**SR. DATA ENGINEER**

**NAVYA SREERAMA  
+1 712-724-8014**

**navya.sreeramasam@gmail.com**

**Career Objective:**

Results-driven and technically skilled Data Engineer with over 10 years of experience in designing, developing, and managing scalable, high-performance data infrastructures across healthcare, banking, and financial services domains. Committed to leveraging cloud-native platforms, distributed data systems, and modern data engineering practices to deliver reliable, secure, and real-time data solutions that empower business intelligence and advanced analytics. Eager to contribute in a collaborative environment where innovation, data accuracy, and performance matter.

**Professional Summary:**

* Accomplished Data Engineering professional with over 10 years of hands-on experience delivering scalable and automated data solutions in enterprise-grade environments.
* Proven expertise in building complex ETL and ELT pipelines using tools like Apache Spark, AWS Glue, and Talend, handling multi-terabyte datasets across structured and semi-structured formats.
* Highly skilled in designing cloud-native data lakes and warehouses using AWS (S3, Redshift), GCP (BigQuery), and Azure Data Lake for optimized analytics and cost efficiency.
* Deep understanding of data architecture, dimensional modeling, and schema design (star, snowflake, normalized/denormalized) for building robust data marts.
* Expertise in batch and streaming data processing using Kafka, Kinesis, and Spark Streaming for real-time data workflows and business alerts.
* Developed and managed enterprise-scale data pipelines that processed millions of daily transactions in industries such as insurance, banking, and asset management.
* Hands-on with data governance, encryption, masking, and PII protection, ensuring regulatory compliance (HIPAA, GDPR, SOX) and end-to-end data lineage.
* Worked closely with DevOps and SRE teams to integrate CI/CD and containerized deployment workflows using Docker, Jenkins, and Kubernetes.
* Adept at optimizing SQL queries, data warehouse tuning, and query performance troubleshooting in Redshift, Snowflake, and BigQuery.
* Delivered robust metadata-driven pipelines, integrating logging, monitoring, alerting, and auto-healing components using tools like Airflow, CloudWatch, and ELK stack.
* Extensive experience building self-service reporting systems and collaborating with BI teams using Tableau, Power BI, and Looker.
* Strong programming background in Python, Scala, and SQL, applying advanced scripting to enable data transformations, automation, and data quality checks.
* Led migration of legacy ETL and warehouse systems to cloud platforms, reducing operational costs and increasing system reliability.
* Designed multi-source data integrations, ingesting and transforming data from REST APIs, SFTP, RDBMS, and SaaS applications like Salesforce and Workday.
* Instrumental in defining data security models, establishing access control policies via IAM, role-based permissions, and encryption at rest/in transit.
* Created and maintained technical documentation, data dictionaries, and lineage maps, facilitating team onboarding and regulatory audit trails.
* Committed to continuous learning, actively exploring evolving data tools (Delta Lake, dbt, Iceberg) and industry best practices for high-impact delivery.

**Technical Skills:**

|  |  |
| --- | --- |
| **Category** | **Tools** |
| **Languages:** | Python, Java, Scala, SQL, Shell |
| **Big Data:** | Hadoop, Spark, Hive, Flink, Kafka, HBase |
| **Cloud Platforms:** | AWS (S3, Glue, Redshift, Lambda, EMR), GCP (BigQuery, GCS), Azure (Data Lake, Synapse, Blob Storage) |
| **ETL Tools:** | Talend, AWS Glue, SSIS, Apache NiFi, Informatica |
| **Data Warehousing:** | Snowflake, Redshift, BigQuery, SQL Server, Teradata |
| **Data Orchestration:** | Apache Airflow, Oozie, Luigi |
| **Streaming:** | Kafka, AWS Kinesis, Spark Streaming |
| **APIs & Integration:** | REST, JSON, XML, Postman, Python Requests |
| **BI & Visualization:** | Tableau, Power BI, Looker |
| **Version Control & CI/CD:** | Git, Jenkins, GitLab CI/CD, Docker, Kubernetes |
| **Monitoring & Logging:** | CloudWatch, ELK Stack, Prometheus, Grafana |
| **Security:** | IAM, KMS, Data Encryption, RBAC, HIPAA, GDPR compliance |
| **Languages:** | Python, Java, Scala, SQL, Shell |
| **Big Data:** | Hadoop, Spark, Hive, Flink, Kafka, HBase |
| **Cloud Platforms:** | AWS (S3, Glue, Redshift, Lambda, EMR), GCP (BigQuery, GCS), Azure (Data Lake, Synapse, Blob Storage) |
| **ETL Tools:** | Talend, AWS Glue, SSIS, Apache NiFi, Informatica |
| **Data Warehousing:** | Snowflake, Redshift, BigQuery, SQL Server, Teradata |
| **Data Orchestration:** | Apache Airflow, Oozie, Luigi |
| **Streaming:** | Kafka, AWS Kinesis, Spark Streaming |
| **APIs & Integration:** | REST, JSON, XML, Postman, Python Requests |
| **BI & Visualization:** | Tableau, Power BI, Looker |
| **Version Control & CI/CD**: | Git, Jenkins, GitLab CI/CD, Docker, Kubernetes |
| **Monitoring & Logging**: | CloudWatch, ELK Stack, Prometheus, Grafana |
| **Security:** | IAM, KMS, Data Encryption, RBAC, HIPAA, GDPR compliance |

