

МЕТОДЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ ДЛЯ ОПТИМИЗАЦИИ ДЕЯТЕЛЬНОСТИ НЕФТЯНЫХ КОМПАНИЙ

METHODS OF BIG DATA ANALYSIS TO OPTIMIZE THE ACTIVITIES OF OIL COMPANIES

Jaf Mohammed Fadhil

Summary. The article describes the application of big data in the oil and gas sector. Their application abroad, and in the domestic market. The article discusses the obstacles that prevent the smooth development of large oil companies. This article also discusses some of the decisions that will help you to find successful techniques and technologies to solve problems.

Keywords: big data, oil and gas sector of the company, holding, the prospects for the ecosystem.

Джаф Мохаммед Фадиль

Аспирант, Санкт-Петербургский политехнический университет Петра Великого (СПбПУ)
hamoorjaf@yahoo.com

Аннотация. В статье рассказывается применение больших данных в нефтегазовом секторе. Применение их за рубежом, и на отечественном рынке. В статье рассматриваются препятствия, которые мешают плавному развитию крупных нефтяных компаний. В данной статье также рассматриваются решения, которые помогут найти успешные методики и технологии решения проблем.

Ключевые слова: большие данные, нефтегазовый сектор, компании, холдинг, перспективы, экосистема.

Конечно, нефтегазовый сектор может и, наверное, должен проявить интерес к выстраиванию практических решений в этой сфере. Сегодня у нас есть (в том числе и в отечественных компаниях) солидные архивы необработанных исторических данных, которые связаны с реализацией производственных процессов и которые при современном уровне технологии их прикладной обработки могут быть превращены в знания.

Кроме того, зависимость перспектив крупных нефтегазовых компаний от конъюнктуры рынка (как мирового, так и национального) более чем очевидна, сама отрасль по понятным причинам всегда была и остается, как говорится, на виду в прессе, и во многих коммерческих отраслях, и в коридорах власти. В таких условиях анализ огромного количества неструктурированной информации, в основном из внешних источников, становится весьма актуальной задачей. А это, в свою очередь, тоже одна из типичных задач технологии Больших Данных.

При всем этом отрасль, о которой мы сейчас говорим, не является драйвером практической реализации концепции BigData и в целом даже не входит в число явных лидеров по количеству проектов соответствующего направления. Хотя некоторые успешные и достаточно продвинутые случаи в мире, безусловно, можно найти. Отечественные компании, как известно, тоже не в первых рядах. Почему это происходит? Кажется, что нет ни одного препятствия, которое можно было бы четко обозначить, но есть целый комплекс причин.

Препятствия на долгом пути

Большинство нефтяных компаний в России—это вертикальные холдинги: горнодобывающие компании и дочерние компании, занимающиеся логистикой и перевозками. Сеть АЗС—это опять же отдельный бизнес. Есть также научно-технические отделы и, конечно же, корпоративный центр. Существует много данных в отрасли, но они управляются различными юридическими лицами. На Конференциях постоянно говорят о сложности вывода данных из так называемых «силосов» предметных областей (финансовых, маркетинговых, производственных и др.) в едином корпоративном поле, в котором должен проводиться фактический анализ.

Но в отечественных холдингах мы сталкиваемся с организационными силосами, в основном из-за вышеупомянутой структурной изоляции, и это гораздо серьезнее. Хотя бы потому, что нет готовых, многократно проверенных и гарантированных успешных методов и технологий решения проблем. В нефтегазовой отрасли никогда не было четкого понимания на уровне высшего руководства, что количество и качество информационных ресурсов (не путать с задачей внедрения ключевых ИТ-систем) влияет или, по крайней мере, в какой-то момент может повлиять на эффективность управления бизнесом. В том же банковском бизнесе это понимание, пусть и где-то неявное, сформировалось задолго до появления BigData. Здесь постоянное совершенствование работы с информацией в целом по компании и улучше-

ние ее управляемости воспринимались во многом как естественно связанные вещи.

В нашей отрасли генеральный директор компании вряд ли захочет делиться данными, поскольку в результате может быть (и скорее всего будет) выявлена недостаточная эффективность управления организацией. И мы собираем данные именно для оптимизации, поэтому будут найдены резервы для повышения эффективности. И в то же время, очень вероятно, что генеральный директор не знал о них. По-хорошему, первое лицо компании должно вести такую деятельность самостоятельно, лично управляя формированием информационного ресурса бизнеса с нетехнологических позиций и определяя возможную переоценку значимости его отдельных составляющих. Но много ли таких руководителей в горнодобывающих компаниях, когда все специалисты сосредоточены в научно-технических центрах? И в какой степени руководители горнодобывающих компаний готовы передать свою эффективность на аутсорсинг научно-техническому центру?

