

Mr. Proper Data Management

Как обеспечить чистоту
данных в огромной
организации



Бауыржан Аубакир

Меня зовут Бауыржан

Jusan Bank – CDO, Kazakhstan (Big Data, DWH, CVM, NBO)

American University of Middle East – Data Science Instructor, Kuwait

Kazakhstan Aselsan Engineering – R&D Engineer

NU – Robotics



Оглавление

Типы неточностей

Способы поиска

Способы предотвращения

Культура чистоты

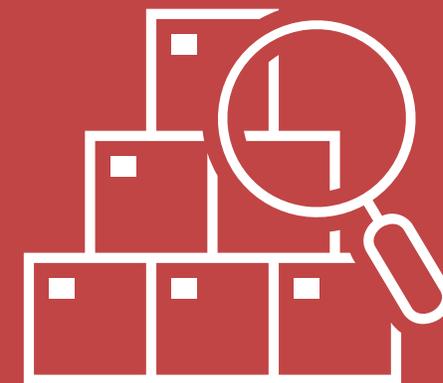
Типы неточностей



Категоризация проблем и неточностей в данных



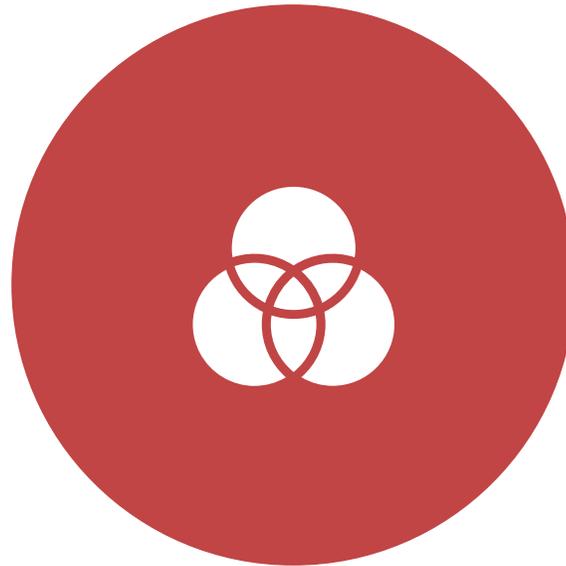
Способы поиска



Способы поиска ошибок в данных



Автоматические



Сравнение систем



Ручной

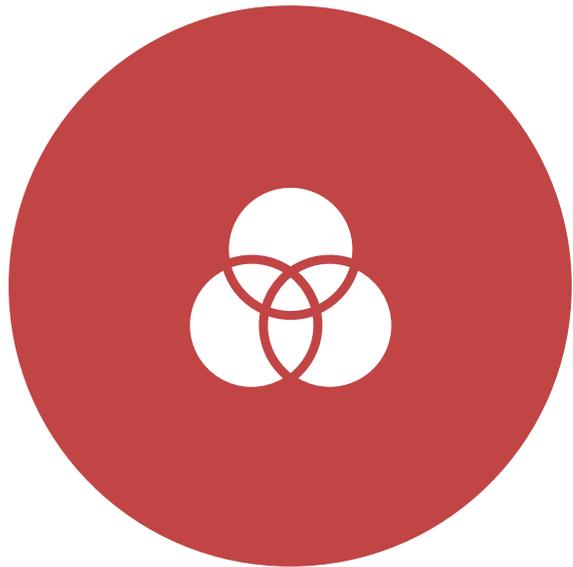


Автоматические

Расчет хэша

Поиск исторических аномалии

Ошибки загрузок



