

# Описание процессов поддержания жизненного цикла Neoflex Datagram

---

## 1. Общие сведения

---

Настоящий документ призван обеспечить поддержку жизненного цикла Neoflex Datagram, в том числе устранение неисправностей, выявленных в ходе эксплуатации программного обеспечения, обновление функционала, описание требований к персоналу, необходимом для обеспечения эксплуатации Neoflex Datagram. Также данный документ регламентирует взаимодействие Заказчика и АО «Неофлекс консалтинг» в процессе эксплуатации программного обеспечения.

## 2. Процессы, обеспечивающие жизненный цикл Neoflex Datagram

---

### 2.1 Поставка Neoflex Datagram

---

Поставка включает в себя электронный носитель, содержащий:

- Инсталляционные файлы Neoflex Datagram;
- Электронный документ «Руководство по установке Neoflex Datagram»;
- Электронный документ «Руководство пользователя Neoflex Datagram».

### 2.2 Установка Neoflex Datagram

---

Установка Neoflex Datagram описывается в отдельном документе «Руководство по установке Neoflex Datagram».

Требования к уровню квалификации специалистов, выполняющих установку Neoflex Datagram:

- Базовые знания администрирования CentOS/Linux.

### 2.3 Эксплуатация Neoflex Datagram

---

Использование Neoflex Datagram требует выполнения следующих видов работ:

- Настройка серверной части программного продукта.

Требования к уровню квалификации специалистов, эксплуатирующих программный продукт:

- Знание SQL, Java, Scala.

### 2.4 Обучение использованию Neoflex Datagram

---

Обучение специалистов возможностям Neoflex Datagram может выполняться:

- Самостоятельно с использованием прилагаемой документации;
- На учебных курсах, организуемых разработчиком АО «Неофлекс консалтинг» (по дополнительному согласованию).

## 2.5 Сопровождение и поддержка

---

Сопровождение и поддержка, выполняемые на этапе эксплуатации включают:

- Консультации по телефону и электронной почте;
- Устранение обнаруженных ошибок при работе с Neoflex Datagram;
- Услуги по модификации и добавлению новых функций в Neoflex Datagram;
- Услуги по настройке Neoflex Datagram на территории Заказчика.

## 2.6 Настройка и добавление новых функциональных возможностей Neoflex Datagram

---

- Услуги по разработке и добавлению новых функций программного обеспечения;
- Настройка Neoflex Datagram для выполнения профильных задач заказчика.

## Приложение

---

Neoflex Datagram - это программная платформа предназначенная для разработки приложений по преобразованию данных. Neoflex Datagram поддерживает как пакетный, так и потоковый режимы обработки данных.

Ядром платформы является сервер метаданных, который обеспечивает хранение и предоставляет инструменты управления хранилищами данных, преобразованиями данных, источниками и приемниками данных, исполняющими средами и т.д.

### Общая информация.

- Поддержка операционных систем на базе Linux;
- Реализация работы программы через веб-интерфейс браузеров Mozilla Firefox и Google Chrome.

### Дополнительная информация

Neoflex Datagram поддерживает полный цикл разработки приложений по преобразованию данных:

- Визуальное проектирование схем преобразования данных;
- Визуальное проектирование потоков управления преобразованиями данных;
- Генерация исходного кода на языке Scala с библиотекой Apache Spark;
- Компиляция и генерация приложения;
- Развертывание приложения на исполняющей среде;
- Планирование исполнения приложения;
- Мониторинг исполнения приложения;
- Инструменты для остановки и перезапуска приложений.

### Устройство программной платформы

Программная платформа разработана на базе архитектуры, управляемой моделью (MDA). Для управления моделями применяется Eclipse Modelling Framework (EMF). Для сохранения моделей применяются PostgreSQL, Hibernate и Teneo. Для валидации моделей и трансформаций model-to-model (M2M) и model-to-text (M2T) применяется Eclipse Epsilon. Внутри платформы используются следующие типы моделей: Authentication, Relational, ETL, Runtime, DWH, UI, Metadata.

