Описание системы

**Релиз 1.0.0.0**

Лист изменений

|  |  |  |  |
| --- | --- | --- | --- |
| **Дата** | **Версия** | **Описание** | **Автор** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**-**

# Общие сведения

## Наименование

Foros (рабочее название)

## Назначение системы/продукта

1. сбор обезличенной информации о пользователях и сегментирование на основании их поведения в интернете;
2. показ наиболее эффективной и рентабельной рекламы на основе собранной информации и условий заданных рекламодателем;
3. Гибкое конфигурирование рекламных сущностей через web интерфейс, удобное отображение детальных статистик
4. выгрузка собранных сегментов аудитории и их анализ.

## Класс системы/продукта

Business Operational.

## Статус

В промышленной эксплуатации.

# Определения и сокращения

|  |  |
| --- | --- |
| *Термин* | *Описание* |
| **SSP (Sell Side platform)** | технологическая платформа, которая позволяет вебмастерам управлять своими площадками и максимизировать доходы в сфере медиа. |
| **SSP\_ID (SSP\_UID)** | Уникальный идентификатор пользователя, используемый SSP (в частности могут быть Unified ID, TDID, IDFA) |
| **RTB (Real Time Bidding)** | технология в индустрии [онлайн](https://ru.wikipedia.org/wiki/Онлайн)-рекламы, которая представляет собой [аукцион](https://ru.wikipedia.org/wiki/Аукцион) рекламных объявлений в реальном времени. |
| **KPI (Key Performance Indicators)** | ключевые показатели эффективности. |
| **CR (Conversion Rate)** | соотношение посетителей сайта, которые совершили "нужное" действие (кликнули по ссылке, проголосовали, купили) к общему числу посетителей сайта. |
| **CTR (Click-Through Rate)** | показатель кликабельности, определяется как отношение числа кликов на баннер или рекламное объявление к числу показов, измеряется в процентах. |
| **DSP (Demand Side Platform)** | технологическая система организации аукциона для рекламодателей, которая торгуется с [SSP](https://ru.wikipedia.org/w/index.php?title=SSP&action=edit&redlink=1) (платформами для RTB-торгов со стороны площадок), управляет несколькими рекламными сетями ([Ad Networks](https://ru.wikipedia.org/w/index.php?title=Ad_Networks&action=edit&redlink=1)) и рекламными биржами ([Ad Exchanges](https://ru.wikipedia.org/w/index.php?title=Ad_Exchanges&action=edit&redlink=1)), обменивается прочими данными в интересах рекламодателя в цифровой экосистеме [RTB](https://ru.wikipedia.org/wiki/Торг_в_реальном_времени). |
| **UID** | Уникальный идентификатор пользователя сети интернет, используемый в системе Foros |
| **URL (Uniform Resource Locator)** | единообразный указатель ресурса в сети Интернет. |
| **Cookie** | набор данных, который сервер посещенного сайта записывает в браузер пользователя |
| **Профилирование** | Сбор характеристик поведения пользователей сети интернет |
| **Click Stream** | Информация о песещении страниц пользователями сети Интернет |
| **CSP (Click Stream Provider)** | Поставщик информации о посещении страниц (SSP, web master, DPI и т п) |

# Архитектура системы

## Описание сущностей системы[[1]](#footnote-2)Untitled 1:Users:jurij_kuznecov:Downloads:Entity+Diagram.png

|  |  |
| --- | --- |
| *Сущность (синонимы)* | *Описание, свойства* |
| **Счет** | Метод учета движения финансовых средств: *рекламодатель, номер счета, остаток на счете и т.д.* |
| **Рекламодатель** | Поставщик рекламы: *имя, рекламная компания и т.д.* |
| **Рекламная Кампания** | Совокупность рекламных мероприятий направленных на достижение маркетинговых целей рекламодателя: *название, рекламодатель, баннер, критерии показа и т.д.* |
| **Баннер** | Графическое изображение рекламного характера: *шаблон, ресурсы и т.д.* |
| **Рекламный показ** | Предъявление рекламного материала пользователю: *баннер, баннерное размещение, состояние, цена, предсказанный CTR, предсказанный Conversion Rate и т.д.* |
| **Пользователь сети Интернет** | Посетитель веб-сайтов: *ID, профиль пользователя, идентификатор в сторонних системах и т.д.* |
| **Рекламное Размещение (Ad Tag)** | Место для размещения рекламного материала на веб-сайте: *публикатор рекламы, URL, width, height, исключительные категории, тип аукциона* |
| **Профиль пользователя** | Набор данных, характеризующий пользователя. (уникальный код, исторические данные о поведении в сети интернет): *сегмент интересов, маркер аномального поведения, последнее время активности, маркеры рекламных показов и т.д.* |
| **Публикатор рекламы (Web издатель)** | Совокупность баннерных размещений, объединенных единым признаком: *имя; комиссия; тип комиссии; агентство; тип аукциона и т.д.* |
| **Сегмент** | *Базовый критерий, характеризующий пользователя или рекламный запрос, используемый для настройки таргетированной рекламы и оценки объема целевой аудитории. Разделяется на виды: поведенческий сегмент интересов, гео сегмент, событийный сегмент,* |
| **Поведенческий сегмент** | Базовый характеристика поведения пользователей привязываемая на основе посещения сайтов: *правила привязывания к пользователю (кол-во срабатываний ключевых слов или url в заданный период времени), списки ключевых слов.* |
| **Гео сегмент** | Сегмент характеризующий запрос (рекламный, cookie sync и т п) на основе IP адреса с точностью до населенного пункта. |
| **Событийный сегмент** | Сегмент привязываемый к пользователю на основе событий : клик на определненной рекламе, срабатывание пикселя, полный просмотр видео рекламы и т п. |
| **Составной сегмент** | Логическое выражение с использованием сущностей “сегмент” и логических операций над ними. Сущность может использоваться для настройки таргетированной рекламы или оценки объема (уники в день, месяц и всего) и свойств целевой аудитории. |

## Пользовательские функции и профили доступа

|  |  |
| --- | --- |
| *ИТ услуга, роль*  *(профиль доступа)* | *Сервисы* |
| **Рекламодатель (Advertiser or Agency)** | * настройка рекламной кампании, учитывая бюджет, требования к содержанию креативов и желаемую целевую аудиторию; * запуск рекламной кампании; * оптимизация рекламной кампании; * пополнение бюджета. |
| **Аккаунт менеджер (AdOperations Manager, AdOp)** | * заведение аккаунта рекламодателя в системе; * настройка рекламных кампаний согласно требованиям рекламодателя; * запуск рекламных кампаний; * оптимизация рекламных кампаний с помощью корректорвки условий показа. |
| **Исследователь поведения пользователей интернета**  **(Channel Manager)** | * Создание сегментов (на основе триггеров: поисковые запросы; ключевые слова на страницах; посещенные url; ключевые слова из url, составные сегменты и т п); * Оптимизация сегментов: * удаление/добавление триггеров; * изменение параметров учета триггеров (время и частотность).   и т п |
| **Web-издатель**  **(Publisher)** | * устанавливает рекламное размещение; * получает оговоренную комиссию в случае показа. |

