

Курс «Хранилища данных»

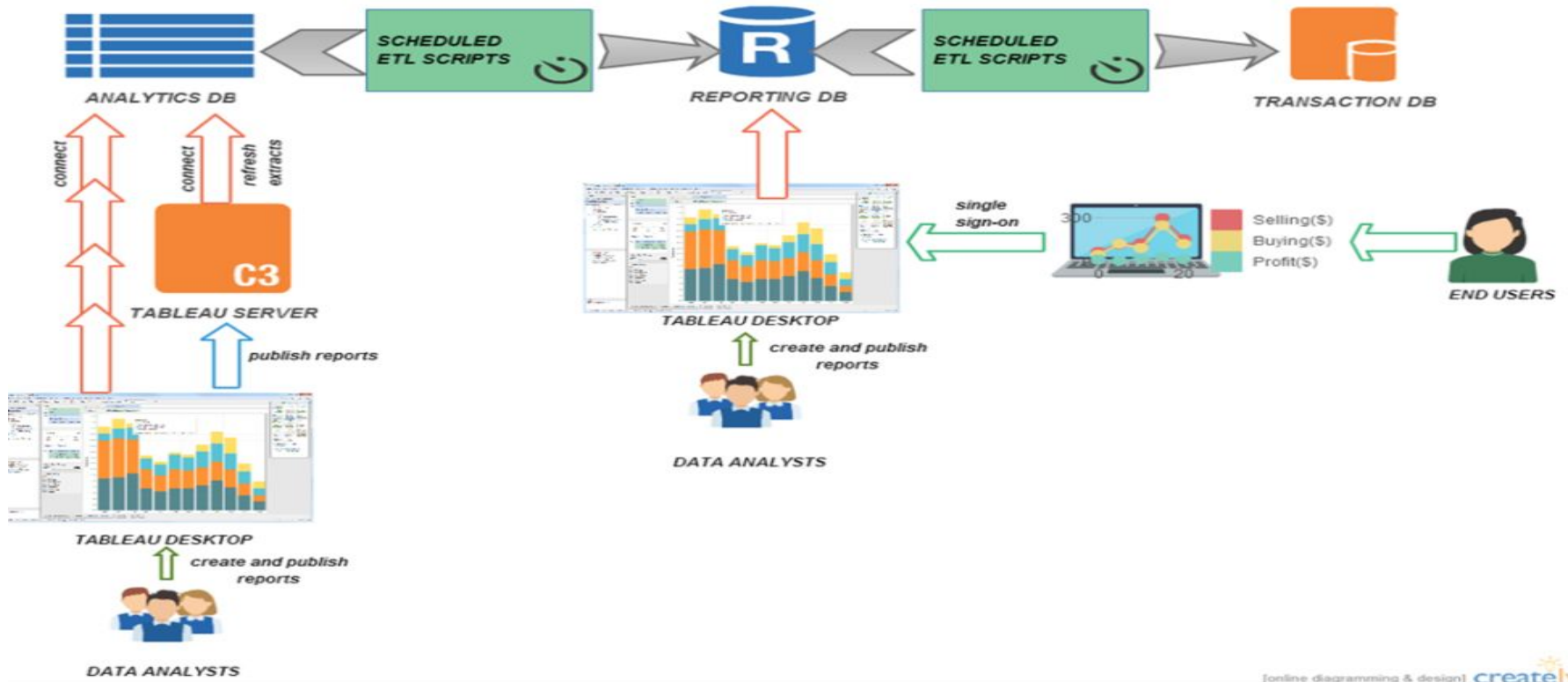
Тема: Tableau

Барабанщиков
Игорь Витальевич

Основные характеристики

- **Tableau является широко используемым ресурсом для визуализации данных и бизнес-аналитики и ориентирован на повышение эффективности аналитического рабочего процесса.**
- **Он может быстро обрабатывать миллионы строк данных, выполнять базовые вычисления и статистику и предоставлять результаты в секундах.**
- **Он может создавать информационные панели с использованием графического пользовательского интерфейса.**
- **Позволяет создавать интерактивные визуализации данных в режиме реального времени, объединяя источники данных, добавляя фильтры и сверяясь с конкретной информацией.**
- **Совместное использование панелей мониторинга не требует программирования, будь то на Tableau Server, Tableau Online или на любом портале или веб-странице.**

