

Steve Liu

Email: steveliu694@gmail.com

Cell: (734) 274-9251

Summary:

- Experienced Data Engineer specialized in big data infrastructure/pipelines/ETL;
- Strong knowledge in Cloud technology, expert in cloud-based Data Lake, Data Warehouse, and Databases;
- Solid understanding of SQL and NoSQL, database design, data governance, data security, and data models;
- Excellent in Python developing and Python-based data tools.

Technical Skills:

Python | Java | SQL | NoSQL | Databricks | Snowflake | DBT | Apache Spark | Pandas | Numpy | PowerBI
REST API | Agile | Talend | CI/CD | Machine Learning | Unit test | Automated testing | ETL/ELT | Scala
AWS S3 | Lambda | API Gateway | Sagemaker | DynamoDB | Step function | Redshift | RDS | Glue | EMR | EC2
Bash | GitHub | Gitlab | Jira | Jenkins | CSV | Parquet | XML | Jupiter notebook | Docker | Kubernetes

Certifications:

AWS Certified Cloud Practitioner, AWS Certified Developer Associate

Professional Experience:

DFLL Consulting, Data Engineer, 07/2023 – Present

- Responsible for architecting, developing, and maintaining scalable and robust high performance data pipelines and infrastructure to integrate diverse data sources and support client's business needs.
- Building and continuously enhancing data pipelines based on a variety of cloud-based services, such as AWS Lambda function, Managed Workflows for Apache Airflow, S3, Redshift, EMR, Glue, DynamoDB, RDS, etc.
- Integrating AWS native services and third-party data tools such Databricks, DBT, Snowflake to cater to a variety of complex use cases.
- Designing and implementing data security features to protect sensitive data and ensure compliance.
- Using low-code tools such as Talend Open Studio for data integration and ETL, importing the raw data from different sources and transforming them to the desired format.
- Designing, implementing, and testing Python, PySpark, and Scala codes and SQL queries as parts of data infrastructure; troubleshooting and resolving production issues in data infrastructure.
- Working with event-driven architecture leveraging Kafka, building and maintaining Kafka connectors to read data from sources to destinations.
- Working closely with business teams and cross-functional teams to understand constantly changing requirements and translate complicated business requirements into technical strategies and data products.
- Writing complex and high-performance SQL queries to generate datasets for business analytics and data QA purposes.
- Leading fellow data engineers on various tech initiatives, training and mentoring junior engineers, giving knowledge transfer on projects and technologies, presenting results to both technical and non-technical audiences.

Amcor PLC, Data Engineer, 09/2019 – 07/2023

- Designed and developed internal Data lake within AWS using S3 as storage and integrated services including Lambda, DynamoDB, RDS, Redshift, API Gateway, resulting in the optimization of data pipelines and improved performance.
- Analyzed complex data sets from different sources and modeled them to meet business requirements; worked with structured and semi-structured data in the format of CSV, Parquet, JSON, and XML.
- Built Data Pipelines using AWS S3, Redshift, Glue, Lambda, and Step Functions; performed ad-hoc data analytics tasks for various business requirements.
- Developed, integrated, and tested Lambda functions for various AWS integrations/architectures.
- Designed, developed and maintained data pipelines using Python and Spark for complex data sets.

- Worked in a cross functional team to support product development, supply chain, and marketing; designed and implemented Python automation tools to greatly speed up complex big data workflows.
- Developed automated testing tools for various projects; tested and deployed containerized applications using Docker and Kubernetes; built CI/CD pipelines with Jenkins to automate building and testing of codes.
- Followed Agile methodology in project planning and developing; communicated with stakeholders to understand requirements and get feedback.
- Performed version control and coding collaboration with Git tools; created and maintained documentations for projects; trained and mentored interns and junior/new team members.

NSK Corporation, Data Engineer, 03/2018 – 09/2019

- Responsible for analytical processing (OLAP) of customers/transactions big data to give data-driven insights on product sales trends of the automotive/machinery product provider.
- Designed and maintained relational database, data warehouse, and ETL pipeline on premises and on cloud (AWS), worked with large multidimensional data from different sources.
- Developed Python applications to automate data processing and product development workflow, greatly accelerated development of products.