Сравнение систем

Сравниваем данные из 2 и более источников

Сравниваем балансы двух систем

Сверяем косвенные данные

Практика



Ручной

Смотрим глазами

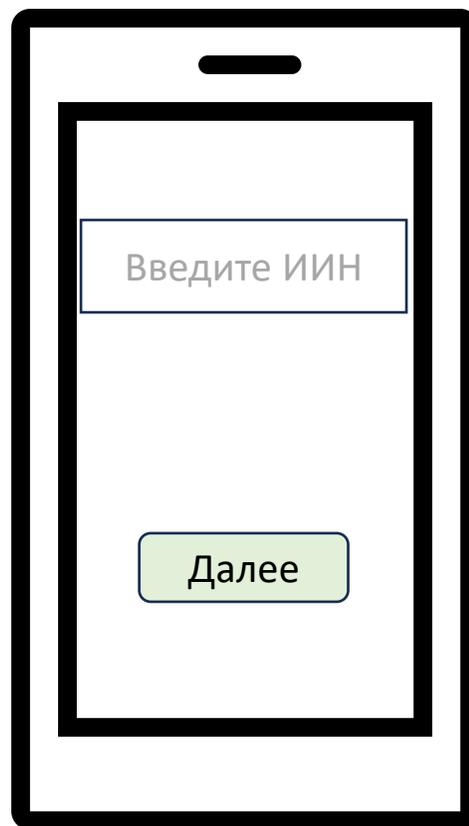
Нанимаем людей, которые мониторят

Способы предотвращения



Способы предотвращения

Интерфейс



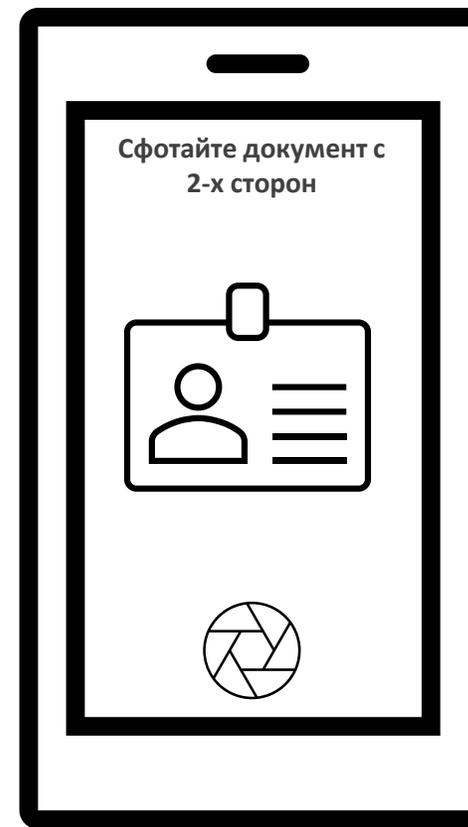
Способы предотвращения

Интерфейс



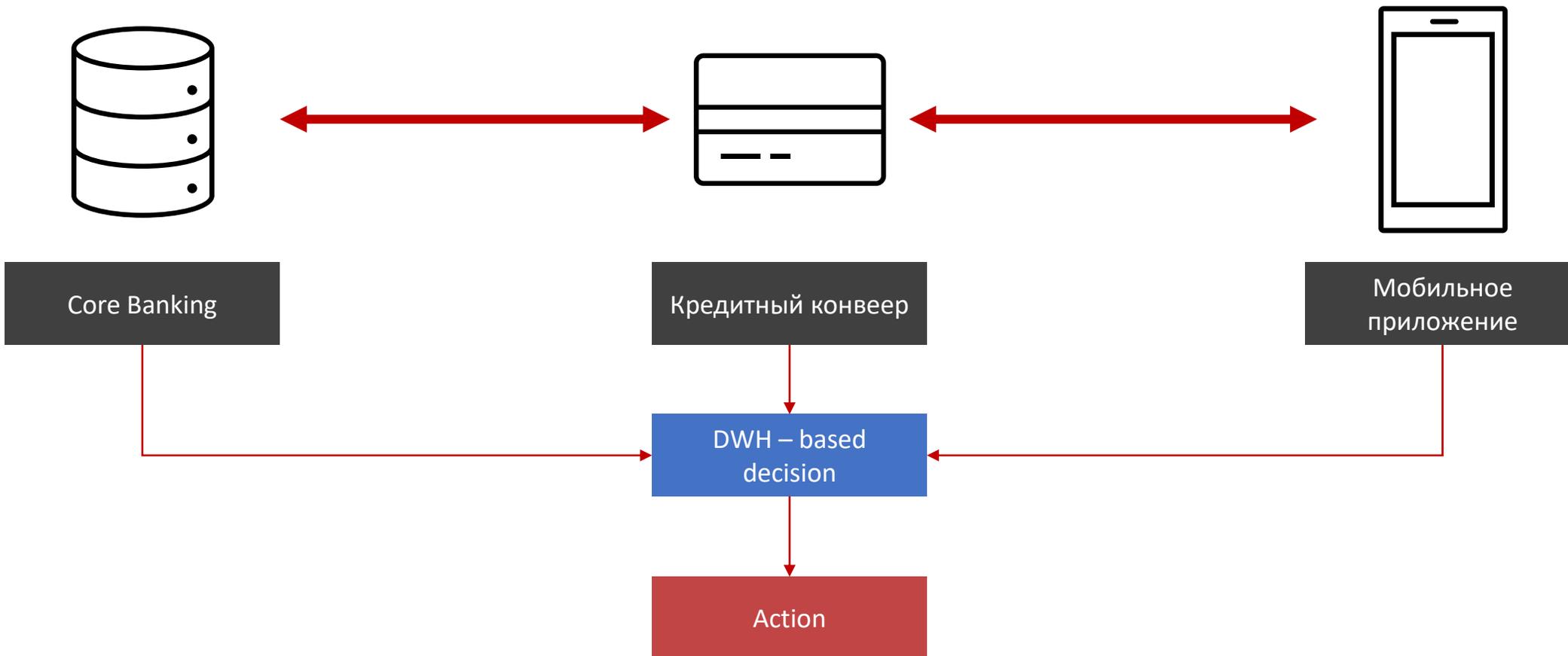
