

Современные методы аналитики и визуализации Business Intelligence Лекция 8

Кирилл Сысоев

Обо мне

5+ лет в Big Data

HSE University

Senior Data Engineer

OneFactor/UZUM Data

Hadoop, Spark, ClickHouse, Kafka, Docker

Python/Scala, SQL



t.me/KRSysoev

krsysoev@edu.hse.ru

Взаимодействие

Общение:

Мой telegram – личные вопросы/консультации/рекомендации

Лекции + ДЗ:

Telegram-чат «НИС Современные методы аналитики и визуализаций» – после лекций буду туда публиковать материалы лекций и описание ДЗ с дедлайном

Сдача ДЗ:

Почта – в установленный дедлайн буду ждать письмо с вложением

BI

Business Intelligence (BI) – вид информационных систем, созданных для помощи аналитику в анализе информации о своей компании и её окружении. Большинство BI-инструментов применяются конечными пользователями для доступа, анализа и генерации отчетов по данным, которые чаще всего располагаются в хранилище, витринах данных или оперативных складах данных.

Важно отметить, что извлечение полезных знаний из накопленных фактических данных с целью принятия эффективных решений для бизнеса находится полностью в компетенции аналитика, а задача BI – предоставить инструментальную возможность к подобному анализу.

Цели

Ознакомиться с функциями и назначением BI-систем, познакомиться с платформой Tableau, возможностями моделирования и построения интерактивных дашбордов.

Что такое BI

По своей сути BI системы – это продвинутое аналитическое программное обеспечение для бизнес-анализа и формирования отчетности. Эти программы могут использовать данные из различных источников информации и предоставлять их в удобном виде и разрезе. В результате аналитик получает быстрый доступ к полной и прозрачной информации о положении дел компании. Особенность отчетов, полученных с помощью BI – возможность самостоятельного выбора аналитиком, в каком разрезе получить информацию.



Что такое BI?

В основе технологии BI лежит организация доступа конечных пользователей и анализ структурированных количественных по своей природе данных и информации о бизнесе. BI порождает итерационный процесс бизнес-пользователя, включающий доступ к данным и их анализ, и тем самым проявление интуиции, формирование заключений, нахождение взаимосвязей, чтобы эффективно изменять предприятие в положительную сторону. BI имеет широкий спектр пользователей на предприятии, включая руководителей и аналитиков.

Задачи BI

Задачи, решаемые BI , классифицируются по сложности реализации:

- Аналитические – вычисление заданных показателей и статистических характеристик бизнес деятельности на основе ретроспективной информации из баз данных
- Визуализация данных – наглядное графическое и табличное представление имеющейся информации
- Добыча знаний – интеллектуальный анализ данных, определение взаимосвязей и взаимозависимостей бизнес-процессов на основе существующей информации. К данному классу можно отнести задачи проверки статистических гипотез, кластеризации, нахождение ассоциаций и временных шаблонов

BI и Хранилища Данных

Хотя многие BI-инструменты имеют встроенный мощный функционал по подготовке и обработке данных перед построением дашборда (агрегация данных, построение модели витрины данных, расчет KPI и пр.), тем не менее, наилучшей практикой является проведение всей подготовительной работы на уровне КХД.

Ключевые преимущества такого подхода:

- Экономия вычислительных ресурсов
- Отсутствие избыточной, лишней информации, что также убыстряет процесс построения отчета
- Строгое регламентирование доступа к данным с разграничением прав
- Разграничение зон ответственности по этапам построения отчетов и дашбордов (формирование ТЗ, насыщение данными, отрисовка и разработка дизайна)

BI и Хранилища Данных

Концепция, методы и средства хранилища данных определяют подходы и обеспечивают интеграцию, очистку, ретроспективное хранение информации, предназначено для анализа, отвечают на вопрос «Как подготовить информацию для анализа?».

Технология бизнес-интеллекта определяет методы и средства доступа и оперативного анализа информации в терминах предметной области.

BI-средства не обязательно должны работать в инфраструктуре хранилища данных, но в этом случае проблема очистки и согласования данных возлагается на них, причем осуществлять эти операции придется на лету или же предварительно, но для обособленного информационного ресурса.

