SANTOSH

Pelham, Alabama**, USA | +1-408-520-9693| santoshav.de2019@gmail.com**

# Professional Summary

Skilled Lead **Data Engineer** bringing about **10+ years**’ experience in the design and **development** of pipelines as well as key skills in the modeling and warehousing of data, **development** of enterprise data needs.

## Proficient in Creating Teradata SQL, MS SQL, POSTGRES SQL, Azure SQL database objects like tables, views, stored procedures, functions and triggers.

* Experience in designing and implementing data engineering pipelines.
* Built data pipelines to load data from various sources to Azure using **Azure Data Factory (ADF)**.
* Involved in **writing complex SQL Queries, Joins, Constraints, indexes, DML to implement the business logic**.
* Proficient in writing **HIVE/IMPALA** queries for handling BIGDATA and **implementing business logic of SDTM conversion** of clinical trials data.
* Supported machine learning models as a Data Engineer by generating pickle files.
* Rich exposure on different ETL tools like SSIS, Trifacta, Tamr , Streamsets etc.
* Implemented ETL jobs in Sqoop to migrate OLTP-RDBMS data into to Hadoop environment both initial and incremental loads.
* Proficient in handling third party **API**(Clarabridge/Calabrio/Pendo) requests and their responses in **JSON/CSV/TXT** file formats.
* Good knowledge on **Clinical trials, SDTM, Banking, Credit Cards and HRM** business units
* Supported Data preparation /Data Sourcing requirements for **Power BI** dashboards.
* Created Elastic search payloads required for frontend applications and created several elastic search indexes for data to be loaded from rdbms using Logstash.
* Proficient in writing automated scripts in Python.
* Data modeling using Star schema (identification of facts and dimensions) and Snowflake schema for designing a Data Warehouse
* **Performance** Tuning of SQL Server objects and maintenance of indexes.
* Good experience in preparing Low level design documents for packages using **SSIS**.
* Created and implemented database **designs** and data models like table design and indexes
* Excellent analytical, problem-solving skills and a motivated team player with excellent inter- personal skills.
* Rich exposure to Software Development Life Cycle (SDLC)
* Hands on experience with both conventional RDBMS (OLTP) and Bigdata environment such as Hadoop ecosystem and Spark.
* Support documentation and change management via **Bitbucket, GIT.**

Skills

|  |  |
| --- | --- |
| * **MS SQL/Teradata/Oracle/ Azure SQL /GCP Big Query/AWS Redshift** Development * **Python** programming proficiency * **MS SQL/SQL tuning** * **SSIS, ADF, TAMR, STREAMSETS, TRIFACTA, LOGSTASH, SQOOP** ETL development * **Hadoop, HIVE, IMPALA** for data analysis and engineering * **ELASTIC SEARCH** development | * **SDLC**, **Agile** Methodologies * Data Engineering, migration, upgrade * **MS Excel** for presenting sample data. * Performance management * Process improvements * **POWER BI** for dashboards * **Apache Airflow, Control-M for orchestration.** * **Cloud Composer** for data pipelines. |

Work History

**Data Engineer** 02/2023 to Till Date

**Flyr Labs**

* Developed high-level design logic and implementation to update the canonical data model (CDM) , Data Quality checks for existing and new clients for the Tickets to Orders project.
* Worked with Cloud Composer (Airflow) and DBT to carry out code production and data backfills.
* Responsible for building end to end data pipelines on Google cloud composer for various data sources like XML, JSON, text files, and other data sources.
* Have built multiple robust and scalable data-pipelines using python scripting.
* Need to support data scientists’ requests by building data pipelines to fuel their ML/AI models.
* Built pipelines in a generalized manner so that they can be re-used for other clients with minimal changes.
* Programming Language: Python, Big query SQL.
* Built custom parsers for reading edifact files which were used extensively in airline industry.
* Technology: Azure Data Factory , Big Query, Cloud Composer, DBT, Airflow ,Azure SQL, Data Engineering.

**Data Engineer** 02/2022 to 02/2023

**Macy’s [for efulgent]**

* Responsible for Data Engineering of lift and shift of existing Teradata Systems to Google cloud and SQL server/SSIS objects to ADF/Azure SQL.
* Design and develop the engineering strategy for Sales related data (Sales/Sale’s Master) migration from TD to BQ including the airflow/ Control-M objects.
* Responsible for lift and shift of 1500+ DAG’s from Airflow V1 to V2 of edw belonging to multiple sub domains of macy’s.
* Need to support data scientists’ requests by building data pipelines to fuel their ML/AI models.
* Built new ADF pipelines for loading data from multiple file systems or types.
* [[1]](#endnote-1)Used Azure Blob/ GCS bucket for handling files of different stages.
* Handled different types of file sources like CSV, TSV , XML, JSON , Excel , binary files while extracting data for my target models.
* Need to prepare Data Quality reports for the tables migrated and monitor the objects on timely manner to check there were no issues or code change happened during migration.
* Worked on Azure logic app for handling few files sources and generative respective notifications based on the file availability.
* Performance tuning of Dataflow jobs by modifying the required configurations used at individual job.

