Svetislav Ćirić

Software Engineer

Belgrade, Serbia (CET) · Remote, overlapping US time zones · hello@ciric.me · ciric.me · github.com/ciric92

SUMMARY

Software engineer with 10+ years building distributed systems, data platforms and AI infrastructure, usually as a founding engineer. Built a cookieless contextual ad platform from its first commit to 1B+ requests a day, the system the company's revenue ran on. I do my best work owning a core system end to end: shipping the first version fast, then keeping it correct as traffic, data and the team grow. Contractor since 2019; my longest engagement lasted four years.

EXPERIENCE

HockeyStack · Senior Software Engineer — Jul 2025 – Nov 2026

B2B marketing and revenue analytics SaaS · Remote

- Owned Data Syncs, the warehouse import/export product used by dozens of enterprise customers. Redesigned its pipeline around composable writers and transformers, so new integrations and custom formats became quick to add instead of yet another copy of drifting code.
- Turned syncs from something data analysts set up for each customer into a self-serve product; support tickets dropped from several a day to about one a week.
- Built and maintained connectors for BigQuery, Snowflake, Redshift, Databricks, S3 and Azure Blob Storage: streaming exports, upserts, schema and type inference.
- Delivered a ~200 TB customer import the existing pipeline couldn't handle: sharded the input across 192 workers on 3 machines and pushed filters agreed with the customer into DuckDB, so less data moved and fewer runs failed.
- Made re-syncs and deletions safe to repeat (idempotent IDs, stale-data rejection, audit log); wrote the design docs and end-to-end test plans.

DeepStructure · Founding Engineer — Oct 2023 – May 2025

Platform for building and running AI workflows (TypeScript SDK, Go runtime) · Remote

- Designed and built the platform's Go runtime, which executed every workflow step; replaced the Python prototype's streaming layer and stayed its primary author until the company closed.
- Made a NATS JetStream event log the source of truth: step outputs materialize into PostgreSQL or SQLite and can be rebuilt by replaying the log after a schema or code change, instead of being wiped.
- Added per-workflow durability modes (transient / resilient / durable) on top of JetStream acks and delivery policies, along with concurrent processing, per-component rate limits, and parent/child run lineage with OpenTelemetry tracing.
- Restructured the runtime from one process per component into one runtime per application with composable middleware handlers; the team built 35+ internal AI applications on it (agents, RAG, Slack, Drive and Notion integrations).

C-WIRE · Principal Engineer — Feb 2020 – Mar 2024

Contextual, cookieless video advertising · Remote

- Founding engineer of the ad platform from its first commit: publisher scraping, contextual matching (Elasticsearch, later embeddings), real-time delivery with auctions, pacing and frequency capping, and Prebid and exchange integrations.
- Built it to serve 1B+ requests per day and 2,000+ campaigns from 200+ advertisers a year; it was the system the company's revenue ran on.
- Cut ad-serving p90 latency from ~150 ms to 20–30 ms with an in-process cache in front of eight Redis-backed caches, a deliberate trade of memory for latency.
- Split search into a hot index (articles from the last 14 days, ~90% of traffic) and a long-tail index, cutting query time by another order of magnitude.
- Replaced my own precomputed cache of article–creative matches with a read-through cache, so targeting changes went live instantly instead of after hours of recomputation.

- Led the event pipeline migration from Pulsar to Redpanda: rewritten in two weeks and rolled out through a canary ad-server deployment, cutting the brokers' CPU use from ~50 cores to 3. Events feed ClickHouse rollups for reporting and billing.
- Set the architecture and monorepo standards a team of ~15 engineers built on; onboarded and mentored new hires and worked directly with AdOps and Sales.

Contract engagements · Small projects alongside my main role — Since 2019
Remote

- GetSetUp: built the first social-graph prototype in Dgraph to recommend classes from attendance.
- Dynaboard: built data connectors for the low-code platform. Claimd: prototyped influencer–brand matching and scheduling.
- Rescued a manufacturer's legacy Java application, cutting its footprint ~100× by removing N+1 database access.

Tamedia (TX Group) · Data Engineer — Aug 2017 – Apr 2019
Swiss media group; central data team · Belgrade, Serbia

- Built a geo-targeted ad server in Go: advertisers drew delivery areas on a map and ads were served by user location.
- Built the group's event collection API (~5,000 requests/s) and the first Kafka → S3 backup (terabytes), feeding cross-site audience analytics for news recommendations, ad targeting and product analytics.
- Built competitor scrapers for the classifieds business and data pipelines for training the content recommender.

Yoterra · Software Developer — Jan 2016 – Aug 2017
Price-comparison and e-commerce data startup · Belgrade, Serbia

- Built distributed ingestion and normalization pipelines (Java, Hadoop, Spark) for 100k+ long-tail products, plus Python scrapers and product-variant detection.

Earlier: Research intern at Hitachi in Japan (2015), prototyping an autonomous-driving ECU with ROS on an NVIDIA Jetson; and at the Mihajlo Pupin Institute (2015) in robotics, building a remote pump-control and data-acquisition system.

PROJECTS

- **gem** (2026, in progress): Cirkon's successor: a cloud ERP in Go and PostgreSQL with an append-only stock ledger and exactly-once projections, built with AI-assisted development and documented in 27 architecture decision records.
- **Cirkon** (2019 – 2022): Retail ERP for a client's store (Go, React/Electron, PostgreSQL on GKE). From first commit to running the live store in under two months; the store's primary ERP for three years, until new fiscalisation rules.

SKILLS

Languages	Go, TypeScript/JavaScript (Node.js and browser, incl. Prebid.js ad scripts), Python, SQL; earlier Java 8 and Scala
Data & storage	PostgreSQL, ClickHouse, Elasticsearch, Redis, DuckDB, SQLite · warehouses: BigQuery, Snowflake, Databricks, Redshift · search: Vespa
Streaming & APIs	Kafka/Redpanda, NATS JetStream, Pulsar · gRPC/ConnectRPC, REST
Infrastructure	Terraform, Kubernetes (GKE, OVH), Docker, GitOps, AWS, GCP, Prometheus/Grafana, OpenTelemetry
Frontend	React; Electron and Wails desktop apps
AI	LLM workflow pipelines, embeddings for contextual matching, RAG; AI-assisted development
Domains	Ad tech, data integration and warehouses, event sourcing and ledgers, ERP

EDUCATION

B.Sc. in Electrical Engineering, University of Belgrade, School of Electrical Engineering, 2015 · GPA 9.57 / 10. Two conference papers on robot localization with machine vision (TELFOR 2015, INFOTEH 2016).