

ROHIT PAUL GERARD NAGARAJAN

Boston, MA | (857) 313 – 0203 | gerardnagarajan.r@northeastern.edu | [GitHub](#) | [LinkedIn](#)

Analytically-minded individual seeking a Data Engineer role with expertise in ETL pipelining, data management, and data integration to build scalable and efficient data architectures with a focus to drive data-driven decision making and optimize business performance.

EDUCATION

Northeastern University

Master of Science in Information Systems | GPA: 3.5/4.0

Relevant Coursework: Data Management & Database Design, Designing Advanced Data Architectures for Business Intelligence

Boston, MA

May 2024

Sathyabama Institute of Science and Technology

Bachelor of Engineering in Electronics and Communication Engineering

Chennai, India

July 2015 – July 2019

TECHNICAL SKILLS

Languages:	Python, SQL, Linux Shell Scripting, PySpark, Pandas and NumPy
Big data:	Apache Hadoop, Apache Spark and HDFS
Databases:	MySQL, PostgreSQL, Microsoft SQL Server, Azure SQL, Oracle SQL and Hive
ETL:	Alteryx Designer, Talend Studio
Data Modeling:	ER/Studio Data Architect
Cloud Services:	AWS S3, AWS Athena, AWS Glue, Microsoft Azure, Google Cloud Platform (Dataproc)
Version control:	Git, GitHub
Business Intelligence:	Tableau, Power BI
Methodologies:	Software Development lifecycle, Normalization techniques, Agile

INDEPENDENT PROJECT

Boston Crime Analysis | SQL, PySpark, Hadoop, Spark, Hive, GCP Dataproc, AWS: S3 Bucket, Athena, Glue, Power BI **Feb 2023**

- Setup a **Dataproc** cluster on **Google Cloud**, extracted Boston Crimes data from **Hadoop** and stored it on **HDFS** in a distributed manner, resulting in a **50% decrease** in data processing time
- Extracted, cleansed, transformed, loaded **300,000 +** records using **PySpark** in a PySpark dataframe and stored the resulting data in a CSV file, reducing data redundancy by **15%** and improving data accuracy by **90%**
- Leveraged **Hive** by creating a database, conducted in-depth **data analysis** through querying, sorting and aggregating the data, revealing key insights that can aid in criminal investigation and prevention
- Automated the entire data processing pipeline using a **Spark-submit** job, saving approximately **2 hours** of manual work and improving overall **efficiency**
- Designed and implemented an end-to-end **data pipeline** in AWS using **S3** for creating **data lakes**, **Athena** for data analysis, and **Glue ETL** jobs for transforming and loading structured data, thereby achieving a reduction in data processing time by **40%**
- Created an interactive **dashboard** and **visualizations** using **Power BI** that helps achieve crime visibility in hot and cold areas and identified safe spots from criminal activities for civilian safety

ACADEMIC PROJECT

Disease Diagnosis and Medic Recommendation System | SQL, Python, Pandas, NumPy, Seaborn

Sep 2022 – Dec 2022

- Led a team of 4 and developed a **MySQL** database that recommends medical professionals for patients based on location, cost, gender, and education
- Gathered structured data from Health websites and Twitter using web-scraping techniques, analyzed and transformed it with Python's **Pandas** and **NumPy** libraries **achieving 98%** data accuracy and data completeness
- Loaded normalized database tables with clean data to eliminate data redundancy, created SQL views **resulting in 100%** database organization and visualized data using **Seaborn** library in Python that boosted our project to receive a **195/200** score

PROFESSIONAL EXPERIENCE

Oracle Financial Services Software Ltd.

Automation Analyst

Chennai, India

Nov 2021 – July 2022

- Successfully delivered various core-banking projects for clients from diverse regions, showcasing expertise in Automation testing, Defect Analysis, Requirements Analysis, and presented RAG status reports to clients with clarity and impact
- Enhanced team performance by **70%** by proposing workflow process changes in Talent connect meetings hosted by senior management, conducted regular **data analysis** on team performance and presented statistics to team lead for further baselining
- Streamlined payments processing, loans and lease management projects for testing team by **developing API Automation** scripts using that resulted in a reduction of manual efforts by a remarkable **80%**

Tata Consultancy Services Ltd.

Data Analyst

Chennai, India

Sep 2019 – Nov 2021

- Delivered core-banking projects with responsibilities such as Requirements analysis, Defect analysis and reporting, Automation testing, providing month-end statistical **data analysis** reports using **Power BI** on key performance metrics of my team
- Reduced unit testing errors by **60%** by conducting regular data quality checks in **Oracle SQL**, monitoring inbound and outbound API calls in Elastic Search **Kibana**, and reporting data inconsistencies to development team
- Led a team of 4 junior associates and helmed on tasks that involved testing and programming **UNIX** scripts for migration of historical data from **talend** server to **Amazon S3 bucket** through Secure File Transfer Protocol and **AutoSys** job scheduling that ultimately resulted in a **90%** storage efficiency
- Streamlined project operations on the Funds Transfer module by automating redundant work in **Selenium**, resulting in a **50% reduction** in manual effort and saving **2 hours** per day
- Successfully prevented post-production mishaps by reporting and triaging data quality issues from partner applications, resulting in a **40% reduction** in incidents reported in JIRA and ServiceNow