Shesha sai Yerragolla

✉ sheshasai72@gmail.com ▯ +1 8154960777 in linkedin.com/in/shesha-sai-

SUMMARY

* 10 years of progressive experience as a Senior ETL Developer, specializing in data analytics, Data Engineer, and large-scale data warehousing.

* Led the design and implementation of scalable ETL solutions on Azure and AWS for leading banking and healthcare institutions, utilizing tools like Azure Data Factory, PySpark, AWS S3, AWS Redshift, AWS Lambda, AWS Glue, and Informatica PowerCenter.

* Developed and maintained data warehouse solutions using MongoDB, Oracle, Cassandra, Amazon DocumentDB, and Amazon RDS for PostgreSQL. Created interactive dashboards using Tableau, PowerBI, and QlikView.

* Implemented Kafka-based and Amazon Kinesis Data Streams-based architectures for real-time processing. Managed Azure SQL Database, MySQL, Amazon RDS for PostgreSQL, and Microsoft SQL Server database environments.

* Utilized Azure DevOps, AWS CodeCommit, and GitHub for version control and automated ETL deployment processes using Jenkins and AWS CodePipeline. Managed Airflow to schedule and monitor workflows.

* Proficient in SQL for comprehensive data analysis across various database management systems including Teradata, Oracle, and SQL Server.

* Extensive experience in data warehousing environments, ensuring efficient data storage, retrieval, and analysis.Skilled in writing SQL scripts specifically tailored for Teradata, facilitating optimized data querying and manipulation within the platform.

* Extensive experience in designing and implementing scalable ETL solutions utilizing Redshift, ensuring efficient data processing and analysis for banking and healthcare institutions.

* Proven track record of successfully implementing SQL scripts for Teradata to efficiently manage and manipulate large datasets, ensuring data accuracy and integrity.

* Demonstrated ability to leverage SQL expertise to optimize data processes and enhance overall system performance within complex data environments.

* Implemented SQL-based performance monitoring and optimization strategies to identify and address bottlenecks in database queries and ETL processes, resulting in a 20% increase in overall system efficiency and data processing speed.

* Utilized R and Python, specifically the Pandas and NumPy libraries, for advanced data analytics. Utilized Apache Hive for querying and managing large data sets. Employed Docker for containerization.

* Utilized SQL data analysis techniques to conduct in-depth exploratory data analysis (EDA) on large datasets, uncovering actionable insights and trends that informed strategic decision-making processes, resulting in a 25% improvement in data-driven strategies and business outcomes.

* Maintained high standards of data quality and integrity. Adopted best practices in data management and security and collaborated closely with cross-functional teams to align data projects with business objectives.

* Provided technical support and guidance to business teams. Demonstrated strong leadership and project management abilities.

EXPERIENCE

Senior ETL Developer | Total Bank Solutions

April 2022 - Present, US, NJ, Hackensack

- Roles & Responsibilities:
- Developed and implemented SQL-based ETL solutions on AWS for a leading banking institution, utilizing SQL queries for data extraction, transformation, and loading.
- Conducted SQL performance tuning and optimization of AWS resources, including scaling Amazon Redshift instances dynamically for efficient handling of peak data loads, resulting in a 25% improvement in data processing speed.
- Utilized SQL for managing and querying data in Hadoop clusters, optimizing HDFS for distributed data storage, and using SQL-based queries for MapReduce parallel data processing, leading to a 30% improvement in data storage and processing efficiency.
- Created SQL-based queries to generate interactive Amazon QuickSight dashboards for analyzing banking transaction data, providing valuable insights into customer spending habits and facilitating informed business decisions based on real-time data, resulting in a 10% increase in data-driven decisions.
- Designed and implemented SQL-based solutions for an Amazon S3-based data lake, enabling the storage and analysis of unstructured banking data, providing real-time analytics and personalized customer experiences, and improving data accessibility by 30%.
- Integrated SQL queries with ETL processes to efficiently ingest and transform large volumes of banking data stored in Amazon S3, resulting in a significant 40% improvement in data processing speed and reduced data latency.
- Developed and maintained SQL-based database schemas in Amazon Redshift for storing transactional data in a banking environment, ensuring high data integrity and reliability, and facilitating efficient data retrieval and analysis, leading to a 20% reduction in data retrieval time.
- Implemented SQL-based solutions for managing and scaling containerized ETL applications using Kubernetes, improving system reliability and resource utilization, leading to a 30% increase in application uptime and a 20% reduction in infrastructure costs.
- Leveraged SQL for managing various AWS services and applications in a Linux environment, enhancing system performance and reliability by 15%.
- Implemented SQL and PL/SQL scripts for data cleansing and transformation, optimizing ETL workflows, and improving data quality and analysis accuracy, resulting in a 30% improvement in data quality scores.
- Configured SQL-based Informatica PowerCenter workflows for extracting, transforming, and loading banking data from heterogeneous sources into the data warehouse, and implemented Informatica Data Quality for enhancing data quality and consistency across banking systems, resulting in a 20% improvement in data consistency scores.