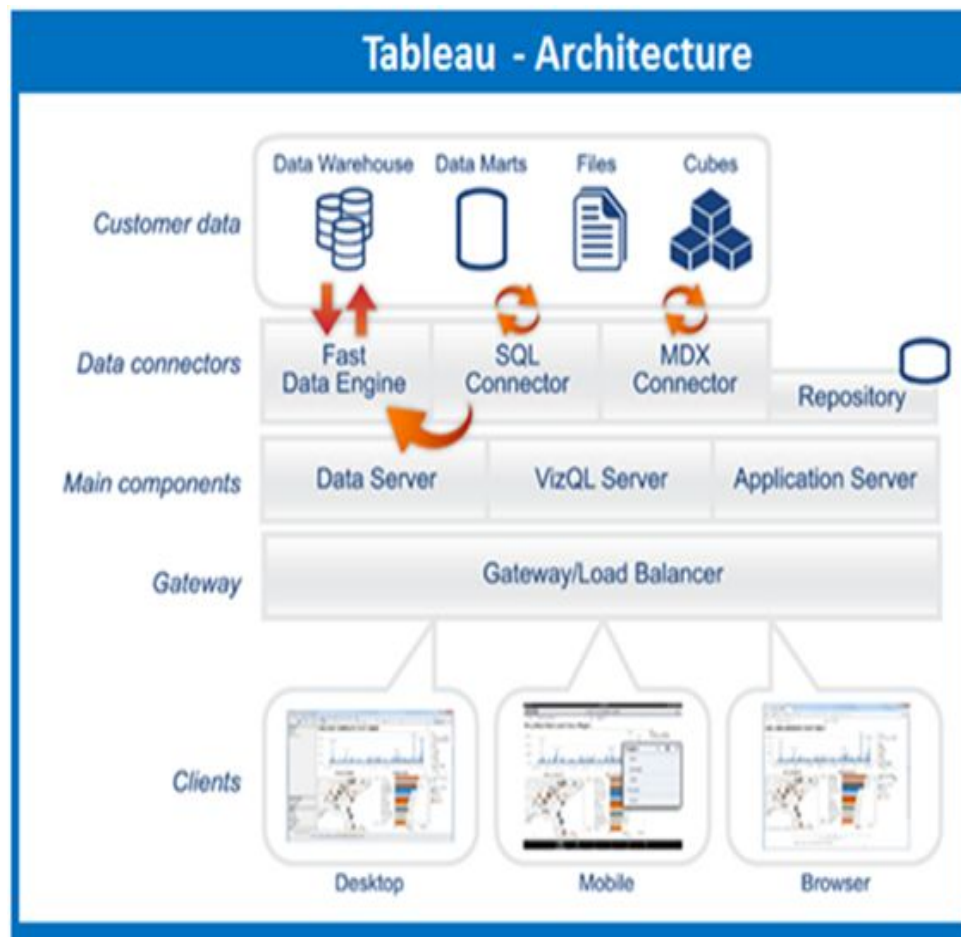
TABLEAU ARCHITECTURE



- Tableau имеет масштабируемую, n-уровневую клиент-серверную архитектуру, которая обслуживает мобильных клиентов, веб-клиентов и настольное программное обеспечение.
- Tableau Desktop - средство разработки и публикации, которое используется для создания общих представлений на сервере Tableau.

