

# Денис Ириняков

MIDDLE DATA ENGINEER · DATA SCIENTIST

Россия

☎ +7 (924) 236-02-83 | ✉ irinyakov2016@yandex.ru | 🏠 karnaksp.github.io/karnaksp | 📺 t.me/calmeds | 🌐 https://github.com/karnaksp

## Summary

Опытный инженер данных с глубокими знаниями в разработке и анализе данных. Специализируюсь на создании эффективных решений для обработки, анализа и визуализации больших объемов данных с использованием современных технологий и методологий. Стремлюсь к оптимизации процессов, повышению производительности и обеспечению высокого качества данных.

Имею бэкграунд в Data science и академической науке, что позволяет эффективно использовать в работе AI-инструменты, статистику и научный подход. Ценю работу в команде профессионалов, решение сложных технических задач.

Активно участвую в хакатонах, митапах и конференциях по DE и DS, проактивен в рабочем комьюнити. Развиваюсь и изучаю новые технологии, чтобы быть в тренде современных решений в области данных.

## Work Experience

### Axenix

Санкт-Петербург, Россия

#### DATA ENGINEER

Авг. 2024 - Настоящее время

- Разработал модульную систему динамической генерации YAML-конфигураций для DAG'ов на основе excel маппингов, что позволило сократить время настройки новых ETL-процессов на 40%.
- Создал комплексный набор инструментов автоматизации разработки: парсер логов Airflow (PowerShell), генератор синтетических данных, инструменты копирования артефактов и командную оболочку для управления Git-процессами, что позволило ускорить процесс адаптации новых сотрудников и улучшить качество работы.
- Оптимизировал генератор etl-потоков совместно с коллегами из других команд: внедрил шаблонизацию, асинхронную обработку и расширил покрытие генерируемых артефактов.
- Автоматизировал процесс тестирования качества данных и создал PL/pgSQL функции для согласующих тестов DataFix, значительно сократив время подготовки.
- Создал и поддерживал хранимые процедуры, функции, витрины и представления в Greenplum для операционных процессов качества данных.
- Обеспечивал интеграцию различных источников данных (ClickHouse, Oracle, Hive) и поддерживал архитектуру Data Vault 2.0.
- Организовал сообщество разработчиков в Telegram и GitLab для обмена инструментами и лучшими практиками.
- Разрабатывал etl-потоки, подготавливал релизы, хотфиксы и датафиксы.

### Spacecode

Краснодар, Россия (удаленно)

#### DATA SCIENTIST

Апр. 2024 - Авг. 2024

- Разработал мультимодального ассистента на базе LLM с использованием моделей HuggingFace, RAG-подхода, LangChain и Unstructured через LLAMA\_CPP/VLLM.
- Создал систему динамического ценообразования с использованием интерпретируемых регрессионных моделей, что, по оценкам, позволит увеличить прибыль сегмента на 15%.
- Использовал Apache Airflow для разработки ETL-пайплайна, обеспечивающего актуальность данных для ML-моделей.
- Внедрил DVC для версионирования данных моделей, что значительно упростило их отслеживание и воспроизведение результатов.

### АСК «ДиректСервис»

Новосибирск, Россия (удаленно)

#### DATA ANALYST

Сен. 2023 - Янв. 2025

- Реализовал с нуля enterprise-level Data Warehouse (DWH) на масштабах десятков терабайт данных, используя PostgreSQL как основное хранилище и ClickHouse как аналитическое хранилище.
- Автоматизировал процесс обновления витрин данных, создав комплексный пайплайн на основе Shell-скриптов и Windows Task Scheduler.
- Настроил механизм миграции данных из ClickHouse в PostgreSQL, обеспечив бесперебойную работу аналитических систем.
- Совместно с системными аналитиками разработал методологии мониторинга и оптимизации производительности DWH.
- Настраивал мониторинг на основе расширений Postgres, ci/cd для верифицированной доставки кода с использованием flyway и Github Actions.

### ООО «ColorDent»

Москва, Россия (удаленно)

#### ML ENGINEER

Июн. 2023 - Дек. 2023

- Разработал ML-классификатор изображений зубов для диагностики заболеваний, используя OpenCV и TensorFlow Lite. Модель достигла точности более 95% на тестовых данных.
- Интегрировал ML-модели в мобильное приложение с использованием Mlflow для управления версиями моделей и DVC для версионирования данных.
- Оптимизировал производительность модели для работы на устройствах с ограниченными ресурсами, что позволило использовать ее на мобильных платформах без значительной потери качества.
- Разработал систему отслеживания метрик обучения моделей и внедрил автоматическое тестирование для обеспечения высокого качества решения.

# Main Education

## ФНЦ Биоразнообразия ДВО РАН

Владивосток, Россия

### Аспирантура

2019 - 2023

- Специальность - экология. Глубокий анализ данных, статистика, математическое моделирование экологических процессов.
- Работа с пространственными и временными рядами для анализа экологических данных.
- Использование языка программирования R, статистических методов и математического моделирования.
- Публикация научных статей в области экологии и анализа данных.

