

## Arya Basu

+1 732 372 3332 | aryabasu2k@gmail.com

Data Blogs: <https://shorturl.at/glwTY>

### Objective:

Accomplished big-picture principal data architect with 25 years of experience. Proven track record in developing, owning, and implementing data strategy that enable accelerated maturity and transformational business growth, aligning and guiding stakeholders, advocating for effective data architecture solutions, and ensuring data accuracy and accessibility. Strategy driven leader with hands-on experience in data architecture and modelling, data engineering & analytics, ML, GenAI, MLOps, and data governance. Not afraid to share well thought out POVs and contrarian views aligned with the strategic vision while striving for consensus in implementation technologies and methodologies.

### Professional Experience:

#### Capital One Financials (2019 – Current) – Director of Data Architecture

**Platform as a Service (PAAS):** Led a cross-functional cloud, data, and AI team of 30+ data professionals to implement a batch and stream processing central data processing and feature engineering platform for all analytics and data science needs, including Customer acquisitions evaluation and fraud detection by enabling batch analytics, real-time and reverse-ETL use cases in credit space. Leveraged Databricks, AWS EMR and Glue clusters, Spark, Kafka, and Kinesis suite.

**Strategy Evangelism and Execution:** Spearheaded the operationalization of the company's end-to-end data transformation and decisioning strategy via effective data product strategy ensuring alignment with strategic goals while actively aligning stakeholders across the organization to ensure a cohesive approach. Written several in-depth blogs on analytical data patterns and maturity journey to help teams and leaders understand data better.

**Data Products and Pipelines:** Spearheaded the enterprise effort in strategizing, building, and deploying a data product model that is MECE at the entity level and architected data ingestion pipelines and platforms to integrate data from various sources, ensuring clean, reliable, and high-quality data product assets.

**Consumer focused Data Systems Design:** Architecture and design data systems and platforms driven by accrual consumer needs and brings data publishers and consumers together via horizontal architectural layers such as self-service, centralized metadata exchanges, data governance etc. alongside the vertical data processing layers to support next-generation data and analytics capabilities.

**Data Architecture and Governance Principles:** Devised and socialized data architecture principles alongside championing Data Governance COE. Identified and addressed resistance or lack of awareness regarding the data vision and strategy, via data blogs and delivering compelling presentations to persuade stakeholders

**Core Enterprise Data Model:** Designed and maintained a comprehensive core enterprise data model, educating stakeholders on its importance and usage.

**Machine Learning Infrastructure as a service:** Built a standardized enterprise MLOps infrastructure for all LOBs with a robust self-service driven workflow layer enabling very low Time-to-market for new machine learning training pipelines

**Data Quality Ownership:** Owned the quality of ingested data, treating data as a product and ensuring its ownership and reliability.

**Finance Domain Solutions:** Developed reusable data products for the finance domain, applicable across the enterprise.

**Data Infrastructure:** Led the construction and deployment of robust, scalable, and secure data infrastructures. Executed metadata catalogue management and governance initiatives to ensure data precision and consistency.

**Cloud native Solutions:** Directed the creation and orchestration of cloud-based data solutions and architectures, ensuring data security and regulatory compliance related data access via proper IAM and semantic layer policies.

**Big Data Technologies:** Extensive experience and application of technologies such as Apache Spark, Spark Structured Streaming, AWS Dynamodb etc. alongside cloud-based innovations, to develop cutting-edge data and analytics solutions.

**Team Leadership:** Built and managed a high-performing data architecture team, leading, motivating, and developing team members.

**Stakeholder Partnerships:** Built and nurtured partnerships with key stakeholders, aligning their data requirements with the overall data strategy.

**Technical Liaison:** Acted as a liaison between technical and non-technical teams, simplifying complex concepts and promoting data-driven decision-making.

**Vendor Management:** Managed vendor relationships, negotiated contracts, and optimized vendor services to achieve cost savings of 20%.

**Collaboration with Engineering Teams:** Worked closely with the data platforms engineering team to implement data technology strategy, ensuring practical and effective architectural decisions.

## **Cognizant Tech Solutions (2015-2019) – Senior Manager Data Engineering and Architecture**

**Data Standards Development:** Developed and implemented data standards and best practices.

**Data Security and Privacy:** Established robust data security and privacy protocols, ensuring regulatory compliance.

**Data Exploration and Analysis:** Explored and analysed data to gain business understanding, establish data quality, summarize results, and communicate insights to stakeholders.

