**Документация, содержащая информацию, необходимую для эксплуатации экземпляра программного обеспечения**

**Модуль «BI аналитика для стриминговых сервисов»**

**г. Москва 2025**

**Содержание**

[1. Введение 3](#_heading=h.h9vqitmibr9s)

[2. Запуск модуля 3](#_heading=h.a2lao0ocdbme)

[2.1. Предварительные требования 3](#_heading=h.ez42fugav00y)

[2.2. Установка и запуск 3](#_heading=h.vb1m1fkmxp6j)

[3. Аналитические инструменты 3](#_heading=h.vwnz9innfno7)

[3.1. Примеры доступных дашбордов 3](#_heading=h.z22y5kwjckll)

[3.1.1. Среднее время просмотра на пользователя 3](#_heading=h.3x157kqwjmdz)

[3.1.2. Click-Through Rate (CTR) в рекомендациях 4](#_heading=h.fibnh55xv85q)

[3.1.3. Число показов видео по категориям 5](#_heading=h.hl4mzzheezjk)

[3.1.4. Click-Through Rate (CTR) в поиске 6](#_heading=h.epagrfz7v1w4)

[3.1.5. Доля сессий с поисковыми подсказками 6](#_heading=h.oi440otrt91i)

[3.2. Средства Adhoc аналитики 7](#_heading=h.c335vwqnxzct)

[3.2.1. Аналитика в ClickHouse 7](#_heading=h.1mvvot7u7622)

[3.2.2. Аналитика в Apache Superset 7](#_heading=h.jqyz40vkl0mi)

[3.2.3. Аналитика в Apache Spark 7](#_heading=h.yla98kmcbx2r)

[4. Работа с системой 7](#_heading=h.xii9ojd8za7h)

[5. Аналитические инструменты 8](#_heading=h.esq55hpfauk0)

[6. Настройка и администрирование 8](#_heading=h.nilgvxm63uw3)

[7. Завершение работы 9](#_heading=h.4zck84w2xiof)

##

## **1. Введение**

Модуль «BI аналитика для стриминговых сервисов» — это инструмент для продуктовой аналитики на видео платформах, таких как Rutube. Он позволяет анализировать данные о пользователях, контенте и их взаимодействиях, обеспечивая глубокий анализ поведения аудитории, эффективности контента и работы алгоритмов.

Модуль интегрируется с модулем «Инфраструктура для сервиса», поисковыми и рекомендательными системами для стриминговых сервисов, получая данные напрямую. Это позволяет быстро адаптироваться к новым источникам информации и приступать к полноценной аналитике с минимальными задержками.

Сервис поддерживает анализ ключевых метрик, таких как время просмотра, CTR, retention rate, а также помогает выявлять тренды, оптимизировать контент и тестировать гипотезы для повышения качества рекомендаций и поиска.ий.

## **2. Запуск модуля**

### 2.1. Предварительные требования

Доступ к Yandex Cloud и соответствующим ресурсам (ClickHouse, DataSphere, Apache Superset).

Настроенная интеграция с модудем «Платформой данных»

Доступ к данным стримингового сервиса (Rutube и др.).

Административные права для настройки BI-инструментов.

### 2.2. Установка и запуск

Развертывание инфраструктуры с помощью Kubernetes (Yandex Managed Service for Kubernetes®).

Запуск контейнеров с ClickHouse, DataSphere и Apache Superset.

Настройка интеграции с источниками данных.

Авторизация в системе и предоставление прав доступа пользователям.

##

## **3. Аналитические инструменты**

#

# **3.1. Примеры доступных дашбордов**

# **3.1.1. Среднее время просмотра на пользователя**

Дашборд показывает среднее время просмотра на пользователя (WTPU, Watch Time Per User). Это метрика, которая измеряет среднее время, проведенное пользователем на платформе за определенный период (например, за день, неделю или месяц). Она рассчитывается как общее время просмотра контента, деленное на количество активных пользователей.



