Saran Kumar Durgam

Senior Data Engineer | Multi Cloud Certified

(940)-290-2722 sarandurgam9@gmail.com www.linkedin.com/in/saran-kumar-5795981b3

SUMMARY

- **Senior Data Engineer** and **ETL Lead** with 9+ years of experience in systems analysis, design, and development across Data Warehousing, Cloud Data Engineering, ETL/ELT pipeline leadership, Data Visualization, Reporting, and Data Quality Solutions, following Agile methodologies.
- Proficient in AWS services (S3, IAM, EC2, EMR, Kinesis, VPC, DynamoDB, Redshift, RDS, Lambda, Athena, Glue, DMS, QuickSight) and skilled in multi-cloud environments including Teradata and Azure Data Factory for large-scale ETL/ELT pipelines, with strong expertise in Informatica PowerCenter and proven experience leading ETL teams in design, development, and optimization.
- Proven success in building advanced analytics and ML-enabled solutions for enterprise-scale environments, with additional multi-cloud experience in AWS and Azure.
- Hand on experience in unified Data Analytics with Databricks, Databrick Workspace UI, Databricks Notebook Management, Delta Lake with Spark SQL.
- Experience in implementing various Big Data Analytical, Cloud Data engineering, and Data Warehouse / Data Mart, Data Visualization, Reporting, Data Quality, and Data virtualization solutions.
- Have proven track record of working as Data Engineer on Amazon cloud services, Google Cloud Platform, Snowflakes, Bigdata/Hadoop Applications, and product development.
- Strong experience with AWS Databases (Aurora, Redshift, DynamoDB, Elastic Cache) and Teradata MPP architecture including BTEQ, FastLoad, Multiload, and advanced indexing.
- Experience in job/workflow scheduling and monitoring tools like Oozie, AWS Data pipeline & Autosys.
- Worked with Azure services like HDInsight, Stream Analytics, Active Directory, Blob Storage, Cosmos DB, Azure Data Lake, Azure Storage, Azure SQL, Azure DW, and Azure Databricks.
- Experience working on creating and running Docker images with multiple micro-services.
- Good experience in deploying, managing, and developing with MongoDB clusters.
- Docker container orchestration using ECS, ALB and Lambda.
- Good experience in building real-time data pipelines using Kafka and Spark Streaming.
- Involved in building Data Models and Dimensional Modeling wif 3NF, Star and Snowflake schemas for OALP and Operational data store applications.
- Experience with Unix/Linux systems with scripting experience and building data pipelines.
- Responsible for migration of application running on premise onto Azure cloud.
- Experience in Cloud Databases and Data warehouses (SQL Azure and Confidential Redshift/RDS).
- Proficiency in multiple databases like MongoDB, Cassandra, MySQL, ORACLE, and MS SQL Server.
- Consulting on Snowflake Data Platform Solution Architecture, Design, Development, and Deployment to promote data-driven culture across enterprises.
- Proficient in batch processing of the data via MongoDB, Solr, Stream processing of the data via Strom API, Java.
- Expertise in the full life cycle of ETL (Extraction, Transformation, and Loading) using Informatica power center, Azure Data Factory and Apache Airflow.
- Played a key role in migrating Cassandra, Hadoop cluster on AWS and defined different read/write strategies.
- Strong SQL development skills include writing Stored Procedures, Triggers, Views, and User Defined functions.
- Experienced in delivering secure, compliant (PCI-DSS, GDPR) cloud-based data engineering solutions for high-volume transactional data in financial and healthcare domains.

