

# SUDHAKAR CHUNDU

Email: chundubabu@gmail.com | Phone: (513) 666-0099 | Location: San Jose, CA

LinkedIn: linkedin.com/in/schundu | Portfolio: sudhakarchundu.org

## SUMMARY

Distinguished Cloud AI Architect and Platform Engineering Leader with 18+ years of hands-on experience including 13+ years in cloud platforms and 5+ years in Linux/Middleware. Expert in building and operating 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 across AWS, Azure, GCP, and OCI. Proven leader managing global teams of 13+ engineers, \$15M+ budgets, and delivering \$8M+ in documented cost savings with 99.97% platform uptime.

## SKILLS

**Cloud Platforms:** AWS (EKS, EC2, Fargate, SageMaker, EMR, Lambda, S3, RDS), Azure (AKS, Azure ML, ExpressRoute, DevOps), GCP (GKE, Cloud Run, Vertex AI), OCI

**GPU/HPC:** Slurm, NVIDIA DGX, CUDA, TensorRT, A100/V100/T4, Triton Inference Server, vLLM, distributed training, fair-share scheduling

**AI/ML:** PyTorch, TensorFlow, MLflow, Kubeflow, SageMaker, Azure ML, Vertex AI, Ray, model serving, MLOps, distributed training

**Containers & Orchestration:** Kubernetes (EKS, AKS, GKE), Docker, Helm, Istio, KEDA, ArgoCD, FluxCD, container security

**Infrastructure as Code:** Terraform, Ansible, CloudFormation, ARM/Bicep, Packer, GitOps, policy-as-code (OPA, Checkov)

**CI/CD & DevOps:** GitHub Actions, Azure DevOps, Jenkins, GitLab CI, ArgoCD, blue-green deployments, canary releases

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

**Observability & SRE:** Prometheus, Grafana, Datadog, ELK Stack, OpenTelemetry, SLI/SLO, chaos engineering, incident management

**Security & Compliance:** SOC2, HIPAA, FedRAMP, ISO 27001, HITRUST, DevSecOps, Vault, Snyk, Veracode, SonarQube

**Programming:** Python, Go, Bash, Java, SQL, YAML

**Operating Systems:** Linux (RHEL, Ubuntu, CentOS), kernel tuning, systemd, performance optimization

## EXPERIENCE

### Distinguished Cloud AI Architect / Director of Platform Engineering

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

- Built and led Infrastructure Engineering team from 0 to 6 members across Cloud, DevOps, Security, and Networking serving clients globally across 8 countries
- Designed and operated Slurm-based GPU compute platform with 65 GPUs across 8 nodes using Slurm, Slinky, and NVIDIA BCM achieving 99.97% uptime for 12+ enterprise clients
- Architected serverless GPU infrastructure using Google Cloud Run and AWS Lambda containers reducing inference costs by 65% while serving 5M+ daily predictions
- Reduced cloud costs by 73% (\$10M to \$2.7M annually) through GPU optimization, fair-share scheduling, reserved instances, and vendor consolidation
- Redesigned multi-cloud architecture across AWS, Azure, GCP, OCI reducing Azure resources from 8,700 to 1,100 (87% reduction)
- Built CI/CD pipelines for 100+ applications using ArgoCD, FluxCD, Azure DevOps, GitHub Actions enabling 50+ daily releases with zero-downtime deployments
- Led SOC2 Type II, ISO 27001, HIPAA, FedRAMP, and HITRUST compliance achieving zero critical findings in audits

- Implemented AIOps frameworks for predictive monitoring and automated root cause analysis reducing incidents by 65% and MTTR by 60%

## **Cloud and SRE Architect - OSDU Data Platform**

**Wipro Technologies** - Seattle, WA / Redmond, WA | February 2020 - October 2023

Clients: BP, Total, Chevron, ExxonMobil

### **AWS Platform - ML Infrastructure and SRE (February 2022 - October 2023)**

- Architected OSDU R3 data platform processing exabytes of seismic data on GPU-accelerated Kubernetes clusters (EKS/Fargate) with Spark, Kafka, and Hadoop
- Delivered GPU cluster management at scale with Kubernetes, Istio, and KEDA ensuring 99.95% availability for ML inference workloads
- Created 50+ Terraform modules for VPC, EKS, RDS, DynamoDB, S3, SageMaker, and IAM reducing deployment time by 80%
- Implemented MLOps best practices including model versioning, automated retraining, and blue-green deployments for 30+ production models

### **Azure Platform - ML and Platform Engineering (February 2020 - February 2022)**

- Engineered scalable AI infrastructure using Azure ML, AKS with GPU node pools, and ExpressRoute for hybrid connectivity
- Deployed and managed 15+ microservices on AKS with Istio service mesh achieving 99.9% availability
- Built Azure DevOps pipelines with Terraform and Bicep IaC improving deployment speed by 80%
- Delivered compliance-as-code achieving SOC 2, HIPAA, and ISO 27001 certifications

## **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 cloud modernization for HIPAA/HITRUST-regulated AI/ML applications to AWS with GPU-enabled EKS clusters
- Designed compliant ML pipeline with end-to-end encryption and audit logging meeting FDA GxP requirements
- Managed \$6M infrastructure budget and implemented Jenkins/Ansible CI/CD pipelines with DevSecOps practices

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

- Managed Kubernetes and Helm deployments on AWS and led cloud migrations to AWS, Azure, and OpenStack
- Pioneered Docker and Kubernetes adoption and implemented CI/CD pipelines with Jenkins and Ansible

### **Solutions Architect - PwC (January 2011 - April 2015)**

- Integrated Chef and Jenkins deployment pipelines and migrated VMware VMs to AWS EC2
- Led infrastructure deployments for 20+ facility buildouts and managed \$15M+ annual budgets

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

- 5 years Linux/Unix administration with WebSphere and WebLogic middleware including kernel tuning and JVM optimization
- Implemented WebSphere clustering, IBM HTTP Servers, and SSO with CA SiteMinder managing 24x7 L3 operations

## **CERTIFICATIONS**

AWS Solutions Architect Professional, AWS Machine Learning Specialty, Azure Solutions Architect Expert, Azure AI Engineer Associate, Google Cloud Professional Architect, GCP Professional ML Engineer, Certified Kubernetes Administrator (CKA), HashiCorp Terraform Associate, Databricks Certified, NVIDIA DLI Infrastructure, TOGAF 9 Foundation, ITIL v4 Foundation, VMware VCP, Cisco CCNP

## **EDUCATION**

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