**SHASHANK TATIPALLI**

Data Engineering

**Phone:** 469-844-7612

**Email:** **shashankt0301@gmail.com**

**Linked IN:** [www.linkedin.com/in/shashank-t-](http://www.linkedin.com/in/shashank-t-)

**Visa Status: GC**

**PROFESSIONAL SUMMARY:**

 Results-driven Data Engineer with 10 years of experience in project development, implementation, deployment, and maintenance. Skilled in Azure Cloud services, including Azure Data Factory, HDInsight, Data Explorer, Event Hub, and Databricks, to build scalable and efficient data solutions. Proficient in handling diverse data serialization formats such as Parquet, Avro, ORC, and JSON to optimize storage and retrieval processes. Adept at designing and implementing robust ETL pipelines and big data architectures to support business intelligence and analytics initiatives. Seeking to leverage expertise in cloud-based data engineering to drive data-driven decision-making and innovation.

* Hand on experience working with Azure HD Insights, Spark and Databricks.
* Experience in importing and exporting data from RDBMS to Hadoop and HIVE using SQOOP.
* Extensive experience in working with HDFS, MapReduce, Pig, Hive, Sqoop, Flume, Oozie, Zookeeper, Kafka, Spark, HUE and Dremio.
* Detail-oriented Data Architect with a passion for clean data solutions, insightful visualizations, and scalable architecture.
* Skilled in data storytelling using Looker, Power BI, and Tableau.
* strong collaborator with product and engineering teams in fast-paced Agile environments
* Proficient in Reltio – Cloud MDM Platform, implementing data governance, entity resolution, and data mastering for enterprise-wide single source of truth.
* Developed Spark jobs using Scala/PySpark and Spark SQL for faster data processing.
* Highly capable of processing large sets of Structured, Semi-structured and Unstructured datasets supporting Big Data applications.
* Integrated Debezium connectors into data pipelines for low-latency data synchronization between source systems and Azure Synapse/Snowflake.
* Tuned and optimized Debezium performance for high-throughput replication and ensured data consistency across distributed systems.
* Extensively designed and developed Snowpipe, task, storeprocs and snowflake integration objects.
* Experience with Snowflake fine tuning snow-sql with cost optimizations.
* Good experience in Hive partitioning, bucketing and perform different types of joins on Hive tables and implementing Hive Sere like JSON, ORC, XML and Parquet.
* Worked on different file formats (PARQUET, ORCFILE, TEXTFILE) and different Compression Codecs (GZIP, BZIP2, SNAPPY, LZO).
* Designed, developed, and maintained highly available clusters and cluster pools on Databricks, ensuring optimal performance and scalability.
* Extensive experience with Snowflake utilities, including Snowpipe, SnowSQL, stored procedures, tasks, and Snowflake integration objects for seamless data processing.
* Proficient in using DBT (Data Build Tool) to build and manage ETL pipelines, enabling efficient data transformations and modeling.
* Hands-on experience with NoSQL databases, including MongoDB and CosmosDB, for handling semi-structured and unstructured data.
* Worked with Databricks Unity Catalog, implementing data sharing, Auto Loader, and Delta Live Tables for efficient data governance and real-time data processing.
* Experience managing Snowflake Multi-Cluster Warehouses, optimizing performance and workload management for high-volume data processing.
* Working experience with Reltio Master data management platform to load and extract data.
* Extensively designed, developed and maintained High availability clusters and cluster pools on Databricks.
* Extensively build and developed ingestion pattern with Azure Data factory by using Integration runtimes, azure Key vaults, azure logic apps, Azure Storage Containers (Blob, ADLS Gen1 and Gen 2), Function Apps, Azure VMs and Azure DevOps Git Repos .
* 8+ years of hands-on experience in Oracle Database Development using PL/SQL and SQL, ensuring optimized query performance.
* Designed and implemented complex stored procedures, functions, triggers, and packages to support business logic and data transformation.
* Expertise in query optimization, indexing strategies, and performance tuning for high-volume transactional and analytical databases.
* Proficient in implementing CI/CD frameworks for data pipelines using tools like Jenkins, ensuring efficient automation and deployment.
* Proficient in using Git for version control, ensuring efficient tracking and management of code changes throughout data engineering projects.
* Conducted code reviews on GitHub, identifying and addressing issues related to code quality, performance, and adherence to coding standards.
* Designed and developed Tableau dashboards and reports to visualize complex data sets, providing actionable insights for business stakeholders.
* Conducted regular quality assurance checks on Tableau reports to identify and rectify data inconsistencies and visualization errors.
* Led the migration of data from PostgreSQL to Snowflake, including schema translation, data type mapping, and performance optimization.
* Designed and implemented robust ETL pipelines to facilitate incremental and full-load migrations with minimal downtime.
* Collaborated with cross-functional teams including DBAs and application developers to validate and reconcile data post-migration.
* Proficient with BI and data visualization tools including Power BI, Google Data Studio, Looker, and Data Studio.
* Hands-on with Talend for ETL processes and Google Suite for collaborative project execution
* Designed and implemented data warehousing solutions using Apache Hive, enabling efficient querying and analysis of large-scale structured and semi-structured data.
* Conducted performance tuning and capacity planning exercises for Kubernetes clusters to handle increasing data processing demands effectively.
* Good experience with various IDEs like Eclipse, IntelliJ.
* Created Git repos and Jenkins jobs for the build and deployment of code to different environments and also to the Docker Container.

**EDUCATION:**

* Bachelors of Computer Science from Institute of Aeronautical Engineer.