**Database Systems Development:** Led teams in database design requirements, data modelling, database performance management, and support for machine learning solutions.

**Metadata Catalogue and data ontology:** Executed a comprehensive metadata management layer with automated self-service layer and workflow capabilities that works as a central repository for the enterprise data entities and data products, data features improving data quality, standardization, and trust. Used ontology and LLM to connect similar entities and data attributes form across LOBs

**Cloud Platforms:** Led a team of data engineers and architects, ensuring the development of scalable and secure data infrastructures using both Azure and AWS clouds

**Big Data Technologies:** Implemented and managed high-performance big data tools and DataLakeHouse architecture like Apache Spark, Kafka, Databricks, Snowflake etc.

**Data Model Design:** Designed and maintained data models to support business intelligence and analytics initiatives.

**NoSQL/Graph Systems:** Worked extensively with schema design and single-table-data-modeling for NoSQL and graph systems like Neo4J, MongoDB, DynamoDB etc.

**Agile/Scrum:** Employed Agile/Scrum development methodologies to enhance team productivity and project delivery.

**Sensitive Data:** Ensured compliance with data security and regulatory requirements; built a data masking and tokenization solution to resolve and clean 64 million PII data points from

a DataLake

**Innovation:** Promoted a culture of innovation and data-driven decision-making across the organization

**Stakeholder Management:** Effectively collaborated with diverse stakeholders to understand their data needs and deliver solutions aligned with strategic objectives

### **Cognizant Tech Solutions (2008-2015) – Manager Data Engineering and Analytics**

**Data Warehouses:** Developed and maintained data warehousing solutions using Redshift and Snowflake

**EDA and DDD:** Worked extensively on distributed and microservices architectures leveraging Event Driven Architecture (EDA) and Domain Driven Design (DDD) principles

**Stream Processing:** Developed streaming applications using Kafka, AWS Kinesis, and other real-time data solutions.

**BI/DW Design Principles:** Applied BI/DW design principles and dimensional modeling to build robust analytics solutions.

**Enterprise Applications:** Worked with SAP, SFDC, and other enterprise applications, leveraging functional and technical knowledge of application data models and domains.

**Visualization Tools:** Utilized Tableau and Power BI for data exploration and visualization, providing actionable insights.

**CI/CD Methodologies:** Employed branching, build, deployment, and CI/CD methodologies such as Octopus and Bamboo.

**Metadata and Governance:** Engaged in metadata, data quality, and model governance practices for banking and healthcare clients

**MLOps and DevOps:** Contributed to the development of DevOps/MLOps processes and frameworks for both Azure and AWS ML platforms

**Cloud Migration:** Led and contributed to data modernization project from on-premises to AWS and Azure cloud platforms for multiple clients

### **PWC India (2006-2007) – Senior Data Engineer**

- Data and ETL consultant and architect – Oracle, SQ Server, Informatica, ERWIN modeler
- Data architecture for complex OLTP and OLAP systems

### **Tata Consultancy Services (2003-2006) – Data Engineer, Data Modeler**

- Data Architect
- Data modeller

### **RS Software India (2000-2003) – Software Engineer**

- Software Engineer
- Data Engineer

### **Certifications & Education:**

- AWS Certified Machine Learning – Specialty
- AWS Certified Solutions Architect Associate
- Bachelor of Technology in Electronics and Computer Engineering (1996-2000)

### **Core Competencies**

- Data and ML Strategy and Architecture
- Data Engineering: Data lake solutions, data analytics, data pipelines, data modelling

- Technologies: Spark, Databricks, Hive, ADLS, PostgreSQL, Azure, AWS,
  - Data Visualization: Tableau, Power BI
  - Programming: Python, PySpark, SQL
  - Agile Methodologies: Scrum, CI/CD (Octopus, Bamboo)
  - Database Systems: NoSQL (Neo4J, MongoDB, DynamoDB), Data Warehouse (Snowflake)\
  - Cloud Data Warehouses - Databricks & Snowflake
  - Data Lake, Data Warehouse, Bigdata, Stream Processing
  - Evert Driven Architecture and stream data processing
-