

Keerthi Rapolu

keerthiofrapolu@gmail.com

+ 1 (281) 802-0700

[linkedin.com/in/keerthi-r-4348833a](https://www.linkedin.com/in/keerthi-r-4348833a)

Location: Houston, TX

SUMMARY:

- Experienced Data Engineer with 13+ years of proficiency in Snowflake, Databricks, Oracle PL/SQL, and Mainframe Technologies.
- Skilled in building production-ready data pipelines and designing high-quality data warehousing solutions on large-scale platforms.
- Expert in migrating data from legacy databases to cloud-based systems.
- Led assessment and proof-of-concept (POC) for mainframe-to-Databricks migration, overseeing Control-M job assessments and Informatica evaluation.
- Experience with Data Migration projects from source (Oracle/SQL Server) to Snowflake.
- Experienced with Snowflake cloud data warehouse and AWS S3 bucket for integrating data from multiple source system which include loading CSV, Parquet and JSON formatted data into snowflake tables.
- Proficient in implementing a diverse range of database objects in Snowflake, encompassing tables, views, sequences, synonyms, stored procedures, functions, tasks, streams, stages, and data sharing configurations, with a focus on meeting business requirements and adhering to best practices.
- Experienced in utilizing dbt (Data Build Tool) for streamlining ETL processes and automating the creation of required tables and views to optimize data modeling and storage for analytics and reporting purposes.
- Utilized PySpark within Databricks to seamlessly migrate mainframe code to modern data environments, ensuring efficient data processing and analysis.
- Efficiently migrated data from Oracle to Snowflake with the seamless integration of Qlik Replicate. Successfully migrated data and code from mainframe systems to Databricks, leveraging delta live tables and automating data pipelines for efficiency.
- Experienced in adeptly crafting sophisticated SQL queries to extract, manipulate, and analyze data with precision and efficiency.
- Proficiently utilized automated testing tools such as Tosca to ensure the integrity and reliability of data. Proficient in developing Python-based scripts for reading/writing data to/from databases and creating automation scripts.
- Skilled in utilizing AWS services such as Redshift, CloudWatch, S3 buckets, Glue, and Lambda for various data management and processing tasks.
- Experience with Apache Airflow for orchestrating and managing complex data workflows, ensuring efficient task scheduling and automation.
- Skilled in utilizing Apache NiFi to efficiently extract data from Oracle sources, transform it into files, and securely transfer them to designated S3 buckets for storage and further processing.
- Experienced in working with Linux environments, proficiently navigating and utilizing its features for various software development and data processing tasks.
- Familiar with data visualization tools such as Tableau and Power BI, enabling effective representation and analysis of data for insightful decision-making.
- Have a comprehensive understanding of Azure cloud services.
- Proficient in PL/SQL, adept at developing and optimizing stored procedures, triggers, functions, and packages to facilitate efficient data manipulation and processing within Oracle databases.
- Experienced in mainframe technologies, including COBOL, JCL, VSAM, and CICS, with a track record of successfully managing and modernizing mainframe systems to meet evolving business needs.
- Experienced in IBM InfoSphere DataStage, proficiently utilizing its capabilities for ETL (Extract, Transform, Load) processes and data integration tasks in enterprise environments.
- Familiar with Informatica, possessing knowledge of its features and functionalities for data integration, ETL processes, and data management solutions.
- Familiar in Apache Kafka for real-time data streaming and event-driven architectures.
- Utilized Fivetran to streamline data integration from various sources into Azure environments.
- Implemented Splink for entity resolution, leveraging AI and ML to improve data accuracy and management.

KEY COMPETENCIES:

Snowflake	Databricks	AWS
Azure	Linux	PL/SQL
Data Build Tool	Pyspark	Tosca
DB2	PostMan	Python
Oracle	Qlik Replicate	SQL
Cobol & JCL	Airflow	Nifi
Datastage	Informatica	Kafka
Power BI	Unix	

EDUCATION:

- **Bachelor of Technology, Electronics and Communication Engineering**
Jawaharlal Nehru Technological University, Hyderabad

CERTIFICATIONS:

SnowPro Core
AWS Certified Solutions Architect – Associate

PROFESSIONAL EXPERIENCE:

Lead Data Engineer
Genpact Ltd.

Client: Kaiser Permanente, Houston, TX

May 2024 – Present

Environment: AWS (Connect, Glue, S3, Lambda, Bedrock, Lex, RDS), SQL, Python.