**CERTIFICATION:**

* AWS DATA ENGINEER ASSOCIATE.
* AZURE FUNDAMENTALS – AZ-104.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Big Data Technologies** | HDFS, Hive, Flume, Oozie, Avro, Hadoop Streaming, Zookeeper, Kafka, Impala, Apache Spark, hue, Ambari. Apache ignite. |
| **Hadoop Distributions** | Cloudera (CDH4/CDH5), Horton Works |
| **Cloud**  | Azure Blob Storage, ADF, Databricks, HDInsight, Azure Functions, Synapse |
| **Languages** | Java, Python, SQL, PL/SQL, HQL |
| **IDE Tools** | Eclipse, IntelliJ, Net Beans, Anaconda |
| **Framework** | Hibernate, Spring, Struts, Junit |
| **Operating Systems** | Windows (XP,7,8), UNIX, LINUX, Mac OS-X, Ubuntu, CentOS |
| **Application Servers** | Tomcat, Web Logic, Web Sphere, Servlets |
| **Reporting Tools/ETL Tools** | Tableau, Informatica, Teradata |

**WORK EXPERIENCE:**

***Client: GM FINANCIAL July 2023 – Till now***

***Role: Data Engineer***

***Responsibilities:***

* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics. Data
* Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure) and processing the data in Azure Data bricks. Experience in Moving Data in and out of Windows Azure SQL Databases and Blob Storage.
* Worked on Developing Data Pipeline to Ingest Hive tables and File Feeds and generate Insights into Cassandra DB,
* Written Scala and python script notebooks for Azure Data bricks transformation task.
* Involved in converting Hive/aw queries into Spark transformations using Spark RDDs and Scala.
* Analyzed the SQL scripts and designed the solution to implement using Scala.
* Designed and optimized scalable data pipelines (batch and streaming) to support ML model training, inference, and real-time analytics.
* Integrated multiple CRM systems (e.g., Salesforce, Zoho) into a centralized data lake, improving customer 360° view across departments.
* Wrote modular Python and R scripts to analyze churn behavior and segment customer lifecycles.
* Worked closely with cross-functional teams, demonstrating strong teamwork in Agile sprint planning and delivery.
* Collaborated with data scientists and MLOps teams to operationalize ML models, ensuring proper versioning, monitoring, and automated retraining.
* Designed and optimized Google BigQuery data models and queries for efficient large-scale data processing.
* Developed serverless data pipelines and ETL workflows using Google Cloud (GCP), Dataflow, and Cloud Functions.
* Experience in Palantir Foundry, working on data integration, pipeline development, and ontology modeling for enterprise analytics.
* Developed RESTful APIs and microservices using Go, Python, and Flask, enabling seamless data access and integration.
* Developed RESTful web services using .NET and C#, enabling secure integration with third-party applications.
* Collaborated on responsive interface design for analytics dashboards using MVC architecture and web frameworks.
* Supported multi-tenant SaaS infrastructure, ensuring data isolation and real-time processing for client-specific insights.
* Applied strong knowledge in data warehouse and data mart architecture, optimizing Snowflake for scalability, query performance, and cost efficiency.
* Utilized Snowflake features like virtual warehouses, Snowpipe, and time travel to manage operational complexities.
* Conducted data quality checks, validation scripts, and implemented rollback strategies for safe migration.
* Provided onshore support for UAT and post-migration troubleshooting, resolving critical issues related to legacy PostgreSQL procedures.
* Documented best practices for future migrations and trained offshore team members.
* Created dynamic reporting solutions using Power BI, and Google Data Studio for cross-functional teams.
* Implemented Talend-based pipelines for specific data integration scenarios.
* Collaborated via Google Suite and Microsoft Office tools for real-time documentation and stakeholder presentations.
* Architected Microsoft Fabric-based data platform, integrating Synapse, ADF, and Power BI for real-time analytics.
* Designed and implemented CDC pipelines using Debezium and Kafka, capturing changes from Oracle and MySQL into Azure Synapse for near real-time reporting.
* Built fault-tolerant Debezium connectors with Kafka Connect for transactional systems, improving system observability and recovery time.
* Led data modeling efforts across Azure Synapse and Snowflake, collaborating with product teams to align business goals with data architecture.
* Played a key leadership role in cross-functional teams to guide end-to-end data integration and delivery in Agile environments.
* Maintained data pipelines leveraging Amazon S3 for large-scale storage and archival of raw and transformed data.
* Utilized Vault for secure storage and rotation of secrets and credentials used in pipeline orchestration.
* Coordinated with stakeholders, demonstrating strong communication skills to translate business requirements into technical solutions.
* Implemented Reltio – Cloud MDM Platform to establish a golden record strategy, improving customer data quality and deduplication.
* Designed data mastering workflows in Reltio, integrating with Azure Synapse, Databricks, and Power BI for enterprise-wide reporting.
* Developed real-time data synchronization between Reltio MDM and Azure Data Lake using Azure Data Factory and API integrations.
* Built fault-tolerant Debezium connectors with Kafka Connect for transactional systems, improving system observability and recovery time.
* Automated data stewardship processes to enable self-service MDM capabilities for business users.
* Integrated ML-based entity resolution models within Reltio for enhanced customer identity matching and householding.
* Designed and managed secure API gateways for data-driven applications using Google Cloud API Gateway and AWS API Gateway.
* Proficient in SQL optimization, enhancing query performance in BigQuery, Redshift, and Oracle databases.
* Implemented ETL/ELT processes for structured and unstructured data using Spark, Airflow, and dbt.
* Good Understanding of Data ingestion, Airflow Operators for Data Orchestration and other related python libraries.
