SHREY SHAH

IT Infrastructure Engineer | Backup & Storage Specialist | AI/ML Engineering (LLM, RAG, Agents) | Automation
Vadodara, Gujarat, India | Shrey6131@gmail.com | +91 8469931021 | LinkedIn | github.com/shrey220201

CAREER SUMMARY

Infrastructure-focused IT Engineer with hands-on enterprise deployment experience across NAS/backup ecosystems (QNAP, NAKIVO, Veeam), virtualization platforms (VMware, Hyper-V, Proxmox), and CCNA-level networking. Promoted from intern to full-time at ITCG Solutions within 2 months based on technical performance and client delivery. Architected a dual-region enterprise IT environment (New York & Hong Kong) as a capstone project, integrating AD, Docker, Nextcloud, Zabbix, and Rocket.Chat. Completed the AI Engineer Core Track (LLM Engineering, RAG, QLoRA, Agents) under industry veteran Ed Donner, and now build applied AI and automation solutions (RAG assistants, agentic workflows, n8n, Microsoft Power Automate) at the intersection of infrastructure and AI/ML engineering.

TECHNICAL SKILLS

Storage & Backup: QNAP NAS/SAN (2–16 bay), RAID (0/1/5/6/10/50/60), SSD Caching, NAKIVO, Veeam, HBS, Storage Migration

Virtualization: VMware vSphere, Hyper-V, Proxmox, NComputing

Networking & Security: Routing & Switching, CCNA-level troubleshooting, Wireshark, Seqrite, Sophos, Firewall Management

Systems & Admin: Windows Server, Active Directory, Exchange Server, DNS, DHCP, Linux, PowerShell, Group Policy

AI & LLM Engineering: LLM Engineering, RAG Pipelines, QLoRA Fine-Tuning, AI Agents (LangGraph), RAGAS Evaluation, Python (Ollama / HuggingFace APIs)

Automation & Integration: n8n, Microsoft Power Automate, REST APIs, API Key & Secrets Management, Workflow Orchestration

Programming: Python, PowerShell, SQL, C/C++, ASP.NET

DevOps & Tools: Docker, Portainer, Zabbix, Nextcloud, Rocket.Chat, BookStack, osTicket, PuTTY

Web: HTML, CSS, JavaScript, Bootstrap, AngularJS, jQuery

PROFESSIONAL EXPERIENCE

ITCG Solutions Pvt. Ltd. — Vadodara, Gujarat Aug 2025 – Present

IT Technical Support Engineer

- Deployed QNAP NAS solutions (2–16 bay) across 10+ enterprise client sites, configuring RAID arrays (0/1/5/6/10), SSD caching, and storage expansion; reduced average setup time through standardized deployment playbooks.
- Integrated NAKIVO Backup & Replication for enterprise clients; managed a 4-server backup environment (SAP SRV, SRV HRD, SRV E-LAB NOTE, SRV TEAM), created repositories, and maintained backup job SLA compliance.
- Resolved a complex multi-layer NAKIVO upgrade failure on ARM64 QNAP TS-431X2 (signature expiry, incompatible package errors, QTS volume root-cause); diagnosed and resolved via SSH, avoiding full data migration.
- Managed and resolved 20+ client tickets/month across storage, backup, NComputing connectivity, and firewall issues with consistent SLA adherence via on-site and remote support.
- Built a RAG-based AI assistant for the support team that retrieves relevant past tickets and suggests troubleshooting steps, reducing time spent searching for resolutions.
- Developed automation workflows in n8n and Microsoft Power Automate to streamline repetitive operational tasks, using API-driven integrations with secure API key handling.
- Executed VM migrations (Hyper-V → Proxmox) and implemented QBelt VPN on QNAP for secure remote client access.
- Investigated Linux storage configuration on an SAP HANA production server; confirmed partition layout and assessed expansion feasibility without disrupting live ERP workloads.

