Satya Vinod Kantipudi

Dallas, TX | +1(972)836-6556 | vinodkantipudi1510@gmail.com | github.com/vinod151019
https://www.linkedin.com/in/svk1510/

EDUCATION

University of Texas at Dallas August 2022 – May 2024
Master of Science in Business Analytics GPA: 3.95/4.0
Relevant Coursework: big data, advanced statistics for data science, business data warehousing, cloud computing, database foundation for business analytics

Amrita School of Engineering August 2017 – May 2021
Bachelor of Technology in Mechanical Engineering GPA: 3.3/4.0
Relevant Coursework: computer science fundamentals

WORK EXPERIENCE

Data Engineer Intern, Center for Brain Health February 2023 – September 2023

- Collaborated with a cross-functional team of 5 to design and implement a company-wide data platform serving over 1M+ users.
- Reduced storage costs by 20% by designing and implementing a lake house architecture on AWS to enable scalable data processing.
- Implemented Apache Hudi open-source technology to manage slowly changing dimensions in the data lake. Enabled handling of 10000 incremental record updates per day without impacting downstream analytics.
- Improved the reliability of data pipelines by 80% using Step Functions and Event Bridge, preventing any data loss for research project.
- Designed and implemented ETL workflows using AWS Glue, reducing data processing time by 2 hours.
- Monitored and troubleshot data pipelines using Cloud Watch, reducing the time to resolve issues by 75%, resulting in researchers being able to get data insights faster and make better decisions.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL
Databases: PostgreSQL, Microsoft SQL Server, NoSQL
Tools and Software: Spark, Hadoop, Hive, Kafka, HBase, Cassandra, Airflow, Snowflake, Tableau
Amazon web services: S3, Lambda, Athena, IAM, RDS, Glue, Step Functions, API Gateway, Redshift, Quick Sight
Interpersonal Skills: Communication, Collaboration

ACADEMIC PROJECT EXPERIENCE

AWS ETL Data Pipeline on YouTube Data

- Developed and deployed a highly efficient data pipeline using AWS Lambda, Glue, and S3, enabling seamless ingestion, processing, and storage of YouTube data. This empowered data-driven insights for strategic decision-making.
- Implemented cost optimization strategies and performance enhancements for data lake architecture, resulting in saving 30+ DPU hours per week.

Built Scalable Cloud Data Pipeline with DBT, Snowflake and Airflow

- Developed modular ETL architecture using DBT to transform 3GB of data from S3 into Snowflake.
- Achieved near real-time data availability by scheduling Airflow to trigger DBT models every 5 mins as new data lands in S3.
- Enabled robust monitoring via Slack alerts and email notifications using AWS SNS for pipeline runs handling daily.

Analysis of Property Value Dynamics in New York City

- Configured a 4-node cluster using 4 laptops to process 5GB data in parallel, implementing a master-slave architecture with Spark for distributed, scalable data analysis of NYC property sales.

EXTRACIRRICULAR & LEADERSHIP ACTIVITIES

AWS Organization at UTD

- Led a team of 8 student volunteers to organize AWS workshops on big data analytics using S3, Glue and Athena for 50+ attendees.
- Conducted 20+ interview preparation workshops for AWS roles, helping students build confidence through practice sessions.