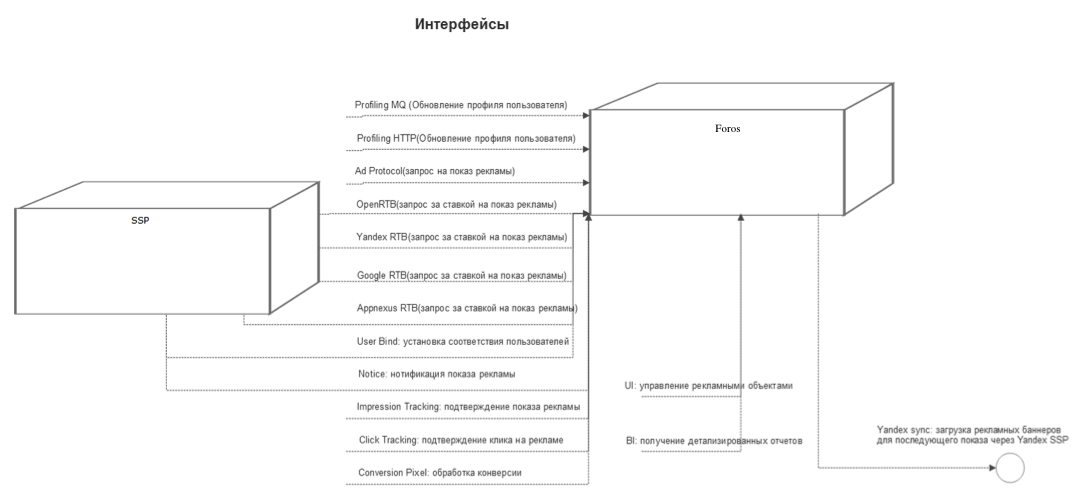
## Интерфейсы

### Реализуемые системой интерфейсы

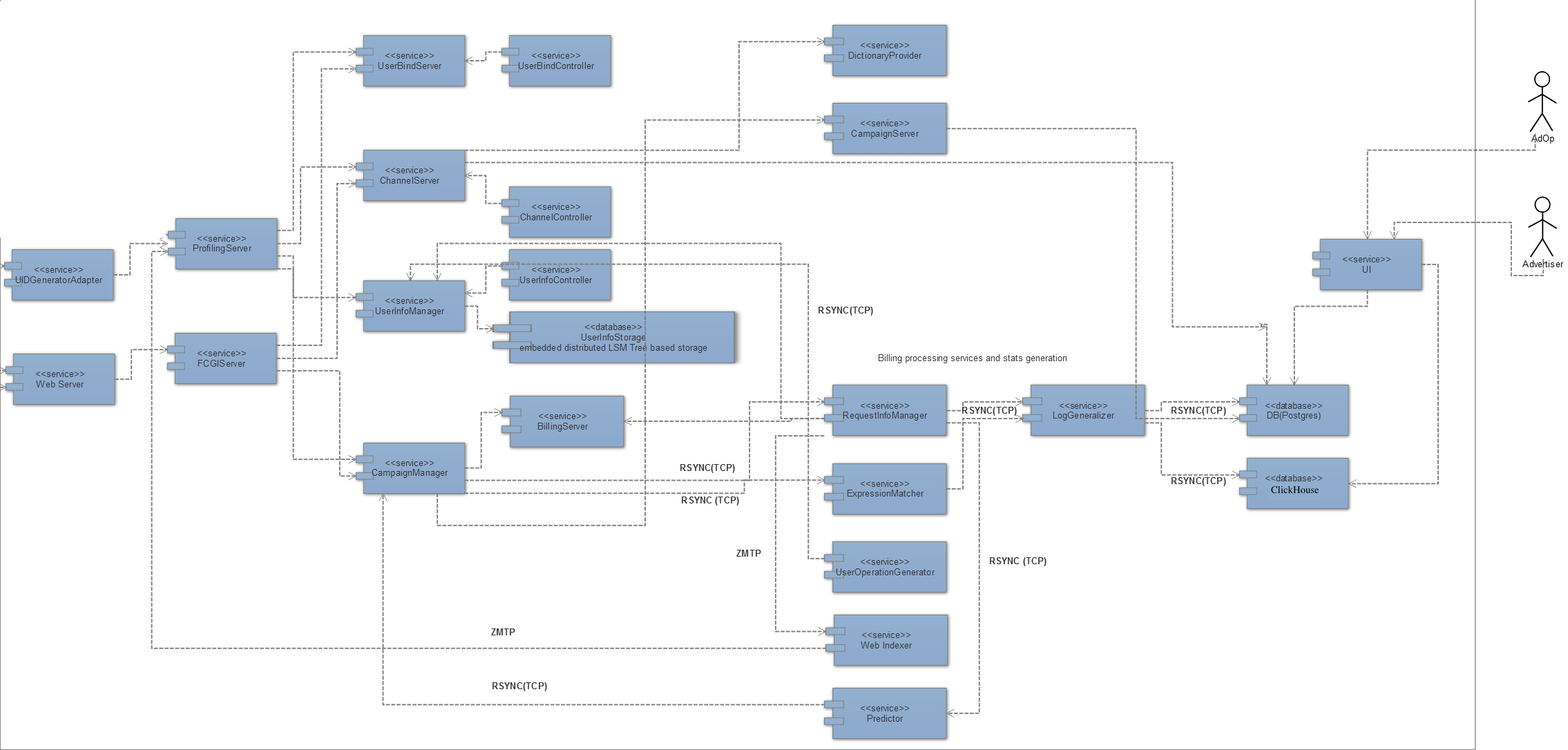
|  |  |  |
| --- | --- | --- |
| *Тип* | *Интерфейс* | *Описание* |
| http/soap | Profiling MQ | Обновление профиля пользователя сети интернет |
| http/soap | Profiling HTTP | Обновление профиля пользователя сети интернет |
| http/soap | Ad Protocol | Запрос на показ рекламы |
| http/soap | OpenRTB | Запрос за ставкой на показ рекламы |
| http/soap | Yandex RTB | Запрос за ставкой на показ рекламы |
| http/soap | Google RTB | Запрос за ставкой на показ рекламы |
| http/soap | Appnexus RTB | Запрос за ставкой на показ рекламы |
| http/soap | User Bind | Установка соответствия SSP UID <=> UID |
| http/soap | Notice | Нотификация показа рекламы (RTB notice) |
| http/soap | Impression Tracking | Подтверждение показа рекламы |
| http/soap | Click Tracking | Обработка клика на рекламе |
| http/soap | Conversion Pixel | Обработка конверсии |
| http/soap | UI | Обеспечение распределённого доступа к управлению рекламными объектами (учетными записями, рекламными кампаниями, баннерами, конверсиями, биллинг) и построение компонентных отчётов по рекламным показам и профилированию пользователей сети интернет |
| http/soap | ВI | Построение более детальных отчётов по рекламным показам и профилированию пользователей сети интернет |

### Используемые системой интерфейсы

|  |  |  |
| --- | --- | --- |
| *Тип* | *Интерфейс* | *Описание* |
| http/soap | Yandex sync | Синхронизация креативов в систему yandex |



