

PROFESSIONAL SUMMARY :

Data Engineer with Overall 3+ years of experience in Data Engineering, Data Warehousing, ETL/ELT, and Business Intelligence. Expertise in AWS, Azure, and GCP, with strong programming and scripting skills in Python, SQL, and Scala. Skilled in designing and optimizing end-to-end data pipelines, real-time streaming architectures, and big data analytics platforms. Extensive experience in Azure Data Factory, Snowflake, Redshift, Hive, Spark, Airflow, and Kafka, implementing highly scalable and resilient solutions for enterprise data processing. Strong background in data modelling, architecture, CI/CD automation, and DevOps practices. Proven ability to manage structured, semi-structured, and unstructured data across cloud and on-premise environments to drive business intelligence and machine learning applications.

TECHNICAL SKILLS

Big Data Technologies: Hadoop, HDFS, Hive, Spark, Airflow, Snowflake, Kafka, MapReduce, Flink, Presto

Cloud Platforms: AWS (Glue, Redshift, Lambda, S3, EMR, Athena, EC2, RDS, VPC, IAM, CloudWatch), Azure (ADF, Synapse, Data Lake, Databricks, Cosmos DB, HDInsight, SQL DW), GCP (BigQuery, DataProc, Cloud Functions)

Programming Languages: Python, SQL, Scala, Java, C++

ETL & Reporting Tools: Apache Airflow, AWS Glue, Talend, Tableau, Power BI, SSIS

Databases: Oracle, SQL Server, MySQL, PostgreSQL, MongoDB, Cassandra, HBase, DynamoDB

DevOps & CI/CD: Jenkins, GitHub, GitLab CI, Docker, Kubernetes, Terraform, Ansible, Bamboo

Machine Learning: PySpark, Pandas, NumPy, Scikit-Learn, MLlib, TensorFlow

Operating Systems: Linux, UNIX, Windows

PROFESSIONAL EXPERIENCE :

Member 1st, PA(Remote)

Data Engineer | Sep 2024 – Present (Contract)

- Designed and implemented end-to-end scalable ETL pipelines to process large-scale financial datasets, integrating Azure Data Factory, Databricks, and Snowflake for advanced analytics.
- Developed Azure Synapse and ADLS-based solutions to optimize structured and unstructured data processing, improving data pipeline efficiency and reducing operational costs.
- Implemented real-time streaming architectures using Apache Kafka, Spark Streaming, and Azure Stream Analytics, ensuring low-latency event-driven data processing.
- Led the migration of on-prem SQL Server datasets to Azure Synapse Analytics and Azure SQL DB, enabling cloud-based high-performance querying and reporting.
- Created an intelligent data foundation accelerator using Azure

Functions and Databricks, streamlining data ingestion, transformation, and orchestration processes.

- Designed feature engineering strategies and Python-based ML pipelines in Azure ML Studio, enhancing predictive analytics for fraud detection and customer insights.
- Developed optimized Redshift and Snowflake data models with partitioning and clustering techniques, improving query performance and analytical workloads.
- Automated ETL pipeline scheduling and monitoring using Apache Airflow, improving data pipeline reliability and reducing manual intervention.

Environment: Azure, Python, Spark, Kafka, Snowflake, SQL Server, ADF, Databricks, Airflow, Power BI, Cosmos DB, PolyBase, Azure Functions

Nationwide, Columbus, OH(Remote)

Data Engineer | Jan 2024 – Aug 2024 (Contract)

- Designed and built scalable ETL workflows using AWS Glue and PySpark to ingest, transform, and process massive datasets from various sources.
- Developed AWS DMS-based solutions for real-time data replication from MariaDB and MySQL into Redshift, ensuring data consistency across environments.
- Integrated Amazon Bedrock Gen AI models for intelligent data enrichment, automating form completion and reducing manual data entry efforts.
- Designed AWS Lambda-based ingestion frameworks to fetch data from Salesforce and other third-party APIs, implementing JWT token-based authentication.
- Implemented optimized Redshift stored procedures and workload management strategies to enhance query execution and performance.
- Developed and governed AWS Lake Formation policies to manage data security, access control, and compliance with enterprise data governance standards.
- Created advanced analytical dashboards using Quick Sight and Redshift Views, enabling business teams to derive meaningful insights from processed data.
- Automated AWS Glue and EMR workflow monitoring using Lambda, SNS, and SQS, proactively identifying and resolving pipeline failures.

Environment: AWS Glue, Redshift, Lambda, S3, EMR, Athena, Python, Snowflake, Tableau, Kinesis, QuickSight, LakeFormation, Iceberg, DynamoDB

Smartron, India

Data Engineer | Feb 2021 – Dec 2022

- Developed ETL workflows using Azure Data Factory and Databricks for seamless cloud-based data transformations and analytics.
- Designed Spark Streaming applications to process real-time Kafka streams, ensuring efficient event processing and alerting mechanisms.

- Built a Redshift Spectrum-based query framework to enable cost-effective analytics on semi-structured and structured data stored in S3.
- Automated AWS infrastructure provisioning with Terraform, streamlining resource allocation and reducing setup times.
- Developed ML models for predictive analytics using PySpark, improving decision-making and business forecasting accuracy.
- Designed and implemented real-time Kinesis-based data ingestion pipelines, ensuring low-latency data availability for analytics teams.
- Migrated SQL Server data warehouses to Hadoop and implemented optimized Hive and Spark queries for performance improvements.
- Automated CI/CD deployment pipelines using Jenkins and GitLab CI to ensure reliable and repeatable infrastructure provisioning.

Environment: Azure, Hadoop, Sqoop, Hive, Spark, Python, GCP, BigQuery, Terraform, Kafka, Jenkins, Redshift Spectrum, Kinesis

Education :

Masters: University: Pace university, New York – 2024

Bachelors: University: Kodad institute of technology and science, India – 2022