

SUDHAKAR CHUNDU

chundubabu@gmail.com | (513) 666-0099 | San Jose, CA | linkedin.com/in/schundu | sudhakarchundu.org

PROFESSIONAL SUMMARY

Distinguished Cloud AI Architect and Platform Engineering Leader with 18+ years of hands-on experience (13+ years in cloud platforms, 5+ years in Linux/Middleware). Expert in building large-scale GPU compute platforms, AI/ML infrastructure, and distributed systems. Skilled in Slurm cluster management, GPU scheduling (NVIDIA A100/V100/T4/DGX), Kubernetes orchestration, and multi-cloud architecture (AWS, Azure, OCI). Proven leader managing global teams of 13+ engineers, \$15M+ budgets, delivering \$8M+ cost savings with 99.97% uptime. Deep expertise in AIOps, MLOps, observability, and compliance (SOC2, HIPAA, FedRAMP). Conference speaker at KubeCon, SREcon, and MLOps World.

TECHNICAL SKILLS

Cloud Platforms: AWS (EKS, EC2, Fargate, SageMaker, EMR, Lambda, S3, RDS, DynamoDB), Azure (AKS, Azure ML, ExpressRoute, DevOps, Cosmos DB), OCI

GPU/HPC & AI/ML: Slurm, Slinky, NVIDIA BCM/DGX, CUDA, TensorRT, A100/V100/T4, PyTorch, MLflow, Kubeflow, SageMaker, Ray, Triton, vLLM

Containers & IaC: Kubernetes (EKS/AKS), Docker, Helm, Istio, KEDA, Terraform (50+ modules), Ansible, CloudFormation, ARM/Bicep, ArgoCD, FluxCD

CI/CD & DevOps: GitHub Actions, Azure DevOps, Jenkins, GitLab CI, blue-green/canary deployments, GitOps, Pipeline-as-Code

Big Data: Databricks, Apache Spark, Flink, Kafka, Hadoop/HDFS, Airflow, Delta Lake, petabyte-scale processing

Observability & SRE: Prometheus, Grafana, Datadog, ELK Stack, OpenTelemetry, Splunk, SLI/SLO, chaos engineering, AIOps

DevSecOps & Compliance: Trivy, Clair, Anchore, Falco, OPA, Kube-bench, OWASP ZAP, Gitleaks, DefectDojo, Sigstore, SonarQube | SOC2, HIPAA, FedRAMP

Programming: Python, Go, Bash, Java, SQL | Linux (RHEL/Ubuntu), kernel tuning, systemd, JVM optimization

PROFESSIONAL EXPERIENCE

Distinguished Cloud AI Architect / Director of Platform Engineering

Trackonomy Systems - San Jose, CA | October 2023 - Present

- Built Infrastructure Engineering team (0 → 6 members) serving Pharma, Airlines, Government, Healthcare, and IoT sectors across 8 countries
- Designed Slurm-based GPU compute platform (65 GPUs, 8 nodes) using Slinky and NVIDIA BCM achieving 99.97% uptime for 12+ enterprise clients
- Architected serverless GPU infrastructure using AWS Lambda containers reducing inference costs by 65% serving 5M+ daily predictions
- Reduced cloud costs 73% (\$10M → \$2.7M/year) through GPU optimization, fair-share scheduling, and reserved instance planning
- Redesigned multi-cloud architecture (AWS, Azure, OCI) reducing Azure resources 87% (8.7K → 1.1K); migrated VMs to AKS/EKS with GitOps
- Built CI/CD for 100+ applications using ArgoCD, FluxCD, Azure DevOps, GitHub Actions enabling 50+ daily zero-downtime deployments
- Led SOC2 Type II, ISO 27001, HIPAA, FedRAMP, HITRUST compliance with Vanta; implemented DevSecOps achieving zero critical findings
- Implemented AIOps for predictive monitoring and automated root cause analysis reducing incidents 65% and MTTR 60%
- Unified 5 monitoring platforms into Datadog; built observability with Prometheus, Grafana, OpenTelemetry, custom GPU exporters

Cloud and SRE Architect - OSDU Data Platform

Wipro Technologies (BP, Chevron, ExxonMobil, Total) - Seattle/Redmond, WA | February 2020 - October 2023