Далее, следует сказать, что нефтяные холдинги явно не являются теми структурами, где использование интернет-пространства (даже с учетом ряда корпоративных ограничений) тесно переплетается с культурой корпоративного управления. Люди на рабочих местах не находятся в сети. Облако, социальные сети и профессиональные интернет-сообщества недоступны пользователям в корпоративной сети. Информационная система тоже не Онлайн: нет Яндекс метрики, Яндекс и Google maps, соответственно, для разработчиков нет доступа к ресурсам GitHub. Кроме того, некоторые данные (например, о недрах) должны храниться определенным образом, и даже внутри компании не все имеют к ним доступ. Многие геологи скептически относятся к ИТ-системам и большим данным, потому что они помнят свои неудачные попытки загрузить файлы размером более двух гигабайт в 32-разрядные операционные системы.

Экосистема бизнеса в нефтяной отрасли состоит из множества нефтесервисных компаний, выполняющих значительную часть работ, и инновационность отрасли в целом резко зависит от технологий (в том числе информационных), используемых такими контрагентами. Предполагается, что законы рынка должны заставить их искать более эффективные решения для увеличения своей прибыли. Было бы логично ожидать, что подрядчики привнесут в отрасль новые технологии управления данными. На самом деле отношение крупных нефтесервисных компаний больше направлено на то, чтобы все подсади на иглу их услуг по высоким ценам: например, стоимость лицензии на одно рабочее место средней конфигурации может составить 20 млн. рублей! Отметим, что функциональность такого рабочего места не может

быть использована на 100%. Поэтому представляется целесообразным перевести некоторые функции такого рабочего места на язык Python и сократить количество используемых лицензий. С другой стороны, небольшие нефтесервисные компании довольны своим монопольным положением. По цене они выигрывают у западных предприятий в разы, оставаясь при этом в качестве, пусть и не на высоком уровне, но в целом на уровне, достаточном для функционирования отрасли в нынешних условиях. Перефразируя вождя, можно сказать, что «и низы могут, и верхи хотят», но в то же время революционных перемен ждать неоткуда.

Однако можно также сказать, что в нефтегазовом секторе существуют определенные предпосылки для внедрения технологий Big Data.

В первую очередь, конечно, речь идет о возможности финансирования. BigData зачастую является очень дорогим удовольствием, а возможности нефтегазового сектора, вопреки распространенному мнению, не безграничны. Но тем не менее они есть, и это ограничение явно не является принципиальным.

В последнее время нефтегазовая отрасль получила значительный импульс в развитии коммуникаций. Между различными дочерними компаниями появились каналы связи с достаточной пропускной способностью. Застройка в основном ведется в труднодоступных местах, но если такие проекты признаются необходимыми, то есть электричество, трубы, дороги. При таком подходе уложить волокно не так уж и сложно, и это делается. Не всегда верно, что существует понимание того, что широкополосная связь имеет решающее значение для бизнеса. Но тут, как говорится, жизнь сделает, и никакого другого варианта уже может и не быть.

Следующим шагом после построения единой корпоративной сети традиционно является попытка получить выгоду от учета нефинансовых ресурсов в системах ERP. А если говорить о нем, то его уже можно рассматривать как первую половину шага к построению бизнеса на основе данных (data driven business). Конечно, бизнес начнет задумываться о BigData на очень зрелых этапах этого пути, и важно, чтобы CIO понимал, что для технологических данных необходимы отдельные репозитории (о чем Gartner, кстати, уже давно говорит). Двести миллионов записей из журналов мониторинга состояния насоса за год не имеет смысла загружать в SAPBW, и для анализа эти данные важны.

Подводя итоги, можно сказать, что в целом ситуация в нашей отрасли не уникальна. Как и везде, существуют организационные, культурные и технологические (в основном связанные с традициями управленческого

взгляда на ИТ, а не на саму ИТ) барьеры для внедрения новых подходов к автоматизации. Есть, конечно, факторы, которые способствуют их развитию в компаниях. Вряд ли возможно точно локализовать точку, в которой бизнес находится на этом пути, и для того, чтобы объективно оценить это, мне кажется полезным заполнить таблицу, которую я привожу ниже. Он не ориентирован конкретно на нефтегазовый сектор и, думаю, может представлять интерес для специалистов, работающих в различных областях.

