|  |
| --- |
| **ОАО "АЛЬФА-БАНК"**  **Утверждено** РУКОВОДСТВО Администратора  Системы сбора внешних данных  **Версия 0.1**  **Москва**  **2018** |

Оглавление

[1 История изменений 4](#_Toc511933599)

[2 Предисловие 5](#_Toc511933600)

[2.1 Термины и сокращения 5](#_Toc511933601)

[2.2 Ссылки 6](#_Toc511933602)

[2.3 Требования к аппаратному и программному обеспечению 7](#_Toc511933603)

[3 Общее описание системы 8](#_Toc511933604)

[3.1 Основные архитектурные принципы 8](#_Toc511933605)

[3.2 Источники данных 8](#_Toc511933606)

[3.3 Модели данных 8](#_Toc511933607)

[4 Состав Системы 9](#_Toc511933608)

[5 Руководство по установке Системы 11](#_Toc511933609)

[5.1 Порядок первоначальной установки системы 11](#_Toc511933610)

[5.2 Порядок обновления основного приложения ETL Системы 12](#_Toc511933611)

[5.3 Порядок установки плагинов Системы 12](#_Toc511933612)

[6 Основное приложение ETL 14](#_Toc511933613)

[6.1 Общее описание 14](#_Toc511933614)

[6.2 Компоненты основного приложения ETL 14](#_Toc511933615)

[6.3 Функции API IPC 14](#_Toc511933616)

[6.4 Процедуры Oracle 15](#_Toc511933617)

[6.5 Журналы логирования 17](#_Toc511933618)

[6.6 Регулярная загрузка 18](#_Toc511933619)

[6.7 Перезагрузка вручную 18](#_Toc511933620)

[6.7.1 Общее описание 18](#_Toc511933621)

[6.7.2 Условия запуска 19](#_Toc511933622)

[6.7.3 Инструкция 19](#_Toc511933623)

[7 Плагины 21](#_Toc511933624)

[7.1 Плагин «DWH Data» 21](#_Toc511933625)

[7.1.1 Общие сведения 21](#_Toc511933626)

[7.1.2 Компоненты плагина 21](#_Toc511933627)

[7.2 Плагин «СПАРК» 21](#_Toc511933628)

[7.2.1 Общие сведения 21](#_Toc511933629)

[7.2.2 Компоненты плагина 21](#_Toc511933630)

[7.3 Плагин «СПАРК» 21](#_Toc511933631)

[7.3.1 Общие сведения 21](#_Toc511933632)

[7.3.2 Список CTL-Потоков (Потоков-Модулей) 22](#_Toc511933633)

[7.3.3 Список REG-Потоков (Регламентных потоков) 23](#_Toc511933634)

[7.3.4 Инициализирующая загрузка 25](#_Toc511933635)

[8 Мониторинг системы 27](#_Toc511933636)

[9 High Availability (встроенные средства для построения отказоустойчивых решений) 28](#_Toc511933637)

[10 Порядок восстановления загрузки после фатального сбоя системы 28](#_Toc511933638)

[11 Порядок установки патчей 29](#_Toc511933639)

[12 Периодические процессы в системе 29](#_Toc511933640)

[12.1 Автоматическая нарезка секций в таблицах 29](#_Toc511933641)

[12.2 Работа с табличными пространствами 29](#_Toc511933642)

[12.3 Backup данных 29](#_Toc511933643)

[12.4 Сбор статистики 30](#_Toc511933644)

[13 Штатный функционал администратора системы 31](#_Toc511933645)

[13.1 Компрессия таблиц 31](#_Toc511933646)

[13.2 Ручная нарезка секций в таблицах 31](#_Toc511933647)

# История изменений

|  |  |  |  |
| --- | --- | --- | --- |
| **Дата** | **Номер версии** | **ФИО автора изменений** | **Описание изменений** |
| 16.04.2018 | 0.1 | Богданов С.Ш. | Начальная версия |

# Предисловие

Документ описывает порядок установки, настройки и администрирования Системы сбора внешних данных (далее ССВД).