CERTIFICATIONS

- Workloads and Machine Learning on Azure
- Google Cloud Foundations
- AWS Certified Machine Learning Engineer Associate (MLA-C01)
- Academy Accreditation - Databricks Fundamentals
- Power BI Essential Training

SKILLS

Programming & Scripting: Python, SQL, R, Java, HiveQL, Shell Scripting, Scala, COBOL
Mainframe & Legacy Systems: IBM Mainframe z/OS, JCL, DB2, IMS DB, CICS, VSAM

Analytics & Reporting: Power BI, Tableau, Business Objects, Crystal Reports, SSIS, SSRS, SSAS, SAS
Big Data Technologies: Apache Spark, Hadoop, Kafka, HDFS, Hive, Pig, HBase, Sqoop, Flume, Airflow, Oozie
ETL Tools: Informatica PowerCenter, Talend, AWS Glue, Azure Data Factory, DBT, StreamSets, SSIS
Cloud Platforms: AWS (EC2, S3, Glue, Athena, DynamoDB, Redshift), Azure (ADF, Databricks, ADLS Gen2), GCP (BigQuery, Dataflow, Pub/Sub)
Databases: Snowflake, Teradata, Oracle, SQL Server, MySQL, PostgreSQL, MongoDB, Cassandra
Other Tools: TOAD, BTEQ, Teradata SQL Assistant, SQL Loader, MS Project, MS Visio
Methodologies: Agile (Scrum), SDLC, Sprint Planning, UAT Support, Code Reviews, Documentation Best Practices

EXPERIENCE

Client: The Cigna Group, Plano, TX **January 2024 – Present**

Role: Senior Data Engineer

Roles and Responsibilities

- Designed and setup Enterprise Data Lake to provide support for various use cases including storing, processing, analytics, and reporting of voluminous, rapidly changing data by using various AWS services.
- Used various AWS services including S3, EC2, AWS Glue, Athena, Redshift, EMR, SNS, SQS, DMS, Kinesis.
- Extracted data from multiple source systems S3, Redshift, RDS and created multiple tables/databases in Glue Catalog by creating Glue Crawlers.
- Created AWS Glue crawlers for crawling the source data in S3 and RDS.
- Created multiple Glue ETL jobs in Glue Studio and then processed the data by using different transformations and then loaded into S3, Redshift, and RDS.
- Created multiple recipes in Glue DataBrew and then used in various Glue ETL jobs.
- Designed and developed ETL processes in AWS Glue to migrate data from external sources like S3, Parquet/Text files into AWS Redshift.
- Used AWS Glue Catalog with Crawler to get the data from S3 and perform SQL query operations using AWS Athena.
- Designed and developed scalable, fault-tolerant Kafka-based data pipelines using Apache Kafka, Kafka Streams, and Kafka Connect for real-time healthcare data ingestion and processing.
- Developed Snowflake views to load the unload data from S3 bucket and deploying the code in Production.
- Written PySpark job in AWS Glue to merge data from multiple tables and in utilizing Crawler to populate AWS Glue Data Catalog with metadata table definitions.
- Used AWS Glue for transformations and AWS Lambda to automate the process.
- Used AWS EMR to transform and move large amounts of data into and out of AWS S3.
- Created monitors, alarms, notifications, logs for Lambda functions, Glue Jobs using CloudWatch.
- Performed end-to-end architecture & implementation assessment of various AWS services like Amazon EMR, Redshift, and S3.
- Used AWS EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.
- To analyze the data vastly, used Athena to run multiple queries on processed data from Glue ETL jobs and then used QuickSight for visualization and reporting.
- Used AWS EMR to transform and move large amounts of data into and out of AWS S3.
- Used DMS to migrate tables from homogeneous and heterogeneous DBs from on-premise to AWS Cloud.
- Created Kinesis Data Streams, Kinesis Data Firehose, and Kinesis Data Analytics to capture and process streaming data and then output into S3, DynamoDB, and Redshift for storage and analysis.
- Created Lambda functions to run the AWS Glue job based on the AWS S3 events.
- Coordinated with legacy system teams to interpret COBOL/JCL-based data feeds from mainframe DB2 sources, creating detailed data mapping specifications for ingestion into AWS and Snowflake pipelines.

Client: Cisco Systems, San Jose, CA **January 2022 - July 2023**

Role: Senior Data Engineer

Project Details:

Roles and Responsibilities

- Led a team of ETL developers (Offshore), conducting code reviews, performance tuning, and enforcing best practices for Informatica PowerCenter mappings and workflows, while ensuring smooth coordination between Offshore and Onsite teams.
- Created scripts to read CSV, JSON, and Parquet files from S3 buckets in Python and load into AWS S3, DynamoDB, and Snowflake.
- Designed and built scalable batch and streaming data pipelines using DBT, PySpark, and AWS Glue, ingesting structured/unstructured data into Snowflake and Redshift.
- Developed DAGs using Apache Airflow to automate daily workflows, data quality checks, and change data capture jobs.

