Informatica Power Center.

Инструкция для разработчиков

Оглавление

[1. Введение 3](#_Toc472624988)

[1.1 Архитектура IPC 3](#_Toc472624989)

[2. Установка IPC 4](#_Toc472624990)

[3.Общая схема работы 6](#_Toc472624991)

[4. Разработка в IPC Designer 7](#_Toc472624992)

[4.1 Создание Source в IPC Designer 8](#_Toc472624993)

[4.1.1 Создать вручную 8](#_Toc472624994)

[4.1.2 Создание Source через импорт таблицы 9](#_Toc472624995)

[4.1.3 Создание Source через копирование. 11](#_Toc472624996)

[4.2 Создание целевой таблицы в Target Designer 12](#_Toc472624997)

[4.3 Создание карты проливки и преобразований в Mapping Designer 13](#_Toc472624998)

[5 Разработка в Workflow Manager 15](#_Toc472624999)

[5.1 Создание потоков 15](#_Toc472625000)

[5.2 Запуск Workflow и постановка на расписание 20](#_Toc472625001)

[6. Мониторинг отработки потов в Workflow Monitor 22](#_Toc472625002)

[7. Работа в IPC Repository Manager 24](#_Toc472625003)

[7.1 Создание/редактирование папок и настройка доступа 24](#_Toc472625004)

[7.2 Экспорт/импорт потоков в Repository Manager 26](#_Toc472625005)

[7.3 Чистка логов с помощю Truncate Log 32](#_Toc472625006)

# Введение

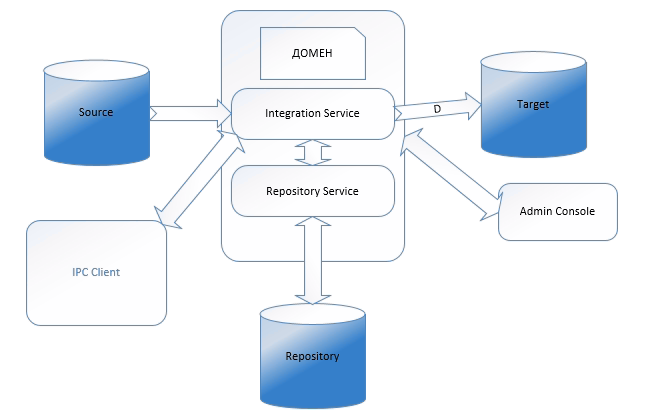
IPC (Informatika Power Center) – средство разработки ETL.

**DWH** – операционное ХД (Oracle).

**ETL** (от англ. Extract, Transform, Load — дословно «извлечение, преобразование, загрузка») — один из основных процессов в управлении хранилищами данных, который включает в себя: извлечение данных из внешних источников; их трансформация и очистка, чтобы они соответствовали потребностям бизнес-модели; и загрузка их в хранилище данных.

IPC представляет собой СОА архитектуру. В основе его лежат веб сервисы IPC.

## Архитектура IPC



Модули клиентской части IPC.



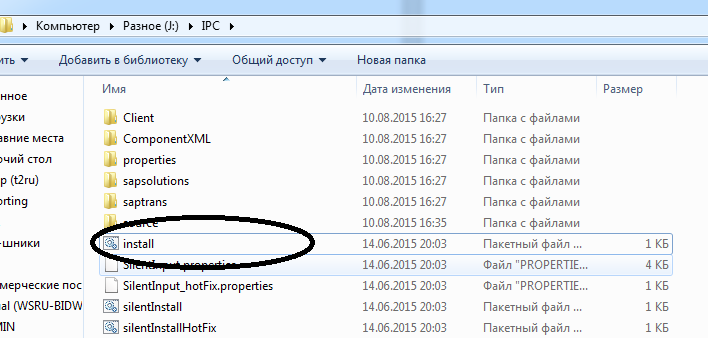
Админ консоль:

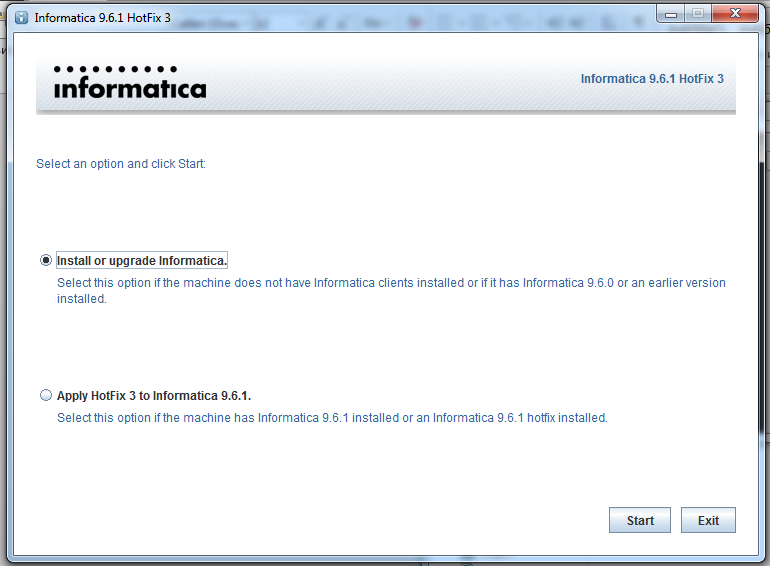


# Установка IPC

Дистрибутив находится в L:\Reporting\distrib\IPC

Запускаем пакетный установщик install.





Выбираем установку, жмем несколько раз Next, и устанавливаем IPC.

После этого обновляем до 9.6.1.

# 3.Общая схема работы

Рассмотрим на примере проливки таблицы TRAFIK\_TYPE. Как разрабатывать ETL процессы в IPC.

Разработаем заливку таблицы из TRAFIK\_TYPE из stage0X в таблицу DWH\_IPC.TRAFIK\_TYPE

1. Создаем таблицу в схеме DWH\_IPC с нужной структурой и предоставляем доступ схеме INFORMATICA.

grant select on stage02.TRAFIK\_TYPE to INFORMATICA;

grant select on stage03.TRAFIK\_TYPE to INFORMATICA;

grant select on stage04.TRAFIK\_TYPE to INFORMATICA;

grant select on stage05.TRAFIK\_TYPE to INFORMATICA;

grant select,insert,update,delete on DWH\_IPC.IPC\_TRAFIK\_TYPE to INFORMATICA;

create table DWH\_IPC.IPC\_TRAFIK\_TYPE

(

trft\_id NUMBER(2) not null,

trft\_name VARCHAR2(50) not null

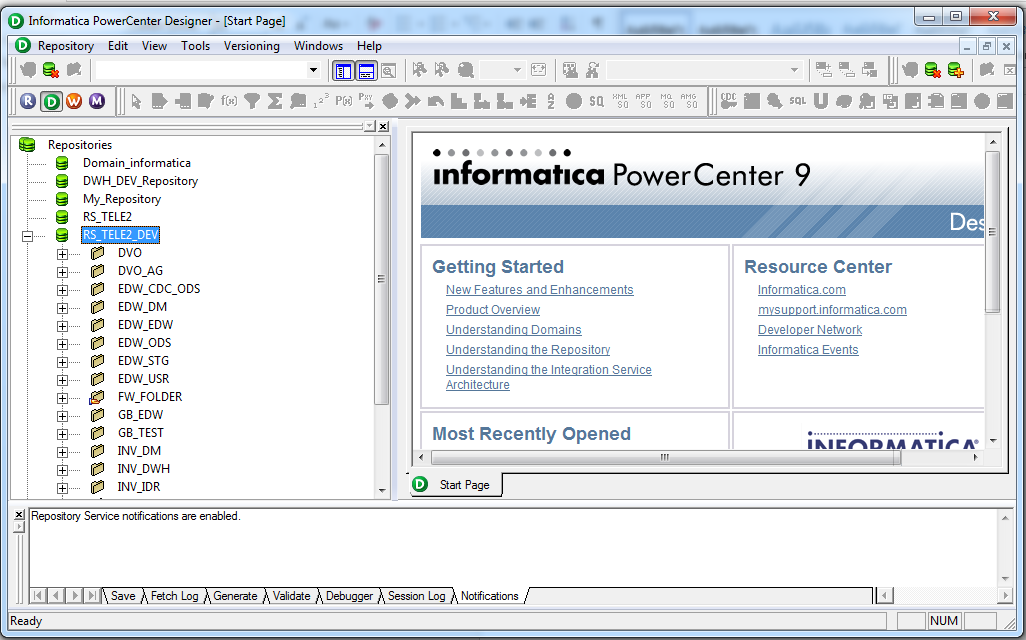
)

INFORMATICA – это техническая учетка, под которым выполняются запросы от сервераа IPC. В данном примере источником и целевой базой является DWH.