- Developed SQL-based solutions within a Flask-based architecture for real-time processing of banking transactions, providing a scalable and efficient solution for handling transactional data streams.

ETL Developer | Finish Line
June 2019 - April 2022, US, IN, Indianapolis

- Roles & Responsibilities:
- Developed a cloud-based ETL solution for a retail chain using SQL-based queries and scripts in AWS Glue, PySpark, and Amazon Redshift, optimizing data workflows and improving efficiency by 35% for timely insights and decision-making.
- Utilized Redshift for optimized data storage and processing, ensuring high performance and scalability of the SQL-based ETL solution.
- Implemented AWS CodePipeline for version control and automated deployment of SQL-based ETL pipelines, reducing deployment time by 40% and doubling the deployment frequency for more agile responses to business needs.
- Optimized AWS infrastructure performance, including dynamic scaling of Amazon Redshift clusters, improving data processing speed by 30% through SQL performance tuning and optimization.
- Managed Amazon RDS database environments with SQL skills, ensuring 99.9% data integrity and reliability for high-stakes financial transactions and critical business operations.
- Developed and maintained SQL-based stored procedures and triggers in Amazon RDS for data transformation and validation, reducing data processing errors by 15% and ensuring accurate and reliable data for business analysis.
- Configured AWS Glue workflows for ETL with SQL queries and scripts, enhancing data quality and consistency across retail systems by 20%.
- Utilized SQL-based data modeling techniques to design and optimize database schemas in Amazon Redshift, ensuring efficient data storage and retrieval for analytical queries, leading to a 25% improvement in query performance.
- Implemented SQL-based data governance policies and procedures to enforce data quality standards and regulatory compliance across the retail chain's data ecosystem, ensuring data consistency and integrity for informed decision-making and regulatory requirements adherence.
- Deployed and managed Hadoop clusters for distributed data storage and processing, utilizing SQL-based queries for data manipulation and optimization, resulting in a 40% improvement in data storage and processing efficiency.
- Implemented AWS CloudWatch for robust SQL-based data pipeline monitoring, reducing troubleshooting time by 30% and ensuring uninterrupted data flows for business operations.
- Leveraged Linux for the deployment and management of various AWS services and applications, enhancing system performance and reliability by 15% with SQL expertise in managing and optimizing AWS resources.

Data Engineer | IQVIA
February 2018 - June 2019, US, NC, Durham

- Roles & Responsibilities:
- Implemented a robust healthcare analytics platform on Azure, utilizing SQL queries and scripts in Azure Data Factory for ETL orchestration and Azure SQL Database for data processing, resulting in a 20% increase in data processing efficiency.
- Managed Azure SQL Database for healthcare data warehousing with SQL expertise, optimizing schemas and utilizing advanced features for enhanced performance and data querying, leading to a 15% increase in query performance.
- Implemented Azure Synapse Analytics for healthcare analytics, leveraging SQL queries for structured and semi-structured data analysis and benefiting from its automatic scaling capabilities, resulting in a 20% improvement in analytics speed.
- Utilized Azure Stream Analytics for real-time processing of healthcare data streams with SQL skills, reducing response time by 30% and enabling timely analysis and response to critical events such as patient monitoring alerts.
- Leveraged Azure Repos for version control and Azure Pipelines for automated deployment, ensuring code integrity and facilitating continuous integration and delivery of healthcare data solutions with SQL proficiency.
- Developed interactive PowerBI dashboards for visualizing healthcare performance metrics and patient outcomes, utilizing SQL queries for data retrieval and aggregation, leading to a 30% improvement in decision-making speed for healthcare providers.
- Enhanced efficiency and data accessibility by integrating Azure Cosmos DB with MongoDB for flexible storage of EHRs, utilizing SQL scripts for data transformation and analysis, resulting in a 25% improvement in structured and semi-structured data analysis.