- Developed Scala scripts, UDFs using both data frames/SQL and RDD in Spark for data aggregation, queries, and writing back into S3 bucket.
- Tuned EMR and Spark cluster configurations for batch processing, leveraging AWS EC2, S3, and YARN for distributed compute efficiency.
- Implemented AWS Lambda functions to run scripts in response to events in Amazon DynamoDB table or S3 bucket, or to HTTP requests using Amazon API Gateway.
- Built automated CI/CD pipelines in Azure DevOps and GitHub Actions for DBT and ETL workflows. Added notifications, test validations, and rollback strategies using YAML configurations.
- Developed automated regression suites for ETL pipeline validation and data dashboard testing using Python and Playwright. Integrated test execution into CI/CD pipelines with email/slack reporting.
- Provided enriched datasets to BI teams using Tableau, enabling dynamic dashboards and KPI tracking for business insights.
- Wrote Spark applications for data validation, cleansing, transformation, and custom aggregation; used Spark engine and Spark SQL for data analysis, providing the data scientists with further analysis-ready datasets.
- Designed and built Kafka-based and Kafka-like ingestion pipelines using Apache Kafka, Kafka Connect, and Kafka Streams to process telemetry and lifecycle analytics data.
- Pulled data from Teradata and Oracle using Sqoop and processed in Spark for ingestion into Snowflake.
- Configured monitoring with AWS CloudWatch and custom Python logging modules to detect pipeline failures and performance degradation.
- Designed normalized and denormalized data models in Snowflake, optimized for analytic workloads with warehouse sizing and query tuning.
- Responsible for the execution of big data analytics, predictive analytics, and machine learning initiatives.
- Experience in data cleansing and data mining.
- Developed internal Python applications for data validation, anomaly detection, and scheduled reporting tasks. Scheduled execution via CloudWatch and orchestrated results back to S3 and BI dashboards.
- Managed access control using IAM roles and tags, ensuring data protection and regulatory alignment.
- Collaborated with cross-functional teams including product owners, QA engineers, and business analysts in Agile sprint cycles. Authored user stories, test scenarios, and acceptance criteria for analytics features.
- Participated in sprint planning, retrospectives, and UAT support to ensure data product delivery matched stakeholder expectations.

Client: Axis Bank, Bengaluru, India **October 2020 – January 2022**

Role: Data Analyst

Project Details:

Roles and Responsibilities

- Design and implement end-to-end data solutions (storage, integration, processing, and visualization) in Azure.
- Developed Python programs for manipulating the data reading from various Teradata and converting them as one CSV file.
- Designed and implemented COBOL and JCL batch processes on IBM Mainframe z/OS to handle high-volume mortgage and banking data from DB2 and IMS DB systems.
- 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, AzureSQL DW, HDInsight/Databricks, NoSQL DB).
- Building, Deployment, Configuration, Management of SPLUNK Cloud instances in a distributed environment which spread across different application environments belonging to multiple lines of business.
- Designed and Developed Informatica mappings and sessions based on business requirements and business rules to load the data from source flat files and RDBMS to target tables.
- Ingested huge volume and variety of data from disparate source systems into Azure Data Lake Gen2 using Azure Data Factory.
- Partnered with business analysts, data architects, and ETL teams to validate mappings using SQL queries, SAS datasets, and COBOL outputs, supporting reconciliation and audit requirements during data migration.
- Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) & Azure SQL DB).
- Collaborated with data teams to showcase the projects KPIs by using Big Data system Azure Data Lake, Scope, and Azure Data Explorer (Kusto).
- Transform data by running a Python activity in Azure Databricks.
- Architect and implement ETL and data movement solutions using Azure Data Factory, SSIS create and run SSIS Package ADF V2, Azure-SSIS IR.
- Worked with Terraform templates to automate the Azure IaaS virtual machines using Terraform modules and deployed virtual machine scale sets in production environment.
- Defined and documented data mapping specifications between mainframe source datasets and Teradata targets, ensuring business rules and mortgage domain logic were accurately applied.
- Built pipelines to move hashed and un-hashed data from Azure Blob to Datalake.

