

# RAHUL JOSHI

+91 74548 47198 | rahuljoshi0925@gmail.com  
linkedin.com/in/rahul-j0shi | github.com/rahul-j0shi

## SUMMARY

Software engineer with **2+ years** of experience, focused on AWS data engineering and backend systems. Architected medallion pipelines for an **estimated 12M daily events and updates**, with dimensional models for cross-domain analytics. Builds reusable workflow automation that reduces custom integration effort.

## WORK EXPERIENCE

### Software Engineer - Enveu | Aug 2024 - Present

- Architected AWS medallion pipelines for **24+ projects** at an **estimated 12M events and updates/day (~24 GB raw JSON/day)**, unifying experience, content, and subscription data in S3 bronze/silver/gold layers.
- Orchestrated Brightcove/Mux/JWX sync and gold aggregation with Step Functions/Lambda checks; modeled star-schema fact and dimension tables for Athena OLAP queries and embedded tenant-filtered Superset dashboards in the CMS.
- Owned and improved Enveu Flow's execution engine (Kafka, Redis, Keycloak, MinIO); built **40+ plugins** with reusable nodes and configurable templates, cutting repeat integration setup from **days to hours**.
- Automated **real-time transcription, translation, and subtitle generation** in Enveu Flow for a customer processing **hundreds of videos/week**, removing manual handoffs across 3 processing stages.
- Built CMS Ops Buddy with Amazon Bedrock, available to **20+ tenants**, automating content lookup and guiding placement with page context, content/screen/widget selection, publishable previews, and human approvals.
- Integrated **2 video platforms** (Wowza and MediaKind) into Experience Cloud; contributed to Wowza Streaming Engine and live-to-VOD workflows in Enveu Flow to automate clip creation from live streams.
- Develop Axis360's subscription management backend and support its first customer onboarding and rollout.

### Data Engineering Intern - HT Media Labs | Apr 2024 - Jul 2024

- Processed XML feeds from 3 OTT partners (Dangal, Chaupal, ZEE) with Python extraction and crawling; reconciled feed discrepancies and corrected OTTplay catalogue records to improve data quality.
- Monitored Apache Airflow jobs and diagnosed ingestion/crawling failures via PuTTY to support catalogue updates.

## TECHNICAL SKILLS

**Languages:** Python, SQL, Java, TypeScript

**Data Engineering:** ETL/ELT, medallion architecture, dimensional modeling, star schemas, data quality

**Analytics and Orchestration:** Online analytical processing (OLAP), Apache Superset, Apache Airflow

**AWS:** Amazon S3, Lambda, Step Functions, Data Firehose, Glue Data Catalog, Athena

**Databases and Storage:** PostgreSQL, MongoDB, Redis, MinIO

**Backend and Messaging:** FastAPI, WebSockets, Kafka, RabbitMQ, Keycloak

**Graph and Vector Databases:** Neo4j, Pinecone, Milvus, Qdrant

**AI and Speech:** Amazon Bedrock, Gemini, Azure AI Search, ONNX Runtime, Kokoro

## PROJECTS

### Shale - LSM Storage Engine (GitHub)

- Built an embeddable Java LSM key-value engine with CRC32C-checked write-ahead log replay for crash recovery and prefix-compressed SSTables for compact storage, with no external runtime dependencies.
- Implemented lock-free reads in a single-writer/multi-reader skiplist using release/acquire publication; reconciled memtable and SSTable scans by version and tombstone to return current values and exclude deleted keys.
- Validated recovery and read correctness with **5,000-operation** differential histories against TreeMap, including restart, WAL truncation, SSTable corruption, concurrency, and golden-format tests.

### Svara - Voice and 3D Avatar System (GitHub)

- Built a FastAPI/WebSocket voice prototype with 3 swappable stages (speech-to-text, LLM, text-to-speech); rebuilding it as an Electron companion for Ubuntu.
- Derived lip-sync timelines from Kokoro ONNX phoneme durations and mapped **15 viseme channels** to 5 VRM mouth expressions, synchronizing avatar animation with the audio clock.
- Tested 4 playback cases (cancellation, late decoding, audio failures, repeated cleanup) to prevent stale audio.

## EDUCATION

**Master of Computer Applications (MCA)** | Birla Institute of Applied Sciences, Bhimtal | 2022 - 2024

**Bachelor of Computer Applications (BCA)** | Amrapali Institute of Applied Sciences, Haldwani | 2019 - 2022