**Data Engineer** 08/2021 to 01/2022

**Eli Lilly**

* I'm responsible for leading the development and maintenance of data wrangling  
  processes, data model designs, and data visualization or dashboards for information  
  integration and exchange initiatives.
* I need to work closely with business SME’s to understand desired product outcomes and design and implement data movement,  
  transformations, and visualizations to answer business questions. A key area of focus will be  
  the review and approval of data movement to ensure consistent validated operational  
  processes across the enterprise.
* Design and develop the whole end to end flow for CRM (Clinical Results Mart). CRM is a data mart for clinical trials of Lilly.
* Used Azure Data Factory for building data pipelines to extract data from various sources.
* Worked on architecture related things of the CRM, which includes data base objects, python scripts, ETL objects.
* Built pipelines using TAMR for the multiple projects inside Lilly for their Analysis/reporting purpose.
* Trained my team mates to get better understanding of SDTM/CDISC terminology used vastly across the clinical trials and its data collection.
* Need to build data quality reports and report gaps in data to form perfect tabular model for SDTM.
* Converted Lilly standards data into SDTM(v3.1.3) standard data using hive/sql scripts
* Built data models in AWS Redshift for storing the output of SDTM converted achieved using TAMR.

**Data Engineer (Sr. Member Technical)** 07/2019 to 07/2021

**Automatic Data Processing (ADP)**

* Designing, developing warehouse, SQL objects and SSIS packages using SQL server 2014/16 Enterprise and Visual Studio 2019.
* Performance tuning of SQL objects, SSIS packages.
* Handling Billions of records of data in SQL server effectively and handling data loads faster handling all typed of analytical data reference, transactional, analytical, aggregations etc..
* Worked on numerous applications in my tenure as Data Engineer at ADP, Pendo (Analytics), Case Search Engine(Search Engine built on .net and Elastic Search) , One&Done(Predictive model) and Clara- bridge(Conversion of call voice data to text and performing sentiment analysis) are a few among them .
* Created python applications to pull data from Elastic Search and create reports in a custom format
* Worked on Hadoop, data load using Sqoop. Data validation using Hive, impala, hue
* Worked on third party API’s like Clarabridge / Clarabrio / Pendo for handling call data and user experience data on the ADP tools.

**Software Engineer** 05/2017 to 7/2019

**GSK [for Modak Analytics]**

SDTM Conversion: SDTM conversion is for converting clinical trials data of GSK internal standards across different geo locations to SDTM standard data for submitting to FDA approval.

* Development regarding converting clinical trials data into industry accepted SDTM standard format.
* Building HIVE & Impala queries for automating the coding techniques
* Performed analysis and presented results using SQL, SSIS, Excel
* Build Scripts to increase the efficiency and to automate the repeatable tasks.
* Built SSIS packages to extract, transform and load data to data warehouse
* Diagnosed and addressed operational data issues.
* Optimized stored procedures and jobs into SQL Server to increase overall performance and time/server use of daily functions.
* Used Tamr tool for data mapping , Cataloguing by training the model for 100+ datasets .

**Program Analyst** 02/2014 to 05/2017

**American Express [ for Cognizant Technology Solutions**]

GCA: Global Credit Administration (GCA) manages the company's credit, collections, and fraud service operations around the world. Global Collections oversees the management of our first-and third-party collections agencies, as well as various specialty partners.

GCA aims to strike the right balance between helping customers in need through a range of workout programs and taking actions to prevent spending that will not be paid back to American Express. In short, our primary focus is to protect and recover American Express' receivable assets, while delivering on our customer commitment.

* Enhancing the existing stored procedure as per the latest requirements and developing stored procedures as per the front-end team requirement.
* Assigning the work from technical perspective to my juniors’ team of size 5 members.
* Metrics collection, analysis & reporting
* Managing the knowledge, process and configuration.
* Built SSIS packages to extract, transform and load data to data warehouse.
* Developing stored procedures and other scripts as per the UI team requirement and interacting with clients on the daily status updates.

# Education

**Bachelor of Technology**: Electronics & Instrumentation Engineering

**Aditya Institute of Technology and Management, JNT University** - Kakinada, Andhra Pradesh

# Certifications

Diagram

Description automatically generatedDiagram, text

Description automatically generated Text, letter

Description automatically generated

1. [↑](#endnote-ref-1)