

H F Labs

Модуль «Метрики качества данных»

«Единый клиент»

1. Описание

2. Основные функции модуля

- 2.1. Помогает оценить, сколько данных пригодно для бизнес-процесса
- 2.2. Реагирует на резкое изменение качества входящих данных
- 2.3. Позволяет проверить качество данных внешнего источника
- 2.4. Показывает источники некачественных данных

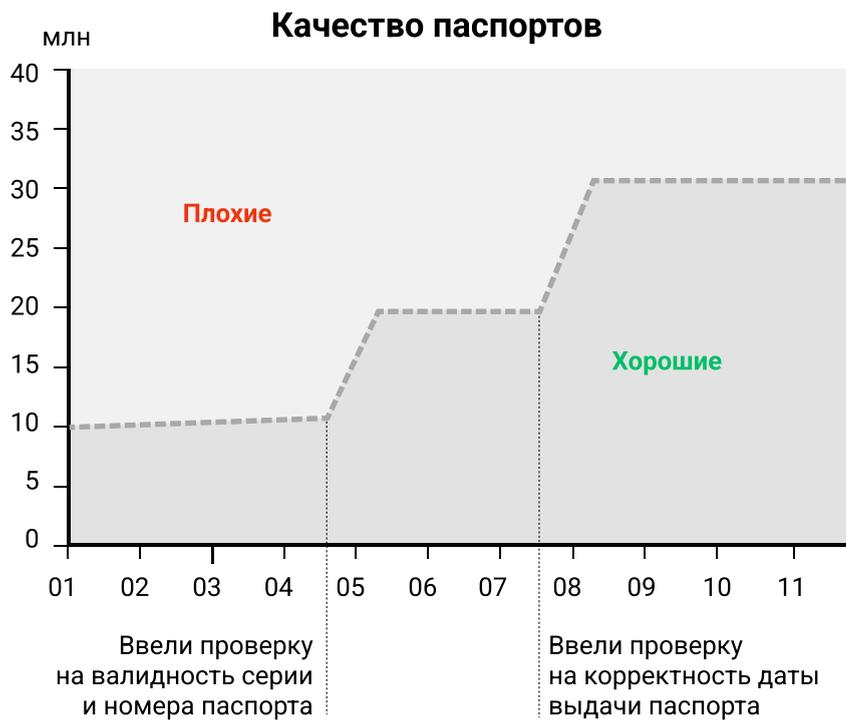
3. Технические особенности

- 3.1. Расчет, настройка и историцируемость метрик
- 3.2. Частота обновления по умолчанию
- 3.3. Хранение
- 3.4. Визуализация
- 3.5. Базовые наборы метрик

4. Ограничения

1. Описание

Позволяет контролировать и оценивать качество данных в «Едином клиенте» и исходных системах, находить источники проблем с данными и отслеживать здоровье интеграций.



1. Описание

«Единый клиент» собирает информацию из исходных систем с разным качеством данных. У каждого отдела компании свои требования к поступающей информации:

- *Директор Семен Семеныч желает видеть, как меняется график прироста клиентов и анализировать его пики и плато.*
- *Аналитик риск-подразделения Лиза хочет понять, почему за последние полгода в базе стало на 30% больше дублей.*
- *Логист Николай собирается проверить, можно ли доставить письма по адресам из нового источника.*
- *Инженер отдела сопровождения Екатерина нуждается в моментальном оповещении о проблемах с интеграциями.*

Модуль «Метрики качества данных» помогает бизнесу решать эти и подобные задачи.

2. Основные функции

2.1. Помогает оценить, сколько данных пригодно для бизнес-процесса

Показывает качество «золотых карточек» в базе. Содержит готовые наборы метрик для оценки:

- динамики изменения количества клиентов,
- степени задублированности базы,
- изменения качества данных во времени.

Позволяет в любой момент добавить новые метрики в соответствии с задачами бизнеса.

2.2. Реагирует на резкое изменение качества входящих данных

Помогает не пропустить сбой в системах-источниках: показывает качество входящих данных в моменте. Поддержка сразу увидит проблемы, возникшие в источнике или интеграции.

2. Основные функции модуля

2.3. Позволяет проверить качество данных внешнего источника

Или отследить начало заполнения какого-то поля в уже интегрированном источнике. Поможет разным отделам понять, будут ли полезны данные из источника: сможет ли маркетинг с их помощью сегментировать клиентов, а логистика — доставить письма.

2.4. Показывает источники некачественных данных

Помогает составить план исправления и отследить, как меняется качество. Покажет, работают ли идеи по улучшению ситуации или нет. Компания может устанавливать KPI по качеству данных и контролировать выполнение показателей.

3. Технические особенности

3.1. Расчет, настройка и историцируемость метрик

- Модуль вычисляет основные метрики по поисковым индексам «Единого клиента», чтобы не нагружать сервер СУБД. Дополнительно метрики можно собирать с помощью SQL-запросов.
- Позволяет настроить наборы метрик, расписание и частоту сбора под потребности заказчика. Новые метрики можно добавить в любой момент времени.
- Хранит историю метрик, что позволяет проанализировать динамику качества данных.
- Имеет настраиваемую политику эскалации.

3.2. Частота обновления по умолчанию

- Метрики стандартизации на входе — раз в 15 секунд. Показывают качество данных в исходных системах. Рассчитываются в момент очистки.
- Метрики качества данных в базе «Единого клиента», метрики задублированности и «серой зоны» — раз в сутки. Позволяют отслеживать ежедневную динамику качества данных. Считаются фоновыми задачами, запускаются в момент простоя приложения. Предагрегированы.

3. Технические особенности

3.3. Хранение

Для хранения и организации информации модуль использует открытое программное обеспечение.

Базово, из коробки:

- Victoria Metrics.

Понимает и может поддерживать:

- Prometheus,
- Influxdb.

3.4. Визуализация

Все метрики визуализируемы. В поставку модуля входит настройка одного дашборда. Заказчик получает доступ ко всем исходным данным метрик для самостоятельного построения дашбордов или может заказать набор дашбордов под свои потребности.

Для визуализации модуль использует открытое программное обеспечение.

Базово, из коробки:

- Grafana,
- Metabase.

Понимает и может поддерживать:

- Superset,
- Redash.



3. Технические особенности

3.5. Базовые наборы метрик

- заполненность полей на источнике,
- реквизиты ФЛ и ЮЛ по кодам качества в объединенной базе,
- реквизиты ФЛ и ЮЛ по маркерам изменений на исходных данных,
- разбивка атрибутов ФЛ и ЮЛ по кодам качества,
- разбивка атрибутов по популярности,
- разбивка атрибутов по типам,
- задублированность по источникам,
- число одиночных и объединенных карточек,
- разбивка серой зоны по правилам.

4. Ограничения

Сбор метрик начинается в момент подключения модуля.

Собрать данные за период времени до подключения невозможно.

H F Labs

✉ hflabs_bdm
📞 +7 495 928-86-41
✉ ask@hflabs.ru
🌐 hflabs.ru

119034, Россия,
Москва, Турчанинов переулок,
д. 6, стр. 2 офис 104,
БЦ «Крымский мост»