



AMBER  
**DATA**

**ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ  
ХАРАКТЕРИСТИК ПРОГРАММНОГО  
ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИЯ,  
НЕОБХОДИМАЯ ДЛЯ ЭКСПЛУАТАЦИИ  
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ**

ПО «ВИ КРУТИЛКА» (AmberData)

Сведения о правообладателе

Общество с ограниченной ответственностью «Амбердата»  
127521, г. Москва, 12-й проезд Марьиной рощи, д.9, корп. 1, пом. VI, ком. 1  
ИНН/КПП: 7703096683/771701001  
тел.: (499) 938-43-78  
e-mail: info@amberdata.ru

# Оглавление

1	Термины и определения.....	2
2	Введение.....	3
3	Основные функции платформы.....	3
3.1	Описание обрабатываемых данных.....	3
3.2	Перечень основных функций.....	3
4	Основные задачи, выполняемые платформой.....	4
5	Структура системы.....	4
5.1	Перечень модулей.....	4
5.2	Архитектура DMP AmberData.....	5
5.2.1	Компоненты платформы.....	5
6	Описание компонентов платформы.....	5
6.1	Spark - Processing.....	5
6.2	Kafka - Processing.....	6
6.3	Upload-Processing.....	6
6.4	DMP-API.....	6
6.5	DMP-UI.....	6
6.6	AWG.....	7
6.7	ADCM.JS.....	7
7	Технологическая основа.....	7
8	Требования к техническим специалистам для разработки и эксплуатации платформы 8	
9	Информация, необходимая для установки Системы.....	9
9.1	Вход в систему.....	9

# 1 Термины и определения

DMP	Data Management Platform, программно-аппаратный комплекс для сбора, анализа и управления данными.
Таргетированная реклама	Реклама, направленная на определённый круг лиц
Деперсонафикация	Искажение персонифицированной информации таким образом, который делает невозможным определение её изначального значения
Обезличенные данные	Данные, хранимые в информационных системах в электронном виде, принадлежность которых конкретному субъекту персональных данных невозможно определить без дополнительной информации.
Дашборд	Доступное для Self-Service аккаунтов подмножество функций из модулей отчеты и мониторинг.
Коды (Тэги и Пиксели)	Работа с устанавливаемыми на площадки медиа-партнеров кодами, обеспечивающими сбор сырых данных о профилях пользователей (например, статистики посещений) и технологические задачи интеграции с платформами медиа-партнеров (например, синхронизацию id профилей пользователей, он-лайн обмен данными).
Таксономии	Справочники сегментов, в терминах которых описываются профили пользователей и аудитории. Управление правилами сегментации, взаимосвязями сегментов.
Аудитории	Подготовка данных о профилях пользователей для медиа-партнеров в роли потребителей данных, управление выгрузкой и лицензиями сегментов во внешние системы медиа-партнеров (DSP, Exchange и т.п.).
Cookie	Небольшой фрагмент данных, отправленный веб-сервером и хранимый на компьютере пользователя
IP-адрес	Уникальный сетевой адрес узла в компьютерной сети
API	Набор программных методов для интеграции с внешними системами
REST, REST API, RESTful API	(англ. Representational State Transfer Application Programming Interface) – стандартизированный метод взаимодействия приложений в рамках архитектуры REST
Мониторинг	Сбор и накопление бизнес-метрик, расчет целевых показателей и анализ отклонений.
Процессинг	обработка данных о профилях пользователей, полученных в рамках интеграции с системами медиа-партнеров в роли поставщиков. Модуль обеспечивает регулярные процессы валидации полученных данных, построения профилей пользователей на основе сырых данных (например, о клик-стрим), нормализации, актуализации и агрегации данных о профилях пользователей, а так же автоматизированное построение Look-a-like моделей.

## 2 Введение

ПО «ВИ КРУТИЛКА» (публичное фирменное наименование – AmberData, DMP AmberData) – DMP-система, многофункциональная технологическая платформа по сбору, обработке, хранению и управлению обезличенными данными об интернет-пользователях. Предметом сбора и анализа в DMP AmberData становятся профили пользователей.

Профиль пользователя – сгруппированные обезличенные данные по конкретному пользователю, которые включают в себя социально-демографические данные (например, пол, возраст и т.д.) и данные по интересам пользователя (например, увлекается Авто, Туризмом и т.д).

Данные по интересам сформированы в вектор интересов пользователя – матрица с весом величины интереса к той или иной теме.