Accomplishments:

- Delivered a POC showcasing AWS Connect’s capabilities, successfully demonstrating intelligent call routing, sentiment detection, and integration with patient data. Earned positive feedback from stakeholders.
- Designed and implemented AWS Glue ETL pipelines to load CSVs from Amazon S3 into RDS MySQL, automating jobs and Modeled data into a star schema (facts & dimensions).
- Developed an Amazon Connect Agent Workspace “Patient 360” dashboard (**POC**), binding contact attributes to display schedules, labs, triage, and billing data using synthetic/de-identified patient records to ensure HIPAA compliance.
- Built AWS Lambda APIs (with VPC access) to fetch patient data from RDS and surface it into Connect flows/widgets; instrumented solutions with CloudWatch for monitoring.
- Designed and documented use cases highlighting Connect’s AI features, including a draft Patient 360 dashboard to visualize patient interactions and outcomes.
- Generated synthetic datasets and performed data modeling to support development of AI-driven analytics such as contact sentiment analysis, call evaluation, and customer-experience scoring.

Lead Data Engineer

Client: Landrys/Tachus, Houston, TX

Aug 2024 – May 2024

Environment: Databricks, Snowflake, Python, SQL, Pyspark, Azure DevOps, Fivetran, Splink, Azure, Azure ADF, SQL Server, SQL Server Integration Services (SSIS),

Accomplishments:

- Developed and optimized Delta Live Tables in Databricks to streamline data processing and improve pipeline performance for real-time streaming data.
- Built and deployed custom Python and PySpark functions for data cleaning, ensuring high-quality, transformed datasets for downstream analytics.
- Engineered and executed complex SQL queries for data transformation, enabling accurate reporting and data insights.
- Automated data pipeline deployment and management using Azure DevOps, reducing manual interventions and enhancing workflow efficiency.
- Integrated data sources into Azure environments by utilizing Fivetran, ensuring seamless and reliable data movement.
- Integrated Snowflake with Azure data pipeline architectures to automate data management tasks and enable continuous, real-time data processing across the organization.
- Designed and implemented scalable ETL pipelines using Snowflake, automating data ingestion, transformation, and loading processes for data lakes and warehouses.
- Designed, developed, and maintained scalable ETL pipelines using Azure Data Factory (ADF) to automate data ingestion, transformation, and loading into data lakes and warehouses.
- Designed and implemented Splink-based entity resolution processes using AI and ML techniques, significantly improving data accuracy and entity matching.
- Applied advanced machine learning techniques within Splink to accurately match and deduplicate records across multiple data sources, improving the integrity and consistency of business-critical data.
- Customized Splink's entity resolution algorithms to handle complex use cases, significantly reducing false positives and enhancing the precision of entity matching.
- Integrated Splink with Azure-based data pipelines, automating entity management tasks and ensuring continuous, real-time data cleansing across the organization.
- Designed and developed ETL processes using SSIS to streamline data integration across multiple systems, enhancing the data flow between on-premises and cloud environments.
- Integrated SSIS packages with SQL Server to manage data migration, synchronization, and integration processes across diverse data sources and systems.

Impetus Technologies

Client: Ally

Dec 2022 – Aug 2024

Module Lead Engineer

Environment: Snowflake, Oracle, AWS, Git-Lab, Qlik Replicate, Data Build Tool, Tosca, Data Stage, SQL, Python

Accomplishments:

- Developed comprehensive strategies aimed at efficiently transitioning on-premise Oracle databases to the Snowflake cloud environment.
- Conducted thorough evaluations of existing database schemas, data structures, and dependencies to identify potential migration challenges and devise appropriate mitigation strategies.
- Directed the migration of data from on-premise Oracle databases to Snowflake using Qlik Replicate, ensuring a seamless transition with minimal disruption to business operations.
- Engineered and deployed various database objects in Snowflake, including tables, views, sequences, Snow SQL synonyms, stored procedures, functions, tasks, streams, stages, regEx, and data sharing configurations, meticulously adhering to business requirements and industry best practices.
- Utilized Snowflake's intuitive web UI to efficiently manage database objects, optimizing data storage and accessibility for streamlined querying and analytics.
- Designed and implemented Snowflake data warehouse structures to effectively accommodate migrated data from Oracle databases, ensuring a smooth transition with minimal disruption to operations.

- Leveraged the capabilities of the Data Build Tool (dbt) to automate data transformation workflows and construct reusable data models within the Snowflake environment, enhancing efficiency and scalability.
- Implemented dbt models for incremental data loading, version control, and documentation, simplifying data pipeline development and maintenance processes.
- Migrated complex PL/SQL database objects, including stored procedures and functions, to Snowflake using dbt and Snowpark, ensuring efficient and seamless data transformation and management.
- Leveraged Snowpark to query and process data within the Snowflake Data Cloud, enhancing data analysis capabilities and operational efficiency.
- Crafted intricate SQL queries tailored to project specifications within the Snowflake environment, optimizing performance and scalability to handle large datasets efficiently.
- Implemented automated testing procedures using Tosca to validate data integrity and ensure compliance with business rules and regulatory requirements.

