# Keerthi Reddy Sure

Bloomington, Indiana
812-369-7971 | surekeerthireddy@gmail.com | https://www.linkedin.com/in/keerthireddysure/

## Skills

**Programming Languages:** PySpark, SQL, Python, PostgreSQL, Hive, Java, Scala
**Databases:** MYSQL, Postgres, Relational Databases, Snowflake, Amazon S3, Aurora, Document DB, Bigquery, MongoDB
**Cloud:** Amazon Web Services(AWS), Google Cloud Platform(GCP), Microsoft Azure
**Developer Tools:** Kafka, Apache, Hadoop, Airflow, Docker, Git, Jenkins, Sonarqube, Tableau, LookerStudio
**Certifications:** AWS Data Engineer (associate), Data Engineering Professional Certificate - IBM

## Education

**Indiana University** **Aug 2022 – May 2024**
*Master of Science in **Data Science** 3.7/4.0* *Bloomington, Indiana*

**Malla Reddy Engineering College for Women** **Aug 2017 – Jul 2021**
*Bachelor of technology in Computer Science* *Telangana, India*

## Experience

**Indiana University** **August 2023 – May 2024**
*Data Engineer* *Bloomington, Indiana*

- Transformed legacy Python scripts into Spark applications, slashing application processing times by 36%, resulting in a significant boost in performance and operational efficiency.
- Developed AWS Lambda functions to automate email notifications using SNS to email students upon data updates, reducing manual tasks and increasing system efficiency by 20%, streamlining communication processes.
- Learned to leverage various cutting-edge technologies such as Spark using PySpark or Scala and AWS Lambda, gaining valuable insights into scalable and efficient data processing, pipelining and automation strategies.

**Opentext** **Jul 2021 – Jul 2022**
*Associate Data Engineer* *Hyderabad, India*

- Architected data collection mechanisms from diverse sources within an enterprise management product on the OT2 Platform. Designed departmental dashboards to streamline document workflow management for individual users, facilitating efficient data warehousing processes.
- Identified and resolved 30+ bugs, enhancing project performance by 2%. Implemented unit testing using JasmineJS, consistently improving performance metrics by at least 3%.
- Pioneered the development of intelligent web forms on OpenText's Appworks Platform, ensuring smooth migration from IBM Designer. Leveraged Java, Data Migration, and Selenium to decrease submission errors by 20%.

**Malla Reddy Engineering College for Women** **Jul 2020 – May 2021**
*Research Assistant* *Bloomington, Indiana*

- Utilized PySpark, Hive, and Airflow to build a data ingestion pipeline, including parallelizing data ingestion, writing data to HDFS as Avro files, creating external and internal tables, and scheduling the pipeline using Airflow..
- Performed data preprocessing tasks such as cleaning, feature scaling, and feature engineering using Python libraries like Pandas and NumPy, and worked with Hive tables, loading data, and writing Hive queries.

## Projects

**Real-Time E-Commerce Data Analytics Pipeline |** *AWS S3, AWS EMR, AWS Redshift, Airflow, Pyspark* **Jun 2023**

- Built a data pipeline using data from a e-commerce company to ingest and process user purchase data and product review data from PostgreSQL database, and create a de-normalized fact table containing user behaviour data divided by date in AWS Redshift for efficient analytical querying and reporting.
- User purchase events and product review data were ingested into AWS S3 as the raw data lake and then processed using AWS EMR before storing them in staging layer in S3 and final moving them into AWS Redshift.
- All the spark scripts and the sql scripts were stored in AWS s3 in script area. These scripts were used for transforming the dimension tables and creating a fact table containing the user behaviour data. This OLAP data warehouse contributed towards offering selective deals to customers based on their past interactions.

**Finhub Streaming Data Pipeline |** *Finhub.io API, Cassandra, Docker, Kafka, Kubernetes* **May 2022**

- Managed the creation of a scalable streaming data pipeline using Kubernetes, Kafka, Spark, and Cassandra, facilitating real-time processing of trading data sourced from Finnhub.io API.
- Engineered a robust data ingestion layer in Python, seamlessly capturing and encoding real-time trading data from Finnhub.io web sockets into Kafka message broker, ensuring data integrity and low-latency processing.
- Implemented Spark Structured Streaming for dynamic data transformation, enabling efficient loading of real-time data into Cassandra tables for persistent storage, and integrated Grafana for intuitive data visualization and monitoring of trading activities.