## 3 Основные функции платформы

### 3.1 Описание обрабатываемых данных

Существуют 3 основных типа данных о профилях пользователей:

- First-Party Data - собственные данные, собранные и используемые на своих площадках.
- Second-Party Data - собственные данные, которые вы используете, но их собрали и подготовили для использования сторонние системы.
- Third-Party Data - данные третьих лиц, которые можно использовать при нехватке своих - для увеличения охвата, или для уточнения собственных данных.

Технически для сбора данных на страницах сайтов и/или в рекламных сообщениях размещают специальные коды тэгов, которые обеспечивают идентификацию профилей пользователей с помощью уникального идентификатора. В дальнейшем происходит синхронизация этих уникальных идентификаторов профилей с уникальными идентификаторами наших партнеров и получение имеющихся у них данных о профилях этих пользователей – таким образом, формируются Third party Data.

Данные партнеров о профилях пользователей передаются специальными файлами в модули DMP, где они обрабатываются, приводятся к правильному виду и подготавливаются для дальнейшего использования в системах потребителей данных.

Для сбора данных в нашей DMP используется партнерские договоры с крупнейшими поставщиками данных в сети Интернет и с провайдерами.

Данные сгруппированы в более чем 2000 различных сегментов (социально-демографические данные, данные об интересах пользователей, прошлые покупки, In-market), собранных в категоризатор – Таксономию DMP.

### 3.2 Перечень основных функций

DMP AmberData выполняет следующие основные функции:

- Агрегация данных;
- Интеграция данных;
- Складирование данных;
- Сегментация данных;
- Создание аудиторий;
- Анализ аудиторий;
- Передача данных во внутренние системы Заказчика и внешние рекламные системы (Google, MyTarget, Яндекс, Adriver и др.);
- Статистика и мониторинг.

## 4 Основные задачи, выполняемые платформой

Платформа AmberData выполняет следующие задачи:

- Сегментация и использование собственных данных (данные заказчиков) о пользователях;
- Анализ и расширение собственных аудиторий;
- Доступ к недостающим аудиторным знаниям, обогащение собственных данных;
- Связь онлайн и офлайн данных;
- Предоставление данных для аудиторного таргетинга в маркетинговых активностях;
- Предоставление данных для динамической адаптации креативных сообщений;
- Предоставление данных для динамической адаптации контента сайта;
- Предоставление данных для допродажи сопутствующих услуг (cross-sell) и перевода на более дорогой тариф (up-sell);
- Предоставление данных для скоринга;
- Формирование кросс-платформенных пользовательских профилей и монетизация знаний;
- Сбор, агрегация и хранение информации о рекламных контактах с целевой аудиторией.

## 5 Структура системы

### 5.1 Перечень модулей

Для выполнения перечисленных выше функций была разработана модульная архитектура продукта DMP AmberData.

DMP AmberData включает в себя модули:

- Коллектор-модуль;
- Модуль сегментации;
- Модуль аналитики;

- Модуль выгрузки сегментов.

## 5.2 Архитектура DMP AmberData

### 5.2.1 Компоненты платформы

- DMP UI — пользовательский интерфейс, включающий модули:
  - Дашборд (Страница, со сводной информация по объектам, входящим в состав проектов);
  - Управление кодами и пикселями через UI DMP в контуре AmberData;
  - Управление таксономиями, категориями и логикой построения сегментов;
  - Управление выгрузкой («лицензированием») сегментов во внешние системы;
  - Интерфейс администрирования.
- DMP-API – Провайдер данных о сущностях, используемых в процессе работы комплексом систем DMP, предоставляющий программный интерфейс для:
  - DMP-UI;
  - Модулей DMP Processing;
  - Трекинг модуля AWG;
  - Внешних систем.
- AWG – high-load сервис для realtime сбора данных и синхронизации идентификаторов
- Библиотека кодов, размещаемых на веб-площадках или встраиваемых во внешние системы. Библиотека доступна в UI DMP в контуре AmberData, также может быть настроена для DMP в контуре СП.
- Spark - Processing – модуль, реализующих сегментацию и агрегирование профилей пользователей на основе правил любой сложности, получаемых от DMP-API, а также подготовку данных, для модуля процессинга «Upload».
- Upload - Processing – модуль выгрузки сегментов во внешние системы.
- Kafka - Processing – модуль для процессинга, мониторинга и валидации данных клик-стрима в Kafka.

## 6 Описание компонентов платформы

### 6.1 Spark - Processing

Компонент, реализующий сегментацию/агрегирование пользователей на основе правил, получаемых от DMP-API, а также подготовку данных, для модуля процессинга «Upload»

Поставляется в виде скомпилированных Java-приложений и Airflow DAG сценария, отвечающим за порядок работы и запуска компонента в среде Airflow.

