RAHUL JOSHI

DISTINGUISHED DATA ENGINEER | DIRECTOR

CONTACT

+240-492-9437
reachrahuljoshi@gmail.com
LinkedIn
Washington DC, USA

EDUCATION

2009 - 2011
IIT KHARAGPUR, INDIA

- Master of Technology, Computer Science & Engg
- GPA: 9.11/10

2002 - 2006
PUNE UNIVERSITY, INDIA

- Bachelor of Engineering
- First Class with Distinction

SKILLS

- Cloud Native Data Platforms
- Multi-Tenant Data Lakes & Lakehouses
- ML & Analytics Environments
- Real Time & Batch Data Engineering
- Data & Analytics for Large Scale Financial Services

LEADERSHIP

- Modern Data Platform Strategy
- Data Architecture Patterns & Governance
- Enterprise Scale Platform Adoption & Modernization
- Cross Functional Domain Alignment & Execution

PROFILE SUMMARY

Innovative and impact-driven data engineering leader with 19+ years of experience architecting modern cloud-native data platforms that power intelligent products, decisioning systems, and analytics at scale. Currently recognized as a Distinguished Engineer at Capital One — an elite technical leadership role held by fewer than 0.5% of associates — specializing in building scalable, secure, reusable modern data platforms and analytics environments. Known for driving technical strategy, cost optimization, and mentoring high-performing engineering communities.

WORK EXPERIENCE SUMMARY

Capital One 2022 - PRESENT

Distinguished Data Engineer & Director

As a Distinguished Engineer at Capital One, I lead data architecture across some of the company's most critical platforms and systems. My work has spanned Enterprise Data Platforms Tech, and currently focuses on Card Tech—where I lead architecture for modern credit card core systems that manage credit card accounts and transactions.

EY 2019 - 2022

Senior Manager , Data & Analytics, Financial Services

As a Senior Manager within EY's Data and Analytics practice for the Financial Services sector, I led high-impact engagements in Banking, Capital Markets, Wealth Management, and Insurance sectors. I specialized in Next-Gen Data Architecture, driving cloud adoption, data lake implementations, and advanced analytics solutions to help clients transform their data ecosystems and drive strategic value.

IBM 2015 - 2019

Sr. Managing Consultant, Big Data & Analytics, Financial Services

Started as a Managing Consultant where I architected and delivered hybrid cloud data engineering solutions, progressed to Sr. Managing Consultant to lead the architecture, engineering, and implementation of large-scale hybrid data platforms for the financial services sectors. Engineered scalable, fault-tolerant data solutions enabling real-time analytics, customer insights, and fraud detection, significantly improving operational efficiency and risk management.

Persistent Systems Limited 2007 - 2015

Architect

Progressed from Engineer to Technical Architect role, leading end-to-end architecture and development of data platforms and cloud-based analytics solutions. Led IBM-partnered platform delivery for U.S. clients, built reusable frameworks and foundational components for Big Data enterprise scale adoption, demonstrating deep engineering expertise and platform vision.

NVIDIA 2006 - 2007

Software Engineer Trainee

Worked on quality assurance and validation of GPU device drivers and media decoders for NVIDIA products, focusing on stability, performance, and compliance.

WORK EXPERIENCE DETAILS

Capital One 2022 - PRESENT

Distinguished Data Engineer | Director Level IC | Enterprise Data Tech → Card Tech

Distinguished Engineer - Strategic Enterprise Leadership

Capital One Distinguished Engineers (DEs) are deep technical experts and enterprise-level visionaries who drive long-term strategy, shape technology platforms, and mentor senior engineering talent. The role demands innovation, strategic influence, and recognized leadership across organizational boundaries. As the Distinguished Engineer representing Card Tech, I specialize in data & analytics platforms, data architecture, and data engineering domains—playing a critical role in shaping Capital One's next-generation data & analytics systems and influencing enterprise platform direction.