BI и Хранилища Данных

Кроме того, есть эффект влияния на производительность и надежность оперативной системы обработки транзакций.

Вот почему хорошей корпоративной практикой является выделение транзакционной и аналитической составляющих и применение для второй различных решений по хранилищу данных.

Основные стыки идут не только на уровне информации, но и на уровне метаданных. В случае хранилища данных можно обеспечить централизованное управление метаданными.

Функциональные Требования

к результатам ВІ аналитики

- Формирование текущих и заключительных отчётов в соответствии со стандартами компании, нормативными актами или потребностями управляющего персонала.
- Систематизация данных ВІ отчётности для задач бизнес-аналитики, например, для поиска слабых мест в организации работ, планирования расходов и т.д.
- Групповая консолидация, бюджетирование, учёт текущих и необходимых ресурсов, создание скользящих прогнозов, упрощение принятия решений.
- Мониторинг ключевых показателей эффективности и производительности, соответствия стандартам и регламентам.
- Информирование руководящих специалистов о ходе выполнения задач, занятости и ответственности сотрудников, предоставление других важных данных для организации работы компании.

Технические Требования

к инструменту BI аналитики

- **Анализ данных.**

Система должна «уметь» в несколько кликов мышью делать выборку из базы, сравнивать информацию по периодам, определять текущий уровень показателей, создавать сценарии и строить прогнозы на будущее.

- **Визуализация данных всеми популярными диаграммами** – гистограммы, секторные диаграммы, воронки, пузырьковые, радар и так далее. Очень желательна быстрая смена вида диаграммы, а также удобный доступ к их выбору.

- **Сбор данных из различных источников.**

Аналитическая система должна подключиться и импортировать все эти виды данных, а затем оперировать ими комплексно как единым целым.

Технические Требования к инструменту BI аналитики

- **Доступ через интернет-браузер.**

Если в работе с приложением можно использовать любое устройство (ПК, планшет, смартфон), это радикально расширяет рабочее пространство. Сотрудники компании смогут работать не только в офисе, но и в командировках, дороге, дома.