*Рис. 1 Дашборд WTPU (Watch Time Per User)*

Дашборд позволяет рассматривать данные в следующих разрезах:

* Период: Фильтрация данных по дате.
* Категория: Выбор одной или нескольких категорий видео.
* Группа: Группа в A/B тестировании.
* Флаг регистрации пользователя: Анализ в разрезе зарегистрированных / незарегистрированных пользователей.
* Свежесть видео: Время после публикации видео на платформе.
* Длина видео: Длина опубликованного видео.
* Устройство: Устройство, с которого происходит просмотр (web браузер, мобильный телефон, планшет или Smart TV).

# **3.1.2. Click-Through Rate (CTR) в рекомендациях**

Дашборд показывает долю кликов (CTR, Click-Through Rate) на определенной рекомендательной поверхности. Это метрика, которая показывает процент пользователей, которые кликнули на определенный элемент рекомендательной поверхности относительно общего числа пользователей, которым этот элемент был показан.



*Рис. 2 Дашборд CTR (Click-Through Rate)*

 Дашборд позволяет рассматривать данные в следующих разрезах:

* Период: Фильтрация данных по дате.
* Категория: Выбор одной или нескольких категорий видео.
* Группа: Группа в A/B тестировании.
* Флаг регистрации пользователя: Анализ в разрезе зарегистрированных / незарегистрированных пользователей.
* Свежесть видео: Время после публикации видео на платформе.
* Длина видео: Длина опубликованного видео.
* Устройство: Устройство, с которого происходит просмотр (web браузер, мобильный телефон, планшет или Smart TV).

# **3.1.3. Число показов видео по категориям**

Метрика "Число показов видео в рекомендациях по категориям" измеряет, сколько раз видео из определенной категории (например, "Сериалы", "Музыка", "Игры") было показано пользователям в рекомендательной ленте за определенный период.



*Рис. 3 Дашборд “Число показов видео в рекомендациях по категориям”*

Дашборд позволяет рассматривать данные в следующих разрезах:

* Период: Фильтрация данных по дате.
* Категория: Выбор одной или нескольких категорий видео.
* Группа: Группа в A/B тестировании.
* Флаг регистрации пользователя: Анализ в разрезе зарегистрированных / незарегистрированных пользователей.
* Свежесть видео: Время после публикации видео на платформе.
* Длина видео: Длина опубликованного видео.

# **3.1.4. Click-Through Rate (CTR) в поиске**

Дашборд показывает долю кликов по поисковым результатам (CTR, Click-Through Rate). Это метрика, которая показывает, какой процент пользователей кликнул на результаты поиска после ввода запроса.



*Рис. 4 Дашборд Click-Through Rate (CTR) в поиске*

Дашборд позволяет рассматривать данные в следующих разрезах:

* Период: Фильтрация данных по дате.
* Группа: Группа в A/B тестировании.
* Флаг регистрации пользователя: Анализ в разрезе зарегистрированных / незарегистрированных пользователей.
* Популярность запрос: Анализ в разрезе по частотности запроса (популярный, середина или хвост).

# **3.1.5. Доля сессий с поисковыми подсказками**

Метрика "Доля сессий с поисковыми подсказками" измеряет, какой процент пользовательских сессий на платформе включает использование поисковых подсказок (автодополнения). Она рассчитывается как отношение числа сессий, в которых пользователи взаимодействовали с подсказками, к общему числу сессий.



*Рис. 5 Дашборд "Доля сессий с поисковыми подсказками"*

Дашборд позволяет рассматривать данные в следующих разрезах:

* Период: Фильтрация данных по дате.
* Группа: Группа в A/B тестировании.

# 3.2. Средства Adhoc аналитики

# 3.2.1. Аналитика в ClickHouse

Агрегированные витрины данных, формируемые модулем «Платформы данных» (также фигурирует в документации как «Инфраструктура для сервиса»), доступным для adhoc аналитики при использовании модуля «BI аналитика для стриминговых сервисов».

ClickHouse — это высокопроизводительная колоночная СУБД, которая идеально подходит для adhoc аналитики по агрегированным витринам данных. Благодаря своей скорости обработки запросов и эффективной работе с большими объемами данных, ClickHouse позволяет аналитикам быстро выполнять сложные запросы к агрегированным данным без необходимости предварительной подготовки. Это особенно полезно для анализа исторических данных, построения отчетов и исследования трендов, где требуется гибкость и оперативность.

