**NAVEEN KRISHNA DEVULAPALLY  
Data Engineer**(217) 441-0047 **|** [naveenkrishnadevulapally18@gmail.com](mailto:naveenkrishnadevulapally18@gmail.com) | [linkedin.com/in/naveenkrishna620/](https://www.linkedin.com/in/naveenkrishna620/) **|** [github.com/naveenkrishna-d](https://github.com/naveenkrishna-d)

**SUMMARY**

* Results-driven Data Engineer with over 4+ years of professional experience in designing, developing, and optimizing data pipelines and architectures.
* Proficient in a wide range of programming languages including Python, Scala, R, and SQL, and experienced with essential data science packages such as NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, and TensorFlow.
* Skilled in ETL and data integration using Informatica, Talend, SSIS, Google Dataflow, AWS Glue and Azure Data Factory (ADF), as well as managing big data technologies within the Hadoop Ecosystem.
* Adept at leveraging cloud technologies from AWS, GCP, and Azure to deliver scalable and efficient data solutions.
* Expertise in database management with PostgreSQL, MS SQL, MySQL, Cassandra, and DynamoDB, coupled with advanced capabilities in data visualization using Tableau, QlikView, and Power BI.
* Proficient in Unix Shell Scripting for automating tasks.
* Committed to continuous improvement and innovation in data engineering, utilizing a diverse array of technologies including version control (Git, GitHub), data cataloging (Alation), data analytics (ThoughtSpot), real-time data integration (Striim), job scheduling (Control-M, Airflow), SQL management tools (MySQL, PostgreSQL, SQL Server), cloud platforms (AWS, GCP, Azure), big data processing (Apache Spark, Hadoop), containerization (Docker, Kubernetes), and CI/CD pipelines (GitLab CI/CD) to ensure robust, maintainable, and high-performance data systems.

**PROFESSIONAL EXPERIENCE**

**KeyGlee Tempe, AZ, USA**

**Data Engineer Feb 2024 – Current**

* Designed and implemented a multi-cloud data pipeline integrating AWS services with Google Cloud Platform to optimize data storage and processing workflows.
* Maintained Amazon DynamoDB tables for real-time data storage needs, ensuring low-latency data access and high throughput performance.
* Leveraged Apache Hadoop ecosystem (HDFS, MapReduce, Hive, HBase) for distributed storage and processing of large datasets, optimizing performance and scalability.
* Produced interactive dashboards and reports by applying Power BI and SSRS, enabling stakeholders to visualize key metrics and trends.
* Orchestrated the cross-cloud data transfer from AWS to GCP using AWS-native tools and Google’s Data Transfer Service, ensuring data consistency and integrity.
* Created and managed Cloud Storage buckets in Google Cloud (data-lake-bronze) as part of a tiered data storage strategy, facilitating efficient data lifecycle management.
* Implemented Cloud Functions in Google Cloud to automate ETL tasks and data transformations, reducing manual effort and speeding up data availability for analytics.
* Automated data workflows and enhanced real-time processing by leveraging AWS Lambda for serverless computing, along with AWS SNS for efficient notification services.
* Monitored data quality and implemented data validation checks using cloud functions, guaranteeing the accuracy of data for business intelligence applications.

**University of Illinois Springfield Springfield, IL, USA**

**Graduate Student Assistant Oct 2022 - Sep 2023**

* Developed an AI-based system to predict student performance and identify at-risk students to provide timely interventions.
* Utilized Scikit-learn and TensorFlow to create predictive models based on student demographic data, academic records, and engagement metrics.
* Collaborated with the university’s administration to gather and preprocess data from various sources, including student information systems, learning management systems, and survey results.
* Deployed machine learning models into production using Azure functions and integrated them with the university's existing systems for real-time predictions. Successfully identified at-risk students with an accuracy rate of over 85%, enabling the university to implement targeted support measures and improve overall student retention rates.
* Developed a recommendation system to suggest personalized interventions and resources for at-risk students, improving their academic outcomes.
* Worked closely with faculty members and academic advisors to refine models and ensure their relevance and accuracy in predicting student performance.
* Ensured the privacy and security of student data by adhering to FERPA guidelines and implementing robust data handling practices.