Key Contributions & Scope:

- Represent Card Core Tech in Capital One's Distinguished Engineer community, bringing Card's perspective into enterprise-wide platform decisions and aligning long-term data architecture with business goals.
- Define and publish reusable architecture patterns and data platform frameworks, adopted across Card and Enterprise Data Tech—accelerating lakehouse adoption and enabling scalable, cloud-native analytics.
- Drive Capital One's cloud data strategy by authoring white papers, internal blogs, and technical presentations; align executives, product, and platform leaders around next-generation architecture initiatives.
- Mentor and develop senior engineering talent, lead peer reviews, and sponsor bar-raising programs across Card's most critical data and ML platforms—elevating technical rigor and design maturity.
- Contribute thought leadership at internal DE Summits and External Forums, shaping the organization's stance on modern data architecture and influencing industry trends beyond Capital One.

Enterprise Data Tech → Card Tech

Key Contributions:

- Leading the data architecture for core systems that manage credit card accounts and transactions—modernizing foundational platforms that power Capital One's Card business, while driving alignment with enterprise data platforms to ensure scalability, governance, reuse, and smooth adoption across Card teams.
- Architected a Machine Learning Zone for production-scale model training, integrated with Capital One's internal ML platform—bringing the entire Card data science and analytics environment into a governed, secure, and enterprise-standard framework for data protection and operational scale.
- Proposed and implemented reusable architecture patterns across the enterprise data lake, Snowflake warehouse, and emerging Lakehouse solutions—accelerating adoption across Card, Auto Finance, and Retail Bank.
- Drove major cloud cost optimization efforts across enterprise data platforms—identifying key engineering opportunities that resulted in over $10M in annual savings.
- Led architecture of Capital One's enterprise-scale data lake (300+ PB) and multi-tenant cloud warehouse (50+ PB), enabling regulatory reporting, business analytics, and decision intelligence for thousands of users and hundreds of applications.
- Mentor senior engineers and serve as a peer reviewer for architecture, design, and code within Capital One's Distinguished Engineer community—shaping technical direction, judging critical initiatives, and driving engineering excellence across teams.

WORK EXPERIENCE DETAILS

EY 2019 - 2022

Senior Manager , Data & Analytics, Financial Services

Key Contributions:

- Architected and implemented a cloud-based enterprise data lake and Customer Analytics Hub for a major U.S. bank using AWS and Snowflake. Enabled ingestion and processing of real-time customer interactions and account updates across multiple lines of business. Supported identity resolution and powered workflow automation through integration with PEGA.
- Migrated critical BI and servicing reporting workloads to Snowflake, including 1,000+ fact and dimension tables and 10,000+ datasets. Improved performance, accuracy, and operational visibility of Auto Servicing dashboards in Power BI.
- Defined and delivered data strategy for a U.S. government-backed financial institution, leading the migration of on-prem workloads to AWS. Architected an AWS-based data lake and Dremio virtualization layer to support CFO attestation, regulatory reporting, and data lineage.
- Designed and implemented an evidence-based resiliency framework for a global investment bank using Neo4j graph modeling. Integrated telemetry, reference data, and technology network relationships to monitor dependencies and inform recovery strategy—reducing systemic operational risk.
- Led internal audit and architecture review engagements for a global financial institution. Assessed enterprise data architecture, analytics workflows, and platform readiness. Identified key risks and influenced audit and data architecture strategy presented at the board level.
- Drove platform modernization and cloud adoption across large number of client engagements. Provided architecture leadership for transformations involving AWS, Snowflake, Databricks, Kafka, and open-source tools—aligning platform capabilities with evolving regulatory and business requirements.