# 4. Разработка в IPC Designer

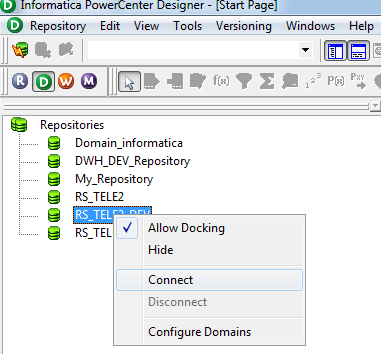
Чтобы настроить механизм преобразования данных, необходимо для начала нарисовать карту преобразования.

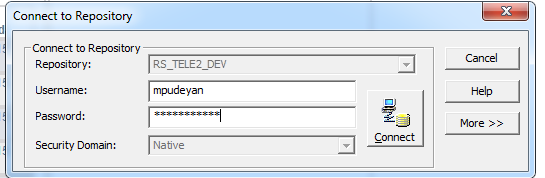
Для этого используется модуль Power Center Designer



Для выполнения проливок, необходимо разработать так называемую карту проливки.

1. Заходим в репозиторий RS\_T2\_DWH\_DEV





## 4.1 Создание Source в IPC Designer

Source – это условная таблица-источник из которой будут выгружаться данные.

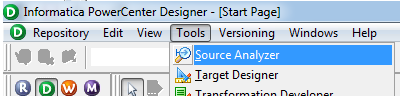
Создадим источник. Для этого заходим в Tools – Source Analyser и открываем рабочее окно разработки источника

Есть несколько способов создать Source:

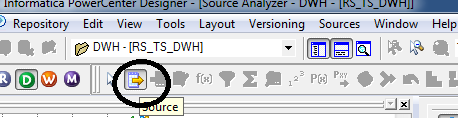
1. Создать вручную
2. Импортировать из базы
3. Скопировать из существующей

### 4.1.1 Создать вручную

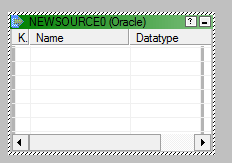
Переходим в рабочее окно создания Source



Нажимаем создать объект



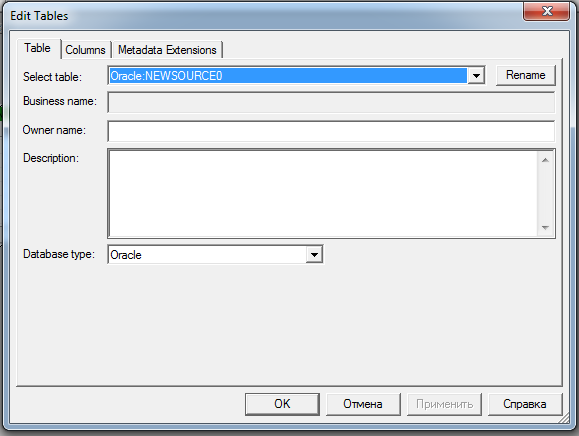
И создаем таблицу

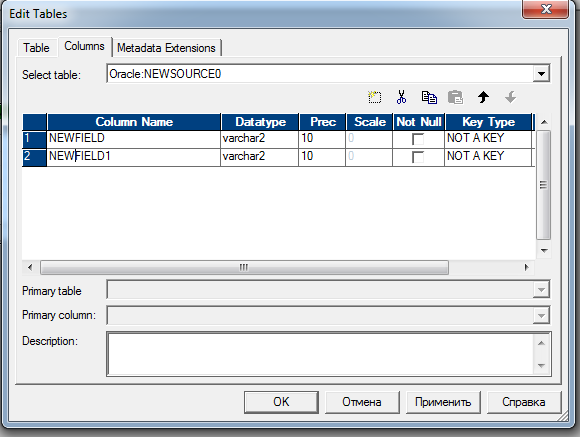


Нажимаем дважды по объекту- переходим в окно настроек:

В первой вкладке Table вводим название таблицы и тип базы данных.

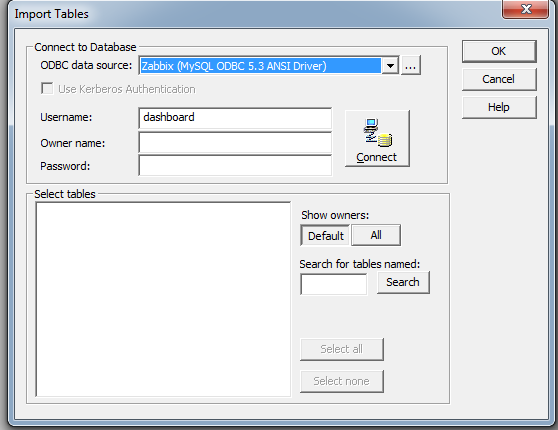
На вкладке Columns редактируем поля в таблице



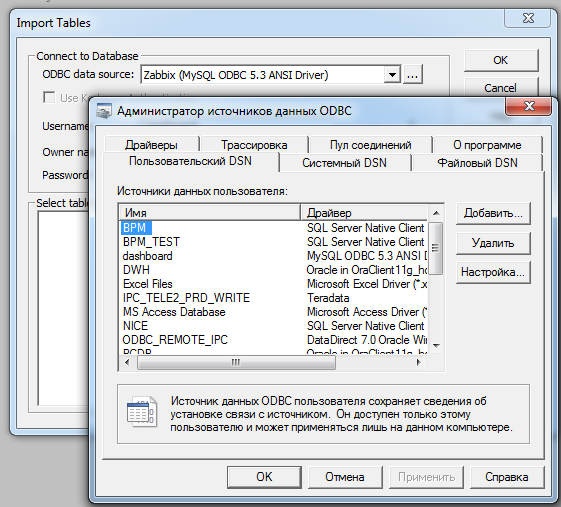


### 4.1.2 Создание Source через импорт таблицы

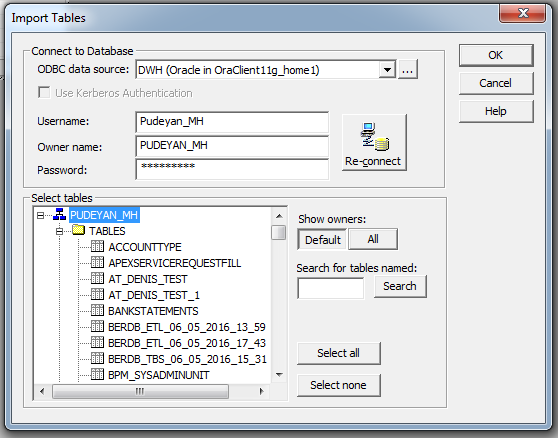
Для этого в пункте меню нажимаем Source – Import From Database



!!Важно понимать, что перед импортом таблицы из базы необходимо настроить ODBC либо другой коннектор на самой локальной машине. И также должны быть установлены нужные драйвера для создания коннектора.



Далее вводим логин и пароль для подключения к базе и вводим название схемы, из которой необходимо импортировть объект.



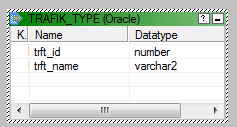
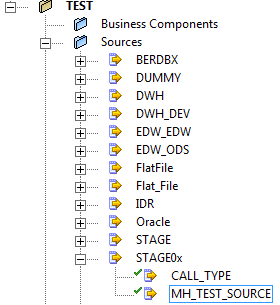


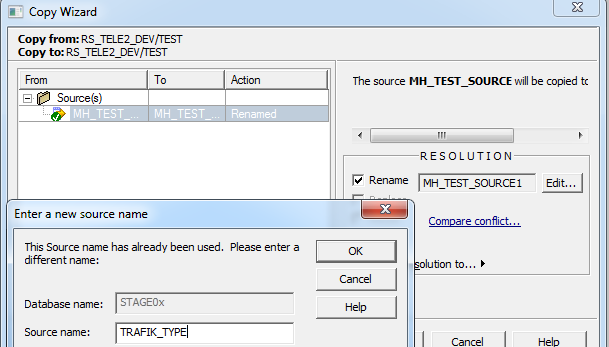
Таблица успешно создана.

