# VENKATA SAI M

venkatasai.work30@gmail.com **|** +1(925)-553-4050

# PROFESSIONAL SUMMARY

* 10+ years of extensive IT experience specializing in **Data Engineering**, **Data Analytics**, and **Business Systems Analysis** across cloud and enterprise environments.
* Expert in **building scalable ETL/ELT pipelines** using **Azure Data Factory, Azure Databricks, Azure Synapse Analytics, AWS Glue, Redshift, Athena**, and **Microsoft Fabric**.
* Certified **SnowPro Core and Advanced Architect Professional** with hands-on experience in **Snowflake**, **DBT**, **Python**, and **AWS S3** integration.
* Strong background in **SQL programming (T-SQL, PL/SQL)**, **Spark SQL**, **PySpark**, and **Python scripting** for data transformation and big data processing.
* Extensive experience designing **medallion architecture** (Bronze, Silver, Gold layers) and implementing **data governance** and **security standards** using **Microsoft Purview**, **Azure Key Vault**, and **Azure Entra ID**.
* Skilled in developing **interactive dashboards** and **reports** using **Power BI**, **SAP Business Objects**, and **ThoughtSpot**.
* Proficient in **cloud-native data architecture** on **Azure** and **AWS**, including hands-on experience with **IAM roles, VPC setups, S3 storage management, and Lambda/Step Functions** automation.
* Deep expertise in **Informatica PowerCenter, IDQ, EDC, Axon, and IDMC/IICS**, including **installation, configuration, migration,** and **ETL optimization**.
* Strong experience in **Continuous Integration/Continuous Deployment (CI/CD)** processes using **Azure DevOps, Git, GitHub, ARM Templates**, and **Terraform**.
* Proficient in **writing complex dynamic SQL queries**, designing **Source-to-Target mappings**, and supporting large-scale **data warehouse solutions** and **business intelligence frameworks**.
* Skilled in documenting **Business Requirements (BRD)**, **Functional Specifications (FSD)**, **Gap Analysis**, and **UML-based process modeling** using **MS Visio**.
* Proven track record of collaborating extensively with **cross-functional teams** in **Agile/Scrum environments** to deliver high-quality, secure, and innovative data solutions.
* Passionate about **cloud data modernization**, **data governance**, **analytics engineering**, and transforming business requirements into actionable, scalable solutions.

# EDUCATION

**Dr. M.G.R University** Aug 2009 - Jul 2013

Bachelor’s Degree in Electronics and Communications Engineering

# TECHNICAL SKILLS

* **Programming Languages**: Python, SQL, Scala, R, HTML, CSS
* **Databases**: Azure SQL Database, Azure Synapse Analytics, Amazon Redshift, MySQL, Oracle SQL, DynamoDB, MongoDB
* **Data Engineering & Processing**: PySpark, Spark SQL, Apache Spark, SQL, T-SQL, PL/SQL, Data Transformation, Data Aggregation, ETL/ELT Pipeline Development
* **Cloud Platforms & Tools**: Microsoft Azure (Data Factory, Databricks, Synapse, S3, Key Vault, Purview, OneLake Lakehouse Architecture), AWS (S3, Glue, Athena, Lambda, RDS, DynamoDB, Redshift), Docker, Jenkins, Git, Terraform, Azure DevOps
* **Big Data Systems**: Apache Kafka, Hadoop, Snowflake, Airflow, Informatica PowerCenter, Spark
* **Business Intelligence & Reporting**: Power BI, Tableau, SAP Business Objects, Looker
* **Data Manipulation & Visualization**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, ggplot2, Excel
* **CI/CD & Version Control**: Git, GitHub, Jenkins, Azure DevOps, ARM Templates
* **Agile & Project Management**: JIRA, Agile Methodologies (Scrum, Kanban)
* **Data Governance & Security**: Microsoft Purview, Azure Entra ID, Azure Key Vault, RBAC, Data Lineage