Способы предотвращения

Интерфейс



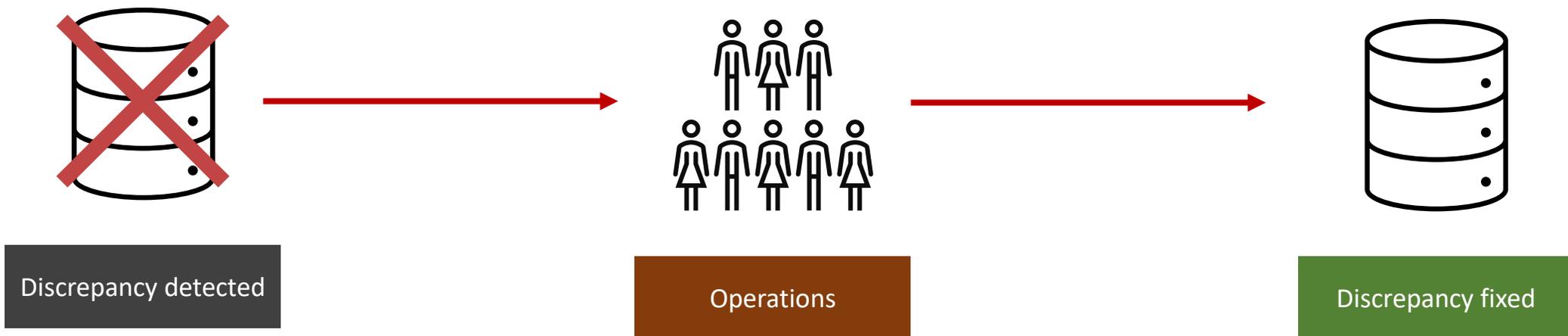
Способы предотвращения

Сравниваем системы



Способы предотвращения

Автоматически репортим аномалии



Культура чистоты



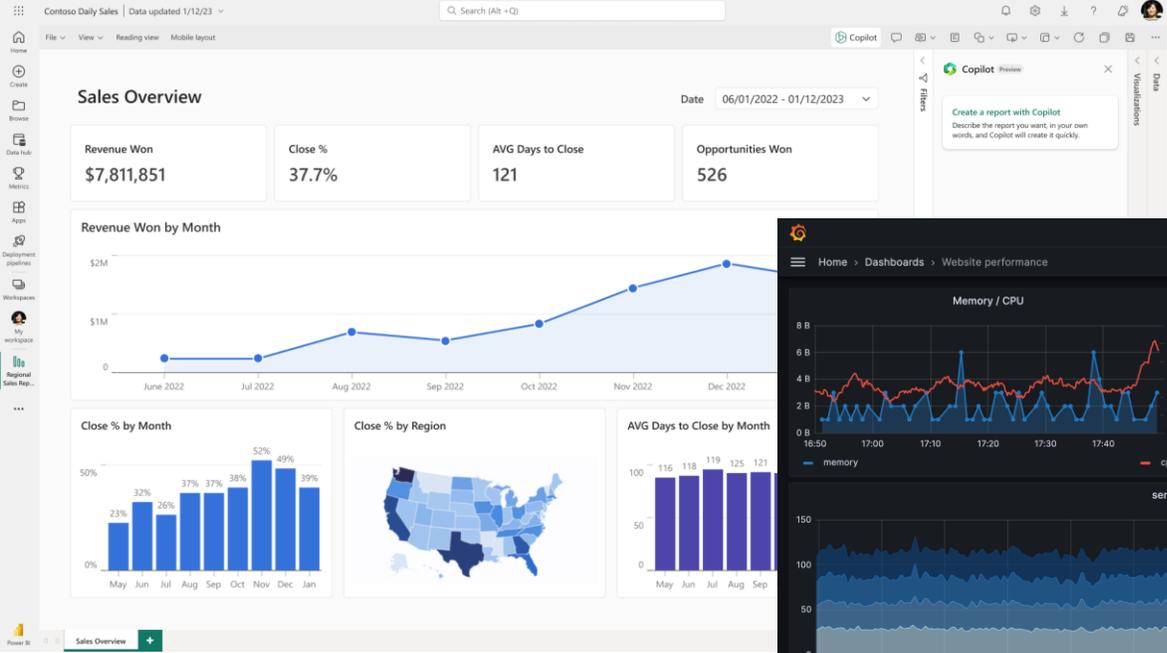
Культура чистоты

1. Демократизация данных
2. Свободный доступ к данным для аналитиков
3. Обеспечение инструментами для работы



