

Дамир Даиров

Москва | dairovdr@yandex.ru | +7 (999) 442-98-23 | tg: @dairovdr

Образование

Санкт-Петербургский государственный университет

2021–2025

Математико-механический факультет

Бакалавр прикладной математики и информатики

Навыки

- **Стек**
 - SQL(в т.ч JOIN-ы, cte, оконные функции, индексы, оптимизация запросов)
 - Python(numpy, pandas)
 - СУБД: ClickHouse, Greenplum, PostgreSQL
 - Git
 - Оркестрация: Apache Airflow
 - BI: Apache Superset
 - Excel, Google Sheets
 - Опыт работы с облачными технологиями
 - Уверенный пользователь MacOS, Linux
- **Математические дисциплины**
 - Теория вероятностей
 - Алгебра и теория чисел
 - Математический анализ
 - Математическая статистика

О себе

Целеустремленный, исполнительный, всегда открыт для критики, не стесняюсь задавать вопросы.

Опыт

- **Intern data engineer Sapiens Solutions**

Декабрь 2025 – Февраль 2026
Организация ELT пайплайна:

 - Выгрузка сырых данных из внешних источников: OLTP базы данных и файлы
 - Создание партицированных таблиц фактов и таблиц справочников распределенных на сегментах в Greenplum
 - Загрузка и трансформация данных в эти таблицы при помощи UDF
 - Создание витрины и загрузка ее в ClickHouse для дальнейшего анализа
 - Передача витрины в Superset и построение дашборда
 - Оркестрация процесса в Airflow

Проекты

- **ETL Pipeline для финансовых данных МОЕХ**

Ссылка на GitHub
Стек: Python, Pandas, Requests, Apache Airflow, ClickHouse

 - Разработал распределенную ETL-систему для сбора рыночных данных с API Московской биржи.
 - Реализовал отдельные DAG'и для акций и облигаций.
 - Организовал хранение в СУБД ClickHouse.

- **Сравнительный анализ СУБД (ClickHouse, Greenplum, Spark)**

Стек: ClickHouse, Greenplum, PySpark, YandexCloud, Docker, TPC-H

- Развернул и настроил кластеры СУБД (ClickHouse, Greenplum) и фреймворка Spark в облаке (Yandex Cloud) с помощью Docker.
- С помощью бенчмарка TPC-H выявил зависимости в производительности инструментов при различных конфигурациях железа (CPU, RAM).
- Выявил сценарии использования: ClickHouse - для быстрых агрегаций, Greenplum - для сложных SQL-запросов, Spark - для масштабируемой обработки.