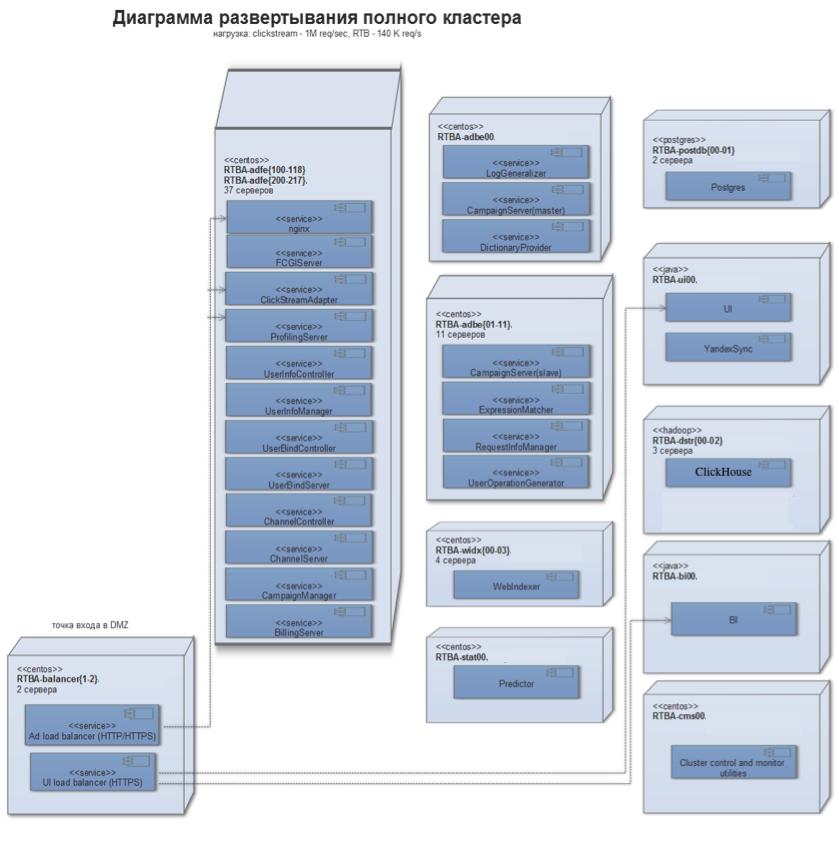
## Модули системы и требования по их развертыванию



Модули описаны в разделе 4.4.2

Т. к. на одном хосте могут размещаться несколько сервисов, требующих соединения между собой : **на всех хостах на которых разворачивается кластер должен быть сконфигурирован loopback интерфейс**.

### Диаграмма развертывания



Сокращение {100-118} подразумевает группу хостов с последовательной нумерацией.

К примеру, «RTBA-adfe{100-118}» включает в себя хосты с следующими именами: «RTBA-adfe100», «RTBA-adfe101», … , «RTBA-adfe118».

Данная конфигурация использует 64 сервера в подсети.

RTBA-adfe : группа хостов, отвечающая за runtime обработку рекламных и профилировочных запросов (ADserver FrontEnd).

RTBA-adbe : группа хостов, отвечающая за постобработку результатов показа рекламы и профилировки (ADserver BackEnd).

RTBA-widx : группа хостов, отвечающая за загрузку содержимого веб-страниц (Web InDeXer).

RTBA-dstr : группа хостов с базой данных для статистик большого размера (Data SToRage)

RTBA-cms : хост для управления кластером (Configuration Management System)

### Модули системы

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Тип компонента* | *Название компонента* | *Назначение и*  *краткое описание компонента* | *Требования к общесистемному ПО и оборудованию* |  |
|  | Predictor(STAT) | Генерация модели предсказания CTR, CR на основе показов | RAM >= 256 Gb,  2 x SATA >= 1.2 Tb  ОС: CentOS 6.9  ПО: perl, python  Библиотеки: sklearn, xgboost, catboost, vanga |  |
| WebIndexer(WI) | Выгрузка html страниц из сети интернет, выделение значимого текста страницы | RAM >= 128 Gb,  SSD >= 700 Gb  ОС: CentOS 6.9  Библиотеки: boost, llibcurl, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ |  |
| CampaignServer (BE) | Загрузка рекламных сущностей из БД, предоставление описания сущностей остальным компонентам кластера | RAM >= 256 Gb, SSD >= 800 Gb,  SATA >= 1.4 Tb или SSD >= 800 Gb  ОС: CentOS 6.9  Библиотеки: ACE-TAO, httpd, apr, boost, GeoIP, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ, xgboost. |  |
| DictionaryProvider(BE) | Предоставление словоформ по слову на основе словарей |  |
| ExpressionMatcher(BE) | Сбор и сведение статистик по сегментам |  |
| RequestInfoManager(BE) | Контроль цикла выполнения показа рекламы. Сведение показа с подтверждением, кликом и конверсией. Исключение повторных и ложных действий. Выделение fraud пользователей. |  |
| UserOperationGenerator(BE) | Синхронизация сегментов заданных в виде списков UID с профилями пользователей |  |
| LogGeneralizer(BE) | Аггрегирование статистик перед записью в postgres, hadoop |  |
| BillingServer(FE) | Контроль бюджетных правил для рекламных сущностей при выборе рекламы | RAM >= 128 Gb, SSD >= 600 Gb  ОС: CentOS 6.9  Библиотеки: ACE-TAO, httpd, apr, boost, GeoIP, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ, xgboost, vanga, catboost. |  |
| CampaignManager(FE) | Выбор рекламы на основе параметров запроса, сегментов работающих в контексте запроса и предсказанных CTR, CR; Логгирование выбора, его подтверждения, клика, конверсии |  |
| ChannelServer(FE) | Загрузка правил отображения url, текста html страницы на сегменты (trigger matching). Реализация этого отображения на этапе выполнения рекламного/профилировочного запроса |  |
| ChannelController(FE) | Контроль состояния ChannelServer сервисов, распределение ресурсов между ними |  |
| UserBindServer(FE) | Отображение SSP\_ID на UID, распределенное хранение данного отображения, резервное копирование, сохранение/чтение отображения на носитель/с носителя |  |
| UserBindController(FE) | Контроль состояния UserBindServer сервисов,  распределение ресурсов между ними |  |
| UserInfoManager(FE) | Хранение меток сегментов привязанных к времени для всех пользователей на носителе; Загрузка исторических правил срабатывания сегментов из БД; Обновление меток и предоставление списка сегментов срабатывающих для пользователя при обслуживании рекламного/профилировочного запроса |  |
| UserInfoController(FE) | Контроль состояния UserInfoManager сервисов,  распределение ресурсов между ними |  |
| ProfilingServer(FE) | Делегирование обработки профилировочного запроса полученного по протоколу ZeroMQ в компоненты реализующие логику: UserBindServer,ChannelServer,UserInfoManager,CampaignManager |  |
| ClickStreamAdapter(FE) | Компонент реализуется в рамках интеграции с ВК.  Основные функции: конвертирование запроса из UID Generator protobuf формата в ZeroMQ формат и распределение (балансировка) запросов между ProfilingServer сервисами |  |
| Ad load balancer | Балансировщик HTTP/HTTPS траффика для nginx, UI, BI сервисов, реализуется ПО haproxy, keepalived | RAM >= 16 Gb  ОС: CentOS 6.9  ПО: haproxy, keepalived |  |
| UI load balancer | Балансировщик HTTP/HTTPS траффика для UI, BI сервисов, реализуется ПО haproxy, keepalived | RAM >= 16 Gb  ОС: CentOS 6.9  ПО: haproxy, keepalived |  |
| Cluster control and monitor utilities | Набор утилит для конроля состояния кластера и управления им | ОС: CentOS 6.9 |  |
| YandexSync | Синхронизация креативов в систему Yandex SSP | RAM >= 16 Gb  ОС: CentOS 6.9 |  |
| Веб-приложения | nginx | Nginx сервер делегирующий обработку запроса в FCGIServer сервис по протоколу FCGI | RAM >= 16 Gb  ОС: CentOS 6.9  Библиотеки: ACE-TAO, httpd, apr, boost, GeoIP, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ, xgboost. |  |
| FCGIServer | Чтение разновидностей RTB запросов (OpenRTB, appnexus, ..). Делегирование обработки рекламного/профилировочного запроса в компоненты реализующие логику : UserBindServer,ChannelServer,UserInfoManager,CampaignManager |  |
| UI | Описание/редактирование и управление рекламными сущностями, получение отчетов | RAM >= 64 Gb,  диски >= 256 Gb  ОС: CentOS 6.9 |  |
| BI | Работа со сложными отчетами | RAM >= 128 Gb,  диски >= 256 Gb  ОС: CentOS 6.9 |  |
| Базы данных | PostgreSQL | Хранение рекламных объектов и предоставление статистик для OLTP | RAM >= 128 Gb, SAS диски >= 3Tb  ОС: CentOS 7.3 |  |
|  | Hadoop | Хранение больших видов статистики для OLAP | RAM >= 128 Gb, диски >= 3 Tb  ОС: CentOS 6.9 |  |
|  | Impala | Хранение больших видов статистики для OLAP | RAM >= 128 Gb, диски >= 3 Tb  ОС: CentOS 6.9 |  |

