Denis **Irinyakov**

MIDDLE DATA ENGINEER · DATA SCIENTIST

Saint Petersburg, Russia

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

Summary

Experienced data engineer with deep knowledge in data development and analysis. Specializing in creating efficient solutions for processing, analyzing and visualizing large volumes of data using modern technologies and methodologies. Striving to optimize processes, improve performance and ensure high data quality.

I have a background in data science and academic science, which allows me to effectively use AI tools, statistics and a scientific approach in my work. I value working in a team of professionals and solving complex technical problems.

I actively participate in hackathons, meetups and conferences on data engineering and data science, and I'm proactive in the work community. I develop and study new technologies to stay up to date with modern data solutions.

Work Experience

Axenix
Saint Petersburg, Russia

DATA ENGINEER · Aug. 2024 - Present

- Developed a modular system for dynamic generation of YAML configurations for DAGs based on Excel mappings, which reduced the time for setting up new ETL processes by 40%.
- Created a comprehensive set of development automation tools: Airflow log parser (PowerShell), synthetic data generator, artifact copying tools, and a command shell for managing Git processes, which accelerated the onboarding process for new employees and improved work quality.
- Optimized the ETL flow generator together with colleagues from other teams: implemented templating, asynchronous processing and expanded the coverage of generated artifacts.
- Automated the data quality testing process and created PL/pgSQL functions for согласующие DataFix tests, significantly reducing preparation time.
- Created and maintained stored procedures, functions, data marts and views in Greenplum for operational data quality processes.
- Provided integration of various data sources (ClickHouse, Oracle, Hive) and maintained the Data Vault 2.0 architecture.
- Organized a developer community on Telegram and GitLab to share tools and best practices.
- Developed ETL flows, prepared releases, hotfixes and data fixes.

Spacecode
Krasnodar, Russia (Remote)

DATA SCIENTIST · Apr. 2024 - Aug. 2024

- Developed a multimodal assistant based on LLM using HuggingFace models, RAG approach, LangChain and Unstructured via LLAMA_CPP/VLLM.
- Created a dynamic pricing system using interpretable regression models, which is estimated to increase segment profit by 15%.
- Used Apache Airflow to develop an ETL pipeline ensuring data freshness for ML models.
- Implemented DVC for model data versioning, which greatly simplified their tracking and result reproduction.

ASK «DirectService»
Novosibirsk, Russia (Remote)

DATA ANALYST · Sep. 2023 - Jan. 2025

- Implemented an enterprise-level Data Warehouse (DWH) from scratch on a scale of tens of terabytes of data, using PostgreSQL as the main storage and ClickHouse as the analytical storage.
- Automated the data mart update process by creating a comprehensive pipeline based on Shell scripts and Windows Task Scheduler.
- Set up a data migration mechanism from ClickHouse to PostgreSQL, ensuring uninterrupted operation of analytical systems.
- Together with system analysts, developed methodologies for monitoring and optimizing DWH performance.
- Set up monitoring based on Postgres extensions, CI/CD for verified code delivery using Flyway and Github Actions.

LLC «ColorDent»
Moscow, Russia (Remote)

ML ENGINEER · Jun. 2023 - Dec. 2023

- Developed an ML image classifier for tooth disease diagnosis using OpenCV and TensorFlow Lite. The model achieved accuracy of over 95% on test data.
- Integrated ML models into a mobile application using MLflow for model version management and DVC for data versioning.
- Optimized model performance for operation on resource-constrained devices, which allowed it to be used on mobile platforms without significant quality loss.
- Developed a system for tracking model training metrics and implemented automated testing to ensure high solution quality.

Main Education

FRC Biodiversity of FEB RAS   *Vladivostok, Russia*

GRADUATE SCHOOL   *2019 - 2023*

- Specialization - Ecology. Deep data analysis, statistics, mathematical modeling of ecological processes.
- Working with spatial and time series for ecological data analysis.
- Using the R programming language, statistical methods and mathematical modeling.
- Publication of scientific articles in the field of ecology and data analysis.

Data education

School of Programming 21   *Novosibirsk, Russia*

FOUNDATION - SCHOOL 21   *2023 - 2025*

- C, bash, Python, SQL, Go, Ruby, algorithms, data structures.
- Deep dive into programming languages and development methodologies.
- Peer-to-peer learning and project activities.

Yandex Practicum   *Online*

DATA ANALYST   *Feb. 2023 - 2024*

- Python libraries: pandas, scipy, numpy, matplotlib, scikit-learn, plotly, sqlalchemy.
- SQL queries, data analysis, machine learning.
- Statistical analysis and data visualization.

T1 Data Engineer School   *Online*

DATA ENGINEER   *2024 - Jun. 2024*

- Data engineer school from T1 holding with a graduation project.
- ETL/ELT processes, data architecture, modern tools.
- Practical projects for processing big data in Greenplum.

Arenadata   *Online*

ARENADATA HADOOP ADMINISTRATOR   *Jan. 2025 - Mar. 2025*

- Operation of Arenadata Hadoop. HDFS, Zookeeper, Grafana, Spark, Kafka.
- Monitoring services of the Hadoop ecosystem.
- Deploying clusters in the Hadoop ecosystem.

Specialized Training

School 21   *Novosibirsk, Russia*

C/C++ DEVELOPMENT   *2023 - 2024*

- Writing standard libraries in C, data structures, algorithms.
- Memory management, writing desktop programs in C++ & QT.
- Projects: 3dViewer, Calculator, Tetris and others.

School 21   *Novosibirsk, Russia*

PYTHON DEVELOPMENT   *2024*

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

School 21   *Novosibirsk, Russia*

SQL & DATABASE   *2024.1 - 2024.2*

- pl/pgsql, functions, procedures, window functions.
- Subqueries, CTE, nested queries, recursive queries.
- Database performance optimization.

School 21   *Novosibirsk, Russia*

DEVOPS & INFRASTRUCTURE   *2024.3 - 2024.4*

- Linux, VM, Docker, CI/CD, bash.
- Prometheus, Grafana, Makefile, Docker, Vagrantfile.
- Containerization and orchestration of applications, CI/CD setup.

Key Projects & Achievements

School 21   *Novosibirsk, Russia*
DATA SCIENCE CLUB LEADER   *Mar. 2023 - Sep. 2024*

- Formed a community within School 21 of 60+ people.
- Conducted educational sessions in a peer-to-peer style.
- Joint participation in hackathons, developing team soft skills.
- Team leadership at hackathons and coordination of project activities.

SmartSocialHack   *Moscow, Russia*
HACKATHON WINNER   *June. 2024 - Jul. 2024*

- First place in a case dedicated to socially significant problems.
- Using LLM, PostgreSQL, FastAPI to create a solution.
- Developing a mobile and web application as a team.
- Integration of PWA technologies to create a comprehensive solution.
- Development based on microservice architecture.

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 and data analysis, TensorFlow Lite, CV, NLP, LLM, Time Series, EDA, Feature Engineering, Data Cleaning, Data Augmen Data Visualization, statistics
AI Workflows & Integration	RAG, Vector DB, MCP, n8n, Copilot, Cursor, OpenRouter, HuggingFace, OpenAI, Ollama

Soft Skills & Business Skills

Leadership & Teamwork	Teamwork, Leadership, Kanban, Agile
Problem Solving	Critical thinking, data analysis, hypothesis generation, A/B testing
Adaptability	Adaptability, high learning ability, following technology trends
Innovation	Innovative thinking, using AI tools, DevOps, modern DB architecture approaches
Data Quality	Data quality, integrity, accuracy, data availability at all stages of the life cycle
Security & Standards	Information security, development standards