

D SUDEEP REDDY

Site Reliability Engineer | DevOps and Cloud Infrastructure

Bengaluru, India • +91 9137804971 • sudeepreddy340@gmail.com • github.com/sudeepr34

PROFESSIONAL SUMMARY

Results-driven Site Reliability Engineer with 3+ years of experience delivering 24x7 high availability and incident response across business-critical, high-traffic fintech environments. Specialized in AIOps, real-time troubleshooting, toil reduction, and zero-downtime infrastructure maintenance. Proven track record architecting multi-region cloud systems, applying AI/ML to incident response and anomaly detection, and deploying transformer-based ML models for proactive issue detection. Proficient in CI/CD management, IaC, and Kubernetes orchestration across AWS, Azure, and on-premises datacenters.

CORE TECHNICAL SKILLS

OS: RHEL, Ubuntu, CentOS, Windows.

Cloud Platforms: AWS (EC2, S3, Route53, IAM, CloudWatch, EMR, VPC), Azure (Cosmos DB, Virtual Networks).

Containers and Orchestration: Kubernetes, Docker, Helm, Drove (PhonePe's container orchestrator).

Infrastructure as Code: Terraform, Ansible, SaltStack.

CI/CD and Deployment: GitLab CI/CD, Jenkins, Release orchestration.

Monitoring and Observability: Prometheus, Grafana, ELK Stack (Elasticsearch, Logstash, Kibana), Splunk, OpenSearch.

Databases: MySQL, Percona, Galera, Cosmos DB, Aerospike.

AI/ML and Advanced Analytics: Transformer Models, LLM Integration, TensorFlow, Predictive Anomaly Detection, Automated Incident RCA.

Scripting and Automation: Python, Shell Scripting, Bash.

Batch and Workflow Orchestration: Apache Airflow, Control-M.

Networking and Load Balancing: TCP/IP, DNS, VPN, BGP, VLAN/VXLAN, Nginx, firewalls, NAT

ITSM and Collaboration: Jira, Confluence, ServiceNow, Agile/Scrum, incident management, RCA documentation

PROFESSIONAL EXPERIENCE

Site Reliability Engineer - Ops | PhonePe Pvt. Ltd. | Dec 2025 – Present | Bengaluru, India

- Primary on-call responder for mission-critical fintech payment platforms serving **700+ million users**, leading incident detection, troubleshooting, and resolution of high-severity production incidents.
- Built and operated a multi-region active-active **MySQL Galera cluster** across **Azure** and two on-premises **datacenters** for high availability (**HA**) and disaster recovery (**DR**); leveraged virtually **synchronous replication** and seamless failover mechanisms to achieve sub-second data synchronization and **zero downtime** during node failures.
- Engineered **NGINX** load-balancing strategies and optimized SQL performance through query profiling and execution plan analysis, reducing **p99 latency by 35%** and achieving **sub-5-minute MTTR**.
- Centralized database replication monitoring using **ELK**, **Grafana**, and **Prometheus**, reducing root-cause analysis time by **70%** and accelerating incident response.
- Managed production **Kubernetes** infrastructure, resolving networking, scheduling, and resource bottlenecks while executing zero-downtime releases and infrastructure changes.
- Designed an **LLM-powered incident RCA pipeline** for automated log summarization, alert correlation, and runbook generation, reducing **MTTR by 35%** during production outages.
- Developed a **Transformer-based anomaly detection model** on Prometheus telemetry to proactively detect latency spikes and infrastructure saturation before customer impact.

- Built and maintained **Docker/Podman** containers, using **GitLab CI/CD** pipelines and **Harbor** (in-house container registry) for image builds and registry management, to standardize **development, testing, and deployment** environments and improve **consistency** across production releases.
- Define and track Service Level Objectives (SLOs) and error budgets across owned services, using **PagerDuty** for **on-call** alerting and **incident escalation**.
- Conduct manual **chaos/failure-injection testing** on staging infrastructure to validate failover behavior and resilience before **production rollout**.

Product and Platform Engineer (SRE) | Altimetrik Pvt. Ltd. (Client: Visa) | Sept 2023 – Dec 2025 | Bengaluru, India (Remote)

- Supported **24×7 production operations** for Visa's **450 PB Hadoop ecosystem** (HDFS, Hive, Presto), ensuring high availability and uninterrupted CI/CD pipeline operations across mission-critical data processing workloads.
- Automated infrastructure provisioning in **AWS** and configuration management using **Terraform** and **Ansible**, **reducing provisioning time by 70%** while improving deployment consistency.
- Optimized **Apache Airflow DAG** scheduling and dependency management, improving workflow **success rates from 88% to 96%** across 500+ daily production pipelines.
- Monitored **YARN ResourceManager** for Spark job queue-level allocation and managed **Kubernetes cluster** health optimizing resource allocation, resolving pending pods, & terminating rogue processes.
- Automated platform health checks, recovery workflows, and incident response using **Python** and **Bash**, reducing manual **operational effort by 60%**.
- Built **Grafana dashboards with Prometheus monitoring** to improve **SLAs/SLOs** tracking and reduce **MTTR by 60%** across big data services.
- Managed **Control-M** batch scheduling, CRON jobs, proactively resolving dependency failures and SLA breach escalations while supporting secure production releases and platform reliability.
- Deployed a **ServiceNow-integrated L1 support chatbot** backed by OpenAI models, fine-tuned on historical ServiceNow tickets, Confluence docs, Jira, and SOPs—serving as the first line of **incident resolution** for users.
- Configured and troubleshooted **Kerberos authentication, LDAP integration, and IAM policies** to secure access across the Hadoop ecosystem, supporting compliance with Visa's **security** standards.

EDUCATION and CERTIFICATIONS

MBA in Data Science | Amity University, 2025 | 8.4/10 CGPA

B.E. in Electronics and Communication | CMR Institute of Technology, Bangalore, 2022

Oracle Cloud Infrastructure (OCI) 2025 Certified DevOps Professional | Oracle Corporation, 2025

NOTABLE AI/ML and TECHNICAL PROJECTS

LLM-Powered Incident RCA Pipeline

- Designed and deployed an incident **RCA** pipeline integrating **Prometheus** and **ELK** log data with an **LLM-based** summarization layer, automatically correlating alerts across services and generating draft **runbook** recommendations for on-call engineers, contributing to a **35% MTTR reduction**.

Transformer-Based Anomaly Detection System

- Engineered a TensorFlow Transformer pipeline over Prometheus **time-series telemetry**, using multi-metric attention to detect **latency spikes** and infrastructure saturation ahead of customer impact; tuned for **94% precision** with **minimal false positives** through iterative threshold calibration.

On-Prem Incident Response Assistant (Gemma 4, Offline LLM)

- Explored running the **Gemma 4** model fully **offline** in a home-lab environment, pairing it with a log streaming layer to build a **proof-of-concept incident response** assistant that summarizes outages and drafts RCA steps without any data leaving the infrastructure.