AWS Platform - ML Infrastructure & SRE (February 2022 - October 2023)

- Architected OSDU R3 data platform processing exabytes of seismic data on GPU-accelerated EKS with Spark, Kafka, and Hadoop
- Delivered GPU cluster management with Kubernetes, Istio, and KEDA ensuring 99.95% availability for ML inference workloads
- Created 50+ Terraform modules (VPC, EKS, RDS, S3, SageMaker); implemented GitOps (ArgoCD/FluxCD) reducing deployment time 80%

- Built CI/CD pipelines with Jenkins and GitHub Actions; containerized applications with Docker; managed secrets with HashiCorp Vault
- Implemented container security with Trivy, Clair, and Anchore Engine for image scanning; Falco for runtime security monitoring
- Enforced policies with Open Policy Agent (OPA); ran Kube-bench for CIS benchmarks; integrated SonarQube for code quality gates

Azure Platform - ML & Platform Engineering (February 2020 - February 2022)

- Engineered AI infrastructure using Azure ML, AKS with GPU node pools, and ExpressRoute for hybrid connectivity
- Deployed 15+ microservices on AKS with Istio service mesh (mTLS, distributed tracing); achieved 99.9% availability via FluxCD GitOps
- Built Azure DevOps and Jenkins pipelines with Terraform/Bicep IaC; automated with Ansible; improved deployment speed 80%
- Implemented DevSecOps: OWASP ZAP for DAST, Gitleaks for secret detection, DefectDojo for vulnerability management
- Deployed Sigstore for artifact signing/verification; OSQuery for endpoint visibility; delivered SOC 2, HIPAA, ISO 27001 compliance

Multi-Cloud Architect / Senior Infrastructure Engineer

Tata Consultancy Services - Multiple Locations | May 2007 - February 2020

Cloud Architect - Harvard Pilgrim Health Care via NTT Data (June 2018 - February 2020)

- Led HIPAA/HITRUST cloud modernization to AWS with GPU-enabled EKS; containerized apps with Docker and Kubernetes
- Built CI/CD with Jenkins; IaC with Terraform and Ansible; implemented HashiCorp Vault for secrets; managed \$6M budget
- Implemented container security scanning with Trivy and Clair; code quality gates with SonarQube; secret detection with Gitleaks

Cloud Senior Engineer - CNA Insurance (May 2015 - June 2018)

- Managed Kubernetes and Docker deployments on AWS; led migrations to AWS, Azure, OpenStack; pioneered container adoption
- Built CI/CD with Jenkins; automated infrastructure with Terraform and Ansible; implemented Open Policy Agent for governance

Solutions Architect - PwC (January 2011 - April 2015)

- Integrated Chef/Jenkins pipelines; migrated VMware to AWS EC2; led 20+ facility buildouts; managed \$15M+ budgets

Middleware Engineer / Systems Administrator - Verizon, Owens Corning (May 2007 - January 2011)

- 5 years Linux/Unix administration with WebSphere/WebLogic; kernel tuning, JVM optimization, HA clustering
- Implemented WebSphere clustering, IBM HTTP Servers, SSO with CA SiteMinder; managed 24x7 L3 operations with strict SLAs

KEY ACHIEVEMENTS

- **Scale:** 10M+ users served | Petabyte-scale data processing | 200+ microservices | 15+ AWS regions | 50+ daily releases
- **Cost Optimization:** \$8M+ annual savings | 73% cloud cost reduction | 87% infrastructure consolidation | 65% inference cost reduction
- **Reliability:** 99.97% uptime | 95% deployment time reduction | 60% MTTR improvement | Zero critical audit findings
- **Leadership:** Built team 0 → 6 | \$15M+ budgets | 13+ global engineers | Speaker: KubeCon, SREcon, MLOps World

CERTIFICATIONS

AWS Solutions Architect Professional | AWS Machine Learning Specialty | Azure Solutions Architect Expert | Azure AI Engineer Associate | Certified Kubernetes Administrator (CKA) | HashiCorp Terraform Associate | Databricks Certified | NVIDIA DLI Infrastructure | TOGAF 9 Foundation | ITIL v4 Foundation

EDUCATION

Bachelor of Engineering, Mechanical Engineering - Acharya Nagarjuna University, 2005