SAI MANIDEEP CHITTIPROLU – Azure Certified Data Engineer Associate

980 895 5999 | saimanideep83@gmail.com | www.linkedin.com/in/saimanideep13 |Charlotte, NC

Accomplished Data Engineer with 5 years of experience in designing, developing, and optimizing scalable data solutions across cloud and on-premises environments. Experienced in leading complex data migration projects, orchestrating efficient ETL workflows and managing high-performance data warehousing solutions. Proficient in AWS, Azure, and various data engineering tools, with a proven track record of improving data processing efficiency, accuracy, and security. Adept at implementing CI/CD pipelines, infrastructure as code practices, and advanced data modeling techniques to enhance overall productivity and code quality. Strong collaborator with cross-functional teams, contributing domain-specific knowledge and technical expertise to drive innovative data analytics solutions

TECHNICAL SKILLS

Programming: Python, R, HTML, CSS, SQL, Git, MySQL, Scala, Oracle DB2, PySpark, MongoDB.
Data Tools & Techniques: Data architecture, Data modeling, integration, Kafka, Microsoft SQL Server, SSIS, SSAS, data governance, data quality, data security principles, OLAP, OLTP, ETL, Tableau, Power Bi, Data mining, Data warehousing, Hadoop, Snowflake, Talend, Advance MS Excel (VLOOKUP, VBA), SAS, Sampling, Hypothesis Testing, Regression, Decision Trees, Churn Analysis, Exploratory Data Analysis.
Cloud & Utilities: AWS Redshift, EMR, Athena, Kinesis, S3, EC2, Glue, Quick Sight, Code Pipeline, Code Build, CloudWatch, SNS, SQS, Cloud Formation, Azure Synapse Analytics, Azure Data Lake, Azure HDInsight, Azure Databricks, Apache Airflow, Apache Hadoop, Apache Kafka.

WORK EXPERIENCE

Data Engineer @ Vision Square Inc   **August 2023 – Present**
- Led the end-to-end migration of large-scale financial data from on-premises systems to AWS cloud infrastructure, ensuring minimal downtime and data integrity.
- Designed and developed scalable and robust data pipelines using AWS Glue, PySpark, and AWS EMR, resulting in improved data processing efficiency and reliability.
- Orchestrated complex ETL workflows with Apache Airflow, automating data extraction, transformation, and loading processes to ensure data accuracy and consistency.
- Managed data warehousing solutions with AWS Redshift, Snowflake, and Snowflake Snowpark, optimizing schema design and query performance for faster data retrieval and analysis.
- Implemented partitioning, indexing, and other performance tuning techniques to enhance data query efficiency.
- Administered Oracle Database, Amazon DynamoDB, Mongo DB, IBM DB2, and Snowflake Snowcloud, ensuring high availability, scalability, and data security.
- Developed and maintained RESTful APIs for seamless data integration with various applications and systems.
- Employed infrastructure as code (IaC) practices using tools like Terraform and AWS CloudFormation to automate the deployment and configuration of data infrastructure.
- Implemented CI/CD pipelines to streamline the development, testing, and deployment of data solutions, enhancing overall productivity and code quality.

Graduate Research Assistant @ University of North Carolina | Greensboro, NC   February 2022 – May 2023
- Processed over 100 GB of university's data using Azure services including Azure Data Lake Storage and Azure Cosmos DB, resulting in improved data handling efficiency and scalability.
- Developed and validated advanced financial models that accurately projected financial outcomes with a forecast accuracy rate of over 95%, leading to informed decision-making and strategic planning.
- Orchestrated and automated data workflows for 10+ heterogeneous data sources using Azure Data Factory pipelines and Databricks clusters, reducing data processing time by 50%.
- Implemented machine learning algorithms and models for various tasks, resulting in a 30% improvement in model performance and scalability compared to previous implementations.
- Designed and developed 5+ RESTful APIs using Azure API Management and ASP.NET Web API, facilitating seamless integration between internal data analytics systems and external applications
- Utilized Apache Kafka to enable real-time data streaming and processing, ensuring timely and accurate analysis of university data and facilitating immediate decision-making processes.
- Contributed domain-specific knowledge and technical expertise to interdisciplinary teams, leading to the successful completion of 5 research projects and the development of innovative analytical methodologies.
- Conducted ETL processes to extract, transform, and load university data using Talend, ensuring data accuracy, consistency, and Hadoop for processing and analyzing large volumes of university data.
- Created interactive dashboards and visualizations using Power BI to present key insights from university datasets to stakeholders.

