ООО «Астра Консалтинг»

Эксплуатация экземпляра программного обеспечения, модуль Astra Apache Airflow для Астра ИС МД

Документация, содержащая описание функциональных характеристик экземпляра программного обеспечения «Astra Apache Airflow», предоставленного для проведения экспертной проверки

Москва, 2026

1 Введение

Документ содержит описание функциональных характеристик программного обеспечения, модуль Astra Apache Airflow (далее Apache Airflow) ПО предоставляет графический интерфейс для создания, выполнения и мониторинга рабочих процессов, позволяет создавать рабочие процессы в форме направленных ацикличных графов (DAG), где каждый узел представляет собой отдельную задачу.

2 Общие сведения о программном обеспечении

Программное обеспечение для создания, выполнения, мониторинга и оркестровки потоков операций «Astra Apache Airflow». содержит следующие модули:

- планировщик (Scheduler)
- executor (Исполнитель)
- база метаданных
- веб-сервер
- работники (Workers)
- триггер (Trigger)

1. Планировщик (Scheduler) — читает расписание каждого DAG и определяет, какие задачи должны быть запущены и когда. Например, планировщик может определить, что ежедневная задача обработки данных должна быть запущена в 11:00 каждый день.
2. Executor (Исполнитель) — компонент, который определяет, как именно задачи будут выполняться. Airflow предоставляет несколько типов исполнителей, которые могут работать в различных средах и конфигурациях.
3. База метаданных — хранит описательную информацию, а не фактическую. Данный тип баз используют для управления, организации и поиска данных. База метаданных в AirflowAA хранит информацию о задачах, их статусе, зависимостях и истории выполнения.
4. Веб-сервер — предоставляет пользовательский интерфейс для мониторинга, управления и запуска задач. Через веб-интерфейс пользователи могут просматривать список задач, проверять их статус и управлять расписанием выполнения.
5. Работники (Workers) — берут на себя выполнение заданий, распределяемых исполнителем.
6. Триггер (Trigger) — механизм, который позволяет запустить процесс при наступлении определённого события или условия, не обязательно связанного с расписанием.

3 Область применения

Модуль Astra Apache Airflow используется для обеспечения работоспособности высоконагруженных сервисов.

- ETL/ELT-конвейеры — оркестровка извлечения данных из сотен источников, их трансформация с помощью Spark/dbt и загрузка в облачные хранилища.
- Конвейеры машинного обучения (MLOps) — автоматизация всех этапов жизненного цикла ML-моделей — от подготовки данных и обучения до деплоя и мониторинга.
- Автоматизация отчётности — периодический запуск SQL-запросов, генерация дашбордов и рассылка PDF-отчётов по расписанию.
- Инфраструктурные задачи — автоматизация бэкапов, управление облачными ресурсами, запуск периодических health-check-систем.

Областью применения настоящего программного обеспечения являются любые сферы государственной или частной деятельности, автоматизирующие свою деятельность (использующие программное обеспечение при ведении деятельности).

4 Языки программирования

Исходный код «Astra Apache Airflow» написан на следующих языках:

- Python
- TypeScript react
- SQL

5 Общее описание функциональных характеристик

5.1 Основной функционал изделия

Программное обеспечение, модуль «Astra Apache Airflow» обладает следующими основными функциями:

- Оркестрация данных — система позволяет определить сложные рабочие процессы как направленные ациклические графы (DAG), где каждая задача имеет чётко определённые зависимости от других. Это обеспечивает правильный порядок выполнения операций и предотвращает возникновение циклических зависимостей.
- Автоматизация ETL-процессов — Airflow может извлекать данные из различных источников — от традиционных баз данных до облачных сервисов, трансформировать их с помощью Python-кода или SQL-запросов, а затем загружать в целевые системы.
- Планирование задач — с помощью гибкого планировщика, который поддерживает как временные расписания, так и запуск по событиям. Можно настроить ежедневные, еженедельные или более сложные расписания выполнения.
- Мониторинг выполнения — через веб-интерфейс, который позволяет отслеживать статус задач в реальном времени, анализировать историю выполнения и быстро выявлять проблемы в рабочих процессах.
- Поддержка отказоустойчивости — встроенные механизмы восстановления: если задача завершается с ошибкой, Airflow может автоматически повторить её выполнение согласно настроенной политике retry. История выполнения сохраняется в базе метаданных, что позволяет восстанавливать процессы после сбоев.

