Cubix BM и новые технологии.

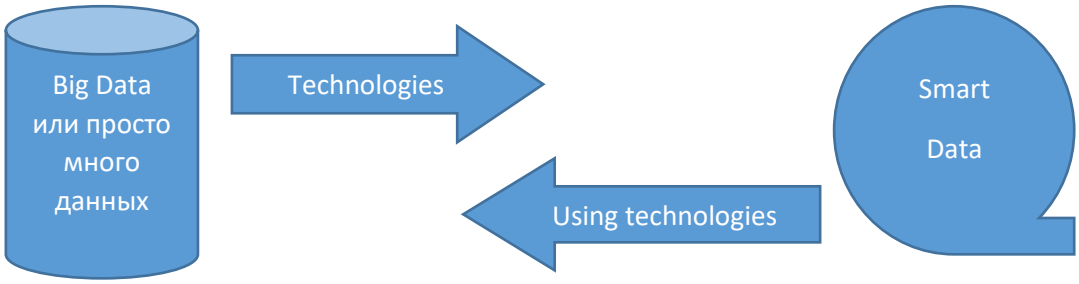
Идея: Бичев К.В. Босенко О.Г.

Автор Бичев К.В., редактор Босенко О.Г. Написана середина 2018 года, редакция июнь 2020 года.

Документ/статья– рассуждение на тему, на истину не претендует. Личное мнение и позиция при создании эффективно работающих систем.

Проблематика

В данный момент (середина 2018) при решении задач извлечения и использования данных есть два взаимно направленных вектора, которые очень стремятся встретится, но находятся в разных плоскостях



Представим **Две силы**, которые давят на задачу эффективности использования данных с разных сторон:

1. Накопившийся объем структурированных и неструктурированных данных, которые несут в себе огромную возможность для анализа, выводов и решений;
2. Необходимость формирования для людей выводов и решений, которые бы учитывали накопленные данные и помогали им принимать обоснованные и эффективные решения.

Кажется, что задача симбиотическая, что и данные хотят быть использованными, и люди хотят обработать и использовать их, но на практике оказывается, что существует одно противоречие, которое зачастую приводит к отсутствию фактических результатов:

- Данные напрямую отвечают на одни вопросы или молчат, а человек задает другие
    - Данные описывают факты реального мира, причем так, как это получилось у них
    - Человек запрашивает значимые **для него** выборки к значимым **для него** данным

То есть возникает ключевой вопрос: «Как преобразовать реальное к значимому или как извлечь значимое?»

Предпосылки.

Рассмотрим два часто используемых подхода для извлечения значимых данных: Агрегацию и Моделирование.

1. Агрегация:
    a. Простой метод, который позволяет сгруппировать данные к виду, который поможет сделать определенные выводы и принять решение;
    b. Главный недостаток, что данные остаются “реальными” и очень далекими от “значимых для принятия решений”;

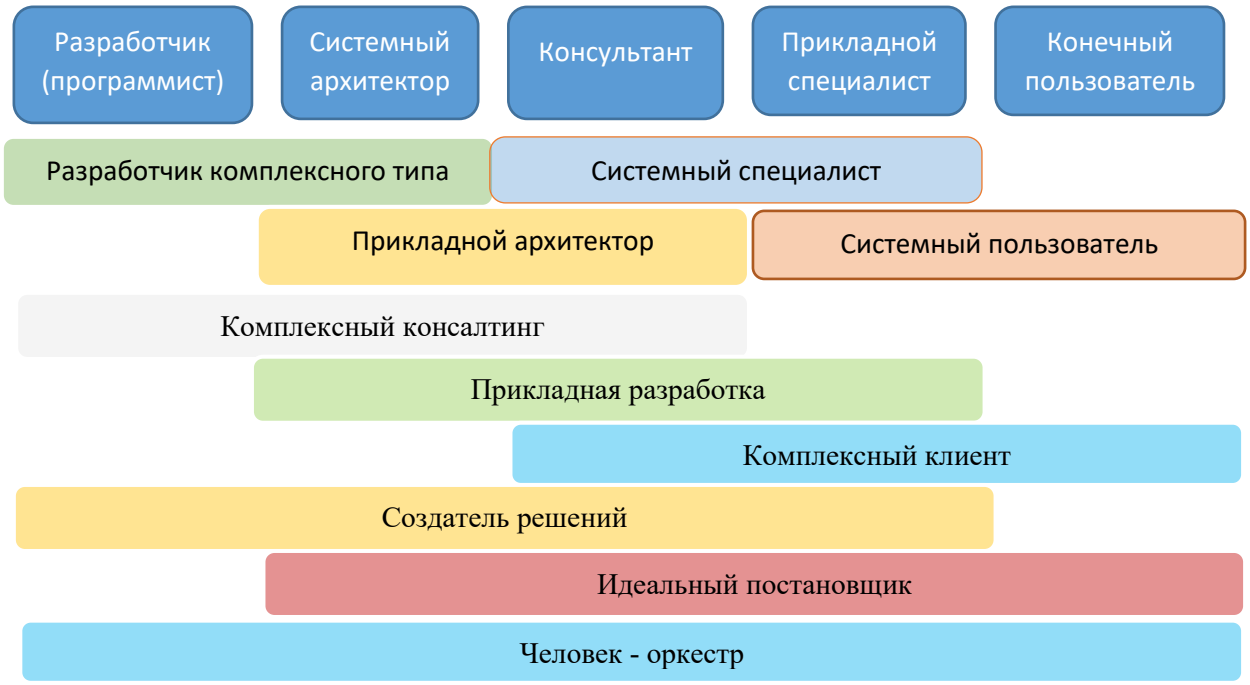
c. При этом предполагается, что главное преимуществом является то, что пользоваться данными может тот же человек, который запрашивает значимую информацию, то есть нет «потерь передачи информации».

