Sergei Volkov

Data analyst/Python developer with an academic background in limnology and hydrophysics, passionate about automation. A long-term fan and user of Linux, Vim, and open-source software. I write efficient code to collect data from APIs, load it into databases, analyze, visualize, and create automated reports with Tableau or other BI tools. I prepare and analyze AB tests, check hypotheses, and describe the research to stakeholders. I value reproducibility (in computing and research), keeping information version-controlled, and writing concurrent, easy-to-read code.

Work experience

Aug 2023 – Feb 2024
Data Scientist, *Crocus Labs GmbH*
Software development, Software architecture, Product design, Project management, Testing, Team leading

- ○ Set up CI/CD and coding conventions
- ○ Designed AWS architecture for IoT, web app, service DB, and ETL
- ○ Developed various Python and Bash scripts
- ○ Created IoT showcase with virtual clients in Python and MQTT broker in separate Docker containers

I was hired to build an ETL for collecting data from smart lamps for greenhouses. However, upon my arrival, a complete redesign of the AWS architecture was necessary, which I made but eventually it was canceled. My employment ended when the company decided to switch to a self-hosted solution from cloud-based.

Jun 2021 – Mar 2023
PhD Student, *Leibniz-Institut für Gewässerökologie und Binnenfischerei (IGB Berlin)*
Research design, Data processing, Fieldwork, DevOps, Academic writing and editing

- ○ Analyzed various data, conducted research, and published papers
- ○ Created complex environments for the MATLAB package (OpenCV and mexopencv built with CMake in Docker, GUI on host machine with remote X session)
- ○ Developed Python computer vision programs and ad-hoc data wrangling frameworks
- ○ Organized fieldwork and participated in lab work
- ○ Administered remote Linux servers

Focused on the hydrophysical aspect of a joint limnological project, with an emphasis on biomass sinking, degradation, and the carbon cycle.

Oct 2020 – Feb 2023
Data QA, *Playrix*
Data analysis, Validation, ETL design, Manual and automated QA

- ○ Wrote complex SQL (various dialects) queries and scripts in Python and PySpark
- ○ Worked with Big Data
- ○ Developed automated data quality evaluation methods
- ○ Implemented several BI algorithms in SQL and Python
- ○ Designed methods for creating ad-hoc data validation SQL queries with metaprogramming
- ○ Managed QA section of the project and led QA team
- ○ Designed and verified BI dashboards

Hired by the data engineering team to check the legacy ETL data and conduct data quality control during migration to a new ETL. Later, responsibilities included developing Tableau dashboards and their sources, and consulting developers on calculating different metrics.

Mar 2020 – Oct 2020	**QA Engineer**, *Playrix* Manual QA, Automation, Teaching, Testing documentation writing, MitM traffic sniffing, Statistical analysis ○ Mocked package loss and network random latency by proxying traffic on a Linux host ○ Data-driven checking of random-based algorithms ○ Developed a Python script for TestRail API
Nov 2019 – Feb 2020	**Data Analyst**, *Playrix* Data analysis, Automation, User tracking events design, Reporting and presenting to stakeholders, AB tests *SQL, Python, Analytics, DataViz, Big Data*
May 2016 – Jul 2020	**Teaching Assistant**, *Petrozavodsk State University* Educational program design, Lecturing, Teaching *Thermodynamics, Heat transfer, and other courses*
Jan 2016 – now	**Research Assistant**, *Northern Water Problems Institute* Data analysis, Fieldwork, Experiment design, Data processing frameworks development, Academic writing, editing, and typesetting

