

Sanjay Dadadahalli Prasanna Kumar

Miami, FL | 786-784-9890 | sanjaydp135@gmail.com | [LinkedIn](#) | [GitHub](#)

EDUCATION

Master of Science, Data Science & AI

Florida International University (GPA: 3.70/4.00)

Coursework: Data Vizualization, Data Mining, Data Science, AI, Data Modeling, Data Analytics

Bachelor of Engineering, ECE

SJB Institute of Technology (GPA: 3.72/4.00)

EXPERIENCE

Data Engineer (GRA)| Florida International University

Dec 2024 – Present

- Developed and deployed **end-to-end data pipelines** using **AWS Glue, Lambda, S3, EMR**, supporting sensor telemetry processing from ESP32/MPU6050 devices, improving real-time analytics availability by **40%**.
- Automated ingestion, cleansing, and transformation of **thermal imaging and environmental datasets** using **Python and SQL**, reducing manual preprocessing time by **60%**.
- Integrated **PostgreSQL** with dashboard tools for live research reporting, enhancing anomaly detection workflows in power electronics and energy systems.

Senior Data Engineer| Brillio Technologies Pvt Ltd

Nov 2021 - Jul 2024

- Built a **scalable global royalty platform** serving **1M+ artists**, increasing data pipeline availability by **30%** using **AWS Glue, Step Functions, Lambda, SQL stored procedures**, and Airflow
- Achieved **75% faster statement generation** by creating **event-driven ingestion workflows** using **S3 triggers** and serverless orchestration.
- Engineered a microservice with **Great Expectations** for data validation, reducing cluster memory usage by **40%** and improving data quality metrics.
- Led the **sync pipeline** between **Amazon Aurora** and **Redshift**, improving **query responsiveness by 50%** for downstream analytics teams.
- Supported stakeholders across finance/legal departments via **interactive Power BI dashboards**, increasing **revenue insight efficiency by 20%**.

Associate Professional Software Engineer| DXC Technologies Pvt Ltd

Jun 2021 - Nov 2021

- Built and supported **ETL pipelines** using **AWS Glue, Lambda, and Step Functions**, ingesting **2M+ daily records** from APIs, S3, and PostgreSQL.
- Automated job scheduling, transformation logic, and pipeline failure alerting using **Airflow** and **CloudWatch**, reducing data downtime by **35%**.

PROJECTS

Real-Time Retail Data Pipeline (Individual)

- Designed and implemented a comprehensive real-time retail data pipeline using Apache Kafka, Spark Streaming, and Airflow, featuring automated data quality validation, real-time monitoring with Prometheus/Grafana, and RESTful API integration for end-to-end retail analytics processing.

Driver Allocation Optimization using Machine Learning (Team Lead)

- Implemented LSTM, XGBoost, and reinforcement learning models to optimize driver allocation, forecasting, and real-time route decisions, reducing idle time by 20% and improving trip efficiency by 10%.

SKILLS

- **Programming Languages:** Python, C++, SQL, R(basics)
- **Big Data & Distributed Systems:** Apache Spark, Hadoop, HDFS, Hive, Kafka, Spark Streaming
- **Orchestration & Workflow:** Apache Airflow, DBT, AWS Step functions
- **Data Quality & Governance:** Great Expectations, Custom Alerting, Metadata Tracking
- **Databases & Query Engines:** PostgreSQL, HiveQL, Spark SQL, Postgres, Data Modelling (Snowflake Schema), NoSQL
- **Cloud & Storage:** AWS (S3, Lambda, EC2, Glue, Athena, EMR, IAM, CloudWatch), Azure(basics)
- **Data Visualization & DevOps Tools:** Power BI, Docker, Kubernetes, Terraform, Git, GitHub Actions, Microsoft Excel, Microsoft Office
- **Other Tools:** Streamlit, Plotly, REST API ingestion

CERTIFICATIONS

•AWS Certified: Cloud Practitioner

•Microsoft Certified: Azure Data Fundamentals