##  EXPERIENCE

**Sr. Data Analyst, Abbott, IL** Mar 2023 - Present

* Actively participated in all project execution phases, including effort estimation, proposal submission, requirement gathering, design, development, testing, deployment, and post-implementation support, ensuring on-time project delivery and a 25% increase in project efficiency.
* Created and executed SQL queries to analyze and troubleshoot issues during migration and integration processes, improving data accuracy and reducing data migration time by 30% while streamlining troubleshooting processes.
* Utilized AWS technologies such as Glue, Step Functions, and Athena for data ingestion, transformation, and analysis, ensuring data accuracy during migration to cloud platforms, resulting in a 40% reduction in data inconsistency and migration errors.
* Analyzed and transformed data in the data lake, leveraging AWS QuickSight, Spark, and Airflow to optimize workflows and provide actionable insights, enhancing operational efficiency by 20% and reducing data processing time by 25%.
* Implemented and supported EMR/EHR/HL7 integrations within Athena Health, ensuring seamless data exchange and compliance with healthcare standards, which resulted in a 15% improvement in system integration and a reduction in integration errors.
* Troubleshot issues with the Athena Health EHR system and provided onboarding support for new practices, ensuring smooth transition and system utilization, contributing to a 20% increase in new practice adoption and reduced transition time by 25%.
* Participated in the build, configuration, and testing of Athena Health features to align system functionality with business needs, ensuring a 30% improvement in system performance and user satisfaction.
* Conducted JAD sessions, interviews, and walkthroughs with stakeholders, product owners, and technical teams to gather and translate requirements into actionable technical specifications, improving requirement clarity and reducing ambiguity by 20%.
* Partnered with business stakeholders to manage requirements reprioritization and ensure alignment with project goals, improving project alignment and reducing scope changes by 15%.
* Developed and managed test scenarios executed unit test scripts, and debugged issues to ensure smooth system transitions, reducing post-deployment issues by 30% and ensuring 100% system functionality at launch.
* Created Requirement Traceability Matrices to track compliance with objectives during and after testing phases, enhancing traceability and ensuring 98% accuracy in meeting project requirements.
* Built automated reports and dashboards using Power BI, Tableau, and AWS QuickSight, enabling data-driven decision-making and reducing reporting time by 35%, improving overall decision-making accuracy by 20%.
* Regularly coordinated with project stakeholders to track project progress, update statuses, and ensure deliverables were met, increasing team collaboration and improving project tracking efficiency by 25%.
* Successfully supported data migration and transformation for healthcare systems, leveraging AWS technologies for efficient integration and compliance, resulting in a 40% improvement in migration efficiency and a 10% decrease in integration errors.
* Played a critical role in troubleshooting and enhancing the Athena Health EHR system, ensuring alignment with business and clinical needs, leading to a 20% improvement in system performance and stakeholder satisfaction.
* Delivered high-quality reports and dashboards that streamlined business operations and improved data accessibility for stakeholders, resulting in a 30% increase in decision-making speed and operational efficiency.

