264 Steeplechase Dr.
Exton PA 19341
GitHub | LinkedIn

VENKATA SIVA SANDEEP YEDLA

Phone: (425)-551-9040
yxv180001@utdallas.edu
www.sandeepyedla.com

EDUCATION

Dallas, TX	**The University of Texas at Dallas**	**Jan 2019 – Dec 2020**

- M.S in Information Technology and Management

Bangalore, **India**	**Amrita Vishwa Vidyapeetham University**	**Aug 2010 – May 2014**

- B.S in Electronics and Communications Engineering

Coursework**:** Distributed Systems, Data Structures & Algorithms, Database Management, Cloud Computing, Big Data, Object-Oriented Programming (Python), Agile Project Management.

Certifications: **AWS Certified Solutions Architect** – Associate (SAA C02)

EMPLOYMENT

Data Engineer Intern	**CBL Solutions** (Irving, TX)	**Fall 2020**

- Designed and implemented a real-time **scalable data pipeline** to process semi-structured data by integrating 390 million raw records from different data sources using **Kafka** and **PySpark** and store processed data in **Redshift**.
- Utilized **Spark Streaming** to consolidate and clean transactional and event data resulting in speed improvements of 20% in the production.

Software Engineer Intern	**Neiman Marcus** (Irving, TX)	**Spring 2020**

- Developed **REST APIs** for segmentation, personalized recommendations and Omnichannel messaging that reduced cart-abandonment rate by 22%, leading to a $0.4 M increase in online sales.
- Optimized an effective **CI/CD pipeline** for MuleSoft applications resulting in an 18% reduction in time to market.

Senior Software Engineer	**Infosys** (Hyderabad, India)	**July 2014 - Oct 2018**

- Designed and developed **ETL workflows** in **Informatica PowerCenter** and automated data gathering from SQL databases through schedulers and event-based triggers, which reduced the team size by 10%.
- Resolved critical performance bottlenecks using **SQL Query tuning techniques** resulting in reduced costs of system resources by 20%.
- Saved 16 hours/week of manual work by creating a reusable object that calculates MD5 used in SCD1 type mappings.

ACADEMIC PROJECTS

Data Pipelines for Music Streaming Company (GitHub)

- Created a **reusable production grade data pipeline** that incorporates data quality checks and allows for easy backfills.
- Developed a data pipeline that automates data warehouse ETL by building custom **Apache Airflow** operators that handle the extraction, transformation, validation and loading of data from **S3** to **Redshift**.

Slack Bot with Python & Django using Events API (GitHub)

- Designed a minimal Slack Bot with **Python Django** based **REST** backend web application using Slack's event-based API.
- Developed a bot logic to perform a specific task in response to an event i.e., when a user greets, the bot will respond with a greeting message.

Building a Modern Datawarehouse using SQL Server, Apache Spark (GitHub)

- Created a data warehouse model that can support data for disparate data sources.
- Built a **dimensional model** to import the data with different shapes into a data warehouse and create processes to transform and load the data using **Apache Spark**.

LANGUAGES AND SKILLS

- **Programming:** **SQL**, **Python**, Scala, Shell Scripting
- **Databases (SQL & NoSQL):** MySQL, MS SQL Server, **Amazon Redshift**, Cassandra, HBase.
- **ETL and BI Skills:** Data Warehousing, Data Modeling, ETL design, Data Mining, Informatica, Tableau.
- **Big Data/Cloud Stack**: Hadoop, HDFS, **Spark**, Airflow, **Hive**, **Kafka**, Sqoop, AWS EMR, Athena, AWS Glue.
- **Others:** Linux, JSON, XML, Git, JIRA, REST API, Jenkins, Django Framework, Docker, Kubernetes.

LEADERSHIP AND AWARDS

- Developed a Python script GitHub to ease the process for about 2000 incoming international students for which, I was recognized as a **Key Contributor** to the Golden Comet award our Organization received among 3000 organizations.
- Received the **Star Team Performer** award in a project team of 35 employees in Infosys for my annual performance.