

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 1. Описание программного обеспечения

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 работы с операционными системами на базе Linux (в том числе с российскими аналогами: AlterOS (рег. номер ПО: 3801), АЛЪТ 8 СП (рег. номер ПО: 4305));
- Реализация работы программы через веб-интерфейс браузеров Yandex (рег. номер ПО: 3722), 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 (open source);
- 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 сохранения моделей применяются Postgres Pro (рег. номер ПО: 104, сайт разработчика: <https://postgrespro.ru>) или PostgreSQL (open source, сайт разработчика: <https://www.postgresql.org>), Hibernate (open source, сайт разработчика: <http://hibernate.org>).

Исполняющие среды

Исполняющие среды Neoflex Datagram базируются на Arenadata Hadoop (рег. номер ПО: 4074, сайт разработчика: <https://arenadata.tech>).

Neoflex Datagram может выполнять запуск приложений на серверах Livy или Oozie. На сервере Livy запускаются приложения из сред разработки/отладки. Сервер Oozie используется для запуска отлаженных 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, Hive, 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;
- Возможность интеграции с Apache Subversion;
- Поддержка иерархии проектов;
- Возможность синхронизации 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.

Приложение 2. Неисправности и способы их устранения

Ошибка идентификации пользователя в программе.

Описание неисправности:

Пользователь не может выполнить вход в программу.

Способ устранения неисправности:

Проверьте параметры подключения к серверу LDAP, указанные в файле `ldap.properties`:

Параметр	Обязательно заполнять	Описание
<code>ldap.domain</code>	Да	Доменное имя Ldap сервера. <i>Для авторизации по LDAP используется <code>userPrincipalName</code> вида: username@domainname.com. Если поле не заполнено, то при авторизации необходимо указывать <code>userPrincipalName</code> полностью. Если <code>domain</code> указан, то допускается ввод только <code>userName</code></i> <i>Пример: <code>ldap.domain=ldapServer.ru</code></i>
<code>ldap.host</code>	Да	Имя хоста Ldap сервера <i>Пример: <code>ldap.host=msk-ldserv1.company.ru</code></i>
<code>ldap.port</code>	Да	Порт Ldap сервера <i>Пример: <code>ldap.port=789</code></i>
<code>ldap.base</code>	Да	Путь к каталогу для поиска пользователей <i>Пример: <code>ldap.base=CN=Users,DC=company,DC=ru</code></i>
<code>ldap.admin</code>	Да	Имя группы пользователей, которым будут предоставлены права администратора
<code>ldap.operator</code>	Да	Имя группы пользователей, которым будут предоставлены права оператора
<code>ldap.user</code>	Да	Имя группы пользователей, которым будут предоставлены права пользователя

Ошибки, возникающие при выполнении трансформаций данных

Batch finished with status dead

Описание неисправности:

Выполнение трансформации данных оканчивается сообщением: "Batch finished with status dead See Batches in Livy Console for details".

Способ устранения неисправности:

- Если источником данных является файл (.csv, .xml, .avro), то необходимо проверить:

1. Корректность данных в поле "Path" в настройках элемента трансформации;
 2. Если файл хранится в файловой системе HDFS, то необходимо проверить настройки доступа к HDFS в объекте Livy Server. А также наличие ошибок и статус HDFS в Arenadata Hadoop.
- Если источником данных является таблица, то необходимо проверить:
 1. Корректность выбора объекта Context в настройках элемента;
 2. Настройки объектов JDBC Connection, Software System, Deployment, которые используются для создания подключения к базе данных.
 - Если источником данных является Kafka, то необходимо проверить параметры подключения к серверу Kafka в настройках элемента трансформации Kafka Source.

Connection refused

Описание неисправности:

Выполнение трансформации данных оканчивается сообщением: "Connection refused".

Способ устранения неисправности:

Проверьте настройки объектов Transformation deployment, который разворачивает трансформацию, и связанного с ним объекта Livy server.

Ошибки, возникающие при работе с проектами

file(s) unzipped. 0 object(s) imported

Описание неисправности:

Выполнение загрузки архива проекта оканчивается сообщением: "file(s) unzipped. 0 object(s) imported".

Способ устранения неисправности:

Проверьте структуру архива проекта. Файлы проекта должны находиться в корневом каталоге архива.

Ошибки, возникающие при работе с датасетами.

null pointer exception

Описание неисправности:

При выполнении параграфа в notebook появляется ошибка: "null pointer exception".

Способ устранения неисправности:

Проверьте значение Cluster в атрибутах workspace.