RAKESH KUMAR REDDY
+1 2169033355
rakesh22rreddy@gmail.com
Data Engineer

PROFESSIONAL SUMMARY:

- A seasoned Data Engineer with over 6 years of experience in managing and processing large datasets using advanced Big Data technologies.
- Proficient in deploying Apache Hadoop, HDFS, MapReduce, and Kafka for efficient data handling and streaming across various platforms.
- Specializes in creating robust ETL pipelines utilizing Apache Airflow, Apache Oozie, and scripting to enhance data integration and workflow automation.
- Expert in cloud data management using AWS S3, AWS Redshift, and Azure services, ensuring scalable and secure data solutions.
- Skilled in data modeling and analysis using Python, SQL, and SSIS, providing comprehensive support for data-driven decision-making.
- Experienced in using Apache Spark and Azure Stream Analytics for real-time data processing and analytics.
- Capable of implementing secure and compliant data environments with Azure Active Directory and Terraform.
- Proficient in machine learning implementations using PyTorch on cloud platforms to develop predictive models and analytics.
- Committed to maintaining data integrity and quality through rigorous validation and testing using industry-standard tools and methodologies.
- Adept at performance tuning and optimization of data-intensive applications to ensure efficiency and reliability.
- Utilizes Git for effective version control and collaboration in software development projects involving large teams.
- Knowledgeable in setting up MongoDB and Azure Cosmos DB for optimal data storage and retrieval operations.
- Experienced in the insurance industry, effectively managing data solutions tailored to specific business needs and compliance requirements.
- Continuously updates skills in the latest data technologies and methodologies, ensuring relevance in a rapidly evolving industry.
- Collaborates effectively with both technical and non-technical stakeholders to translate complex data needs into actionable IT strategies.
- Regularly participates in Agile and Scrum sessions, contributing to continuous improvement and iterative development.
- Develops and maintains comprehensive documentation for data processes and systems to support audit and compliance requirements.
- Strong communicator, adept at presenting complex data concepts and solutions in clear, understandable terms.
- Engages in professional development opportunities to remain at the forefront of technology advancements in data engineering.

- Implements advanced security features to safeguard sensitive data across multiple platforms and environments.
- Dedicated to delivering high-quality and timely project outcomes, prioritizing client satisfaction and operational excellence.
- Focused on leveraging extensive industry experience to drive innovations and improvements in data management and analytics.

TECHNICAL SKILLS:

- **Big Data Processing :** Apache Hadoop, HDFS, MapReduce, Apache Spark
- **Data Streaming :** Apache Kafka
- **Cloud Services :** AWS S3, AWS Redshift, Azure SQL Database, ADLS, Azure Cosmos DB
- **ETL Tools :** Apache Oozie, Apache Airflow, Azure Data Factory (ADF), SSIS
- **Scripting & Automation :** Python, Terraform
- **Data Warehousing :** Snowflake, AWS Redshift
- **Machine Learning :** PyTorch
- **Version Control :** Git
- **Database Management :** MongoDB, Azure SQL Database, Azure Cosmos DB

PROFESSIONAL EXPERIENCE:

Client: Global Atlantic, New York, NY **Dec 2022 to Till date**

Role: Data Engineer

Roles & Responsibilities:

- Currently engaged in integrating Apache Spark with Azure SQL Database to optimize data processing capabilities in the insurance domain.
- Utilized Azure Data Lake Storage (ADLS) to effectively manage large-scale data storage needs for comprehensive data analysis.
- Developed and maintains ETL workflows using Azure Data Factory, ensuring seamless data integration and management.
- Implemented real-time data analytics solutions using Azure Stream Analytics to provide actionable insights for business decision-making.
- Managed secure data environments using Azure Active Directory, ensuring compliance with industry-standard security protocols.
- Designed and executes data migration scripts using Terraform, enhancing data agility and system reliability in cloud environments.
- Optimized data storage and retrieval operations using Snowflake, improving performance in cloud-native data warehousing.
- Deployed machine learning models using PyTorch to facilitate predictive analytics and risk assessment in the insurance sector.
- Automated routine data operations using Python scripts to improve efficiency and reduce manual intervention.
- Streamlined version control processes with Git, enhancing collaboration and code management in project development.
- Designed efficient SQL database schemas to enhance data accessibility and maintain data integrity across multiple platforms.
- Ensured high standards of data integrity and accuracy are maintained throughout all stages of data processing.

- Monitored and troubleshoots Azure-based data solutions, ensuring operational stability and efficiency.
- Implemented robust data backup and recovery protocols to safeguard data against potential threats and losses.
- Ensured the quality and validation of data before processing, prioritizing accuracy and reliability.
- Coordinated with cross-functional teams to align project goals with business objectives, ensuring a cohesive approach.
- Documents data processes and system changes to maintain compliance with regulatory standards and facilitate audits.
- Participated actively in Agile sprints, focusing on continuous improvement and iterative development processes.
- Developed custom data solutions tailored specifically to the needs of the insurance sector, enhancing operational effectiveness.
- Resolved data ingestion and pipeline issues promptly, maintaining system performance and reliability.
- Lead knowledge sharing sessions on Azure technologies, enhancing team capabilities and system efficiencies.
- Enhances system performance through strategic scripting and data handling techniques, focusing on optimization and scalability.