Data Engineer @ Infosys Pvt Ltd | Hyderabad, India   September 2019 –November 2021
- Designed and deployed AWS-based ETL pipelines for 10+ diverse medical device data sources, facilitating the ingestion and processing of over 500 GB of medical records data daily.
- Engineered scalable data transformation processes using AWS Lambda, Apache Spark, SQL, Databricks and AWS Glue, resulting in a 40% improvement in data quality and integrity.
- Implemented advanced data modeling techniques, resulting in a 40% improvement in query performance for multidimensional medical device datasets, enabling efficient data retrieval system for stakeholders.
- Utilized AWS Redshift and AWS Athena to analyze over 100 GB of medical device data, enabling stakeholders to uncover actionable insights and trends with 15% increase in operational efficiency.
- Orchestrated automated deployment and scaling of data processing infrastructure, reducing deployment time by 60% and enabling seamless scalability to accommodate fluctuating data volumes.
- Implemented robust data governance and security measures, resulting in a 30% reduction in data security incidents and ensuring compliance with industry regulations such as HIPAA and GDPR.

- Applied R and Python for hypothesis testing, regression analysis, decision trees, and churn analysis on medical device datasets, deriving actionable insights using Tableau and Power BI.
- Implemented continuous integration and continuous deployment (CI/CD) pipelines using Azure DevOps, Git, and Jenkins, automating the build, test, and deployment processes for data analytics solutions.
- Collaborated with development and operations teams to implement infrastructure as code (IaC) using Terraform and ARM templates, enabling repeatable deployment of infrastructure components.
- Participated in Agile development sprints, collaborating with cross-functional teams to deliver data analytics solutions on schedule, leveraged advanced MS Excel features for data manipulation and analysis.

EDUCATION

University of North Carolina **|** Greensboro, NC **January 2022 – December 2023**
Master of Science in Computer Science with **3.68 GPA**

R.V.R & J.C College of Engineering **|** Guntur, India **June 2016 – June 2020**
Bachelor of Engineering in Electronics & Communications with **3.2 GPA**

PROJECTS

- **Drowsiness Detection System:** Developed ML models for early detection of driver drowsiness, analyzing over 10 million data points from RGB video sequences, achieving a detection accuracy of 90%.
- **Heart Disease Prediction Analysis:** Utilized Support Vector Machine, Naïve Bayes, and Decision Tree algorithms to predict heart disease symptoms – analyzing dataset comprising 10,000 patient records.
- **Customer Satisfaction Analysis:** Evaluated feedback data from 5,000 customer surveys and created Tableau dashboards to track customer satisfaction levels, with 20% improvement in satisfaction scores.
- **Aspect Based Sentiment Analysis On Restaurant Reviews**: Developed a sentiment analysis model to evaluate customer reviews, focusing on specific aspects of the dining experience, Utilized BERT, NLTK, Word2Vec, GloVe, and lemmatization techniques to preprocess and analyze textual data, achieving accurate sentiment categorization.
- **Spotify ETL Pipeline**: Created an ETL pipeline for top chart song analysis using Azure services, and used Azure Storage Account, Data Factory, Databricks, Synapse Analytics, and Power BI for data processing, storage, and visualization, conducted comprehensive data analysis to identify trends and insights from Spotify's top charts.