### Размещение системы по элементам ИТ инфраструктуры

|  |  |  |
| --- | --- | --- |
| *Элемент ИТ инфраструктуры* | *Тип и название модуля* | *Описание деталей установки и конфигурации* |
| RTBA-adfe{100-118} | <<service>> nginx, <<service>> FCGIServer, <<service>> ProfilingServer, <<service>> UserInfoController, <<service>> UserInfoManager, <<service>> UserBindController, <<service>> UserBindServer, <<service>> ChannelController, <<service>> ChannelServer, <<service>> CampaignManager, <<service>> BillingServer | yum install foros-config-server-central |
| RTBA-adfe{200-203} | <<service>> nginx, <<service>> FCGIServer, <<service>> ProfilingServer, <<service>> UserInfoController, <<service>> UserInfoManager, <<service>> UserBindController, <<service>> UserBindServer, <<service>> ChannelController, <<service>> ChannelServer, <<service>> CampaignManager, <<service>> BillingServer | yum install foros-config-server-central |
| RTBA-adfe{204-211} | <<service>> nginx, <<service>> FCGIServer, <<service>> ProfilingServer, <<service>> UserInfoController, <<service>> UserInfoManager, <<service>> UserBindController, <<service>> UserBindServer, <<service>> ChannelController, <<service>> ChannelServer, <<service>> CampaignManager, <<service>> BillingServer | yum install foros-config-server-central |
| RTBA-adfe{212-213} | <<service>> nginx, <<service>> FCGIServer, <<service>> ProfilingServer, <<service>> UserInfoController, <<service>> UserInfoManager, <<service>> UserBindController, <<service>> UserBindServer, <<service>> ChannelController, <<service>> ChannelServer, <<service>> CampaignManager, <<service>> BillingServer | yum install foros-config-server-central |
| RTBA-adfe214 | <<service>> nginx, <<service>> FCGIServer, <<service>> ProfilingServer, <<service>> UserInfoController, <<service>> UserInfoManager, <<service>> UserBindController, <<service>> UserBindServer, <<service>> ChannelController, <<service>> ChannelServer, <<service>> CampaignManager, <<service>> BillingServer | yum install foros-config-server-central |
| RTBA-adfe{215-217} | <<service>> nginx, <<service>> FCGIServer, <<service>> ClickStreamAdapter, <<service>> ProfilingServer, <<service>> UserInfoController, <<service>> UserInfoManager, <<service>> UserBindController, <<service>> UserBindServer, <<service>> ChannelController, <<service>> ChannelServer, <<service>> CampaignManager, <<service>> BillingServer | yum install foros-config-server-central |
| RTBA-adbe00 | <<service>> LogGeneralizer, <<service>> CampaignServer(master) | yum install foros-config-server-central |
| RTBA-adbe{01-11} | <<service>> DictionaryProvider, <<service>> CampaignServer(slave), <<service>> ExpressionMatcher, <<service>> RequestInfoManager, <<service>> UserOperationGenerator | yum install foros-config-server-central |
| RTBA-widx{00-03} | <<service>> WebIndexer | yum install foros-config-webindex |
| RTBA-stat00 | <<service>> Predictor | yum install foros-config-server-central |
| RTBA-postdb{00-01} | <<postgres>> Postgres | Установить пакеты на postdb00(master) :  yum install foros-pgadm  yum install foros-pgadm  yum install foros-config-pgadm-bp  yum install foros-pgdb  yum install foros-config-pgdb-bp  yum install foros-config-datastore-clickhouse-load  yum install foros-datastore-clickhouse-load  yum install foros-config-datastore-mgr  Установить пакеты на postdb01(standby) :  yum install foros-pgadm  yum install foros-pgadm  yum install foros-config-pgadm-bp  yum install foros-config-pgadm-standby  yum install foros-pgdb  yum install foros-config-pgdb-bp  Разархивировать следующие файлы в указанные директории под пользователем uiuser:  pgdata.tar.gz /u01/foros/pgadm/var/pgdata  stat.tar.gz /opt/foros/pgadm/var/tablespaces/stat  u03\_tablespace.tar.gz /u03/tablespace  u02\_tablespace.tar.gz /u02/tablespace  00000001000027\*.bz2 /opt/foros |
| RTBA-ui00 | <<java>> UI, <<java>> YandexSync | yum install foros-config-ui-bp |
| RTBA-dstr{00-02} | <<hadoop>> Hadoop, <<hadoop>> Impala | 1. Развернуть кластер: ansible-playbook -i hosts site.yaml 2. Установка пакетов:    1. yum install foros-datastore-adm    2. yum install foros-config-datastore-adm    3. yum install foros-config-datastore |
| RTBA-bi00 | <<java>> BI | 1. Разархивировать biserver-ce.tar в /u01/foros/bi/biserver-ce 2. Установить пакеты:   yum install foros-biadm  yum install foros-config-biadm  yum install foros-bi-mondrian |
| RTBA-cms00 | <<service>> Cluster control and monitor utilities | yum install foros-config-server-central-mgr |
| RTBA-balancer{1-2} | <<service>> Ad load balancer,  <<service>> UI load balancer | yum install foros-loadbalancer  service keepalived start |

Старт кластера осуществляется после выполнения инструкций “Описание деталей установки и конфигурации”.

Все необходимые библиотеки и ПО не указанные в инструкциях устанавливаются по RPM зависимостям.

Для старта кластера используются следущие команды на сервере cms00:

cmgr -f pgadm start

cmgr -f pgdb start

cmgr -f datastore-adm start

cmgr -f dstr start

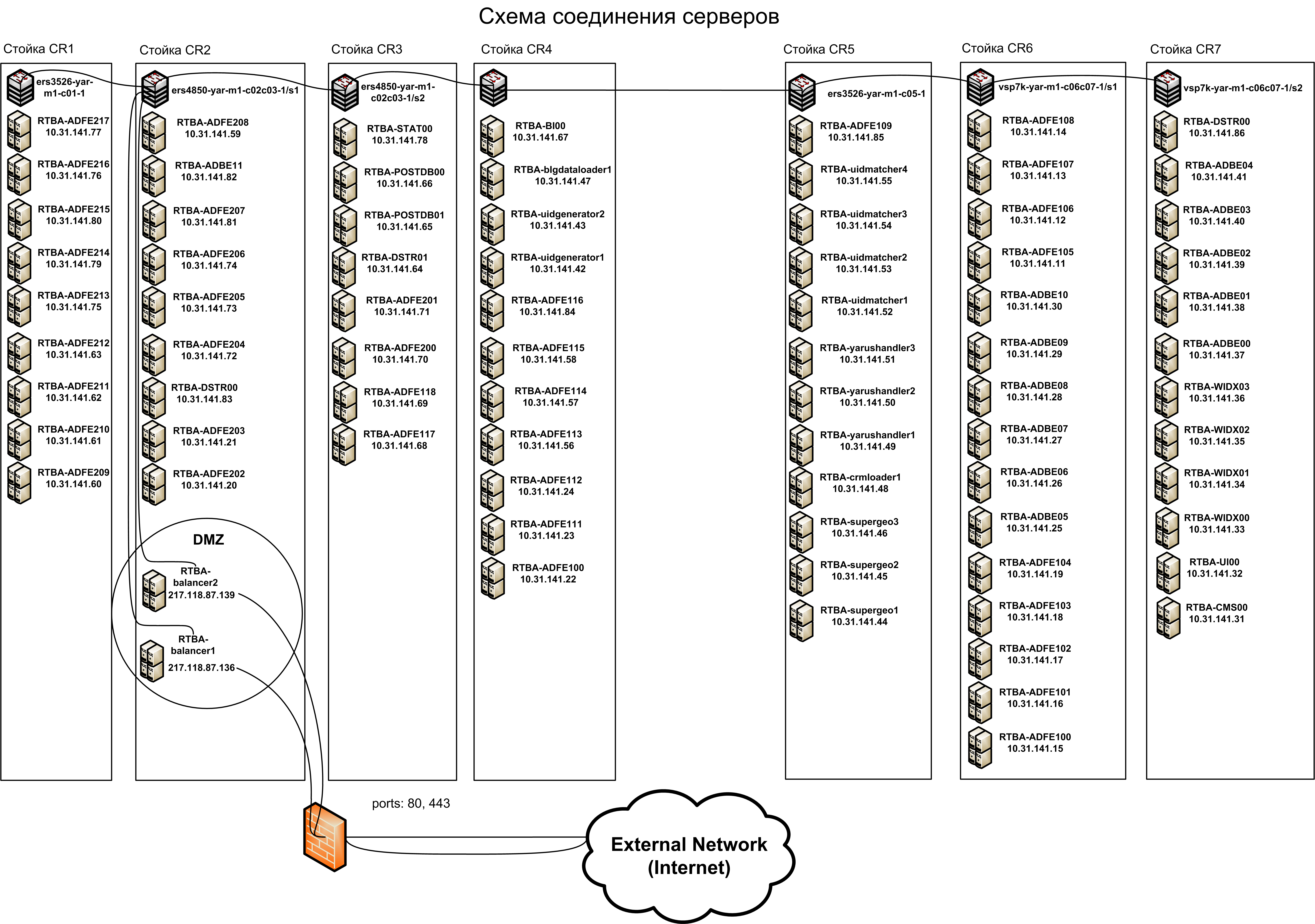
cmgr -f foros-bi start

cmgr start ui

cmgr start bi

cmgr start adserver

### Сетевая архитектура

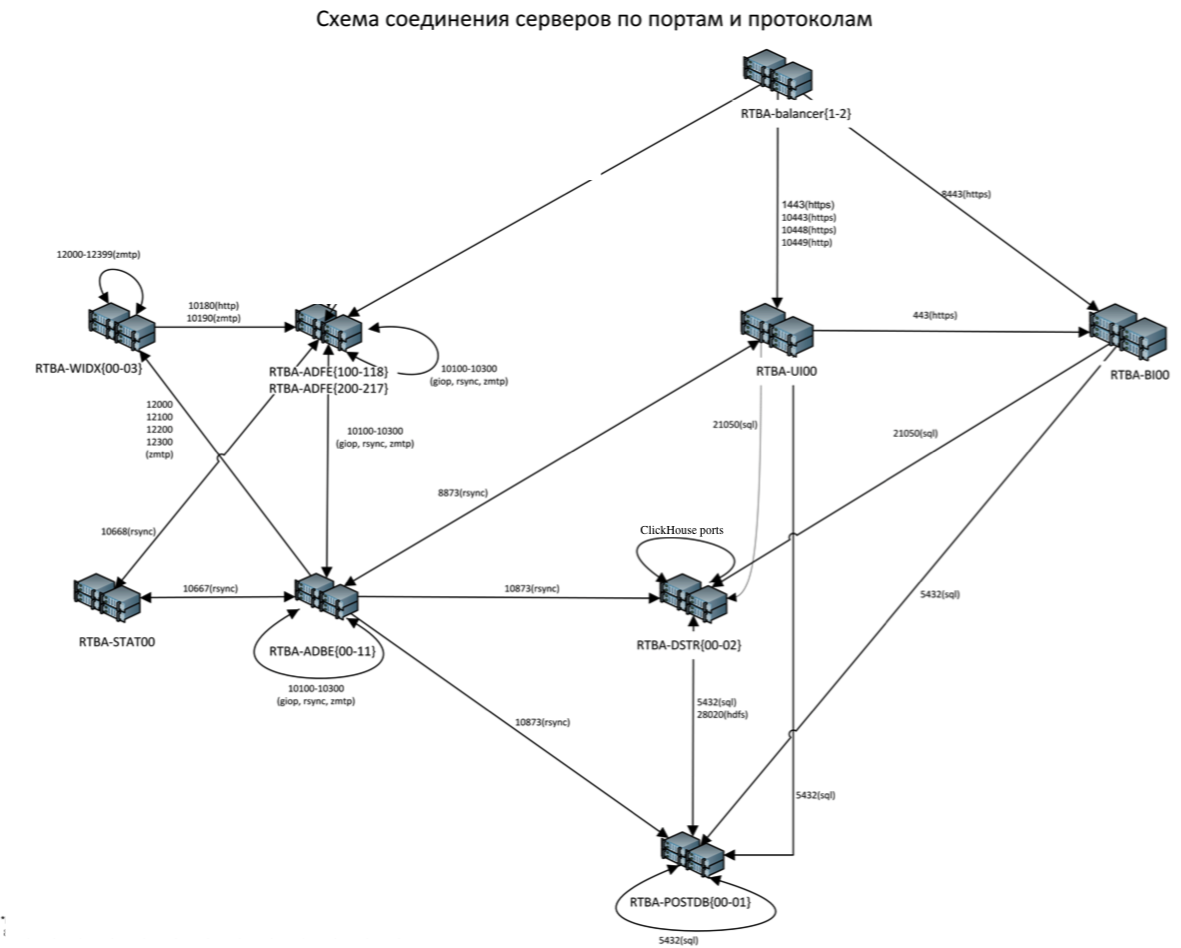


Все точки входа в кластер из сети Интернет размещены на хостах RTBA-balancer1 и RTBA-balancer2