Архитектура Tableau

- Пользовательские источники данных
- Коннекторы данных
- Основные компоненты
- Балансировщик нагрузки
- Клиентские приложения

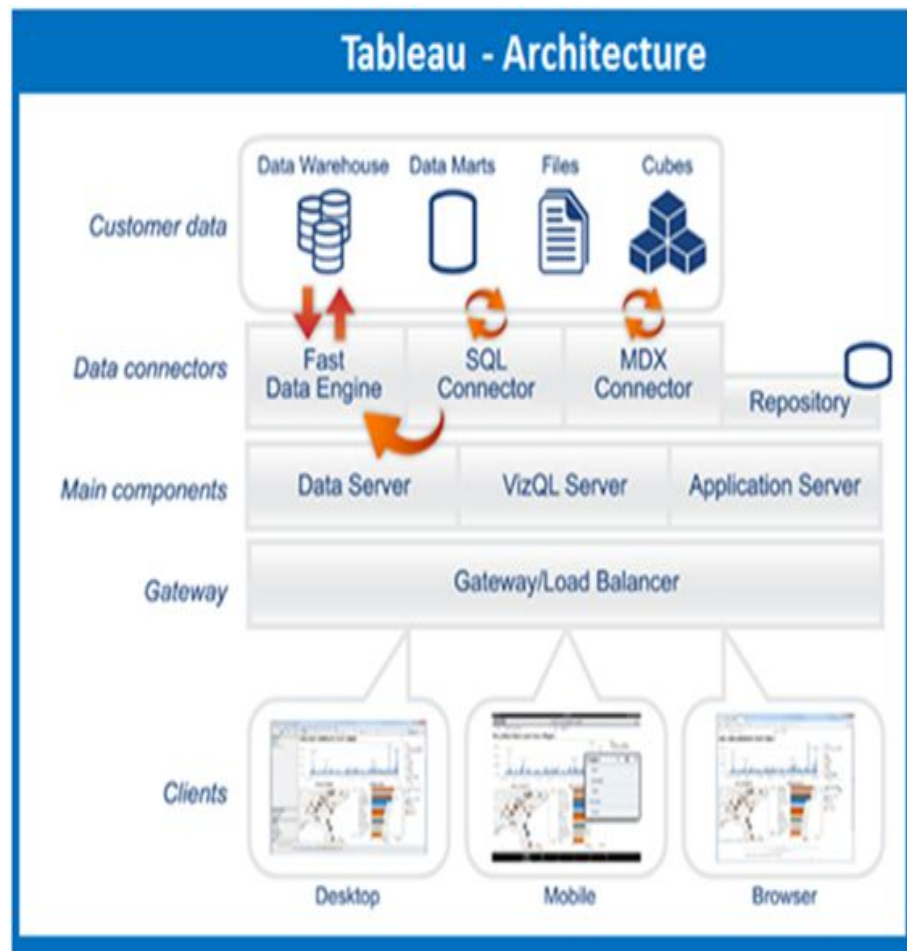


Уровень подключения данных

- В состав Tableau входят оптимизированные соединители данных для баз данных, таких как Microsoft SQL Server, Oracle, Teradata, Vertica, Cloudera, Hadoop и многие другие.
- Существует также подключение через ODBC.
- Не нужно получать все данные в памяти до тех пор, пока вы не решите ими воспользоваться.

