

GOKULNATH RAGHAVAN RAGHAVAN

Email: Gokulnath.raghavan@gmail.com | Phone: +91 8073180406

SUMMARY

AWS Certified Cloud Practitioner and Microsoft Certified Azure Fundamentals (AZ-900) professional with 16+ years of experience designing and implementing scalable, secure, and cost-effective cloud solutions. Expertise in cloud architecture, data engineering, and ETL framework design across AWS, Azure, and on-premises environments. Proven track record of delivering complex cloud migration and modernization projects for global clients that reduce operational costs and improve performance. Skilled technical manager with experience leading data engineering teams to deliver enterprise-scale cloud solutions.

PROFESSIONAL EXPERIENCE

Ernst & Young (EY)

Manager - Cloud Solutions & Data Engineering | Jan 2025 - Present

Enterprise Data Warehouse Migration & Orchestration Framework

- **Led and managed** a team of data engineers to design and develop a metadata-driven AWS ETL orchestration framework replacing legacy Oracle Data Integrator workflows
- **Architected** a fault-tolerant, restartable ETL pipeline using AWS Step Functions, Lambda, Glue, S3 and Redshift
- **Designed and implemented** a multi-layered data processing pipeline (Raw → Normalized → Stage → Intermediate → DW) with comprehensive error handling and retry mechanisms
- **Developed** a sophisticated control framework in Python that enables:
 - Tier-based table processing with dependency management
 - Automatic failure recovery with configurable retry limits
 - Detailed execution logging and performance monitoring
 - Automatic record counting and data quality validation
 - Dynamic restart capabilities for failed jobs
- **Created** a modular, configuration-driven approach using Python, Lambda, and Step Functions that reduces code redundancy and improves maintainability

- **Implemented** comprehensive error handling with MAX_RETRIES logic, ensuring pipeline reliability while preventing infinite failure loops
- **Established** S3 partitioning strategies and optimized Redshift tables, resulting in 50% faster query performance
- **Mentored** junior engineers on AWS best practices, serverless architecture patterns, and ETL design principles
- **Conducted** code reviews and architectural assessments, ensuring high-quality deliverables and knowledge sharing
- **Reduced** operational costs by 30% compared to on-premises solution while improving data accessibility

IBM India Pvt Ltd

AWS Solutions Architect | Apr 2022 - Jan 2025

Data Governance and Regional Compliance: EIG to DDEP Workflow Optimization

- **Designed and implemented** a sophisticated data filtering model to exclude China customer data from enterprise workflows
- **Applied filtering mechanisms** at the Enterprise Information Governance (EIG) layer before transmission to Data Driven Enterprise Platform (DDEP)
- **Ensured compliance** with data privacy regulations while maintaining system performance
- **Optimized data pipelines** to incorporate filtering logic without performance degradation
- **Implemented** comprehensive testing, documentation, and monitoring procedures

Azure Metadata-Driven Test Data Generation Framework

- **Designed and implemented** a flexible metadata-driven framework for generating test data on Azure
- **Utilized** Azure Data Factory for orchestration and Azure Functions for serverless execution
- **Developed** Python scripts to parse metadata from attribute, rule, dataset, and process specification sheets
- **Implemented** dynamic data generation logic based on configurable rules and data types

- **Improved** testing efficiency by automating the creation of realistic and compliant test datasets

IQVIA Japan Data Curation Project

- **Led architecture** for pharmaceutical sales data analytics platform using AWS Glue, S3, and Athena
- **Developed** PySpark ETL jobs processing diverse data sources (CSV, Excel) into a unified analytical platform
- **Implemented** robust data validation and error handling mechanisms ensuring data quality
- **Created** Athena tables and views enabling efficient querying of curated datasets
- **Optimized** Spark jobs resulting in 50% faster processing and 30% cost reduction

Merck Control Tower Project

- **Delivered** high-quality architecture designs as Lead AWS Solution Architect
- **Led** root cause analysis forums to identify and resolve technical issues
- **Implemented** AWS Well-Architected Framework, streamlining the architecture design process
- **Proposed** feasible solutions for Warehouse Capacity Optimization project based on collected requirements

Infosys Technologies Limited

Cloud Data Architect | Mar 2015 - Apr 2022

Data Driven Enterprise Platform

- **Designed** end-to-end solutions for data ingestion to Hadoop clusters using DDEP framework
- **Integrated** Apache Atlas with EMR for data governance and lineage tracking
- **Performed** data migration while maintaining integrity and compliance
- **Developed** ETL jobs and custom scripts for data transformation
- **Deployed** code using Azure DevOps Pipeline for consistent releases

