# Vinay Kumar N

*Senior Data Engineer at Finoa*
*Software Engineering; Distributed Systems*
*AWS Certified Solutions Architect - Professional*

narayanavinaykumar@gmail.com
+49-162-1518203 / +91-987-237-5950
/vinay26k /Vinay26k
/vinay.dev Credly/vinay26k


## Summary


**Senior Data Engineer** with **7+ years** building data platforms, governance systems, and internal tools across fintech, banking, retail, and consulting. **First data hire at Finoa**, working directly with the CTO to define the data strategy and build a modern platform from architecture through adoption.


***Core Competencies***

- **Data Platforms & Architecture**: Building lakehouse, streaming, and analytics platforms from scratch and aligning them to business goals.
- **Data Engineering & Governance**: Designing scalable pipelines, metadata systems, access controls, and observability frameworks.
- **Platform Thinking**: Building cost-aware, maintainable systems with a bias for in-house tooling, open standards, and low vendor lock-in.

## Skills & Tools

- **Languages & Frameworks** - Python, SQL, FastAPI, Postgres
- **Data Platforms** - Apache Spark, Apache Kafka, Apache Iceberg, Delta Lake, Snowflake, Amazon Redshift
- **Cloud & Platform** - AWS, Azure, GCP, Docker, Kubernetes, Terraform (IaC), Git
- **AWS Stack** - EKS, ECS, MSK, Lake Formation, Glue, EventBridge, S3, Aurora
- **Engineering Practices** - Data Modeling, ETL/ELT, Streaming Pipelines, Lakehouse Architectures, Data Governance, Observability
- **AI Engineering** - RAG, LangChain, Agentic Workflows, MCP
- **Open Source** - Contributions to open source, exploring Rust & TypeScript

## Experience

**Finoa** Berlin/Germany, DE
*Senior Data Engineer; Data Platform* *Oct 2024 - Present*

- **First Data Hire**: Defined the data vision, roadmap, and operating model for analytics adoption across business teams in direct collaboration with leadership.
- **Data Platform**: Designed and built Finoa's data platform from scratch across ingestion, storage, transformation, and governance. Core stack includes
  - AWS EKS, Apache Spark, Apache Iceberg, AWS MSK, S3, Amazon Aurora PostgreSQL, REST APIs, and Snowflake
  - Serverless architecture - AWS ECS, Python, ADBC, AWS Glue, AWS EventBridge, Redshift
- **Multi-Agent Developer Workflows**: Built multi-agent developer workflows to accelerate feature development and testing across the SDLC. Defined role-specific agents, guardrails, and orchestration workflows, and now building a higher-level agent to coordinate end-to-end development across multiple repositories.
  - Python, FastMCP, AWS Kiro, Amazon Q Developer
- **Platform Adoption & Operations**:
  - Partnered with Platform, Engineering, SecOps, Compliance, and business teams to onboard assets and automate operational workflows
  - Built ingestion and transformation pipelines, Lake Formation controls for data platform governance, observability metrics, and deployment processes
  - Tracked platform adoption with usage metrics and user feedback to improve usability
- **Business Impact**:
  - Reduced platform costs for a single source by **$33k** (**43%**) by replacing managed components with in-house pipelines on Kubernetes
  - Established a centralized, queryable data layer that replaced spreadsheet-driven reporting and enabled more reliable self-serve analytics for business teams

**N26** Berlin/Germany, DE

*Data Engineer; Data Governance* *Apr 2023 - Aug 2024*

- **Data Engineering**: Built scalable metadata and data pipelines on Apache Airflow and an in-house orchestrator running on Kubernetes. Developed custom connectors, containerized execution, and data quality observability around accuracy, completeness, and consistency.
- **Data Discovery Platform**: Designed and implemented a data discovery platform that improved access to data assets across the organization and reduced time-to-discovery for business users.
- **Data Governance Frameworks**: Contributed to the design and implementation of data governance frameworks, including automated access management controls and the generation of audit reports to support Security, Audit & Compliance requirements.
- **Execution Scope**:
  - Built Airflow pipelines for metadata ingestion and integrated ingestion and transformation jobs into the in-house orchestrator
  - Developed Python packages and frameworks for governance automation and data observability
- **Business Impact**:
  - Data discovery platform implementation reduced time required for data access by **40-50%**
  - **200+** MAU actively using the data discovery platform
  - Automated access request assessment and validation, removing **60-80 minutes** of manual effort per request