### 4.1.3 Создание Source через копирование.

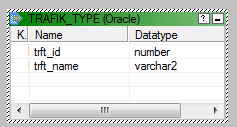
Для этого выделяем любой объект в дереве Source и нажимаем ctrl-c и ctrl-v.



Дизайнер предложит скопировать объект и переименовать. Соглашаемся.

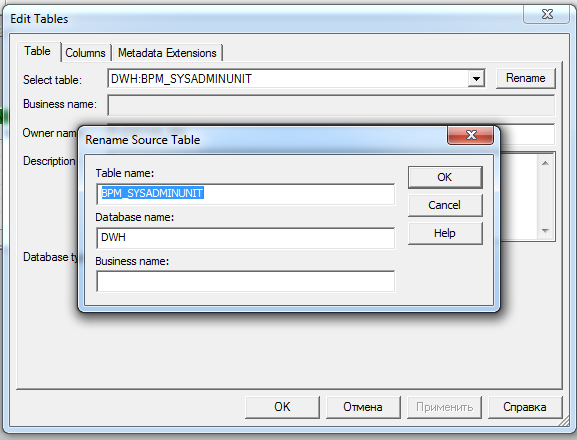


Перетаскиваем созданный объект на экран и в свойствах объекта редактируем набор полей. В итоге должен получится такой источник



Сохраняем проект нажав ctrl-s

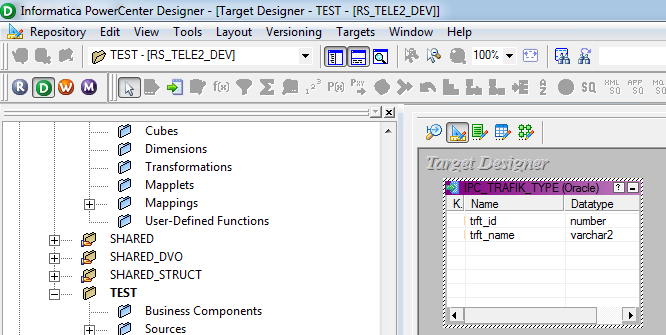
!!!Также при переименовывании объекта важно указать правильный Database name. Как видно на скриншоте выше все Source разбиты по группам и базам согласно их расположению.



## 4.2 Создание целевой таблицы в Target Designer

Создание целевых таблиц в Target Designer аналогичен созданию Source с той лишь разницей, что не нужно редактировать название базы.

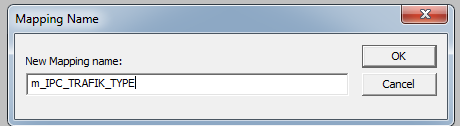
Заходим Tools – Target Designer для создания целевого объекта . Повторив действия описанные в пунктах выше создаем целевой объект в папке Target



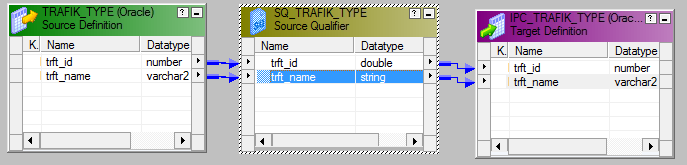
## 4.3 Создание карты проливки и преобразований в Mapping Designer

После создания объекта Target необходимо настроить схему проливки в Mapping Designer.

Для этого заходим в Tools – Mapping Designer. И потом нажимаем Mappings- Create и создаем новый маппинг.



После этого необходимо перетащить созданные объекты из source и targets в маппинг дизайнер и соединить поля между собой. В итоге должно получится следующее.

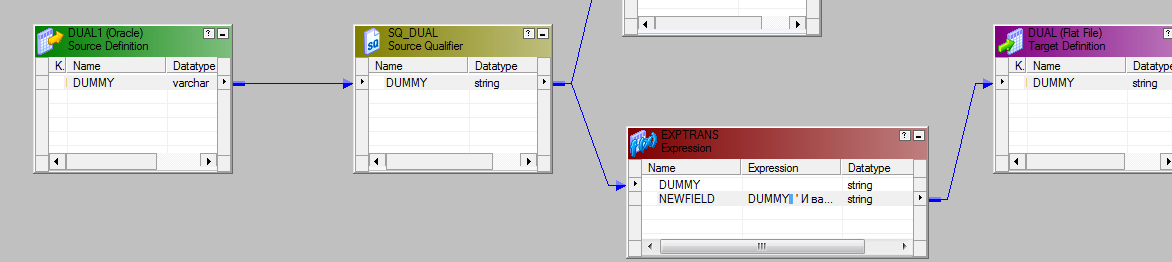


Также в карте проливок можно добавлять различные преобразования. У любой трансформации есть входные и выходные порты (поля) которые участвуют в преобразованиях.

Рассмотрим основные преобразования:

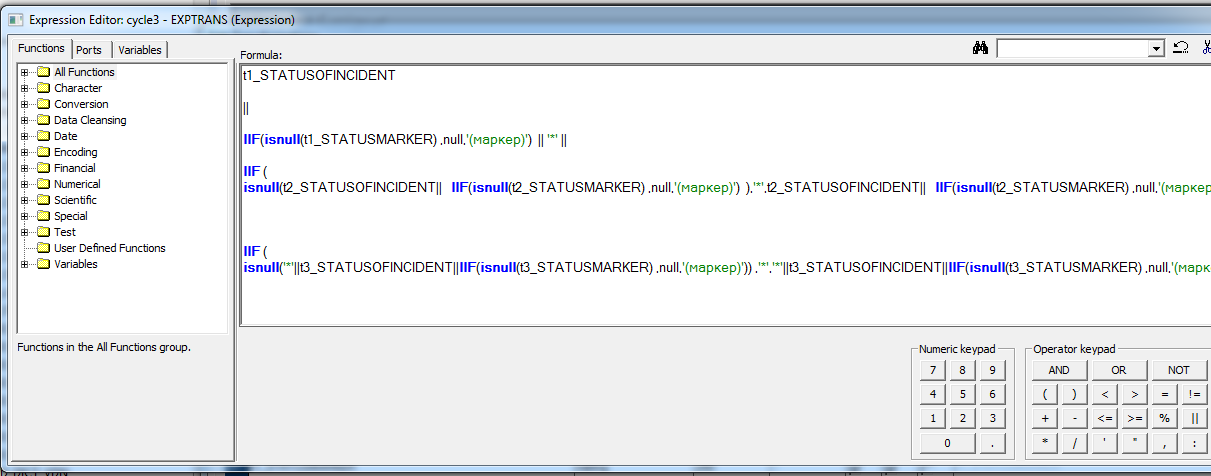
- Update Strategy – позволяет указать DML операцию которую необходимо выполнить с данными. Аналогичен Insert Update Delete.

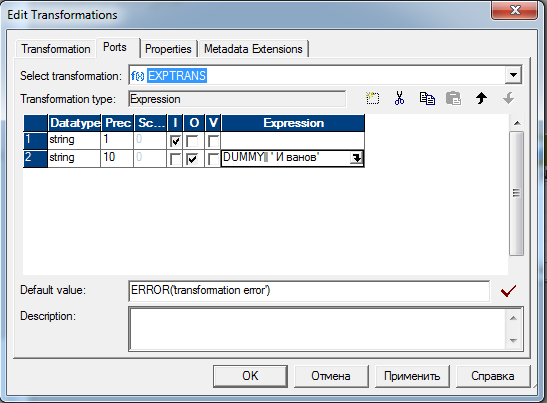
 - Expression Transformation - аналогичен преобразованиям, выполняющимся в разделе select в sql-запросах.



Есть возможность добавлять любые преобразования (например case, decode, присвоение и т.д.)

Ниже пример использования:





 Filter Transformation – трансформация позволяющая наложить фильтр на данные. Аналогичен условию where.