* Used Spark-Streaming APIs to perform necessary transformations.
* Designed and implemented scalable data pipelines using Azure Data Factory (ADF), Databricks, and Palantir Foundry, optimizing ingestion and transformation processes.
* Created and reviewed business requirement documents (BRDs), functional design documents (FDDs), and test cases, collaborating closely with stakeholders.
* Ensured data integrity and consistency through detailed validation processes and reconciliation reports.
* Developed reusable job frameworks and configuration-driven integrations using Foundry’s Code Workbooks and Ontology Management to streamline enterprise data workflows.
* Deployed and managed cloud-native data solutions on AWS (S3, Redshift, Glue), Azure (Data Factory, Synapse), and GCP (BigQuery, Dataflow, Dataproc).
* Leveraged distributed computing technologies like Hadoop, Kafka, and Apache Flink for high-throughput data ingestion and processing.
* Expertise in Extract, Transform, and Load (ETL) processes for Oracle Autonomous Data Warehouse (ADW) in Oracle Cloud Infrastructure (OCI).
* Proficient in PL/SQL, SQL, and Oracle database development, ensuring efficient data transformations and high-performance querying.
* Independently designed, developed, and optimized ETL workflows for large-scale data processing with minimal supervision.
* Strong understanding of dimensional modeling and data warehousing best practices, ensuring efficient data storage and retrieval.
* Translated business and functional requirements into technical specifications for data warehouse design and implementation.
* Developed data pipelines and performance tuning strategies for Oracle ADW, optimizing query execution and resource utilization.
* Experience in performance tuning, query optimization, and data partitioning to improve data retrieval efficiency.
* Proficient in Oracle Business Intelligence Enterprise Edition (OBIEE) for data visualization and reporting.
* Integrated data from multiple cloud and on-prem sources, leveraging Foundry’s Pipelines and Data Lineage Tools for transparent, auditable transformations.
* Designed and implemented data pipelines using Python & PySpark to process large-scale data on Azure.
* Developed automated Python scripts to integrate Snowflake, Synapse, and Power BI for data transformation.
* Deployed machine learning models using Azure ML and Databricks for predictive analytics.
* Created Python-based ETL frameworks for data ingestion from APIs and flat files into Snowflake.
* Developed and optimized queries in Foundry’s SQL Transformations, improving performance and reducing compute costs.
* Utilized SAS for data processing, statistical analysis, and advanced reporting.
* Applied analytical techniques to optimize business intelligence reports and forecasting models.
* Developed Neo4J graph models for complex relationship analysis and data discovery in financial applications.
* Enhanced security and access controls using Foundry’s Role-Based Access Management (RBAC) for data governance and compliance.
* Collaborated with data scientists and analysts to implement machine learning-ready datasets in Foundry for advanced analytics.
* Designed and developed a microservices-based architecture using Spring Boot and PostgreSQL, enhancing system scalability and performance.
* Built dynamic and responsive front-end components using React.js and Next.js, improving user experience.
* Integrated CMS solutions to streamline content management and enhance workflow automation.
* Optimized MySQL and PostgreSQL database queries, improving data retrieval efficiency.
* Worked with SQL Server, MySQL, and Oracle databases, implementing query optimization, indexing, and stored procedures for efficient data processing.
* Implemented replication and backup strategies in MySQL and MongoDB for high availability and disaster recovery.
* Developed reusable job frameworks and configuration-driven integrations using Foundry’s Code Workbooks and Ontology Management to streamline enterprise data workflows.
* Implemented SAP HANA integration with Azure Synapse Analytics to enhance financial reporting and supply chain analytics.
* Enhanced security and access controls using Foundry’s Role-Based Access Management (RBAC) for data governance and compliance.
* Developed and optimized queries in Foundry’s SQL Transformations, improving performance and reducing compute costs.
* Implemented CI/CD pipelines using Jenkins, GitHub Actions, and Azure DevOps to automate code testing, deployment, and monitoring.
* Integrated Git and Bitbucket for version control, ensuring smooth collaboration and continuous integration.
* Configured unit tests and regression tests within CI/CD pipelines to detect issues early and ensure code stability before deployment.
* Utilized Terraform and AWS CloudFormation to automate infrastructure provisioning and updates, ensuring consistency across environments.
* Worked on SAP ECC data extraction using BAPI, IDocs, and OData services to enable enterprise-wide analytics.
* Collaborated with data scientists and analysts to implement machine learning-ready datasets in Foundry for advanced analytics.
* Developed and optimized machine learning models using Azure Databricks, leveraging Bayesian Machine Learning techniques.
* Implemented Machine Learning Operations (MLOps) pipelines for automated model deployment and monitoring.
* Developed PySpark-based ETL transformations for processing large datasets efficiently.
* Implemented SAP HANA integration with Azure Synapse Analytics to enhance financial reporting and supply chain analytics.
* Developed and optimized queries in Foundry’s SQL Transformations, improving performance and reducing compute costs.
* Developed machine learning data pipelines using Databricks, Synapse, and Snowflake, optimizing performance for data science teams.
* Enhanced security and access controls using Foundry’s Role-Based Access Management (RBAC) for data governance and compliance.
* Collaborated with data scientists and analysts to implement machine learning-ready datasets in Foundry for advanced analytics.
