Muhammad Salehi

Principal Engineer | DevOps, Site Reliability (SRE), Platform & Infrastructure Engineering

salehi1994@gmail.com · linkedin.com/in/s4l3h1 · github.com/salehi · salehi.github.io

Summary

DevOps and infrastructure engineer who designs, builds and operates production systems on bare metal, virtualized and containerized platforms. Led infrastructure transformations from manual SSH deployments and single-VPS setups to automated GitLab CI/CD, Docker and Kubernetes delivery on privately managed, highly available infrastructure. Deep experience with high-availability MySQL (Percona XtraDB Cluster, ProxySQL, Keepalived), VMware vSphere, Linux, networking and security for fintech, cryptocurrency exchange and high-traffic platforms serving millions of requests per hour. Also a backend engineer and technical lead (Python, Django, Go) who has led full platform rewrites and live, zero-loss migrations of terabyte-scale data.

Skills

- **Infrastructure & Virtualization:** Linux, Bare Metal, VMware vSphere, vCenter, vMotion, Veeam Backup & Replication, OpenNebula, Private Cloud, Data Center Colocation, Cost Optimization (FinOps), Capacity Planning, Right-Sizing
- **Containers & Orchestration:** Docker, Docker Compose, Kubernetes
- **CI/CD & Tooling:** GitLab, GitLab CI/CD, Sonatype Nexus, Preview Environments, Release Management
- **Databases & Storage:** MySQL, Percona XtraDB Cluster (Galera), ProxySQL, Elasticsearch, ClickHouse, MinIO (S3), Backup & Disaster Recovery
- **Networking & Security:** TCP/IP, DNS, TLS, HTTP, Nginx, HAProxy, Keepalived, MikroTik, FortiGate, ModSecurity (WAF), CDN, WebRTC, STUN/TURN, Request Signing (HMAC), Bot Mitigation, Red Teaming, Phishing Simulation, Penetration Testing, Security Hardening, Firewalls
- **Observability:** Prometheus, Grafana, Graylog, Netdata, Sentry (APM), Monitoring & Alerting
- **Data Platform:** Kafka, Splunk, Metabase
- **AI / Machine Learning:** PyTorch, TensorFlow, CUDA, Hugging Face (Datasets, Transformers), LoRA Fine-Tuning, YOLO, Computer Vision, Ollama, FastFlowLM, Local LLM Inference (GPU and NPU), Speech-to-Text (STT), Text-to-Speech (TTS)
- **AI-Assisted Engineering:** Claude Code, LLM-Driven Development and Code Review, Agentic Workflows
- **Programming & Backend:** Python, Django, Django REST Framework, Go, FastAPI, Starlette, REST APIs, Next.js, React Native, Modular Service Architecture, Twelve-Factor App, Clean Code, Technical Leadership
- **Reliability:** High Availability, Automatic Failover, Load Balancing, SPOF Reduction, Performance Tuning, Incident Response

Experience

Lead DevOps Engineer | Lendo, formerly Iran Renter (Fintech: Crowdfunding & Digital Lending)
Full-time · Apr 2020 – Sep 2026

