Naga Venkata Satya Narayana, Gurram   phone: **+91-7660045631**
8.5 yrs exp with Data warehousing projects   email: **nvsngurram@outlook.com**

Career Objective:
      To leverage my expertise in software development and contribute to innovative projects that utilise my skills, with a focus on enhancing user experiences and driving business success.

Experience Summary:
- 6+ years of experience with ETL tools IBM Infosphere DataStage, SSIS.
- 2+ years of experience Reporting tools Tableau and Python.
- 8+ years of experience with SQL, Shell Scripting.
- 1+ years of experience in Databricks, Python and PySpark.
- 1+ years of experience in Machine Learning with NLP, supervised and unsupervised learning.
- Worked on Implementing end-to-end data pipelines for diverse business applications.
- Experience in data extraction, transformation and loading across different relational databases and file systems.
- Experience in building visualisations to study the Insights of data with attractive charts in BI tool.
- Good at fine tuning the model, deploying model using MLOps principles.
- Written scripts to cleanse the data, to do event waits, to run the etl jobs through, perform ftp process.
- Proficient in writing SQL to read data from multiple tables, stored procedures, and cursors.
- Followed Agile methodologies like Scrum and Kanban in software development process.
- Adhering the organisation standards in writing code and deploying it into higher environments.

Skills:
- ETL - IBM Infosphere DataStage 11.5/11.7, SSIS 2017/2019, Databricks
- Reporting - Tableau Desktop, Tableau Server last used 2019.4.
- Databases - Oracle, SQL Server, DB2, Netezza and MySQL
- Programming - Java, Python, Shell scripting
- IT Service Management - Service now
- Scheduling tool - Auto sys, control-M, Prefect

Education:
- In 2010, I have completed schooling from Sri Ravindra High School – Proddatur with 80%.
- In 2012, I completed Pre–University from Sri Chaithanya Junior College - Vijayawada with 83%.
- In 2016, I graduated from the Annamacharya Institute of Technology and Sciences Rajampeta with a bachelor's degree in computer science and engineering with 75%.

Professional Summary:
- I started career at Atos | Syntel in 2017 and worked till March 2020.
- I worked in HCL Technologies from April 2020 to February 2022.
- At present I am working in NTT Data Information processing services from February 2022 to till date.

Work Summary:
Organisation: NTT Data (Feb 2022 to till date)

Internal POC: Automate the staffing process to post the Job Description into careers portal and third party portals to get candidate profiles, once we get profiles parse them to extract skills set, contact information, and save details for further analysis. Using chat prompt, we interact with process to execute posting JD into portal, get suitable candidates from resume corpus and suggest suitable candidates available from the corpus.

Technical skills: Python, Azure services, NLP technique, model building, model testing, MLOps.

Responsibilities:

- Conducted feasibility assessments for integrating with enterprise applications by analyzing and validating available API endpoints to securely post and retrieve application data.

- Designed and implemented NLP pipelines using Python and Azure services (such as Azure Functions, Azure Cognitive Services, and Azure Storage) to perform Named Entity Recognition (NER) and relationship extraction, enabling automated document parsing and generation of structured JSON objects containing the extracted information.

- Fine-tuned pre-trained language models downloaded from Hugging Face and LangChain community models, adapting them to domain-specific datasets to improve performance on tasks such as entity extraction, text classification, and document understanding.

- Designed and implemented a vector database (using tools such as FAISS, Pinecone, or similar) to store contextual embeddings, and performed systematic data splitting into training, validation, and test sets to support robust model training and evaluation.

- Evaluated model performance using standard machine learning metrics, including accuracy, precision, recall, and F1-score, and iteratively refined models based on these metrics to achieve improved predictive quality and reliability.

- Worked on Azure AD tenant configuration, application registration, and access management, including the creation and configuration of service principals and API permissions, to securely integrate with SharePoint for operations such as page and list creation, list item updates, and automated content management workflows.

- Implemented and optimized machine learning models using algorithms such as XGBoost for tasks including classification and regression, leveraging its gradient boosting framework to improve model accuracy, handling of imbalanced data, and feature importance analysis to support better decision-making and model interpretability.

Project 1 Description: **Client: Humana**

This project is all about ensuring the data of specific business processes are accurate, consistent, and dependable across its shareable business entities. It uses IBM Infosphere MDM to master data and helps in ensuring specified objectives.

Technical skills: IBM Infosphere DataStage, SSIS, Oracle, MSSQL and Shell Scripting, Databricks, Azure Data Lake Service.

Responsibilities:
- Involved in PI meetings understand the scope of project and commit for the necessary improvement features.
- In sprint planning divide the feature into multiple stories and commit for the stories which can be completed in a sprint time.
- Built the connection interface between database servers and ETL projects using odbc connection drivers.
- Build the end-to-end data pipelines in terms of extracting data, cleansing, and loading into target process.
- Created data quality rules to validate the accuracy of data that is flowing in the database objects and integrate them in ETL flow.
- Created the ETL pipeline using python notebooks in data bricks to read ADLS files and write into Catalog tables and save staging files in DBFS location.
- Worked on getting files from FTP server (Windows machine) to ADLS location.
- Worked building workflows, parameter initialization and scheduling the created workflows.
- Addressed the bugs that are raised while testing related to data quality and transformations.
- Have experience with Prefect orchestration tool to schedule workflows.

Project 2 Description: **Client: Humana**

This project is all about implementing Data governance strategies to one of Humana's business processes. As part of it I have worked on building data pipelines and quality rules on the data we used to process.