Шлюз / балансировщик нагрузки

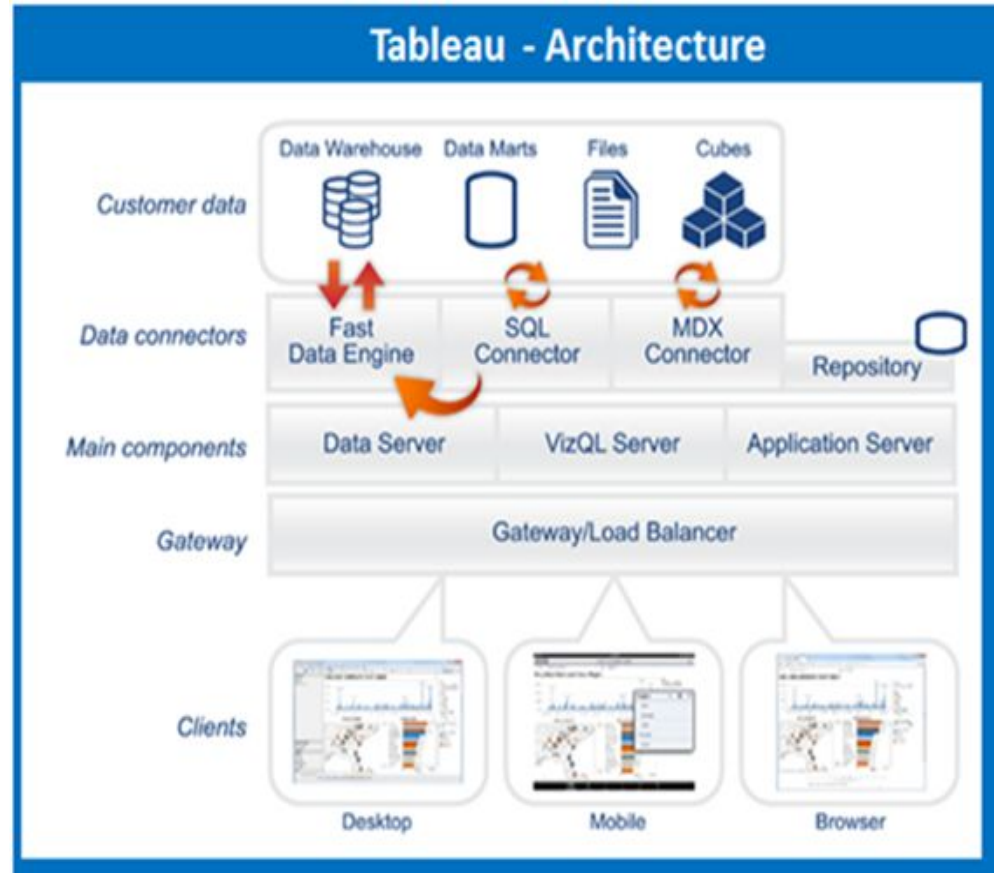
- Это основной сервер таблиц, который передает запросы другим компонентам.
- Запросы, поступающие от клиента, сначала направляются на сервер шлюза, а затем направляются к соответствующим процедурам.



Уровень клиента

Последний уровень показывает через какие устройства пользователи могут работать с Tableau:

- Desktopные приложения,
- мобильные приложения,
- браузер.



Два режима для взаимодействия с данными

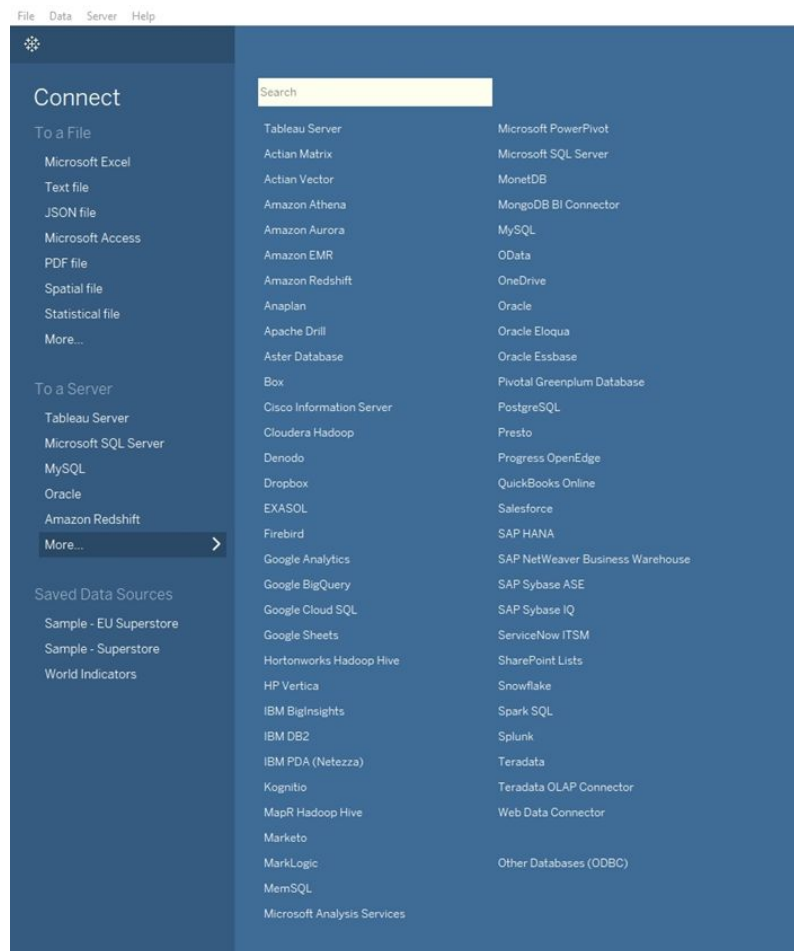
1. Extract – предварительное извлечение со сжатием всех данных из источника на диск и загрузка в оперативную память. Возможно инкрементальное обновление. Хорошо работает, когда данных не очень много (в пределах объёма оперативной памяти).