Промышленность
«точка кристаллизации»

Известно, что концепция Bigdata очень «демократична» по отношению к различным функциональным областям корпоративного управления: ее можно успешно применять буквально везде — на предпроизводственном этапе, в самом производстве, в логистике, продажах и маркетинге. В то же время BigData, как известно, одинаково применима во всей вертикали управления — от анализа параметров цехового оборудования в режиме реального времени до создания информационной основы для принятия стратегических решений по развитию бизнеса. Еще раз повторю, все это потенциально может быть реализовано с помощью больших данных в крупных нефтегазовых компаниях, и важно расставить приоритеты, а точнее даже объективно спрогнозировать «точки кристаллизации», с которых начнется процесс.

Мне кажется, что инженеры-разработчики полей более готовы использовать BigData. Особенно если они уже имели опыт работы с суррогатными моделями и понимали преимущество аналитических функций, использующих подходы машинного обучения в дополнение к физическим моделям. Бурение и добыча по отношению к данным аналогичны добыче: все понимают важность прогнозирования надежности, моделирования наземных и подземных работ на основе накопленных данных.

В разведочно-эксплуатационном контуре (апстрим) также важны методы распознавания образов на службе геомеханики, когда по картине шлейфа можно сделать выводы о физическом состоянии образца породы. Уже несколько лет все признают, что переход на цифровые методы анализа ядра приведет к необходимости анализа огромных объемов данных и позволит получить показатель проницаемости каждого конкретного куска горной породы, даже если он имеет неправильную форму.

В цепочке обработки, логистики и продаж (Downstream), на мой взгляд, следует ожидать прорыва от использования умной « клиентской аналитики. Образу говоря, несложно представить, что вас встретят по названию АЗС на входе, если вы уже заправились

на ней, и предложат специальные предложения по СМС. Я думаю, что когда это станет очень важным для нас на практике, мы сможем многое перенять из опыта компаний в так называемых клиенто ориентированных отраслях.

Интересна одна из интерпретаций будущего развития событий, согласно которой массового прихода BigData в нефтегазовую отрасль может и не произойти. Например, получится примерно то же, что и с цифровой телефонией (тональный набор) на стационарных линиях в России: с аналоговой телефонии сразу перескочили на мобильную. В российской нефтяной отрасли не будет проектов BigData, а будут ИТ-инструменты, использующие подходы BigData, о которых не будут знать геологи, разработчики месторождений и ИТ-специалисты. На вероятность такого сценария указывает, например, тот факт, что в 2015 году BigData была выведена из печально известного Гипецикла аналитической компании Gartner. Место этого термина заняли когнитивные технологии. И один из них особенно важен для нефтяной промышленности — это машинные интерфейсы в целом и автоматизация процесса бурения в частности. Задача такой автоматизации требует глубокого анализа данных в режиме, близком к реальному времени. Данных здесь очень много, и на их основе необходимо строить прогнозную аналитику. Актуальность анализа возможности использования стека hadoop при решении данной задачи неоспорима.

Одной из важнейших бизнес-задач нефтегазовой или горнодобывающей компании является пополнение минерально-сырьевой базы (МСБ) [2]. Проще говоря, обеспечение наличия достаточных резервов для поддержания уровня производства компании. Геологоразведочные работы проводятся для пополнения МСП.

Важнейшие бизнес-процессы в нефтегазовой геологоразведочной компании [3], составляющие жизненный цикл геологоразведочных работ (рис. 1.), он:

- ◆ оценка перспективных территорий;
- ◆ разработка программы геологоразведочных работ;
- ◆ управление разведкой;
- ◆ анализ результатов геологоразведочных работ.

Задача выявления перспективных направлений и сфер интересов компании тесно связана с обработкой огромных объемов разнородных данных. Специалистам компании предстоит изучить весь огромный массив информации, зачастую противоречивой и неполной, и решить, стоит ли инвестировать в ту или иную территорию, стоит ли покупать тот или иной лицензионный участок.

Для оценки всего набора данных необходимо создать хранилище данных.



Рис. 1. Жизненный цикл бизнес-процесса воспроизводства минерально-сырьевой базы (ВМСБ)

В предыдущие годы это решалось путем создания хранилища в реляционной базе данных для цифровых структурированных данных, неструктурированные данные размещались в файловых каталогах. СУБД имеет один существенный недостаток-жесткую привязку структуры хранения к структуре исходных данных. Из-за высокой неопределенности поиска данных использование SQL-запросов к жестко структурированной базе данных неэффективно.