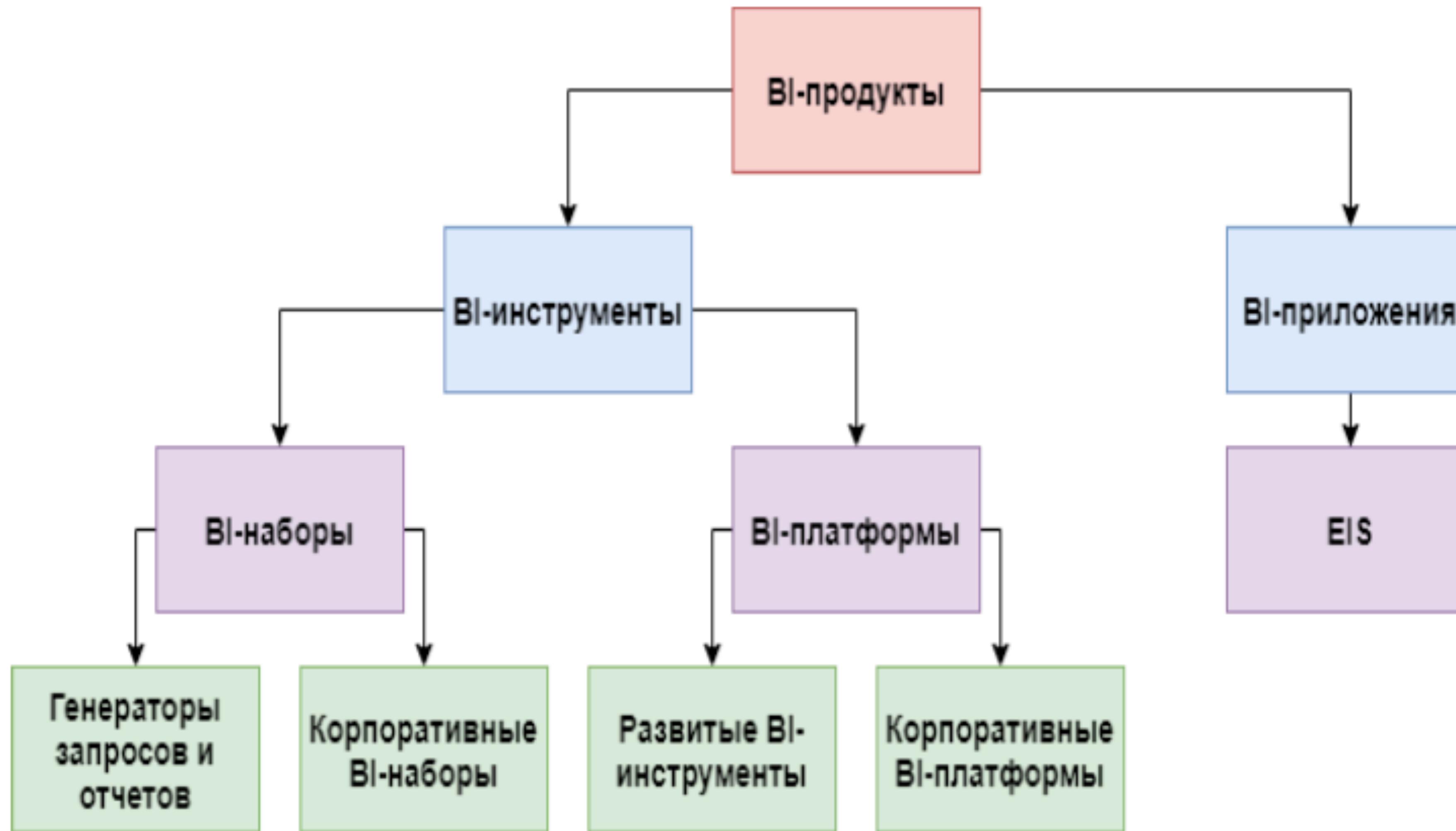
- **Настраиваемый внешний вид.**

Чем больше возможностей к изменению внешнего вида (настройка цветовой гаммы под брендбук, произвольное позиционирование элементов на листе, поддержка различных денежных и других форматов, настройка легенды и пр.), тем эффективней аналитики выполняют свои рабочие задачи.

- **Разграничение прав доступа.**

Это очевидный, но критично важный аспект. В идеале ИТ-решение должно полностью и буквально отражать всю структуру полномочий, действующую в компании.

Классификация BI-продуктов



Инструменты генерации запросов и отчетов

Генераторы запросов и отчетов – типично «настольные» инструменты, предоставляющие пользователям доступ к базам данных, выполняющие некоторый анализ и формирующие отчеты. Запросы могут быть как незапланированными (ad hoc), так и иметь регламентный характер. Настольные генераторы запросов и отчетов расширены также некоторыми облегченными возможностями OLAP. Развитые инструменты этой категории объединяют в себе возможности пакетной генерации регламентных отчетов и настольных генераторов запросов, рассылки отчетов и их оперативного обновления, образуя так называемую корпоративную отчетность (corporate reporting). В ее арсенал входят сервер отчетов, средства рассылки, публикации отчетов в Web, механизм извещения о событиях или отклонениях (alerts). Характерные представители – Crystal Reports, Cognos Analytics и OpenText Analytics Suite.

Корпоративные BI-наборы

EBIS (Enterprise BI Suites) – естественный путь для предоставления BI-инструментов, которые ранее поставлялись в виде разрозненных продуктов. Эти наборы интегрируются в наборы инструментов генерации запросов, отчетов и OLAP. Корпоративные BI-наборы должны иметь масштабируемость и распространяться не только на внутренних пользователей, но и на ключевых заказчиков, поставщиков и др.

Продукты BI-наборов должны помогать администраторам при внедрении и управлении BI без добавления новых ресурсов. Из-за тесного родства Web и корпоративных BI-наборов некоторые поставщики описывают свои BI-наборы как BI-порталы. Эти портальные предложения обеспечивают подмножество возможностей EBIS с помощью Web-браузера, однако поставщики постоянно увеличивают их функциональность, приближая ее к возможностям инструментов для «толстых» клиентов. Типичные EBIS поставляют Business Objects и Cognos.