Культура чистоты

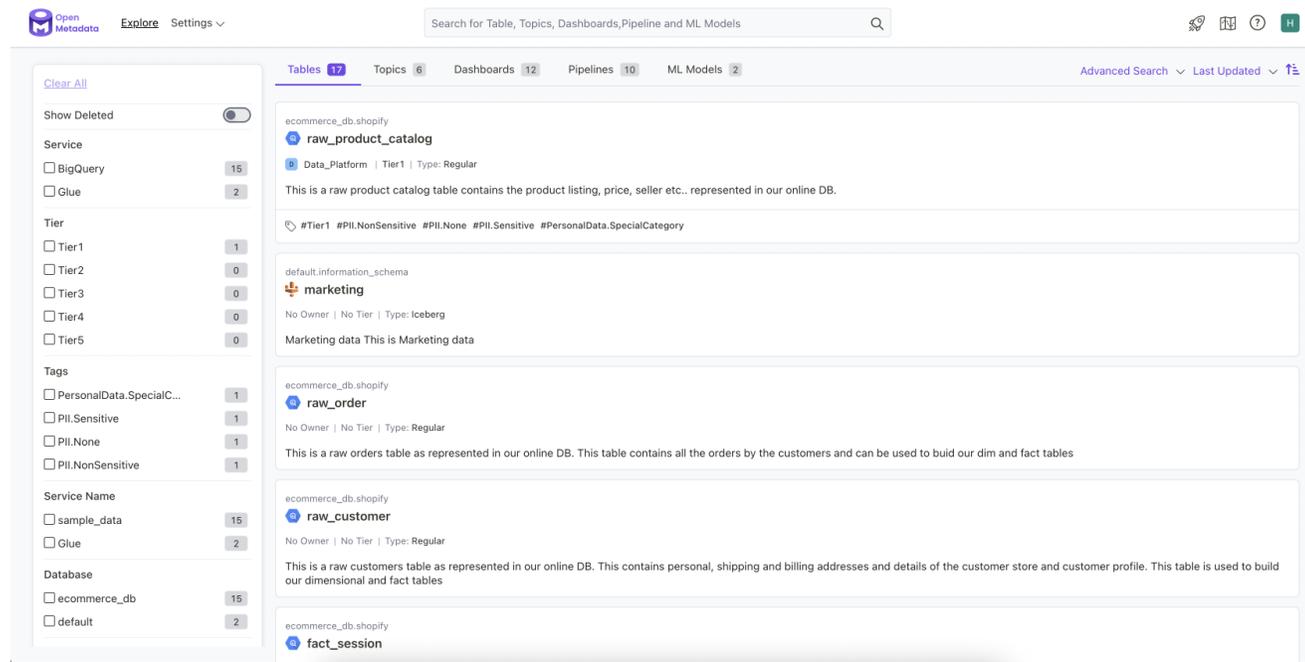
Access to Dashboards



Культура чистоты

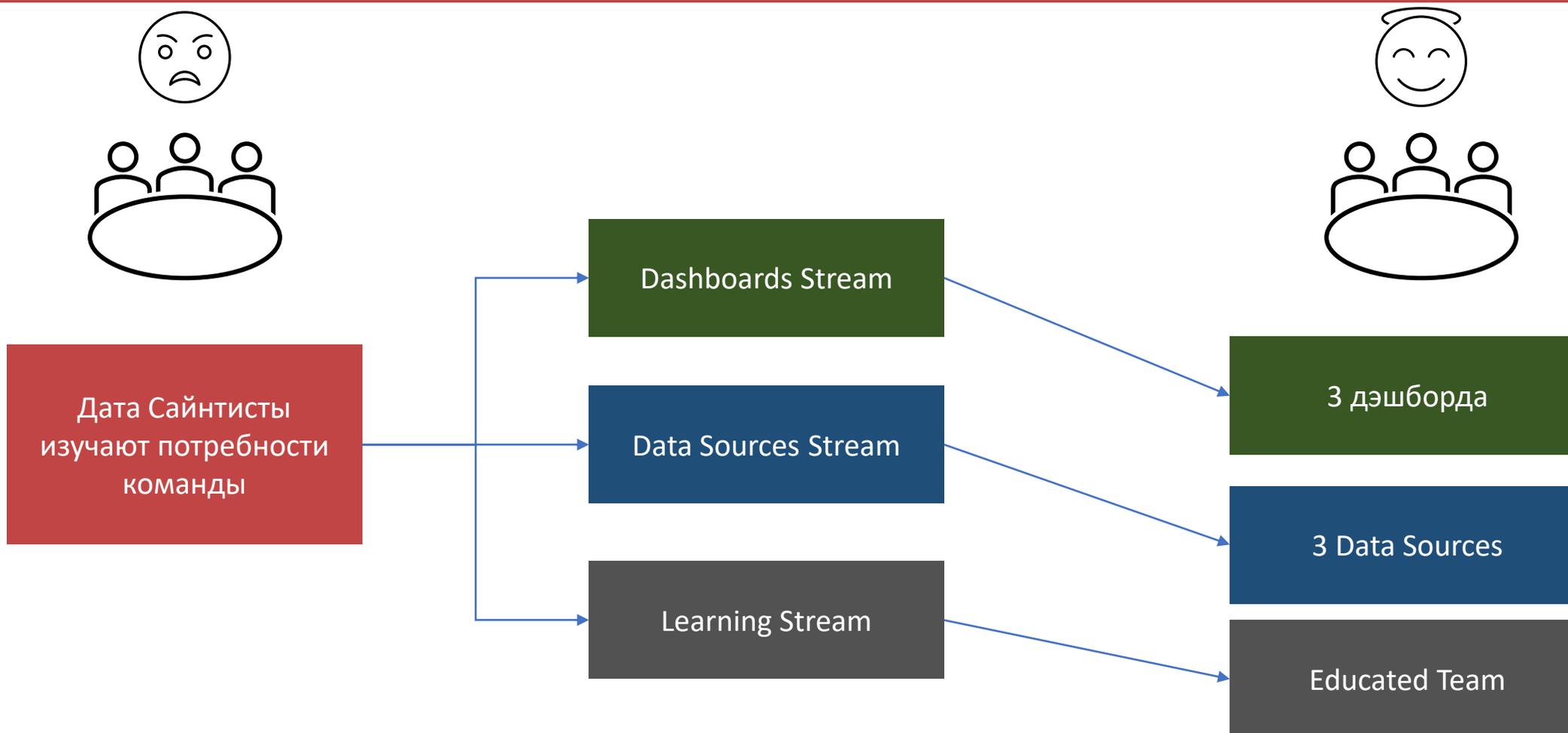
Access to Meta Data

1. Описание данных в одном месте
