

# 1 Описание базы данных

## Термины, которые необходимо определить до этого

Выборка, системное поле, первичный ключ, внешний(вторичный) ключ, репозиторий, ftp-server, система(), пользователь, администратор системы, привилегированный пользователь, цепочка расчетов, конфигурационный файл скрипта.

## Таблица Users

Назначение таблицы.

Предназначена для хранения информации о пользователях репозитория QSAR: имени пользователя в системе, пароля пользователя(храниться в незашифрованном виде), e-mail пользователя, на который отправляются уведомления, информации о том, что пользователь является администратором системы или привилегированным. Также является таблицей, содержащей информацию о пользователях, имеющих доступ на ftp-сервер.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер пользователя, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле UserName - имя пользователя в системе, состоящее из 40 символов(так же является именем для входа в систему).
3. Поле FirstName - фамилия пользователя для обращения, состоит из 40 символов.
4. Поле SecondName - имя пользователя для обращения, состоит из 40 символов.
5. Поле Password - пароль пользователя для доступа в систему, состоит из 40 символов.
6. Поле Email - e-mail пользователя для получения уведомлений
7. Поле Admin - флаг, является ли данный пользователь администратором системы.
8. Поле Premium - флаг, является ли данный пользователь привилегированным.

## **Таблица Molecules**

Назначение таблицы.

Предназначена для хранения информации об описаниях молекул: номенклатурное название молекулы, smiles для данной молекулы, путь до файла с описанием, дату внесения описания в систему. Также таблица содержит ссылку на информацию о пользователе, внесшего это описание в систему.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер молекулы, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле User\_ID - идентификационный номер пользователя, внесшего молекулу в систему(системное поле, является внешним ключом для связи с полем ID таблицы Users).
3. Поле Molecule Name - номенклатурное название молекулы, состоит из 1000 символов.
4. Поле Path To File - полный путь до файла с описанием на сервере, состоящий из 255 символов.
5. Поле Input Date - дата внесения описания молекулы в систему.
6. Поле Smiles - содержит значение smiles данной молекулы, состоит из 20000 символов

## **Таблица ActivityType**

Назначение таблицы.

Предназначена для хранения информации о типах активностей, расчеты которых могут быть произведены при помощи данных, содержащихся в системе. Также таблица содержит ссылку на информацию о пользователе, внесшего информацию об активности в систему.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер типа активности, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Activity Name - имя типа активности в системе, состоящее из 40 символов.

3. Поле User\_ID - идентификационный номер пользователя, внесшего тип активности в систему(системное поле, является внешним ключом для связи с полем ID таблицы Users).
4. Поле Input Date - дата внесения типа активности в систему.

### **Таблица Activity**

Назначение таблицы.

Предназначена для хранения значений активности заданного типа для молекул, чье описание было введено в систему в качестве обучающей выборки(значение активности взято из внешнего источника) или для которых расчеты были произведены при данных, содержащихся в системе. Для активностей, принадлежащих описанию молекулы из обучающей выборки, обязательно наличие данных об источнике значения активности. В случае активности, чье значение было рассчитано при помощи данных системы, в качестве информации об источнике в таблицу будет помещаться значение "QSAR".

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер активности, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Molecules\_ID - идентификационный номер описания молекулы, значением активности которого является данная(системное поле, является внешним ключом для связи с полем ID таблицы Molecules).
3. Поле ActivityType\_ID - идентификационный номер типа активности, которому принадлежит данное значение(системное поле, является внешним ключом для связи с полем ID таблицы ActivityType).
4. Поле Value - значение активности, действительное число.
5. Поле Info - данные об источнике информации об активности, строка, состоящая из 256 символов.

### **Таблица Selections**

Назначение таблицы.

Предназначена для хранения информации о выборках, сформированных пользователями: ссылки на описания молекул, включенных в выборку, ссылку на тип активности, расчет значений которой будет производиться при обработке данной выборки, а также имя выборки в системе.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер выборки, состоящий из 40 символов (системное поле, первичный ключ).
2. Поле Molecules\_ID - идентификационный номер молекулы, значением активности которой является данная (системное поле, является внешним ключом для связи с полем ID таблицы Molecules).
3. Поле ActivityType\_ID - идентификационный номер типа активности, для расчета которой сформирована выборка (системное поле, является внешним ключом для связи с полем ID таблицы ActivityType).
4. Поле Selection Name - имя выборки в системе, состоящее из 40 символов.
5. Поле Input Date - дата формирования выборки.

