**Shashank Shashishekhar Reddy**

San Jose, California | 5108927191 | shashankshashishekharreddy@gmail.com | [LinkedIn](https://www.linedin.com/in/shnkreddy) | [GitHub](https://www.github.com/shnkreddy98)

**Work Experience**

|  |  |
| --- | --- |
| **Data Engineer | Kantar** | Apr 2020 - Apr 2022 |

• Built and optimized ETL pipelines using Apache Airflow and Apache Spark, integrating real-time data streaming with Apache Kafka and automating ETL tasks with AWS Lambda

• Applied machine learning models for sales predictions, leveraging metrics like ad interaction and conversion rates, and integrated data quality checks for data integrity across ingestion, transformation, and storage

• Developed scalable data transformation and enrichment processes with schema validation, error handling, and unit testing of individual pipeline components

• Created data workflows for internal modeling and delivered actionable insights for clients like ING Bank Australia, Itau Bank Brazil, Pandora UK, and Heineken Africa

• Designed forecasting models for holiday effects to optimize inventory, resulting in a 12% increase in holiday sales

• Implemented RBAC using AWS IAM to secure sensitive data and ensure GDPR and CCPA compliance

• Designed ROI models to optimize ad spend, increasing sales by 34%

|  |  |
| --- | --- |
| **Data Engineer | The Sparks Foundation** | Apr 2018 - Mar 2020 |

• Engineered ETL pipelines using Apache Airflow to handle large-scale integration of NOAA and NASA datasets, improving efficiency by 25% through parallel processing and error-handling mechanisms for fault tolerance

• Automated data transformation processes using Python and SQL, reducing manual intervention by 40%

• Collaborated with cross-functional teams to optimize PostgreSQL and MySQL database performance, resulting in a 30% reduction in data retrieval times and a 20% reduction in processing time

• Developed real-time dashboards and visualizations, providing insights for stakeholders, including the implementation of a Star Schema to analyze NOAA Climate Data

• Applied AI models for predictive analytics to forecast extreme weather events and developed a recommendation system for NASA Earth Observation Data using collaborative filtering and neural networks

• Designed CI/CD pipelines using Jenkins and GitLab CI to automate ETL processes, including automated unit tests, Dockerized deployment of Apache Beam jobs, and continuous monitoring via CloudWatch

**Skills**

Python | SQL/NoSQL | Java | R | C | C++ | Bash | JavaScript | Scikit-learn | Keras | PyTorch | Delta Lake | Apache Iceberg | MLOps | AWS | S3 | Redshift | RDS | GCP | Azure | Snowflake | BigQuery | Databricks | dbt | Apache Hadoop | HDFS | Hive | Sqoop | HBase | MySQL | PostgreSQL | MongoDB | Elastisearch | Apache Airflow | Kafka | Spark | PySpark | Terraform | Docker | Kubernetes | Jenkins | Prometheus | Grafana | Presto | Flume | Dask | CI/CD Pipelines | Azure Blob Storage | GraphQL | Power BI

**Academic Projects**

|  |  |
| --- | --- |
| [**Your Own Cabs Analysis**](https://github.com/shnkreddy98/YourOwnCabsAnalysis) | Jan 2024 - May 2024 |

• Built scalable data pipelines to analyze booking and clickstream data for a cab service, ingesting data into Hadoop HDFS and AWS RDS via Sqoop, increasing throughput by 15%

|  |  |
| --- | --- |
| [**Spar Nord Bank ETL**](https://github.com/shnkreddy98/SparNordBankETL) | Jan 2023 - May 2023 |

• Designed an ETL process to move transactional data from MySQL RDS to Amazon Redshift and performed data analysis to optimize ATM refill processes, reducing refilling costs by 20%

**Education**

|  |  |
| --- | --- |
| Masters | Data Analytics | San Jose State University | Jan 2023 - Dec 2024 |

|  |  |
| --- | --- |
| PG Diploma | Data Science | IIIT | Aug 2020 - Sep 2021 |

|  |  |
| --- | --- |
| Bachelors | Computer Science | VTU | Aug 2016 - Aug 2020 |