Technical skills: IBM Infosphere DataStage, IBM Infosphere Quality stage, IBM Infosphere Metadata Asset Manager, IBM Infosphere Information Governance Catalogue, Oracle, MSSQL, Netezza, DB2, Shell Scripting.

Responsibilities:
• Involved in PI meetings understand the scope of project and commit for the necessary improvement features.
• In sprint planning divide the feature into multiple stories and commit for the stories which can be completed in a sprint time.
• Build the end-to-end data pipelines in terms of extracting data, cleansing, and loading into target process.
• Created the data quality rules to validate the accuracy of data that is flowing in the database objects.
• Worked on Importing metadata assets into IMAM and share to IGC for building data lineage and data governance reports.
• Scheduled the auto import process through internal schedulers using shell script commands,
• After each enhancement, the code published into multiple environments like INT, QA, PROD. Supported production failures and re-deployed working code in production.

Organisation: HCL Technologies (April 2020 to Feb 2022)   **Client: ING Netherlands**
Project Description:

This project Implements and maintains the Global Data Management strategies across its data sharing platforms. In this project I majorly worked on collecting data from data lake and sharing it to same after performing necessary transformations which are required for downstream applications.

Technical skills: IBM Infosphere DataStage, Azure CI/CD pipelines, GIT, Oracle, MSSQL and Shell Scripting.

Responsibilities:
• Analyzing the user stories related to data indigestion.
• Worked on input data sources like csv, Json, and database tables.
• Preparing the source & target table structures using mapping document.
• Worked on transformations like type conversions, string manipulations, aggregations, pivoting, stage valuations etc.
• Extracted the Json response into individual files by writing script in python.
• Worked on implanting historization logic with help of scd type 2-dimensional model.
• Worked on facts and dimensions and star schema.
• Ingested data from .csv files into data lake also worked on dl-to-dl loading.
• Used azure git repos for doing versioning with the help of istool and git commands.
• Created deployment bundle with IGC using export and import commands of istool.
• Getting the created deployment bundle into azure artifacts and making use of azure pipelines for the deployment of code into higher environment data stage servers.
• Triggering jobs using DS job commands from unix command line for unit testing.
• Written complex queries for validating transformation logic after loading into tables.

Organisation: Atos | Syntel (April 2017 to March 2020)   **Client: State Street**

Project Description: (April 2018 to March 2020)

To Implement GRC Framework for one of the State Street Business processes with structured workflows. We collect data from respective applications and generate reports which are used by business users for studying risk exposure on business. It deals with reporting operational risk information to business entities, senior management, and the board of directors.

Technical skills: IBM Infosphere DataStage, Oracle, MSSQL, Shell Scripting and Tableau.

Responsibilities:
• Analyzing user requirement and mapping logic from fsd documents.
• Designed parallel jobs with the finalized transformation logic.
• Designed the sequence, batch, and main flow jobs as per dependency logic in requirement.
• Loaded the flat files data into tables by applying the transformation logic.
• Used information server manager for packaging the developed jobs and deploying them into different environments.
• Implemented transformations like aggregations, date transformations, calculative columns.
• Scheduled the jobs using running director to run jobs and load the data at specified time.
• Analyzed the log information for fixing the runtime errors.
• Extensively used sql queries in validating transformation logic on target table columns.
• Developed effective reports and dashboards in tableau based on requirement by using calculative fields, parameters, sets, dual axis, lod's, actions, and filters etc.
• By making use of marks card provided necessary information by making use of color, shapes, labels, and sizing also inserted sheets in tool tip information.
• Created visualization like bar chart, donut chat, pie-chat, bubble chart, trend charts and effective dashboards etc., then published into work environments like DEV, SQA, UAT and PROD.

Project Description: (June 2017 to April 2018)
This project is all about maintaining the various levels of risk data that are evaluated upon running designed data principles to evaluate the risk and provides information for making decisions.

Technical skills: IBM Infosphere DataStage, Oracle, MSSQL, Shell Scripting and Tableau.

Responsibilities:
• Analyzing user requirement from FSD Documents and user inputs from meetings.
• Build interactive dashboards by connecting with different data sources like db connections, flat files, web connectors etc.
• Formatted the reports and dashboards like background changes, object arrangement, use of buttons & providing navigations, number formatting and tooltip etc.
• Used date functions, string functions in both sql and tableau for validating data and making custom calculations in reports.
• Created visualization like bar chart, donut chat, pie-chat, bubble chart, trend charts and effective dashboards etc., then published into work environments like DEV, SQA, UAT and PROD.
• Having experience on tableau server to scheduling reports which are using live connections to refresh the data at specified time.
• Used the reference lines in tableau under worksheets for representing the specific information.
• Assigned custom images for representing data in reports using shapes option in worksheet.
• Published the developed reports into respective environments with user permissions.

<u>Training: (April 2017 to June 2017)</u>

In training I got trained on Full Stack Development with java with courses like core java, J2ee, Unix, sql, angular etc., after completing the courses we are asked to solve one case study to build an application where user can check the availability of table in a restaurant and book of a specific time.

Responsibilities:

- Attended the instructor led training and cleared all the exams conducted by the training department.
- Prepared the design document for the given case study.
- In the meetings we discussed implementing the case study and getting the inputs on technology stack to use
- A group of training mates shared our work and completed the case study.

Declaration:

  I hereby declare that the information that I have furnished is authentic and true to the best of my knowledge.

<div align="right">G.N.V Satya Narayana</div>