

# SHIVANAND KUMAR

SENIOR DATA ENGINEER | AGENTIC AI & DATA PLATFORM AUTOMATION | SQL & DATABRICKS

kumarshivanand7@gmail.com | Bengaluru, India | linkedin.com/in/shivachauhan

## PROFESSIONAL SUMMARY

Senior Data Engineer with 5 years of experience delivering enterprise data platforms and AI-driven engineering automation across HP and Amazon. Built production pipelines, migration and observability frameworks, and agentic contract-generation workflows using SQL, Python, Databricks, PySpark, and AWS, with additional LLM-data and evaluation experience.

## CORE SKILLS

**Agentic AI & Automation:** Agentic workflows, RAG, vector databases, knowledge graphs, GitHub Copilot agent skills, structured generation and validation, data-contract automation, Microsoft Waza evaluation sets, LLM evaluation

**Data Engineering:** Advanced SQL, Python, PySpark, ETL/ELT, distributed processing, data modeling, pipelines, migration, data quality

**Data Platforms & Cloud:** Databricks, Delta Lake, Unity Catalog, Spark, AWS S3, Glue, Redshift, lakehouse systems

**Reliability & Analytics:** Observability, SLA monitoring, performance tuning, BI datasets, KPI design, reporting automation

## PROFESSIONAL EXPERIENCE

### HP | Research & Development Center | Bengaluru

**Data Scientist II** | Data Engineering & AI Platform Automation | Jun 2024 - Present

- Led end-to-end design and rollout of an **AI-driven data-product contract automation framework across 47 Databricks pipelines**, reducing manual documentation effort from days to hours.
- Owned the **five-phase agentic generation and validation control loop** for configurations, schemas, data sources, target DDLs, and 20+ transformations; integrated **Microsoft Waza evaluation sets** into the Data Quality framework for repeatability and regression checks before human approval.
- Built and deployed an **AI-powered KLO Agent across 54 production pipelines** using **RAG, a vector database, and a knowledge graph** to retrieve operational context, correlate failure signals, identify likely root causes, and recommend corrective actions; reduced recurring manual monitoring by **85-90%**.
- Led Splunk deprecation and logging standardization through a reusable migration and code-refactoring framework; migrated **40 production pipelines with zero migration-related runtime failures**, cutting manual effort by ~90%.
- Optimized firmware-log and Cloudflare telemetry pipelines through partitioning, distributed PySpark processing, join optimization, and caching, delivering a **~75% runtime reduction** while lowering compute overhead.
- Led Spark-based deduplication for the Hermes platform, reducing data volume from **21B to 400M rows** while improving downstream performance and infrastructure efficiency.
- Designed lakehouse observability for pipeline health, SLA tracking, proactive monitoring, and incident diagnosis, contributing to **~\$1.55M in annual infrastructure savings**.

### Amazon | Oct 2022 - Jun 2024

**Business Analyst** | Jul 2023 - Jun 2024 | Bengaluru

- Owned analytics for IXD-AMXL supply-chain systems supporting **15M+ users**; translated operational requirements into production SQL pipelines and BI datasets for dashboards and leadership decision-making.
- Automated DRR and SSD metric generation and validation using SQL and Python, cutting recurring reporting effort by **~70%**.
- Built Dynamic CPT transformations in SQL, encoding complex operational rules and reducing runtime by **~30%**.

**Recognition:** Director-level "Minesweeper" and "Automation Expert" awards for automation impact.

**Machine Learning Data Associate II** | Oct 2022 - Jul 2023 | Hyderabad

- Curated and quality-checked supervised fine-tuning datasets for LLMs through annotation, validation, and structured dataset-quality workflows.
- Evaluated prompt-response pairs for relevance, factual consistency, instruction adherence, and overall response quality; identified recurring failure patterns to improve dataset consistency and support LLM training and evaluation cycles.

### BYJU'S | Oct 2021 - Oct 2022

**Centre Head - Operations**

- Led a **16-member cross-functional team** across centre operations, scheduling, delivery, and performance management.
- Used SQL dashboards to drive **~30% quarterly growth** and achieve 100% centre ratings across North India.

## EDUCATION & CERTIFICATIONS

**M.Tech, AI & Data Science Engineering** | Indian Institute of Technology (IIT) Patna | Jul 2026 - Present

**B.E. in Mechanical Engineering** | University Institute of Engineering and Technology (UIET), Panjab University, Chandigarh | 2016 - 2020

**Certifications:** DeepLearning.AI - Supervised ML | Google - AI Essentials & Data Analytics | HackerRank - SQL (Advanced)