Rakesh Kumar Reddy
+1 216903335
LinkedIn:LinkedIn/In/Rakesh
rakesh22rreddy@gmail.com
Senior Data Engineer

PROFESSIONAL SUMMARY:

- 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

PROFESSIONAL EXPERIENCE:

Client: German Town Technologies, Portsmouth, VA Apr 2023 to till date

Role: Sr Data Engineer

Roles & Responsibilities:

- Implemented data pipelines for large-scale healthcare data analysis using AWS EMR and Snowflake, enhancing data accessibility and analysis capabilities.
- Developed predictive models to improve patient care outcomes, utilizing Python and SQL in conjunction with collaborative analytics tools.
- Engineered real-time data streaming solutions with Apache Kafka and AWS S3 to facilitate instant access to healthcare data for timely decision-making.
- Automated data processing workflows using Airflow, increasing operational efficiency and reducing manual errors in data handling.
- Developed complex data pipelines using Databricks, optimizing data flow from multiple sources into structured and unstructured data lakes.
- Leveraged Apache Spark on Databricks for large-scale data processing, achieving significant performance improvements and cost reductions.
- Spearheaded the integration of PySpark with AWS services (S3, EMR, Redshift) to enhance data storage and retrieval processes.
- Utilized PySpark with Kafka to build real-time data streaming applications, improving data availability and decision-making capabilities.
- Implemented machine learning algorithms with PySpark MLlib to predict trends and behaviors from large datasets, enhancing business strategies.
- Led projects that integrated data from various sources into MongoDB, enhancing data storage and retrieval processes for critical healthcare applications.
- Utilized Python to create scripts that automated routine data cleaning and preparation tasks, streamlining the data pipeline.
- Implemented data integration solutions using AWS Glue to consolidate disparate data sources, simplifying data management and improving data quality.
- Enhanced data security and compliance in Databricks environments by implementing robust security measures and monitoring tools.

- Trained and mentored junior data engineers and analysts on best practices and efficient use of Databricks in data engineering projects.
- Collaborated with cross-functional teams to define data requirements and deliver scalable data solutions using Databricks.
- Optimized SQL queries and database schemas in PostgreSQL and Teradata, significantly improving performance and scalability of healthcare data applications.
- Deployed DBT for data transformation tasks within cloud environments, enabling more efficient data modeling and reporting.
- Coordinated with clinical staff to leverage analytics in operational and patient care strategies, using Pandas to analyze and interpret data.
- Configured and managed data streaming architectures using Apache Kafka, ensuring robust data flow and integration for real-time analytics.
- Fostered a collaborative environment using Databricks Notebooks, facilitating sharing insights and working closely with data scientists to refine analytics models.
- Implemented data governance practices using Databricks Unity Catalog, ensuring compliance with data security and privacy standards.
- Conducted data migration projects to AWS EMR, ensuring seamless transitions with minimal downtime and no data loss.
- Developed and maintained robust data security protocols using AWS technologies, safeguarding sensitive healthcare data against potential breaches.
- Integrated machine learning algorithms with existing data systems using Python and Pandas, enhancing predictive analytics capabilities.
- Enhanced data visualization and reporting capabilities using Python and SQL, enabling healthcare professionals to access tailored dashboards.
- Automated various data processes using Python scripting, significantly reducing the time required for data preparation and loading.
- Performed extensive data cleaning and normalization to improve the quality of healthcare analytics, using Python and SQL. Architected a data lake solution utilizing AWS S3 as the primary storage, paired with AWS Glue and Athena for efficient serverless querying.
- Enhanced data security for AWS S3 storage by implementing comprehensive bucket policies, IAM roles, and enforcing encryption standards both in transit and at rest.
- Automated data transformation and cleansing processes using Python scripts and AWS Lambda, storing outputs in AWS S3 to ensure high-quality data for analytics.
- Managed large datasets using MongoDB, optimizing data architecture for better performance and scalability in healthcare applications.
- Integrated real-time and batch data processing using Apache Kafka and AWS EMR, balancing workload and improving processing time.
- Deployed and optimized DBT models in cloud environments, enhancing data transformation and load processes for complex healthcare datasets.
- Implemented automated monitoring with Prometheus and Grafana in CI/CD workflows, enabling proactive issue detection and resolution in data pipelines.
- Utilized Ansible for configuration management, automating the setup and maintenance of Hadoop clusters integrated with CI/CD pipelines.
- Architected a CI/CD strategy for a multi-cloud environment using Spinnaker, improving deployment strategies across AWS platform.
- Assisted in the development of a collaborative analytics platform, facilitating data sharing and decision-making across healthcare teams.
- Implemented robust backup and disaster recovery solutions using AWS technologies, ensuring high availability and data integrity.
- Conducted performance tuning of healthcare databases in PostgreSQL and Teradata, ensuring optimal performance during critical data operations.
- Streamlined data ingestion and integration using AWS Glue and Python, enhancing the speed and efficiency of data flows into healthcare analysis systems.