## Термины и сокращения

|  |  |  |
| --- | --- | --- |
| **Понятие** | **Сокращение** | **Определение понятия** |
| Банк |  | ОАО «Альфа-банк» |
| Система сбора внешних данных, External data collection system | ССВД,  EDCS | Информационная система, предназначенная для сбора информации из внешних источников и ее преобразования для удобного использования другими системами Банка. |
| Область временного хранения плагина СCВД | Staging, STG | Область данных, предназначенная для: хранения данных, выгруженных из исходных систем; очистки и трансформации данных; для подготовки их к загрузке в базу данных плагина; хранения вспомогательных данных, используемых в ходе загрузки |
| Плагино-ориентированная ахитектура |  | Архитектура приложений, в которой определяется основное приложение с общим набор функций и сервисов, для которого могут быть реализованы функциональные расширения (плагины). |
| Плагин |  | Независимо поставляемый программный модуль, динамически подключаемый к основной программе и предназначенный для расширения её функциональных возможностей. |
| DWH |  | Корпоративное хранилище данных |
| Hadoop Distributed File System | HDFS | Распределенная файловая система Hadoop |

## Ссылки

|  |  |
| --- | --- |
| Ссылка | Описание |
| [1] | [ОТАР «Система сбора внешних данных»](http://confluence.moscow.alfaintra.net/x/OwxeBg) |
| [2] | [Техническая спецификация «ETL Management»](http://confluence.moscow.alfaintra.net/x/aKflCg) |
| [3] | [Документация ССВД](http://confluence.moscow.alfaintra.net/x/-yd2BQ) |
| [4] |  |
| [5] |  |
|  |  |
|  |  |
|  |  |

## Требования к аппаратному и программному обеспечению

Системное программное обеспечение:

* ETL-инструмент - Informatica PowerCenter Big Data Management 10.2 with Real-Time Add-on,
* СУБД - Oracle 12с R1 и выше.

Минимальные требования к аппаратному обеспечению соответствуют требованиям к программному обеспечению.

Детальный сайзинг (Capacity Planing) оборудования не является предметом данного документа.

# Общее описание системы

## Основные архитектурные принципы

1. Система разработана в концепции плагино-ориентированной архитектуры для обеспечения единства принципов реализации вне зависимости от разнообразия внешних источников, состава команд разработки.
2. В качестве основного приложения используется продукт ETL Management 3.0, которое обеспечивает необходимые и достаточные сервисы для приложений, реализованных в виде плагинов.
3. Для каждого внешнего источника реализуется приложение-плагин, которое использует функции и сервисы основного приложения в своей работе.
4. Все приложения, как основное, так и приложения плагинов, реализуются в рамках одного набора технологий.

Концептуально архитектура ССВД представлена в рамках документа [ОТАР «ССВД»](#otar_ssvd).

## Источники данных

Источниками данных для Системы являются внешние поставщики данных, перечень и характеристики которых определяются в рамках документа [ОТАР «ССВД»](#otar_ssvd).

## Модели данных

1. Модель данных основного приложения определена в рамках документа [Техническая спецификация «ETL Management»](#cts_etl)
2. Модели данных плагинов определены в соответствующих разделах [документации Системы сбора внешних данных](#edsc_docs).

# Состав Системы

Компоненты Системы делятся на классы по своему назначению:

* Системные компоненты определяют платформу Системы.
* Компоненты основного приложения предназначены для обеспечения базового функционала и обсулживания Системы.

