

Денис Ириняков

MIDDLE DATA ENGINEER · DATA SCIENTIST

Россия

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

Summary

Data engineer с опытом в программной разработке, анализе данных и машинном обучении. Работаю на стыке BI и ML: проектирую витрины и ETL/ELT-пайплайны, развиваю data catalog, строю инженерные контуры для AI-продуктов и обеспечиваю качество данных в modern data stack.

Имею бэкграунд в Data Science и академической науке, что позволяет эффективно использовать в работе AI-инструменты, статистику и научный подход. Сильные стороны: системная инженерия данных, глубокое погружение в бизнес-домен, автоматизация рутинных процессов, документация и доведение решений до production-эксплуатации.

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

DDX Fitness

Россия

DATA ENGINEER

Окт. 2025 - Наст. время

- Работаю на стыке **BI** и **ML**-команд: развиваю витрины, пайплайны и инженерный слой данных для аналитики, моделей и AI-продуктов.
- Орkestрирую **ETL/ELT-процессы**: регулярные и инкрементальные загрузки, обновление витрин, интеграцию продуктовых и внешних источников, контроль расписаний и SLA.
- Отвечаю за **качество данных**: проверки полноты, дубликатов и расхождений, мониторинг ключевых витрин, алертинг, разбор инцидентов и оперативные исправления.
- Развиваю **data catalog** и metadata governance: табличные контракты, семантический слой, теги, процессы ревью, валидации и синхронизации метаданных.
- Участвую в разработке **AI-тренера**: подготовка единого датасета упражнений, обогащение и нормализация данных.
- Проектирую миграции и синхронизацию данных между аналитическими и продуктовыми контурами: идемпотентные загрузки, dry-run/apply/backfill, rollback/restore и SQL-аудит результата.
- Разрабатываю внутренние сервисы и API для экспертного ревью и операционной работы с датасетами.
- Строю ingestion-пайплайны от API и файловых источников до аналитического слоя: инкременты, SCD2/merge, контракты типов, нормализация и подготовка сырых данных.
- Поддерживаю ML-направление: подготовка фич, inference-процессы, витрины для моделей и аналитиков, воспроизводимость расчётов.
- Оформляю инженерную документацию, runbook'и, E2E/smoke-сценарии, SQL-проверки и MR-процессы, снижая зависимость команд от ручных знаний.

- Участие в Big Tech проектах банковского сектора в архитектуре **Data Factory**.
- Работа в распределённой экосистеме данных, объединяющей десятки доменов и команд, в условиях высокой нагрузки и требований к масштабируемости.
- Разработка и поддержка ядра хранилища данных (DWH) в рамках **лейкхаус-архитектуры**, обеспечение согласованности, качества и доступности данных для аналитики (**Greenplum**).
- Интеграция данных из разнородных источников: **data lakes** (S3, HDFS, Hive, Delta Lake), аналитических платформ (ClickHouse), data platforms и других систем.
- Формирование **ETL/ELT конвейеров** передачи данных с использованием декларативного подхода к динамической генерации DAG'ов для Airflow (YAML + Python).
- Поддержка и развитие существующей архитектуры **Data Vault 2.0**.
- Разработка в долговременном релизном цикле на разных средах (от dev до prod), выполнение hotfix.
- Мониторинг потоков данных (Grafana, Prometheus), контроль их качества и полноты, оперативное устранение ошибок путём выполнения DataFix.
- Разработал модульную систему динамической генерации YAML-конфигураций для DAG'ов на основе Excel-маппингов, что позволило сократить время настройки новых ETL-процессов на 40% (Python).
- Создал комплексный набор инструментов автоматизации разработки: парсер логов Airflow (PowerShell), генератор синтетических данных, инструменты обработки артефактов и взаимодействия с микросервисами (Python), а также командную оболочку для управления Git-процессами (bash), что ускорило адаптацию новых сотрудников и уменьшило количество дефектов на code-review.
- Интегрировал в процесс командной разработки **code-review**, peer-review и pre-push проверки.
- Оптимизировал генератор ETL-потоков совместно с коллегами из других команд: внедрил шаблонизацию, асинхронную обработку и расширил покрытие генерируемых артефактов.
- Автоматизировал процесс подготовки артефактов для датафиксов на PL/pgSQL (проверки на согласованность, ожидаемые результаты), сократив время согласования DataFix на 80%.
- Создавал и поддерживал хранимые процедуры, функции и представления в **Greenplum**.
- Организовал междоменное сообщество разработчиков в Telegram для обмена инструментами и лучшими практиками.

Spacecode

Краснодар, Россия (удаленно)

- Разработал **мультимодального LLM-ассистента** с использованием RAG (LangChain, Unstructured, HuggingFace, Llama.cpp/vLLM), построив архитектуру хранения на **ChromaDB** (горячее хранилище) и **S3** (холодное хранилище).
- Построил систему **динамического ценообразования** на основе интерпретируемых моделей; генерация фич осуществлялась с помощью **PySpark** из сырых слоёв DWH и Data Lake.
- Оркестрировал **ETL/ML-пайплайны** через **Apache Airflow**, интегрируя запуск Spark-джобов на кластере и контроль качества данных.
- Внедрил **DVC** для версионирования датасетов и моделей, связав его с артефактами Spark и метаданными из Airflow.
- Настроил **DVC + Git workflow** для управления артефактами ML, что сократило время отладки и аудита на 70%.

АСК «ДиректСервис»

Новосибирск, Россия (удаленно)

- Реализовал с нуля **enterprise-level Data Warehouse (DWH)** на масштабах десятков терабайт данных, используя **PostgreSQL** как основное хранилище и **ClickHouse** как аналитическое.
- 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.
- Настроил мониторинг на основе расширений **Postgres** и внедрил **CI/CD** для верифицированной доставки кода с использованием **Flyway** и **GitHub Actions**.

ООО «ColorDent»

Москва, Россия (удаленно)

- 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 её 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

- 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 - Июнь. 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

Новосибирск, Россия

Руководитель DATA SCIENCE клуба

Мар. 2023 - Сен. 2024

- 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.

SmartSocialHack

Победитель хакатона

Москва, Россия

Июнь. 2024 - Июл. 2024

- 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.

Technical Skills

System Fundamentals	Linux, WSL, grep, env, Makefile, cron, systemd, git, ssh, rsync
Programming Languages	Python, SQL + PL/pgSQL, C, Go, Shell/Bash
Favorite Tools	ClickHouse, OpenMetadata, dbt, Airflow, PostgreSQL, Docker, GitLab
Data Storage & Databases	PostgreSQL, ClickHouse, Greenplum, MySQL, SQLite, Neo4j, ChromaDB, S3, Iceberg, Delta Lake, OpenMetadata
Python Stack	uv, pip, pandas, polars, scipy, numpy, matplotlib, scikit-learn, CatBoost, plotly, SQLAlchemy, asyncio, FastAPI, Flask, Pytest
DevOps & Infrastructure	Docker, Docker Compose, VMBox, CI/CD, GitHub Actions, GitLab Runner, Vagrant, Ansible, Consul, Prometheus, Grafana
Data Pipeline	Airflow, dlt, PeerDB, Kafka, Redis, Spark, Trino, Flink, dbt, FDW, pg_cron
Data Architecture	Data Lake, Lakehouse, DWH, Data Catalog, Metadata Management, Data Contracts, Semantic Layer, Data Vault 2.0, Anchore, Kimball, scd1/2/3
Machine Learning	CV, NLP, LLM, Time Series, EDA, Feature Engineering, Data Cleaning/Augmentation, Data Visualization
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