

# Shesha sai Yerragolla

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

## SUMMARY

With 10 years of progressive experience as a Senior ETL Developer, I specialize in data analytics, Data Engineering, and large-scale data warehousing. I have 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 Glue, and Informatica PowerCenter for ETL orchestration. I have developed and maintained data warehouse solutions using MongoDB, Oracle, Cassandra, Amazon DocumentDB, and Amazon RDS for PostgreSQL, ensuring data integrity, reliability, and efficient retrieval and analysis of structured and semi-structured data. I have created interactive dashboards using Tableau, PowerBI, and QlikView for analyzing transaction data, healthcare performance metrics, and business insights. I have implemented Kafka-based and Amazon Kinesis Data Streams-based architectures for real-time processing of transactions and healthcare data streams. I have managed Azure SQL Database, MySQL, Amazon RDS for PostgreSQL, and Microsoft SQL Server database environments and developed and maintained Oracle Database systems for optimized data storage. I have utilized Azure DevOps, AWS CodeCommit, and GitHub for version control and automated ETL deployment processes using Jenkins and AWS CodePipeline. I have managed Airflow to schedule and monitor workflows, enhancing the efficiency and reliability of ETL processes. I have utilized R and Python, specifically the Pandas and NumPy libraries, for advanced data analytics, enabling the extraction of meaningful insights from complex datasets. I have utilized Apache Hive, a data warehouse infrastructure built on top of Hadoop, for querying and managing large data sets, allowing for the handling of big data and the execution of queries in a distributed storage environment. I have employed Docker, a leading platform in containerization, to package applications with their dependencies and libraries, ensuring consistency across multiple development, testing, and production environments, leading to streamlined deployment and scaling processes. I have maintained high standards of data quality and integrity by implementing rigorous validation and cleaning processes. I have adopted best practices in data management and security, including access controls, encryption, and regular audits, to protect sensitive information and comply with regulatory standards. I have collaborated closely with cross-functional teams to align data projects with business objectives, fostering a data-driven culture within the organization. I have provided technical support and guidance to business teams on data-related matters, helping them understand and utilize data effectively for decision-making. In addition to my technical skills, I have demonstrated strong leadership and project management abilities. I have successfully led cross-functional teams and managed multiple high-stakes projects concurrently, ensuring timely and efficient delivery of data solutions.

## EXPERIENCE

### Senior ETL Developer | Total Bank Solutions | US, NJ, Hackensack | April 2022 - Present

- Roles & Responsibilities:
- Led the development and implementation of an Azure-based ETL solution for a leading banking institution, leveraging Azure Data Factory and Pyspark to orchestrate data workflows and Azure SQL Database for data storage and processing, resulting in a **20% increase** in data processing efficiency.
- Implemented Azure DevOps for version control and CI/CD pipelines, facilitating automated testing and deployment of Pyspark ETL jobs, resulting in a **50% reduction** in deployment time.
- Conducted performance tuning and optimization of Azure resources, including scaling Azure SQL Database instances dynamically to handle peak data loads efficiently, reducing peak load processing time by **30%**.
- Created interactive Tableau dashboards for analyzing banking transaction data, providing insights into customer spending habits, and enabling executives to make informed business decisions, leading to a **15% increase** in data-driven decisions.
- Designed and implemented a MongoDB-based data warehouse solution for storing and analyzing unstructured banking data, enabling real-time analytics and personalized customer experiences, improving customer experience scores by **25%**.
- Integrated MongoDB with ETL processes to ingest and transform large volumes of banking data efficiently, improving data processing speed by **40%**.
- Developed and maintained Oracle database schemas for storing transactional data in a banking environment, ensuring data integrity and reliability, reducing data anomalies by **20%**.
- Implemented Oracle PL/SQL and Pyspark scripts for data cleansing and transformation, optimizing ETL workflows and improving overall data quality, leading to a **30% improvement** in data quality scores.
- Configured and optimized Informatica PowerCenter workflows for extracting, transforming, and loading banking data from heterogeneous sources into the data warehouse, increasing data load efficiency by **35%**.
- Implemented Informatica Data Quality for cleansing and standardizing customer data, ensuring accuracy and consistency across banking systems, improving data consistency scores by **20%**.
- Implemented Kafka-based streaming architecture for real-time processing of banking transactions, enabling instant fraud detection and prevention, reducing fraud incidents by **50%**.
- Managed Airflow to schedule and monitor workflows, enhancing the efficiency and reliability of ETL processes, improving workflow reliability by **30%**.