2. Live – Tableau запрашивает из источника только те данные, которые запросил пользователь (например, генерируются SQL запросы в случае реляционной БД). Это предпочтительный вариант, когда у вас есть DWH с объёмом данных 1 Тб и выше. Но есть существенный нюанс – структуры хранения данных должны быть хорошо оптимизированы для быстрого чтения.

Источники данных

В качестве источника данных платформа может использовать:

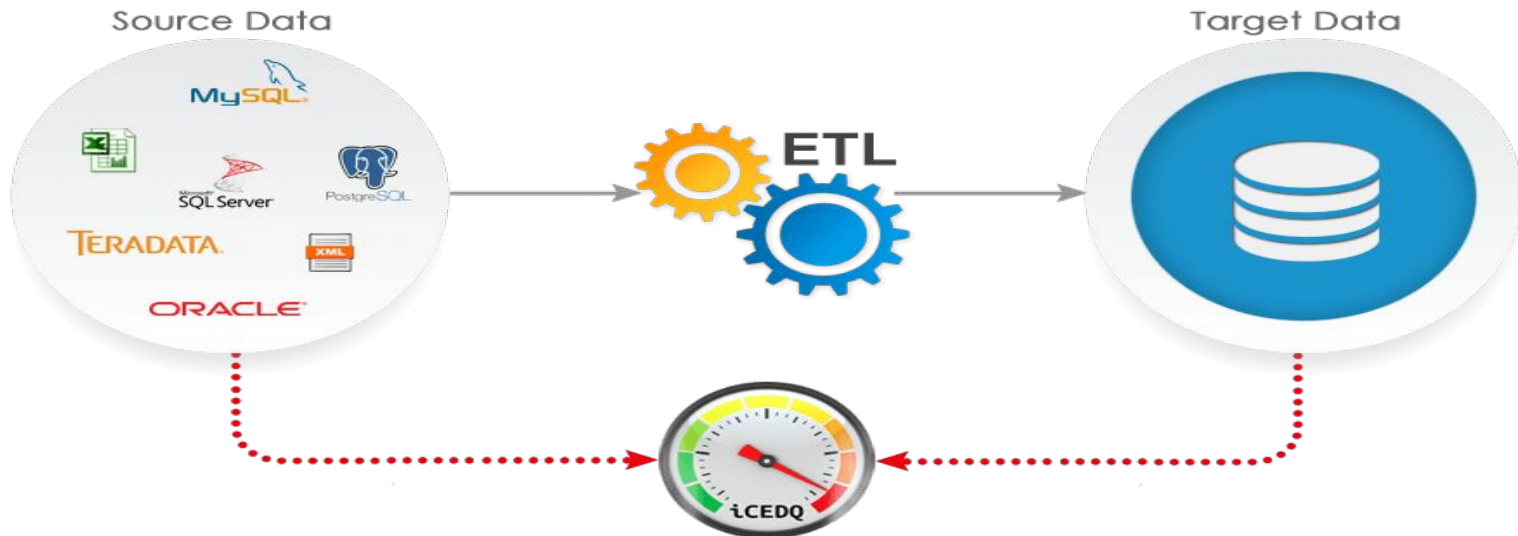
- **Облачные системы:** Google BigQuery, Windows Azure и т.д.
- **Реляционные системы:** SQL Server, DB2, Oracle и т.д.
- **Файловые системы** Excel, CSV и пр.
- Любые другие источники, которые используют программный интерфейс для доступа к базам