AWS Data Lake Implementation

- **Architected** ETL workflows using AWS managed services (Lambda, Glue, Step Functions, SNS, SQS)

- **Designed and developed** efficient workflow orchestration for ETL processes
- **Created** a data quality framework using DDQ library for Spark, ensuring data integrity
- **Implemented** serverless data processing patterns reducing infrastructure costs

CloudBees Flow Setup on GCP

- **Administered** CloudBees Flow deployment on Google Cloud Platform
- **Automated** the installation process using Terraform
- **Deployed** Docker images to GKE cluster improving scalability and resilience
- **Explored and implemented** CI/CD features of CloudBees Flow

Disaster Recovery for MariaDB

- **Developed** Python scripts to capture data differences between main and backup MongoDB/MariaDB instances
- **Set up** development environments on AWS platform for team productivity
- **Implemented** database synchronization mechanisms ensuring data consistency

Marketing Data Lake (Data Reactor)

- **Collected** data from various sources and loaded it into Marketing data lake
- **Served** as product release manager for the data ingestion framework enhancement team
- **Designed** scalable data collection and transformation processes

HCL Technologies

Cloud and Big Data Architect | Jan 2014 - Mar 2015

Flight Science Datamart

- **Developed** prototypes supporting architecture design technology stack
- **Indexed** artifacts using Apache Solr and stored them in HDFS
- **Implemented** Apache Nutch for efficient crawling and data discovery
- **Enabled** "free text search" functionality for artifacts from Solr admin UI

Analytics Innovation Projects

- **Created** predictive analytics prototypes for toner usage patterns using R and Excel
- **Utilized** models including classical multiplication, artificial neural networks, and ARIMA for time series forecasting
- **Implemented** Hive on Hadoop for reporting and data generation

Cognizant Technology Solutions

Senior Hadoop Developer | Jan 2011 - Nov 2013

Infrastructure Monitoring with Nagios

- **Set up** Nagios in CentOS 6.3 for production operations monitoring
- **Configured** service monitoring, persistent MySQL database, and alert notification systems
- **Implemented** comprehensive infrastructure monitoring improving system reliability

Distributed Data Processing Benchmarking

- **Compared** distributed data processing using Hadoop versus HPCC solutions
- **Configured** AWS EC2 instances and deployed a 4-node Hadoop cluster
- **Created** performance benchmarking reports for technology selection decisions

Tata Consultancy Services

Developer | Jun 2008 - Dec 2010

PubSubHub Client APIs

- **Developed** reusable publish/subscribe APIs for content migration
- **Led** development effort and created coding guidelines and standards
- **Implemented** Jersey API for RESTful services

New Generation CRM Integration

- **Integrated** web-based systems using Oracle Fusion Middleware components
- **Created** composite services and adapter services for application integration
- **Set up** automated testing for services deployed in WebLogic app servers

EDUCATION

B.E. Computer Science

PSG College of Technology, Coimbatore, Tamil Nadu

SKILLS

- **Cloud Platforms:** AWS (EC2, S3, Lambda, Glue, Step Functions, Athena, EMR, Redshift, QuickSight), Azure (VM, Functions, Data Factory, Synapse, Blob Storage)
- **Data Processing:** Spark, PySpark, Hive, Hadoop, Kafka, Kinesis, Airflow
- **Databases:** Redshift, MySQL, PostgreSQL, DynamoDB, MongoDB, HBase
- **Programming:** Python, Java, Scala, Shell Scripting
- **Infrastructure as Code:** CloudFormation, Terraform, CDK
- **Containers & Orchestration:** Docker, Kubernetes, ECS, EKS
- **Analytics & ML:** Data modeling, ETL design, Data lakes, Data warehousing

CERTIFICATIONS

- AWS Certified Cloud Practitioner
- Microsoft Certified: Azure Fundamentals (AZ-900)
- IBM Enterprise Design Thinking Practitioner
- IBM Cloud Pak for Data V3.5.x Essentials

ACHIEVEMENTS

- Reduced ETL processing times by 50% and operational costs by 30% on cloud migration projects
- Won Bronze Medal representing Tamil Nadu in the National Pickleball Tournament 2024
- Received the Oscar of the Month award (June 2013) for delivering bug-free quality products
- Awarded Bright Spark Q1 2013 for proactive technology recommendations
- Won various corporate badminton and table tennis tournaments