**PROFESSIONAL SUMMARY:**

**Rakesh Kumar Reddy Poreddy**

**+1 2169033355**

LinkedIn:[LinkedIn/In/Rakesh](http://www.linkedin.com/in/rakesh-reddy-a76792146) **[rakesh22rreddy@gmail.com](mailto:mail@gmail.com) Senior Data Engineer**

* Accomplished Senior Data Engineer with over 9+ years of experience in data-intensive environments.
* Expert in SQL database development and optimization, with extensive experience in MySQL and PostgreSQL, enhancing data management and query performance.
* Skilled in using Amazon Redshift for building scalable, high-performance data warehouses, enabling effective big data analytics.
* Experienced in implementing real-time data processing systems using Apache Kafka, supporting timely data analysis and decision-making.
* Proficient in deploying and managing cloud-based data solutions on AWS, leveraging services such as AWS S3 and AWS DynamoDB for data storage and retrieval.
* Adept at developing and tuning data pipelines using Apache NiFi and Informatica, ensuring accurate and efficient data flow across systems.
* Advanced user of data visualization tools including Power BI and Tableau, providing insightful and actionable business intelligence.
* Skilled in the use of Git for version control, maintaining code integrity and supporting collaborative development environments.
* Experienced with Apache Hadoop for managing large datasets, improving data processing and storage capabilities.
* Proficient in using Azure SQL Database and Azure Cosmos DB for creating and maintaining scalable and secure cloud databases.
* Skilled in deploying and managing Azure HDInsight, enhancing big data analytics capabilities with a managed cloud service.
* Utilizes Snowflake schema in Azure HDInsight for efficient data storage and fast query performance in cloud environments.
* Experienced with stream analytics using Azure Stream Analytics and PostgreSQL, enabling real-time data processing and insights.
* Proficient in data integration and workflow automation using Terraform, optimizing cloud infrastructure and data operations.
* Utilizes Databricks for complex data processing and machine learning tasks, integrating seamlessly with existing data architectures.
* Advanced skills in deploying machine learning models using PyTorch, enhancing predictive analytics and decision support systems.
* Experienced in automating data quality checks and data integration tasks with AWS Quick Sight and Python scripting.
* Proficient in agile project management, utilizing tools like JIRA to ensure efficient delivery of data engineering projects.
* Skilled in the implementation of collaborative analytics solutions using Pandas and DBT, fostering team-based data analysis.
* Expertise in data streaming and processing using Apache Kafka integrated with AWS EMR, ensuring efficient data handling and analysis.
* Expertise in data modeling and warehousing in Databricks, utilizing best practices for schema design and data organization.
* Engaged in continuous learning and application of new Databricks features and updates, keeping the data architecture modern and efficient.
* Played a key role in the strategic planning and execution of data governance and data lifecycle management using Databricks.
* Advanced knowledge in managing and querying MongoDB databases, supporting dynamic data storage needs in high-growth environments.
* Skilled in the strategic use of Terraform for infrastructure as code, automating cloud resource provisioning and management.
* Utilizes Apache Airflow for orchestrating complex data workflows, ensuring precise and reliable automation of data processing tasks.
* Proficient in data backup and recovery procedures, employing robust strategies to ensure data integrity and availability.
* Adept at developing and implementing comprehensive data security measures, ensuring compliance with industry standards and regulations.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Programming Languages:** | Python, SQL, Scala |
| **Databases:** | MySQL, PostgreSQL, Teradata, MongoDB, Azure Cosmos DB, AWS DynamoDB |
| **Tools:** | Talend, Apache Airflow, Apache Kafka, AWS Glue, Informatica, Apache NiFi |
| **Platforms:** | AWS, Azure, GCP |
| **Visualization:** | Power BI, Tableau, QlikView, AWS Quick Sight |
| **Technologies:** | ETL, Data Streaming, Machine Learning, Automation, Agile, Version Control (Git,  SVN) |
| **Frameworks:** | DBT, Pandas, Kubeflow |
| **Other:** | Terraform, Docker, Apache Hadoop, Snowflake |
| **Certifications** | Snowflake,Databricks,AWS |

**PROFESSIONAL EXPERIENCE:**

**Client: German Town Technologies, Portsmouth, VA Apr 2023 to till date Role: Sr Data Engineer**

**Roles & Responsibilities:**

• Implemented large-scale data pipelines using AWS EMR and Snowflake, improving healthcare data accessibility.