 - Aggregator Transformation – трансформация аггрегации. Аналогичен инструкции Group By

- Lookup Transformation – трансформация отбора. Аналогичен выполнению подзапроса в инструкции SELECT

 - Sequence Generator Transformation - Аналогичен обычному сикнсу (счетчику)

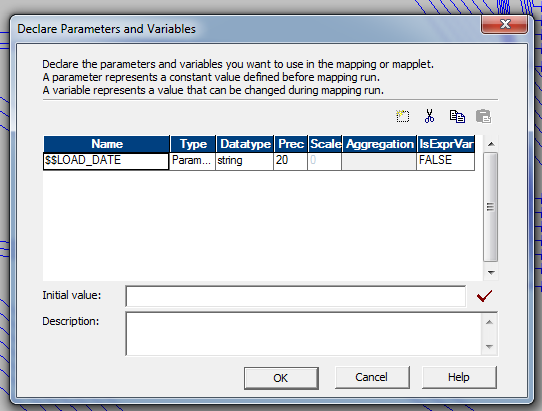
 - Joiner Transformation – аналогичен обычному джойну. Причем в опциях можно выбирать как правый, так и левый джойн.

 - Sorter Transformation –трансформация сортировки. Аналогичен инструкции ORDER BY

 Union Transformation – трансформация объединения. Аналогичен Union All.

Также при разработке потоков можно создавать переменные, в которые записываются значения

Для этого нажимаем Mapping –Parameters and Variables



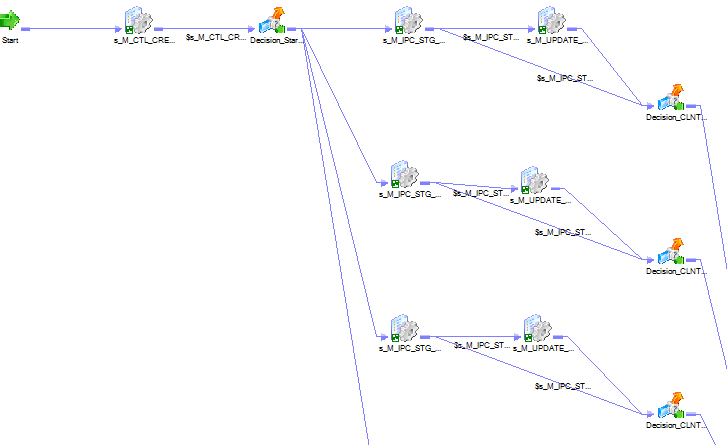
Сами значения переменных проставляются из файла параметров в Workflow Manager

Сохраняем проект нажав crtl-c.

# Разработка в Workflow Manager

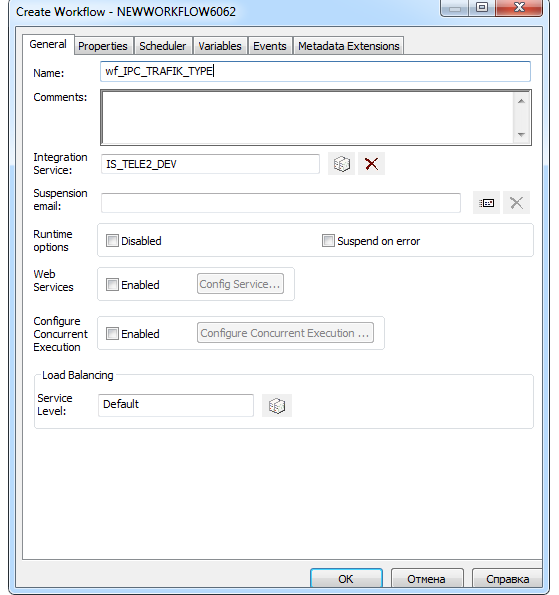
## 5.1 Создание потоков

Модуль Worflow Manager позволяет создавать уже сами потоки, настраивать коннектора к базам а также настраивать расписания для потоков.



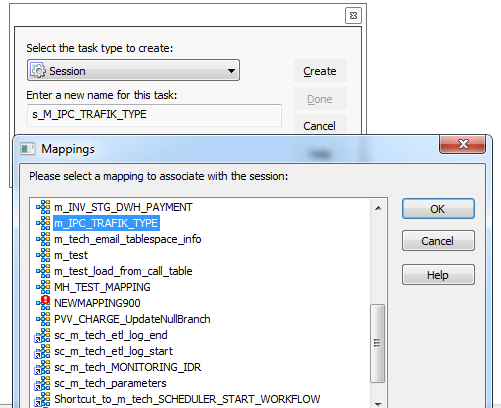
Заходим в Workflow Manager нажав  в Designer. Подключаемся к репозиторию и Tools- Workflow Designer

Нажимаем Workflows-Create и создаем новый Workflow

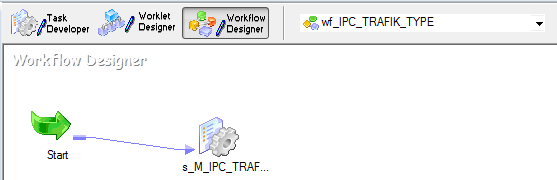


Теперь надо создать сессию, запускающую нашу созданную карту. Для этого Tools – Create Session.

И привязываем созданный ранее маппинг к данной сессии.

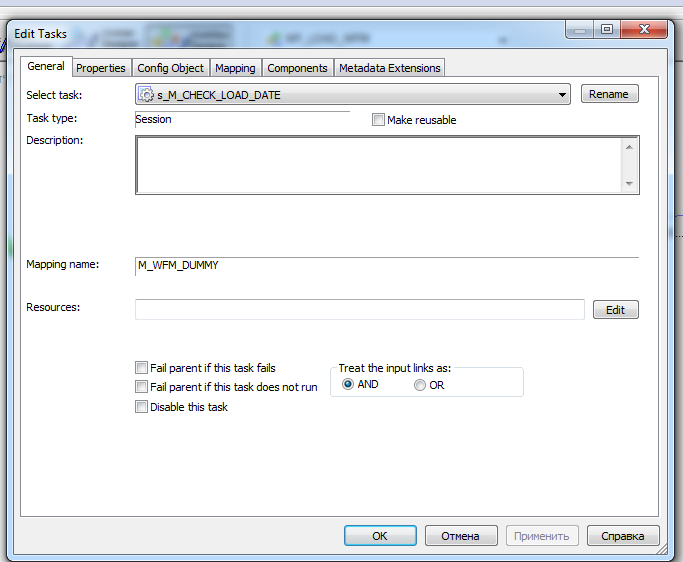


Далее соединяем между собой start и сессию нажав Tasks – Link Task



После этого заходим в настройки сессии (правой кнопкой мыши по сессии - Edit) в вкладку Mapping и настраиваем параметры.

Рассмотрим основные вкладки в сессии:



Вкладка General позволяет ввести основные настройки:

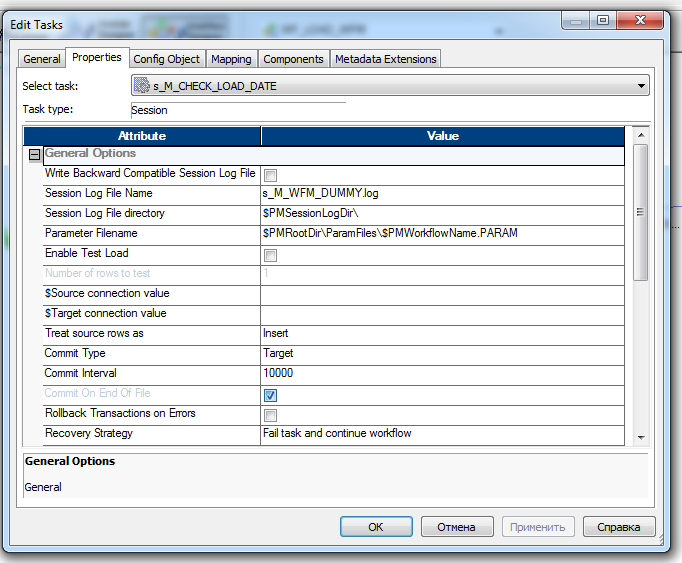


Fail parent if this task fails - опция принуждает упасть весь Workflow при падении данной сессии.

