Johanan Abhishek Prabhurai
apjohanan@gmail.com | +1-(782)-234-5476 | https://www.linkedin.com/in/pjabhishek

SKILLS

Python, C#, Java, SQL, MySQL, PostgreSQL, MongoDB, Git, Jira, Jenkins, Airflow, Object Oriented Programming, Design Patterns, SOLID Principles, Agile, Test-Driven Development. AWS: EC2, S3, RDS, Athena, Redshift, Glue, Lambda, DynamoDB, SNS, SQS, CloudWatch.

EXPERIENCE

Software Engineer, Magnet Forensics **June 2022 – June 2023**

- Expanded and updated features in Magnet AXIOM Cypher, a digital forensics solution, within 100k+ lines C# codebase, used by ~4000 government agencies across 93 countries.
- Improved data processing speeds by 90% for report generation and analysis, by integrating with an internal tool that generates more than 200 JSON files, by collaborating with a cross functional team of 7 and communicating with other teams.
- Wrote unit tests and functional tests using XUnit for features developed using WPF, MVVM, LINQ in an Agile environment using continuous integration.
- Supported customers, which involved troubleshooting, debugging on production environments, and building beta versions using Azure CI/CD pipeline to address certain use cases.
- Coordinated with the product owner, product manager, testers, and the team to understand user stories, scenarios and requirements. Actively contributed to design planning, retrospectives, code reviews in an AGILE environment.

Software Engineer Coop, Magnet Forensics **January 2022 – April 2022**

- Collaborated with a cross-functional team of 8 to develop frontend and backend logic for a feature to process files, for report generation and analysis by integrating with a third-party library, contributing to $4.0 million revenue increase.
- Created modular and reusable C# code for front-end and back-end logic using WPF, MVVM, and a test-driven approach (TDD), accompanied by documentation that adhered to the company's standards, to tackle small tasks, bug fixes showcasing the ability to write production ready quality code.

Data Engineer, Capgemini **July 2019 – November 2020**

- Developed an ETL pipeline using big data solutions - Hadoop, Hive, Spark, Airflow to migrate a project from Oracle database to Hadoop by analysing and translating PLSQL queries to PySpark code, to process ~400GB of data. This doubled process speeds and increased accuracy of reports generated by 10% for data visualization.
- Reduced processing time by 33% (6 hours to 4 hours) of the ETL pipeline through the identification and execution of parallel data processing optimizations within the Airflow workflow using Python in a Unix environment.
- Delivered an ETL data pipeline, using Python, AWS cloud platform that consists of S3 for storing raw and processed files, Glue for processing and Redshift for advanced analytics, resulting in 98% accuracy of reports.
- Coded a Python framework using Boto3 for automating AWS Athena's and AWS Redshift's databases, tables, structures, and automated tests, resulting in a 30% decrease in hands-on tasks.
- Participated in a team of 11 to understand requirements, design planning, data modelling. Mentored juniors through KT sessions.

Junior Data Engineer, Capgemini **June 2018 – June 2019**

- Accelerated test execution and management efficiency by 90% in Zephyr Enterprise by creating a Python library for Zephyr Enterprise APIs, using it to develop a desktop application for automation.
- Engineered an ETL pipeline in AWS using Python, PySpark, S3, Glue, and Athena that processed 50+ CSV files, with ~100GB data in ~15 minutes with an 87% accuracy rate for data analytics.
- Built bash/shell scripts in Unix environment for data ingestion and scheduling jobs using Autosys to process data. Created multiple SQL queries for testing and validation of raw records with processed records.
- Worked in a team of 2 for planning the development, testing and release stages. Presented end-end workflow explanations to clients.
- Received training in Informatica (ETL tool), BI Tools, and Python3, successfully completing the program, by placing in top 10% and transitioning into a full-time role.

PROJECTS

Distributed Database GitHub link

- A distributed database and management system created using Google Cloud Platform and Java's Data Structures. Architected the core communication framework (unique among 95 other peers) for clients across diverse locations, implemented logic for SQL queries, along with a concurrency control locking technique.

EDUCATION

Master of Applied Computer Science **January 2021 – May 2022**
Dalhousie University GPA - 4.2 / 4.3

Bachelor of Technology in Computer Science and Engineering **September 2014 – April 2018**
CVR College of Engineering