2. Моделирование:
   a. Более сложный метод, который путем очистки, фильтрации, трансформации, агрегаций, группировок, модификаций, распределений и т.д. дает возможность получить значимые данные;
   b. Главный недостаток, что процесс комплексный, многоэтапный и при переходе между участниками данные теряются или падает их качество («потери передачи информации»);
   c. Главное преимущество - позволяет получить **результат**: **значимые данные**

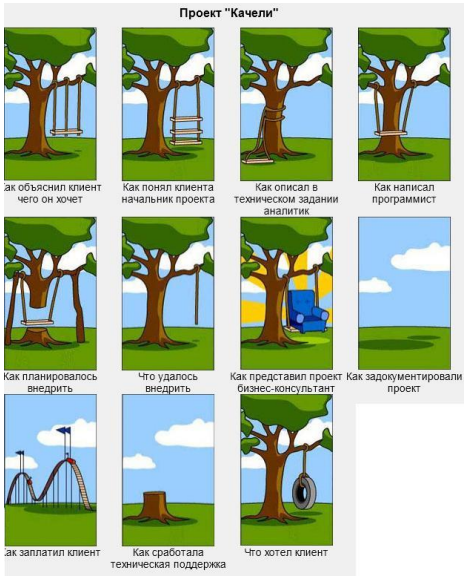
Немного технологии: Какие программные методы являются адептами этих подходов?!

1. Агрегационный подход сейчас применяют так называемые business intelligence платформы (BI), которые позволяют собирать данные из многих источников, агрегировать их и очень красиво представлять. Данный подход был очень востребован в период вакуума предложений машинного анализа, так как это намного больше, чем ничего. BI и сейчас востребован, так как комплексные решения второго типа еще не распространены.
2. Комплексный подход реализуют так называемые платформы бизнес-моделирования (EPM, CPM, BPM) (далее для простоты CPM - Corporate Performance Management), в которых строится модель модификации исходных “реальных данных” в значимые данные. В данный момент платформы этого класса приобретают все большую популярность, так как могут дать лучший результат по сравнению с классической BI платформой. Кроме того, CPM решения могут быть успешно переплетены с BI решениями (полученные значимые данные подаются как реальные с понятным красивым представлением и готовые к использованию)

Вернемся теперь к ключевой проблеме комплексного подхода «потери передачи информации». В данный момент проект по внедрению CPM выглядит схематично из следующих участников (верхний уровень на схеме), причем это - самый простой вариант, на практике подключаются разработчики системы интеграции данных (ETL), программисты и аналитики по различным продуктам, прикладные специалисты разных функциональных направлений и прочие, и прочие.



Результат получается, как на знаменитой картинке «Проект «Качели».



Есть более развернутые примеры проекта «Качели» с подключением открытого кода и т.д., но идея остается одинаковой (можно найти в интернете в разных вариациях и на разных языках):

Главные проблемы получения знаний из данных:

1. ***«Потери передачи информации»***
2. ***Слабые и узкие специалисты***

В этой связи очень высокий спрос в последнее время на комплексных специалистов, которые совмещают в своих навыках умения нескольких базовых ролей (ниже первого ряды на схеме). Такие специалисты показывают намного более высокие результаты.

Отсюда возникает логичное решение «минимизировать количество ролей» и использовать архитектурный подход при решении задач преобразования "реальных данных к значимым». Прикладной Архитектурный подход в обработке данных для задач повышения эффективности управления бизнеса собственно и получил у нас название "Бизнес Моделирование"

Может быть это Решение!