|  |  |  |  |
| --- | --- | --- | --- |
| № | Группа компонентов/ Компонент | Описание | Платформа (размещение) |
| 1. | Системные компоненты | | |
| 1.01 | Oracle Database 12с R1 и выше | РСУБД, используемая в качестве основной платформы для баз данных DWH и витрин данных | HP-UX, Linux x86 |
| 1.02 | Informatica Big Data Management 10.1 with Realtime Add-on и выше | Программное обеспечение, используемое в качестве основной платформы для организации ETL- процессов | Linux |
| 2. | Системные компоненты | | |
| 2.01 | DOMAIN\_BDM\_PROD | Схема БД Oracle, в которой размещаются метаданные домена Informatica BDM. | Oracle |
| 2.02 | REP\_IPC\_PROD | Схема БД Oracle, в которой размещаются метаданные репозитория Informatica Power Center . | Oracle |
| 2.03 | MOD\_DVLP\_PROD | Схема БД Oracle, в которой размещаются метаданные Informatica Model Repository Service. | Oracle |
| 2.04 | DIS\_DVLP\_PROD\_BDA11 | Схема БД Oracle, в которой размещаются метаданные Informatica Data Integration Service. | Oracle |
| 2.05 | DIS\_DVLP\_PROD\_BDA21 | Схема БД Oracle, в которой размещаются метаданные Informatica Data Integration Service. | Oracle |
| 2.06 | INT\_SSVD\_EVENTS | PowerCenter Integration Service, организованный для исполнения потоков отслеживания событий и генерации флагов в рамках одного репозитория IPC. | IPC |
| 2.07 | INT\_SSVD\_UNICODE | PowerCenter Integration Service, организованный для исполнения функциональных потоков ССВД с поддержкой Unicode в рамках одного репозитория IPC. | IPC |
| 2.08 | INT\_SSVD\_HDFS | PowerCenter Integration Service, организованный для исполнения функциональных потоков ССВД с поддержкой Unicode и подключения к Hadoop в рамках одного репозитория IPC. | IPC |
| 2.09 | INT\_SSVD | PowerCenter Integration Service, организованный для исполнения функциональных потоков ССВД, не требующий поддержки Unicode и подключения к Hadoop, в рамках одного репозитория IPC. | IPC |
| 3. | Компоненты основного приложения ETL (состав компонентов определен в разделе 6.2) | | |
| 4. | Компоненты плагина (для каждого плагина возможен свой набор) | | |
| 4.01 | ED<Код плагина>\_DICT | **Схема** БД Oracle, в которой размещаются прикладные структуры данных для ручного ввода соответствующего плагина. | Oracle |
| 4.02 | ED<Код плагина>\_DICT | **Tablespace** БД Oracle, в которой размещаются файлы данных ручного ввода соответствующего плагина. | Oracle |
| 4.03 | ED<Код плагина>\_DATA | **Схема** БД Oracle, в которой размещаются прикладные структуры публикационного слоя соответствующего плагина. | Oracle |
| 4.04 | ED<Код плагина>\_DATA | **Tablespace** БД Oracle, в которой размещаются файлы данных соответствующего плагина. | Oracle |
| 4.05 | ED<Код плагина>\_STG | **Схема** БД Oracle, в которой размещаются прикладные структуры стейджинг-слоя соответствующего плагина. | Oracle |
| 4.06 | ED<Код плагина>\_STG | **Tablespace** БД Oracle, в котором определяются файлы данных стейджинг области соответствующего плагина. | IPC |
| 4.07 | ED<Код плагина>\_STG | **Папка** IPC, в которой размещаются прикладные функции ETL соответствующего плагина. | IPC |

# Руководство по установке Системы

## Порядок первоначальной установки системы

1. Выполняется установка сервера Oracle Database в соответствии с руководством по установке Oracle Database Installation Guide
2. Выполняется установка компонентов платформы Informatica, входящих в состав системы, в соответствии с руководством по установке Informatica Installation and Configuration Guide
3. Выполнить настройку серверной части Informatica, в том числе:
   1. В каталоге $PMRoot Dir[[1]](#footnote-1) создать папки:

* ETL\_Config
* ETLLogDir
* Flags
* Scripts
* Param
  1. Для папки $PMRootDir/ETLLogDir дать права на запись группе/пользователю, под которым будут запускаться потоки в IPC
  2. В папку $PMRootDir/ETL\_Config выложить файлы и затем внести настройки (обозначены таблице ниже).