* Worked with SQL Server, MySQL, and Oracle databases, implementing query optimization, indexing, and stored procedures for efficient data processing.
* Implemented replication and backup strategies in MySQL and MongoDB for high availability and disaster recovery.
* Developed reusable job frameworks and configuration-driven integrations using Foundry’s Code Workbooks and Ontology Management to streamline enterprise data workflows.
* Integrated data from multiple cloud and on-prem sources, leveraging Foundry’s Pipelines and Data Lineage Tools for transparent, auditable transformations.
* Enhanced security and access controls using Foundry’s Role-Based Access Management (RBAC) for data governance and compliance.
* Developed and optimized queries in Foundry’s SQL Transformations, improving performance and reducing compute costs.
* Collaborated with data scientists and analysts to implement machine learning-ready datasets in Foundry for advanced analytics.
* Worked with spark to consume data from Kafka and convert that to common format using Scala.
* Worked in Agile development environment in sprint cycles of two weeks by dividing and organizing tasks.
* Designed and implemented **scalable data pipelines** using **Azure Data Factory (ADF), DBT Labs, and Databricks**, optimizing ingestion and transformation processes.
* Developed **reusable job frameworks** and **configuration-driven integration** to enhance **scalability and maintainability.**
* Managed ingestion of **structured and unstructured data** from **ADLS, Salesforce, SharePoint, and external partner data sources (IHME, WASDE, etc.).**
* Managed and optimized MongoDB databases, focusing on query performance, indexing, and high-availability configurations.
* Implemented MongoDB replication and sharding strategies to ensure data availability and scalability.
* Designed Cosmos DB solutions with partitioning strategies and consistency tuning for distributed applications.
* Designed and optimized MySQL databases, improving query performance through indexing and partitioning.
* Implemented replication and backup strategies in MySQL for high availability and disaster recovery.
* Developed and maintained stored procedures, triggers, and views for efficient data processing.
* Tuned MySQL queries using EXPLAIN plans, optimizing performance for large datasets.
* Integrated MySQL databases with cloud platforms such as AWS RDS and Azure Database for MySQL.
* Implemented **ELT processes, source-to-target mapping, and transformation logic in DBT Labs, ADF, Databricks Notebooks, and Snowflake.**
* Designed and optimized **Client Tables and External Tables** for structured/unstructured data in a **lakehouse architecture.**
* Integrated **Databricks Unity Catalog** for efficient data governance and security.
* Collaborated with **data scientists, analysts, report developers, and infrastructure teams** for seamless data workflows.
* Co-developed **CI/CD pipelines using GitHub Actions and Azure DevOps**, streamlining deployments and version control.
* Built visual reporting assets in Looker and Power BI to support regulatory and actuarial teams.
* Ensured secure, clean access and governed datasets using DevOps principles and CI/CD practices.
* Contributed in a high-paced environment, consistently delivering within sprint cycles alongside product and QA teams.
* Developed and optimized Snowflake Data Warehouse solutions, improving query performance by 40%.
* Implemented Snowflake Streams, Tasks, and Snowpipe for real-time data ingestion and processing.
* Automated **source-to-target mappings, data pipelines, and data lineage** in **Collibra**, ensuring compliance and traceability.
* Developed a microservices-based backend using C# and .NET, ensuring high performance and maintainability.
* Built RESTful APIs and integrated with PostgreSQL and MySQL for efficient data storage.
* Develop and maintain various data ingestion pipelines as per the design architecture and processes: source to landing, landing to curated & curated to process.
* Managed Microsoft SQL Server environments, including backups, user management, and disaster recovery planning.
* Developed and optimized T-SQL procedures for complex data transformations and analytics workflows.
* Played a Sr. Engineering role guiding junior developers and maintaining high standards in code quality and architecture.
* Use various types of activities: data movement activities, transformations, and control activities; Copy data, Data flow, Get Metadata, Lookup, Stored procedure, Execute Pipeline
* Dealt with Storage services named Data Lake Storage Gen 1 and Gen 2for hosting CSV, JSON and Parquet files and managing access across storage accounts.
* Write Data bricks notebooks (Python) for handling large volumes of data, transformations, and computations
* Work with various file formats: flat-file TXT & CSV; parquet & other compressed formats
* Build Delta Lake for the curated layer, maintain high-quality data available for the teams: data scientists, finance etc.
* Utilize Azure’s ETL, Azure Data Factory (ADF) services to ingest data from legacy disparate data stores - SAP (Hana), SFTP servers & Cloud era Hadoop’s HDFS to Azure Data Lake Storage (Gen2)
* Developed mapping document to map columns from source to target.
* Created azure data factory (ADF pipelines) using Azure blob.
* Performed ETL using Azure Data Bricks. Migrated on-premises Oracle ETL process to Azure Synapse Analytics.
* Worked on python scripting to automate generation of scripts. Data curation done using azure data bricks.
* Worked on azure data bricks, PySpark, HDInsight, Azure ADW and hive used to load and transform data.
* Implemented and Developing Hive Bucketing and Partitioning.
* Applied indexing strategies and warehouse cost optimization techniques to reduce compute costs by 30%.
* Developed RESTful APIs using Java 13+ and Spring Boot to interact with Snowflake data models.
* Implemented GraphQL APIs for efficient querying of large datasets.
* Built AWS Lambda functions with Java to automate data transformation processes.
* Implemented Kafka, spark structured streaming for real time data ingestion.
* Used Azure Data Lake as Source and pulled data using Azure blob.

