



# Linus Valiukas

Lithuania • [linas@valiukas.dev](mailto:linas@valiukas.dev) • +370 687 65870 • [LinkedIn](#) • [GitHub](#) • [valiukas.dev](http://valiukas.dev)

WRITTEN BY HUMAN, NOT AI

Accomplished backend lead software engineer with 15+ years of experience building high-scale distributed systems. Proven track record in platform engineering, including reducing AWS costs by \$500k/year at Automattic and scaling data pipelines for MIT/Harvard. Expert in Python and AWS, with a focus on building resilient, high-throughput infrastructure that aligns technical execution with commercial growth.

## EXPERIENCE

### **Founder**, aero.zip - Remote

2024 - 2025

Devised, developed and marketed a novel WeTransfer-like privacy oriented file transfer service ([Link](#)). The service features end-to-end encryption, 2 Gbps speeds, virtually unlimited number of files/file sizes, real-time downloads (one can start downloading before the upload is finished), resumable transfers, passkey/OPAQUE authentication. The project reached 1000 MAU in 2 months.

### **Senior Software Engineer**, Automattic Inc. - Remote

2022 - 2024

Employed by the core developer of WordPress CMS and owner of WordPress.com. Primarily worked on a team assigned to develop and maintain a WordPress backup/restore solution (7+ year old codebase, ~420K LoC) ([Link](#)):

- Inventorized a ~2.08 PB Amazon S3 store, conceived and executed a way (including both technical challenges and the red tape) to optimize storage and data retrieval, ended up reducing the monthly cost by about ~75% (\$52.4k -> \$13.4k), thus saving the company ~\$500k/year.
- Eliminated the number one churn reason for the product by designing, convincing stakeholders of, leading a team and co-implementing a bespoke binary protocol to upload/download file archives over HTTP(s). Said solution made WordPress website restores up to 26 times faster, allowed users to restore their websites more seamlessly as restores could be done over plain HTTP(S), reduced CPU, RAM and I/O resources on both sides of the interaction, and simplified the restore process in general.
- Containerized the main application including 22 dependent services, in order to automate integration testing and increase development speed and isolation; introduced disciplines of unit and integration testing to teammates.
- Debugged GnuPG (an encryption tool) with strace/GDB, discovered performance bottlenecks, suggested an effective workaround.
- Upgraded a 15+ year old, ~520K LoC PHP application with zero tests from PHP 7.4 to PHP 8.1.

### **Senior Software Engineer / Tech Lead**, MIT Media Lab - Remote

2012 - 2022

Architected and scaled one of the world's largest open-source news ecosystems in partnership with Harvard University. Managed the full technical lifecycle: from physical server provisioning to high-level system architecture and team leadership.

#### • **System Modernization & Leadership:**

- *Legacy Transformation:* Led the 10-year evolution of a ~100K LoC monolith, transitioning the stack from legacy Perl to a modern, distributed Python architecture.

- *Workflow Orchestration*: Replaced brittle cron-based systems with Temporal.io, implementing state-of-the-art asynchronous orchestration to handle complex, long-running data ingestion workflows.
- *Funding & Impact*: Directed a nimble dev team whose technical output secured \$30M+ in funding (Gates Foundation, Ford Foundation) and supported research featured in the New York Times and cited in 600+ scholarly articles.
- **Big Data & AI Infrastructure:**
  - *Massive Scale Data*: Engineered and maintained a 30+ TB PostgreSQL cluster. Implemented sharding (Citius), table partitioning, and query optimization to support real-time global news analysis
  - *AI/NLP Pipeline*: Built a multi-lingual NLP pipeline for 20+ languages. Implemented early AI features including word2vec model generation and automated audio/video transcription for podcasts using Google Speech-to-Text.
  - *Search Optimization*: Managed a high-throughput Solr cluster for full-text search across hundreds of millions of unstructured news articles.
- **Full-Stack Platform Engineering (The "Everything" Role):**
  - *DevOps Revolution*: Modernized the entire development lifecycle by introducing Git (from SVN), Docker containerization, Ansible for server provisioning, and automated CI/CD pipelines.
  - *Cloud Migration*: Executed a seamless migration of petabyte-scale data from MongoDB GridFS to Amazon S3, optimizing both cost and retrieval speed.
  - *Mentorship*: Acted as the primary technical mentor for new backend hires and supervised multiple Google Summer of Code interns.