Этим недостатком лишено хранилище на базе решения NoSQL в сочетании с поисковой системой. Использование баз данных «ключ-значение» обеспечивает единое индексирование всего массива данных, независимо от их структуры. Использование поисковой системы, такой как Apache Lucene (<http://lucene.apache.org/>), как связующий элемент между объектом интереса и массивом информации предоставляет широкие возможности для поиска необходимой информации в условиях большой неопределенности.

В задаче выявления перспективных направлений технология текстового майнинга предоставляет возможность поиска литературных источников по заданной территории и заданной теме, поиска схожих проектов и геологических результатов, схожих идей для разведки и т.д. примером таких инструментов является RapidMiner (<http://rapidminer.com/>) или HP IDOL. Благодаря этой технологии для анализа становятся доступны значительно большие объемы источников данных, а глубина



Рис. 2. Взаимодействие элементов Big Data

литературы, отчетности и документации увеличивается за меньшее время.

На этапе формирования геологоразведочной программы важным элементом является оптимизация этой программы по различным критериям с учетом:

- ◆ стратегические планы компании;
- ◆ имеющиеся производственные ресурсы компании;
- ◆ доступны на рынке сервисные компании, бесплатные ресурсы компаний;
- ◆ сезонные и климатические особенности исследуемых регионов;
- ◆ финансово-экономические условия в компании;
- ◆ конъюнктура;
- ◆ геологические условия изучаемых регионов, запасы;
- ◆ ограничительные условия лицензионных соглашений и др.

При наличии таких разнообразных критериев лучше всего подходят эволюционные методы вычислений [3], особенно генетические алгоритмы. Многокритериальная оптимизация с использованием эволюционных методов (генетических алгоритмов) хорошо решается с использованием одной из технологий Big Data — Data Mining tools.

Отдельной задачей является оценка рисков проектов, включенных в программу геологоразведки [2].



Рис. 3. Пример последовательности внедрения технологии Big Data для решения задач ГРП

