MOHAK CHUGH

Software Engineer at Amazon

+91 9810178257 · me.mohakchugh@gmail.com · https://www.linkedin.com/in/mohakchugh · Delhi NCR

Work Experience

Amazon
Gurugram, Haryana, India · Hybrid

Software Development Engineer 2 – [Level 5 | Current]
10/2024 – Present

- Developed APIs to serve Amazon Financial data for Transactional and Non-Transactional collections use-cases with **100% reconciliation** coverage.
- Developed a DFS-based sorting algorithm and self-healing reconciliation for employee data backfills (processing **5MM+ employees**), ensuring grouped hierarchy updates and **zero data loss within 10min SLA**, which **eliminated loss of access** issues during any Amazon re-orgs.
- Worked on **RDS performance optimizations** bringing down query performance times from **20mins to under 30s** (used vertical partitioning of tables, creating new indexes and deprecating irrelevant indexes), **impacting 200+ downstream ETL pipelines** and increasing data freshness from **4 hours to 30 mins** SLA. Led development of internal UI tool for People Experience org (REST/RPC APIs, ABCDA pattern, SEL/Xray tracing, IAM auth; Cypress UI **integ tests to 80% coverage**, telemetry alarms, security review), accelerating team velocity from 4 points/25 days to deliver critical delayed **project 1 week ahead of deadline.**
- Led design and API contracts for new Candidate Search service (post-legacy system deprecation for hiring dashboard migration) and re-architecture of high-volume Search Populations API (response time from **5s to under 200ms** P99), via RCA/POCs/team reviews/customer signoffs, enabling Q1 2026 launches.
- Served as Operational Excellence (OE) Champion, analyzing 6 months of ticket data (75 SEV2s/473 SEV3s) to prioritize 20+ backlog items (alarm tweaks, retries, performance tuning), **reducing SEV2s from 21 to 3, latency breaches from 26 to 1, and SAS risks from 104 to 6** while leading security reviews, dashboards, and retrospectives.
- Worked on employee data to power employee segmentation for payroll use-case, ingesting more than **10TB data per day**, vending out at more than **1000 TPS** via APIs

Used: Java, AWS EMR (spark), AWS Big Data, APIGateway, AWS Lambda, DynamoDB, Scala, Hadoop, Glue, Athena, Spark, Advanced SQL, Python, AWS Step Functions, AWS DMS, AWS Lambda

Software Development Engineer 1 – [Level 4]
12/2022 - 09/2024

- Built a **Spark**-based **big data reconciliation engine** capable of processing over **1 billion records daily** across **SQL and NoSQL sources**, ensuring **100% coverage** of internal financial data flowing through the central data lake. Enabled seamless integration with data pipelines via custom hooks.
- Led the creation of an award-winning **RAG-based LLM QnA bot, engaging 300+ users** and processing 5K questions daily; leveraged advanced prompting techniques and guardrails, resulting in **93% accuracy** from rich text documents. This was published in AWS Blog posts as one of the first AI blogs in Amazon
- Designed and deployed **Airflow-based email notification system** for Workforce Capacity Planning dashboards using Anaplan to S3 exports with Hive partitioning, SES throttling/randomization, and reconciliation (notifying 300+ analysts), ensuring timely validations within 5 business days without exceeding AWS limits.
- Led **operational readiness reviews** (ORR) for **reconciliation tool** (**100B records/day, SOX/GRC complian**t), and ingestion service as sole SDE, automating job scheduling, unifying CDK pipelines (infra as code), implementing **DAG** based alarms and retries in **Airflow**, optimizing, and building customer dashboards/runbooks to complete GRC compliance and eliminate operational gaps.

Used: Java, LLM ClaudeV2.1, AWS cloud, Streamlit, DynamoDB, PostgreSQL (RDS), ECS Fargate, Apache Airflow, Agentic AI, Python, Lambda, Vector Knowledge Bases, OpenSearch (Elastic Search), Embedding Models, LLM Fine Tuning

Data Engineer 1 – [Level 4]
09/2021 - 12/2022

- Engineered AI-driven headcount management tools by leveraging **100+ work driver datasets**, enabling precise ML models for scalable headcount predictions; boosted operational efficiency by forecasting resource allocation with **95% accuracy**. Integrated functionalities like email notification systems, auth management and ML data pipeline orchestration using **Airflow**.
- Launched e-invoicing compliance for Amazon by enabling VAT IDs encryption for Brexit, AU, CA regions, impacting **3,000+ global vendors.**
- Collaborated with senior data engineers to re-architect the **financial analytical mesh data lake**, consolidating databases and automating schema-level data sharing across Redshift clusters. Designed and delivered a secure solution meeting **1-second SLAs, achieving 90% test coverage** and improving organizational productivity through efficient, security-reviewed data storage.
- Mentored 3 interns and got 2 converted to full time engineers.

Used: Airflow, Redshift, AWS EMR (spark), Scala, Hadoop, Glue, Athena, AWS Serverless, CI-CD pipelines, AWS Big Data, AWS S3, AWS Glue and Lake Formation. Amazon Quicksight, Apache Hive, Apache Airflow, Apache Presto

Education

Maharaja Surajmal Institute of Technology
- Bachelor of Technology - Computer Science | 8.7 | University Rank 3 in Final year
08/2017 - 12/2021

Publications

Benchmarking geospatial database on Kubernetes cluster	— Analysis and Optimizations of PostgreSQL/PostGIS in Kubernetes for improved geospatial query performance.
Generative AI Q&A chat assistant using Amazon Bedrock	— Built a custom RAG based QnA system for a niche legal use case which gave accuracy of 96% as compared to 49% from of the shelf tools

Languages and Skills

Work Domains:	AWS Cloud (**Development** and **Deployment** for **Serverless** and Server based solutions), Applied Artificial Intelligence - **LLM RAG** Architecture and Natural Language Processing (**NLP**), **Big Data** (Analytical and Transactional) and Advanced System Design on AWS
Languages:	**Python**, **Java**, NodeJS, Angular, **Advanced SQL**
Tools and Technologies:	API Development, Distributed Systems (**Spark**) and Data pipeline management via **Airflow**, Data Structures and Algorithms

Achievements

1st Runner's up in Digital India Makeathon
- Certificate Link: https://drive.google.com/file/d/1YEMPcosx8COkOxqFn3vXHiugsIIhkZMX/view

Soft Skills Certification from AICTE
- Certificate Link: https://drive.google.com/file/d/1aBjeD3gCGHphhdfNAyZrXADcOZyBjLQ-/view

University Rank 3 in BTech Final year
- Certificate Link: https://www.linkedin.com/feed/update/urn:li:activity:7055093560970256384/