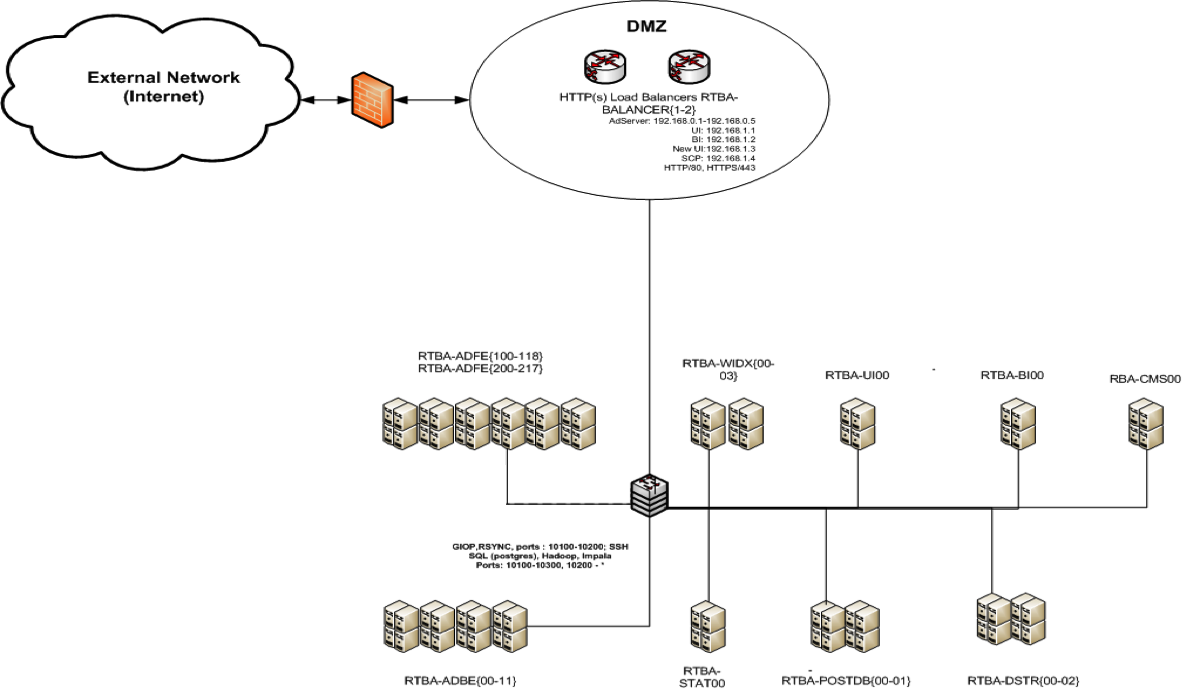
Для полноценного запуска необходимо 4 адресса:

* 3 IP адрессов для обслуживания RTB, возврата рекламы и ее ресурсов. Используемые порты: 80(HTTP), 443(HTTPS)
* 1 IP адресс для UI и BI. Используемые порты: 80(HTTP), 443, 10443, 10444(HTTPS)

Внутренние адресса хостов RTBA-balancer1 и RTBA-balancer2 на данный момент неизвестны – будут использованы текущие внутренние адресса.



Логическая схема соединений серверов



# Требования к техническим и программным средствам

|  |  |  |
| --- | --- | --- |
| *Продукт или его компоненты* | *Аппаратные требования* | *Програмные требования* |
|  |  |  |
| Predictor(STAT) | RAM >= 256 Gb,  2 x SATA >= 1.2 Tb | ОС: CentOS 6.9;  ПО: perl, python,  sklearn, xgboost |
| WebIndexer(WI) | RAM >= 128 Gb,  SSD >= 700 Gb | ОС: CentOS 6.9;  ПО: boost, llibcurl, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ |
| CampaignServer (BE) | RAM >= 256 Gb, SSD >= 800 Gb,  SATA >= 1.4 Tb или SSD >= 800 Gb | ОС: CentOS 6.9;  ПО: ACE-TAO, httpd, apr, boost, GeoIP, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ, xgboost. |
| LogGeneralizer(BE) |
| DictionaryProvider(BE) |
| ExpressionMatcher(BE) |
| RequestInfoManager(BE) |
| UserOperationGenerator(BE) |
| BillingServer(FE) | RAM >= 128 Gb, SSD >= 600 Gb | ОС: CentOS 6.9;  ПО: ACE-TAO, httpd, apr, boost, GeoIP, libevent, net-snmp, xerces-c++, libxslt, libxml, tokyocabinet, ZeroMQ, xgboost. |
| CampaignManager(FE) |
| ChannelServer(FE) |
| ChannelController(FE) |
| UserBindServer(FE) |
| UserBindController(FE) |
| UserInfoManager(FE) |
| UserInfoController(FE) |
| ProfilingServer(FE) |
| nginx(FE) |
| FCGIServer(FE) |
| Ad load balancer | RAM >= 16 Gb | ОС: CentOS 6.9;  ПО: keepalived |
| UI load balancer |
| Cluster control and monitor utilities |  | ОС: CentOS 6.9; |
| YandexSync | RAM >= 64 Gb | ОС: CentOS 6.9;  ПО: java 1.8 |
| UI | ОС: CentOS 6.9;  ПО: cronie, logrotate, rsync, python, postgresql94, postgresql94-libs, python-psycopg2, net-snmp-subagent, curl, procps, httpd, mod\_ssl, mod\_tomcat-connector, glassfish, java-1.7.0-oracle, java-1.7.0-oracle-devel, java-oracle-tzdata >= 2015a, which, memcached, net-tools, libNLPIR-license, libNLPIR, GeoIP-City-CSV, python-jinja2, libxslt, openssl |
| BI | RAM >= 128 Gb | ОС: CentOS 6.9;  ПО: procps, curl, java-1.8.0-oracle, java-1.8.0-oracle-devel, java-oracle-tzdata >= 2015a, postgresql94-server, postgresql94-libs, postgresql94, net-snmp-subagent, syslog-ng, zenoss |
| PostgreSQL | RAM >= 128 Gb, SAS диски ~3Tb | ОС: CentOS 7.3;  ПО: postgresql94, postgresql94-server, postgresql94-contrib, postgresql94-plperl, postgresql94-plpython, perl-DBI, perl-DBD-Oracle-11g, oracle-instantclient-11g-basic, cronie, crontabs, croniter, python-crontab, python-dateutil, net-snmp-subagent, procps, pgbadger, python-jinja2 |
| ClickHouse | SATA или SAS диски ~3Tb | ОС: CentOS 6.9;  ПО: python-psycopg2, oozie, java-1.7.0-oracle-devel, java-1.7.0-oracle, java-oracle-tzdata >= 2015a, clickhouse |

|  |  |  |
| --- | --- | --- |
| *Рабочее место пользователя* | *Аппаратные требования* | *Програмные требования* |
| **Рекламодатель** |  | *Один из Web браузеров c выключенным AdBlock плагином:*  Google Chrome 16+  Firefox 8+  Internet Explorer 10+  Opera Desktop  Yandex Browser  Safari 6+ |
| **Аккаунт менеджер** |
| **Исследователь поведения пользователей интернета** |
| **Web-издатель** |

# Функциональные требования

Таргетинг рекламы

* Гео таргетинг (страны, регионы, города)
* Устройства
* Браузеры
* Операционные системы
* Black list, White list площадок для показа рекламы (реализуется через составные сегменты)
* Список SSP
* Подтип показа (in app, video pre-roll, middle-roll …)
* Частота показов: максимум показов на уника, максимум показов в заданный период, минимальное время между показами (полная гарантия исполнения, даже при конкурентном поступлении запросов за ставкой для уника).

Возможность редактирования критериев таргетинга (включая содержимое сегмента) с гарантией вступления в силу новых настроек в течение 10 мин (за исключением отказа БД).

Отсечение дубликатов событий для рекламного показа и контроль их корректности (включая notice url).

Сведение конверсионных пикселей и показов/кликов рекламы.

Полная гарантия корректности и полноты биллинговой статистики при отказе элементов системы.

Полный контроль перевыполнения бюджетных ограничений – т е отключение отгрузки компании в момент поступления события, которое выполняет бюджет (нужно понимать что для CPC компаний отключение произойдет в момент клика, что не гарантирует дальнейшее отсутствие кликов на уже совершенных показах).