Загрузка данных

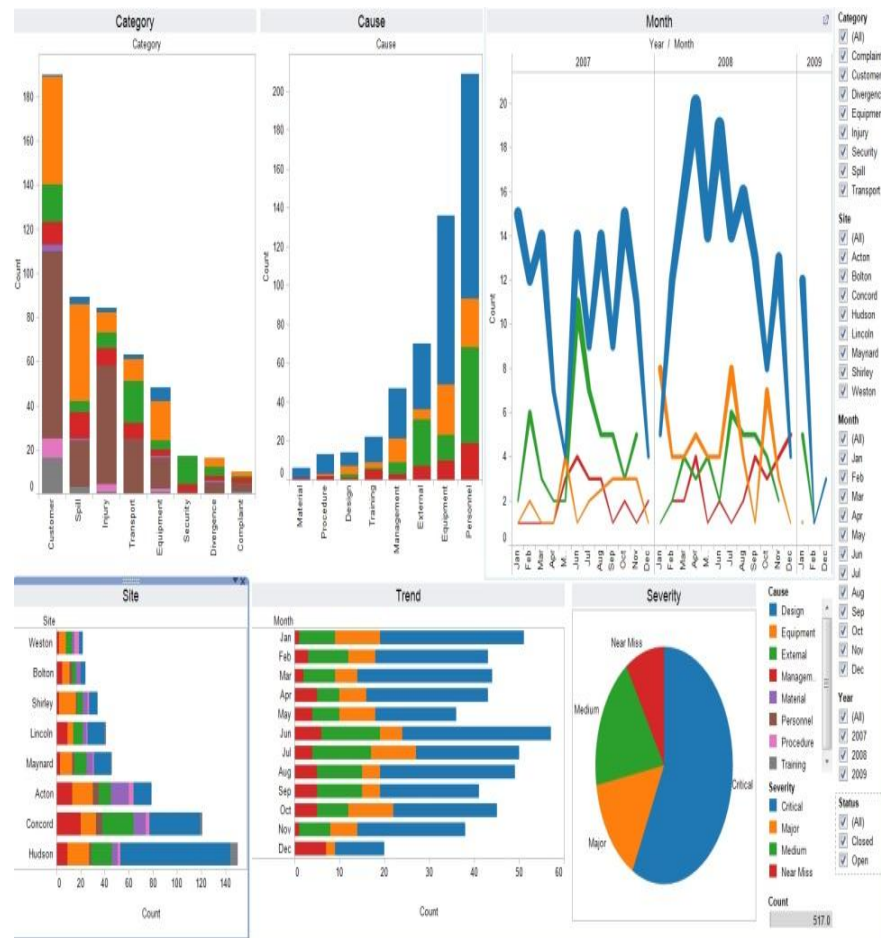
В Tableau не встроен собственный ETL-обработчик, поэтому **нельзя** грузить и обрабатывать любые данные.

Tableau требует уже хороших данных.