**Environment**: ADF, Data bricks and ADL Spark, Hive, HBase, Sqoop, Flume, ADF, Blob, cosmos DB, Map Reduce, HDFS, Cloud era, SQL, Apache Kafka, Azure, Python, power BI, Unix, SQL Server.

***Client:*** ***STATE OF NEW MEXICO Sep 2021 to June 2023***

***Role: Data Engineer***

***Responsibilities:***

* Coordinated with business customers to gather business requirements. And also interact with other technical peers to derive Technical requirements and delivered the BRD, MDD and other documents.
* Partnered with business and subject matter experts to understand the technology landscape and produce integrated end-to-end solution that manifest building of centralized Data lake in Azure Data lake (ADLS) and Cloud based Snowflake Database.
* Developed clean, modular data pipelines using PySpark and ADF; ensured high code maintainability and team collaboration.
* Created Looker dashboards for fraud detection KPIs; enabled real-time monitoring for fraud analysts.
* Applied statistical modeling techniques to identify anomalies and trends in transactional data.
* Integrated MongoDB into enterprise applications, optimizing schema design, indexing for low-latency access, and replication strategies for high availability.
* Implemented Talend for automating ETL of telecom metadata.
* Designed Looker dashboards for fraud detection and operational insights.
* Enhanced cross-team collaboration with Microsoft Office tools and Google Suite.
* Migrated on-prem data to AWS Redshift and S3, reducing storage costs by 40%.
* Developed AWS Glue jobs (PySpark-based) to transform raw data and store it in AWS Redshift for analytics.
* Automated data pipeline orchestration with AWS Step Functions and EventBridge for workflow scheduling.
* Utilized Athena and Redshift Spectrum for cost-efficient SQL queries on large datasets stored in AWS S3.
* Ensured data security by implementing AWS KMS encryption, IAM roles, and S3 bucket policies.
* Designed and Developed medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB)
* Data migration experience from on-premise to cloud by using command line tools like AZCopy, Az PowerShell and Azure CLI.
* Applied DevOps best practices to implement CI/CD pipelines for automated deployment of data pipelines in ADF and Databricks.
* Developed anomaly detection models using R and statistical methods to proactively flag fraudulent patterns in telecom usage.
* Ensured attention to detail in schema design and version control, maintaining clean and auditable data pipelines.
* Implemented end-to-end Azure Data Factory workflows, reducing data processing time by 60%.
* Designed telecom-specific fraud detection models using ontology modeling and AI-driven insights.
* Participated in product roadmap discussions, ensuring data capabilities aligned with evolving platform needs.
* Collaborated with business units and product managers to gather data requirements and ensure alignment with delivery timelines.
* Tuned SQL queries and transformations using Toad for Oracle and DB2 for high-performance workloads.
* Demonstrated effective team leadership during high-priority initiatives involving migration and cloud modernization.
* Led the migration of legacy MDM systems to Reltio – Cloud MDM Platform, ensuring seamless data integration with Azure services.
* Designed data ingestion pipelines for structured/unstructured MDM data using Azure Synapse, ADF, and dbt.
* Configured Reltio Data Model and Relationship Graphs for hierarchical data representation and analytics.
* Established data governance policies in Reltio MDM and Microsoft Purview, ensuring data accuracy, lineage tracking, and compliance.
* Optimized ETL workflows between Reltio and Azure Synapse, reducing data processing time by 40%.
* Strong interpersonal skills, collaborating effectively with cross-functional teams in Agile and Kanban environments.
* Led and participated in Kanban-based development cycles, ensuring efficient task tracking and sprint planning.
* Excellent technical writing skills, creating detailed documentation for APIs, data models, and workflow automation.
* Worked on Palantir Foundry for data governance, ontology modeling, and scalable pipeline orchestration.
* Built and maintained secure, high-performance data pipelines in Palantir, integrating with cloud-based data platforms.
* Implemented machine learning workflows in Palantir, Google Cloud AI, and TensorFlow, supporting data-driven decision-making.
* Developed cloud-native applications in a fast-paced, innovative environment, leveraging Ruby and JavaScript to automate key business processes.
* Implemented microservices architecture using Spring Boot, ensuring high availability and seamless scalability.
* Engineered React.js and Next.js front-end solutions, delivering an intuitive and responsive UI.
* Designed and optimized data reporting dashboards with MySQL and PostgreSQL, enabling real-time analytics and decision-making.
* Deployed applications using Docker and Kubernetes, streamlining microservices management and improving scalability.
* Integrated AWS CloudWatch, Datadog, and Splunk into CI/CD workflows for real-time monitoring and alerting, ensuring quick response to failures.
* Designed CI/CD workflows for Azure Synapse, Snowflake, and AWS Redshift, optimizing data pipeline deployments.
* Implemented automated security scans and policy checks in CI/CD pipelines to ensure compliance with industry standards.
* Used SAS for trend analysis, predictive modeling, and performance optimization.
* Applied data analytics techniques to enhance data-driven decision-making processes.
* Led Snowflake performance tuning initiatives, optimizing complex queries and materialized views.
* Built Terraform scripts for Snowflake infrastructure automation, enabling seamless deployment.
* Designed and implemented data governance policies for compliance and security.
* Developed Python-based data validation scripts for ensuring data integrity in Snowflake.
* Integrated Snowflake with cloud storage and external services for cross-cloud data sharing.
* Migrated on-premise data to GCP BigQuery using Python-based data pipelines.
* Developed Python scripts for data validation and transformation across multiple cloud storage layers.
* Implemented ML-based anomaly detection in streaming data pipelines with Apache Flink.
* Developed Neo4J-based graph solutions to support fraud detection and customer relationship analysis.
* Designed and implemented GraphQL APIs for optimized querying of complex datasets, reducing response time and enhancing data retrieval efficiency.
