

# Manzana CDP&BI. Инструкция по установке

## Оглавление

Оглавление .....	2
1. Глоссарий .....	3
2. Инструкция по установке .....	4

## 1. Глоссарий

Термин	Определение
Apache AirFlow	Открытое программное обеспечение для создания, выполнения, мониторинга и оркестровки потоков операций по обработке данных.
Apache Kafka	Распределенный программный брокер сообщений с открытым исходным кодом, разрабатываемый в рамках фонда Apache на языках Java и Scala.
Apache Superset	Открытое программное обеспечение для исследования и визуализации данных, ориентированное на большие данные
BI	Инструмент визуализации данных
CDP	Продукт, в котором хранится информация о клиентах, чеках, бонусах, купонах и прочая информация. Источник данных для Manzana BI
ClickHouse	Аналитическая СУБД
Manzana CDP&BI	Программное обеспечение, предоставляющее пользователям удобный доступ к аналитическому контенту для визуализации данных и интерактивного анализа покупательского поведения и маркетинговых акций.
Yandex Data Transfer (PaaS)	Сервис, предназначенный для логического переноса данных между источниками и приемниками. В качестве источников и приемников данных могут выступать СУБД, объектные хранилища или брокеры сообщений.
Yandex Object Storage (S3)	Масштабируемое облачное хранилище
СУБД	Система управления базами данных

## 2. Инструкция по установке

Решение Manzana CDP&BI - программное обеспечение, предоставляющее пользователям удобный доступ к аналитическому контенту для визуализации данных и интерактивного анализа покупательского поведения и маркетинговых акций.

В качестве СУБД решение использует аналитическую СУБД ClickHouse. Загрузка данных происходит батчевым либо потоковым способом. Целевая архитектура предполагает развертывание решения на PaaS серверах в облаке либо на виртуальных машинах операционной системы Linux.

Ключевые компоненты решения Manzana CDP&BI:

1. Yandex ClickHouse
2. Apache Superset
3. Apache AirFlow
4. Apache Kafka
5. Yandex Data Transfer (PaaS)
6. Yandex Object Storage (S3)

Порядок развертывания компонент, следующий:

1. Установка Yandex ClickHouse
2. Развертывание структуры Manzana CDP
3. Установка и настройка вспомогательных компонентов загрузки данных (AirFlow, Kafka, PaaS, S3)
4. Установка и настройка компонентов визуализации данных