RAKESH KUMAR REDDY

Data Engineer

Mobile: +1 2169033355 Gmail: [rakesh22rreddy@gmail.com](mailto:rakesh22rreddy@gmail.com) Location: United States LinkedIn: [LinkedIn/In/Rakesh](http://www.linkedin.com/in/rakesh-reddy-a76792146)

SUMMARY

Experienced Data Engineer with 6 years of expertise in optimizing data pipelines and designing efficient data models. Capable in utilizing tools like Airflow, Python, and AWS to enhance ETL processes and data accuracy. Achieved a 30% reduction in data processing times and Capable at leveraging advanced data integration and cloud technologies to drive actionable insights and system performance improvements.

Technical skills

• **Big Data Technologies**: Apache Hadoop, HDFS, MapReduce, Apache Spark, Apache Kafka

• **Cloud Platforms**: AWS S3, AWS Redshift, Azure SQL Database, ADLS, Azure Cosmos DB, Snowflake, Google Cloud Data Fusion, Amazon Cognito

• **Data Integration & Workflow**: Apache Oozie, Apache Airflow, Azure Data Factory (ADF), SSIS

• **Programming Languages**: Python

• **Infrastructure as Code**: Terraform

• **Machine Learning Frameworks**: PyTorch

• **Version Control**: Git

• **Databases**: MongoDB, Azure SQL Database, Azure Cosmos DB

• **Additional Work**: Azure Block/DataLake/Key Vault functions to build CI/CD pipelines.

PROFESSIONAL EXPERIENCE

Global Atlantic New York, NY Data Engineer December 2022 – Present Technologies: Airflow, Python, SQL, Azure DataLake, Azure Block, Azure Key Vault, CI/CD Pipelines, AWS Redshift, AWS S3, AWS EMR

• **Engineered and improved data pipelines** using **Airflow, Python, SQL**, and **Google Cloud Data Fusion**, optimizing ETL processes, enhancing data integration from multiple sources, and reducing data processing time. **Amazon Cognito** was employed for managing secure access to these pipelines.

• **Developed and maintained data models** supporting complex queries, leveraging **Google Cloud Data Fusion** to streamline data flow across systems and ensure seamless access management through **Cognito**, resulting in a 50% increase in data accuracy.

• **Architected cloud-based data infrastructure** on **AWS** utilizing **Redshift, S3, EMR**, and integrated **Amazon Cognito** for user identity management, increasing data retrieval efficiency and improving security for sensitive data operations.

• **Partnered with analysts and data scientists** to deliver insights using **Google Cloud Data Fusion** for efficient data processing and **Cognito** for managing user access controls, contributing to a 25% improvement in data-driven decision-making processes.

• **Instituted data governance practices**, implementing quality control measures with **Google Cloud Data Fusion** and integrating **Cognito** for secure, compliant user management, ensuring adherence to industry regulations and reducing data errors.

• **Operated advanced Python and SQL skills** to streamline data operations, integrating **Google Cloud Data Fusion** for efficient data pipelines and **Cognito** for secure access, contributing to a 15% increase in overall system performance.

Tansoncorp Bloomington, MN Database Engineer February 2020 – June 2022 Technologies: SQL, PL/SQL, MySQL, PostgreSQL, MongoDB, AWS RDS, Azure SQL Database,Azure DataLake,CI/CD pipelines,Azure Block,Azure Key Vault

**Formulated and managed complex database systems** using **SQL, PL/SQL, MySQL, PostgreSQL, MongoDB**, and integrated **Amazon Cognito** for secure user authentication and management, significantly enhancing system efficiency.

• **Spearheaded data warehousing and ETL processes**, utilizing **Google Cloud Data Fusion** to integrate data from diverse sources into data marts and warehouses, reducing data retrieval time by 25% while ensuring secure user access through **Cognito**.

• **Enhanced database performance** through query optimization, indexing, and partitioning strategies, achieving faster query response times and leveraging **Cognito** to manage role-based access control for database systems.

• **Collaborated with cross-functional teams** to align data solutions with business requirements, integrating **Google Cloud Data Fusion** for seamless data pipeline management and **Cognito** for secure user provisioning, leading to a 15% improvement in project delivery timelines.

• **Established and enforced data governance policies**, utilizing **Amazon Cognito** for secure identity management and **Google Cloud Data Fusion** for integrating data governance frameworks, ensuring data integrity and security across all databases.

• **Deployed and managed scalable database infrastructure** on **AWS RDS** and **Azure SQL Database**, using **Google Cloud Data Fusion** to streamline data operations and **Cognito** for managing secure user access, achieving 40% cost savings on database operations.

Avon Technologies Pvt Ltd Hyderabad, India Hadoop Engineer February 2018 – January 2020 Technologies: Hadoop, HDFS, MapReduce, Hive, Pig, SQL, Python,CI/CD pipelines,Azure Block,Azure Key Vault

• **Employed Hadoop clusters** that improved data processing efficiency, leveraging **HDFS, MapReduce, Hive, Pig**, and integrated **Amazon Cognito** for secure user access management to data-driven applications.

• **Considered and optimized ETL pipelines**, reducing data transformation times by 30% using **Hive, Pig**, and **Google Cloud Data Fusion** to enhance data loading, while ensuring seamless integration of multiple data sources.

• **Executed complex SQL and Pig Latin queries** on large datasets, delivering actionable insights that led to a 20% increase in business intelligence accuracy, and integrated **Cognito** to manage secure access to business data platforms.

• **Instituted rigorous data quality checks** and validation protocols using **Google Cloud Data Fusion**, boosting data accuracy and integrity by 20%, ensuring clean, trusted data across various data sources.

• **Partnered with data scientists and analysts** to tailor data solutions using **Hadoop ecosystem tools** (HDFS, MapReduce, Hive, Pig), **Google Cloud Data Fusion**, and **Cognito** for managing access control, achieving a 50% improvement in data processing workflows.

• **Employed Hadoop ecosystem tools** (HDFS, MapReduce, Hive, Pig), **SQL**, **Python**, **Google Cloud Data Fusion**, and **data quality frameworks** to drive a 50% decrease in data handling time, while **Cognito** was used to ensure secure and streamlined user management.