

Sai Ganesh Tanuku Data Engineer

✉ saiganesh@mymailsbox.net ☎ 660-909-2694 📍 Buford, Atlanta in LinkedIn

Dynamic Data Engineer with 4+ years of expertise in designing and optimizing data pipelines. Proficient in Python, Hadoop, Hive, Spark, SQL, and AWS. Specialized in data engineering and analytics platforms, adept at translating business requirements into efficient technical solutions for impactful visualization reporting and data quality enhancements.

Skills

Programming Languages: (Python, R, Java, Scala, SQL, Unix/Shell Scripting),
Big data Tools: (Hadoop, HDFS, MapReduce, Hive, Sqoop, Oozie, EMR, Spark, Pig, lambda functions, Zookeeper),
Frameworks & Tools: (NumPy, Scikit-learn, Pandas, Seaborn, Pyspark, Power BI, Tableau, Git, GitHub, MS Excel, MS Access, PySpark, Airflow, ETL, Jira, Informatica power center, Waterwall, Agile/Scrum, Jenkins, Docker, Confluence),
Datastores and Cloud: (Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), MySQL, Oracle, MongoDB, Teradata, PostgreSQL, Cassandra, Apache Kafka, Snowflake, DataBricks)

Professional Experience

Data Engineer, IQVIA 08/2023 – Present | Remote, USA

- Implemented PySpark and AWS Glue pipelines to integrate patient electronic health records (EHR) and management systems with Amazon Redshift, reducing record retrieval time by 20%.
- Developed a significant analytics platform, achieving a 20% improvement in client decision-making efficiency, by leveraging Amazon S3 for data storage, AWS Glue for data processing, and Amazon QuickSight for advanced analytics
- Facilitated a 15% boost in healthcare operational efficiency by architecting a comprehensive data platform, utilizing Amazon EC2 for computing power, Amazon RDS for database management, and AWS Lambda for serverless data processing workflows.
- Enabled 40% faster decision-making by integrating large-scale medical and geographic datasets, utilizing AWS Glue for ETL processes and Amazon Redshift for scalable data warehousing, enhancing insights and operational efficiency.
- Automated regulatory reporting and data governance processes using Python, PySpark, AWS Lambda, and AWS Glue, focusing on patient billing, insurance claims, and healthcare compliance.
- Strengthened data protection measures for patient privacy compliance (e.g., HIPAA) by leveraging Amazon Kinesis and AWS security services in healthcare information systems.

Data Engineer, IKS Health 12/2019 – 07/2022 | Hyderabad, India

- Revolutionized patient data management by architecting a robust cloud infrastructure on Azure, integrating Azure Data Lake Gen 2 and Cosmos DB. Significantly enhanced data integrity and accessibility across diverse data types.
- Led Azure Synapse Analytics pipeline development and optimized ETL with Azure Data Factory and Databricks, achieving a 50% increase in efficiency, reducing processing times by 35%.
- Pioneered storage optimization in Azure by transitioning to Parquet and Avro formats, resulting in a 30% reduction in cloud storage costs. Implemented robust data security measures, reducing unauthorized access incidents by 50%.
- Automated data pipelines with Azure PowerShell, PySpark-based Python notebooks, and dynamic Azure SQL objects, enhancing report generation and supporting ML engineers and data analysts.
- Ensured compliance and safeguarded sensitive patient information for over 10,000 records daily. Implemented advanced encryption and access logging techniques, resulting in a 50% reduction in unauthorized access incidents.

Data Engineer, R1RCM 06/2019 – 11/2019 | Hyderabad, India

- Implemented Azure Data Factory to extract and load JSON data from REST APIs, resulting in a 20% improvement in data accuracy specifically within Azure SQL Server.
- Executed the migration of 100 TB on-premises data to Snowflake Data Warehouse via Azure Data Factory also Implemented DBT and Airflow DAGs for intricate data transformations.
- Managed CI/CD pipelines with GitHub, Jenkins, Terraform, and developed scalable ML pipelines for real-time analysis.

Data Engineer, NTT DATA 03/2018 – 05/2019 | Chennai, India

- Transformed Revenue Cycle Management with T-SQL, cutting data redundancy by 25% and driving data-driven decisions.
- Ensured 100% HIPAA compliance with strict SQL Server security and automated SSRS reporting for multiple AR teams.
- Achieved a 20% reduction in query execution time, tailoring database performance to evolving healthcare standards and developing custom reporting tools for five AR teams.
- Optimized operational workflows and instilled a culture of data-driven excellence, driving substantial improvements in Revenue Cycle Management and data security protocols.

Education

Master of Science, University of Central Missouri 08/2022 – 12/2023 | Warrensburg, USA
Industrial Technology

Bachelors in Technology, Jawaharlal Nehru Technological University 09/2013 – 05/2017 | Kakinada, India
Mechanical Engineering

Certificates

- Microsoft Certified: Azure Data Engineer Associate
- Microsoft Certified: Azure Data Fundamentals
- Microsoft Certified: Power BI Data Analyst Associate