- Joined the fintech startup Iran Renter as a crowdfunding platform. After it was acquired by a bank, the company became the bank's digital lending contractor, delivering loans to consumers.
- Grew the infrastructure from two VPS instances into privately owned, highly available production infrastructure with zero-downtime requirements and up to 100,000 concurrent users, through both the startup and post-acquisition phases.
- Cut infrastructure costs by 70% in three ways: moving CDN delivery onto self-hosted infrastructure, replacing paid object storage with self-hosted MinIO S3, and right-sizing over-provisioned VMs based on measured average CPU and RAM usage per workload.
- Migrated source control from GitHub to self-hosted GitLab and built GitLab CI/CD pipelines that automated build, test and deployment for every service. Put the platform behind a CDN to reduce its attack surface.
- Procured, provisioned and colocated eight dedicated bare-metal servers. Built an OpenNebula private cloud running 60–70 VMs for databases, analytics, Metabase and internal department workloads.
- Migrated the private cloud from OpenNebula to VMware vSphere and vCenter, using vMotion to live-migrate every workload with no downtime. Introduced Veeam Backup & Replication.
- Ran offensive security testing against the company's own systems:
  - An authorized phishing simulation with RedProxy, a self-built man-in-the-middle red-team tool (github.com/salehi/RedProxy). It exposed staff credentials and led to company-wide security improvements.
  - An SMS-bombing test that used up the SMS budget in one hour and bypassed IP-based rate limits through public proxies.
- Designed and built a request-signing security layer. The frontend and backend share a secret that is never transmitted, and every request carries a time-limited HMAC-style signature of its body. A middleware between HAProxy and the backend drops unsigned or expired requests before they reach the application. This stops bots and replay attacks without

the CAPTCHAs or IP rate limits that block legitimate mobile users behind carrier CG-NAT.
- Secured the perimeter and web applications with a FortiGate 100D firewall, a ModSecurity web application firewall (WAF) on Nginx, and MikroTik routing. Redesigned upstream connectivity to cut internet hops to end users.
- Deployed a three-node Percona XtraDB Cluster (multi-writer MySQL) for high availability of financial data, with layered backups: VM-level Veeam backups plus encrypted, off-site, file-level replication of MinIO S3 object storage.
- Built observability with Prometheus and Grafana metrics, Graylog centralized logging and Netdata host monitoring. Set up HAProxy and Nginx for load balancing and reverse proxying.
- Advised the data team on analytics architecture, including ClickHouse, Kafka and Splunk data pipelines, and introduced Metabase for self-service BI.
- Mentored two engineers to take over full operational ownership of the infrastructure, then handed it over to them.

Principal Engineer | Pingi (Cryptocurrency Exchange)
Contract · Jul 2024 – Apr 2025

- Assembled and led a cross-functional team (backend, frontend and Android) to rewrite a cryptocurrency exchange that was split across two tangled, coexisting Django codebases (v2 and v3). Delivered a unified Django 4 platform (v4) covering the user, admin and system sides in four months.
- Designed a modular architecture of independent domain services: authentication, trading, wallet and balance services, plus a blockchain wrapper service that monitors multiple cryptocurrency nodes over RPC. Followed Twelve-Factor App and Single Responsibility principles.
- Cut the production database from 2 TB to 800 GB (60% smaller) and live-migrated all transaction data to the new v4 schema through Kafka, with no service interruption.
- Brought every API endpoint under 400 ms (as fast as 10 ms) by finding bottlenecks with Sentry performance monitoring (APM) and fixing their root causes: Django ORM query optimization and annotations, and simplified business flows.
- Turned a codebase nobody dared to change into clean, maintainable backend, frontend and Android code that junior developers and AI-assisted workflows can safely extend.
- Built CI/CD pipelines and hardened production security, including firewall configuration and access controls.

Technical Consultant | GAMA EdTech
Contract · Jan 2020 – Mar 2024

- Designed and deployed a five-node Percona XtraDB Cluster (Galera) in multi-writer mode for a write-heavy monolithic PHP platform serving millions of requests per hour.
- Built a highly available database access layer with two ProxySQL instances and Keepalived, providing load balancing and automatic failover across all writable nodes.
- Planned the placement of application, database and supporting services across bare-metal servers to minimize network round trips and end-to-end latency.
- Led MySQL performance tuning and capacity planning for peak-period traffic, using Prometheus, Grafana and Netdata monitoring.