Fail parent if this task does not run - опция принуждает упасть весь Workflow если данная сессия не запустилась

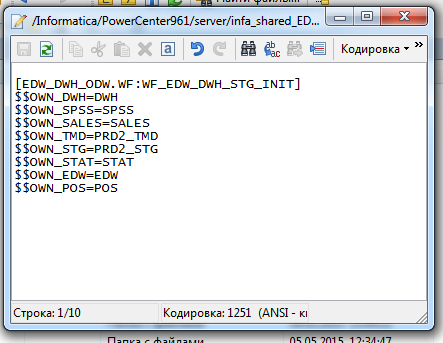
Disable this task – выключает сессию.

Вкладка Properties позволяет ввести основные параметры сессии



Рассмотрим основные:

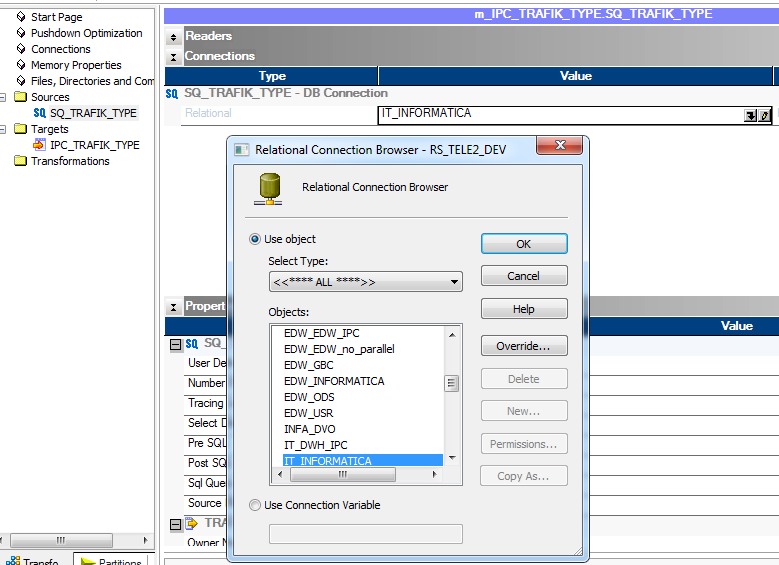
Parameter Filename – здесь вводится дирректория к файлу где прописаны значения переменных для маппингов. Сам файл создается на сервере IPC. Пример текстового файла.



Commit Interval - интервал через который выполняется commit

Вкладка Config Object – позволяет ввести конфигурационные прамаетры. Один из основных параметров Stop on Error – кол-во reject-ов строк, после которого будет выполнено падение сессии (0 - никогда).

Для подключения источника и таргета используется вкладка Mapping.



Для подключения к базе DWH используется объект Connection «IT\_Informatica»

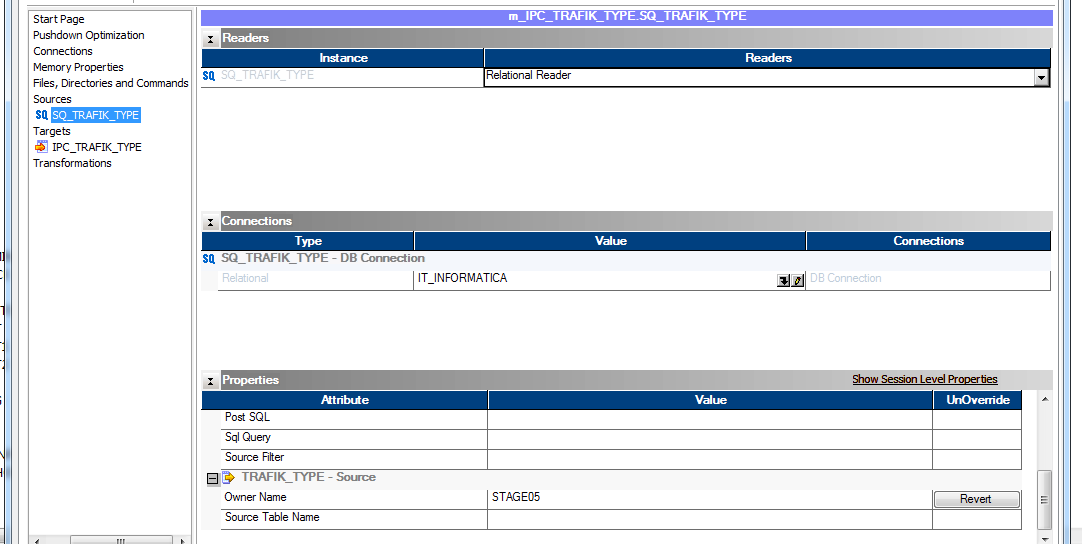
Также для корректной отработки запросов необходимо указать в настройках сессии схему и таблицу из которой надо забирать данные и куда заливать.

Для этого заходим в настройки сессии – вкладка mapping –source и в разделе properties в поле owner указываем схему источника .

Потом повторяем операцию для папки Targets - IPC\_TRAFIK\_TYPE-Properties – и в разделах - Table Name Prefix и Target Table Name вводим схему и название целевой таблицы.

Также в Source и Target есть параметры PreSQL и PostSQL позволяющие прописать тестовый скрипт для запуска в базе.

Параметр SQL Query в Source также позволяет вместо таблицы-источника в базу пробросить запрос для отработки.



## 5.2 Запуск Workflow и постановка на расписание

Существует 2 способа поставить Workflow на расписание:

- Reusable – когда, для постановки на расписание используется стандартный Schduler с преднастроенными параметрами запуска.

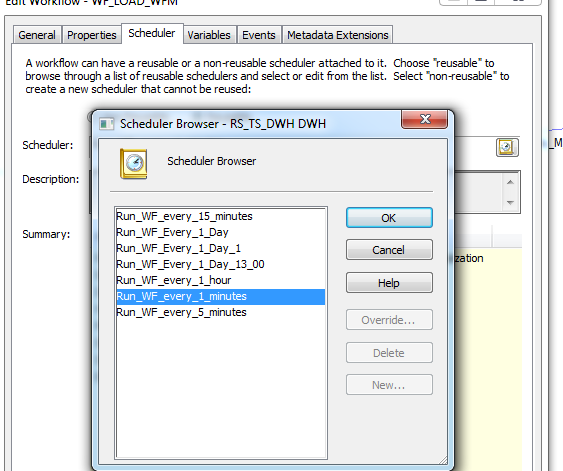
-Non reusable - пользователь сам выбирает расписание для запуска.

!! При постановке на расписание всегда выбираем вариант с reusable. На них завязан механизм ежедневного мониторинга отработки потоков.

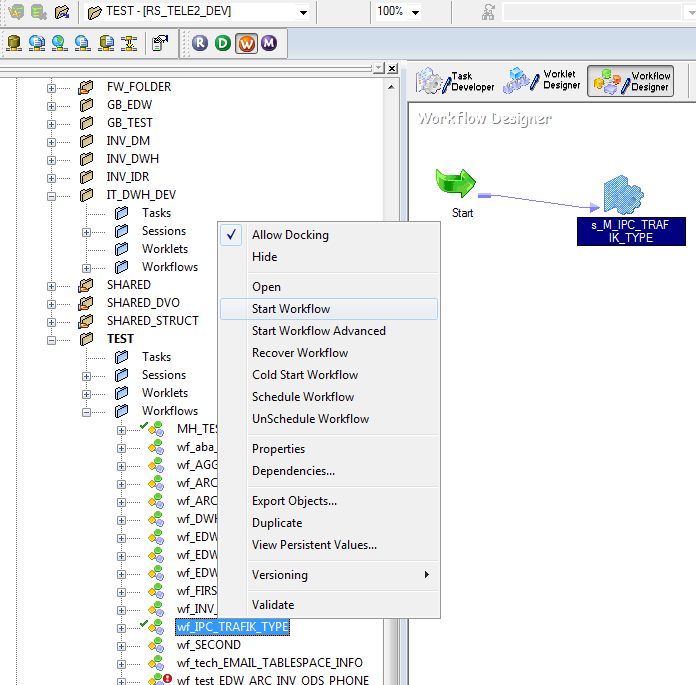
В IPC уже есть набор преднастроенных расписаний. Все из них можно посмотреть в базе

select \* from dwh\_ipc.CTL\_SHEDULERS t

Для постановки на расписание выбираем Scheduler и выбираем нужное расписание.