2. Удобный поиск
3. Версионность метаданных
4. Возможность коллаборации

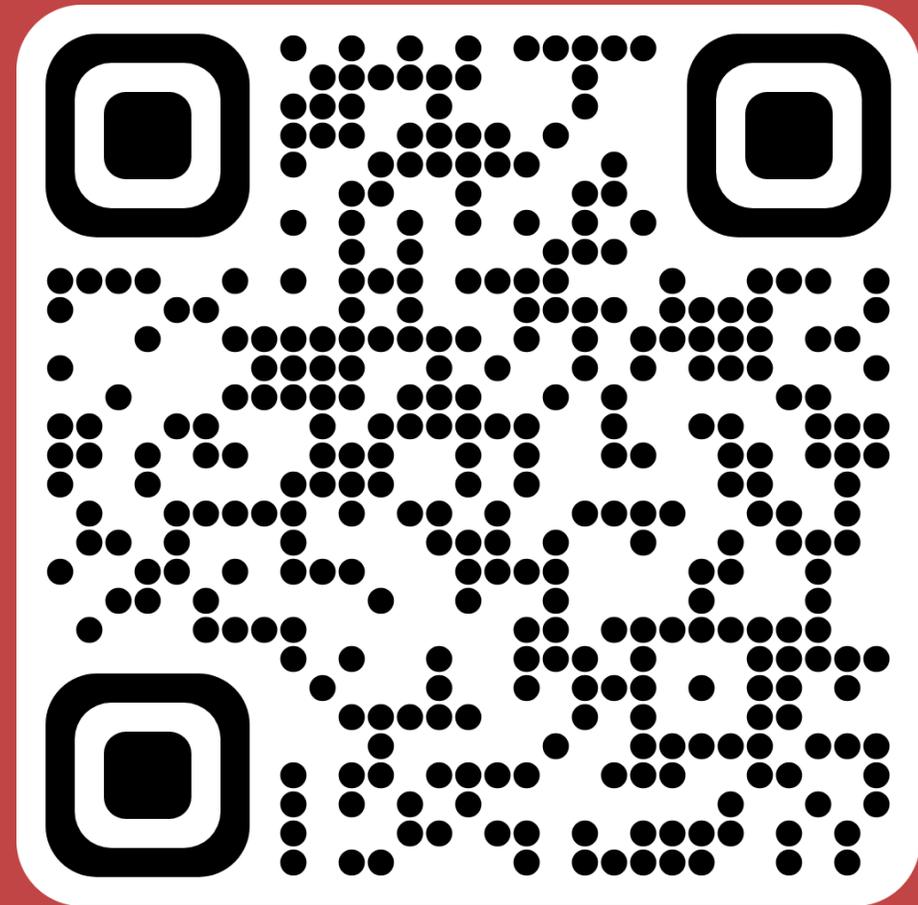


Культура чистоты

Data Partnership



Рахмет!



linkedin