# Data education

## Школа программирования 21

Новосибирск, Россия

### Основа - Школа 21

2023 - 2025

- C, bash, Python, SQL, Go, Ruby, алгоритмы, структуры данных.
- Погружение в языки программирования и методологии разработки.
- Peer-to-peer обучение и проектная деятельность.

## Яндекс Практикум

Онлайн

### Аналитик данных

Фев. 2023 - 2024

- Python библиотеки: pandas, scipy, numpy, matplotlib, scikit-learn, plotly, sqlalchemy.
- SQL запросы, анализ данных, машинное обучение.
- Статистический анализ и визуализация данных.

## Школа дата инженера T1

Онлайн

### DATA ENGINEER

2024 - Июн. 2024

- Школа дата инженера от холдинга T1 с выпускным проектом.
- ETL/ELT процессы, архитектура данных, современные инструменты.
- Практические проекты по обработке больших данных в Greenplum.

## Arenadata

Онлайн

### ARENADATA HADOOP ADMINISTRATOR

Янв. 2025 - Мар. 2025

- Эксплуатация Arenadata Hadoop. HDFS, Zookeeper, Grafana, Spark, Kafka.
- Мониторинг сервисов экосистемы Hadoop.
- Разворачивание кластеров в экосистеме Hadoop.

# Specialized Training

## Школа 21

Новосибирск, Россия

### C/C++ DEVELOPMENT

2023 - 2024

- Написание стандартных библиотек на C, структуры данных, алгоритмы.
- Работа с памятью, написание десктоп программ на C++ & QT.
- Проекты: 3dViewer, Калькулятор, Тетрис и другие.

## Школа 21

Новосибирск, Россия

### PYTHON DEVELOPMENT

2024

- Dockerfile, Makefile, Ansible, Neo4j, Python.
- unittest, pytest, subprocess, bs4, dotenv, scipy, plotly.
- grpc, sqlalchemy, asyncio, fastapi, aiogram.

## Школа 21

Новосибирск, Россия

### SQL & DATABASE

2024.1 - 2024.2

- pl/pgsql, функции, процедуры, оконные функции.
- Подзапросы, CTE, вложенные запросы, рекурсивные запросы.
- Оптимизация производительности баз данных.

## Школа 21

Новосибирск, Россия

### DEVOPS & INFRASTRUCTURE

2024.3 - 2024.4

- Linux, VM, Docker, CI/CD, bash.
- Prometheus, Grafana, Makefile, Docker, Vagrantfile.
- Контейнеризация и оркестрация приложений, настройка ci/cd.

## Key Projects & Achievements

### Школа 21

Руководитель DATA SCIENCE клуба

- Формирование сообщества в стенах школы 21 на 60+ человек.
- Проведение учебных занятий в стиле пир-ту-пир.
- Совместное участие в хакатонах, развитие софт скиллов команды.
- Тимлидерство команды на хакатонах и координация проектной деятельности.

Новосибирск, Россия

Мар. 2023 - Сен. 2024

### SmartSocialHack

Победитель хакатона

- Первое место в кейсе, посвященном социально значимым проблемам.
- Использование LLM, PostgreSQL, FastAPI для создания решения.
- Разработка мобильного и веб-приложения в команде.
- Интеграция PWA технологий для создания комплексного решения.
- Разработка на основе микросервисной архитектуры.

Москва, Россия

Июнь. 2024 - Июл. 2024

## Technical Skills

System Fundamentals	Linux, WSL, grep, env, Makefile, cron, systemd, git, ssh, rsync
Programming Languages	Python, SQL + PL/pgSQL, C, Go, Shell/Bash
Data Storage & Databases	PostgreSQL, ClickHouse, Greenplum, MySQL, SQLite, Neo4j, ChromaDB, S3
Python Stack	uv, pip, pandas, polars, scipy, numpy, matplotlib, scikit-learn, plotly, SQLAlchemy, asyncio, FastAPI, aiogram, selenium, pytest, pydantic, etc.
DevOps & Infrastructure	Docker, CI/CD, YAML, GitHub Actions, GitLab Runner, Vagrant, Ansible, Consul, Prometheus, Grafana
Data Orchestration	Apache Airflow, Kafka, Redis, Docker Swarm, VMVirtualBox
Data Architecture	Data Lake, Lakehouse, DWH, Data Vault 2.0, Anchor, Kimball, scd1/2/3
Machine Learning	ML и анализ данных, TensorFlow Lite, CV, NLP, LLM, Time Series, EDA, Feature Engineering, Data Cleaning, Data Augmentation, Data Visualization, статистика
AI Workflows & Integration	RAG, Vector DB, MCP, n8n, Copilot, Cursor, OpenRouter, HuggingFace, OpenAI, Ollama

## Soft Skills & Business Skills

Leadership & Teamwork	Командная работа, Лидерство, Kanban, Agile
Problem Solving	Критическое мышление, анализ данных, выдвижение гипотез, A/B тестирование
Adaptability	Адаптивность, высокая обучаемость, следование трендам технологий
Innovation	Инновационное мышление, использование AI-инструментов, DevOps, современные подходы к архитектуре БД
Data Quality	Качество данных, целостность, точность, доступность данных на всех этапах жизненного цикла
Security & Standards	Информационная безопасность, стандарты разработки