Environment: Apache Spark, Azure SQL Database, Azure Data Lake Storage, Azure Data Factory, Azure Stream Analytics, Azure Active Directory, Terraform, Snowflake, PyTorch, Python, Git, SQL.

Client: Tansoncorp, Bloomington, MN **Feb 2020 to Jun 2022**

Role: Database Engineer

Roles & Responsibilities:

- Managed large-scale data storage using AWS S3, ensuring efficient data retrieval strategies tailored to client needs.
- Developed robust data warehousing solutions using AWS Redshift, enhancing analytics capabilities for business intelligence.
- Designed and executed ETL pipelines leveraging AWS EMR and Apache Airflow, optimizing data flow and processing.
- Utilized Hadoop YARN for effective resource management in large Big Data clusters, enhancing operational efficiency.
- Configured AWS Schema Conversion Tool for seamless database migration projects, ensuring data consistency across systems.
- Programmed complex data transformations utilizing Python and SQL to support advanced analytics and reporting needs.
- Implemented Apache Kafka for data ingestion, facilitating real-time data feeds and streaming capabilities.
- Ensured data accuracy and consistency through SSIS-based transformations, maintaining high standards of data quality.
- Automated routine data operations to enhance process efficiency and accuracy, reducing manual efforts and errors.
- Collaborated closely with cross-functional teams to ensure IT strategies are aligned with broader business objectives.
- Optimized MongoDB schema designs for performance improvements, enhancing data handling and query response times.
- Documented technical specifications and project progress comprehensively, maintaining clear and effective communication.
- Conducted thorough testing to validate data integrity and quality, ensuring reliable and accurate data outputs.
- Integrated security protocols for data transactions, safeguarding sensitive information against unauthorized access.
- Refined data processing workflows to achieve operational excellence, focusing on efficiency and effectiveness.
- Adapted quickly to rapidly changing technology requirements, maintaining relevance and competitiveness in the IT services domain.
- Ensured robust data backup and recovery systems were in place, providing resilience against data loss scenarios.
- Maintained high levels of team collaboration and communication, fostering a productive and inclusive work environment.
- Developed comprehensive data handling and processing guidelines, standardizing procedures across the

organization.
- Led initiatives to enhance data security and compliance, ensuring adherence to legal and regulatory requirements.

Environment: AWS S3, AWS Redshift, AWS EMR, Apache Airflow, Hadoop YARN, AWS Schema Conversion Tool, Python, SQL, Apache Kafka, SSIS, MongoDB.

Client: Avon Technologies Pvt Ltd, Hyderabad, India **Feb 2018 to Jan 2020**
Role: Hadoop Engineer
Roles & Responsibilities:

- Engineered scalable data processing solutions using Apache Hadoop, enhancing data handling capabilities across distributed systems.
- Implemented automated data workflows with Apache Oozie, streamlining operations and improving process efficiency.
- Managed extensive data storage requirements using HDFS, ensuring data availability and reliability for large-scale applications.
- Utilized MapReduce to facilitate efficient data processing across distributed environments, improving throughput and performance.
- Developed comprehensive data ingestion solutions utilizing Apache Hive and Pig, enhancing data accessibility and manipulation.
- Scripted automation tasks in Python to streamline data workflow processes, reducing manual intervention and increasing accuracy.
- Managed version control effectively using Git, facilitating team collaboration and maintaining a robust codebase.
- Applied industry-standard SQL practices to ensure robust data manipulation and retrieval across various data platforms.
- Integrated Apache Kafka with the Hadoop ecosystem to enhance data streaming and real-time processing capabilities.
- Leveraged Apache Oozie for automating complex data processing jobs, increasing operational efficiency and reliability.
- Enhanced data query capabilities using optimized Hive scripts, improving data access and analysis across business units.
- Prioritized data security and integrity in all processing tasks, ensuring the protection of sensitive information.
- Implemented robust data backup and recovery strategies, safeguarding critical data assets against potential threats.
- Participated in project planning and scoping sessions, aligning data strategies with business objectives and requirements.
- Maintained high standards of data quality and accuracy, implementing rigorous validation and testing procedures.
- Developed and maintained comprehensive documentation for data processes and systems, supporting audit and compliance efforts.
- Contributed to team training and development, sharing knowledge and best practices in Big Data technologies.
- Focused on continuous improvement and adaptation to new technologies, enhancing data solutions and operational practices.

Environment: Apache Hadoop, Apache Oozie, HDFS, MapReduce, Apache Hive, Apache Pig, Python, Git, Apache Kafka, SQL.

Education:

- masters in computer science from lindsey wilson college