SQL Developer | Infra Soft Tech
January 2016 - September 2017, IN, MAHARASHTRA, Pune

- Roles& Responsibilities:
- Developed and maintained Oracle Database systems for optimized data storage, managing over 500GB of data. Utilized PL/SQL for advanced database programming and data manipulation, improving data processing efficiency by 30%.
- Implemented data solutions using Informatica PowerCenter for effective ETL processes, processing over 1TB of data daily. Integrated Apache Kafka for real-time data streaming and processing, handling over 10,000 events per second.
- Utilized R for data analysis and insights generation, improving model accuracy by 20%. Managed Jenkins for continuous integration and deployment of data projects, deploying over 100 jobs across development, testing, and production environments daily.
- Developed data visualizations and dashboards using Tableau, serving over 50 users across the organization. Employed Docker for containerization of data applications, managing over 50 containers for various data applications.

- Applied analytics and visualization techniques to support business decision-making, influencing over 10 major business decisions per quarter. Enhanced SQL skills for complex querying and data manipulation, handling over 1000 queries per day.
- Collaborated with teams to align data projects with business objectives, working with over 5 teams across the organization. Maintained high standards of data quality and integrity, reducing data errors by 15%. Implemented best practices in data management and security, passing 100% of security audits.
- Environment: Oracle Database, SQL, Informatica PowerCenter, Apache Kafka, R, Jenkins and Docker

Data Analyst | CSS Corp Pvt Ltd
March 2014 - January 2016, IN, TAMIL NADU, Chennai

- Roles & Responsibilities:
- Developed and maintained Oracle Database systems for optimized data storage, managing over 500GB of data, resulting in improved data processing efficiency by 30% through the utilization of PL/SQL for advanced database programming and data manipulation.
- Implemented data solutions using Informatica PowerCenter for effective ETL processes, processing over 1TB of data daily, and integrated Apache Kafka for real-time data streaming and processing, handling over 10,000 events per second.
- Utilized R for data analysis and insights generation, improving model accuracy by 20%, and managed Jenkins for continuous integration and deployment of data projects, deploying over 100 jobs across development, testing, and production environments daily.
- Developed data visualizations and dashboards using Tableau, serving over 50 users across the organization, and employed Docker for containerization of data applications, managing over 50 containers for various data applications.
- Applied analytics and visualization techniques to support business decision-making, influencing over 10 major business decisions per quarter, and enhanced SQL skills for complex querying and data manipulation, handling over 1000 queries per day.
- Developed and maintained web applications using Django framework for optimized data management and user interaction, integrating with Oracle Database systems to handle over 500GB of data. Implemented Django ORM for efficient database operations, resulting in improved data processing efficiency by 30% through advanced database programming and data manipulation with PL/SQL.
- Collaborated with teams to align data projects with business objectives, working with over 5 teams across the organization, while maintaining high standards of data quality and integrity, reducing data errors by 15%, and implementing best practices in data management and security, passing 100% of security audits.

EDUCATION
*Bachelor of Technology (B.Tech) in Computer Science from JNTUH, Hyderabad, Telangana, India. - 2014
2014

- LinkedIN:

SKILLS
TECHNICAL SKILLS:

*Data Warehousing: Amazon Redshift, Amazon RDS, Oracle Database, Azure SQL Database.

*Big Data Technologies: AWS Glue,AWS S3, AWS Redshift,AWS Lambda, PySpark, Azure Data Factory.

*ETL Tools: AWS Glue, PySpark, Informatica PowerCenter, Azure Data Factory.

*Programming Languages: Python, R.

*Cloud Platforms: AWS, Azure.

*Data Streaming: Amazon Kinesis, Kafka, Azure Stream Analytics.

*Data Visualization & BI Tools: Amazon QuickSight, PowerBI.

*Database Systems: Amazon DynamoDB, Amazon RDS, Oracle Database, Azure Cosmos DB (compatible with MongoDB),SQL scripts for Teradata.

* Project Management: Agile methodologies, Workflow management (Airflow).