

Астра ИС МД (Инфраструктурные Сервисы)

Модуль Astra Apache Airflow

**Документация, содержащая информацию, необходимую для
эксплуатации экземпляра программного обеспечения,
предоставленного для проведения экспертной проверки**

Содержание

1 Область применения.....	3
1.1 Архитектура системы.....	3
2 Подключение к модулю Astra Apache Airflow	4
2.1 Доступ к Apache Airflow.....	4
2.2 Доступ через CLI.....	4
2.3 Работа с DAG.....	4
3 Основные команды Airflow.....	5
3.1 Запуск DAG:	5
3.2 Приостановка DAG:.....	5
3.3 Возобновление DAG:.....	5
3.4 Управление задачами.....	5
4 Мониторинг и статистика.....	6
4.1 Мониторинг	6
4.2 Просмотр логов.....	6
5 Резервное копирование и восстановление.....	7
5.1 Резервное копирование.....	7
5.2 Восстановление из копии:	7
6 Конфигурация.....	8
6.1 Основные параметры.....	8
Приложение А Часто задаваемые вопросы.....	9
Приложение Б Инструкция по установке экземпляра программного обеспечения	
10	
Приложение В Инструкция по проверке экземпляра программного обеспечения	13
Приложение Г Инструкция по подключению к стенду AstraAA (Standalone)	14

1 Область применения

Настоящий документ содержит описание функциональных характеристик экземпляра программного обеспечения, модуль Apache Airflow, далее Apache Airflow, развернутого для экспертной проверки.

Система используется для планирования задач, управления рабочими процессами и мониторинга выполнения задач.

Документ предназначен для системных администраторов и разработчиков, осуществляющих эксплуатацию и обслуживание системы.

1.1 Архитектура системы

1.1.1 Программное обеспечение

Apache Airflow

Версия: 3.1.7

Назначение: платформа оркестрации и планирования рабочих процессов (DAG)

1.1.2 Компоненты системы

- Airflow Webserver (веб-интерфейс управления)
- Scheduler (планировщик задач)
- Executor: LocalExecutor
- Metadata Database: SQLite
- DAG каталоги: ~/airflow/dags
- Логи: ~/airflow/logs

1.2. Характеристики системы

Таблица 1—Характеристики системы

Параметр	Значение
Версия Airflow	3.1.7
Web интерфейс порт	8080
Executor	LocalExecutor
База метаданных	SQLite
Каталог DAG	~/airflow/dags
Каталог логов	~/airflow/logs

2 Подключение к модулю Astra Apache Airflow

2.1 Доступ к Apache Airflow

Web интерфейс продукта доступен по адресу:

http://localhost:8080

Вход выполняется с использованием учетной записи администратора.

2.2 Доступ через CLI

Основные команды:

```
airflow version  
airflow dags list  
airflow dags show <dag_id>
```

2.3 Работа с DAG

DAG файлы размещаются в каталоге:

~/airflow/dags

После добавления файла DAG он автоматически подхватывается scheduler.

3 Основные команды Airflow

3.1 Запуск DAG:

```
airflow dags trigger <dag_id>
```

3.2 Приостановка DAG:

```
airflow dags pause <dag_id>
```

3.3 Возобновление DAG:

```
airflow dags unpause <dag_id>
```

3.4 Управление задачами

3.4.1 Просмотр задач:

```
airflow tasks list <dag_id>
```

3.4.2 Ручной запуск задачи:

```
airflow tasks run <dag_id> <task_id> <execution_date>
```

Управлять DAG можно используя интерфейс пользователя (UI)

4 Мониторинг и статистика

4.1 Мониторинг

Мониторинг выполняется через Web UI.

Доступны представления:

- DAG Dashboard
- Graph View
- Grid View
- Task Logs

4.2 Просмотр логов

Логи хранятся в каталоге:

~/airflow/logs

5 Резервное копирование и восстановление

5.1 Резервное копирование

Файл базы SQLite:

~/airflow/airflow.db

5.1.1 Создание копии:

```
cp ~/airflow/airflow.db backup_airflow.db
```

5.2 Восстановление из копии:

```
pkill -f airflow  
cp backup_airflow.db ~/airflow/airflow.db
```

6 Конфигурация

Основной файл конфигурации:

~/airflow/airflow.cfg

6.1 Основные параметры

```
executor
```

```
dags_folder
```

```
base_log_folder
```

```
sql_alchemy_conn
```

Приложение А Часто задаваемые вопросы

Как добавить DAG

Поместить файл в ~/airflow/dags

Как проверить систему

```
airflow dags list
```

Где находятся логи

~/airflow/logs

Приложение Б Инструкция по установке экземпляра программного обеспечения

Версия экземпляра: – 3.1.7

Тип развертывания: – односерверный режим (LocalExecutor + SQLite)

Назначение: – программная платформа для оркестрации, планирования и мониторинга рабочих процессов (DAG).