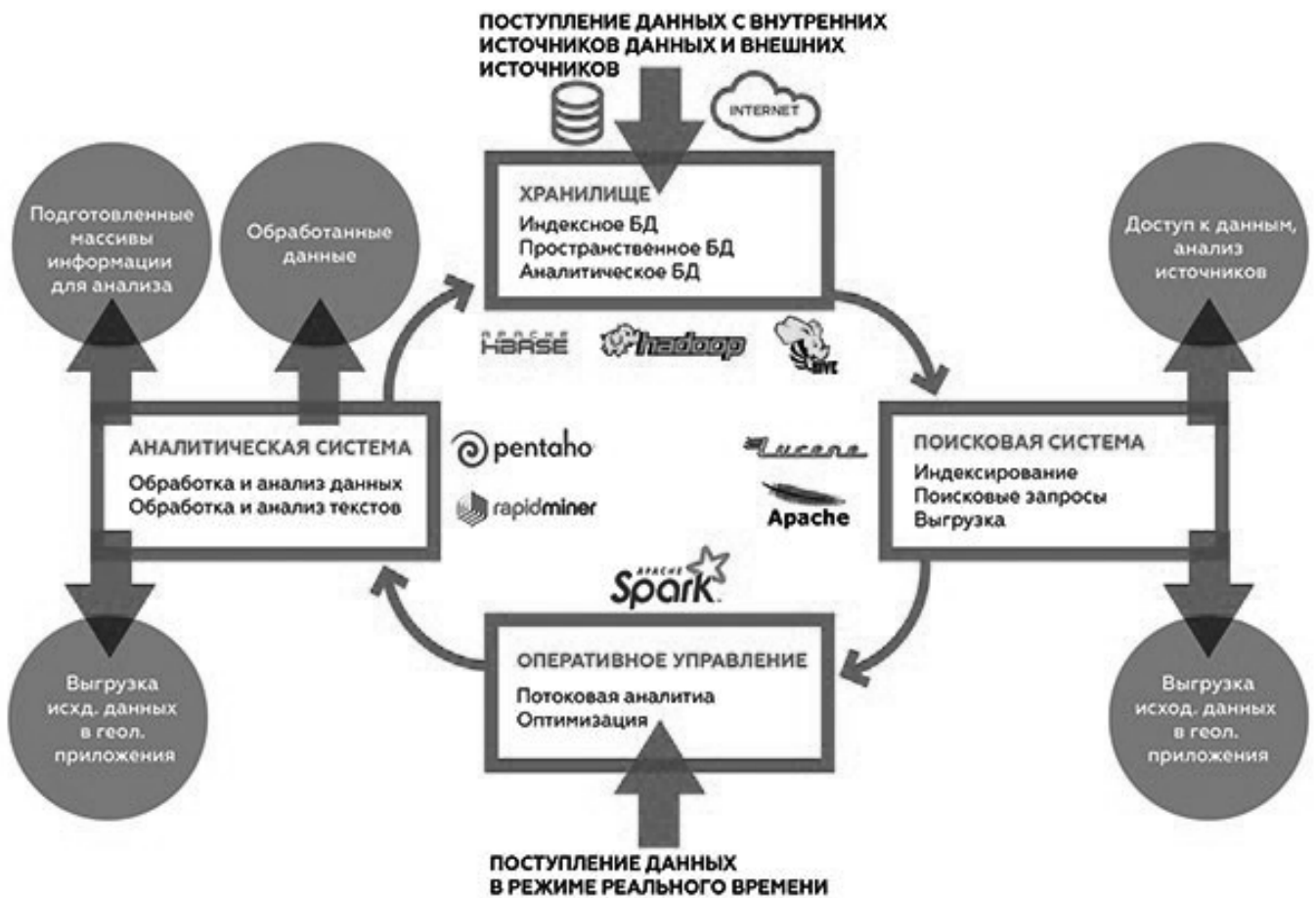


Рис. 4. Пример архитектуры системы и взаимодействие с внешними системами

И в этой задаче существенную помощь могут оказать технологии «Больших Данных». Ведь за счет гораздо более быстрой обработки, в то же время, средства работы со скорингом (риском) [2] и прогностическими моделями [2] позволяют рассчитать гораздо большее количество вариантов. Такие системы, как RapidMiner или Pentaho Business Analytics (<http://www.pentaho.com>), в сочетании с Apache Hadoop может значительно сократить время оценки рисков и помочь более точно сформировать программу разведки.

На этапе управления геологоразведкой необходимо обеспечить своевременное получение данных, оперативный мониторинг операций, оперативный контроль операций в процессе бурения. База данных NoSQL [3] в сочетании с поисковой системой отлично справляется с обеспечением своевременного получения данных. Через систему сканеров («пауков») обеспечивается постоянный обзор источников данных: датчиков и САУ, баз данных, пультов оператора, серверов управления. Данные хранятся, индексируются, классифицируются и сра-

зу же доступны для анализа, что дает значительные преимущества по сравнению с традиционными системами хранения на основе СУБД, требующими полной системы загрузки данных.

Благодаря использованию Hadoop, можно обеспечить функционирование постоянных моделей в производстве, а именно:

- ◆ текущие скоринговые и прогнозные модели для управления бурением, разведкой на шельфе и другими рисковыми операциями;
- ◆ постоянные прокси-модели геологической среды.

При оценке качества и полноты полученных результатов большое значение имеет глубина оценки данных, которая тесно связана с сопоставлением данных между собой. Благодаря высокой скорости выполнения поисковых запросов можно проводить более сложные сравнения массивов данных за меньшее время.

Активно развивается технология обработки сейсмических данных с использованием Hadoop. Такие компании, как Chevron, Shell, Yandex Terra, US Seismic успешно использовали распределенную вычислительную систему MapReduce и ее программные средства Hadoop для обработки полевых сейсмических данных.

При изучении нефтегазовых прогнозов и аналогичных задач можно использовать инструменты текстового майнинга и Data Mining, позволяющие находить аналогии по заданным закономерностям. Другими словами, можно искать подобные геологические структуры по аналогии.

Внедрение технологий Big Data в геологоразведочную деятельность компании не является одновременным процессом. Успешная реализация требует последовательности, в которой каждый последующий шаг основывается на результате предыдущего. Для риса. 3 показан пример последовательной реализации. Первым шагом внедрения является создание хранилища данных, где важно интегрировать данные из существующих систем, наладить непрерывное пополнение данных и контроль качества. Второй шаг – создание распределенной вычислительной и запоминающей системы. Третий этап – формирование аналитической части системы.

Технологии «больших данных», безусловно, не способны заменить существующие традиционные системы обработки, анализа и моделирования. Для риса. 4 показана схема взаимодействия с существующим ИТ-ландшафтом. В то же время использование этих технологий в сочетании с традиционными технологиями позволяет более эффективно работать с полученными данными, экономить время и деньги компании.

ЛИТЕРАТУРА

1. <https://www.iemag.ru/analytics/detail.php?ID=35805>
2. <https://www.eprussia.ru/epr/297-298/7809824.htm>
3. <http://www.olap.ru/home.asp?artId=3102>
4. <http://samag.ru/archive/article/3135>

© Джаф Мохаммед Фадиль (hamoorjaf@yahoo.com).

Журнал «Современная наука: актуальные проблемы теории и практики»