

Ankit Rai

📞 9082809759 ✉️ ankitrai0396@gmail.com [🌐 linkedin.com/in/ankitrai047](https://www.linkedin.com/in/ankitrai047)

Education

Mukesh Patel School of Technology Management & Engineering, NMIMS University

Master of Computer Application

Aug. 2022 – April 2022

Mumbai, Maharashtra

Thakur college of Science and Commerce, Mumbai University

Bachelor of Science

July 2016 – May 2019

Mumbai, Maharashtra

Functional Areas

- Data Ingestion
- Data Processing
- Storage
- Reporting

Experience

Acrotrend Solutions

Software Engineer Trainee

April 2022 – Dec 2024

Navi Mumbai, Maharashtra

- Designed and deployed ETL solutions using AWS Glue to transfer data from on-premises databases to an Amazon S3 data lake.
- Contributed to multiple proof-of-concept on AWS during the onboarding process.
- Developed and implemented PySpark scripts for incremental data imports and quality assessments, automating data validation processes within the S3 Data Lake using AWS Glue Jobs.
- Cataloged transformed data with crawlers and automated ETL processes by scheduling jobs and crawlers through workflow features.
- Monitored ETL processes through CloudWatch logging and alerts, and automated cloud infrastructure deployment using AWS CloudFormation.
- Collaborated with cross-functional team to refine AWS-recommended data quality rules to include business-specific logical checks, identifying and addressing data gaps effectively.
- Deployed data quality frameworks across Dev, UAT, and Prod Environments.
- Developed Power BI dashboards to visualize data quality trends and insights, enabling stakeholders to easily understand data quality metrics and make informed decisions.

Projects

Spotify ETL Pipeline / Python, AWS Lambda, S3, Snowflake, Power BI

January 2024

- Automated the ETL process to extract, transform, and load data from the Spotify API, improving analysis of user listening data.
- Developed a data pipeline using AWS Lambda, S3, and Snowflake to efficiently extract and process Spotify API data.
- Transformed and enriched the data using Python before loading it into Snowflake for optimized querying and data analysis.
- Generated insights by designing a Power BI dashboard to visualize key music trends and user preferences.

RX Data Lake / Python, AWS Glue, S3, Athena, CloudWatch, IAM, Power

April 2022 – Dec. 2023

Client: Reed Exhibition

- Implemented a data quality framework in AWS Data Lake, improving data accuracy by 85% and automating assessments across all the data stored in data lake.
- Optimized ETL job execution efficiency by implementing parallelism and data partitioning strategies, achieved reduction in processing time.
- Integrated data from on-premises sources into the data lake and streamlined data access via Amazon Athena for ad-hoc queries.

Technical Skills

Programming Languages: Python (Pandas, NumPy, Pyspark), SQL

Developer and Visualization Tools: VS Code, Jupyter Notebook, Power BI

Cloud Platforms: AWS (S3, DMS, Glue, Athena Lambda, CloudWatch, Step function, IAM), Snowflake

Databases: SQL Server, DynamoDB, Snowflake

Data Modelling: Dimensional Modeling, ER Modelling, Star Schema, Snowflake Schema

Certifications / Extracurricular

- Completed AWS Certified Data Engineer Associate 2024 Hands On (Udemy Certificate)
- Completed Mastering Data Modelling Fundamentals (Udemy Certificate)
- Conducted Knowledge Transfer Sessions Within a Team of 8 Members.