GOKUL RAJA
SENIOR CLOUD SOLUTION ARCHITECT | CLOUD PLATFORM | BACKEND & DATA ENGINEERING

PHONE: +49 152 58450297 | **EMAIL:** gokul.227.r@gmail.com
LINKEDIN: linkedin.com/in/gokul-07 | **GITHUB:** github.com/gokul-227 | **PORTFOLIO:** gokul-portfolio-227.web.app
WORK AUTHORIZATION: EU Blue Card– No Sponsorship Required
LOCATION: Berlin, Germany | Open to relocate | **AVAILABILITY:** Immediate Joiner | Open to Full-time / Contract Roles
LANGUAGES: **ENGLISH** – Native/ Fluent, **GERMAN** – Level B1 (Actively Learning)

SUMMARY

Senior Cloud Solution Architect with 8+ years of experience designing and delivering enterprise cloud platforms, distributed systems, backend applications, and data solutions across Banking, Healthcare, Finance, and Retail domains. Strong expertise in solution architecture, system design, microservices, serverless, event-driven architecture, APIs, data platforms, and cloud migrations. Hands-on experience across backend engineering, data engineering, DevOps, Infrastructure as Code, CI/CD, cloud security, monitoring, and enterprise integrations. Experienced in translating business requirements into scalable technical solutions, leading technical delivery, conducting architecture reviews, and collaborating with engineering, security, product, and business stakeholders.

CORE SKILLS

AWS: Airflow, Amazon Q, API Gateway, Athena, Aurora, Cloud9, CloudFormation, CloudFront, CloudTrail, CloudWatch, Cognito, CodeBuild, CodeCommit, CodeDeploy, CodePipeline, DMS, Direct Connect, DocumentDB, DynamoDB, EBS, ECR, ECS, EFS, EMR, EventBridge, Fargate, Glue, IAM, Kinesis, Kafka, Lake Formation, Lambda, QuickSight, RDS, Redshift, Route53, SageMaker, S3, SES, SNS, SQS, Step Functions, Secrets Manager.
AZURE: Azure Data Factory, Synapse Analytics, Azure Functions, Azure Blob Storage, Azure SQL, API Management, Service Bus, Key Vault, Azure Monitor.
GCP: BigQuery, Cloud Storage, Pub/Sub, Cloud Functions, Cloud Run, API Gateway, Cloud Build, Cloud Logging, IAM.
PROGRAMMING & SCRIPTING: Python, PySpark, Scala, SQL, Bash, Shell Scripting, Linux Commands.
DATABASES & WAREHOUSES: MySQL, PostgreSQL, SQL Server, Oracle, MongoDB, Snowflake.
BACKEND DEVELOPMENT: REST API, SOAP API, FastAPI, Flask, Microservices, API Development & Integration, Serverless Applications, Authentication, Event-Driven Services, Postman.
DATA ENGINEERING: ETL / ELT, Data Lakes, Lakehouse, Apache Iceberg, Star Schema, Dimensional Modeling, Data Vault, Metadata Management, Schema Design, Data Quality, DBT, Databricks.
TESTING & VALIDATION: Unit, Integration, System, Regression, ETL Testing, Data Validation, Data Quality
DEVOPS: Terraform, Jenkins, Git, GitHub, GitLab, Bitbucket, Docker, Kubernetes, CI/CD, Infrastructure as Code, Cloud Monitoring, Logging, Release Automation, Jira, Confluence, Agile / Scrum.
OTHERS: Solution Architecture, System Design, Distributed Systems, Cloud Migration, High Availability, Disaster Recovery, Security, Governance, Scalability, Performance Optimization, Cost Optimization, Technical Leadership, Team Mentoring, Stakeholder Management, Client Management, Architecture Reviews, Code Reviews, UAT, Testing & Troubleshooting, Technical Documentation.

PROFESSIONAL EXPERIENCE

Senior Cloud Solution Architect – NatWest Group. **MAR 2024 – MAR 2026**
Project: Cloud Data Archival Platform
Description: Designed & built a next gen AWS cloud-native archival platform replacing Informatica ILM to automate enterprise data archival, retention, validation, and lifecycle governance across distributed databases.

