

Raj Nodagala Data Engineer

✉ rajnodagala@careerwebmail.com ☎ 9377276806 📍 Dayton, OH 🌐 LinkedIn

Profile

Experienced Data Engineer with over 6 years of expertise in data pipelines, databases, and cloud technologies. Proficient in Hadoop, Spark, Kafka, SQL/NoSQL databases, and data visualization tools. Implements CI/CD pipelines for efficient data processing. Skilled in Python programming with NumPy, Pandas, and Matplotlib for data analysis. Ensures data security, governance, and compliance. Strong focus on high availability through redundancy protocols and failover mechanisms in data systems.

Skills

Languages: Python (Pandas, NumPy, Scikit-Learn, Matplotlib, PySpark), R, SQL, NoSQL, Java, Scala, HTML, CSS

Data Processing & Streaming: Apache Kafka, Airflow, Snowflake, T-SQL, PL-SQL, SSIS, SSRS, ETL, MS Excel, Power BI, Azure SQL, Azure Data Factory (ADF), Google Analytics, Databricks, Hadoop, Synapse Analytics

Others Competencies: Git, GitHub, Docker, CI/CD pipelines, Agile, Jenkins, Power BI, Tableau, IntelliJ, Machine Learning, KNN, Logistic Regression, Decision Trees, JIRA

Database: MS SQL, MySQL, PostgreSQL, MongoDB

Professional Experience

Data Engineer, Goldman Sachs

08/2023 – Present | Remote, USA

- Initiated 5 ETL (Extract, Transform, Load) pipelines using Apache Spark and NiFi to process financial data from varied sources like banking systems, payment gateways, and transactional databases where Apache Spark allows fast and efficient data processing while Apache NiFi simplifies the automation of data activity between systems.
- Designed 10 robust data schemas for banking products, including loan portfolios and customer accounts, using Oracle databases to improve scalability and security by integrating databases with RDS and cloud-based solutions to handle the needs of the financial industry.
- Operated Python's SQLAlchemy library to interact with the databases and perform 20 data manipulation tasks.
- Exploited Python scripts for data extraction, transformation, and loading tasks, seamlessly merging them with SQL databases like PostgreSQL to optimize data processing and analysis workflows and leveraged Python libraries like Pandas for data manipulation, NumPy for numerical operations, and Matplotlib for data visualization.
- Fused accurate banking transactions and customer records and automated 50 data validation and quality assurance processes using Informatica Data Quality for data governance and to secure the reliability of critical financial data.
- Used Tableau, a powerful data visualization and business intelligence tool, to create 8 interactive dashboards for monitoring data quality metrics.
- Optimized 12 data processing jobs for performance using technologies like Apache Flink for real-time analytics, Apache Hadoop for batch processing, and Apache Cassandra for high availability in banking systems.
- Clutched agile methods and JIRA, a popular project control tool, to plan, track, and deliver projects efficiently.

Data Engineer, CISCO

02/2021 – 05/2022 | Bangalore, India

- Spearheaded the development of a cutting-edge CI/CD pipeline, streamlining the deployment of data pipelines and ensuring seamless integration and efficient delivery of data products resulted in a significant 30% reduction in deployment time and a 50% increase in pipeline reliability.
- Crafted a cloud-native solution utilizing S3 for data storage, Glue for ETL, Athena for serverless querying, and Redshift for data warehousing to enabled the effective display of financial data to stakeholders through BI dashboards and reports.
- Implemented Apache Kafka as a real-time data streaming platform, leveraging Kafka Connect to integrate diverse data systems and setting custom Kafka producers and consumers using Python ensuring reliable data ingestion and processing, enhancing responsiveness and scalability by 40%.
- Developed an ETL solution using AWS Glue, Python, and SQL, optimizing data transformation processes to achieve a 20% reduction in processing time and a 15% improvement in data quality.
- Initiated a data governance framework to enhance data security and compliance, resulting in a 25% reduction in data breaches and a 30% increase in regulatory compliance adherence that framework included encryption protocols, access controls, and regular audits to safeguard sensitive information.

Data Engineer, Coriant

10/2017 – 01/2021 | Bangalore, India

- Established and handled a diverse range of databases, including Microsoft SQL Server, data lakes, Synapse Analytics, and blob storage through meticulous adherence to best practices, I ensured data integrity, security, and optimal performance across these platforms.
- Enforced ETL pipelines leveraging Apache Kafka, SSIS, ADF, and data migration techniques to allow for consistent storage and analysis across platforms, ensuring data quality and accessibility.
- Crafted agile data models for efficiency and compliance, seamlessly migrating existing models to adapt to evolving storage technologies and business requirements, thereby maintaining uninterrupted operations.
- Optimized data warehouse solutions and designed using Snowflake, Azure SQL, and Synapse Analytics, including migration strategies from on-premises to cloud-based environments assuring superior performance and reliability.
- Created Tableau visualizations to extract insights from diverse data sources, such as data lakes, Synapse Analytics, and blob storage to facilitate existing Tableau dashboards and reports, ensuring uninterrupted access to critical insights amid the organization's evolving data infrastructure.

Data Engineer, Coriant

03/2017 – 09/2017 | Chennai, India

- Prepared an ETL pipeline architecture that integrated Python, Apache (Kafka, Airflow and Data Factory) to architecture scalable and efficient workflows, ensuring seamless data processing and analytics this pipeline was planned to handle large volumes of data, ensuring that supply chain operations ran smoothly and efficiently.
- Executed robust data quality checks and monitoring agents using Apache NiFi and Monitor these tools help real-time monitoring and alerting, ensuring that any issues were identified and resolved promptly and with Azure Databricks for real-time data processing and analytics, enabling agile decision-making and strategic planning.
- Automated data lineage tracking and documentation using Atlas and Data Catalog and also containerized ETL workflows with Docker and Kubernetes, easing seamless deployment and management across diverse environments including on-premises and multi-cloud setups.

Education

Master of Science, University of Dayton

05/2022 – 05/2024 | Ohio, USA

Computer Science

Bachelor of Technology, Jawaharlal Nehru Technological University

03/2013 – 03/2017 | AP, India

Computer Science and Engineering