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 using Terraform and Python, enhancing data consistency and reducing manual intervention across insurance datasets.
- Improved data quality and integration using Informatica and Apache NiFi, which facilitated accurate reporting and analytics in financial services.
- Implemented real-time data streaming solutions with Databricks Structured Streaming and Apache Kafka to enhance data availability and decision-making processes.
- Leveraged Docker and Kubernetes to create reproducible environments for data pipelines, facilitating consistent deployments across development, testing, and production stages.
- Developed Terraform scripts to automate the provisioning of cloud infrastructure on AWS, ensuring scalable and resilient data engineering solutions.
- Led the migration of legacy data systems to Hadoop-based platforms, leveraging PySpark for efficient data transformation and aggregation.
- Utilized Databricks SQL Analytics to provide actionable insights through dashboards and visual reports, enhancing business decision-making.
- Led migration projects to Databricks, achieving seamless transitions from legacy systems and enabling cloud-based data analytics capabilities.
- Utilized Databricks MLflow to manage the machine learning lifecycle, including experimentation, reproducibility, and deployment of ML models.
- Deployed machine learning models for insurance risk assessment using PyTorch, integrated with AWS Quick Sight for dynamic visualization.
- Optimized costs by analyzing and reconfiguring data storage and retrieval practices in AWS S3, applying lifecycle policies to transition to cost-effective storage tiers.
- Leveraged AWS S3 event notifications to automate and trigger downstream processing in AWS Lambda, enhancing responsiveness in event-driven data architectures.
- Created a secure, scalable multi-tenant data storage framework using AWS S3, improving data access control, governance, and compliance with international standards.
- Managed Agile project implementations, utilizing JIRA for project tracking, ensuring timely delivery of data enhancements and upgrades.
- Developed cost-effective solutions on Databricks by optimizing cluster management and leveraging spot instances.
- Ensured data security by implementing ACLs and role-based access control within Databricks environments.
- Designed data pipelines that combined AWS Glue with Python scripting to automate and simplify data ingestion and integration processes.
- Conducted data quality assessments using SQL and Python, ensuring high data integrity and supporting compliance with regulatory standards.
- Optimized data storage and processing using AWS DynamoDB, improving performance and scalability for large insurance datasets.
- Implemented version control best practices using Git, enhancing collaboration and code management across the data engineering team.
- Enhanced data visualization capabilities using AWS Quick Sight, enabling stakeholders to derive actionable insights from complex datasets.
- Streamlined data migration processes using Terraform, ensuring efficient and error-free transitions between different storage platforms.
- Facilitated the integration of structured and unstructured data using Apache NiFi, supporting comprehensive analytics in insurance applications.

- Developed real-time data feeds using Apache Kafka, enabling immediate data availability for timely decision-making in risk management.
- Orchestrated data workflows using Informatica, automating complex transformations and loading processes in the insurance data environment.
- Designed and implemented robust backup and disaster recovery strategies using AWS technologies, ensuring data availability and continuity.
- Automated repetitive data processing tasks using Python, significantly improving efficiency and reducing processing times.
- Configured AWS S3 for optimized data storage and retrieval, effectively managing large volumes of insurance transaction data.
- Conducted performance tuning of AWS DynamoDB instances, optimizing response times and resource utilization for critical applications.
- Implemented security measures in data handling and storage using AWS security tools, safeguarding sensitive insurance customer data.
- Utilized Apache NiFi for efficient data ingestion and distribution, enhancing the flow of information across insurance processes.
- Enhanced team collaboration and project management using Agile methodologies and JIRA, improving overall project visibility and tracking.
- Developed and maintained documentation for data processes and systems using Confluence, ensuring knowledge sharing and continuity.
- Supported the training of team members on new technologies and data processes, fostering a culture of continuous learning and 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 using Apache NiFi, optimizing the flow of financial data for analytics purposes.
- Configured AWS DynamoDB to provide scalable and reliable data storage solutions tailored for the financial sector.
- Utilized Terraform to automate cloud infrastructure deployments, enhancing operational agility and system reliability.
- Deployed data applications in Docker containers to improve scalability and manageability of financial services applications.
- Developed Python scripts for data transformation, streamlining the processing tasks within the financial data environments.
- Integrated AWS S3 for secure and scalable data storage, supporting extensive data management solutions in finance.
- Employed Git for version control, ensuring the integrity and collaboration in the development of financial data projects.
- Orchestrated automated data pipelines using Informatica, improving data quality and efficiency in data handling.
- Designed and implemented robust data pipelines in Databricks, integrating Apache Spark for real-time data processing and analytics.
- Skilled in using Databricks MLflow to manage machine learning lifecycle, including experimentation, reproducibility, and deployment.
- Leveraged AWS QuickSight for developing insightful dashboards and visualizations, aiding financial decision-making processes.
- Configured AWS DynamoDB for high-performance data operations, enhancing data retrieval and storage in financial applications.