## Freelance Developer

2005 - 2021

### Client Project Highlights:

- **Fotonija** (Lithuanian language software company)
  - Built a proprietary first-of-its-kind Lithuanian language support module for an open-source proofreading software featuring a spellchecker, a Hunspell-based morphological analyzer, and support for identifying most frequent Lithuanian language errors. The module is being used to analyze various Lithuanian language corpora. ([Link](#))
  - Wrote iOS and OS X versions of a leading English-Lithuanian dictionary (no longer available on the App Store).
- **Institute of Lithuanian Language** - Wrote a Qt-based Lithuanian language dictionary app for Windows, macOS, and Linux.
- **Merakas** (a transport technology company) - Developed a website that provides real-time public transport data for ~10 cities using client-side JavaScript to reduce hosting costs. ([Link](#))

## PUBLICATIONS & PRESENTATIONS

- Roberts, H., Bhargava, R., Valiukas, L., Jen, D., Malik, M.M., Bishop, C.S., Ndulue, E.B., Dave, A., Clark, J., Etling, B. and Faris, R., 2021, May. Media Cloud: Massive open source collection of global news on the open web. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 15, pp. 1034-1045). Available at: <https://ojs.aaai.org/index.php/ICWSM/article/view/18127> (Accessed: 22 January 2026).

- Dadurkevičius, V., Ralys, D.A., Samuilis, A., Vaičiulis, J., Ralienė, F. and Valiukas, L., 2020. Lithuanian Spelling Checker V. 1.0. 45 for macOS. Available at: <http://hdl.handle.net/20.500.11821/22> (Accessed: 22 January 2026).
- Dealing with Gigantic Tables (extended version): @ PGConf APAC 2019, Singapore ([Summary](#), [Video](#))
- Dealing with Gigantic Tables @ PostgresConf Silicon Valley 2018, San Jose, CA ([Summary](#))

## PERSONAL PROJECTS

- **iOS Keyboard** - Developed a third-party Lithuanian keyboard layout for iOS that replicates the standard system keyboard, featuring support for adjustable height, keys for popular emoticons / Emoji, and a bug report system. The app earned 400+ paid downloads in its first month. ([App Store link](#))
- **Google Summer of Code 2011 Project** - Developed an open source wxWidgets-based C++ wrapper over Cocoa Touch (iOS) that implemented a subset of wxWidgets API for touch-based devices. The API enabled wxWidgets users to port their C++ applications to iOS devices, and the interface could be reused to make wxWidgets available to other mobile platforms.
- **Public Transportation** - Published iOS / Java ME applications for using public transportation timetables while offline. iOS versions of the apps contain automatic timetable data updates, an offline trip planner, and a full city map.
- **Fervor** - Built software library for inclusion in Qt-based applications to enable autoupdates. ([Repository](#))
- **FreeBSD ports** - Ported [Quake III](#), [KCHM](#) (ebook reader), and [tspec2](#) (IPv6 tunnel utility) to FreeBSD (~2005).

## SKILLS

- **Programming:** Python, TypeScript/JavaScript, C/C++, Java, Qt, Perl (+ gradually porting legacy Perl apps to Python), PHP, Bash, some Rust
- **Databases:** PostgreSQL (schema design, maintenance, query optimization, PITR backups, streaming / logical replication, updates, sharding with Citus Data), MySQL/MariaDB, MongoDB
- **OS:** Linux (administration, performance benchmarking, debugging), macOS, FreeBSD
- **Web:** HTML 5, CSS, React/Svelte 5
- **Mobile:** Objective-C (iOS SDK), Android SDK / NDK
- **Cloud & Infrastructure:** AWS (migrations, EC2, S3, networking, pricing estimates), Google Cloud (Compute Engine, Cloud Storage, Elastic Container Registry, Speech-to-Text, networking, pricing estimates)
- **Platform & Workflow Management:** Temporal.io, Prefect, Uber Cadence, Netflix Conductor, Airflow
- **Tooling:** Docker, Ansible, ELK stack, RabbitMQ, CI (GitHub Actions, Travis, CircleCI, Jenkins), Git, Solr, LXC, LaTeX, some Kubernetes

## EDUCATION

**BA, Journalism**, Vilnius University - Vilnius, Lithuania (2012)