Возможность редактирования и создания новых шаблонов для рекламных показов (реализуется текстовыми шаблонами с использованием оговоренного набора токенов).

Сейчас реализованы шаблоны: native, web push, banner, overlay banner, custom html, VAST, VAST Wrapper.

Исполнение механизма предсказания KPI критериев при наличии достаточного объема информации (в частности критично для CPC компаний, т к CTR используется для расчета закупочной цены).

Полный набор механизмов синхронизации пользователей с SSP и между ними: инициируемый SSP, синхронизация на выкупе, синхронизация на редиректе (актуально для web push SSP), синхронизация инициируемая платформой с подбором не синхронизированных SSP для конкретного пользователя.

Исполнение 99% процентов RTB запросов с задержкой не превышающей 80 ms, и устойчивость данного механизма к отказу компонент или оборудования (обеспечивается реализацией внутрикластерного ограничения на исполнение запроса).

Исполнение cookie sync и событийных запросов(подтверждение показа, клик и т п) с задержкой не превышающей 1 s.

Возможность обновления ПО кластера в режиме Zero Downtime (полное исполнение функциональных требований в каждой точке времени).

Возможность горизонтального расширения кластера с пропорциональным ростом пропускной способности кластера.

Выгрузка и индексация Click Stream по ключевым словам.

В разработке: механизм автоматического расширения сегментов с url списками через look like по содержимому сайтов (ключевым словам).

# Опыт эксплуатации

Кластер из 36 хостов эксплуатировался под максимальной нагрузкой 120К rps/sec RTB запросов + 80K rps/sec CSP запросов.

Использовалось 1.6 Gbit/sec входного канала.

При суммарном объеме таблиц метчинга (отображение SSP ID в UID): 1.4 Tb.

При суммарном объеме профилей пользователей: 9.2 Tb данных с зеркалированием (4.6 Tb чистых данных).

Среднее время ответа на RTB запросы (вход и выход из баласировщика): 24 ms.

Кол-во запросов с временем ответа >50 ms не превышало 0.4% (внутреняя отсечка = 60 ms).

# Требования к безопасности

Система хранит информацию о пользователях как совокупность сегментов, к которым он относится, что не позволяет получить конкретной информации о посещенных URL. Целостность данных обеспечивается контрольными суммами. Данные хранятся в двух экземплярах на разных хостах.

Доступ к кластеру, хранящему профили пользователей, осуществляется только для привилегированных пользователей через VPN.

|  |  |  |
| --- | --- | --- |
| *Роль/профиль доступа* | *Полномочия* | *Ответственность* |
| **Рекламодатель (Advertiser or Agency)** | * настройка рекламной кампании * управление рекламной кампанией * пополнение бюджета |  |
| **Аккаунт менеджер (AdOperations Manager, AdOp)** | * заведение аккаунта рекламодателя в системе * настройка рекламных кампаний согласно требованиям рекламодателя * управление рекламными кампаниями | * контроль соответствия настроек рекламных компаний договорам с рекламодателями * оповещение рекламодателя о неоптимальности настроек * оказание поддержки рекламодателю |
| **Исследователь поведения пользователей интернета**  **(Channel Manager)** | * Создание и редактирование аудиторных сегментов | * поддержка актуальности базы описаний сегментов |
| **Web-издатель**  **(Publisher)** | * описание рекламных размещений |  |

# Требования к доступности

Класс продукта: Business Operational.

Требования к доступности: 24x7;

Система реализует обновление по схеме Zero Downtime, поэтому плановых перерывов в обслуживании не предполагается.

Система не имеет «single points of failure», за исключением подключения данной системы к сети интернет, если не предусмотрен резервный канал.

В случае недоступности системы максимальный риск – невыполнение обязательств по отгрузке рекламы.

# Описание основного сценария функционирования.

* Система получает информацию о поведении пользователя от SSP и CSP(click stream provider) в следующем виде:
* данные о времени посещения страницы и URL страницы;
* геопривязка пользователей;
* привязка пользователей к сегментам внешних систем.
* На основе данных, полученных от SSP и CSP, пополняются профили пользователей. Пользователи привязываются к аудиторным сегментам на основе следующих факторов:
* частоты и количества посещений заданных URLs;
* геопривязки пользователя;
* частоты и количества посещений страниц с заданным контентом;
* браузера и операционной системы, используемых пользователем;
* частоты и количества поисковых запросов;
* выполнения логических выражений над уже существующими сегментами;
* и т.д.
* Рекламодатель вносит в систему рекламные кампании и конфигурирует обязательные критерии показа:
  + - бюджет рекламной кампании и план его реализации;
    - частота показов креатива уникальному пользователю;
    - целевая аудитория (набор аудиторных сегментов);
    - геоположение;
    - минимально допустимые показатели KPI (минимальный CTR, BR, Avg Session Duration, CR); данная настройка возможна только если существует regression модель по требуемым показателям (модель взята из схожей компании, произведен холодный или смещенный старт компании и расчитана новая модель)
    - и т.д.
* Для каждой рекламной кампании формируется набор аудиторных сегментов в соответсвии с требованиями и целями рекламодателя. На начальных этапах это делается операторами системы, после сбора достаточного количества информации система делает это автоматически. Подбор аудиторных сегментов производится таким образом, чтобы наиболее эффективно достичь целей рекламной кампании. Критериями эффективности (KPI) могут быть CTR, BR, Avg Session Duration или CR. Цель рекламной кампании и KPI формулирует рекламодатель.
* При внесении в систему новой рекламной кампании загружаются ее креативы, предоставленные рекламодателем.

Основной сценарий показа рекламы описан ниже:

1. Пользователь посещает веб страницу со встроенным рекламным размещением.

2. Web-издатель или SSP транслирует запрос к системе Foros.

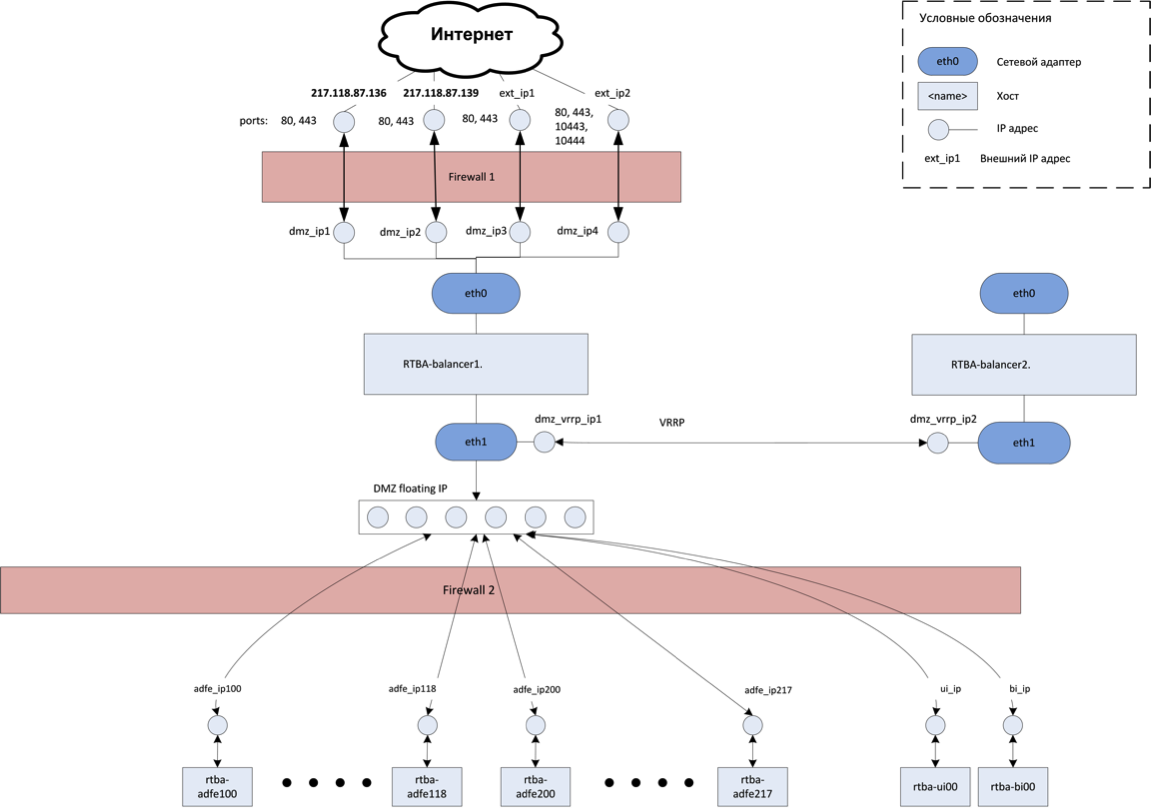
3. Система определяет возможность и показывает наиболее рентабельный рекламный контент в текущий момент времени в зависимости от предыдущего поведения пользователя.

