

Айдар Алиев — Data Engineer



Инженер данных / Аналитик данных




Python | SQL | Linux

Airflow | PostgreSQL | pandas | PySpark | scikit-learn | Bash | R

Работаю из Москвы

 [Скачать полное резюме](#) |  [Веб-версия резюме](#)

-
- Более 14 лет опыта работы с данными
 - Промышленный опыт построения ETL на **Airflow** + **PostgreSQL** для крупных проектов (ВТБ)
 - Уверено владею **Linux**, **SQL** и **Python**, есть опыт CI/CD и контейнеризации
 - Ищу развитие в Data Engineering, также открыт к **ML Engineering** и **Data Analytics**

-
- **Jupyter** ( **Python**) **ноутбуки**  с проектами по анализу данных, машинному и глубокому обучению
 - **Проекты на PostgreSQL**  (сложные запросы, пользовательские типы данных, PL/pgSQL, плагины, проектирование БД, триггеры, оптимизация)
-

Иннотех, Группа компаний

Москва • Ноябрь 2022 — настоящее время

Ведущий российский интегратор IT-решений для корпоративного сектора

Разработчик ETL-процессов (Data Engineer)

Стек: ETL/ELT, **Python**, **SQL**, **Airflow**, **PostgreSQL**, PL/pgSQL, Spark, git, Confluence, Maven

Зона ответственности:

- Автоматизация ETL-процессов для проекта ВТБ на стеке **Airflow** + **PostgreSQL**.
- Рефакторинг, доработка и исправление ошибок внутренних фреймворков (**Python**, **SQL**).
- Обновление, создание новых и исправление ошибок в отчётах (**Airflow**, **PostgreSQL**, YAML, xml).
- Написание инструкций для разработчиков по алгоритмам работы с внутренними фреймворками (xml).
- Создание и регистрация .jar (Scala, Spark, Maven) модулей для загрузки больших (~100-150 Gb) таблиц в staging слой DWH команды.
- Поставка всех выполненных работ до production среды в рамках CI/CD процесса (git, sfera, Liquibase, YAML).

Примеры решённых задач:

- Автоматизировал обслуживание ODS слоя DWH (удаление устаревших данных), устранив необходимость в дополнительном железе (**Airflow**, **PostgreSQL**).
- Разработал DAG для ETL/ELT FDW таблиц объёмом до 700 млн строк (150 GB) в рамках DWH формирования регуляторной отчётности. Использовал PL/pgSQL процедуру, вызываемую из DAG'a **Airflow**.
- Внедрил новый этап логирования и модуль, повысив прозрачность и удобство поддержки процессов; спроектировал схему хранения логов, написал рутины на PL/pgSQL и триггеры, а также модуль **Python**.
- Создал модуль, используемый в 100+ DAG'ах **Airflow** команды; отрефакторил существующий код, применил подход DRY.
- Реализовал декоратор авторизации на **Python** с несколькими уровнями вложенности; используется ~в 90% DAG'ов команды.

Институт молекулярной генетики, НИЦ Курчатовский институт

Москва • Сентябрь 2020 — настоящее время

Ведущий российский научный центр в области молекулярной генетики

Аналитик данных / Биоинформатик

Стек: **Python**, R, bash, **Linux**, tabix, vcftools, bcftools, plink2, multiprocessing, FOSS

Зона ответственности:

- Обработка больших массивов биомедицинских данных (50–60 млн строк) в **Linux**-среде (**pandas**, multiprocessing, bash, CLI утилиты).
- Прунинг, импутация и анализ данных о мутациях с использованием **Python** и специализированных утилит.
- Интеграция C/C++ инструментов (tabix, vcftools, bcftools и др.) в пайплайны на **Python** и R.
- Автоматизация процессов, аналитические отчёты, контроль качества данных.

Примеры решённых задач:

- Спроектировал и оптимизировал пайплайн для анализа 50+ млн записей на **Python** и bash в **Linux**, реализовал многопоточность.
- Интегрировал C++ утилиты, ускорив расчёты с ~36 часов до ~5 часов при обработке VCF-файлов до 150 GB.
- Интегрировал расчёт статистик для анализа мутаций, связанных с ОНМК.
- Разработал систему автоматизированных отчётов (bash + R), снижающую ошибки первого рода.

Департамент здравоохранения города Москвы

Москва • Август 2021 — Ноябрь 2022

Крупнейшая медицинская организация, управляющая цифровыми сервисами и данными здравоохранения Москвы

Аналитик данных

Стек: ETL/ELT, **Python**, **SQL**, ClickHouse, **Airflow**, **pandas**, openpyxl, requests, matplotlib, seaborn

Зона ответственности:

- Автоматизация загрузки и обработки данных из ЕМИАС и S3 в DWH на ClickHouse **SQL** (**Airflow**).
- Разработка управленческих отчётов и дашбордов (**pandas**, ClickHouse **SQL**, matplotlib, seaborn).
- Расчёт метрик эффективности для медицинских учреждений.
- Автоматизация обновления справочников и поддержка ad-hoc аналитики.

