

Neeraj Bhatyal

Enterprise Data Architect | AI & Forward Deployed Engineering (FDE) | AWS Cloud | Data Governance | SOC 2 Type II Readiness

New York, NY · [linkedin.com/in/neerajbhatyal](https://www.linkedin.com/in/neerajbhatyal) · neerajbhatyal.com/contact

Summary

- Enterprise Data Architect with more than two decades of experience designing, architecting and delivering enterprise-scale data, cloud, analytics and AI platforms.
- Leads enterprise data governance, data inventory, data flow mapping and data lineage initiatives, and drives SOC 2 Type II readiness from current-state assessment and gap analysis through control mapping and remediation.
- Strong AWS cloud experience across S3, EC2, EKS, Lambda, DynamoDB, RDS, Redshift and IAM, delivering secure, scalable and high-performance data solutions.
- Hands-on with generative AI, LLMs, RAG, AI agents, LangChain, LangGraph, vector databases, Amazon Bedrock, SageMaker, Databricks AI and MLOps; applies AI-assisted and forward deployed engineering to deliver measurable outcomes.
- Proven expertise in big data technologies including Spark, Kafka, Hadoop, Hive, HBase, Cassandra and Databricks, building data lakes, lakehouses, ingestion frameworks and distributed processing systems.
- Technology leader experienced in leading global teams, driving platform modernization and partnering with engineering, security, compliance and business stakeholders.
- Author of *The Enterprise Data Architect in the Age of AI* and the forthcoming two-volume *The Forward Deployed Engineer*.

Experience

Enterprise Data Architect · Hitachi Digital Services

North America · Oct 2026 – Present

- **Enterprise data architecture & governance:** lead enterprise-wide data inventory, data flow mapping and data lineage initiatives to establish end-to-end visibility, traceability and governance of sensitive and restricted data throughout its lifecycle.
- **SOC 2 Type II readiness:** drive readiness initiatives including current-state assessments, gap analysis, control mapping, remediation planning and cross-functional coordination to align data architecture, security practices and operational processes with compliance requirements.
- **Data lineage & medallion architecture:** establish data movement tracking and lineage across bronze, silver and gold layers, strengthening traceability, quality, access controls and governance.
- **AWS cloud data architecture:** use AWS and modern data platforms to support secure data ingestion, processing, storage, access management and metadata management.
- **AI & forward deployed engineering:** apply AI-assisted engineering and FDE principles to translate business, data and compliance requirements into practical technical solutions, working closely with stakeholders to identify gaps, accelerate implementation and deliver measurable outcomes.
- **Data security, privacy & compliance:** strengthen governance of sensitive data through classification, cataloging, metadata management, access controls and documented data flows.
- **Stakeholder leadership & documentation:** partner with engineering, security, compliance and business teams to define architecture standards, document current and target states, and drive initiatives from assessment through implementation.

Vice President · JPMorgan Chase & Co.

New York, NY · Jun 2017 – Sep 2026

- Led architecture, strategy and implementation of enterprise-scale big data and cloud data platforms using Hadoop, Spark, Kafka, Hive, HBase, Databricks and AWS.
- Designed and implemented an enterprise data lake on Cloudera Hadoop with reusable ingestion frameworks for structured, semi-structured and unstructured data; built a Cloudera security microservice in Java Spring Boot for secure, scalable ingestion.
- Designed a Databricks lakehouse integrating marketing, campaign, survey, billing, email and operational data; scaled it to 50+ data sources and 100+ production pipelines for payment-marketing analytics, and automated data preparation that saved thousands of analyst hours.
- Built a cloud cost-optimization analytics solution combining AWS CloudWatch, Databricks system tables and platform telemetry; analyzed EC2, EKS and Databricks workloads and eliminated unnecessary compute, saving hundreds of thousands of dollars.
- Supported machine learning model development, deployment and optimization across distributed Spark environments, and evaluated and built proofs of concept for generative AI, LLMs, RAG and AI agents.
- Defined architecture standards, engineering practices and modernization roadmaps, and led globally distributed engineering teams.

Programmer Analyst · Selsoft Inc.

Allen, TX · Oct 2014 – Jun 2017

- Led big data consulting and solution delivery with Hadoop, Spark, Kafka, Hive, HBase and Cassandra.
- Migrated enterprise pipelines from Greenplum to Spark on Cloudera Hadoop and built a metadata-driven ingestion framework that accelerated data onboarding.
- Migrated KNN models to Python and Spark MLlib and built clickstream analytics solutions.

Lead Software Engineer · Torres Networks

Bengaluru, India · Sep 2013 – Sep 2014

- Designed a network intelligence platform capturing and analyzing network traffic for mobile analytics and BI, with Hadoop MapReduce ingestion pipelines.
- Developed a 3GPP Release 10 compliant PCRF solution with node-locked licensing.

Senior Associate · Cognizant Technology Solutions

Pune, India · May 2012 – Sep 2013

- Built an automated testing framework for LTE network components, automating execution across multiple LTE interfaces.

Senior Development Consultant · CSG International

Bengaluru, India · May 2011 – May 2012

- Developed policy and charging solutions for telecom operators, and an Android demo application for sales teams.

Senior Subject Matter Expert · Amdocs

Pune, India · Jun 2008 – May 2011

- Implemented and customized Amdocs Enabler for global telecom providers; integrated it with Vertex tax calculation and built Oracle-to-SAP interfaces.

- Built an inventory control system for a semiconductor manufacturer and features for engine-testing and simulation data platforms.

Skills

Data architecture & governance	Enterprise data architecture, data governance, data lineage, data inventory and flow mapping, medallion architecture, data classification and cataloging, metadata management, access controls, SOC 2 Type II readiness, gap analysis and remediation
AI & machine learning	Generative AI, LLMs, RAG, AI agents, LangChain, LangGraph, Amazon Bedrock, SageMaker, Databricks AI, vector databases, MLOps, forward deployed engineering
Cloud	AWS (S3, EC2, EKS, Lambda, DynamoDB, RDS, Redshift, Route 53, IAM)
Big data	Databricks, Spark, Kafka, Hadoop, Hive, HBase, Cassandra, Pig, Oozie, YARN, ZooKeeper
Languages	Python, Java, Scala, C++, TypeScript, Bash, Perl, TCL/TK
Databases	Oracle, Greenplum, Teradata, MySQL, SQLite, TimesTen
Frameworks & DevOps	Spring Boot, Angular, Kubernetes, Git, Bitbucket, Jenkins, Maven, Gradle, SBT, CI/CD

Education

Bachelor of Engineering in Computer Science · **Government College of Engineering & Technology, University of Jammu**

Jammu, India · 2001 – 2005

Books

- *The Enterprise Data Architect in the Age of AI*: architecting the intelligent enterprise through cloud, lakehouse, data products and generative AI (free PDF).
- *The Forward Deployed Engineer*, Volumes 1 and 2: foundations, agents, MCP and engineering in the field (forthcoming).