***Note:*** *Объекты являются общими, могут меняться с течением времени и не входят в дистрибутив текущей системы, поэтому экспорт описанных ниже объектов необходимо проводить из системы контроля версий, которая используется на текущий момент при разработке Хранилища Данных. Необходимо использовать последние версии объектов.*

|  |  |
| --- | --- |
| Объект, который необходимо выложить по адресу *$PMRootDir/ETL\_Config* | Настройка, которую необходимо внести после установки объекта |
| com\_inf\_domain.cfg | Изменить имя домена IPC, в котором расположен репозитарий IPC.  *Пример: DOMAIN\_DWPROD* |
| com\_inf\_service.cfg | Изменить наименование интеграционного сервиса IPC.  *Пример: INT\_DW\_PROD* |
| com\_inf\_service\_unicode.cfg | Изменить наименование интеграционного сервиса UNICODE.  *Пример: INT\_DW\_PROD\_UNICODE* |
| com\_inf\_user.cfg | Изменить имя пользователя, из-под которого будут запускаться потоки в репозитарии IPC.  *Пример: etladmin* |
| com\_inf\_pwd.cfg | Изменить пароль для пользователя IPC, , из-под которого будут запускаться потоки в репозитарии IPC.  *Пример: etladmin\_pwd~01* |
| com\_ora\_connect.cfg | Изменить пароль для пользователя УМ и имя репозитария:  *Пример: пароль@DW\_INF\_PROD* |
| com\_um\_connect.cfg | Изменить имя пользователя УМ, пароль для пользователя УМ и имя репозитария:  *Пример: UM/пароль@DW\_INF\_PROD* |

1. Устанавливается дистрибутив последней версии ETL Management 3.0 в соответствии с правилами Релизной политики Системы.

*TODO: В последствии в соответствии с рукодством по установке ETL Management 3.0*

1. Устанавливается патч обновления последней версии ETL Management 3.0 в соответствии с правилами Релизной политики Системы и инструкцией Release Notes.

## Порядок обновления основного приложения ETL Системы

1. Устанавливается патч обновления последней версии ETL Management 3.0 в соответствии с правилами Релизной политики Системыи инструкцией Release Notes.

## Порядок установки плагинов Системы

1. Каждый плагин устанавливаются патчом в соответствии с правилами Релизной политики Системы и инструкцией Release Notes.
2. Для каждого плагина выполнить настройку серверной части Informatica, в том числе:
   1. В каталоге $PMRootDir создать папку $PMRootDir/ED<Код плагина>\_STG, которая определяется корневой папкой настроек плагина.
   2. В каталоге $PMRootDir/ED<Код плагина>\_STG создать папки:

|  |  |
| --- | --- |
| Папка файловой системы | Назначение папки |
| $PMRootDir/ED<Код плагина>\_STG | Корневая папка настроек плагина |
| $PMRootDir/ED<Код плагина>\_STG/Download | Папка, в которую загружаются запакованные файлы от внешнего поставщика данных. |
| $PMRootDir/ED<Код плагина>\_STG/ETL\_Config | Папка, в которой хранятся конфигурационные файлы плагина, а также файлы сертификатов доступа к внешнему источнику (если требуются для плагина) |
| $PMRootDir/ED<Код плагина>\_STG/ETLLogDir | Папка для журналов логирования работы sh-скриптов плагина |
| $PMRootDir/ED<Код плагина>\_STG/Param | Папка для файлов параметров ETL-процесов плагина |
| $PMRootDir/ED<Код плагина>\_STG/Scripts | Папка для sh-скриптов ETL-процесов плагина |
| $PMRootDir/ED<Код плагина>\_STG/SessLogs | Папка для журналов работы сессий ETL-процесов плагина |
| $PMRootDir/ED<Код плагина>\_STG/SrcFiles | Папка для сохранения распакованных файлов, являющихся источниками для ETL-процесов плагина |
| $PMRootDir/ED<Код плагина>\_STG/TgtFiles | Папка для сохранения файлов, являющихся приемниками для ETL-процесов плагина |
| $PMRootDir/ED<Код плагина>\_STG/WorkflowLogs | Папка для журналов работы потоков ETL-процесов плагина |

1. При необходимости переноса накопленных данных из пилотной среды SSVDPLT, он осуществляется также в рамках патчей, оформленных по правилам Релизной политики Системы и инструкцией Release Notes.

# Основное приложение ETL

## Общее описание

Основное приложение ETL использует продукт [ETL Management 3.0](#cts_etl) и обеспечивает необходимые и достаточные сервисы для приложений, реализованных в виде плагинов.