• Developed predictive models with Python and SQL to enhance patient care outcomes.

• Built real-time data streaming with Apache Kafka and AWS S3 for faster decision-making.

• Automated data workflows with Airflow, increasing efficiency and reducing errors.

• Optimized Databricks workflows with Apache Spark, reducing costs and improving performance.

• Integrated PySpark with AWS (S3, EMR, Redshift) for better data storage and processing.

• Developed real-time applications using PySpark and Kafka for enhanced data availability.

• Led data integration projects with MongoDB for healthcare applications.

• Automated data cleaning and preparation using Python.

• Improved data management using AWS Glue for consolidated data sources.

• Enhanced Databricks security and compliance through robust monitoring.

• Trained junior engineers on Databricks best practices.

• Collaborated with teams to deliver scalable data solutions and optimized database performance.

• Deployed DBT for efficient cloud-based data transformations.

• Managed real-time data streaming with Apache Kafka for analytics.

• Implemented governance using Databricks Unity Catalog for data security.

• Conducted seamless AWS EMR migrations with no data loss.

• Enhanced data security and compliance using AWS tools.

• Automated data processes using Python and AWS Lambda for high-quality analytics.

• Deployed monitoring tools and CI/CD strategies across cloud environments.

**Environment:** AWS EMR, Snowflake, Python, SQL, Databricks, Apache Kafka, AWS S3, Airflow, MongoDB, AWS Glue, PostgreSQL, Teradata, DBT, Pandas.

**Client: QBE, New York, NY Jan 2021 to Mar 2023 Role: Data Engineer**

**Roles & Responsibilities:**

• Automated data transformations with Terraform and Python, improving data consistency and reducing manual intervention.

• Enhanced data quality using Informatica and Apache NiFi for accurate reporting in financial services.

• Implemented real-time streaming with Databricks and Kafka, improving data availability for decision-making.

• Used Docker and Kubernetes to create reproducible environments for consistent data pipeline deployments.

• Automated AWS infrastructure provisioning with Terraform for scalable solutions.

• Led legacy system migration to Hadoop with PySpark for efficient data transformation.

• Provided business insights through Databricks SQL Analytics, enhancing decision-making.

• Managed machine learning lifecycle with Databricks MLflow and deployed models using PyTorch.

• Optimized AWS S3 storage, reducing costs with lifecycle policies and improving responsiveness with Lambda.

• Developed secure multi-tenant data storage in AWS S3 for better governance and compliance.

• Managed Agile projects with JIRA, ensuring timely delivery of data solutions.

• Streamlined data ingestion using AWS Glue and Python for automated processes.

• Optimized AWS DynamoDB for improved scalability and performance.

• Developed real-time data feeds with Kafka for immediate decision-making.

• Automated workflows with Informatica, improving data transformation efficiency.

• Implemented disaster recovery strategies and enhanced security using AWS tools.

• Facilitated team collaboration and project management using Agile and JIRA.

• Trained team members on new technologies, fostering continuous improvement.

**Environment:** AWS Glue, AWS S3, AWS DynamoDB, Databricks, Terraform, Python, Informatica, Apache NiFi, PyTorch, AWS QuickSight, JIRA, SQL, Git, Apache Kafka, Confluence.

**Client: PNC Bank, Pittsburgh, PA Oct 2018 to Dec 2020 Role: Data Engineer**

**Roles & Responsibilities:**

• Designed data ingestion solutions with Apache NiFi, optimizing financial data flow for analytics.

• Configured AWS DynamoDB for scalable, reliable storage in the financial sector.

