

AMEY MEHER

+1 (919) 897-0346 | avmeher@ncsu.edu | LinkedIn | GitHub

EDUCATION

North Carolina State University - Raleigh

August 2022 - May 2024

Master's, Computer Science

GPA: 4

University of Mumbai

August 2015 - May 2019

Bachelor's, Computer Engineering

PROFESSIONAL EXPERIENCE

Barclays

Data Engineer

July 2019 - July 2022

- Led development for 4 ETL projects on managing data pipelines; being responsible for design, code, unit & integration tests, migration to Prod with CI/CD, compliance checks, project documentations and handover to support team.
- Achieved 99% optimization in Hadoop and Spark jobs used for extracting risk metrics from datasets having millions of records; solving a scaling bottleneck in a critical service which was used to calculate risk scores of customers.
- Integrated data from 50+ tables for creating JSON of over 12000 complex elements. Developed automated jobs to call APIs with these JSONs as payload for parallel testing of a migration project, which reconciled data between 16 different tables
- Collaborated with data engineers, product managers and data scientists, representing key insights with data visualization (Power BI) of production data in an intuitive way, which helped us secure stakeholder's signoff for a major project.

IBM

Software Engineer Intern

May 2023 - August 2023

- Implemented a highly scalable RESTful web service using Flask in a microservices architecture for synchronization of metadata in MongoDB backend using Kafka and a stream processor reducing manual efforts of 40+ hours per week.
- Designed optimized InfluxDB queries on 3 years of minute-level data to derive metrics for efficient job scheduling.
- Developed an unsupervised clustering model to analyze incidents and automate resolution steps, thereby reducing manual efforts.

SKILLS

Programming Languages: Python, Java, SQL, Scala, Rust, C++

Data Engineering: Spark, Hudi, Hadoop, Hive, Kafka, Airflow, ETL/ELT, Oracle, Power BI, MongoDB, Superset, Postgres

Web Services: Microservice Architecture, REST APIs, Flask, Spring, Postman, Docker, Kubernetes, Git

AWS: S3, EC2, Lambda, DynamoDB, Athena, Redshift, API Gateway, IAM

Azure: Data Factory, Databricks, Data lake storage Gen 2, Synapse Analytics, HDInsight

Courses: DBMS, Automated Learning & Data Analysis, Machine learning on graphs, Data Structures & Algorithms, OOP Methodology

PROJECTS

• Apache Hudi Lakehouse ETL

Created ETL processes for source agnostic data ingestion, CDC workload, and async clustering job for removing small file issue

• Stock Price Prediction using Graph Neural Networks

Developed a stock price prediction system using GCN and LSTM models. Implemented using PyTorch and TensorFlow frameworks

• Android application based on Proximity reminder : Oncue

Awarded first place for amassing 500+ downloads on Playstore for a Hackathon in collaboration with Facebook developers' circle

PUBLICATIONS

- "Error mitigation of Hamiltonian Simulation from an Analog based compiler (SimuQ)" paper accepted at the 2024 QCE IEEE conference
- "Application for Counterfeit Detection in Supply Chain using Blockchain Technology" paper published in 2021 ICAC3 conference
- "Canteen Automation System" paper published in 2021 IJRASET journal