## Компоненты основного приложения ETL

|  |  |  |  |
| --- | --- | --- | --- |
| № | Группа компонентов/ Компонент | Описание | Платформа (размещение) |
|  | EDSC\_META | **Схема** БД Oracle, в которой размещаются структуры метаданных, настроек, журналов процессов, а также программный код, используемый стандартным API Системы . | Oracle |
|  | EDSC\_META | **Tablespace** БД Oracle, в котором определяются файлы данных основного приложения ETL | Oracle |
|  | EDSC\_META | **Папка** IPC, в которой размещается стандартный API основного приложения ETL. | IPC |

## Функции API IPC

Функции API IPC размещаются в папке EDSC\_МЕТА. Для этой папки определено свойство IsShared = Yes.

Определены следующие маппинги:

|  |  |  |
| --- | --- | --- |
| **Маппинг** | **Входные параметры** | **Описание** |
| m\_INIT\_JOB | $$P\_LOADING\_ID *decimal(28)*  Параметр принимает значение ID (идентификатора) исполнения Управляющего потока, в котором регистрируется исполнение функционального. | Предназначен для регистрации исполнения функции ETL в журнале работы функциональных потоков в соответствии с правилами заполнения атрибутов.  Вызывает процедуру INIT\_JOB пакета EDSC\_META.API\_ETL. |
| m\_COMPLETE\_JOB | $$P\_JOB\_ID *decimal(28)*  Параметр принимает значение ID идентификатора исполнения функции ETL. | Предназначен для регистрации завершения исполнения функции ETL в журнале работы функциональных потоков в соответствии с правилами заполнения атрибутов.  Вызывает процедуру COMPLETE\_JOB пакета EDSC\_META.API\_ETL. |
| $$P\_STATUS\_CODE *string(25)*  Параметр принимает значение статуса, с которым завершается исполнение функции ETL.  Значение по умолчанию: SUCCEEDED |
| m\_INIT\_CTL | $$P\_REG\_LOADING\_ID *decimal(28)*  Параметр принимает значение ID исполнения Регламентного потока (reg\_loading\_id), если Управляющий поток исполняется в его составе. | Предназначен для регистрации начала работы Управляющего потока.  Вызывает процедуру INIT\_CTL пакета EDSC\_META.API\_ETL. |
| m\_COMPLETE\_CTL | $$P\_LOADING\_ID *decimal(28)*  Параметр принимает значение ID ( loading\_id), с которым исполняется Управляющий поток. | Предназначен для регистрации завершения работы Управляющего потока.  Вызывает процедуру COMPLETE\_CTL пакета EDSC\_META.API\_ETL. |
| $$P\_STATUS\_CODE *string(25)*  Параметр принимает значение статуса, с которым завершается исполнение Управляющего потока.  Значение по умолчанию: SUCCEEDED |
| m\_INIT\_REG | Параметры отсутствуют | Предназначен для регистрации начала работы Регламентного потока.  Вызывает процедуру INIT\_REG пакета EDSC\_META.API\_ETL. |
| m\_COMPLETE\_REG | $$P\_REG\_LOADING\_ID *decimal(28)*  Параметр принимает значение ID ( loading\_id), с которым исполняется Регламентный поток. | Предназначен для регистрации завершения работы Регламентного потока.  Вызывает процедуру COMPLETE\_REG пакета EDSC\_META.API\_ETL. |
| $$P\_STATUS\_CODE *string(25)*  Параметр принимает значение статуса, с которым завершается исполнение Регламентного потока.  Значение по умолчанию: SUCCEEDED |
| m\_CREATE\_PARAMFILE | Параметры отсутствуют | Предназначен для формирования файла параметров функционального, управляющего или регламентного потока соответственно. |
| m\_LOG\_PARAMPREVVALUE | $$P\_JOB\_ID *decimal(28)*  Параметр принимает значение ID идентификатора исполнения функции ETL. | Предназначен для сохранения в журнале значений параметров исполненных потоков. |