### **Таблица MoleculesInSelections**

Назначение таблицы.

Это вспомогательная таблица. Она предназначена для хранения информации о вхождении описания молекулы в ту или иную выборку.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер соответствия, состоящий из 40 символов (системное поле, первичный ключ).
2. Поле Molecules\_ID - идентификационный номер молекулы (системное поле, является внешним ключом для связи с полем ID таблицы Molecules).
3. Поле Selections\_ID - идентификационный номер выборки (системное поле, является внешним ключом для связи с полем ID таблицы Selections).

### **Таблица Chains**

Назначение таблицы.

Предназначена для хранения списка цепочек расчетов, сформированных пользователями системы.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер цепочки, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Name - имя цепочки в системе, состоящий из 70 символов.

### **Таблицы ConvertMolecules**

Назначение таблицы.

Предназначена для хранения информации о промежуточных результатах преобразования mol файла в формат, удобный для поиска особых точек: ссылки на этап цепочки расчетов, на котором этот результат был получен, ссылки на выборку, которая обрабатывалась в данном расчете.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер результата, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Path To File - полный путь до файла с результатом на сервере, состоящий из 255 символов. Может быть пустым, в случае если сохранялась только информация о промежуточных результатах.
3. Поле ScriptsInChains\_ID - идентификационный номер скрипта в цепочке, с помощью которой этот результат был получен результат(системное поле, является внешним ключем для связи с полем ID таблицы ScriptsInChains).
4. Поле Selection\_ID - идентификационный номер выборки, которая обрабатывалась(системное поле, является внешним ключем для связи с полем ID таблицы Selections).
5. Поле Input Date - дата внесения результатов в систему.

### **Таблицы SpecialPoints**

Назначение таблицы.

Предназначена для хранения информации о промежуточных результатах поиска особых точек: ссылки на этап цепочки расчетов, на котором этот результат был получен, ссылки на выборку, которая обрабатывалась в данном расчете.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер результата, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Path To File - полный путь до файла с результатом на сервере, состоящий из 255 символов. Может быть пустым, в случае если сохранялась только информация о промежуточных результатах.
3. Поле ScriptsInChains\_ID - идентификационный номер скрипта в цепочке, с помощью которой этот результат был получен результат(системное поле, является внешним ключем для связи с полем ID таблицы ScriptsInChains).
4. Поле Selection\_ID - идентификационный номер выборки, которая обрабатывалась(системное поле, является внешним ключем для связи с полем ID таблицы Selections).
5. Поле Input Date - дата внесения результатов в систему.

### **Таблицы Classification**

Назначение таблицы.

Предназначена для хранения информации о промежуточных результатах маркировки и кластеризации особых точек: ссылки на этап цепочки расчетов, на котором этот результат был получен, ссылки на выборку, которая обрабатывалась в данном расчете..

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер результата, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Path To File - полный путь до файла с результатом на сервере, состоящий из 255 символов. Может быть пустым, в случае если сохранялась только информация о промежуточных результатах.

3. Поле ScriptsInChains\_ID - идентификационный номер скрипта в цепочке, с помощью которой этот результат был получен результат(системное поле, является внешним ключем для связи с полем ID таблицы ScriptsInChains).
4. Поле Selection\_ID - идентификационный номер выборки, которая обрабатывалась(системное поле, является внешним ключем для связи с полем ID таблицы Selections).
5. Поле Input Date - дата внесения результатов в систему.

### **Таблицы Discriptors**

Назначение таблицы.

Предназначена для хранения информации о промежуточных результатах формирования списка дескрипторов: ссылки на этап цепочки расчетов, на котором этот результат был получен, ссылки на выборку, которая обрабатывалась в данном расчете.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер результата, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Path To File - полный путь до файла с результатом на сервере, состоящий из 255 символов. Может быть пустым, в случае если сохранялась только информация о промежуточных результатах.
3. Поле ScriptsInChains\_ID - идентификационный номер скрипта в цепочке, с помощью которой этот результат был получен результат(системное поле, является внешним ключем для связи с полем ID таблицы ScriptsInChains).
4. Поле Selection\_ID - идентификационный номер выборки, которая обрабатывалась(системное поле, является внешним ключем для связи с полем ID таблицы Selections).
5. Поле Input Date - дата внесения результатов в систему.

### **Таблицы Matrix**

Назначение таблицы.

