ROHIT PAUL GERARD NAGARAJAN

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

SUMMARY

Experienced data professional skilled in data engineering, data analysis, and business intelligence. Strong technical skills in Python, SQL, ETL, and BI tools such as Tableau and Power BI. Proven track record of delivering successful projects and upscaling team performance

EDUCATION

Northeastern University **Boston, MA**
Master of Science in Information Systems | GPA: **3.5/4.0** **May 2024**
Relevant Coursework: Data Management & Database Design, Designing Advanced Data Architectures for Business Intelligence

Sathyabama Institute of Science and Technology **Chennai, India**
Bachelor of Engineering in Electronics and Communication Engineering July 2015 – July 2019

TECHNICAL SKILLS

Languages/Databases:	Python, SQL, MySQL, PostgreSQL, MS SQL Server, Oracle SQL, NoSQL, MongoDB, Snowflake, Hive
Libraries & Version Control:	Pandas, NumPy, PySpark, Git, GitHub
Tools/Platforms:	Talend, Alteryx, ER/Studio, Tableau, Power BI, Hadoop, Spark, HDFS, JIRA
Cloud Services/Frameworks:	AWS S3, AWS Athena, AWS Glue, Microsoft Azure, GCP, Software Development Lifecyle, Agile

PROFESSIONAL EXPERIENCE

Oracle Financial Services Software Ltd. **Chennai, India**
IT Analyst **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
- Improved team performance by **70%** through workflow process changes proposed in Talent connect meetings and regulated **Data quality** checks before every project sprint
- Streamlined payments processing, loans and lease management projects for testing team by developing **REST API Automation** scripts using that resulted in a reduction of manual efforts by a remarkable **80%**

Tata Consultancy Services Ltd. **Chennai, India**
Data Analyst **Sep 2019 – Nov 2021**

- Analyzed requirements, conducted root-cause analysis, performed testing, and generated monthly statistical **data analysis** reports using **Power BI** for core-banking projects
- 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, programmed **shell scripts** for migration of historical data from **Talend** server to **Amazon S3 bucket** resulting in a **90% storage** efficiency
- Streamlined project operations on the Funds Transfer module by **automating** redundant work in **Selenium**, resulting in a **50% reductio**n 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**

INDEPENDENT PROJECTS

Data Pipeline for Social Media Analytics *NumPy, Pandas, TextBlob, MongoDB, SQL, Snowflake, Power BI, Dax, Airflow* **Mar 2023**

- Built a twitter scraping **bot** in **Python** that scrapes **5000 unstructured** tweet data, cleansed the data using **NumPy**, **Pandas**, ensuring high data quality and accuracy
- Utilized **MongoDB** as a **NoSQL** database to store clean data, performed analytical queries through **Mongo Shell**, and conducted **sentiment analysis** using **TextBlob** Machine learning library, revealing valuable insights into user sentiment
- Loaded the processed data into a **Snowflake** Data warehouse, transformed the data using **SQL functions**, and discovered meaningful insights
- Designed interactive dashboards in **Power BI**, employed **DAX** formulas for sentiment remark calculations, providing even deeper **insights** into user behavior and automated the entire **data pipeline** using **Apache Airflow** resulting in an efficient and reliable data processing

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

- Employed **PySpark**, **Hadoop**, **Spark**, and **GCP Dataproc** to extract and transform over **300k+** Boston Crimes records, reducing data redundancy by **15%** and improving data accuracy by **90%**, automated the entire processing pipeline through **Spark-submit**
- Created and loaded a **Hive data warehouse** to conduct in-depth **data analysis** through querying, sorting, and aggregating the data, revealing key **insights** for criminal investigation and prevention
- Designed and implemented an end-to-end **data pipeline** in **AWS** using **S3 Bucket**, **Athena**, **Glue**, achieving a reduction in data processing time by **40%** and created an interactive **dashboard** for crime visibility using **Power BI**

Disease Diagnosis and Medic Recommendation System *SQL, Python, Pandas, NumPy, Seaborn* **Sep 2022 – Dec 2022**

- Developed a **MySQL** database that recommends medical professionals for patients by leading a team of 4, utilizing **SQL**, **Python**, Pandas, NumPy, and Seaborn
- Analyzed and transformed **structured data** gathered from Health websites and Twitter using **web-scraping** techniques, achieving **98% data accuracy** and **completeness** with **Pandas** and **NumPy** libraries
- Loaded **normalized** database tables with **clean data**, eliminated **data redundancy**, created **SQL views** resulting in **100% organization**, and visualized data with **Seaborn** library, further resulting in a score of **195/200** for our project