Technical Skills

Development	*Python*: NumPy, Pandas, SQLAlchemy, Matplotlib, concurrency, Plotly
	SQL: Analytic/window functions, query optimization
	Data: Spark, Snowflake, dbt, Tableau, Looker, Plotly Dash, Spreadsheets
	NoSQL: DynamoDB, MongoDB
	Linux: Bash, system management, SSH, Nix
	Containers: Docker, systemd-nspawn, LXC, Kubernetes, chroot
	IaC: Terraform, Pulumi, Nix
	General: Git, Vim, Emacs, Jupyter
	CI/CD: GitHub Actions, TeamCity, GitLab
	Cloud: AWS, GCP
	Typesetting: Markdown, LaTeX, Pandoc
	Other: Julia, MATLAB, minor knowledge of JS, C/C++
Analytical	AB test design and analysis
	Cohort analysis
	Time-series analysis
	Descriptive and inferential statistics
	Machine learning basics
	Data wrangling
Research	*Fieldwork*: Organizing, sampling, working with probes, logistics, hand and power tools
	Writing: Academic writing, editing, publishing, proofreading
	Data processing: Writing ad-hoc frameworks, automation, visualization

Communication skills

- ○ Presenting
- ○ Communicating with stakeholders
- ○ Public speaking
- ○ Teaching/Mentoring
- ○ Feedback sharing
- ○ Active listening
- ○ Remote team building

Organisational / managerial skills

- ○ Agile, Kanban, Scrum
- ○ Managing small teams of developers
- ○ Task management and distribution
- ○ Project management

Projects

2015 – 2017   **Lake Onego: Life Under the Ice**, *NWPI*
Research and fieldwork
Joint interdisciplinary project to study the under-ice life

2021 – 2023   **Lake Pycnoclines Trap Organic Particles Forming Hot Spots of Accelerated Carbon Cycling in the Water Column (PycnoTrap)**, *IGB Berlin*
Joint interdisciplinary project researching sinking and degradation of biomass in the turbulent environment (lake water column)

Awards - Certifications - Licenses

2017   **Machine Learning**, *Stanford University on Coursera*
Certificate available at https://www.coursera.org/account/accomplishments/verify/U4VGJWXFGMS9

Voluntary

2013 – 2015   **Garbage Collection**, *On multiple sites near Onego Lake*, Petrozavodsk, Russia

2011 – now   **OpenStreetMap**, *Data surveys and contribution*, Republic of Karelia, Russia

Education

2010 – 2015   **Specialist Degree in Energy Supplies**, *Petrozavodsk State University*

Thesis

2015   **Noniterative Heat Exchanger Calculation**
Supervisor: Professor Sergei Bogdanov

The thesis proposes a direct analytical calculation procedure for heat exchanger design, avoiding the use of optimization algorithms

Personal

Citizenship   **Russian**

Languages   **Mother tongue: Russian**
English: C1
German: B1

Driving license **European Driving License: B, C**

Publications

2021 **Full Reynolds Stress Tensor of Convective Turbulence Estimated with Paired Acoustic Doppler Current Profilers**, *S. Bogdanov, G. Kirillin, S. Volkov, G. Zdorovennova*
DOI:10.1002/essoar.10507975.1

2019 **Fine scale structure of convective mixed layer in ice-covered lake**, *S. Volkov, S. Bogdanov, R. Zdorovennov, G. Zdorovennova, A. Terzhevik, N. Palshin, D. Bouffard, G. Kirillin*
DOI:10.1007/s10652-018-9652-2

2019 **Under-ice convection dynamics in a boreal lake**, *D. Bouffard, G. Zdorovennova, S. Bogdanov, T. Efremova, S. Lavanchy, N. Palshin, A. Terzhevik, L. Vinnå, S. Volkov, A. Wüest, R. Zdorovennov, H. Ulloa*
DOI:10.1080/20442041.2018.1533356

2019 **Structure and dynamics of convective mixing in Lake Onego under ice-covered conditions**, *S. Bogdanov, G. Zdorovennova, S. Volkov, R. Zdorovennov, N. Palshin, T. Efremova, A. Terzhevik, D. Bouffard*
DOI:10.1080/20442041.2018.1551655

2018 **Albedo of a Small Ice-Covered Boreal Lake: Daily, Meso-Scale and Interannual Variability on the Background of Regional Climate**, *G. Zdorovennova, N. Palshin, T. Efremova, R. Zdorovennov, G. Gavrilenko, S. Volkov, S. Bogdanov, A. Terzhevik*
DOI:10.3390/geosciences8060206