ClickHouse поддерживает стандартный SQL, что делает его удобным для аналитиков, а также предоставляет возможности для работы с материализованными представлениями и предварительно агрегированными данными. Это позволяет ускорить выполнение запросов к часто используемым витринам, сохраняя при этом возможность глубокого анализа.

# 3.2.2. Аналитика в Apache Superset

Apache Superset — это современный инструмент для визуализации и анализа данных, который отлично подходит для adhoc аналитики. Он предоставляет интуитивный интерфейс для создания интерактивных дашбордов и отчетов, что позволяет аналитикам и бизнес-пользователям быстро исследовать данные без необходимости написания сложных запросов. Superset поддерживает множество источников данных, включая ClickHouse, что делает его универсальным решением для работы с агрегированными витринами.

С помощью Apache Superset пользователи могут легко создавать визуализации, выполнять фильтрацию данных "на лету" и проводить глубокий анализ с помощью встроенного SQL-редактора. Это особенно полезно для adhoc аналитики, где требуется гибкость и оперативность.

# 3.2.3. Аналитика в Apache Spark

При необходимости пользователи имеют возможность работать с сырыми даными, используя Apache Spark – мощный фреймворк для распределенной обработки данных. Благодаря архитектуре, ориентированной на распределенные вычисления, Apache Spark может обрабатывать данные на сотнях узлов, что делает его идеальным решением для задач, требующих высокой производительности и масштабируемости.

Apache Spark поддерживает работу с сырыми данными в различных форматах (например, JSON, CSV, Parquet) и предоставляет инструменты для их очистки, трансформации и анализа. Это позволяет аналитикам и инженерам данных работать с необработанными данными напрямую, не требуя предварительной агрегации.

## **4. Работа с системой**

Авторизация в системе через корпоративные учетные записи.

Доступ к дашбордам с основными метриками и отчетами.

Работа с WTPU Dashboard:

Выберите период анализа (день, неделя, месяц).

Используйте фильтры (категория видео, возраст контента, длина видео, устройство просмотра).

Интерпретируйте результаты для оценки вовлеченности пользователей и улучшения стратегии контента.

Фильтрация данных по различным параметрам (категория контента, временные диапазоны, пользовательские сегменты).

Формирование пользовательских запросов через SQL-редактор для Adhoc-анализа.

Сохранение и экспорт данных в виде отчетов (CSV, JSON, Excel, PDF).

Доступ к дашбордам с основными метриками и отчетами.

Фильтрация данных по различным параметрам (категория контента, временные диапазоны, пользовательские сегменты).

Формирование пользовательских запросов через SQL-редактор для Adhoc-анализа.

Сохранение и экспорт данных в виде отчетов (CSV, JSON, Excel, PDF).

##

## **5. Аналитические инструменты**

WTPU (Watch Time Per User) Dashboard:

Позволяет фильтровать данные по дате, категории контента, длине видео и другим параметрам.

Используется для анализа эффективности видео и привлечения пользователей.

Интегрирован в Apache Superset и доступен для Adhoc-анализа.

ClickHouse – высокопроизводительная колоночная СУБД для Adhoc-аналитики.

Apache Superset – инструмент BI для визуализации и построения дашбордов.

DataSphere – среда для интерактивного анализа данных.

Apache Spark – распределенная обработка больших объемов данных.

## **6. Настройка и администрирование**

Настройка источников данных (интеграция с платформой данных).

Конфигурация Apache Superset (создание дашбордов, управление отчетами).

Оптимизация ClickHouse (индексация данных, настройка запросов).

Мониторинг работы системы (Prometheus, Grafana)

Обновление и масштабирование (управление контейнерами, авто-масштабирование в Kubernetes).

## **7. Завершение работы**

Завершение сеанса работы пользователя (выход из системы).

Остановка BI-системы при необходимости (выключение контейнеров с ClickHouse, Apache Superset и др.).

Архивирование данных и отчетов.

Мониторинг логов для выявления возможных ошибок.