OLAP или развитые BI-инструменты

Средства OLAP позволяют исследовать данные по различным измерениям. Пользователи могут выбрать, какие показатели анализировать, какие измерения и как отображать в кросс-таблице, обменять строки и столбцы «pivoting», затем сделать срезы и вырезки («slice&dice»), чтобы сконцентрироваться на определенной комбинации размерностей. Можно изменять детальность данных, двигаясь по уровням с помощью детализации и укрупнения «drill down/ roll up», а также кросс-детализации «drill across» через другие измерения.

OLAP-серверы оптимизированы для многомерного анализа и поставляются с аналитическими возможностями. Они обеспечивают хорошую производительность, но обычно требуют много времени для загрузки и многомерной модели. Поставляются, как правило, с возможностью «reach-through», позволяя перейти от агрегатов к деталям в РБД. Представители: Cognos TM1, Essbase, SAS OLAP и др.

BI-платформы

BI-платформы предлагают наборы инструментов для создания, внедрения, поддержки и сопровождения BI-приложений. Имеются насыщенные данными приложения с «заказными» интерфейсами конечного пользователя, организованные вокруг специфических бизнес-проблем, с целевым анализом и моделями.

Хотя EBIS являются высоко функциональными средствами, но они не имеют такого большого значения, как BI-платформы. Зато BI-платформы обычно не так функционально полны, как корпоративные BI-наборы. При выборе BI-платформ нужно учитывать следующие характеристики: модульность, распределенную архитектуру, поддержку стандартов XML, OLE DB for OLAP, LDAP, COM/DCOM и обеспечение работы в Web. Они должны также обеспечивать функциональность, специфическую для бизнес-интеллекта, а именно: доступ к БД, манипулирование многомерными данными, функции моделирования, статистический анализ и деловую графику. Эту категорию продуктов представляют фирмы Microsoft, SAS, ORACLE, SAP и другие.

BI-приложения

В приложения бизнес-интеллекта часто встроены BI-инструменты (OLAP, генераторы запросов и отчетов, средства моделирования, статистического анализа, визуализации и data mining). Многие BI-приложения извлекают данные из ERP-приложений.

BI-приложения обычно ориентированы на конкретную функцию организации или задачу, такие как анализ и прогноз продаж, финансовое бюджетирование, прогнозирование, анализ рисков, анализ тенденций, «churn analysis» в телекоммуникациях и т.п. Они могут применяться и более широко как в случае приложений управления эффективностью предприятия (enterprise performance management) или системы сбалансированных показателей (balanced scorecard).

Разведка данных и иные формы анализа данных

Разведка данных (data mining) представляет собой процесс обнаружения корреляции, тенденций, шаблонов, связей и категорий. Она выполняется путем тщательного исследования данных с использованием технологий распознавания шаблонов, а также статистических и математических методов. При разведке данных многократно выполняются различные операции и преобразования над сырьими данными (отбор признаков, стратификация, кластеризация, визуализация и регрессия), которые предназначены:

- для нахождения представлений, которые являются интуитивно понятными для людей, которые, в свою очередь, лучше понимают бизнес-процессы, лежащие в основе их деятельности;
- для нахождения моделей, которые могут предсказать результат или значение определенных ситуаций, используя исторические или субъективные данные.