- Managed agile project cycles using JIRA, ensuring timely delivery of data engineering projects in the finance domain.
- Implemented data validation and testing using PyTorch, ensuring accuracy and reliability of predictive models in finance.
- Applied data encryption and security measures in AWS, ensuring compliance with financial regulations and data privacy standards.
- Developed and maintained data warehouses using AWS technologies, facilitating complex financial analyses and reporting.
- Implemented automated monitoring and alerting mechanisms in Databricks, ensuring high availability and performance of data processes.
- Developed custom UDFs (User Defined Functions) in Databricks for specific business logic integration, enhancing data transformation capabilities.
- Utilized Tableau for complex data visualization tasks, enhancing the presentation and accessibility of financial data.
- Enhanced data workflows with Apache Kafka, improving the real-time data streaming capabilities in financial operations.
- Configured automation frameworks using Terraform, streamlining infrastructure management for financial services.
- Applied continuous integration and deployment practices using Docker, enhancing the reliability of financial data applications.
- Utilized AWS DynamoDB streams to capture real-time changes in financial data, enhancing data accuracy and timeliness.
- Orchestrated data migrations to AWS cloud environments, ensuring seamless transitions and minimal downtime.
- Enhanced operational efficiency and data processing using AWS Glue, automating data integration tasks in the financial domain.

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 enhancements using Python and SQL, utilizing Apache Hive for big data management and insights.
- Designed visualizations in QlikView and Tableau, facilitating insightful business reports and dashboards that drove strategic decisions.
- Managed version control for data projects using Subversion (SVN), ensuring code integrity and facilitating collaborative development.
- Supported business intelligence initiatives with robust data models using Power BI, enhancing reporting capabilities across departments.
- Leveraged Apache Hadoop for efficient processing of large datasets, improving data access and analysis for client projects.
- Developed and maintained data warehouses using SQL, ensuring structured data storage and efficient data retrieval.
- Configured and utilized Apache Hive to handle complex data queries, optimizing data operations and analytics.
- Implemented data governance and compliance measures, ensuring data integrity and security were maintained.
- Conducted data visualization workshops using Tableau and QlikView, enhancing team skills and data presentation techniques.
- Optimized data retrieval processes using custom SQL queries, reducing processing times and improving response rates.

- Performed data migrations and integrations using Apache Hadoop, ensuring data consistency and accuracy.
- Developed SQL scripts for database management and report generation, enhancing operational efficiency and data usability.
- Utilized Power BI to develop interactive and automated reporting solutions, increasing data accessibility for non-technical users.
- Assisted in database tuning and performance optimization, ensuring high performance and reliability of data operations.
- Engaged in agile project management practices, utilizing JIRA to manage tasks and monitor project progress.
- Delivered comprehensive data analysis reports to stakeholders, providing insights that influenced key 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 SQL databases using MySQL and PostgreSQL, enhancing data operations for business analytics and reporting.
- Utilized Talend for data integration and transformation, supporting analytics and business intelligence activities.
- Managed data backups and recovery procedures using Git, ensuring data integrity and high availability.
- Implemented dashboarding solutions with Power BI, delivering actionable insights to enhance business decision-making.
- Conducted performance tuning on SQL databases to ensure optimal operation and access speeds.
- Developed and maintained documentation for database designs and data management processes.
- Designed and executed SQL queries for data analysis and reporting, supporting various business units.
- Collaborated with business analysts to understand data requirements and deliver tailored database solutions.
- Participated in data migration projects, ensuring seamless data transfers with minimal downtime.
- Assisted in the setup and configuration of PostgreSQL databases, optimizing settings for performance and security.
- Developed custom data extraction and reporting tools using Python, enhancing data accessibility and user engagement.
- Trained junior developers in database management and ETL processes, fostering skill development within the team.
- Monitored and resolved database performance issues, ensuring stable and efficient data operations.
- Engaged in project meetings to provide updates on database health and data management strategies.

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

Education:
guru nanak institution technical campus