ITCG Solutions Pvt. Ltd. — Vadodara, Gujarat Jun 2025 – Aug 2025

IT Technical Support Intern

- Completed hands-on training across enterprise solutions (QNAP NAS, NAKIVO, NComputing); participated in client site NAS installations, troubleshooting, and data recovery.
- Assisted senior engineers in backup monitoring, POCs, and client demos; supported CCNA-level networking tasks including routing, switching, and Wireshark packet analysis.

Nupur Software — Vadodara, Gujarat Jan 2022 – Mar 2022

Frontend Development Intern

- Developed the Hunt and Hire payroll system frontend using HTML, CSS, Bootstrap, jQuery, and AngularJS.

URO Web-soft Solutions — Vadodara, Gujarat Jun 2021

Web Development Intern

- Built job-oriented skills in ASP.NET (C#), SQL, and frontend web development in a professional client delivery setting.

KEY PROJECTS

Runbook Copilot — RAG & Agent-Based Infrastructure Incident Triage System | *Python, Ollama, Qdrant, LangGraph, HuggingFace, Modal, FastAPI*

- Built a retrieval-augmented, agent-based system for infrastructure incident triage over real Windows Server, Active Directory, Hyper-V, and Proxmox documentation, grounding remediation guidance in verified source content instead of model recall alone.
- Designed a confidence-gated LangGraph agent (runbook search, system status check, human escalation) with automatic escalation on destructive commands and low-confidence retrieval; adversarial testing exposed 7 of 12 gate bypasses (incl. prompt injection), all subsequently closed.
- Fine-tuned a Llama 3.2 1B incident classifier with QLoRA (4-bit NF4) on Modal GPUs; traced an initial 86.6% result to a labeling defect rather than data scarcity, and fixed it to reach 95.5% held-out accuracy using a leak-proof, source-level split.
- Built a RAGAS evaluation harness against real-world-grounded test tickets to measure answer quality rather than assume it.
- Architected a local-first, open-source stack (Ollama, Qdrant, HuggingFace) with an optional token-authenticated, GPU-backed Ollama deployment on Modal, switchable via a single environment variable; added a FastAPI chat UI with multi-turn session memory. Fully reproducible from a public GitHub repo.

Capstone: Dual-Region Enterprise IT Infrastructure (New York & Hong Kong)

- Architected and deployed a full dual-region enterprise IT environment simulating 300+ user scale across two geographic sites.
- Stack: Active Directory, Exchange Server, Apache, Docker, Nextcloud (file sharing), Rocket.Chat (messaging), Zabbix (monitoring), osTicket (helpdesk), AWS integration, and high-availability design.
- Implemented site-to-site DR strategy, failover procedures, and multi-region network design to meet enterprise RTO/RPO targets.

QNAP NAS & NAKIVO Backup Deployment for SMB Clients

- Configured NAS devices (2–16 bay) with RAID setups, SSD tiering, and NAKIVO integration for disaster recovery across multiple SMB client environments.
- Documented deployment procedures and runbooks; reduced onboarding time for repeat client configurations.

Network Design — NinjaPostal Corp (Tokyo, Toronto, Vancouver)

- Designed a secure, scalable multi-office network architecture with inter-site routing, VPN, and future-ready capacity planning.

BookStack Wiki & Domain Integration

- Deployed BookStack open-source wiki with SQL backend and Markdown integration; enforced LAN homepage redirection and GPO-based policy across domains for standardized access control.

EDUCATION

Post-Graduate Certificate in Network & System Administration Apr 2025

Conestoga College, Kitchener, ON, Canada | GPA: 3.0/4.0

Bachelor of Engineering in Computer Engineering Dec 2023

ITM (SLS) Baroda University, India | GPA: 7.79 / 10

CERTIFICATIONS

- NAKIVO Technical Professional — NAKIVO
- Google Cybersecurity Professional Certificate — Google / Coursera
- AI Engineer Core Track: LLM Engineering, RAG, QLoRA, Agents — Udemy, Instructor: Ed Donner (33.5 hrs, Aug 2026)
- CCNA (Cisco Certified Network Associate) — Cisco, Trained
- Python Fundamentals; Time Series Analysis with Python; PHP Development