## Процедуры Oracle

Процедуры API Oracle, используемые в процессах установки и эксплуатации, определены в следующем реестре:

|  |  |  |  |
| --- | --- | --- | --- |
| № | **Имя пакета/процедуры** | Тип | **Описание** |
|  | **API\_METADATA** |  | **Пакет процедур управления метаданными и настройками** |
|  | DEL\_MD\_IPCOBJECTS | API | Процедура удаления метаданных для объекта IPC |
|  | UPSERT\_MD\_IPCOBJECTS | API | Процедура изменения метаданных для объекта IPC |
|  | DEL\_MD\_PARAMETERS | API | Процедура удаления метаданных для параметра |
|  | UPSERT\_MD\_PARAMETERS | API | Процедура изменения метаданных для параметра |
|  | DEL\_MD\_IPCOBJPARAMS | API | Процедура удаления метаданных определения значений параметра |
|  | UPSERT\_MD\_IPCOBJPARAMS | API | Процедура изменения метаданных определения значений параметра |
|  | UPSERT\_MD\_TABLES | API | Процедура добавления/изменения метаданных таблицы Системы BI |
|  | REMOVE\_MD\_TABLES | API | Процедура физического удаления сведений о таблице Системы BI из метаданных |
|  | UPSERT\_MD\_COLUMNS | API | Процедура добавления/изменения метаданных oб атрибуте таблицы Системы BI |
|  | REMOVE\_MD\_COLUMNS | API | Процедура физического удаления сведений об атрибуте таблицы Системы BI из метаданных |
|  | …<in progress> |  |  |
|  | IMPORT\_METADATA | API | Процедура импорта файла метаданных, который входит в состав патча. |
|  | EXPORT\_METADATA | API | Процедура экспорта файла метаданных в заданном формате для включения в патч. |
|  | **API\_ETL** |  | **Пакет процедур журналирования исполнения функций ETL, Управляющих, Регламентных и циклических потоков** |
|  | INIT\_JOB | API | Процедура инициализации функции ETL |
|  | COMPLETE\_JOB | API | Процедура регистрации завершения работы функции ETL |
|  | INIT\_CTL | API | Процедура инициализации Управляющего потока модуля ETL |
|  | COMPLETE\_CTL | API | Процедура регистрации завершения Управляющего потока модуля ETL |
|  | INIT\_REG | API | Процедура инициализации Регламентного потока ETL |
|  | COMPLETE\_REG | API | Процедура регистрации завершения Регламентного потока ETL |
|  | INIT\_CYCLE | API |  |
|  | COMPLETE\_CYCLE | API |  |
|  | **API\_USERPARAMS** |  | **Пакет процедур управления пользовательскими настройками параметров** |
|  | SET\_USERPARAM | API |  |
|  | …<in progress> |  |  |
|  | **API\_EVENTS** |  |  |
|  | …<in progress> |  |  |
|  | **API\_SQL** |  | **Пакет процедур для исполнения и журналирования динамических SQL запросов** |
|  | LOG\_SQL | API | Процедура логирования динамических SQL-запросов, исполняемых в рамках PL/SQL процедур и функций. |

## Журналы логирования

Структуры журналов логирования определены в документе [Техническая спецификация «ETL Management»](#cts_etl).

Перечень журналов:

|  |  |  |
| --- | --- | --- |
| № | **Таблица** | **Назначение** |
|  | EDSC\_META.LOG\_LOADING | Журнал работы управляющих и регламентных потоков.  Предназначен для протоколирования работы управляющих и регламентных потоков. |
|  | LOG\_JOB | Журнал работы функциональных потоков.  Предназначена для протоколирования работы функциональных потоков. |
|  | LOG\_PARAMACTIVEVALUE | Журнал значений параметров исполняемых потоков.  Предназначена для регистрации значений параметров, для исполняемых в текущий момент функциональных, управляющих, регламентных потоков. Значения записей журнала могут использоваться для вычисления значений параметров в рамках текущего исполнения. |
|  | LOG\_PARAMPREVVALUE | Журнал значений параметров исполненных потоков.  Предназначен для протоколирования значений параметров, с которыми выполнялись функциональные, управляющие, регламентные потоки. Значения записей журнала могут использоваться для вычисления значений параметров для последующих запусков потоков. |
|  | LOG\_EVENTRAISE | Журнал регистрации событий для исполнения потоков |
|  | LOG\_EVENTFLAG | Журнал регистрации флагов и генерации флагов для EventWait |
|  | LOG\_SQL | Журнал исполнения динамических SQL-запросов.  Предназначен для обязательного протоколирования SQL-запросов, которыe выполняются в рамках PL/SQL процедур загрузки или запросов определения значений параметров |
|  | LOG\_INSTALL | Журнал регистрации обновлений метаданных и системных настроек. |
|  | LOG\_ERRORS | Журнал регистрации ошибок предназначен для логирования программных исключений/ошибок, возникающих в процессе исполнения процедур и функций PL/SQL. |