* Developed Java-based microservices for data processing and ETL orchestration.
* Implemented large-scale data processing using AWS Lambda, S3, and DynamoDB.
* Built and optimized REST APIs for real-time analytics and reporting applications.
* Worked with GraphQL to enable flexible and efficient data retrieval from Snowflake datasets.
* Implemented Data Loss Prevention (DLP) frameworks across cloud environments, securing structured and unstructured data.
* Integrated MongoDB into enterprise applications, optimizing schema design and indexing for high-performance queries.
* Built machine learning pipelines using Azure Databricks, leveraging PySpark and pandas for feature engineering and model training.
* Designed and developed serverless ETL pipelines using AWS Glue, Step Functions, and Lambda to automate data ingestion.
* Optimized AWS Redshift clusters using WLM, Sort Keys, and Dist Keys, reducing query execution time by 30%.
* Built real-time event-driven architecture using AWS Kinesis Firehose and processed streaming data using Apache Flink.
* Built scalable ETL workflows in Foundry and SAP Data Services to automate data ingestion and quality assurance processes, improving data reliability.
* Designed and optimized Palantir Foundry Object Models and Ontologies to support financial and supply chain analytics.
* Developed Custom Python Transformations in Foundry and SAP HANA Stored Procedures to extend native capabilities and implement complex business logic.
* Implemented supply chain analytics dashboards in Power BI using SAP HANA datasets for procurement and logistics optimization.
* Ensured data integrity and security by implementing Palantir Foundry’s Data Governance Policies and Compliance Controls.
* Led the integration of Palantir Foundry with Snowflake and Azure Synapse Analytics, ensuring seamless data processing across cloud environments.
* Built scalable ETL workflows in Foundry to automate data ingestion and quality assurance processes, improving data reliability.
* Designed and optimized Palantir Foundry Object Models and Ontologies to support financial and life insurance analytics.
* Developed Custom Python Transformations in Foundry to extend native capabilities and implement complex business logic.
* Ensured data integrity and security by implementing Palantir Foundry’s Data Governance Policies and Compliance Controls.
* Developed and optimized JDBC-based data access layers for enterprise applications.
* Implemented connection pooling techniques using JDBC to enhance database efficiency and reduce latency.
* Developed and tested Java applications utilizing JDBC for seamless interaction with MySQL and Oracle databases.
* Integrated JDBC with Hibernate and Spring frameworks for efficient ORM and transaction management.
* Debugged and optimized SQL queries executed via JDBC, reducing query execution time by 30%.
* Led data pipeline development and data governance strategies for **telecom fraud detection,** integrating **ontology modeling.**
* Worked on **automobile industry data integration,** enhancing **real-time analytics and machine learning model deployment.**
* Migrated large-scale **GCP data environments** to Azure and AWS, ensuring minimal downtime and data integrity.
* Developed **machine learning data pipelines** using **Databricks, Synapse, and Snowflake**, optimizing performance for data science teams.
* Drove excellence in Data and Business Intelligence (BI) development using SAP DataSphere, S/4HANA Analytics, and Core Data Services (CDS) Views.
* Worked with SAP BTP DataSphere to integrate Data Lakes and cloud-based solutions for enterprise-wide data accessibility.
* Designed and developed advanced data modeling solutions to enhance decision-making across Finance, Sales, Supply Chain, and Operations.
* Implemented ETL workflows and data pipelines, ensuring smooth data integration from various sources into SAP DataSphere.
* Developed cutting-edge data visualization techniques to provide actionable business insights.
* Ensured the scalability and robustness of data systems by staying updated with emerging technology trends.
* Optimized CDS Views and SQL queries to enhance reporting and analytical performance.
* Collaborated with business users and cross-functional teams to gather requirements and build tailored data solutions.
* Led Change Data Capture (CDC) implementations to enable real-time data updates across SAP and non-SAP systems.
* Designed and implemented data warehousing strategies for improved performance and governance.
* Implemented **data security automation** across **AWS, Azure, and Snowflake** with **SailPoint integration.**
* Designed and Developed data pipelines for provisioning as well as orchestration of various Data science projects for predictive and descriptive modeling
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards
* Developed pipelines by connecting to various source systems like Oracle , SQL Server , Teradata
* Developed Spark applications on Databricks using Python, Scala and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data
* Developed rules for implanting Hot and Cold storage / cache policies
* Analyzed the hive SQL scripts and designed the solution to implement using PySpark on Databricks.
* Creating, validating and reviewing solutions and effort estimate of converting existing workloads from classic to ARM based Azure Cloud Environment
* Used Control-M as a scheduling tool to handle event based scheduling and time based scheduling.
* Worked on GIT to maintain source code in Azure Git and GitHub repositories for Azure ADF pipelines and Azure Databricks
* Developed workflows to automate Real time and batch processing in Databricks
* Loaded data into snowflake tables by using COPY INTO Command.
* Build Store procs on Snowflake to automate batch loads and scheduled by using Tasks.
* Build Snow pipe for real time streaming and used stream objects
* Build logic App and functional App for building additional functionality like running Power BI Extracts, extracting Data from share point
* Worked on DBT connection setup with snowflake and have implemented ETL pipelines in snowflake with DBT
* Used DBT to test the data (schema tests, referential integrity tests, custom tests and ensure data quality
* Used DBT to debug complex chains of queries. They can be split into multiple models and macros that can be tested separately
* Schedule and automate DBT runs to keep data models up-to-date and synchronized with source systems.
* Worked on GIT to maintain source code in Azure Git and GitHub repositories for Azure ADF pipelines and Azure Databricks.