4. При необходимости система сводит профили пользователей данного SSP с профилями пользователей других SSP на основе пользовательских cookie.

* Оптимизация рекламной кампании (в том числе подбор эффективных сегментов аудитории, форматов рекламы, геопривязок и т.д.) может проходить как в ручном, так и в автоматическом режиме при помощи регрессионных моделей (xgboost, catboost, vanga и их ансамбли).

# Приложения

## Описание внешнего доступа к кластеру



**Настройки балансировщика**

**dmz\_ip\_1, порты: 80,443**

порт 80 балансируется на RTBA-adfe{100-118,200-217}:10181

порт 443 балансируется на RTBA-adfe{100-118,200-217}:10146

привязанные домены:

* bidders.rtb.{DOMAIN}.ru - RTB Protocols; запросы от SSP систем

**dmz\_ip\_2, порты: 80,443**

порт 80 балансируется на RTBA-adfe{100-118,200-217}:10180

порт 443 балансируется на RTBA-adfe{100-118,200-217}:10145

привязанные домены:

* ssp1.rtb.{DOMAIN}.ru - Ad Protocol, User Bind; запросы за рекламой и синхронизация пользователей с SSP)
* ssp2.rtb.{DOMAIN}.ru - Video Ad Protocol (CORS enabled); запросы за видео рекламой

**dmz\_ip\_3, порты: 80,443**

порт 80 балансируется на RTBA-adfe{100-118,200-217}:10182

порт 443 балансируется на RTBA-adfe{100-118,200-217}:10147

привязанные домены:

* media.rtb.{DOMAIN}.ru(ports: 80, 443) - Ad Content, Video Ad Content

**dmz\_ip\_4, порты: 80,443,10443,10444**

порт 443 направляется на RTBA-ui00:9443

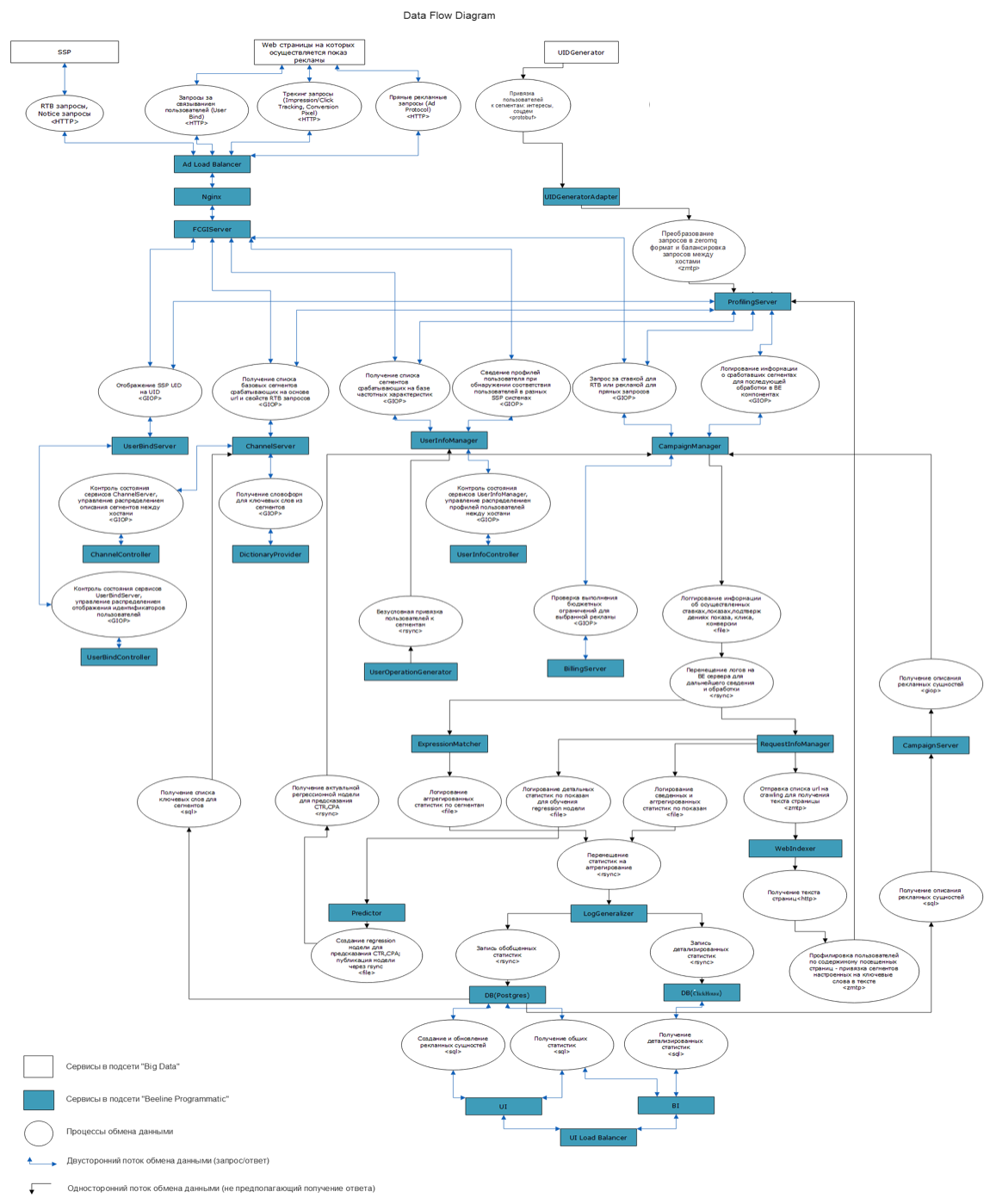
порт 10443 направляется на RTBA-bi00:8443

порт 10444 направляется на RTBA-ui00:10448

привязанные домены:

* ui.rtb.{DOMAIN}.ru(port: 443) - UI : интерфейс управления рекламными сущностями
* ui.rtb.{DOMAIN}.ru(port: 10443) - BI : интерфейс доступа к детальным отчетам
* ssp3.rtb.{DOMAIN}.ru(port: 10444) – содержимое рекламы для preview в UI

## Диаграмма потоков данных



1. TO DO [↑](#footnote-ref-2)