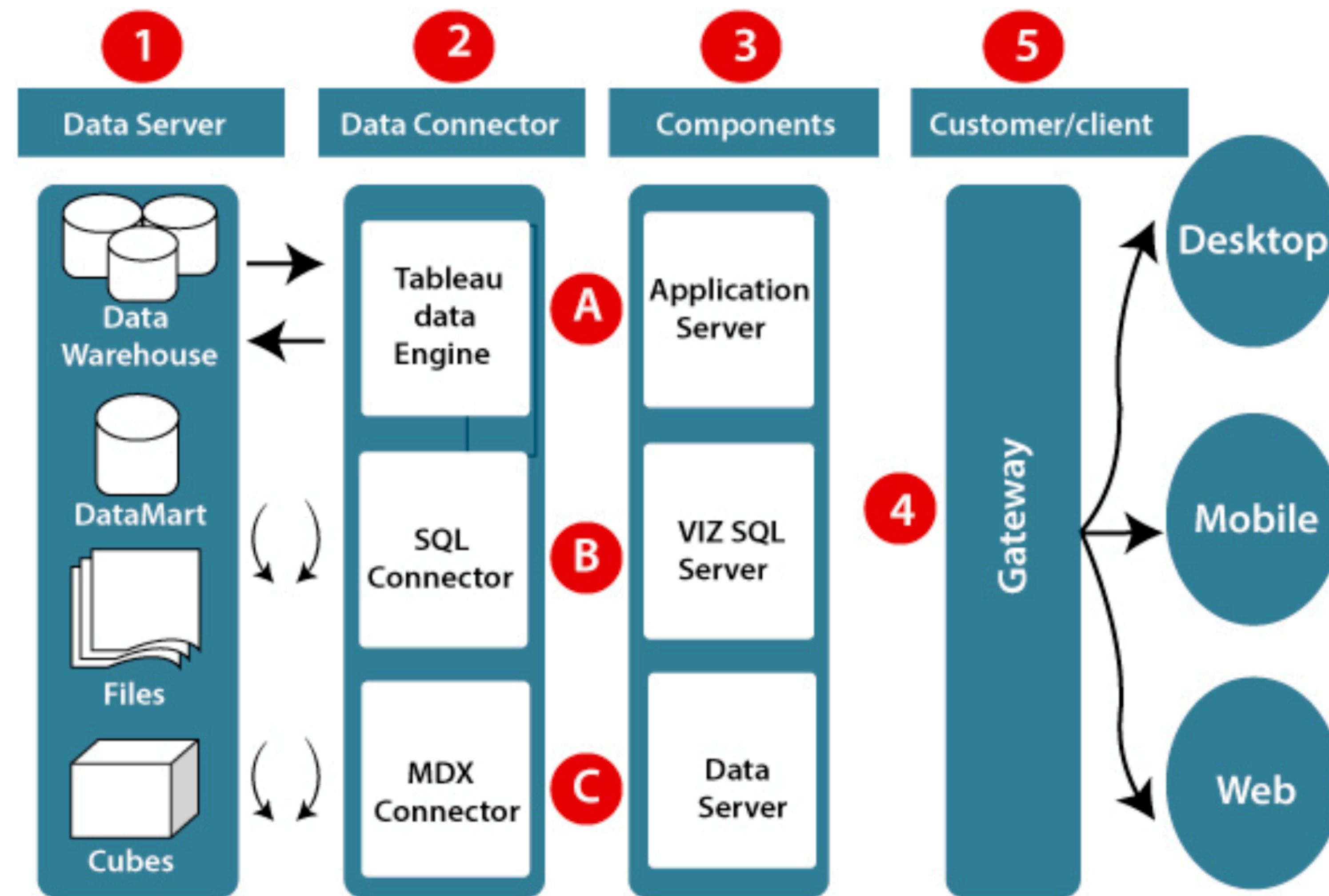
Разведка данных и иные формы анализа данных

В отличие от использования OLAP разведка данных в значительно меньшей степени направляется пользователем, вместо этого полагается на специализированные алгоритмы, которые устанавливают соотношение информации и помогают распознать важные (и ранее неизвестные) тенденции, свободные от предвзятости и предположений пользователя.

Кроме перечисленных инструментов, в состав BI могут входить следующие средства анализа: пакеты статистического анализа и анализ временных рядов и оценки рисков; средства моделирования; пакеты для нейронных сетей; средства нечеткой логики и экспертные системы.

Дополнительно нужно отметить средства для графического оформления результатов: средства деловой и научно-технической графики; «приборные доски», средства аналитической картографии и топологических карт; средства визуализации многомерных данных.

Архитектура Tableau



Преимущества Tableau

- Поддержка LIVE соединение с источником
- Поддержка технологии Data Blending
- Поддержка мобильных устройств
- Легкая интеграция с Big Data платформами
- Совместная работа над отчетом в реальном времени
- Бесшовная интеграция с R
- Множество встроенных инструментов для импорта данных из различных источников

Недостатки Tableau

- Конструктор загрузки начинает вызывать проблемы, когда нужно построить довольно сложную модель данных
- Своеобразная система построение дашбордов на базе отдельных листов с графиками
- Высокий уровень требований к компетенциям аналитика для объединения большого числа разных источников
- Данные необходимо предварительно структурировать

Спасибо за внимание!