

Job duties

- Lead the design, development, and optimization of data ingestion pipelines for on-premise and cloud-based data lake solutions.
- Drive initiatives to migrate and integrate data from various structured and unstructured sources into Snowflake, Databricks, and AWS platforms.
- Collaborate with cross-functional teams to implement ETL frameworks using Spark and AWS services (EMR, StepFunctions, Lambda, S3, SNS, SQS) to support payments and merchant analytics.
- Architect and build scalable data sharing capabilities using Snowflake and data lake solutions.
- Automate the generation of data ingestion pipelines to improve efficiency and maintainability.
- Design and develop dashboards for cluster monitoring, resource utilization, and cost monitoring in Snowflake and Databricks environments to provide insights for performance optimization and capacity planning.
- Collaborate with data scientists and analysts to integrate machine learning models and advanced analytics capabilities into data processing workflows on Databricks clusters.
- Architect and implement disaster recovery strategies to ensure high availability and business continuity for critical data stored in Snowflake, Databricks, and AWS services.
- Lead cost optimization initiatives by analyzing cloud usage patterns, identifying opportunities for resource optimization, and implementing cost-saving measures across Snowflake, Databricks, and AWS environments.
- Design and implement CI/CD pipelines using IaC to automate the deployment of services and infrastructure components on AWS.
- Mentor junior team members and provide technical leadership in data engineering best practices, software design principles, and cloud architecture.