Системные требования

- 2–4 vCPU
- 4–8 ГБ оперативной памяти
- не менее 10 ГБ свободного дискового пространства
- Программные требования:
- Операционная система Linux (Ubuntu 22.04 / Debian 12 / Astra Linux)
- Python версии 3.10 или 3.11
- Пакеты python3-venv, python3-dev, build-essential

Программные требования

- Операционная система Linux (Ubuntu 22.04 / Debian 12 / Astra Linux)
- Python версии 3.10 или 3.11
- Пакеты python3-venv, python3-dev, build-essential

Подготовка среды

Перед установкой необходимо обновить систему:

```
sudo apt update  
sudo apt upgrade -y
```

Установить необходимые зависимости:

```
sudo apt install -y python3-venv python3-dev build-essential libssl-dev  
libffi-dev
```

Создание отдельного пользователя для запуска продукта

```
sudo useradd -m airflow  
sudo passwd airflow  
sudo su - airflow
```

Создание виртуального окружения

Создать виртуальное окружение Python:

```
python3 -m venv ~/airflow_venv
```

Активировать его:

```
source ~/airflow_venv/bin/activate
```

Обновить менеджер пакетов pip:

```
pip install --upgrade pip
```

Установка Apache Airflow 3.1.7

Задать переменные среды:

```
export AIRFLOW_VERSION=3.1.7

export PYTHON_VERSION=$(python -c "import sys;

print(f'{sys.version_info.major}.{sys.version_info.minor}')")

export

CONSTRAINT_URL="https://raw.githubusercontent.com/apache/airflow/constraints-
${AIRFLOW_VERSION}/constraints-${PYTHON_VERSION}.txt"
```

Выполнить установку:

```
pip install "apache-airflow==${AIRFLOW_VERSION}" --constraint
"${CONSTRAINT_URL}"
```

Настройка Apache Airflow

Задать рабочую директорию:

```
export AIRFLOW_HOME=~/.airflow

mkdir -p $AIRFLOW_HOME
```

Инициализировать базу данных (будет автоматически создан файл SQLite):

```
airflow db migrate
```

Настройка режима выполнения

Открыть файл конфигурации:

```
nano $AIRFLOW_HOME/airflow.cfg
```

Найти параметр:

```
executor = SequentialExecutor
```

И заменить его на:

```
executor = LocalExecutor
```

Проверить строку подключения к базе данных. Она должна использовать SQLite, например:

```
sql_alchemy_conn = sqlite:///home/airflow/airflow/airflow.db
```

Путь может отличаться в зависимости от значения AIRFLOW_HOME.

Создание учетной записи администратора

Выполнить команду:

```
airflow users create --username admin --firstname Admin --lastname User
--role Admin --email admin@example.com
```

Запуск компонентов системы

Для работы Airflow необходимо запустить планировщик и веб-сервер.

Запуск планировщика (scheduler):

```
airflow scheduler
```

В отдельной сессии терминала запустить веб-интерфейс:

```
airflow webserver --port 8080
```

После запуска доступ к веб-интерфейсу осуществляется по адресу:

http://<имя_сервера>:8080

Приложение В Инструкция по проверке экземпляра программного обеспечения

Цель проверки — подтверждение корректности установки и работоспособности экземпляра.

Проверка запуска сервисов

```
ps aux | grep airflow
```

Должны быть запущены процессы:

- airflow scheduler
- airflow webserver
- airflow worker (при использовании CeleryExecutor)

Проверка подключения к базе данных

```
airflow db check
```

При корректной настройке ошибки подключения отсутствуют.

Проверка выполнения тестового DAG

Создать файл `$AIRFLOW_HOME/dags/expert_test_dag.py` со следующим содержанием:

```
from airflow import DAG
from airflow.operators.bash import BashOperator
from datetime import datetime

with DAG(dag_id='expert_test_dag', start_date=datetime(2024,1,1),
schedule=None, catchup=False) as dag:
    task = BashOperator(task_id='print_date', bash_command='date')
```

Активировать DAG через веб-интерфейс и выполнить вручную. Ожидаемый результат: статус выполнения — Success.

Проверка логирования

Проверить каталог `$AIRFLOW_HOME/logs/`.

Должны присутствовать журналы выполнения DAG и задач.

Критерии успешной проверки

- Компоненты запускаются без ошибок.
- Веб-интерфейс доступен.

Приложение Г Инструкция по подключению к стенду AstraAA (Standalone)

Адрес веб-интерфейса

После запуска экземпляра Apache Airflow веб-интерфейс системы доступен через веб-браузер по адресу:

http://<адрес_сервера>:8080

При доступе непосредственно с сервера, на котором установлен Apache Airflow, можно использовать адрес:

http://localhost:8080

Подключение осуществляется с использованием веб-браузера

Аутентификация пользователя

В режиме Standalone при первом запуске Apache Airflow автоматически создаётся учетная запись администратора.

Для входа в веб-интерфейс используются следующие учетные данные:

Логин: admin

Пароль автоматически генерируется системой и выводится в консоли при запуске команды:

```
airflow standalone
```

После успешной аутентификации пользователь получает доступ к веб-интерфейсу Apache Airflow и может выполнять просмотр и управление рабочими процессами (DAG).