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

- Roles & Responsibilities:
- Spearheaded the development of a cloud-based ETL solution for a prominent retail chain, utilizing Azure Data Factory and Azure SQL Database for efficient data storage and processing.
- Achieved a **50% reduction** in deployment errors, enhancing overall system stability. Processed and transformed 1TB of retail data daily, ensuring timely insights and decision-making.
- Implemented Azure DevOps for streamlined version control and automated deployment of ETL pipelines, enhancing team collaboration. Reduced deployment time by **40%** through automation, improving development efficiency.
- Executed performance optimization strategies on Azure infrastructure, including dynamic scaling of Azure SQL Database instances. Achieved a **20% increase** in database performance through dynamic scaling.
- Designed and deployed interactive Tableau dashboards for retail analytics, empowering stakeholders with actionable insights. Provided insights into sales trends and customer behavior, leading to a **15% increase** in revenue.
- Engineered a Cassandra-based data repository for storing and analyzing large volumes of retail transaction data. Handled over **100 million** transactions daily, ensuring scalability and performance.
- Managed MySQL database environments for retail inventory management systems, ensuring data integrity and reliability. Maintained **99.9%** database uptime, minimizing disruptions to inventory management operations.
- Developed MySQL stored procedures and triggers for data transformation and validation, optimizing ETL processes. Reduced data processing time by **30%** through optimized transformations.
- Implemented NiFi-based data ingestion pipelines for real-time streaming of retail sales data and orchestrated NiFi dataflows for data enrichment and transformation. Processed **10,000 transactions** per second in real-time, providing up-to-date insights.
- Utilized GitHub for version control and collaborative development of ETL codebase, ensuring code quality. Managed **500+ commits** per month, maintaining codebase integrity and version history.
- Automated ETL deployment processes using Jenkins, improving deployment reliability and reducing manual intervention.

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

- Roles & Responsibilities:
- Designed and implemented a scalable healthcare analytics platform on AWS, utilizing AWS Glue for ETL orchestration and Amazon Redshift for data processing, resulting in a **20% increase** in data processing speed.
- Developed interactive PowerBI dashboards for visualizing healthcare performance metrics and patient outcomes, facilitating data-driven decision-making for healthcare providers, leading to a **15% improvement** in patient outcomes.
- Implemented Amazon DocumentDB (compatible with MongoDB) for flexible storage of electronic health records (EHRs), ensuring efficient retrieval and analysis of structured and semi-structured healthcare data, reducing data retrieval time by **25%**.
- Managed Amazon RDS for PostgreSQL databases for healthcare data warehousing, optimizing schemas and utilizing advanced features for enhanced performance and data querying, resulting in a **30% increase** in query performance.
- Leveraged Amazon Redshift for healthcare analytics, benefiting from built-in support for structured and semi-structured data formats and automatic scaling capabilities, leading to a **40% reduction** in data processing time.
- Implemented Amazon Kinesis Data Streams for real-time processing of healthcare data streams, enabling timely analysis and response to critical events such as patient monitoring alerts, reducing response time by **50%**.
- Utilized AWS CodeCommit for version control and AWS CodePipeline for automated deployment, ensuring code integrity and facilitating continuous integration and delivery of healthcare data solutions, leading to a **35% increase** in deployment efficiency.
- Engineered an end-to-end ETL pipeline using AWS Glue to extract data from various sources, transformed it to fit operational needs, loaded into Amazon Redshift, allowing for complex queries and data models visualized through PowerBI, leading to a **30% increase** in decision-making efficiency.

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

- 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 Dockers.

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

- Roles & Responsibilities:
- Utilized Microsoft Excel for data analysis and reporting, processing over **10,000 rows** of data daily. Employed R and Python (Pandas, NumPy) for advanced data analytics, improving model accuracy by **20%**.
- Developed interactive dashboards and reports using QlikView and Tableau, serving over **50 users** across the organization. Managed Microsoft SQL Server databases for data storage and retrieval, handling over **1TB** of data. Utilized Apache Hive for querying and managing large data sets, processing over **500GB** of data daily.
- Employed Git for version control in data projects, managing over **1000 commits**. Translated business requirements into technical documentation and solutions, producing over **200 pages** of documentation. Applied analytics skills to derive insights from complex data sets, leading to a **15%** increase in business efficiency.
- Collaborated with cross-functional teams on data-driven projects, working with over **5 teams** across the organization. Maintained high data quality standards and best practices, reducing data errors by **30%**. Engaged in continuous improvement of data processes and systems, improving system performance by **25%**.
- Participated in training and professional development in data technologies, completing over **50 hours** of training. Provided data support and guidance to business teams, supporting over **100 users** across the organization.
- Environment: Microsoft, R, Python, Qlikview, Tableau, Microsoft SQL Server, Apache Hive, Git.

---

## **EDUCATION**

### **\*Bachelor of Technology (B.Tech) in Computer Science from JNTUH, Hyderabad, Telangana, India. – 2014 | 2014**

- LinkedIn:

---

## **SKILLS**

### **TECHNICAL SKILLS:**

- \*Data Warehousing: Azure SQL Database, Amazon Redshift, Oracle Database, Microsoft SQL Server, PostgreSQL.
- \*Big Data Technologies: Apache Hive, Hadoop, Pyspark.
- \*ETL Tools: Informatica PowerCenter, Azure Data Factory, SSIS, Pyspark, Airflow, AWS Glue.
- \*Programming Languages: Python, R, Scala
- \*Cloud Platforms: Azure, AWS
- \*Data Streaming: Apache kafka, Kinesis, NiFi.
- \*Data Visualization & BI Tools: QlikView, Tableau, Power BI.
- \*Database System: MongoDB, Cassandra, Couch DB, My SQL, Oracle Database, Microsoft SQL Server.
- \*Version Control System: GitHub, Azure DevOps
- \*Data Analytics & Machine Learning: Matplotlib, Numpy, Seaborn, Scikit, Keras, Tensorflow, Deep Learning.
- \*CI/CD Tools: Jenkins, Azure DevOps
- \*Project Management: Agile methodologies, Workflow management (Luigi)
- \*Miscellaneous: Business requirement documentation, Data modeling, Streaming, Extracting from multiple sources