, которая куда не делась и не денется. И если меня в очередной раз спросят, не уйдет ли в прошлое старая технология с появлением новых, я отвечу, что рынок определяют не мода и новизна, а реальные запросы потребителей.

С тех пор как мир осознал, что данные являются ценным ресурсом, который имеет стоимость и может приносить прибыль, качество аналитических инструментов постоянно повышается и они становятся все более разнообразными. А совместное использование этих инструментов предоставляет более широкие возможности, чем выбор чего-то одного. ■