Так как повысить эффективность управления бизнесом? Приведем пример наших «трех китов» для решения этой задачи:

1. Образование: Все больше кафедр в отраслевых вузах с прикладным уклоном (автор документа получал высшее образование по специальности «Прикладная информатика в экономике» отраслевого ВУЗа), дополнительное образование по расширению ролевой принадлежности, перепрофилирование и т.д.
2. Бизнес: Все больше появляется должностей, в которых совмещаются роли, например, «Руководитель бизнес-процессов» (на схеме «Комплексный клиент»), сотрудник, основными обязанностями которого является разработка схем и регламентов отношений людей и данных во всей цепочке бизнес-процесса компании, и т.д.
3. Программное обеспечение: Создание программных продуктов, а чаще групп программных продуктов или комплексных платформ, например, платформа Cubix BM, которая позволяет специалисту роли «Консультант» (а, следовательно, «Системный специалист» и главное «Комплексный клиент») самостоятельно разрабатывать модели модификации реальных данных в значимые

Философия и стратегия первых двух направлений не являются предметом этой статьи, поэтому их можно опустить, хотя практические наработки в них также имеются из консалтингового опыта. Рассмотрим программное обеспечение и его тренды в общем и в частности Cubix BM, как пример платформы создания CPM решений.

Cubix BM – комплексная платформа для бизнес-моделирования.

Из вышеизложенного следует, что ключевой задачей Cubix BM является исключение из процесса создания решения ролей программиста и IT- архитектора(системного архитектора) и передачи функций от них к консультанту и/или предметному специалисту. У консультанта нет навыков

программирования и построения информационных систем, при этом он мыслит системно и может переложить реальный мир в сущности, параметры и т.д., поэтому инструмент должен:

1. Позволять достигать результата без непосредственного программирования задачи.
2. Исключить классический архитектурный подход к данным, в том числе и к BigData (таблицы, строки, столбцы), в котором не мыслит предметный специалист, а использовать метод объектов, через которые можно описать реальный мир с необходимой детализацией.
3. Позволять описывать связи между этими объектами
4. Поддерживать изменчивость решения
5. Содержать в себе все преимущества результатов, получаемых программистами и архитекторами, а именно:
    a. Разграничение прав участников процесса.
    b. Сценарность и вариативность процесса (то есть описание порядка, например, «каждый день в 8 утра загрузить данные по продажам, провести 22 модификации и подготовить результат»)
    c. Администрирование (например, «этот человек больше здесь не работает» или «эти данные признаны на значимыми, больше их в активных процессах не использовать»)
    d. И т.д.

Изучив актуальные методы и инструменты для решения задач была выбрана архитектура, которая представляет симбиоз нескольких технологических подходов.

Если документ/статья рассматривается, как теоретический, то можно читать дальше, но для того, чтобы дальнейшее предметное представление было более полным и понятным, нужно предварительно ознакомиться с основами создания моделей в Cubix BM и объектами, которые для этого используются, иначе понимание, где конкретно применяется технология, будет размыто.

Почти Big Data.

При обработке больших объемов данных используются несколько ключевых подходов, рассмотрим два часто используемых на практике:

1. **Приведение больших данных к менее большим данным без потери качества.**
    a. Например, контроллер токарного станка сообщает количество оборотов раз в секунду. Для задачи отказоустойчивости это важно, для дальнейшего анализа в процессе получения данных они сразу укладываются в объект поминутно в трех значениях (Max, min, avg), что приводит к 60\3=20 кратному уменьшению объема
2. **Уплотненное хранение данных**
    a. В платформе два состояния данных «архив» и «актив». В состоянии «архив» данные находятся в сжатом виде, описание хранится отдельно. Например, тот же токарный станок, мы знаем, что количество оборотов в секунду – это целое число от 0 до 12000, на каждое надо 4 байта и они будут идти последовательно в описании 1 секунда-одно значение, 1 минута-60 секунд, 1 час-60 минут и т.д. То есть данные в состоянии «архив» это 4байта+4 байта+4 байта и т.д., когда нам нужно значение от 12.12.2012 15:22:22 мы считаем на каком оно месте и забираем.
    b. В состоянии «актив» данные находятся в таком же виде только в адресном пространстве, например, 22-я секунда, 22-я минута и т.д. = 7000 в оперативной памяти, тем самым достигается возможность ввода и модификации данных в состоянии «актив»
    c. За счет оптимизации процесса «актив» <-> «архив» достигается беспрецедентное сочетание «реальный объем-объем хранения» и «объем-скорость»

3. Данный подход является самым современным в текущих подходах к Big Data (это подтверждают последние конференции по этой проблематике). В данный момент Cubix BM может отдавать и принимать данные в таком «плотном» виде. К сожалению, пока ни один из наших партнеров к такому подходу технически не готов, в результате чего для взаимодействия с внешними системами для импорта\экспорта используется классическая «ключ\значение», но мы убеждены, что наш подход будет востребован уже в ближайшее время

Некоторое отступление от темы: “Block Chain” подход.