**Data Engineer, L&T, TX** Jan 2021 - Feb 2023

* Conducted data profiling and mapping using Talend, Alteryx, and Informatica, analyzing functional and non-functional data elements, which improved data accuracy and consistency during the migration process, reducing data errors by 20% and ensuring a seamless transition to the target data environment.
* Proficient in reverse engineering from platforms like Qlik Sense, Alteryx, and Cobol code, successfully analyzing existing solutions to migrate data from on-premises warehouses to Google Cloud Platform (GCP), reducing migration time by 25% and ensuring optimized performance for future workloads.
* Utilized Agile project management tools like Jira and Trello to create and manage user stories, track progress, and visualize project workflows. This enhanced team collaboration, reduced project delays by 15%, and ensured timely delivery of key features in an Agile environment.
* Stayed up to date on the latest Azure services and features, proactively advising stakeholders on opportunities to optimize business processes, leading to a 20% improvement in productivity by implementing Azure-based solutions tailored to business needs.
* Demonstrated expertise in DAX (Data Analysis Expressions) within Power BI to create advanced calculations, complex measures, and dynamic reporting functionalities, improving reporting accuracy by 30% and enabling more insightful decision-making for business stakeholders.
* Hands-on experience with Power BI, including the web service and data flows, ensuring seamless data integration, collaboration, and sharing across teams, which facilitated real-time reporting, improving decision-making speed and efficiency across departments by 25%.
* Developed visually compelling and insightful Power BI reports and dashboards, leveraging advanced visualization techniques and interactive features to communicate key findings, improving business transparency by 40% and enabling data-driven decisions across leadership.
* Employed custom visuals and themes to tailor reports to business requirements, aligning visuals with company branding guidelines, resulting in a 15% increase in user engagement and satisfaction by ensuring the presentation of data was both functional and aesthetically aligned.
* Implemented efficient data preparation and transformation workflows using Power Query in Power BI Desktop and Azure Data Factory, enhancing data processing speed by 20% and reducing manual intervention by 35%, ensuring timely and accurate reporting for business stakeholders.
* Integrated Azure services like Azure Data Lake Storage, Azure SQL Database, Azure Analysis Services, and Azure Machine Learning with Power BI, creating end-to-end BI solutions that ingested, processed, analyzed, and visualized data seamlessly, reducing report generation time by 30%.
* Imported and cleansed data from sources like Hadoop, Spark, and cloud storage services onto SQL Server and Azure using ETL tools like Apache Nifi and Kafka, optimizing performance across platforms and reducing data load times by 25%, improving data management efficiency.
* Utilized Power BI's scheduling and automation features to automate report generation and data refresh processes, reducing manual effort by 40% and ensuring data accuracy by keeping reports up to date with the latest data, enhancing operational efficiency.
* Proficient in writing complex SQL queries to extract, transform, and analyze data from Azure databases and data warehouses, optimizing query performance and reducing query execution time by 20%, ensuring efficient data retrieval for reporting and analysis.
* Applied data wrangling techniques to clean, reshape, and combine data from multiple sources, ensuring data quality and consistency, which improved data integrity across various business intelligence platforms by 30% for more reliable decision-making.
* Worked with big data technologies like Hadoop, Spark, and NoSQL databases (Cassandra, MongoDB), handling large volumes of structured, semi-structured, and unstructured data, ensuring scalability and performance, leading to a 25% increase in big data processing efficiency.
* Familiar with integrating Epic with other healthcare systems, ensuring seamless data exchange and promoting continuity of care through improved interoperability, enhancing healthcare delivery and improving patient outcomes by 20%.

**Sr. Data Analyst, St of GA, GA** Sep 2018 - Dec 2020

* Led complex data analysis and geospatial analytics projects end-to-end, including effort estimation, proposal submission, requirements gathering, development, and production deployment, achieving a 30% increase in project throughput and consistent on-time delivery.
* Conducted advanced data analysis using SQL (Snowflake), Python, and GIS tools to uncover actionable insights from structured and spatial datasets, supporting strategic business decisions and driving a 15% improvement in operational efficiency.
* Automated interactive dashboards and KPIs using Power BI and Tableau, enabling real-time insights and increasing stakeholder engagement by 25% through intuitive data visualizations aligned with business performance metrics.
* Utilized JAD sessions, stakeholder interviews, and cross-functional walkthroughs to gather and clarify data requirements, reducing ambiguity and improving requirement clarification efficiency by 25%.
* Executed statistical modeling and geospatial data mining with Python and R, delivering predictive insights that improved decision accuracy by 20% and contributed to a 10% increase in business profitability.
* Crafted complex SQL queries in a cloud environment (e.g., Snowflake) to extract and analyze data trends, patterns, and KPIs, informing marketing, credit, and operational strategies in alignment with business objectives.
* Managed data workflows and quality from acquisition through preprocessing, ensuring accuracy and governance compliance, resulting in a 40% reduction in project delays and improved data confidence.
* Performed end-to-end UAT and GUI testing for geospatial applications and dashboards, improving system usability, reliability, and cross-platform performance, leading to a 30% boost in user satisfaction.
* Collaborated with product owners, data stewards, and technical teams to track project progress, escalate issues, and conduct sprint reviews, enhancing cross-functional communication and reducing project lag by 20%.
* Conducted gap analysis on spatial and non-spatial datasets, identifying enhancement opportunities and guiding product improvements that increased data product quality by 15%.

