

Shashank Shashishekhar Reddy

San Jose, California | (510) 892-7191 | shashankshashishekharreddy@gmail.com

Summary

Data Engineer with comprehensive expertise in administering and enhancing data applications to support high-throughput environments, especially in semiconductor processes. Familiar with statistical data analysis and proficient in creating data solutions that align with evolving manufacturing requirements. Possess strong programming skills in Python and Java, collaborating effectively with cross-functional teams to ensure optimally functioning data systems and insightful analytics.

Work Experience

Data Engineer | Kantar

Apr 2020 - Apr 2022

- Administered and optimized data infrastructure, enhancing data visualization applications for analytical use in production processes
- Updated and maintained models for key manufacturing statistics, ensuring relevancy with current product flow dynamics
- Trained end-users and provided support to facilitate effective use of data applications, resulting in improved departmental workflows
- Executed rigorous data quality checks during transformation and storage, ensuring high data integrity across systems
- Conducted in-depth statistical analysis and reporting which informed critical decision-making, leading to a marked increase in operational efficiency

Data Engineer | Sparks Foundation

Jul 2019 - Mar 2020

- Led the establishment of ETL pipelines integral for large-scale data integration within manufacturing frameworks, boosting process efficiency
- Worked closely with engineering teams to optimize database performance, enhancing data retrieval and analytical capabilities
- Automated complex data transformation and visualization processes using Python and SQL, significantly reducing workload
- Supported special projects pertaining to customer-oriented data inquiries, improving overall client satisfaction with data insights

Data Engineer | Sparks Foundation

Apr 2018 - Jun 2019

- Facilitated the development of scalable solutions for data extraction and transformation across multiple sources in the production sector
- Performed advanced data analysis to identify trends in production data, enhancing operational efficiency and decision-making
- Assisted teams in streamlining SQL queries, which significantly improved data processing times during analytics tasks

Skills

Programming Languages: Python, Java, C++, SQL, R, Bash, JavaScript, VBA Cloud & Data Platforms: AWS (S3, Redshift, RDS), GCP, Azure, Snowflake, dbt, Databricks, BigQuery, Apache Hadoop Ecosystem (HDFS, Hive, Sqoop, HBase) Databases: MySQL, PostgreSQL, MongoDB, Elasticsearch Tools: Apache Airflow, Kafka, Spark (PySpark, Spark Streaming), Terraform, Docker, Kubernetes, JIMP, Grafana, Tableau, Delta Lake, Apache Iceberg, Great Expectations, Presto, Flume, Jenkins, Dask Machine Learning Frameworks: Scikit-Learn, Keras, PyTorch Certifications: AWS Cloud Practitioner | Validation Code: QD8E074C82EQ1WCE.

Academic Projects

Booking and Clickstream Analysis

Jan 2024 - May 2024

- Developed robust data pipelines utilizing Kafka and Spark to handle vast amounts of clickstream data, enabling refined analytics
- Leveraged Hadoop for data storage and real-time processing capabilities, effectively increasing throughput by 15% for actionable insights
- Conducted comprehensive analyses on customer metrics using Hive tables, facilitating informed production decisions

ATM Refilling Optimization

Jan 2023 - May 2023

- Engineered a dynamic ETL process to migrate transactional data from MySQL RDS to Amazon Redshift, enhancing performance metrics
- Executed statistical analysis in Redshift, impacting resource allocation strategies and reducing operational costs by 20%

Education

Masters in Data Analytics | SJSU
PG Diploma in Data Science | IIIT
Bachelor's in Computer Science | VTU

Jan 2023 - Dec 2024
Oct 2023 - Sep 2021
Jul 2016 - Aug 2020