**Name**

**Email Address**

**Phone Number**

### **DATA ENGINEER**

**Professional Summary**

* Data Engineer with **10 years of professional experience** designing, developing, and deploying scalable, end-to-end data pipelines and analytics solutions across multiple enterprise environments.
* Architected robust ETL frameworks using Apache Spark, Python, and SQL for high-volume data processing and analytics readiness.
* Automated batch and real-time data workflows with Airflow and Kafka, ensuring reliability, consistency, and operational efficiency.
* Optimized cloud data warehouses and data lakes on AWS, GCP, and Azure for performance, storage efficiency, and query optimization.
* Developed modular and reusable Python and Scala scripts for standardized data transformations and enrichment across multiple pipelines.
* Migrated legacy on-premise data systems to cloud-native architectures while ensuring data integrity, minimal downtime, and operational continuity.
* Implemented monitoring and alerting solutions with Grafana, Prometheus, and Stackdriver to proactively manage data pipelines.
* Integrated CI/CD workflows using Jenkins, GitLab CI, and Terraform to automate deployment and version control of ETL and infrastructure pipelines.
* Built event-driven streaming pipelines using Kafka, Kinesis, and Spark Streaming for real-time analytics and operational reporting.
* Enforced data quality standards with automated validation, anomaly detection, and reconciliation scripts to ensure high data reliability.
* Collaborated with cross-functional teams to translate business requirements into scalable, maintainable, and secure data solutions.
* Leveraged containerization and orchestration tools like Docker and Kubernetes for fault-tolerant, scalable data processing workloads.
* Developed and maintained comprehensive metadata, lineage, and audit tracking to support governance and compliance requirements.
* Evaluated emerging technologies and implemented solutions to optimize data engineering processes, reduce manual efforts, and improve operational efficiency.
* Mentored and guided junior engineers in best practices, code quality, and modern data engineering methodologies.

## **Technical Skills**

|  |  |
| --- | --- |
| **Category** | **Skills** |
| **Programming & Scripting** | Python (Modular scripting, Data Validation), Scala, SQL (Query Optimization, DDL/DML), Bash/Shell |
| **Big Data & Processing** | Apache Spark (Batch/Streaming), Databricks, Apache Beam, Dataflow, Spark Streaming |
| **Cloud Platforms** | AWS (S3, Redshift, Lambda, Kinesis), GCP (BigQuery, Pub/Sub, Dataflow), Azure (Azure Data Lake) |
| **Data Warehousing & DBs** | Snowflake, AWS Redshift, Google BigQuery, SQL Server, S3 (Data Lake), Parquet, JSON, CSV |
| **Orchestration & Messaging** | Apache Airflow (DAGs, Monitoring), Apache Kafka, GCP Pub/Sub, AWS Kinesis |
| **DevOps & Infrastructure** | Terraform (Infrastructure as Code), Docker (Containerization), Kubernetes (Orchestration) |
| **CI/CD & Version Control** | Jenkins, GitLab CI/CD, Git |
| **Data Modeling & Tools** | DBT (Data Build Tool), ETL/ELT Frameworks, Data Modeling (Dimensional/Star Schema) |
| **Monitoring & BI** | Grafana, Prometheus, Stackdriver, Tableau, Power BI, Looker |

### **PROFESSIONAL EXPERIENCE**

**Client: Chewy, Dania Beach, FL**

**Role: Principal Data Engineer  
Environment:** AWS, Snowflake, Python, Spark, Airflow, Kafka, DBT, Terraform, Jenkins, Grafana, Kubernetes

**Responsibilities:**

* Designed and implemented highly scalable, automated ETL pipelines for ingestion, transformation, and storage of enterprise datasets.
* Architected Snowflake-based data warehouses optimized for query performance, storage efficiency, and analytics readiness.
* Built real-time streaming pipelines using Kafka and Spark Streaming to support operational and analytics workloads.
* Developed reusable Python and DBT modules for standardized data transformations and enrichment.
* Automated CI/CD pipelines using Jenkins and Terraform for seamless deployment of ETL workflows and cloud infrastructure.
* Containerized data processing workflows with Docker and orchestrated with Kubernetes for fault-tolerant and scalable operations.
* Monitored pipeline performance, executed proactive troubleshooting, and implemented alerting using Grafana and Prometheus.
* Designed data validation and reconciliation frameworks to maintain high data quality and reliability.
* Collaborated with analytics teams to ensure data pipelines meet business requirements and operational needs.
* Optimized Snowflake schemas, partitioning, and query performance to enhance data access and reporting efficiency.
* Implemented secure data access, encryption, and governance policies in line with organizational standards.
* Developed event-driven workflows for ingestion, transformation, and reporting of high-volume datasets.
* Conducted technical reviews, mentoring, and knowledge-sharing sessions to promote best practices across the engineering team.
* Evaluated emerging data engineering technologies and integrated relevant solutions to improve pipeline efficiency and scalability.