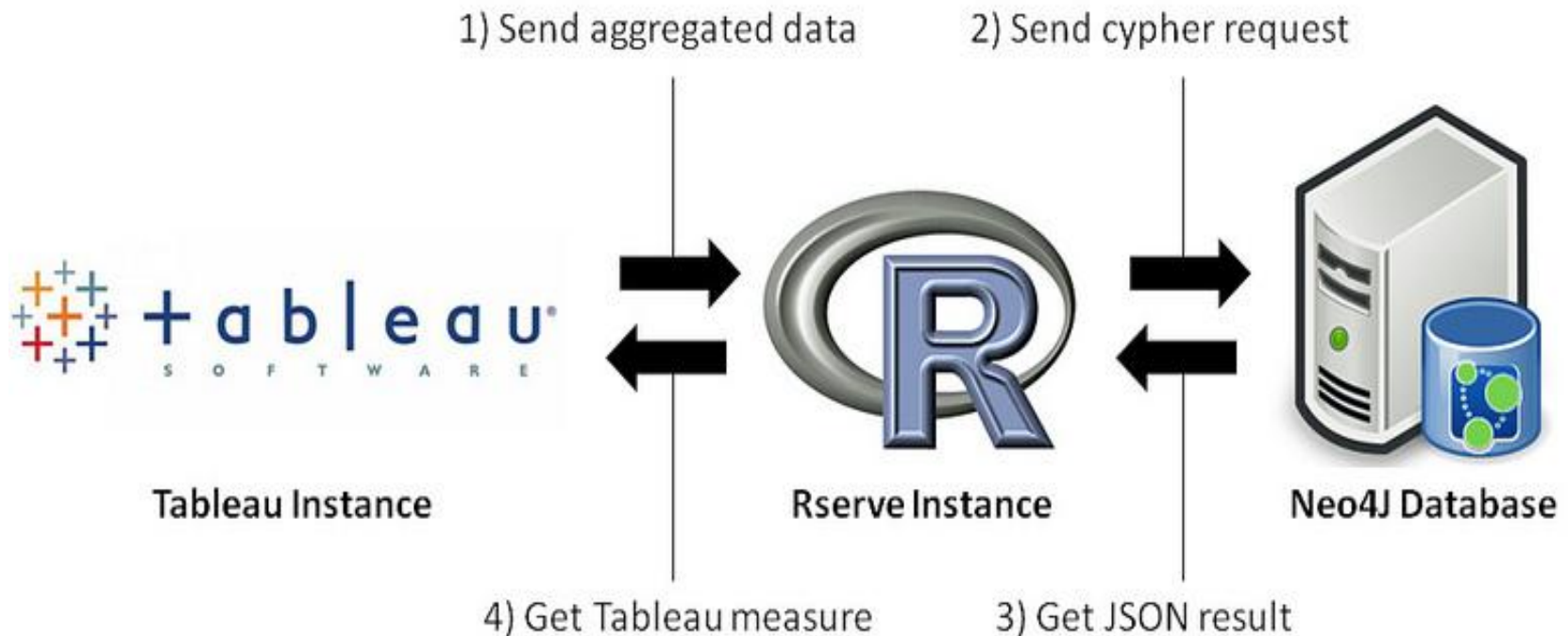
Визуализации данных

- Tableau позволяет создавать удобные визуализации данных и дашборды.
- Это можно делать в desktop-клиенте или в браузере.



Интеллектуальный анализ данных

Клиенты Tableau могут использовать расширенные возможности интеллектуального моделирования при помощи интеграции языком программирования R.



Взаимодействие с языком R

- в Tableau можно использовать R-функции и модели путем создания новых вычислительных полей, которые динамически вызывают R-машину и передают значения в R.
- Затем результаты возвращаются в Tableau для использования визуализационной машиной Tableau.

Достоинства

1. Отличные возможности для распространения отчетов и информационных панелей.
2. Поддержка более 30 типов данных.
3. Смешивание данных из разных источников.
4. Интеграция с R.
5. Самое активное сообщество пользователей, которые создают тысячи обучающих видео, блогов и форумов.
6. Легкость реализации сравнения по периодам
7. Визуализация данных.

Недостатки

1. Нельзя сравнивать агрегированные и не агрегированные данные.
2. Не обрабатывает иерархические связи (за исключением дат).
3. Нет глобального поиска
4. Оптимизация производительности является сложной задачей.
5. Отсутствие возможности «рисовать» отчёты строгой формы
6. Отсутствие локализации на русский язык
7. Требуется предварительной обработки данных (данные должны быть структурированы)
8. Создавать визуализации в «тонком» клиенте менее удобно, чем в варианте desktop.

Кто использует Tableau в России

1. МегаФон
2. S7 Airlines
3. НОРНИКЕЛЬ
4. MOMRIInstitute
5. РИА Новости
6. GlowByte
7. Ростелеком
8. Ростех
9. Газпром медиа
10. Ай-Теко
11. hh.ru
12. СОГАЗ
13. Билайн
14. ЭФКО
15. Вимм Билль Данн

Итоги

- **Tableau – один из лидеров** мирового рынка BI.
- Программное обеспечение Tableau **популярно в России.**



tableau