**Professional Experience:**

**Sr. Data Engineer | BCBS | VA**  **May 2023 – Till Date**

**Responsibilities:**

* Designed and implemented cloud-native, event-driven data pipelines in AWS Glue and Spark, processing healthcare claims and eligibility data at scale.
* Integrated multiple data feeds (HL7, EDI 837, X12) into a centralized data lake on AWS S3, enabling downstream BI and actuarial analysis.
* Implemented data encryption, access policies (IAM), and row-level security to safeguard PHI and ensure HIPAA compliance.
* Developed CDC-based ETL workflows using AWS DMS and Lambda, maintaining data freshness and low-latency ingestion.
* Worked with Snowflake to design data marts and optimized views for care management and utilization review teams.
* Created metadata-driven orchestration in Apache Airflow, allowing reusable DAGs and dynamic task branching.
* Enabled real-time claims adjudication alerts via Kafka streams integrated with Spark Structured Streaming.
* Built data profiling and quality frameworks using Deequ and PyDeequ, ensuring conformance to validation rules.
* Tuned SQL queries in Snowflake, reducing report generation time and eliminating performance bottlenecks.
* Automated data reconciliations between source (Oracle) and target (S3/Redshift) using custom Python scripts.
* Configured S3 lifecycle policies to archive processed data, optimizing storage cost while maintaining compliance.
* Led weekly stakeholder syncs with the Data Governance and Security teams to review lineage and audit controls.
* Conducted DR drills and backup validation using AWS Backup and CloudFormation scripts.
* Delivered Tableau dashboards using clean data sets for executive-level healthcare utilization insights.
* Integrated FHIR APIs and internal clinical systems into the data lake for population health analytics.
* Created Lambda-based alerting system for pipeline failures and anomalies with push notifications.
* Used AWS Glue Catalog and Lake Formation for centralized schema and permission management.
* Managed data onboarding for new providers with schema evolution support and automated validations.
* Created detailed SOPs, pipeline diagrams, and lineage maps for knowledge transfer and compliance audits.

**Environment:** AWS Glue, Spark, Redshift, Airflow, Snowflake, Kafka, Python, S3, Tableau, Lambda, Deequ, HL7, EDI, Jira, Git.

**Sr. Data Engineer |BMO | NJ March 20 – April 22**

**Responsibilities:**

* Developed enterprise-grade data pipelines in Apache Spark and Scala, processing credit, deposits, and risk data from legacy mainframe and modern banking platforms.
* Led data warehouse modernization by migrating batch ETL workflows from SSIS to Google Cloud BigQuery with dbt for modeling.
* Built real-time fraud detection streams using Kafka, enabling downstream applications to monitor suspicious activity instantly.
* Ingested data via REST and SOAP APIs from credit scoring systems (FICO, Experian) into the central data lake.
* Built robust reconciliation logic in Python to match GL data against operational systems and flag discrepancies.
* Implemented multi-layered security using VPC SC, encryption with KMS, and access control with service accounts and IAM roles.
* Reduced daily ETL execution time by refactoring Spark jobs with broadcast joins and partition pruning.
* Created modular and reusable DAGs in Airflow, reducing ETL development cycles and increasing job transparency.
* Configured BigQuery partitioning and clustering, optimizing query performance and cost for regulatory reports.
* Worked with Tableau and Looker to create interactive dashboards for finance and operational risk stakeholders.
* Established a fully automated testing framework with Pytest for validating transformation logic in pre-production.
* Migrated 30+ workflows to CI/CD pipelines using GitLab and Terraform, enabling infrastructure as code (IaC).
* Designed archival strategies for long-term data retention using GCS and automated archival routines.
* Developed dbt models with snapshots and incremental strategies, improving auditability and traceability.
* Enabled lineage tracking via integration with DataHub and Airflow metadata APIs.
* Integrated metrics logging into Prometheus and Grafana for performance monitoring and pipeline visibility.
* Orchestrated source-to-target mappings and created high-level designs (HLDs) for internal architecture reviews.
* Led data engineering workshops for upskilling business teams on cloud tools and analytics readiness.
* Contributed to regulatory reporting projects such as Basel III and CCAR using cleansed, governed datasets.