**Client : UPS Capital, Atlanta, GA  
Role: Data Engineering Consultant  
Environment:** GCP (BigQuery, Dataflow, Pub/Sub), Python, SQL, Apache Beam, Airflow, Docker, Kubernetes, Looker

**Responsibilities:**

* Developed end-to-end ETL pipelines using Dataflow and Apache Beam for high-volume structured and semi-structured datasets.
* Implemented automated scheduling and monitoring of data workflows with Airflow to ensure reliability and consistency.
* Built streaming pipelines using Pub/Sub to process real-time event data and integrate it with analytics platforms.
* Designed and deployed BigQuery data models for reporting, analytics, and machine learning consumption.
* Containerized ETL jobs using Docker and orchestrated workloads using Kubernetes for scalability and fault tolerance.
* Developed Python utilities for reusable data transformations, validations, and enrichment across pipelines.
* Implemented monitoring dashboards and alerting using Grafana, Stackdriver, and Prometheus to maintain operational efficiency.
* Optimized queries, data partitioning, and processing workflows for faster execution and reduced costs.
* Ensured robust data quality by implementing validation, reconciliation, and anomaly detection frameworks.
* Collaborated with cross-functional teams to understand requirements and deliver scalable and maintainable data solutions.
* Automated CI/CD deployment pipelines using GitLab CI/CD and Terraform for consistent and reliable releases.
* Conducted root cause analysis for pipeline failures and implemented long-term corrective actions.
* Provided technical guidance on data governance, security, and operational best practices.
* Mentored junior engineers, reviewing code and promoting adherence to modern engineering standards.

**Client: Giant Eagle, Pittsburg , PA**

**Role: Lead Data Engineer  
Environment:** AWS, S3, Redshift, Spark, Python, Airflow, Kafka, Terraform, Docker, Tableau

**Responsibilities:**

* Architected and deployed cloud-based data lakes to consolidate enterprise datasets for analytics and operational reporting.
* Designed and implemented event-driven streaming pipelines using Kafka and Spark Streaming to process real-time data streams.
* Automated ETL workflows with Airflow, implementing monitoring, error handling, and alerting mechanisms.
* Optimized Spark jobs and Redshift schemas to improve processing efficiency and reduce execution time.
* Developed reusable Python modules for ETL operations, ensuring consistency and maintainability across pipelines.
* Implemented Terraform scripts for infrastructure provisioning and version-controlled deployments.
* Containerized ETL services with Docker for scalable and repeatable deployments across environments.
* Integrated BI tools like Tableau with cloud data warehouses for interactive reporting and insights generation.
* Developed automated anomaly detection and reconciliation processes to maintain data integrity.
* Collaborated with analytics and data science teams to implement pipelines supporting model training and analytics workflows.
* Conducted data profiling, lineage analysis, and metadata management to improve governance and traceability.
* Monitored workflow execution and system performance proactively, resolving bottlenecks efficiently.
* Migrated legacy batch processes to modern cloud pipelines to reduce operational overhead and enhance scalability.
* Mentored team members and conducted knowledge sharing sessions on modern data engineering practices.

**Client: IBing Software Solutions Private Limited Hyd India**

**Role: Senior Data Engineer  
Environment:** Azure, Databricks, Python, Scala, Spark, Airflow, SQL Server, GitLab CI, Power BI

**Responsibilities:**

* Designed, built, and optimized ETL pipelines for ingestion and transformation of large-scale enterprise datasets from diverse sources.
* Developed Spark-based batch and streaming workflows to process structured and semi-structured data efficiently.
* Automated pipeline scheduling and orchestration using Airflow DAGs with retry and alert mechanisms.
* Built modular Python and Scala scripts for reusable data transformation logic across pipelines.
* Integrated Azure Data Lake with Databricks for centralized storage, processing, and analytics readiness.
* Implemented CI/CD pipelines using GitLab CI for automated deployment of ETL jobs and infrastructure scripts.
* Optimized SQL queries, Spark transformations, and data models for improved processing efficiency and query performance.
* Developed data validation frameworks to enforce quality, detect anomalies, and reconcile data across sources.
* Implemented logging, monitoring, and alerting using Grafana and Stackdriver to ensure pipeline reliability.
* Collaborated with business and analytics teams to understand requirements and translate them into scalable solutions.
* Conducted root cause analysis for data discrepancies and implemented long-term corrective solutions.
* Implemented secure data access policies and encryption to maintain compliance with organizational standards.
* Migrated on-premise SQL Server workloads to cloud-based platforms with minimal downtime and risk.
* Provided technical guidance and mentoring to junior engineers on ETL design, coding standards, and pipeline optimization.

**Client: Amigos Software Solutions, Hyd, India  
Role: Data Engineer  
Environment:** AWS (S3, Redshift, Lambda), Python, SQL, Airflow, Spark, Docker, Jenkins, Git, Tableau

**Responsibilities:**

* Designed, developed, and maintained ETL pipelines to ingest, transform, and load large volumes of structured and semi-structured data across multiple sources.
* Built automated data validation and reconciliation scripts in Python to ensure accuracy, consistency, and reliability of enterprise datasets.
* Architected cloud-based Redshift data warehouses for centralized storage, query optimization, and reporting efficiency.
* Developed Spark batch processing jobs to handle high-volume JSON, CSV, and Parquet datasets efficiently.
* Implemented Airflow DAGs for scheduling complex workflows with dependency management and failure recovery.
* Containerized ETL workflows with Docker for reproducible deployments across development, QA, and production environments.
* Integrated CI/CD pipelines using Jenkins and Git to automate ETL job deployment and version control.
* Created dashboards in Tableau to visualize data trends, operational metrics, and pipeline health for stakeholders.
* Optimized SQL queries and Spark transformations for faster processing and reduced resource usage.
* Migrated legacy ETL processes to cloud-native workflows, reducing manual intervention and operational overhead.
* Monitored ETL pipelines using Airflow and Grafana dashboards, identifying and resolving bottlenecks proactively.
* Developed reusable Python modules for standard ETL operations to ensure maintainability and scalability.
* Collaborated with analytics teams to ensure data structures met downstream reporting and business requirements.
* Conducted metadata management and lineage tracking to improve data governance and traceability.