## **Исполняющие среды**

Исполняющие среды Neoflex Datagram базируются на Apache Spark.

Neoflex Datagram может выполнять запуск приложений на серверах Apache Livy или Apache Oozie. На сервере Apache Livy запускаются приложения из сред разработки/отладки. Сервер Apache Oozie используется для запуска отложенных приложений на рабочей среде.

## **Дизайнер трансформаций**

Дизайнер трансформаций - интерфейс для визуальной разработки схем преобразования данных.

В дизайнере трансформаций поддерживается широкий спектр источников/приемников данных:

- RDBMS источники/приемники данных использующие соединение JDBC (включая хранимые процедуры);
- Иерархические источники/приемники: XML, AVRO и JSON;
- Специфические форматы файловой системы HDFS: ORC, PARQUET;
- Источники/приемники данных: CSV, Apache Hive, Apache Kafka.

Типы преобразований данных:

- Широкий набор операций реляционной алгебры: join, sort, aggregation, union, selection, projections, pivot, explode arrays, sequence generation;
- Специфические для Spark трансформации: Spark SQL - выполняет произвольные SQL запросы к потокам данных;
- Алгоритмы машинного обучения с использованием Spark MLlib (decision trees, SVM, logistic regression и т.д.);
- Jboss Rules (Drools) - система управления бизнес правилами.

Основные возможности:

- Поддержка типов данных полей: STRING, DECIMAL, INTEGER, DATE, TIME, DATETIME, BINARY, BOOLEAN, LONG, FLOAT, DOUBLE;
- Поддержка типов данных STRUCT и ARRAY;
- Отслеживание происхождения полей потока данных;
- Частичное выполнение преобразования с просмотром промежуточных результатов;
- Просмотр сгенерированного кода приложения, его редактирование и запуск на исполнение;
- Валидация трансформации на основе базы данных часто повторяемых ошибок;
- Поддержка Spark Catalyst Optimizer.

## **Дизайнер Workflow**

Дизайнер Workflow - интерфейс для визуальной разработки потоков управления последовательностями преобразований данных.

Основные возможности:

- Создание потоков управления для параллельного или последовательного исполнения преобразований данных, а также потоков управления с возможностью настройки условий для запуска преобразований;
- Универсальные элементы управления преобразованиями: shell scripts и java scripts;
- Возможность создания потоков управления последовательностями преобразований с использованием вложенных объектов Workflow;
- Возможность настройки исполнения Workflow по расписанию или по событиям файловой системы.

## **Безопасность**

- Централизованная аутентификация пользователей с использованием корпоративного сервера каталогов (LDAP);
- Ролевая авторизация. Возможные роли: developer, operator, viewer;
- Шифрование паролей доступа к внешним системам;
- Использование алгоритма аутентификации Kerberos для подключения к исполняющим средам.

## **Версионность и teamwork**

- Блокировка одновременных обновлений;
- Интеграция Apache Subversion;
- Поддержка иерархии проектов;
- Синхронизация с системой контроля версий TortoiseSVN для выбранного объекта или проекта;
- Защищенный от обновлений код (определяемый пользователем) сохраняется при обновлении версии метаданных.

## **Поддержка рабочих сред**

Поддержка цикла разработки: разработка->тестирование->Запуск на рабочей среде.

- Импорт/экспорт метаданных;
- Перенос метаданных между средами как полный, так и отдельного проекта;
- Перезапись URL-адресов, паролей и т.д. при переносе в новую среду.

## **Дополнительные инструменты**

- Консоль HDFS: просмотр, сохранение файлов из/в файловой системы HDFS;
- Консоль Livy: просмотр задач на сервере Livy, просмотр журналов, отмена задачи;
- Консоль Oozie: обзор задач workflow и координатора на сервере Oozie, просмотр журналов, отмена или перезапуск задач;
- Обзорщик объектов: просмотр дерева объектов метаданных.