Akash Reddy Maligireddy

+1 602-758-9848 | akashreddymaligireddy1@gmail.com | https://linkedin.com/in/AkashReddy_ | https://github.com/AkashReddy22 |Tempe, AZ

A passionate Data Engineer with 3 years of experience committed to building data intensive systems for advanced analytics

EDUCATION

Master of Science in Computer Science - Big Data Systems — *May 2024*
Arizona State University, Tempe, AZ — GPA: 4.0/4.0

Master of Science in Economics and Bachelor of Electronics Engineering — *June 2022*
Birla Institute of Technology and Science, Pilani, India — GPA: 8.25/10

TECHNICAL SKILLS

Programming/ Scripting: Python, SQL, Spark, Scala, Java, JavaScript, R, SAS, HTML, CSS, Bash, Unix
Technologies: Hadoop, Docker, DBT, Azure Databricks, Apache Kafka, Apache Spark, HDFS, Apache Airflow, Postman, AWS, RDS, EC2, S3, EMR and Redshift, Kubernetes, Airbyte, Neo4j, Flask, MySQL Server, pipeline infrastructure, Django
Tools, Databases, and Other: PostgreSQL, MySQL, NoSQL, MongoDB, Ajax, Git, Terraform, SQL Server, Ansible, Pandas, NumPy, Scikit-learn, Power BI, MacOS, Linux/Unix, YAML, JSON, Tableau, Data Governance, Data Warehouse

EXPERIENCE

Data Engineer Intern, *Ganges Systems*, *Irving, TX* — *07/2023 – 05/2024*

- Collaborated in the MDM Field Expansion project, enhancing the system's capacity to handle cost attributes up to **$9 million** for various products.
- Designed data pipelines using **Python**, **Airflow** to extract material data from **PostgreSQL**. Built a data warehouse in **AWS Redshift,** and loaded the data, optimizing the loading time by **80%**.
- Built data models on **dbt** using **SQL** and **Redshift**, and test **YAML** files to for validating data integrity, and consistency in an agile environment.

Data Engineer, *School of Mathematics and Statistics, Arizona State University, AZ* — *04/2023 – 05/2024*

- Architected a full stack **ETL**-**Web** data pipeline to collect student lab preferences, using **Flask**, **Bootstrap**, **JS, AJAX API**, and **MongoDB** for storage, and deployed the app on **Apache server**, reducing manual process by 90%**.**
- Integrated the Preference application with a live dashboard built using **JS,** improving search time by 100%.
- Designed a new **MySQL** database for backend using data modeling, migrated the data from MS Access to MySQL.
- Engineered a **Full-Stack** Web Application with **Flask**, **React JS**, **MySQL** and **REST API** to perform **CRUD** functions using interactive GUI and database optimizations, improving the efficiency by 95%.
- Secured the applications using **PyCAS**, improving security and adhering to ASU data governance policies.

Software Development Intern, *Tesser Insights, Atlanta, GA* — *05/2023 - 07/2023*

- Performed hyperparameter optimization for the Random Forest Classifier on **Azure Databricks** reducing runtime by 35%.
- Optimized the run-time of base **PySpark** code that automatically detects the feature importance of datasets by 30%.

Data Engineer, *JPMorgan Chase & Co, Bangalore, India* — *08/2021 - 06/2022*

- Implemented a batch processing **ETL** pipeline using **SAS, Spark** and analyzed risk profiles for 1 million credit default customers to enroll them into a repayment program that yielded a projected annual savings of $2.5 million.
- Refined an **ETL** pipeline using **Spark**, **Python** to automate reporting of customer risk data improving efficiency by 30%.
- Performed data analysis on impact of the US covid reliefs using **SQL** which impacted 30K customers and $1M, and reported using multiple graph **Tableau** Dashboard.

PROJECTS

AWS- Elastic Image Recognition Application (IaaS) — *08/2023 - 12/2023*

- Orchestrated a multi-tiered face recognition application on **AWS** infrastructure, utilizing **EC2** instances, **SQS** queues, **VPC**, and a custom auto-scaling algorithm, classifying 1000 images under 5 minutes with an accuracy of **95%**.
- Handled concurrent multiple requests, configured EC2 instances for the App Tier with custom AMIs, integrating a deep learning model and enabling auto-scaling, and S3 for input and output storage.

Hybrid Cloud- Smart Classroom Assistant with Kubernetes, AWS, OpenFaas — *08/2023 - 12/2023*

- Implemented a local **Kubernetes** cluster using **Minikube**, orchestrating the deployment of customized **docker** images for **OpenFaaS** functions, with FFmpeg and face_recognition library.
- Seamlessly integrated **AWS S3, AWS Lambda** within the hybrid cloud setup, ensuring a scalable object storage, elastic cloud service, and function based serverless computing to perform face recognition on input videos.

NYC Taxicab Cloud ETL pipeline — *01/2023 - 05/2023*

- Orchestrated a cloud ETL data pipeline using **Kubernetes** to process and analyze real-time New York taxi data.
- Streamed the data as messages through **Kafka,** performed graph modelling and then load into **Neo4j** database, and utilized **Graph Data Science** to find the shortest paths from different locations under 2 minutes.