IBM 2015 - 2019

Managing Consultant → Senior Managing Consultant - Big Data & Analytics
Apr 2015 - Apr 2019

Key Contributions:

- Architected and delivered a hybrid cloud data lake for one of the largest U.S. auto insurers in collaboration with IBM. Integrated 100+ source systems, ingesting 5B+ records daily and supporting 25,000+ batch and streaming jobs. Enabled unified consumption-ready data across cloud and on-prem for analytics, ML, and regulatory reporting.
- Designed a centralized, distributed, highly available data transport platform using Apache Kafka, enabling secure, reliable delivery of 4B+ daily events across 100+ topics and 7,500+ partitions. Supported streaming data flows across legacy and modern systems with high resiliency.
- Developed scalable Lambda and Kappa architecture-based pipelines for near real-time business analytics and operational decisioning. Enabled timely insights for customer service, operations, fraud, and marketing.
- Built Spark and MapReduce pipelines to support 10,000+ daily data engineering workloads. Enabled business-critical use cases like SVOC (Single View of Customer), executive dashboards, and fraud detection analytics.
- Reduced customer service call volume by deploying predictive models for call intent (e.g., one-time payment, ID card delivery) and automating 1.5M+ IVR actions. Achieved over 60% acceptance and 80% task completion rate.
- Delivered predictive address change notifications to over 1M customers using USPS data and service event history, with 75%+ completion rate through mobile channels.

WORK EXPERIENCE DETAILS

- Developed a real-time, exploratory fraud analytics store supporting interactive analysis on terabytes of historical and streaming data. Enabled detection of anomalous patterns and accelerated fraud response across lines of business.

- Led client-facing technical engagements and cross-functional teams, serving as trusted advisor to enterprise architects, data leaders, and operations executives. Delivered measurable value in cost, efficiency, and platform scalability.

Persistent Systems Limited 2007 - 2015

Software Engineer → Architect - Big Data & Analytics
Feb 2017 - Apr 2015 | Pune, India → Washington, DC, USA

Key Contributions:

- Led architecture and delivery of a next-generation hybrid data lake platform in partnership with IBM, enabling real-time data processing and advanced analytics for a major U.S. insurance client. Served as the primary technical liaison between Persistent and IBM stakeholders in the U.S.

- Built and scaled Persistent's Big Data Analytics Library (PEBAL)—a reusable set of analytics accelerators and pipelines supporting user profiling, brand sentiment analysis, recommendation engines, and structured/unstructured data integration.

- Designed and delivered analytics solutions for large-scale consumer-facing platforms:
  - Built a brand management and user segmentation solution for a top Canadian wealth and banking institution.
  - Developed recommendation systems and behavior analytics pipelines for a major U.S. online retailer.
  - Implemented Twitter feed analytics pipelines and dashboards for ShareInsights.org and Indian Premier League (IPL), enabling real-time social listening and marketing insights.

- Delivered core components for Microsoft's big data ecosystem as part of a strategic engineering engagement—ported Apache Hadoop to Windows, enabling the launch of HDInsight, Microsoft's Hadoop-based PaaS on Azure.

- Engineered Microsoft's ODBC Driver for Apache Hive, implementing support for both HiveServer1 and HiveServer2 to enhance SQL-based access to big data.

- Built bi-directional Sqoop connectors for seamless integration between Apache Hadoop and Microsoft SQL Server / Parallel Data Warehouse (PDW), accelerating enterprise data migration and hybrid analytics adoption.

- Developed software products in data security and information retrieval space for early stage startups, including:
  - Voltage SecureMail plugin for Microsoft Outlook/Outlook Express, enabling secure messaging.
  - A vertical search engine for an e-commerce platform, supporting faceted navigation and product discovery.

- Mentored junior engineers and led cross-functional teams, contributing to Persistent's growth in Big Data and Analytics competencies and deepening cloud integration capabilities across client programs.

NVIDIA 2006 - 2007

Software QA Engineer Trainee - GPU Device Drivers & Decoders

Key Contributions:

- Worked on quality assurance and validation of GPU device drivers and media decoders for NVIDIA graphics products. Focused on ensuring stability, performance, and compliance through rigorous testing of video rendering pipelines, hardware acceleration features, and multimedia components across platforms.
- Validated GPU device drivers and multimedia decoders across hardware configurations, operating systems, and rendering pipelines—ensuring compliance with DirectX and OpenGL standards.
- Collaborated with driver and hardware teams to troubleshoot bugs, escalate firmware/driver integration issues, and deliver high-stability releases aligned with new GPU product launches.