Impetus Technologies

Client: Norfolk Southern

Jan 2022 - Dec 2022

Module Lead Engineer

Environment: Databricks, Pyspark, JCL, COBOL, Python, Kafka, SQL, AWS, DB2, Azure

Accomplishments:

- Conducted comprehensive analysis of legacy systems, including JCL, COBOL, and DB2, to identify data and code components for migration to Databricks platform.
- Designed and implemented Delta Live Tables in Databricks to enable real-time data processing and analytics, leveraging the power of Delta Lake for ACID transactions and data versioning.
- Designed and implemented data pipelines using PySpark and Databricks Delta Lake, orchestrating data ingestion, transformation, and storage operations.
- Refactored and modernized COBOL codebase to leverage Python and PySpark functionalities within the Databricks environment, optimizing performance and maintainability.
- Collaborated with data architects to design and implement data lake architecture, consisting of gold layer for raw, unprocessed data and silver layer for curated and transformed datasets.
- Implemented data transformation workflows to populate the silver layer with cleansed and standardized data, ensuring consistency and accuracy for downstream analytics and reporting. Leveraged Delta Live Tables and Databricks notebooks to perform real-time analytics on streaming data, enabling stakeholders to gain timely insights and make data-driven decisions.
- Implemented unit tests and validation checks to ensure code quality and reliability throughout the migration process.
- Integrated Kafka streaming platform with Databricks for real-time data ingestion and processing, enabling near-real-time analytics and insights generation.
- Designed and implemented Delta Live Tables in Databricks for real-time data processing and analytics, enabling continuous data ingestion and analysis.

Capricorn Software India Pvt. Ltd

Client: Meridian Cooperative

Feb 2019 - Jan 2022

Senior Software Engineer

Environment: Snowflake, Oracle, AWS, SQL, Nifi, Python, Glue, Lambda

Accomplishments:

- Designed and implemented data migration processes from Oracle to Snowflake using Snowpipe, ensuring efficient and real-time data ingestion into the Snowflake data warehouse.
- Utilized Apache NiFi to orchestrate the creation of files from on-premise Oracle sources and transfer them to Amazon S3 for storage and further processing.
- Implemented Airflow for workflow orchestration, scheduling, and monitoring of data migration tasks, ensuring seamless and automated execution of data pipelines.
- Gathered requirements from product-application team members, utilizing JIRA and Confluence for documentation and tracking of project progress.
- Created and managed database objects such as Databases, Schemas, Tables, Views, Sequences, Stored Procedures, File Formats, Storage Integration, Stages, and Snowpipe in Snowflake, aligning with business requirements.
- Created Lambda functions to trigger AWS Glue jobs based on S3 events.

- 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.
- Interacted with onsite teams via Email and Slack calls to clarify design requirements and ensure alignment with project objectives.
- Developed and executed complex SQL queries as per project needs, optimizing performance and scalability to handle large datasets effectively.
- Conducted end-to-end tests to validate the functionality of data migration processes, ensuring all changes met user requirements and adhered to quality standards.

Capricorn Software India Pvt. Ltd

Client: SEDC

Dec 2016 - Jan 2019

Software Engineer

Environment: Oracle, MF Cobol, .Net, Linux

Accomplishments:

- Successfully translated complex design specifications into detailed technical documents, facilitating effective communication and collaboration between development teams and stakeholders.
- Played a key role in gathering and analyzing requirements, contributing to the successful delivery of projects within specified timelines and budget constraints.
- Developed code for new features, enhancements, and bug fixes across multiple technologies including Oracle, MF Cobol, .Net, and Linux environments.
- Developed and implemented shell scripts to automate tasks using Unix
- Conducted thorough unit testing of developed code to validate functionality, identify defects, and ensure compliance with quality standards and coding best practices.

Blueliner Software Services Pvt. Ltd.

Client: Mediterranean shipping company

Jan 2012 - Nov 2016

Software Engineer

Environment: COBOL, JCL, DB2, VSAM, SPUFI, TSO/ISPF, SQL, SPUFI, CICS, CHANGE MAN, XPEDITOR AND FILE-AID

Accomplishments:

- Successfully developed and tested COBOL programs integrated with DB2 databases, contributing to the timely delivery of project milestones.
- Ensured compliance with coding standards and best practices, resulting in high-quality and maintainable codebase.
- Collaborated effectively with team members to troubleshoot issues and resolve technical challenges during the development and testing phases.
- Conducted comprehensive unit testing of developed programs to verify functionality, identify defects, and ensure code quality and reliability.