# Плагины

## Плагин «DWH Data»

### Общие сведения

Плагин «DWH Data» предназначен для загрузки данных из DWH для параметризации запросов к внешним сервисам в рамках сбора информации о клиентах Банка

### Компоненты плагина

|  |  |  |  |
| --- | --- | --- | --- |
| № | Группа компонентов/ Компонент | Описание | Платформа (размещение) |
|  | IDDWH\_DATA | **Схема** БД Oracle, в которой размещаются таблицы данных плагина. | Oracle |
|  | IDDWH\_STG | **Схема** БД Oracle, в которой размещаются структуры метаданных, настроек, журналов процессов, а также программный код, используемый стандартным API Системы . | Oracle |
|  | IDDWH\_STG | **Tablespace** БД Oracle, в котором определяются файлы данных основного приложения ETL | Oracle |
|  | EDSC\_META | **Папка** IPC, в которой размещается стандартный API основного приложения ETL. | IPC |

## Плагин «СПАРК»

### Общие сведения

### Компоненты плагина

## Плагин «ФНС»

### Общие сведения

### Список REG-Потоков (Регламентных потоков)

## Плагин «Google»

### Общие сведения Плагин предназначен для забора данных из Google Big Query. При загрузке осуществляется перенос файлов json сформированных BQ в Hadoop HDFS при помощи консольных утилит входящих в google BQ framework

### Компоненты плагина

|  |  |  |  |
| --- | --- | --- | --- |
| № | Группа компонентов/ Компонент | Описание | Платформа (размещение) |
|  | EDGOOGLE\_STG | **Папка** IPC, в которой размещаются компоненты приложения ETL | IPC |

### Список REG-Потоков (Регламентных потоков)

|  |  |  |
| --- | --- | --- |
| Потоки загрузки | Пререквизиты | Описание |
| WF\_REQ2ETL\_GA2HDFS\_ALFABANK\_2018\_ga\_sessions | Нет | Логи посещений клиентов сайта Альфа БанкЗагрузка осуществляется на T-1 |
| WF\_REQ2ETL\_GA2HDFS\_ALFABANK\_2018\_CostData | Нет | Расходы на интернет рекламу Загрузка осуществляется на T-1 |

### Перезагрузка данных за период Перезгрузка данных осуществляется с помощью штатных возможностей УМ путем вставок параметров в таблицу PARAM\_OVERRIDE Пример скрипта для перезагрузки потоков за произвольный период:



# Мониторинг системы

Мониторинг системы может осуществляться с помощью средств:

1. Informatica Power Center Workflow Monitor. Возможные действия:

* Просмотр статусов загрузки потоков, сессий, command-task
* Осуществление запусков потоков, сессий, восстановление запусков

1. Программное обеспечение для работы с Oracle DB (например, PL/SQL Developer, TOAD). Возможные действия:

* Просмотр и анализ запущенных sql-запросов
* Работа с сессиями Oracle DB (анализ производительности, отключение сессии)
* Работа с таблицей LOG\_SQL, которая расположена в EDSC\_META. В таблицу производится логирование результатов выполнения процедур обновления публикационного слоя
* Просмотр [журналов основного приложения ETL](#_Журналы_логирования)

# High Availability (встроенные средства для построения отказоустойчивых решений)

В данном разделе описаны действия, которые необходимо предпринимать в случаях аварийного завершения потока загрузки или его отдельных модулей.