Теперь у нас есть все необходимое для запуска потока. Для этого нажимаем правой кнопкой мыши на Workflow и start Workflow

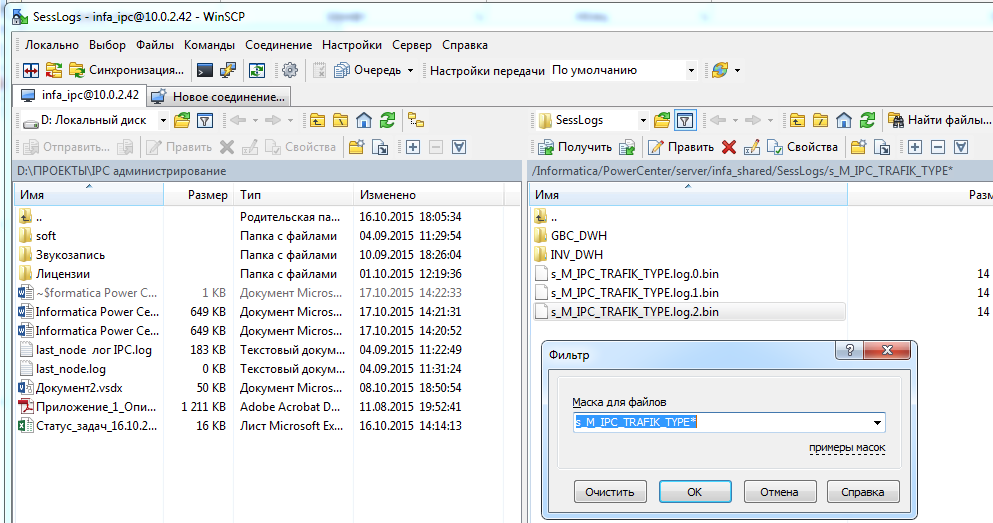


Также при желании можно посмотреть лог отработки сессии. Для этого можно зайти в настройки Sessions - Properties – Session Log File Name там указано название файла которое создается при каждой отработке запросов. Сами логи находятся в директории /Informatica/PowerCenter/server/infa\_shared/SessLogs.

Зайти туда можно с помошью WinCSP.

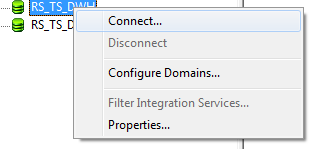
Сервер (зона) с инсталляцией IPC на DWH: 10.0.2.20 и 10.0.2.22

infa\_ipc / \*\*\*\*\*\*\*\* (под этим пользователем стартуют процессы IPC)



# Мониторинг отработки потов в Workflow Monitor

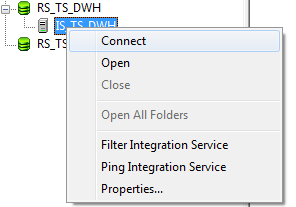
Модуль Workflow Monitor предназначен для проверки отработки потоков и анализа ошибок по которым произошли падения. Для мониторинга подключаемся к нужному репозиторию.



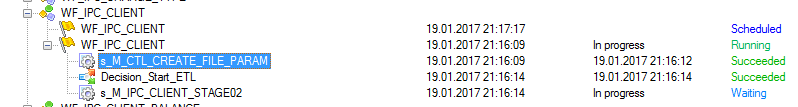
Далее есть два вариант открытия IS сервиса Open и Connect

Open – открывает все имеющиеся потоки и логи по ним.

Connect – позволяет мониторить текущие процессы в реальном времени



Все логи и статусы отработок отображаются в правой части экрана.



Worflow Monitor позволяет управлять всеми потоками в репозитории. Достаточно нажать правой кнопкой мышь на нужном потоке.

Рассмотрим основные команды:

Restart – перезапускает поток.

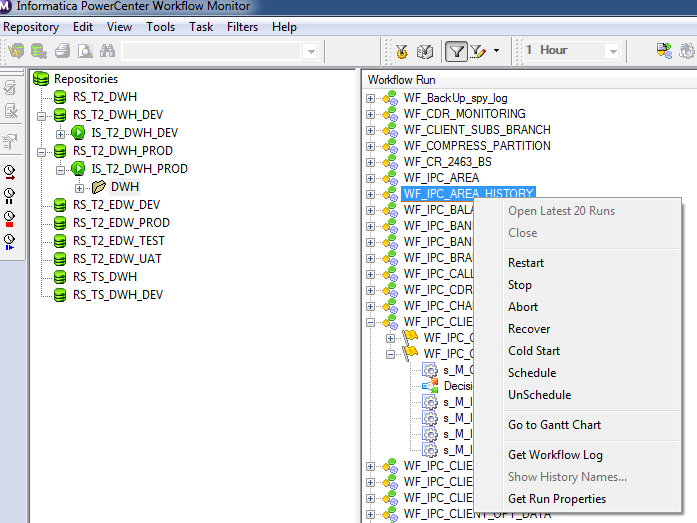
Stop – остановка потока. Аналогичен нажатию стоп в PL/SQL Devloper

Abort- принудительная остановка потока

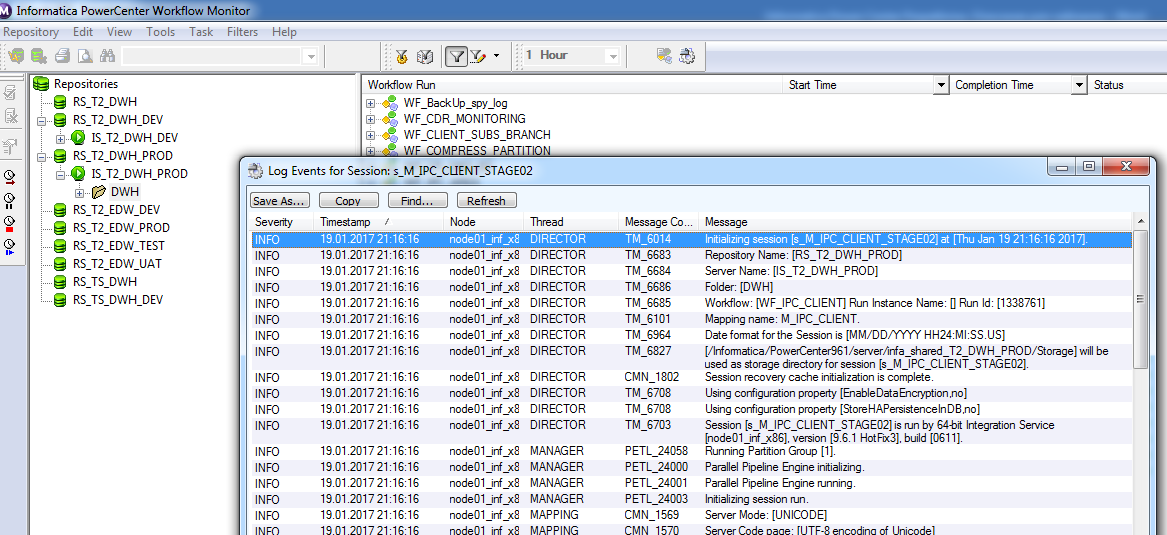
Recover – возобновить отработку потока с той сессии, с которой произошло падение.

Sсhedule- поставить на расписание

UnSchedule – снять с расписания.



Также нажав правой кнопкой мыши на сессии – существует пункт Get Session Log , который позволяет просмотреть подробный лог отработки сессии. Все ошибки в случае падения находятся здесь.

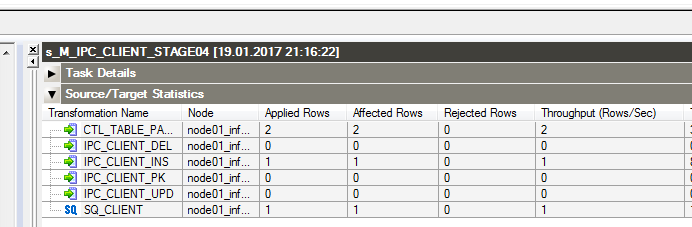


Также справа внизу экрана можно просмотреть статистику отработки потоков:

- количество выгруженых строк

- количество примененных строк

- количество rejected строк



# Работа в IPC Repository Manager

Модуль IPC Repository Manager предназначен для следующих целей:

- создание и редактирование папок для потоков

- настройка доступа и разграничение прав в папках.