 **Data Engineer, TJX Companies, MA** Jul 2016 - Aug 2018

* Implemented medium to large-scale BI solutions using Azure Data Factory, Azure Databricks, SQL Data Warehouse, Azure Data Lake, and Cosmos DB, leading to a 30% improvement in data processing efficiency, enabling faster insights for business decision-making across departments.
* Provisioned Azure resources like ADF, Databricks, ADLS Gen2, and linked services using ARM Templates and Terraform, reducing provisioning time by 25%, ensuring streamlined infrastructure management, and enabling quicker scaling of cloud-based data solutions for ongoing projects.
* Developed and managed ADF pipelines for ingesting data from various sources like SQL Server, REST APIs, SFTP, and others, automating data workflows and reducing manual effort by 40%, which significantly improved data integration speed and enhanced reporting efficiency for stakeholders.
* Transformed large volumes of raw data using Azure Databricks, optimizing ETL workflows and improving data processing speeds by 30%. This enhanced data readiness for analytics and reporting, providing faster insights and actionable recommendations for business growth.
* Built efficient and scalable ETL pipelines using Python, PySpark, and Scala, which streamlined the process of data organization and transformation, reducing pipeline processing time by 25%. This allowed for quicker data availability and more timely business insights.
* Created, scheduled, and optimized data flows in ADF, ensuring timely data refreshes and reducing refresh time by 20%. This led to more reliable, up-to-date data being available for real-time decision-making and enhanced operational efficiency for business teams.
* Worked with various data sources, including SQL Server, Salesforce, XML, JSON, and CSV, using appropriate ingestion techniques in ADF. This improved data integration efficiency by 30%, enabling the seamless merging of disparate data sources for more comprehensive analytics.
* Adopted CI/CD standards using Azure DevOps and GitHub, implementing branching in the development environment and pull requests for streamlined code merging, reducing deployment errors by 25%, ensuring high-quality code, and improving collaboration within development teams.
* Collaborated with security and governance teams to define and implement comprehensive data protection policies, ensuring that all data handling and storage practices met organizational and regulatory compliance requirements, maintaining data integrity and security.

**Data Analyst, Tecordeon Soft Sol, India** Jan 2014 - May 2016

* Provided complete Software Development Lifecycle (SDLC) from business analysis to development, testing, deployment, and documentation, ensuring seamless project execution and a 25% reduction in project timeline through streamlined processes and clear documentation.
* Leveraged Azure SQL Database and Azure Data Lake Storage for storing and processing large volumes of structured and unstructured data, ensuring scalability and performance, which led to a 30% increase in data processing efficiency and scalability.
* Performed requirements gathering and analysis, estimated project timelines, prepared schedules, design documentation, conducted design reviews, and facilitated development, testing, and deployment of application enhancements, improving project planning accuracy by 20%.
* Performed data mapping between source systems to target systems, created class diagrams, ER diagrams, and utilized SQL queries to filter data, resulting in a 15% improvement in data alignment and reducing mapping errors by 25%.
* Analyzed source data and made decisions on appropriate extraction, transformation, and loading (ETL) strategies, reducing data processing time by 20% and ensuring efficient data flow and data quality.
* Conducted performance tuning of SQL queries and data processing pipelines in Azure SQL Database and Azure Data Lake Storage, optimizing query execution times and resource utilization, leading to a 40% improvement in processing speed and 30% reduction in resource consumption.
* Performed financial reporting, including key performance indicators (KPIs) and variance analysis, providing insightful information on revenue trends, and enabling management to make 15% more informed decisions on financial strategy.
* Implemented integration of Python scripts within Power BI for advanced analytics, leveraging Python libraries such as pandas, NumPy, and scikit-learn for complex data manipulation, statistical analysis, and machine learning, enhancing analytical capabilities and reporting accuracy by 25%.
* Utilized Power BI Service for publishing and sharing reports with stakeholders, enabling collaborative data analysis and decision-making, increasing report accessibility by 30% and improving team collaboration by 20%.
* Assigned a data collection report list, scheduled meetings with data collection program representatives, and set meeting agendas, improving team coordination and ensuring a 20% increase in data collection efficiency.
* Used Informatica transformations such as Source Qualifier, Aggregators, Lookups, Filters, and Sequence, optimizing ETL workflows, which reduced data load times by 30% and enhanced processing speed.
* Provided performance tuning of user queries by analyzing explain plans, recreating user driver tables with right primary indexes, scheduling statistics collection, and secondary join indexes, leading to a 25% improvement in query performance.
* Provided training and support to end-users on Power BI usage and best practices for data analysis and reporting, empowering them to leverage data-driven insights, increasing user adoption and self-service reporting by 20%.
* Reviewed existing Teradata scripts and provided valuable input to improve efficiency, leading to a 15% reduction in execution time and enhancing overall system performance.