• Automated cloud deployments with Terraform, enhancing agility and reliability.

• Deployed data applications in Docker containers for better scalability and manageability.

• Developed Python scripts to streamline data transformation tasks.

• Integrated AWS S3 for secure, scalable data storage in finance.

• Utilized Git for version control, ensuring collaboration in financial data projects.

• Automated data pipelines with Informatica, improving data quality and efficiency.

• Implemented robust pipelines in Databricks with Apache Spark for real-time processing.

• Managed the machine learning lifecycle with Databricks MLflow.

• Created dashboards and visualizations with AWS QuickSight for financial decision-making.

• Configured AWS DynamoDB for high-performance data operations.

• Managed agile projects with JIRA for timely data engineering deliveries.

• Ensured model accuracy with data validation using PyTorch.

• Applied AWS data encryption for compliance with financial regulations.

• Developed data warehouses using AWS for complex financial reporting.

• Automated monitoring and alerting in Databricks for process reliability.

• Created custom UDFs in Databricks for enhanced data transformations.

• Used Tableau for advanced financial data visualizations.

• Improved real-time data streaming with Apache Kafka.

• Streamlined infrastructure management with Terraform automation.

• Implemented CI/CD practices with Docker for reliable data applications.

• Used DynamoDB streams for real-time financial data tracking.

• Managed data migrations to AWS with minimal downtime.

• Automated data integration with AWS Glue for improved processing.

**Environment:** Apache NiFi, AWS Glue, DynamoDB, Terraform, Docker, Python, Git, Informatica, AWS S3, QuickSight, JIRA, PyTorch, Tableau, Apache Kafka, and AWS DynamoDB.

**Client: Exl service.com (I) Pvt.Ltd, India Apr 2017 to Jun 2018 Role: Data Quality Analyst**

**Roles& Responsibilities:**

• Analyzed data quality and implemented improvements using Python, SQL, and Apache Hive for big data insights.

• Designed visualizations in QlikView and Tableau, driving strategic decisions through insightful reports.

• Managed version control with Subversion (SVN) for collaborative data project development.

• Enhanced reporting with Power BI by developing robust data models.

• Leveraged Apache Hadoop for processing large datasets, improving data access and analysis.

• Developed and maintained SQL-based data warehouses for efficient data storage and retrieval.

• Configured Apache Hive for optimized data queries and analytics.

• Implemented data governance to ensure integrity and security.

• Conducted Tableau and QlikView workshops to improve data visualization skills.

• Optimized SQL queries, improving data retrieval speed and response times.

• Managed data migrations and integrations using Apache Hadoop, ensuring accuracy.

• Developed SQL scripts to enhance database management and reporting.

• Created interactive Power BI reports, improving data accessibility for non-technical users.

• Tuned databases for better performance and reliability.

• Managed tasks and projects using agile practices and JIRA.

• Delivered data analysis reports to stakeholders, influencing business strategies.

**Environment:** Python, SQL, Apache Hive, QlikView, Tableau, Subversion (SVN), Power BI, Apache Hadoop, JIRA.

**Client:HighRadiusTechnologies,India Jul 2015 to Mar 2017 Role: SQL Developer**

**Roles & Responsibilities:**

• Developed and managed SQL databases (MySQL, PostgreSQL) for business analytics and reporting.

• Utilized Talend for data integration, supporting business intelligence.

• Managed data backups with Git, ensuring data integrity and availability.

• Created Power BI dashboards for actionable business insights.

• Optimized SQL database performance for improved efficiency.

• Documented database designs and data management processes.

• Wrote SQL queries for data analysis and reporting.

• Collaborated with business analysts to provide database solutions.

• Led seamless data migration projects.

• Configured PostgreSQL databases for optimal performance and security.

• Developed Python tools for data extraction and reporting.

• Trained junior developers in database management and ETL processes.

**Environment:**

MySQL, PostgreSQL, Talend, Git, Power BI, Python.

**Education:**

Guru Nanak Institution Technical Campus