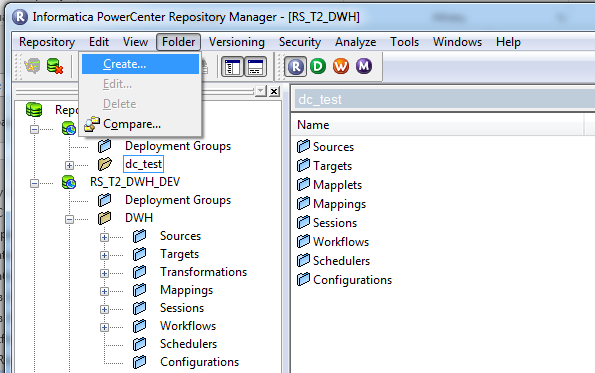
- экспорт/импорт потоков и зависимостей по ним

- чистка логов в Repository Manager

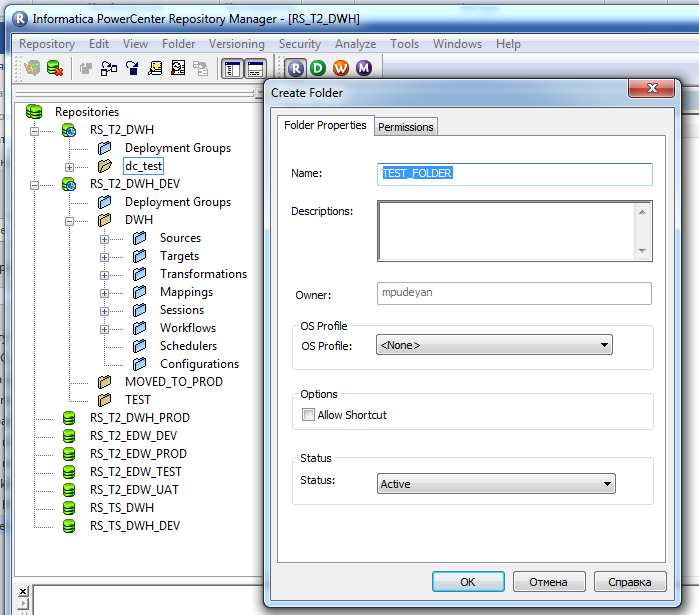
## 7.1 Создание/редактирование папок и настройка доступа

Папка - это некое рабочее пространство в репозитории, предназначеное разграничить доступ к потокам, и логически разделить какие-либо проекты.

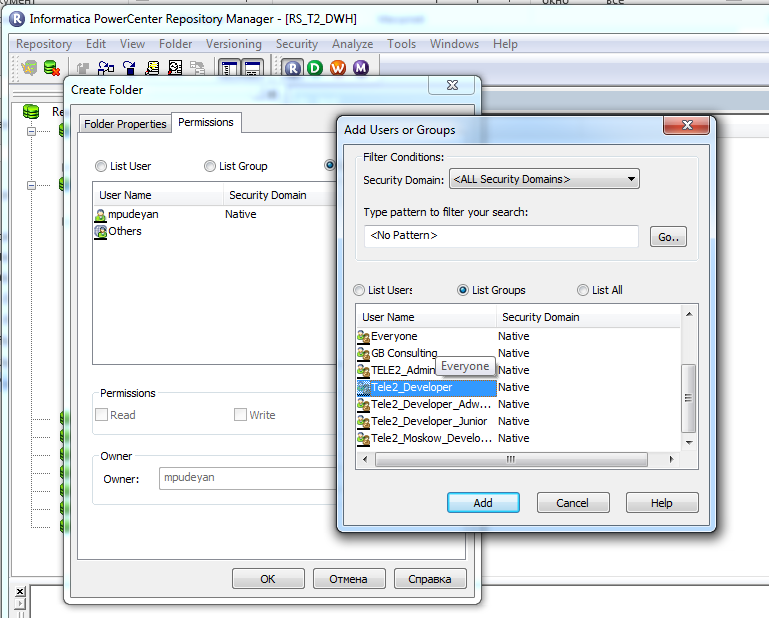
Для создания папки нажимаем Folder - Create и переходим в меню редактирования.



На вкладке Folder Properties вводим название папки и переходим в меню Permissions для настройки прав доступа.



Доступ к папке можно предоставить как отдельным пользователям, так и группе, в которую входят пользователи.



Существует 3 типа доступа к папке:

Read – предоставляет доступ только на просмотр объектов в папке

Write – предоставляет доступ на редактирование объектов в папке

Execute – предоставляет доступ на запуск и постановку на расписание потоков в папке



Щелкаем на нужной группе и проставляем права.

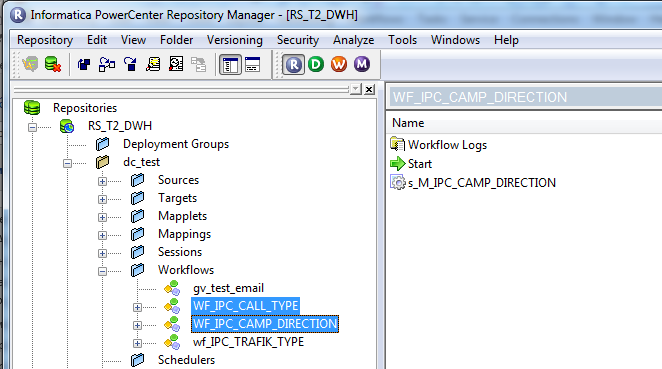
## 7.2 Экспорт/импорт потоков в Repository Manager

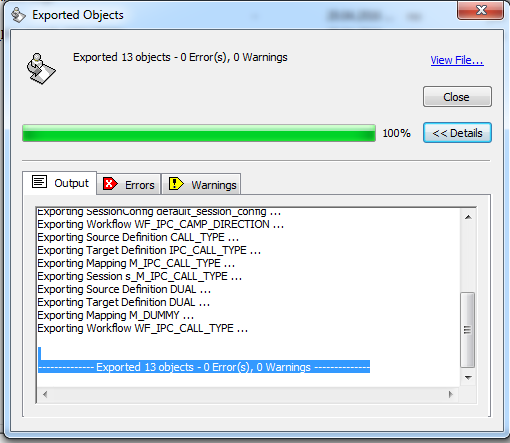
Инструмент экспорт/импорт объектов предназначен для переноса потоков и всех его зависимых объектов из папки в папку / из репозитория в репозиторий.

Делается это следующим образом:

Зажав CTRL выделяем потоки, которые необходимо экспортировать. Нажав Repository- Export Object выбираем папку и название файла для сохранения его в XML. После этого появляется диалоговое окно для экспорта. После выполнения экспорта необходимо проверить что выгрузка прошла успешно. На экране появится лог следующего вида.

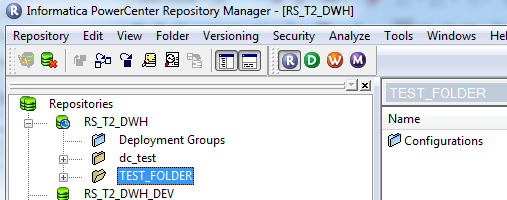
-------------- Exported 13 objects - 0 Error(s), 0 Warnings --------------



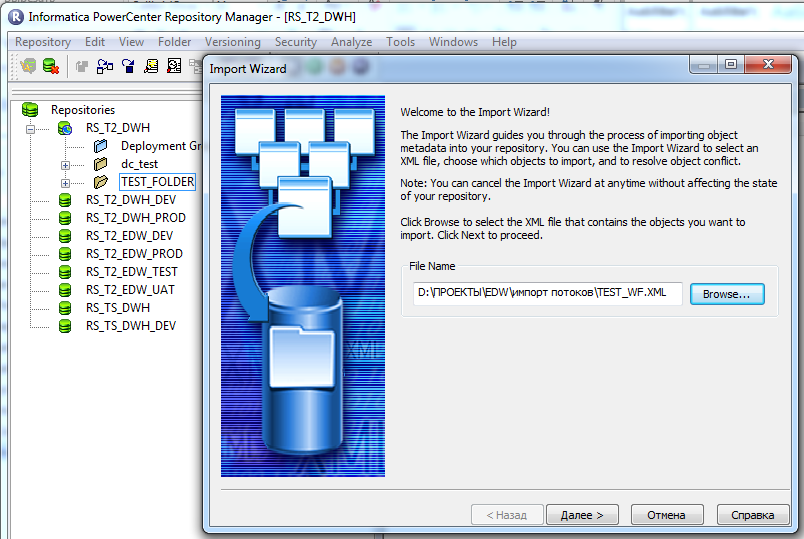


Для импорта WF :