- Building/Maintaining Docker container clusters managed by Kubernetes Linux, Bash, GIT, Docker and Utilized Kubernetes and Docker for the runtime environment of the CI/CD system to build, test, deploy.
- Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database, and SQL Data Warehouse environment.
- Developed MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a Python-based distributed random forest via Python streaming.
- Release pipelines use Azure AD Application Registration service principal in the Azure DevOps service connections for authentication to Azure.
- Worked on data profiling and data validation to ensure the accuracy of the data between the warehouse and source systems.
- Used SQLs to test various reports and ETL load jobs in development, QA, and production.

Client: Landmark Group, Bengaluru, India **August 2019 – August 2020**

Role: Data Engineer

Roles and Responsibilities

- Implemented Apache Airflow for authoring, scheduling, and monitoring Data Pipelines.
- Designed several DAGs (Directed Acyclic Graphs) for automating ETL pipelines.
- Performed data extraction, transformation, loading, and integration in data warehouse, operational data stores, and master data management.
- Setting GitLab repository and Runner for build automation.
- Automation Code version management using GitLab.
- Built data pipelines in Airflow in GCP for ETL-related jobs using different Airflow operators.
- Experience in GCP Dataproc, GCS, Cloud Functions, BigQuery.
- Experience in moving data between GCP and Azure using Azure Data Factory.
- Used Cloud Shell SDK in GCP to configure the services Data Proc, Storage, BigQuery.
- Developed and deployed outcomes using Spark and Scala code in Hadoop cluster running on GCP.
- Migrated an entire Oracle database to BigQuery and used Power BI for reporting.
- Built a program with Python and Apache Beam and executed it in Cloud Dataflow to run data validation between raw source files and BigQuery tables.
- Created Python scripts to ingest data from on-premises to GCS and built data pipelines using Apache Beam and Dataflow for data transformation from GCS to BigQuery.
- Developed ELT jobs using Apache Beam to load data into BigQuery tables.
- Built data pipelines in Airflow in GCP for ETL-related jobs using different Airflow operators.
- Developed PySpark programs, created DataFrames, and worked on transformations.
- Transformed and analyzed the data using PySpark, Hive, based on ETL mappings.
- Developed PySpark scripts that run on MSSQL table pushes to Big Data where data is stored in Hive tables.
- Used cloud shell SDK in GCP to configure the services Data Proc, Storage, BigQuery.
- Worked on architecting the ETL transformation layers and writing Spark jobs to do the processing.
- Implemented one-time data migration of multi-state-level data from SQL Server to Snowflake by using Python.
- Worked on Confluence and Jira.

Client: Citibank, Bengaluru, India **January 2016 – July 2019**

Role: **Data Engineer**

Roles and Responsibilities

- Designed and developed ETL mappings using Informatica PowerCenter to integrate data from flat files, XML, and DB2 into Teradata.
- Extracted data from DB2 mainframes and loaded into Teradata using FastLoad, MultiLoad, and T-Pump and developed BTEQ scripts for complex business logic.
- Built and optimized SQL scripts, indexes, and queries to support data extraction and transformation for analytics workloads.
- Created and tuned workflows, sessions, and parameters in Informatica for efficient ETL execution.
- Developed Teradata and SQL-based solutions for high-volume data processing, ensuring data quality and performance
- Built and maintained SQL scripts, indexes, and complex queries for data analysis and extraction.
- Performed unit, integration, and system-level performance testing; associated with production support team in various performance-related issues.
- Created mappings using Aggregator, Expression, Joiner, Filter, Sequence, Procedure, Connected & Unconnected Lookup, Filter, and Update Strategy transformations using Informatica PowerCenter designer.
- Provided production support by monitoring the processes running daily.
- Worked on mapping parameters and variables for calculations done in aggregator transformation.
- Created, scheduled, and configured workflows, worklets, and sessions using Informatica Workflow Manager.