**Environment**: Azure Cloud, Azure Data Factory, Azure HDInsight’s, Azure Blob Storage, Azure Data Explorer, Azure Eventhub, Databricks, SQL Server, Oracle, Teradata, Workload Automation scheduler, DBT (Data Build Tool)

***Client: Albert Energy Regulator, Calgary AB. June 2020 to Aug 2021***

***Role: Data Engineer***

***Responsibilities:***

* Develop, design data models, data structures and ETL jobs for data acquisition and manipulation purposes.
* Develop deep understanding of the data sources, implement data standards, maintain data quality and master data management.
* Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in Azure Databricks.
* Developed high-performance data pipelines using Talend, Spark, and DataStage.
* Designed reporting solutions with Power BI, Google Data Studio, and Looker.
* Used Microsoft Excel and Google Sheets for pre-analysis and QA collaboration.
* Implemented Copy activity, Custom Azure Data Factory Pipeline Activities.
* Primarily involved in Data Migration using SQL, SQL Azure, Azure Storage, and Azure Data Factory, SSIS, PowerShell.
* Worked on SQL Server Integration Services (SSIS), SSAS, SSRS, T-SQL skills, stored procedures, triggers.
* Architect and implement ETL and data movement solutions using Azure Data Factory, SSIS create and run SSIS Package ADF V2 Azure-SSIS IR.
* Implemented Two-Factor Authentication (TFA) using Google Authenticator, SMS, and email-based OTP verification.
* Developed customer-facing modules using .NET and C#, ensuring a smooth user experience and reliable backend logic.
* Implemented responsive design principles across dashboards to support all screen sizes using Bootstrap and custom media queries.
* Developed and exposed secure REST APIs for internal tools and external clients, enabling seamless data exchange between CRM and analytics layers.
* Integrated Zendesk ticketing data into a central data warehouse for churn prediction and SLA tracking.
* Conducted end-to-end root cause analysis on data pipeline failures and dashboard discrepancies, improving incident resolution time by 45%.
* Delivered optimized, scalable MVC-based solutions for interactive reporting and insights.
* Integrated CRM and transaction data into a centralized web-based SaaS platform, improving marketing and customer outreach workflows.
* Designed and tested interfaces using MVC and REST APIs for efficient access to customer data profiles.
* Integrated Reltio MDM API services with Azure Functions and Power BI, enabling real-time master data access for analytics.
* Enhanced match & merge logic in Reltio to ensure consistent and accurate entity resolution across multi-cloud environments.
* Administered Amazon Aurora clusters, optimized for cost and performance in a hybrid cloud environment.
* Automated database tasks using Shell Scripts, improving deployment times and reducing manual interventions.
* Led continuous improvement initiatives that reduced ETL runtime by 35% via advanced tuning and process automation.
* Developed custom workflows and UI extensions within Reltio MDM, improving data stewardship efficiency by 30%.
* Built end-to-end automation for MDM data ingestion, processing, and validation using Airflow, Azure DevOps, and GitHub Actions.
* Implemented data quality dashboards in Power BI and Reltio Insights to monitor data completeness and integrity.
* Automated AWS data workflows using Python and Boto3 for S3, Redshift, and Glue.
* Designed and optimized Snowflake data models, leveraging Python scripts for schema evolution.
* Integrated machine learning models within AWS Lambda for fraud detection and real-time analytics.
* Designed and integrated TFA mechanisms for securing sensitive data and user authentication in web applications.
* Developed custom security policies and authentication modules using TFA for compliance with security standards.
* Integrated TFA with IAM solutions across AWS, Azure, and GCP for enhanced cloud security.
* Configured and tested multi-factor authentication (MFA) protocols for enterprise applications, ensuring compliance with regulatory requirements.
* Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).
* Expert in developing JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data. Expert in using Databricks with Azure Data Factory (ADF) to compute large volumes of data.
* Performed ETL operations in Azure Databricks by connecting to different relational database source systems using jdbc connectors.
* Developed Python scripts to do file validations in Databricks and automated the process using ADF.
* Developed an automated process in Azure cloud which can ingest data daily from web service and load in to Azure SQL DB.
* Developed and implemented SAP DataSphere-based data solutions to support business intelligence and analytics.
* Optimized Core Data Services (CDS) Views for efficient data retrieval and reporting.
* Designed and built data pipelines integrating SAP S/4HANA Analytics with cloud-based storage solutions.
* Worked with ABAP and SQL to enhance SAP DataSphere data models and improve performance.
* Led data warehousing initiatives, ensuring data consistency and quality across the enterprise.
* Implemented Change Data Capture (CDC) mechanisms to facilitate real-time data synchronization.
* Developed interactive data visualizations for Finance and Supply Chain analytics, enhancing operational efficiency.
* Designed and executed ETL workflows that streamlined data movement across SAP and external platforms.
* Designed and developed serverless ETL pipelines using AWS Glue, Step Functions, and Lambda to automate data ingestion.
* Optimized AWS Redshift clusters using WLM, Sort Keys, and Dist Keys, reducing query execution time by 30%.
* Integrated AWS S3 with Snowflake using external stages for efficient data loading and processing.
* Built real-time event-driven architecture using AWS Kinesis Firehose and processed streaming data using Apache Flink.
* Implemented AWS CloudFormation to automate infrastructure provisioning for data pipelines.
* Developed Streaming pipelines using Azure Event Hubs and Stream Analytics to analyze data for dealer efficiency and open table counts for data coming in from IOT enabled poker and other pit tables.
* Analyzed data where it lives by Mounting Azure Data Lake and Blob to Databricks.
* Used Logic App to take decisional actions based on the workflow.
* Developed custom alerts using Azure Data Factory. SQLDB and Logic App.
* Developed Databricks ETL pipelines using notebooks. Spark Data frames. SPARK SQL and python scripting.
* Developed complex SQL queries using stored procedures, common table expressions (CTES), temporary table to support Power BI reports.
* Worked with enterprise Data Modeling team on creation of Logical models.
* Implemented complex business logic through T-SQL stored procedures. Functions. Views and advance query concepts.
* Development level experience in Microsoft Azure providing data movement and scheduling functionality to cloud-based technologies such as Azure Blob Storage and Azure SQL Database.
* Independently manage development of ETL processes - development to delivery.

**Environment**: Cloudera Manager (CDH5), Azure, PySpark, HDFS, NiFi, Pig, Hive, S3, Kafka, Scrum, Git, Sqoop, Oozie. PySpark, Informatica, Tableau, OLTP, OLAP, HBase, Cassandra, Informatica, SQL Server, Python, Shell Scripting, XML, Unix

***Client: Publix, Lakeland, FL. April 2018 to May 2020***

***Role: Data engineer***

***Responsibilities:***

* Design and implement end-to-end data solutions (storage, integration, processing, visualization) in Azure and Databricks.
* Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure
* Develop conceptual solutions & create proof-of-concepts to demonstrate viability of solutions
* Collaborate with application architects and DevOps
* As a part of center of excellence member, Identified and implemented best practices, tools and standards
* Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, Databricks, SQL Database and SQL Data warehouse environment.
* Created ELT Data Strategy in Synapse using ADF, Polybase (External Tables) and Stored Procedures.
* Created various optimal data distributions like Hash, Round-robin and Replicated in Synapse.
* Used Azure DevOps for continuous integration and continuous deployment of ADF and Databricks scripts.
* Identify potential problems and recommend alternative technical solutions
* Participating in Technical Architecture Documents, Project Design and Implementation Discussions
* Migrate data from traditional database systems to Azure big data platforms
* Recommend, Design and construct policies and standards that impact infrastructure operations and services but also improve overall business performance.
* Creating Databricks notebooks using SQL, Python and automated notebooks using jobs.
* Create, read, write, overwrite and append DeltaLake tables (delta format) by creating DataFrames using Azure Databricks DataFrames Reader.
* Transform the raw data and clickstream data to meaningful information to support business decisions.
* Migration of on-premises data (Oracle, SQL Server, DB2 and Teradata) to Azure Data Lake Store (ADLS Gen2) using Azure Data Factory (ADF V2).
* Automated jobs using different triggers like Events, Schedules and Tumbling in ADF.
* Work closely across Support teams, Solution Architecture teams and peers to establish and follow best practices while solving customer problems
* Provided guidance to developers for developing the reusable and performance improving scripts.
* Worked with QA teams to resolve any issues and helped them to pass all the test cases.
* Worked with Development team in deployments and had a successful deployment without any major issues and validated the first run BAUs.

**Environment**: PySpark, Map Reduce, HDFS, Sqoop, flume, Kafka, Hive, Pig, HBase, SQL, Shell Scripting, Eclipse, SQL Developer, Git, SVN, JIRA, Unix.