## 6.2 Kafka - Processing

Модуль для процессинга и валидации данных стрима в Kafka, поступающие от AWG. Поставляется в виде скомпилированных Java-приложений.

К модулю прилагается документация, описывающая требования для запуска и информацию по конфигурации модуля.

## 6.3 Upload-Processing

Модуль выгрузки сегментов в рекламные партнерские системы.

Поставляется в виде скомпилированных Java-приложений и Airflow DAG сценария, отвечающим за порядок работы и запуска модуля в среде Airflow.

К модулю прилагается документация, описывающая требования для запуска и информацию по конфигурации модуля.

## 6.4 DMP-API

Провайдер данных о сущностях, используемых в процессе работы комплексом систем DMP, предоставляющий программный интерфейс для:

1. Интерфейс DMP-UI;
2. Компоненты Spark Processing;
3. AWG;

Поставляется в виде Docker image.

Для работы API необходима база PostgreSQL.

Прилагается документация, описывающая стек используемых технологий, требования для запуска и информацию по конфигурации модуля, спецификацию авторизации, описания точек API для получения данных, а также схему его взаимодействия с другими компонентами и схему модели данных приложения.

## 6.5 DMP-UI

Приложение, реализующее интерфейс для взаимодействия пользователя с DMP-API.

Поставляется в виде Docker image с обфусцированным javascript/typescript кодом приложения.

К модулю прилагается инструкция пользователя.

## 6.6 AWG

Приложение для разметки пользователей с помощью Cookies, а также логирования click-stream запросов.

Поставляется в виде Docker image со скомпилированным кодом приложения.

## 6.7 ADCM.JS

JS библиотека, для трекинга пользователей, передающая данные в AWG.

Поставляется в виде Docker image с обфусцированным javascript/typescript кодом приложения.

К модулю прилагается инструкция пользователя, объясняющая, как js код устанавливается на клиентских площадках и какие механизмы сбора данных он имеет.

## 7 Технологическая основа

Платформа AmberData DMP построена на продуктах и технологиях с открытым исходным кодом, распространяемых под лицензией Apache 2.0.

При разработке модулей и компонентов решения используются следующие языки программирования:

- Java;
- Scala;
- JavaScript;
- Python;
- Clojure;
- SQL;
- Bash.

Основные компоненты, отвечающие за обработку больших объемов данных, реализованы с использованием языка Java 8, Scala, Clojure.

Загрузка, обработка и хранение данных осуществляется на платформах Apache Hadoop, содержащая, как минимум, следующие компоненты:

- HDFS;
- YARN;
- Hive;
- Spark v > 3;
- Airflow;



- PrestoDB / PrestoSQL;
- Jupyter.

Оркестрация приложений в докере реализована на базе Kubernetes.

Все собираемые и хранимые DMP данные, предоставляются и принимаются в обезличенной форме («Обезличенные данные»). Невозможно однозначно идентифицировать субъектов данных, полученных в результате обезличивания без применения дополнительной информации.

Обеспечена защита данных на следующих уровнях:

- Application (OAuth2),
- HTTPS для доступа из вне к DMP-UI.

## **8 Требования к техническим специалистам для разработки и эксплуатации платформы**

Требования к программистам и системным администраторам:

1) Знание технологий:

- Kubernetes;
- Apache Hadoop;
- Spark;
- Hive;
- PrestoDB;
- Airflow;
- Docker;
- Nginx;
- PostgreSQL;
- Clickhouse.

2) Знание языков программирования:

- Python;
- Bash;
- SQL.

3) Отличное знание Linux.

Требования к специалисту техподдержки сервиса:

- Знание Javascript/html на высоком уровне,
- Знание технических особенностей браузеров и механизмов работы с куками,
- Уверенное владение Python 3.6 - 3.8.

## 9 Информация, необходимая для установки Системы

Клиентская часть приложения не требует проведения установки. Работа осуществляется посредством web-браузера.

### 9.1 Вход в систему

Для осуществления входа в проект необходимо выполнить следующие действия:

1. Запустить браузер.
2. Указать следующий URL <https://my.amberdata.ru/> в адресной строке браузера.
3. В открывшемся окне авторизации (1) ввести логин и пароль, предоставленные администратором системы.
4. Нажать кнопку «**Войти**»



Рисунок 1 – Окно авторизации

В случае успешной аутентификации пользователю открывается экранная форма «**Последние добавленные коды**» и «**Последние добавленные категории**».