5.2 Базовые (основные) модули

Программное обеспечение, модуль «Astra Apache Airflow» содержит следующие основные модули:

- Планировщик (Scheduler) — отслеживает все задачи и управляет расписанием, определяет, когда каждое задание должно быть выполнено.
- Исполнитель (Executor) — отвечает за запуск задач, распределяет их по рабочим узлам и следит за тем, чтобы нагрузка распределялась равномерно.
- Веб-интерфейс (Web UI) — наглядно показывает структуру DAG, текущее состояние задач и логи выполнения. Через Web UI можно вручную перезапускать задачи, настраивать расписание или проводить диагностику проблем.
- Работники (Workers) — берут на себя выполнение заданий, распределяемых исполнителем.
- Триггер (Trigger) — механизм, который позволяет запустить процесс при наступлении определённого события или условия, не обязательно связанного с расписанием.

6 6. Используемые технические средства

6.1 Аппаратные требования

6.1.1 Требования к серверу для «Astra Apache Airflow»

Процессор:

- Минимальные: 2 ГГц, 2 ядра
- Рекомендуемые: 4 ГГц, 4 ядра

Оперативная память:

- Минимальные: 4 Гб
- Рекомендуемые: 8 Гб

Дисковое пространство:

- Минимальные: 10 Гб (предпочтительнее SSD)
- Рекомендуемые: 50 Гб (предпочтительнее SSD)

6.1.2 Требования к АРМ:

Процессор:

- Минимальные: 2,4 ГГц, 2 ядра
- Рекомендуемые: 4 ГГц, 4 ядра

Оперативная память:

- Минимальные: 8 Гб
- Рекомендуемые:16 Гб

Дисковое пространство:

- Минимальные: 50 Гб (SSD)
- Рекомендуемые: 100 Гб (SSD)

Монитор:

- Диагональ: не менее 16”
- Разрешение экрана: не менее 1024*768

Скорость подключения к сети: не менее 2 Мбит/с.

6.2 Программные требования

6.2.1 Операционные системы:

- Операционная система Linux (Ubuntu 22.04 / Debian 12 / Astra Linux)

- Python версии 3.10 или 3.11
- Пакеты python3-venv, python3-dev, build-essential

Рекомендованная ОС:

Astra Linux Special Edition — перед установкой выполните инструкции из статьи Подготовка сервера с ОС Astra Linux.

Использование на других платформах может потребовать локальных изменений.

База данных:

Astra Apache Airflow требует базу данных для хранения своих данных. Рекомендуется использовать PostgreSQL.

АРМ:

Для работы с системой требуется наличие одного из браузеров:

- Google Chrome, версия не ниже 109
- Яндекс.Браузер, версия не ниже 23.7.2.767 (64-bit)
- Mozilla FireFox, версия не ниже 102

Примечание: Astra Apache Airflow может также использовать дополнительные ресурсы для выполнения задач и оркестрации процессов. Ресурсы будут зависеть от объема задач и рабочих процессов, которые необходимо автоматизировать.

7 Входные и выходные данные

Входными данными являются параметры конфигурации серверов, на которых будет реализована автоматизация.

Выходными данными могут быть:

- отчеты о завершении исполнения заданий;
- состояние системы.

8 Подготовка к работе

Состав работ по подготовке Astra Apache Airflow к использованию изложен в Инструкции по установке и настройке, включая описание параметров конфигурационных файлов.