***Client: VIZIENT INC, IRVING, TX. Sep 2016 to Mar 2018***

***Role: Hadoop Developer / Big data Developer.***

***Responsibilities:***

* Used Sqoop to import and export data between Hadoop and relational databases.
* Utilized Sqoop's built-in support for Kerberos authentication to securely transfer data between Hadoop and external systems.
* Developed and optimized Hive queries using techniques like partitioning, bucketing, and indexing.
* Created and managed Hive tables, views, and functions to organize and manipulate data efficiently.
* Developed Spark applications using Spark Core, Spark SQL, and Spark Streaming.
* Worked with Spark SQL for querying structured data in Hadoop.
* Monitored Oozie workflows for troubleshooting and optimization.
* Designed and developed automated data ingestion pipelines using Sqoop and Apache Oozie, allowing seamless data movement from various data sources to Hadoop clusters on a scheduled basis, enhancing data availability for analytics teams.
* Integrated Oozie with other Hadoop ecosystem tools, such as Sqoop and Hive, to create end-to-end data processing pipelines.
* Created reports, dashboards, and visualizations based on SQL query results using tools like Tableau, Power BI, or Excel.
* Conducted data validation and integrity checks using SQL constraints, data profiling, and data cleansing techniques to ensure data accuracy and consistency.
* Collaborated with cross-functional teams to define key performance indicators (KPIs) and metrics, and developed SQL queries to measure and monitor performance.
* Conducted ad hoc data analysis using SQL to answer specific business questions and provide data-driven recommendations.
* Configured Flume to stream data into HDFS and Hive using HDFS Sinks and Hive sinks.
* Designed and implemented MapReduce-based large-scale parallel relation-learning system.
* Developed an application for filtering out data from consumer database and creating a deduped dataframe to store the latest and greatest data in the Hive tenants.
* Created several Databricks Spark jobs with PySpark to perform several tables to table operations.
* Developed automatic job flows and ran through Oozie daily and when needed which runs MapReduce jobs internally.
* Designed interfaces using Boom where quality and performance were of utmost importance.

**Environment**: Apache Sqoop, Apache Hive, Apache Spark, Apache Oozie, Apache HDFS, Hadoop, SQL, Tableau, Power BI, Microsoft Excel, Jenkins, Git, DevOps Pipelines, Flume, Data Ingestion Pipelines, MapReduce, PySpark, Databricks, Boom, Kerberos Authentication, Data Lake, KPIs, Metrics, Ad hoc Data Analysis, Data Validation, Data Migration.

***Client: VIZIENT INC, IRVING, TX June 2015 – Aug 2016***

***Role: ETL Developer / Datawerehouse Developer***

***Responsibilities:***

* Expertise in complete SDLC (Software Development Life Cycle) on various application domain involving different technologies varying from open systems to Data warehousing on AIX and Win NT platforms.
* Strong knowledge in Business Intelligence, Data warehousing and ETL Configuration support.
* Maintained data integrity by performing validation checks.
* This project involves application development and maintenance of new and existing application. Activities involved are starting from analysis, requirement gathering till code deployment to production.
* Optimized queries tuned and monitored server performance.
* Involved in Fine-tuning of database objects and server to ensure efficient data retrieval.
* Monitored performance and optimize SQL queries for maximum efficiency.
* Imported data from various sources into a Staging area, merges the data into the OLTP data model, and presents it to all levels of the over-600 users.
* Mentoring and training team members about the technical architecture and development process being used in the project.
* Developed, monitored and deployed SSIS packages and Conversant with software development life cycle (SDLC).
* Utilized SSIS (SQL Server Integration Services) to produce a Data Mart for reporting.
* Implemented SSIS data transformation scripts, created maintenance procedures and provided data integrity strategies.
* Responsible for creating the reports basing on the requirements using SSRS 2008.
* Create Minor development documents and Release Notes for all development activities
* Gathered business requirements, definition and design of the data sourcing and data flows, data quality analysis, and working in conjunction with the data warehouse architect on the development of logical data models.
* Deployed the generated reports on to the server.
* Assisted in migrating DE normalized report DB into a Star Schema design.
* Actively supported business users for change requests.
* Administered and enhanced SQL Server Integration Services (SSIS) data loads.
* Data blending implemented on EMPI databases and generated interactive dashboards
* Created ad-hoc reports to users in Tableau by connecting various data sources.
* Published Workbooks by creating user filters so that only appropriate teams can view it.

**Environment**: MSSQL Server 2012/2008, Visual Studio 2010. Integration Services (SSIS 2005/2008/2012) Business Intelligence Development Studio (BIDS), Crystal Reports, VB, C#, Office 2007, MS Access, ETL, MS Excel, Tableau, TFS (Team Foundation Server 2010).

***Client: Arrow Electronics Limited, Englewood, CO. Feb 2014 – May 2015***

***Role: ETL Developer***

***Responsibilities:***

* Expert in creating database objects such as tables, procedures, triggers, constraints, indexes and views.
* Used SQL profiler to view indexes performance to mostly eliminate table scan.
* Used highly complex T-SQL Queries and SQL Scripts to perform efficient data load based on complex Business rules.
* Involved in Normalization and De-Normalization of existing tables for faster query retrieval
* Created various DTS packages for data transfer and migration.
* Involved in analyzing and designing disaster recovery / replication strategies with business managers to meet the business requirements.
* Creation/ Maintenance of Indexes for fast and efficient reporting process.
* Created reports using stored procedure that involved in scheduling, creating snapshots and subscriptions for the reports (SSRS).
* Designed and developed matrix and tabular reports with drill down, drill through using SSRS.
* Created ad-hoc reports using SQL server 2005 Reporting Services (SSRS).
* Responsible for testing to optimize the procedures and triggers to be used in production.
* Involved in project planning also schedule for database module with project managers.
* Wrote T-SQL statements using DML and DDL for retrieval of the data.
* Created stored procedures, triggers in SQL Server.
* Created database objects like tables, views, indexes etc.
* Created and modified stored-procedures, triggers, T-SQL.
* Extract data from database and export to Excel files for Data Analyst to analyze the data.
* Scheduled Packages using SQL Server Agent Jobs to execute at specified time on weekly and monthly basis.

**Environment**: MS SQL Server 2008/20005, MS power point, MS Access, MS Excel, Windows 97/2000/XP.