Sarath Varma Chitraju Data Engineer

schitraju@myworkmails.com 240-284-9977 Frederick, MD LinkedIn

Profile

An adept Data Engineer with a 6 year background in crafting, building, and upholding expandable data solutions. Possesses advanced proficiency in AWS Cloud, alongside expertise in data modeling, ETL processes, and data warehousing strategies. Exhibits robust programming acumen in Python and SQL, coupled with specialized skills in data manipulation, transformation, and optimization. Demonstrates a proven capability to efficiently collaborate with multifaceted teams and provide top-notch data solutions that align with business needs.

Skills

Languages (Python, R, SQL, NoSQL, Scala, HTML, CSS)

Data Processing & Streaming (PySpark, Apache Kafka, Pandas, NumPy, Scikit-Learn, Matplotlib, SpaCy, Airflow, Snowflake, T-SQL, PL-SQL, SSIS, SSRS, ETL, MS Excel, Azure SQL, Azure Data Factory (ADF), Google Analytics, Databricks, Hadoop, Synapse Analytics)

Database (MS SQL, MySQL, PostgreSQL, MongoDB, Dynamo DB)

Other Tools (Git, GitHub, Docker, CI/CD pipelines, Agile, Jenkins, Power BI, Tableau, IntelliJ, Machine Learning, KNN, Logistic Regression, Decision Trees, JIRA, Django, Flask)

Professional Experience

Data Engineer, *SiriusXM* 01/2023 – present | Remote, USA

- Designed and implemented ETL pipelines using Python, SQL, Apache Kafka, and Apache Airflow, integrating 8 data sources into AWS cloud services to enhance shipping and distribution demand forecasting accuracy.
- Utilized Databricks Unified Analytics Platform to harness the power of Apache Spark for scalable and efficient data processing within ETL workflows, enhancing overall pipeline performance and throughput.
- Employed Python libraries such as pandas and NumPy within Databricks notebooks for streamlined data manipulation and analysis, optimizing data processing speed and efficiency across the ETL lifecycle.
- Implemented Apache Airflow workflows to orchestrate complex data transformations and scheduling tasks, achieving 20% faster processing times for predictive modeling and analysis on AWS S3 and Amazon RDS.
- Collaborated closely with data scientists to develop and validate demand forecasting models, utilizing advanced statistical analysis, machine learning algorithms, and Spark-based computations within Databricks, leading to a 15% enhancement in supply chain logistics optimization and forecast accuracy.
- Maintained operational dashboards using Amazon QuickSight in Databricks, enabling stakeholders to make informed decisions and improving operational efficiency and customer satisfaction.
- Managed critical infrastructure components, including Apache Kafka clusters, AWS Glue, and Databricks clusters, to enable real-time data processing and scalable analytics aligned with business requirements.
- Ensured high data integrity, security, and compliance in Databricks environment with monitoring, maintenance, and documentation, bolstering trust in analytics-driven decision processes and achieving a 9.0 data integrity rating.

Data Engineer, *Genpact* 04/2018 – 06/2021 | Hyderabad, India

- Evaluated Apache Airflow DAGs to orchestrate tasks and workflows for data processing in AWS S3 buckets. Demonstrated expertise with diverse file formats like Parquet, ORC, JSON, CSV, etc. Utilized AWS Glue to ingest data into Amazon S3 from various sources.
- Devised and implemented a highly scalable and fault-tolerant data processing pipeline using Spark and Amazon Elastic Kubernetes Service (EKS), resulting in a 30% reduction in data processing time and enabling the organization to handle a 50% increase in data volume.
- Created complex ETL (Extract, Transform, Load) jobs to extract data from various sources such as Amazon RDS, PostgreSQL, and loaded them into target databases like Amazon Redshift. Proficient in writing SQL queries, subqueries, and joins to generate Stored Procedures.
- Developed Amazon QuickSight reports for sales, finance, and marketing teams to support predictive analysis. Engaged in performance tuning and optimization of long-running Spark jobs and queries (Amazon Redshift/SQL).
- Transferred NoSQL databases to Amazon DynamoDB or Amazon DocumentDB, enabling the handling of more than 2.5x spikes in transaction volume without extensive pre-planning or downtime, maintaining near-100% uptime.
- Optimized ETL processes to reduce processing times and improve scalability using techniques like parallel processing, distributed computing, and caching within the AWS ecosystem.

Data Engineer, *Tech Mahindra* 10/2016 – 03/2018 | Hyderabad, India

- Designed and orchestrated a sophisticated ETL pipeline architecture tailored for supply chain management, seamlessly integrating Apache Airflow, Python, Apache Kafka, and Azure Data Factory. This optimized workflow scalability and efficiency while overseeing 20 distinct data sources.
- Implemented rigorous data quality checks and monitoring mechanisms leveraging Apache NiFi and Azure Monitor. This ensured the integrity and availability of supply chain data, aligning with industry standards and processing an impressive 60,000 records per hour.
- Leveraged Azure Databricks for real-time data processing and analytics within the supply chain domain. This empowered agile decision-making and strategic planning, thereby enhancing logistics and inventory management. This included handling data streams of 7 TB per week.
- Automated data lineage tracking using Apache Atlas and Azure Data Catalog, guaranteeing regulatory compliance and enhancing data governance transparency. This meticulous tracking extended to overseeing 2400+ transformations.
- Containerized ETL workflows with Docker and Kubernetes, enabling seamless deployment and management across diverse supply chain environments. This included on-premises and multi-cloud setups, deploying an impressive 40 containers per day.
- Employed Azure Monitor and Prometheus for monitoring purposes, ensuring a remarkable 9.0 uptime. This proactive approach identified and resolved supply chain infrastructure issues, maintaining continuous operations and minimizing disruptions.

Education

Masters in computer, *University of Central Missouri* 08/2021 – 12/2022 | Warrensburg, MO, USA
Information systems & Information Technology

Certificates

AWS Certified Developer - Associate