Shiva Pericherla Data Engineer

✉ shiva@mymailsbox.net   📞 9379516640   📍 Dayton, Ohio   in Linkedin

Profile

Data Engineer with 3 years of experience specializing in data pipeline design, development, and optimization. Proficient in ETL processes, SQL, Python, Apache Hadoop, and Apache Spark. Skilled in collaborating with cross-functional teams to gather business requirements and translate them into technical solutions.

Skills

Languages (Python, R, SQL, NoSQL)

Data Processing & Streaming (PySpark, Apache Kafka, Pandas, NumPy, Scikit-learn, Matplotlib, Airflow, Snowflake, ETL, MS Excel, Data Flow, Hadoop, Databricks)

Frameworks & Tools (Power BI, Tableau, Flask, Django, MS SQL, MySQL, PostgreSQL, MongoDB, Git, GitHub, Docker, CI/CD pipelines, AWS, Azure Data Bricks, Jenkins, Statistics, Power BI, HTML, CSS, Agile, Google Analytics)

Other Technologies (AWS Redshift, SSIS, Azure Data Lake, Hadoop, Amazon Neptune, AWS Glue, Azure Synapse Analytics, Apache Airflow, Apache Spark, PySpark)

Professional Experience

Data Engineer, *Capgemini*   01/2023 – present │ Remote, USA
- Architected Apache Spark-based data pipelines using Python for real-time financial data processing, ensuring accuracy and reliability; integrated Talend ETL with SQL for seamless data modeling and integration.
- Integrated Apache Kafka for real-time data streaming and message queuing, while leveraging machine learning in Apache Spark pipelines for predictive analytics and fraud detection.
- Developed Google BigQuery analytics solutions tailored for 50+ analysts, employing Python and SQL for efficient ETL processes, data transformation, and loading, optimizing performance and resource utilization.
- Engineered interactive Tableau dashboards integrated with SQL, leveraging MS Excel for meticulous data analysis and validation, facilitating informed decision-making and actionable insights for stakeholders.
- Implemented a robust real-time fraud detection system utilizing Python, SQL, and cutting-edge data warehousing technologies, significantly reducing financial losses and bolstering customer trust and security measures.
- Demonstrated advanced proficiency in MS Excel for data validation and analysis, while leveraging SQL within data warehousing environments to enhance scalability, performance, and operational efficiency.

Data Engineer Intern, *Anblicks*   06/2022 – 04/2023 │ Remote, USA
- Designed and implemented end-to-end Python and SQL-based ETL pipelines for data extraction, transformation, and loading into AWS Redshift and Snowflake data warehouses, ensuring data quality and integrity.
- Engineered scalable data processing systems using Apache Spark and Hive to handle large volumes of data, optimizing system performance and resource utilization for enhanced efficiency.
- Collaborated with data analysts and business stakeholders to define data requirements and develop data-driven solutions aligned with strategic business objectives, fostering a culture of data-driven decision-making.
- Created interactive dashboards and reports using Tableau and Power BI, translating raw data into actionable insights and effectively communicating key information to stakeholders for informed decision-making.

Data Engineer, *Genpact*   11/2020 – 08/2021 │ Hyderabad, India
- Designed and Implemented high-throughput data ingestion processes, handling 2 million real-time events daily from diverse sources such as weblogs, user activity, sales transactions, and product catalog data.
- Architected and executed a scalable data storage solution using Amazon S3 and Apache Cassandra, seamlessly accommodating a 200% increase in data volume, ensuring optimal performance throughout the project lifecycle.
- Engineered data processing and transformation pipelines with Apache Spark, achieving a 95% data quality improvement by implementing cleansing and enrichment measures for incoming data.
- Integrated disparate data sources into a unified model using Talend ETL, breaking down data silos and enhancing accessibility for analytics, thereby fostering a more cohesive and comprehensive data environment.
- Managed AWS-based data pipeline and analytics platform, ensuring 99.9% uptime and achieving a 30% reduction in incident resolution time through robust monitoring with Grafana and Prometheus.

Education

Master of Science, *Indiana Wesleyan University*   01/2022 – 04/2023 │ Indiana, USA
Information Technology - Project Management

Bachelor of Technology, *Institute of Aeronautical Engineering*   06/2017 – 06/2021 │ Hyderabad, India
Mechanical Engineering