Отказоустойчивость загрузки реализуется инструментами IPC:

* На каждом из потоков (включая управляющие) включено свойство Suspend on error. Это свойство позволяет перезапускать сессии (tasks) в статусе Failed без полного перезапуска потока.
* На каждом из потоков (включая управляющие) включено свойство Enable HA recovery. Это свойство позволяет перезапускать сессии (tasks) в статусе Stopped, Aborted и Terminared без полного перезапуска потока.

Порядок восстановления загрузки:

1. В случае, если сессия (task) завершилась с ошибкой, т.е. перешла в статус Failed, необходимо устранить проблему и перезапустить поток, в котором была запущена сессия путем Recover workflow. При этом статус потока Suspended поменяется на Running.
2. Если поток, который не является управляющим, завершился с одним из статусов Stopped, Aborted или Terminated, необходимо его перезапустить путем Recover workflow.
3. Если управляющий поток завершился с одним из статусов Stopped, Aborted или Terminated, необходимо:
4. Вручную выявить все не управляющие потоки, которые запускались внутри данного УП и завершились со статусом отличным от Success.

Это можно сделать либо с помощью визуального анализа запуска управляющего потока в IPC Workflow Monitor

1. Выполнить остановку всех полученных потоков методом Stop
2. Перезапустить данный управляющий поток путем Recover workflow

# Порядок восстановления загрузки после фатального сбоя системы

<in progress>

# Порядок установки патчей

Порядок установки патчей диктуется политиками Банка.

# Периодические процессы в системе

## Автоматическая нарезка секций в таблицах

Нарезка секций в таблицах производится автоматически с помощью «механизма автонарезки». Система не входит в поставляемую сборку модулей Хранилища и не описывается детально в данном документе. Дополнительная информация о системе описана в документе [[7]](#url7).

В стандартном случае должен проводится периодический запуск «механизма автонарезки» с следующими параметрами:

|  |  |
| --- | --- |
| **Период запуска** | **Глубина нарезки** |
| Вне загрузки модулей витрины 1 раз в неделю | *sysdate+30*  (на 30 дней «вперед» относительно sysdate |

## Работа с табличными пространствами

Работа с табличными пространствами производится автоматически с помощью «механизма автоматического добавления датафайлов». Система не входит в поставляемую сборку модулей Хранилища и не описывается детально в данном документе.

В стандартном случае должен проводится периодический запуск механизма, который добавляет дата-файлы в табличные пространства с следующими параметрами:

|  |
| --- |
| **Период запуска** |
| 2 раза в час |

## Backup данных

## Сбор статистики

Сбор статистики по таблицам Oracle производится с помощью планировищка задач cron в операционной системе семейства UNIX, который запускает методы для сбора статистики в Хранилище Данных. Планировщик не входит в поставляемую сборку модулей Хранилища и не описывается детально в данном документе.

В стандартном случае должен проводится периодический запуск сбора статистики в любое время вне загрузки модулей витрины для схем, перечисленных в таблице:

|  |  |  |  |
| --- | --- | --- | --- |
| **Схема** | **Таблицы** | **Период запуска** | **ESTIMATE\_PERCENT** |
| EDSC\_META | Все таблицы указанной схемы | Вне загрузки модулей витрины 1 раз в день | 0.1 |
| ED\*\_DATA | Все таблицы указанных схем | Вне загрузки модулей витрины 1 раз в день | 0.1 |
| ED\*\_STG | Все таблицы указанных схем | Вне загрузки модулей витрины 1 раз в день | 0.1 |
| ED\*\_DICT | Все таблицы указанных схем | Вне загрузки модулей витрины 1 раз в день | 0.1 |
| IDDWH\_DATA | Все таблицы указанных схем | Вне загрузки модулей витрины 1 раз в день | 0.1 |
| IDDWH\_STG | Все таблицы указанных схем | Вне загрузки модулей витрины 1 раз в день | 0.1 |

# Штатный функционал администратора системы

## Компрессия таблиц

## Ручная нарезка секций в таблицах

1. $PMRootDir – параметер PowerCenter Integration Service, в котором определено полное имя папки файловой системы UNIX для интеграционного сервиса [↑](#footnote-ref-1)