

Guido Maria Serra

Principal Infrastructure Engineer | AI/ML Systems & GPU Platforms

Berlin, Germany

+49 173 7349 795 | guido@whitehat.berlin | linkedin.com/in/zeph1ro | github.com/zeph

Profile

Principal Infrastructure Engineer with ~30 years of experience building and operating large-scale distributed systems, from early Linux infrastructure to modern GPU-based Kubernetes platforms for AI/ML workloads. Specialized in stabilizing critical infrastructure, optimizing costs (e.g., \$300K/year savings at OLX), and coaching SRE/DevOps teams through operational challenges.

Highlights: Operated infrastructure at 55k-host scale (12 DCs); coached teams through production crises.

Selected Engagements

PlayStation Now — Global Infrastructure Operations

Operated multi-datacenter platform (55k hosts, 12 DCs); built Ceph storage, PKI-based SSH. Designed custom metrics DSL (TextX) integrated with Grafana. Gentoo maintainer for Ganeti/KVM.

HelloFresh — Security & Infrastructure Scaling

Took over security function in 2 weeks; stabilized and scaled SIEM, SSO alerting, identity validation.

OLX Group — Infrastructure Cost & Reliability

Led APM strategy across 4 hubs. CDN/SSL/DNS vendor selection. NewRelic vendor manager (\$2M budget). Load testing and performance validation (custom tooling).

Professional Experience

Principal Infrastructure & Security Consultant (Fractional) 2025–present

Parallel engagements: DFND Security, Forward Earth, 3T Labs

- WAF / edge security (Fastly Signal Sciences) deployments for EMEA clients.
- ISO 27001 certification and compliance automation (Vanta).
- Infrastructure security assessments and remediation across multiple client environments.

Software Engineer | FinTech (stealth) 2025

Tokenized financial systems (Python, TypeScript, microservices).

- Developed backend components for tokenized financial systems.
- Evaluated and applied AI-assisted coding workflows to accelerate iteration cycles.
- Performed load testing and reverse engineering of tools/libraries to validate system behavior.

Head of SRE | Octostar 2024–2025

GPU Kubernetes platform for ML workloads; ClickHouse/MinIO data platform.

- Designed GPU-backed Kubernetes clusters for ML training/inference.
- Optimized cost and performance for ML workloads (mixed on-prem/cloud GPU nodes).
- Validated workload portability across local Kubernetes (kind/minikube) and cloud (Hetzner).

DevOps Lead (Advisory) | Contensi 2024–present

CDKTF infrastructure automation; mentoring junior engineers.

Staff SRE (Fractional) | Prima Assicurazioni 2022–2023

Coached 12-engineer SRE team out of operational deadlock; AWS multi-region rearchitecture (London, Barcelona, Milan).

Cloud Architect | ClovrLabs 2021–2023
 Built 5-person security ops team; Kubernetes/GitOps platform (ArgoCD, Traefik, Datadog).

Site Reliability & Security Engineer | Max Planck Institute 2022–2024
 Hybrid OpenStack/Kubernetes; 1.5PB storage; federated IAM (LDAP/SAML/OAuth2).

Security / DevOps Engineer | HelloFresh 2020–2021

DevOps Engineer | Mobivia / ATU 2018–2022
 4-year modernization: Cloudflare/Traefik migrations, Keycloak SSO, ArgoCD GitOps.

DevOps Team Lead | Urban Mobility Int. (WeShare/VW) 2019
 1-week pre-launch engagement; GCP Functions and Kubernetes.

Start-up CTO | SiWeGO (Trento, Italy) 2017–2019
 Solution design for low-OPEX operations; algorithmic design of graph overlaps; suppliers management.

Senior Performance Engineer | OLX Group 2016–2018

Site Reliability Engineer | Sony PlayStation (Gaikai → PlayStation Now) 2013–2016

Additional Engagements (Selected)

SkyCharge — Extended Debian-based ANSI C codebase (ARM / cross-compilation)
Bella & Bona — Architecture review for investor due diligence

Earlier Career (2008–2013)

Rocket Internet — Chief Troubleshooter & Senior Data Architect (2013) | Berlin
txtr (3M) — Backend QA Manager (2011–2012) | Berlin
ProfitBricks (→ IONOS) — Senior System Engineer (2011) | Berlin
Nokia Maps (Gate5 → HERE) — Senior System Engineer (2010–2011) | Berlin
Vodafone Group — QA Architect (2008–2010) | Düsseldorf

Foundational Era (1996–2008)

Solution Architect for SMEs in Italy. LDAP/VoIP/firewall solutions; patches for GOsa², PHP Seagull, GNU TLA. Clients: ISPs, banks, telcos. Stager under Alessandro Rubini (embedded Linux). Lab Assistant at Politecnico di Milano (2002). Active in FOSDEM since 2007.

Core Technical Domains

Distributed Systems & Infrastructure	Large-scale Linux, Ceph, distributed storage, fleet management
Cloud & Platform Engineering	AWS, GCP, Kubernetes, Terraform, GitOps
Observability & Reliability	Prometheus, Grafana, ELK, OpenTelemetry, SLO/SLI
Security & Compliance	WAF, IAM/SSO, ISO 27001, SIEM
AI/ML Workloads	GPU environments, data pipelines, local/cloud portability
Programming	Go, TypeScript, PHP, SQL; hands-on, strongest in debugging and systems analysis
Low-Level Systems	C, embedded Linux, kernel-level performance

Education & Credentials

B.Eng. Computer Science (final year) — Politecnico di Milano | 2000–2008, resumed 2025
Training: Red Hat Ceph Storage (CEPH125), 2016; ISO 27001 / Vanta Auditor, 2025

Additional

Languages: Italian (Native), English (Fluent), German (Working)
Security: DEFCON CTF finalist; iCTF World Champion; SOC GOON (2025)
Open Source: FSFE Fellow; co-founder of Politecnico Open *nix Labs (POuL); wicd Debian package maintainer
Licenses & Interests: Driving (A, B); Offshore Sailing (racing cert., maritime radio); Skiing; Trekking