Backend Developer | Bookapo
Contract · Oct 2021 – Dec 2022

- Co-led (team of two) a full rewrite of a legacy Joomla/PHP platform, which PHP developers had been unable to extend, into Django REST Framework (backend), Next.js (web frontend) and a new React Native mobile app for Android and iOS. Delivered in three months.
- Designed a new relational schema and migrated 3 GB of production data with no data loss, validated through a dedicated month of end-to-end testing.
- Kept the v1 API working during the transition through MySQL stored functions that rebuilt the legacy Joomla responses as JSON and were served through Django.
- Added a time-series database to track user activity and reading time.
- Built Persian full-text book search on Elasticsearch with the ParsiAnalyzer plugin.
- Built the CI/CD pipelines. Deployed on Docker Compose, then migrated to Kubernetes.

DevOps Engineer | AzPezeshk
Contract · Jan 2020 – Dec 2022

- Built the company's DevOps foundation from scratch: self-hosted GitLab, Sonatype Nexus artifact repository and GitLab CI/CD pipelines.
- Containerized the microservices and automated deployments with Docker Compose across separate staging and production environments, behind Nginx and HAProxy with Prometheus, Grafana and Graylog monitoring.
- Deployed WebRTC infrastructure with STUN/TURN servers for real-time audio and video.

DevOps Engineer | Snappfood
Full-time · Jan 2018 – Dec 2019

- Replaced manual SSH/git-pull and legacy Capistrano deployments with GitLab, GitLab CI/CD and Docker, and

containerized the entire Snappfood application stack.
- Automated Docker image builds and Docker Compose deployments across bare-metal production infrastructure.
- Designed a per-branch preview environment system that deployed each team's feature branch to its own subdomain. This enabled the release flow: feature review to shared staging to CTO-approved production release, fully automated by pipelines.
- Cut the developer test database from about 30 GB to about 1.5 GB (95% smaller) by removing obsolete tables and truncating unneeded data, which sped up onboarding and test environment setup.
- Operated a five-node Percona XtraDB Cluster supporting production database workloads.
- Tuned and operated Elasticsearch for Persian-language search using the ParsiAnalyzer plugin. Later took over the plugin, extended it and republished it as open source (see Projects).
- Reduced single points of failure with HAProxy and Nginx load balancing, managed the CDN configuration, and built infrastructure and application monitoring with Prometheus, Grafana, Graylog and Netdata.

Projects

- **Applied AI & Local LLM Lab (personal R&D):** Fine-tuned English language models with LoRA on Hugging Face datasets using PyTorch on NVIDIA GPUs. Train computer-vision models for personal habit tracking. Run self-hosted LLMs with Ollama on GPU and FastFlowLM on an AMD XDNA NPU, combined with STT and TTS into a local voice-enabled conversational avatar.
- **Human Body Detection (Sharif University of Technology AI course):** Built a YOLO-based computer-vision model to detect human bodies in images.
- **ParsiAnalyzer (github.com/salehi/ParsiAnalyzer):** Maintain and extend an open-source Persian text-analysis plugin for Elasticsearch (Java, Apache-2.0), continuing Nariman Esmailyfard's original project. Added release builds for every Elasticsearch 7.x, 8.x and 9.x minor version, built in Docker and published through GitHub Actions. Improved search quality with a new dictionary lemmatizer, a hardened light stemmer that stops over-stemming, Arabic normalization, and a full stop-word list (replacing the original 9 words). All linguistic data moved into configurable files, with three ready-made analyzers. Used in production at Snappfood and Bookapo.
- **MeshCheck:** Distributed network reachability and diagnostics platform with remote ping, TCP, HTTP, DNS, TLS and SMTP checks, run by a Go node agent that speaks a custom agent protocol.
- **NetScope (net.meshcheck.io):** Network intelligence service for IP, ASN, CIDR, country and organization lookups, backed by an offline GeoIP database.
- **S3 Proxy:** Starlette-based S3 proxy with AWS Signature V4, multiple backends, high-availability load balancing, health checks and streaming.
- **Nginx CORS Gateway:** Containerized Nginx reverse-proxy gateway with origin validation, CORS preflight handling, configurable upstreams and health endpoints.
- **AutoSSH Container:** Docker image for persistent SSH and SOCKS5 tunnels with automatic recovery and health checks.