**Environment:** Apache Spark, Scala, BigQuery, dbt, Kafka, GCS, Airflow, Python, GitLab, Terraform, Tableau, Prometheus, Looker, REST APIs.

**Data Engineer | Aflac Asset Management | NY Jan 2018 – Feb 2020**

**Responsibilities:**

* Developed robust ETL frameworks in Python and Spark to process daily positions, trades, and portfolio data from Bloomberg, FactSet, and internal trading systems.
* Designed and maintained a secure AWS S3-based data lake to store both structured and unstructured asset management data, enabling scalable analytics for investment research.
* Created custom ingestion modules for pulling market data and pricing feeds via REST APIs and FTP from third-party vendors like ICE, Morningstar, and Reuters.
* Built reusable transformation pipelines in AWS Glue, standardizing asset classification, performance metrics, and risk attribution data across portfolios.
* Automated reconciliation between trade blotter systems and custodial data using SQL Server and Python scripts, reducing errors in reporting and audit delays.
* Worked on ingesting fixed-income instrument data from Bloomberg and mapping it to internal classifications using fuzzy logic and rule-based transformations.
* Built dimension and fact tables in Redshift, supporting OLAP-style querying for asset allocation and portfolio exposure analysis.
* Tuned long-running queries and optimized joins, partitioning, and vacuuming strategies in Redshift to enhance reporting speed and reduce storage cost.
* Led the data lineage and traceability effort, building audit trails from data sources to BI dashboards for regulatory and internal compliance.
* Delivered Power BI dashboards for front-office portfolio managers, showing real-time views of NAV trends, cash flow projections, and risk exposure.
* Designed and implemented streaming pipelines using AWS Kinesis for market ticker data, delivering sub-second refresh times on dashboards.
* Integrated data pipelines with enterprise data catalog tools to ensure consistent metadata tagging and schema registration.
* Applied role-based access control (RBAC) and KMS encryption to control access to sensitive investment and client data across cloud services.
* Established CI/CD pipelines via Jenkins and Git, with unit tests for transformation logic and alerting on pipeline failures.
* Conducted quarterly data quality reviews with asset managers, addressing anomalies using automated profiling tools and rule-based alerts.
* Wrote SQL-based macros to calculate asset performance, net flows, and benchmark comparisons in Redshift and Python notebooks.
* Authored documentation for technical workflows and business data mappings, supporting onboarding and handovers to other teams.

**Environment:** AWS Glue, Spark, Redshift, Python, SQL Server, Power BI, Jenkins, Git, Kinesis, Docker, S3, Bloomberg APIs, REST APIs, Agile.

**Jr. Data Engineer |Micro land | India May 14 – Sep 17**

**Responsibilities:**

* Supported senior engineers in designing ETL jobs using SQL Server Integration Services (SSIS) to extract and process data from multiple legacy systems.
* Created scheduled data loads for operational reporting by writing T-SQL queries, stored procedures, and views in SQL Server.
* Helped develop Python scripts for automating daily, weekly, and monthly data pulls from internal flat files and external APIs.
* Maintained data flow from on-premise MySQL and Oracle databases to a centralized reporting system using SSIS packages.
* Created lookup tables, constraints, and indexes to improve query performance and ensure data integrity.
* Built basic dashboards and visualizations using Excel and Tableau to support internal operations reporting.
* Conducted manual and automated data quality checks for key metrics, such as sales performance and SLA compliance.
* Monitored job failures and prepared RCA (Root Cause Analysis) reports to aid in debugging and future prevention.
* Supported a migration project moving flat-file-based reporting to relational models in SQL Server.
* Wrote automation scripts in PowerShell and Python for file movement, cleanup, and data staging processes.
* Helped maintain version control of scripts and job configurations using TFS (Team Foundation Server) and Git.
* Gained hands-on exposure to basic concepts of data modeling, indexing, and performance tuning, laying the foundation for future growth.

**Environment:** SSIS, SQL Server, T-SQL, MySQL, Oracle, Python, PowerShell, Excel, Tableau, Git, TFS, Agile.