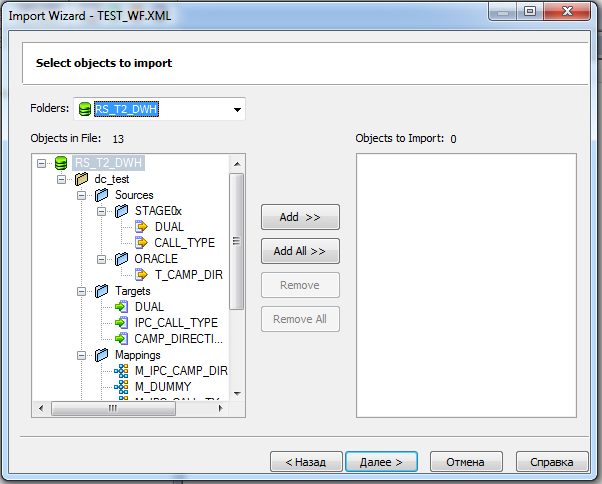
- открываем нужный репозиторий и выделяем папку в которой необходимо развернуть потоки.



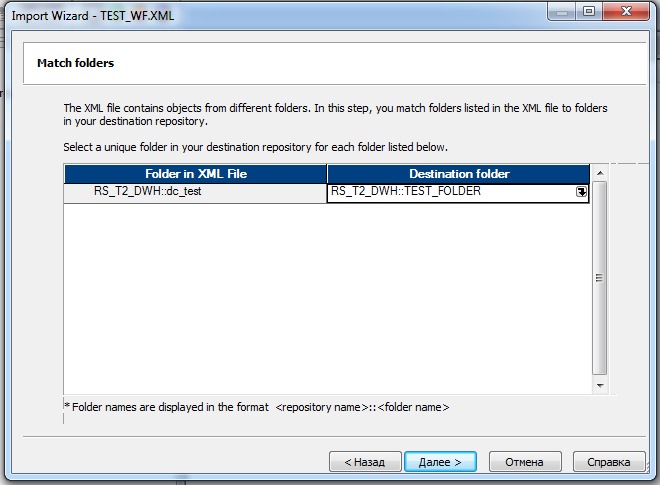
Нажимаем Repositroy – Import Object выбираем XML файл для загрузки и открываем мастер импорта файлов.



Добавляем все объекты для импорта кроме объекта Tasks- default session config

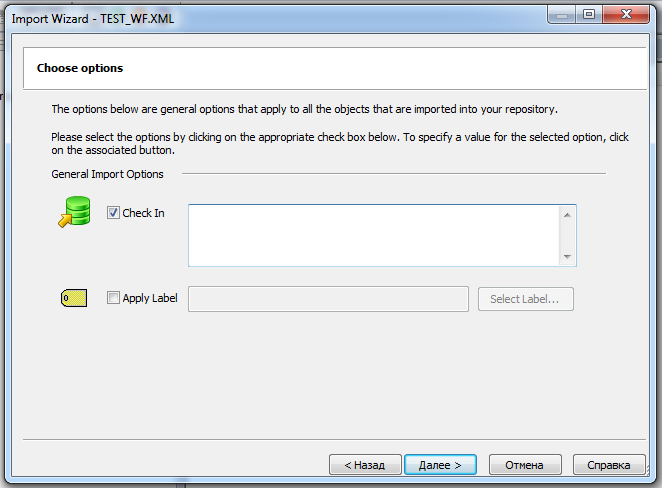


Выбираем папку для импорта



Далее мастер предлагает с. Check In – это аналог SVN только в самой информатике.

Если Check In галочка отсутствует, то поток будет заблокирован под тем пользователем , под которым выполняется импорт потоков.



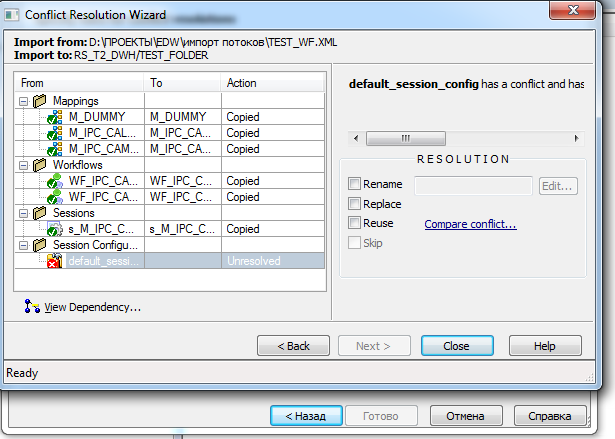
Далее в случае возникновения таких конфликтов, как указано ниже необходимо выбрать действие для мастера:

Rename – переименовать объект

Replace – заменить объект

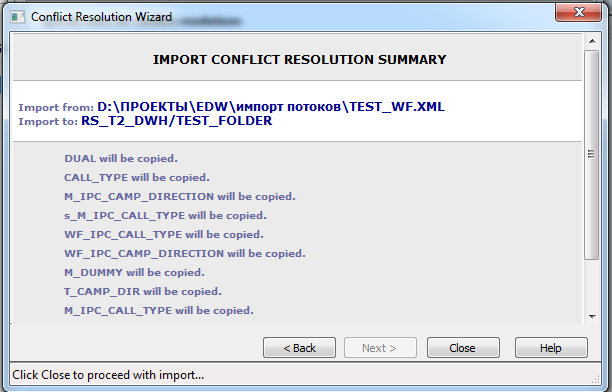
Reuse – не заменять, использовать тот , который уже есть в папке

Skip - пропустить

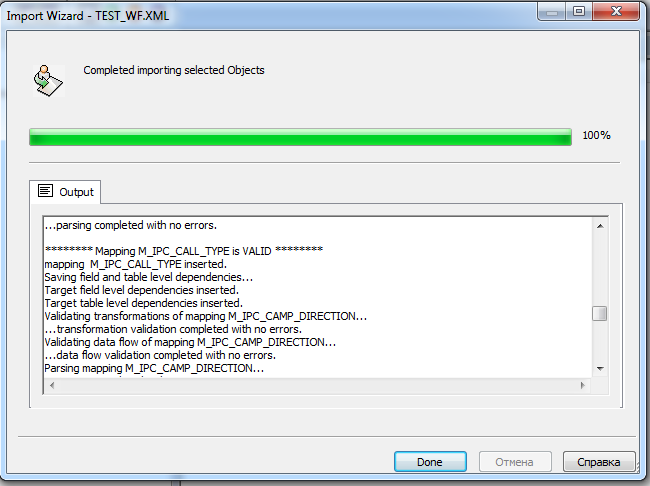


**!! Очень важно при выполнении команды Replace понимать, где и в каких потоках используется объект иначе есть вероятность поломать уже существующие потоки, которые используют этот объект . Поэтому пункт Replace не используем. Вместо нее переименовываем объект.**

Далее нажимаем Close и нажимаем Import.



После окончания импорта потоков необходимо проверить, что нет никаких ошибок и предупреждений. Для этого все логи в разделе Output копируем в текстовый редактор и ищем наличие Error и Warning.

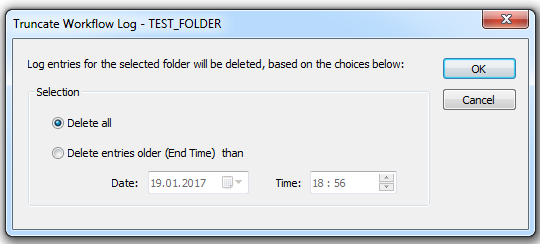


Далее нажимаем DONE. Импорт успешно завершен.

## 7.3 Чистка логов с помощю Truncate Log

Инструмент чистка логов позволяет чистить логи в объектах метаданны в DWH.

Для можно нажать Edit – Truncate Log



Далее выбирается дата ниже которой выполняется чистка.

!Вручную использовать чистку логов нет необходимости, потому что в каждом репозитории развернут поток Workflow Truncate Log выполняющий чистку.