

# SANJEEV RAMASAMY SEENIVASAGAMANI

315-278-0599 | [sanjeevsramasamy@gmail.com](mailto:sanjeevsramasamy@gmail.com) | [www.linkedin.com/in/sanjeev-rs/](http://www.linkedin.com/in/sanjeev-rs/) | [rssanjeev.github.io](https://rssanjeev.github.io)

## TECHNICAL SKILLS

**Languages:** SQL, Python, PySpark, R, Bash | **Data Engineering:** SQL Server, Snowflake, AWS (EC2, S3, EMR, RDS, Redshift), Hive, | **Data Transformation & Orchestration:** Dimensional Modeling (Kimball), Star Schema, dbt, Prefect | **Visualization & Tools:** PowerBI, Tableau, Git, (CI/CD) CircleCI, JIRA, Confluence, Agile/Scrum

## PROFESSIONAL EXPERIENCE

### KAGR | Foxborough, MA

**Data Engineer IV** | Jul 2023 – Present

- Engineered a unified dbt framework on Snowflake, replacing 45+ monolithic SQL views and disparate JSON orchestrations; this transformation slashed client onboarding timelines from ~9 months to 2 sprints.
- Architecting a custom dbt parser to enforce Object-Oriented Programming (OOP) principles; this initiative targets a further reduction in implementation time to just 2 days by abstracting business logic into composable instances.
- Optimized historical seat-level transaction pipelines by refactoring FACT table logic, reducing processing runtime by 75% and significantly lowering compute costs.
- Developed a cross-client internal Python package utilizing Prefect for event-based orchestration, reducing manual intervention and cutting average pipeline runtime by 30%.

### Klearnow.AI | Remote

**Senior Data Engineer** | Nov 2022 - May 2023

- Led the re-design of a PySpark ingestion pipeline processing high-volume shipment data into a Redshift One-Big-Table (OBT) architecture via multi-node AWS EMR clusters.
- Slashed shipment tracking dashboard latency from 48 minutes to 17 minutes by optimizing data transformations and resolving bottlenecks in the critical path.
- Reduced Data Warehouse CPU utilization by 22% through the implementation of materialized views and upstream data validation frameworks (Great Expectations), eliminating data redundancy.
- Architected a Python pipeline to aggregate and resolve disparate people identities (contacts, interactions) from HubSpot, enabling a unified view that drove an 18% increase in customer retention.

### AstraZeneca | Boston, MA

**Data Engineer** | Nov 2020 – Nov 2022

- Engineered a secure AWS pipeline for sensitive PII/PHI clinical data, implementing strict Data Governance controls to accelerate R&D drug development.
- Migrated petabyte-scale patient data from legacy on-prem systems to Snowflake, feeding Qlik dashboards that expanded patient recruitment capabilities by 40%.
- Spearheaded data integrity initiatives by implementing automated anomaly detection scripts, ensuring 99.9% accuracy for FDA-compliant reporting.
- Designed dimensional data models to predict clinical trial deviations, integrating disparate data sources to support advanced predictive analytics.

### Marathon Energy | Syracuse, NY

**Data Analytics Intern** | May 2019 - Dec 2019

- Developed Python-Selenium web scrapers to automate data extraction from utility portals (National Grid), fueling PowerBI dashboards for the revenue team.
- Improved data refresh rates by 94% (from weekly to near real-time), enabling the sales team to react instantly to market rate changes.

### Latentview Analytics | Chennai, India

**Data Analyst** | Jan 2018 - July - 2018

- Partnered with the Data Science team to analyze quarterly customer conversion rates, proposing strategy adjustments that increased sales and retention by 12%.
- Executed end-to-end descriptive analysis on website traffic data to optimize ad placement strategies for high-value e-commerce clients.

### Cognizant Technology Solutions | Chennai, India

**Data Engineer** | Aug 2014 – Jan 2018

- Led a cross-functional team in optimizing international transaction data processing using Hadoop ecosystem tools to enhance regulatory compliance.
- Reduced service downtime by 85 minutes per day by implementing a distributed caching system for HDFS, accelerating daily foreign exchange data availability.

## EDUCATION

**Master of Science**, Applied Data Science, Syracuse University

2020

**Bachelor of Engineering**, Electrical & Electronics Engineering, Anna University

2014