SUHAS RAMESH
Address: 13797 Lazy Creek Rd, Colorado springs, Colorado - 80921
Email: suhas.ramesh.111@gmail.com **LinkedIn:** Suhas Ramesh **Contact number:** (+1) (469)733-8191

EDUCATION

Master of Technology in Computer Science ***Aug'23– May'25***
University of Texas at Dallas GPA: 3.89
Bachelor of Technology in Computer Science Engineering ***Jul'19– Jul'22***
Reva University, Bangalore GPA: 3.96

SKILLS:

- Programming & Concepts: Python, Java, SQL, PySpark, Kafka, Hadoop (HDFS, MapReduce), Hive, ELK Stack (Elasticsearch, Logstash, Kibana), Pandas, NumPy
- Tools & Technologies: MongoDB, Oracle, Cassandra, HBase, Redis

WORK EXPERIENCE

Teaching Assistant, Department of Computer Science, University of Texas at Dallas, Dallas ***Aug'24– Present***

- Taught Big Data concepts to 100+ students, focusing on Hadoop ecosystem architecture (distributed, fault-tolerant, scalable).
- Guided students in designing HDFS-based pipelines, teaching data partitioning and replication for high availability and fault recovery.
- Developed labs on MapReduce and Hive, showcasing key-value operations (e.g., reduceByKey, join) for large-scale processing.
- Assisted with PySpark RDD workflows, explaining lineage-based fault tolerance and lazy evaluation of transformations.
- Mentored students on HBase and Cassandra, integrating NoSQL with collaborative filtering pipelines from external data sources, improving project outcomes by 20%.

Software Development Engineer, Software AG, Bengaluru ***Jul'22–Jun'23***

- Optimized SQL/Oracle tasks with shell scripting, cutting release times by 50%.
- Led "Smart Feedback" project using MongoDB and Redis, processing real-time customer data with variety and velocity.
- Mentored an intern on MS Teams and GIT, aiding their team integration.

INTERNSHIP

Software Development Engineer Intern, Software AG, Bengaluru ***Feb'22– Jun'22***

- Analyzed data with Pandas and NumPy, supporting development of data-driven features.
- Reduced bugs by 15% with testing, leveraging SQL for data validation.
- Adopted Agile workflows, speeding up product iterations

ACADEMIC PROJECTS

Sentiment Analysis on Social Media Data to Predict the Outcome of Soccer Matches ***Aug'23– Dec'23***

- Developed a data pipeline with Python, scraping 500+ Reddit comments per query using BeautifulSoup and Requests, filtering out malformed data.
- Implemented a scalable storage system with MongoDB for persistent data and Redis for caching, handling high-velocity social media inputs.
- Managed large datasets from Kaggle and betting sites, ensuring consistency for downstream processing.

Real-time Reddit Comment Word Count with Kafka, Spark, and Elastic Stack ***Feb'24– Apr'24***

- Engineered a streaming pipeline with Kafka and PySpark to process Reddit comments in real time from a specified subreddit.
- Configured Kafka topics (reddit_comments, wordcount) to ingest comments via PRAW and output word counts, using Spacy for entity extraction.
- Deployed ELK Stack with a custom Logstash filter to ingest data into Elasticsearch and visualize word frequencies on a Kibana dashboard.

DATAPOOL – KWIC index for search engines ***Aug'24– Dec'24***

- Built a search engine with MongoDB and Python, optimizing storage and retrieval for large-scale, distributed datasets.
- Designed a concurrent access system using the Facade pattern and BM25 ranking, improving query performance under multi-user loads.

Ad Click Prediction Using Machine Learning: Comparative Analysis, Tuning and Model Evaluation ***Aug'24– Dec'24***

- Engineered a data pipeline in Python to preprocess a 10,000-row ad click dataset, handling missing values with statistical imputation and KNN techniques.
- Optimized data quality by cleaning and encoding features (e.g., one-hot encoding for categorical variables), enabling efficient downstream ML processing.