Примеры решённых задач:

- Автоматизировал ETL плоских файлов до 20 Гб в ClickHouse, обеспечив стабильные обновления (**Airflow**).
- Автоматизировал ежедневные отчёты по KPI, освободив ~25 часов/неделю (**pandas**, **Airflow**, openpyxl).
- Автоматизировал ведение справочников адресов, сэконобив ~5 часов/неделю.
- Создал пайплайн контроля качества витрин данных, сократив подготовку на ~7 часов.

Работа в лабораториях, преподавание

Москва / Новая Зеландия / Амстердам / Бишкек • Август 2011 — Сентябрь 2020

Исследовательская и образовательная деятельность в международных лабораториях и вузах

Биоинформатик (ETL/ELT-пайплайны, анализ данных, моделирование, преподавание)

Стек: Python, Bash, R, Java, **Linux**, C/C++ build tools, scikit-learn, multiprocessing, FOSS

Зона ответственности:

- Моделирование и анализ биологических данных с использованием **Python**, Java, R и Bash.
- Разработка и поддержка аналитических пайплайнов (**Linux**, **pandas**, seaborn, scikit-learn, multiprocessing).
- Преподавание основ **Python** и анализа данных студентам.
- Интеграция CLI-утилит (C/C++/Fortran/R) в существующие пайплайны.

Примеры решённых задач:

- Смоделировал изменения в человеческом геноме при расселении людей по Океании (Massey University, New Zealand).
- Обучил >60 студентов основам **Python** (АУЦА, Бишкек).
- Автоматизировал обработку сложных данных с помощью CLI-утилит и скриптов.
- Создал модель биореактора для переработки дихлорметана.
- Участвовал в международных коллаборациях с современными методами статистики и визуализации.

Год	Учебное заведение	Специальность и квалификация
2014	Massey University, Новая Зеландия	Master in Computational Biology
2011	МГУ им. М.В. Ломоносова (ФФМ)	Лечебное дело (врач)
2011	МГУ им. М.В. Ломоносова (ВМиК)	Разработчик (доп. квалификация)

- **2024** — **SQL** для разработчиков (Яндекс Практикум)
- **2023** — Специалист по Data Science (Яндекс Практикум)
- **2020** — **Python** and Flask Bootcamp (Udemy)
- **2018** — Data Science Math Skills, Statistics with R, OOP in Java, REST APIs, Linear Algebra, Calculus (Coursera/edX/Stanford)

- **Языки:** Python, SQL, R, Java
- **Модули Python:** multiprocessing, os, requests, json
- **Базы данных и хранилища:** PostgreSQL, ClickHouse, S3
- **ETL и Workflow:** Airflow, Cron, Prefect 2, PySpark, Spark
- **Анализ данных:** pandas, matplotlib, seaborn, scikit-learn, Jupyter, Excel
- **API и Web:** REST, Flask, requests, json
- **DevOps:** git, Docker, CI/CD

- **ОС: Linux** (Ubuntu, Tuxedo OS), Windows, MacOS
-

- Учился программировать на ВМиК МГУ параллельно с дипломом врача (ФФМ МГУ).
 - После выпуска занимался сначала биоинформатикой, затем полностью перешёл в ETL и Data Engineering.
 - Уверенно владею **Python** и **SQL**. Стремлюсь углубить свои знания при помощи онлайн обучения.
 - Имею 3 года промышленного опыта разработки ETL процессов (**Airflow**, **pandas**, **PostgreSQL**, ClickHouse).
 - Сторонник непрерывного обучения. Сейчас прохожу курс по инженерии данных на Яндекс Практикум. Поступил в магистратуру ВШЭ по инженерии данных (обучение онлайн).
 - Слежу за новостями в области Data Science через подписку на medium.
 - Изучал основы теории игр и эволюционные вычисления в Universiteit van Amsterdam и Vrije Universiteit (Амстердам).
 - Анализировал структурированные данные и моделировал биологические системы.
 - Преподавал **Python** и основы анализа данных студентам.
 - Уверенно работаю в командной строке **Linux**, пишу скрипты на bash.
 - Открыт к предложениям Data Engineer, ML Engineer и Data Analyst.
-

 **Веб-версия резюме: инженер-данных.рф**

 **[Скачать полное резюме \(PDF\)](#)**