SANJEEV RAMASAMY SEENIVASAGAMANI

315-278-0599 | sanjeevsramasamy@gmail.com | www.linkedin.com/in/sanjeev-rs/ | rssanjeev.github.io

TECHNICAL SKILLS

Languages: SQL, Python, PySpark, R **| Databases & Tools:** SQL Server, Snowflake, AWS (EC2, S3, EMR, RDS, Redshift), Hive, PowerBI, Tableau **| Data Modeling & Orchestration:** OLTP, OLAP, dbt, Fact & Dimensional modelling, Prefect **| CI/CD & Collaboration:** Git, CircleCI, JIRA, Confluence, Agile

PROFESSIONAL EXPERIENCE

KAGR
Data Engineer 3, Product Data Team   **Jul 2023 – Present**
Pipeline Work:
- Designing a nonprofit organization's ticketing data model spanning across 1,100 schools & 500,000 college student athletes who compete annually in college sports.
- Pioneered Prefect automation by utilizing event-based triggers to create a cross-client internal python package; reduced runtime by 30%.
- Revamped a client's historical seat level transactions per event per season FACT table logic; reduced pipeline runtime by 75%.

Klearnow.AI   **Nov 2022 - May 2023**
Senior Data Engineer, Data & AI Team
Pipeline Work:
- Led the data engineering team on re-designing a PySpark pipeline that loads shipment & merchandise data into Redshift warehouse's one-big-table via a multi-node EMR cluster.
- Collaborated with the AI team in building a data pipeline used for real time shipment tracking and analysis, by ingesting data extracted from third party shipment contracts.
- Reduced shipment tracking dashboard latency from 48 mins to 17mins, by identifying and resolving bottlenecks and inefficient code practices across the entire data pipeline.
- Architected & developed HubSpot API pipeline in Python to extract customer interaction data like contacts, emails, calls, notes, etc. and harness new avenues of insight. This increased customer retention by 18%.

Data Modeling:
- Reduced CPU utilization on the data warehouse by 22% by introducing materialized views and data validation before writing data into the respective schemas. This also fixed a multitude of data quality issues such as duplicates, data inaccuracy & inconsistency.

AstraZeneca   **Nov 2020 – Nov 2022**
Data Engineer, Enterprise Data Analytics Team
Pipeline Work:
- Engineered a data pipeline that ingests global rare disease clinical trials data that enhanced the R&D team's drug development, keeping data accuracy as the primary business goal while avoiding data redundancy.
- Developed a pipeline that migrated study & patient data from multiple sources into our Snowflake warehouse and in turn to our Qlik dashboards through AWS EC2 and S3 buckets. This opened the Clinical Trials to a larger population of patients based on the insight generated by the symptomology dashboards.

Data Modeling:
- Spearheaded a cross-functional data wrangling & integrity effort to predict deviation from other similar clinical trials by extracting data from various sources.

Marathon Energy   **May 2019 - Dec 2019**
Data Analytics Intern
Pipeline Work:
- Designed and implemented data pipelines with Python-Selenium web-scraper for revenue driving teams to improve data governance. These pipelines fueled PowerBI dashboards with data extracted from sources like National Grid and exported it to our local servers. This also improved the data refresh rate by 94%.

Latentview Analytics   **Jan 2018 - July - 2018**
Data Analyst, Analytics Team
Data Analysis:
- Surveyed stockholders, conceived ideas as a part of the data science team that predicts quarterly customer conversion rate and propose strategies to improve it.
- Established in-house methods to extract results from end-to-end descriptive analysis that administered Ad placement on the customer's website which increased sales and customer retention by 12%.

Cognizant Technology Solutions   **Aug 2014 – Jan 2018**
Data Engineer, Engineering Team
Pipeline Work:
- Led a cross-functional team that utilizes international customer transactions and global monetary data to enhance data accuracy & compliance through data warehousing techniques & complex queries.
- Reduced service downtime by implementing a caching system that helped warm up HDFS database with daily foreign exchange data on service startup by 85 minutes per day.

EDUCATION

Master of Science, Applied Data Science, Syracuse University   **2020**
Bachelor of Engineering, Electrical & Electronics Engineering, Anna University   **2014**