Предназначена для хранения информации о промежуточных результатах формирования матрицы из дескрипторов: ссылки на этап цепочки расчетов, на котором этот результат был получен, ссылки на выборку, которая обрабатывалась в данном расчете.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер результата, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Path To File - полный путь до файла с результатом на сервере, состоящий из 255 символов. Может быть пустым, в случае если сохранялась только информация о промежуточных результатах.
3. Поле ScriptsInChains\_ID - идентификационный номер скрипта в цепочке, с помощью которой этот результат был получен результат(системное поле, является внешним ключем для связи с полем ID таблицы ScriptsInChains).
4. Поле Selection\_ID - идентификационный номер выборки, которая обрабатывалась(системное поле, является внешним ключем для связи с полем ID таблицы Selections).
5. Поле Input Date - дата внесения результатов в систему.

### **Таблица FilesFormatList**

Назначение таблицы.

Предназначена для хранения информации о типах файлов, используемых в расчетах: условное имя для данного типа файлов и шаблон, под который подпадает имя файлов данного типа.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер типа файла, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Name - имя типа файла в системе, состоящий из 25 символов.
3. Поле Templates - шаблон названия файла, состоящий из 40 символов.
4. Поле DataTable - ???

## Таблица ScriptList

Назначение таблицы.

Предназначена для хранения информации о скриптах, внесенных в систему, с помощью которых совершаются расчеты на тех или иных этапах цепочки: имя скрипта, имя для выгрузки из репозитория, версия(в рамках скриптов с одинаковым названием), ссылки на входные и выходные типы файлов, ссылки на пользователя, внешнего скрипт в систему. Также содержит информацию о том, является ли данный скрипт исполнимым под OS Linux.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер скрипта, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Name - имя скрипта в системе, состоящий из 25 символов.
3. Поле SVN Name - имя для загрузки скрипта из системы контроля версий, состоящий из 255 символов.
4. Поле Version - версия скрипта, существующая в системе контроля версий, целое положительное число.
5. Поле Input File Format\_ID - формат файлов с входными данными(системное поле, является внешним ключем для связи с полем ID таблицы FilesFormatList).
6. Поле Output File Format\_ID - формат файлов с выходными данными(системное поле, является внешним ключем для связи с полем ID таблицы FilesFormatList).
7. Поле Linux Script - флаг, обозначающий возможность выполнения скрипта под ОС linux.
8. Поле Input Date - дата внесения скрипта в систему.
9. Поле User\_ID - идентификационный номер пользователя, внесшего скрипт в систему(системное поле, является внешним ключем для связи с полем ID таблицы Users).

### **Таблица ConfigList**

Назначение таблицы.

Предназначена для хранения информации о конфигурационных файлах к скриптам, с параметрами(данными) из которых осуществляется запуск этапов из цепочки расчетов.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер конфигурационного файла, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Name - имя конфигурационного файла в системе, состоящий из 40 символов.
3. Поле SVN Name - имя для загрузки конфигурационного файла из системы контроля версий, состоящий из 40 символов.
4. Поле Version - версия конфигурационного файла, существующая в системе контроля версий, целое положительное число.

### **Таблица Chains**

Назначение таблицы.

Предназначена для хранения информации о цепочках расчетов, сформированных и зарегистрированных в системе.

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер цепочки, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле Name - имя цепочки в системе, состоящий из 40 символов.

### **Таблица ScriptsInChains**

Назначение таблицы.

Предназначена для хранения информации о вхождении определенного расчетного скрипта в цепочку на определенном ее этапе: ссылки на скрипт, с помощью которого будет производиться расчет на данном этапе цепочки, ссылки на конфигурационный файл, с параметрами из которого скрипт будет производить расчет в рамках данной цепочки, ссылки на цепочку, в рамках которой будет производиться расчет, строку для запуска скрипта, и номер этапа расчета в цепочке .

Описание полей таблицы.

1. Поле ID - уникальный идентификационный номер соответствия, состоящий из 40 символов(системное поле, первичный ключ).
2. Поле ConfigList\_ID - идентификационный номер конфигурационного файла с которым связан скрипт(системное поле, является внешним ключем для связи с полем ID таблицы ConfigList).
3. Поле Script\_ID - идентификационный номер скрипта(системное поле, является внешним ключем для связи с полем ID таблицы Scripts).
4. Поле Chain\_ID - идентификационный номер цепочки, к которой принадлежит скрипт(системное поле, является внешним ключем для связи с полем ID таблицы Chainy).
5. Поле Etap Count - порядок скрипта в цепочке, натуральное число.