





ARUNIT BAIDYA

430 Centre Street, Boston, MA, 02130

 [Github](#)  [LinkedIn](#)  +16179355115  baidyaarunit@gmail.com

Education

Northeastern University, Khoury College of Computer Sciences

Aug 2024 – April 2026

Master of Science, Computer Science

Boston, MA

Anna University, PSG College Of Technology

Aug 2018 – July 2022

Bachelor of Engineering, Computer Science and Engineering

Coimbatore, India

Skills

- **Technical Skills:** Java, Python, C, Go, API Development, object-oriented programming, object-oriented design, PostgresDB, RestAPIs, SQL, Spark, relational databases, MVC, design patterns, Operating systems, Git, Backend, troubleshooting, AWS, Docker
- **Software Development:** CI/CD, Agile Development, Test Driven Development, Code Reviews,
- **Collaboration:** Excellent communication skills, cross-functional collaboration, domain knowledge interest

Work Experience

Arcesium

Software Engineer

Jul 2022 - Jul 2024, Bangalore

- **Met client SLOs** (max 10 min time to output for datasets with 1,000–100,000 records) by optimizing SQL operations, debugging **Python & Java** code using **Spark**, *and improving process times by 40%*.
- **Designed partitioning strategy** for constant read/write performance in Delta-Tables with historical data.
- Developed a **config-driven data-filtering** module in Java and Spark using partition pruning for **static/dynamic filters**, *reducing compute resource usage and improving workflow efficiency by 30%*.
- **Optimized resource provisioning buckets** using ArgoWorkflow, Kubernetes & AWS, adopted in 10+ workflows.
- Implemented **SQL** for SCD type 2 merge, enabling bi-temporality & historical querying, state of data at point in time.
- **Created LLD documentation** with business logic translation, *reducing* cross-functional team *inquiries by 50%*.

Software Engineering Intern

Feb 2022 - May 2022, Remote

- **Increased ETL throughput by 70%** via preprocessing stage to split heterogeneous data for parallel processing.
- **Built data ingestion module** supporting multiple input types with schema validation for data consistency.
- **Boosted ETL reliability** by engineering module in python to extract malformed data to global kickouts delta-table.
- **Enhanced ETL observability** with efficient run history logging and storing execution stats in Delta-Tables.

Software Engineering Intern

May 2021 - Jul 2021, Remote

- **Automated PDF Parsing in python** to identify and parse tabular information using OpenCV and Camelot.
- Implemented **CRON processing** to parse PDFs mapped against filters & annotations, *saving 4 hours work/week*.

Projects

Image Processing Application (Java) ([Course Project](#))

October 2024 – November 2024, Boston

- Image processing application in **Java**, accepting inputs from CLI, txt script file, or GUI to process transformations on images. Project **intended to demonstrate programming design principles** learnt in graduate course work.
- Following **SOLID principles** using passive MVC model, functional interfaces, higher order functions, factory & command design patterns, and test-driven development.

Qthreads – Cooperative user space thread library (C) ([Course Project](#))

November 2024, Boston

- A **user-space thread library** in C, like the POSIX threads library, implementing functions to create threads, yield threads, put threads to sleep, provide mutexes and condition variables for managing concurrency in a **cooperative threading** model.

Simple Linux Shell (C) ([Course Project](#))

October 2024, Boston

- A Linux shell in C with command execution for internal & external commands, piping, & I/O redirection using **fork**.