- Architected an AWS-native archival platform replacing Informatica ILM, reducing annual licensing cost by ~35%.
- Built PySpark AWS Glue pipelines processing 2–3 TB of data per day from multiple on-prem RDBMS systems.
- Established Athena Iceberg tables enabling schema evolution and improving query performance by ~40%.
- Orchestrated archival, validation, and purge workflows using Airflow, reducing manual operations by ~60%.
- Delivered governance & security using Lake Formation, IAM, KMS across multi-account AWS environments.

Project: Customer Identity & Access Data Platform

Description: Modernized CIAM analytics platform by consolidating identity, authentication, and application log datasets into a unified AWS data lake supporting enterprise reporting.

- Consolidated 10M+ customer records & 500M+ authentication events into centralized S3 data lake architecture.
- Built incremental PySpark Glue pipelines maintaining 13 months of historical data in Athena.
- Designed Athena analytical schema powering 15+ enterprise dashboards used by analytics and marketing teams.
- Automated infrastructure provisioning using Terraform and Jenkins CI/CD across multiple environments.
- Implemented CloudWatch monitoring and SES alerts, reducing pipeline failures by ~30%.

Senior Cloud Data & Backend Engineer – Infosys. **OCT 2021 – MAR 2024**

Project: GT Connect Healthcare Data Platform

Description: Migrated legacy healthcare application AMT to scalable AWS serverless architecture enabling automated ingestion, validation, and API-driven access to healthcare transaction data.

- Delivered serverless data architecture supporting healthcare data workflows, improved scalability & reliability.
- Built PySpark AWS Glue pipelines ingesting, validating 1000s of GB per day into RDS databases.
- Introduced event-driven pipelines using EventBridge & Lambda triggers, reduced manual efforts by 70%
- Developed REST APIs enabling secure search and retrieval across millions of healthcare transaction records.
- Standardized deployments using CodePipeline & CloudFormation, ensuring consistent infrastructure provisioning.

Project: Data Ingestion & Analytics Platform

Description: Built centralized AWS ingestion platform integrating learning and training datasets from multiple enterprise systems into unified analytics environment.

- Established multi-layer data lake architecture implementing raw, refined, and conformed data layers.
- Built distributed PySpark ETL pipelines ingesting 1000s of files daily from Excel, CSV, and database sources.
- Automated ingestion triggers using Lambda and EventBridge, enabling near real-time data availability.
- Delivered API services exposing curated datasets used by analytics and reporting teams.
- Implemented CI/CD pipelines using CodePipeline and CloudFormation, reducing deployment effort by 50%.

Cloud Data & Backend Engineer – IBM. **APR 2019 – OCT 2021**

Project: Enterprise Database Archival Solution

Description: Developed AWS-based archival framework enabling migration of historical enterprise datasets into cost-efficient cloud storage.

- Designed archival architecture migrating multi-year enterprise datasets from on-prem RDBMS into S3 storage.
- Built distributed PySpark AWS Glue pipelines processing 100s of GB per archival cycle.
- Implemented automated purge framework using Fargate &Lambda, reducing database storage by 25%.
- Developed serverless APIs allowing operational teams to trigger archival workflows.
- Implemented centralized monitoring using CloudWatch dashboards and DynamoDB logs.

Project: Healthcare History Data Migration

Description: Migrated large historical healthcare claims datasets from legacy mainframe systems into AWS environment for analytical reporting & regulatory compliance using ETL automation

- Processed and migrated millions of healthcare history data from mainframe systems into AWS S3 storage.
- Built ETL pipelines using Syncsort ETL and processed millions of records per batch.
- Created Athena analytical tables enabling SQL-based reporting across historical datasets.
- Implemented reconciliation workflows ensuring 99.9% data accuracy during migration.
- Supported UAT validation, documentation, and stakeholder walkthroughs during migration lifecycle.

Software Engineer Intern– IPrimed Education Solutions **APR 2018 – APR 2019**

- Assisted in backend development, automation scripting, and application testing for enterprise projects.

EDUCATION

- Bachelor of Technology (B. Tech) – Information Technology **-** Anna University, India **JUL 2015 – APR 2019**