**Walmart Global Tech India** Remote/ Bangalore, IN

*Software Engineer III; Data Engineering* *Nov 2021 - Mar 2023*

- **Data Processing Pipelines at Scale**: Built data processing pipelines on GCP using Dataproc for compute and BigQuery & GCS for storage across multiple data formats.
  - Python packaging and development for common utilities
  - PySpark and Python pipelines for ingesting raw data and publishing refined datasets
  - Integration of data science related processing scripts
  - Observability & logging of data processing pipelines
  - Recovery from point-of-failure based on pipeline requirements
- **Self-Serve Data Platform**: Developed multiple backend iterations for a self-serve data platform with connectors to internal Walmart data sources, enabling analysts and scientists to access data without waiting for dedicated data engineering support. Stack: FastAPI, SQLAlchemy
- **Business Impact**:
  - Standardized new pipelines using reusable templates, reducing duplicated implementation effort
  - Self-serve platform reduced the wait for data engineer allocation, which previously averaged a sprint cycle

**Neudesic Technologies Pvt Ltd** Remote/ Hyderabad, IN

*Senior Data Engineer; Big Data Engineer* *Nov 2020 - Nov 2021*

- **Azure Migration & Streaming**: Migrated workstation telemetry processing from an on-premise Hadoop/Cloudera stack to Azure and built **5** Spark Structured Streaming jobs using Databricks and Azure Event Hubs.
- **Data Integration Patterns**: Designed and implemented data movement patterns across cloud and on-premise systems, including
  - GCP BigQuery to ADLS Gen2 - Databricks as compute
  - Hive to ADLS Gen2 - Kafka MirrorMaker + Azure Event Hubs as streaming platforms, Databricks as compute
  - Apache NiFi to ADLS Gen2 (& vice versa) - Azure Event Hubs as a streaming platform, Databricks as compute
  - ADLS Gen2 to QlikView - Databricks as compute
- **Platform Enablement**: Built a logging framework on Azure Application Insights for Databricks jobs and enabled QlikView reporting on refined ADLS Gen2 data through **external tables**.
  - Article - **Azure Databricks QlikView Integration**
- **Business Impact**: Improved pipeline efficiency to support **4 runs/day**, up from **2 runs/day**, at the same operating cost.

**MAQ Software** Hyderabad, IN
*Software Engineer II* *Sep 2018 - Oct 2020*

- **Azure Migration**: Migrated multiple projects from on-premise VM-based systems to the Azure data stack using
  - Azure Data Factory - SSIS replacement & Jobs orchestration
  - Azure Databricks - compute for jobs execution
  - Azure Function App - replacement for our custom SSIS module
  - Azure Logic Apps - to trigger processing based on some events
  - Azure Event Hubs - streaming platform
  - Azure SQL Database - replacement for on-premises SQL Servers
  - Azure Synapse Analytics (formerly Azure SQL Data Warehouse) - replacement for on-premises SQL data marts and PDWs
- **Data Engineering Utilities**: Built reusable Databricks utilities for large-scale fuzzy string matching across **80M x 20K** records, parallel database interactions, ADLS Gen2 writes, row-count validation, and schema verification.
  - Article - **big-data-fuzzy-matching-in-databricks**
- **Stream Analytics - Real-Time**: Built a real-time data integration solution using the Cosmos DB change feed with
  - Azure Functions to process change feed events from Cosmos DB and ingest them into Event Hubs
  - Azure Stream Analytics to process Event Hubs data and load a Power BI dataset
  - Power BI dashboards refreshed in real time for operational insights
- **Advanced Data Solutions**: Built a Neo4j-based recommendation engine PoC selected for further development, deployed a custom Neo4j image on Azure Container Services, and delivered a time-series forecasting project focused on event-driven prediction and causality analysis.

## Certifications

[Recent ↑]

- ☆ AWS Certified Solutions Architect - Professional
- ☆ AWS Certified Data Engineer - Associate

[ordered by earliest ↓]

- ✹ Exam 761: Querying Data with Transact-SQL
- ✹ Exam 778: Analyzing and Visualizing Data with Power BI
- ✹ Exam 767: Implementing a Data Warehouse
- ✹ Exam 768: Developing SQL Data Models
- ✹ Microsoft Certified: Azure Developer Associate
- ✹ MCSA: SQL 2016 Business Intelligence Development - Certified 2019
- ✹ Neo4j Certified Professional
- ✹ Python for Data Science
- ✹ Applied Data Science I: Scientific Computing & Python (with honors)
- ✹ Applied Data Science II: Machine Learning & Statistical Analysis (with honors)

## Education

**International Institute of Information Technology** Online/ Bangalore, IN
*Executive PGP in ML & AI; GPA: 3.82/4* *Sept 2021 – Nov 2022*

**Lovely Professional University** Punjab, IN
*Bachelor Of Technology Computer Science; GPA: 8.09/10* *Aug. 2014 – May 2018*

**Sri Chaitanya Junior College** Andhra Pradesh, IN
*Higher Secondary (MPC); Percentage: 91%* *June 2014*

**Ravindra Bharati School** Andhra Pradesh, IN
*Secondary School (SSC); GPA : 9.8/10* *May 2012*