Satya Vinod Kantipudi

Dallas, TX | +1(972)836-6556 | Email: vinodkantipudi1510@gmail.com

SUMMARY

*"Dynamic Individual with 1 year of experience as a **Data Engineer** with expertise in designing **Big Data** intensive applications using Cloud Data Engineering, Data Modeling, Data Warehousing, ETL, Hadoop, Reporting and BI"*

EXPERIENCE

Data Engineer, Center For Brain Health   Feb 2023 – Jan 2024

- Architected and deployed Amazon RDS **PostgreSQL** database to manage analytics data, achieving enhanced query performance and ensuring high availability.
- Migrated **50TB** of data from Amazon RDS to Amazon S3 using DMS, reducing migration time from Hours to 20 minutes.
- Designed and implemented 15+ scalable **ETL** jobs leveraging **Apache Spark**, transforming data across 100+ tables from S3 raw zone to refined zone in batch mode.
- Achieved a **20%** reduction in storage cost by implementing a Lakehouse architecture on AWS, enhancing data processing speed.
- Improved query performance in **Amazon Redshift** up to 75% by implementing complex CTEs to optimize complex multi-table joins. Resulted in a 2x reduction of query run times by preprocessing and filtering data prior to the main query execution.
- Enhanced data efficiency by employing **Apache Hudi** for UPSERTS and incremental processing, reducing data latency by 50% and enabling real-time Change Data Capture (CDC) for a 50 TB data lake.
- Collaborated with cross-functional teams to create over 10 interactive **Key Performance Indicators** (KPIs) reports in Amazon **Quicksight**, providing actionable insights to product managers and stakeholders.

TECHNICAL SKILLS & CERTIFICATION

Programming languages: Python, R, SQL
Databases: PostgreSQL, Microsoft SQL Server, Cassandra (NoSQL), Amazon Redshift
Big Data Technologies: PySpark, Hadoop, MapReduce, Yarn, Hive, Kafka, HBase
ETL and Visualization: Airflow, PowerBI, Tableau, QuickSight
Amazon Web Services: S3, Lambda, Athena, IAM, RDS, Glue, Step Functions, API Gateway, Cloudwatch, EC2
Other skills: data modeling, data warehouse, data lakes, ETL, ELT, OLAP, OLTP, agile methodologies
Certifications: AWS Solutions Architect Associate, Databricks Lakehouse Fundamentals, Snowflake Hands on Essentials

ACADEMIC PROJECTS

Real-Time Sentiment Analysis Pipeline
- Architected an end-to-end real-time data analytics pipeline using TCP/IP sockets for data streaming, Apache Kafka for ingestion, Apache Spark for processing, OpenAI's ChatGPT for sentiment analysis, and Elasticsearch for storage and querying.

Built a Pipeline for Reddit Data Analytics
- Developed full-stack data pipeline orchestrated by Airflow to extract Reddit data, store in S3, transform with Glue and Athena, and load into Redshift data warehouse for analytics.

Predicting Product Sales for Big Mart Retail Stores
- Built machine learning models in Python to predict product sales at BigMart stores, achieving top accuracy with a Generalized Additive Model (R2 of 0.62) after substantial data analysis, cleaning, and evaluation of multiple algorithms.

Big Data to Uncover Patterns in New York City's Property and Crime Data
- Implemented a distributed analytics pipeline with PySpark and Tableau that identified connections between population, crime rates, home values, and unemployment in New York City to extract key insights for real estate investment decisions.

Building a Scalable ETL Pipeline for YouTube Data Analytics on AWS
- Spearheaded the development of a comprehensive ETL pipeline using AWS Athena, Glue, and Lambda in Python, facilitating efficient analysis and categorization of YouTube video data for optimized content strategy.

EDUCATION

Master of Science Business Analytics, The University of Texas at Dallas   **GPA: 3.8/4**   Aug 2022 – May 2024
Coursework: Big Data, Advanced Statistics, Data Warehousing, Cloud Computing, DBMS, Predictive Analytics, Econometrics and Time series, Machine Learning, Natural Language Processing (NLP)

ACHIEVEMENTS

Awarded The Dean's Scholarship in Business Analytics department, ranking among the top recipients out of 1,000 applicants.
Leadership - Led a team of 8 student volunteers to organize AWS workshops on big data analytics for 50+ attendees.
Published "Artificial Intelligence Based Fault Diagnosis of Automobile Gearbox" in International Research Journal of Engineering and Technology, Volume: 08 Issue: 11 Nov 2021.