В текущем виде технологический подход Cubix BM не может претендовать на «чистый» Block Chain, но это связано исключительно с отсутствием в этом необходимости. В нашем понимании (может это неверно, но мы практики и для углубления в теорию не хватает времени) существует два основных подхода к управлению распределенными данными:

1. Классический «древний» подход управления распределенными данными, когда данные находятся в разных местах, а единая система управления знает об адресах всех участников и данных, которые в них содержатся.
2. Block Chain, когда каждый следующий объект знает только о полезных ему соседях и ничего о внешнем мире. И нет ни одного участника, который знает всю картину.

У нас используется смешанный подход:

1. Каждый объект знает кто он, какие в нем данные, как они считаются и модифицируются, кто является его источником данных и по каким правилам их забрать, то есть каждый конечный куб - это звено цепи. Если даже на некоторое время выйдет из строя система управления цепью, то каждое звено будет работать как надо и результат от этого не пострадает.
2. Существует центральный процесс управления всеми звеньями, он их активирует, закрывает, опрашивает и т.д., так же есть база знаний об объектах, которая знает какой объект активен и что он делал и делает. Также «Кубы» в состоянии «архив» лежат в центральном хранилище, но единственный способ модифицировать объект «архив» - это процедура сохранения объекта «актив», тем самым поддерживается целостность данных.
3. В итоге объектная модель Cubix BM может быть определена как нарочито ограниченный block chain, что дает ему необходимую гибкость при сохранении устойчивости и качества данных внутри каждого основного информационного объекта – многомерного «Куба».

А если - Artificial Intelligence (AI)?!

Искусственный интеллект в общем понимании, как технология в Cubix BM, пока отсутствует, также как и практическое применение AI в общем понимании вообще. В настоящее время научить машину напрямую мыслить экономическими сущностями (как, впрочем, и многими другими) пока не удалось, поэтому выделяют несколько направлений машинного обучения, в которых есть явные практические достижения. Конкретные решения можно получить и на Cubix BM:

1. Предиктивное решение. В основе бизнес моделирования, как подхода, лежит предсказательная логика. Любой бизнес хочет быть устойчивым к внешним и внутренним событиям, в связи с этим строятся модели на основании исторических данных и логики построения. Cubix BM, как инструмент построения таких моделей, подходит полностью, так как:
    a. Предиктивные модели характеризуются высокой степенью волатильности. Таким образом с течением времени приходят не только новые данные, но и сигналы для изменения алгоритмов, а изменение работающих готовых больших моделей – это одно из преимуществ платформы Cubix BM;

b. Низкий уровень определенности на ранних этапах моделирования. Часто в процессе запуска моделирования понятна только конечная цель и объем различных данных, все картина модификации не видна, и она вырисовывается в процессе внедрения решения. Cubix BM позволяет запустить решение очень быстро и без использования больших ресурсов. По мере роста модели система масштабируется, подключаются новые аппаратные средства и все это практически «на лету».

2. Программное обучение. В данный момент еще не накоплен достаточный объем данных по практическому использованию платформы Cubix BM, но это не мешает уже сейчас работать над «помощником» построения моделей, который в процессе создания моделей будет предлагать варианты развития модели для ускорения и увеличения эффективности использования платформы. Но данный функционал еще не прошел должного тестирования и в качестве распространяемого решения не предлагается.

Промежуточные Итоги…

При выборе подхода к разработке всегда есть несколько вариантов, при этом выделяются два основных:

1. Упрощение объекта моделирования, но и уменьшение потенциала эволюционного развития, (например, за объект принимаем таблицу, ее потенциал ограничен; принимаем СУБД, потенциал выше, но и управление сложнее и т.д.)
2. Усложнение объекта моделирования, но и увеличение потенциала. (например, Cubix BM, где объектом является Куб, в который входят его аналитики, входящие связи, данные и т.д., и все это неразрывно связано и контролируется системой управления неразделимо)

Объединение данных и алгоритмов модификации этих данных в единый объект позволило развивать платформу в обоих направлениях: и внутрь, и наружу.

Выбранные технологические решения стали модными, когда уже были первые прототипы и внедрения Cubix BM. Это подтверждает верность выбранного пути в целом и технологического в частности.

Текущий результат уже является конкурентоспособным в сравнении с платформами создания CPM/BPM решений от ведущий производителей ПО в мире, а в потенциале Cubix BM станет флагманом данного направления ПО.