**Accenture | Client: CNA Insurance Hyderabad, India**

**Application Development Analyst Jun 2020 - Jul 2022**

* Supported the migration of an on-premises legacy data warehouse to Google Cloud Platform, ensuring data integrity through validation checks, directly contributing to improved data accessibility for business operations.
* Assisted in designing data pipelines for streaming and batch processing, utilizing Google Cloud Pub/Sub and Cloud Dataflow under senior supervision.
* Implemented the setup and ongoing maintenance of a Big Query-based data lake, transforming and staging 10TB+ datasets to enable 25% quicker data retrieval for analytics purposes.
* Engineered ETL scripts to automate data extraction from diverse sources such as sales, marketing, and manufacturing databases, impacting over 50 different data streams.
* Developed stored procedures in MS SQL to fetch data from different servers using FTP and processed files to update tables, improving the accuracy of data
* Developed and coordinated Control-M workflows to schedule and automate ETL jobs to kick off on time for data availability
* Migrated an Oracle database to BigQuery using Striim for real-time data, and configured GCP services including Dataflow, Storage, and BigQuery using cloud shell SDK, enhancing data availability, accessibility, and ensuring reliable, scalable data processing.
* Built data pipelines in airflow on GCP for ETL related jobs using different airflow operators, ensuring efficient data processing and performing automation
* Orchestrated the establishment of continuous data synchronization protocols between cloud and on-premises environments, achieving data accuracy and supporting seamless monthly data conversion and validation cycles.
* Performed rigorous data quality checks, maintaining accuracy of data in the staging and reporting layers of the data warehouse, thus boosting data reliability for strategic decision-making.
* Refined SQL queries and data modeling, contributing to the development of both star and snowflake schema designs in a cloud environment.
* Contributed actively in agile team meetings, influencing the planning and execution of data engineering projects, which led to a 10% improvement in project delivery timelines.
* Enforced stringent data governance and security measures in collaboration with the data operations team, ensuring compliance with company policies and data protection regulations.

**KPMG India**

**Data Analyst Aug 2019 - May 2020**

* Achieved improvement in operational efficiency by conducting in-depth data analysis and creating visualizations with Python, NumPy, and Pandas, resulting in more informed business decisions.
* Increased reporting accuracy by developing and maintaining interactive Tableau dashboards, enhancing the organization's data-driven decision-making process.
* Reduced data retrieval time by utilizing SQL Server to design and implement complex ETL processes, ensuring data integrity and accuracy across multiple databases.
* Improved data quality and consistency by performing comprehensive data cleaning and wrangling tasks, facilitating more accurate and reliable analysis.
* Enhanced data accessibility and management efficiency by designing and implementing data warehousing solutions using best practices, enabling more effective data analysis and reporting.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming Languages:** | Python, Scala, R, SQL |
| **Packages:** | NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow |
| **ETL and Data Integration Tools:** | Informatica, Talend, SSIS, Azure Data Factory (ADF) |
| **Big Data Technologies:** | Hadoop Ecosystem (HDFS, Sqoop, HBase, Hive, MapReduce), Apache Spark, Kafka |
| **Reporting Tools:** | Tableau, QlikView, Power BI, MS Excel |
| **Database:** | PostgreSQL, MS SQL, MySQL, Cassandra, DynamoDB |
| **Cloud Technologies:** | AWS (S3, DynamoDB, Glue, Lambda, SQS, SNS, EMR, Kinesis), GCP (Cloud Storage, PubSub, Big Query, Dataflow), Azure (Data Factory, Data Lake, Azure Files) |
| **Scripting and Development:** | HTML, CSS, JavaScript, Unix Shell Scripting |
| **Version Control:** | Git, GitHub |
| **Tools:** | Alation, ThoughtSpot, Striim, Control-M, Airflow, TOAD, SecureFX, SecureCRT, SQL Server  Management Studio, Oracle SQL Developer, VS Code |
| **IDE’s:** | PyCharm, Jupyter Notebook |
| **Operating Systems:** | Windows, Linux |

**EDUCATION**

**University of Illinois Springfield, IL**

Masters of Science in Data Analytics Aug 2022 - Dec 2023

**Sreenidhi Institute of Science & Technology Hyderabad, IND**

Bachelors in Computer Science Aug 2016 - May 2020

**PROJECTS**

**Food Calorie Estimation Model**

* Developed food calorie estimation tool using Convolutional Neural Network by Image identification and classification of food items using CNN deep learning techniques

**Data encryption and decryption**

* Performed secret data transmission of text and image files using ‘AES’ algorithm in Python

**Graduate admission prediction**

* Implemented seven regression models and compared model performance including Lasso, Ridge, Elastic Net, Random Forest, Gradient Boosting, Linear SVM, and Radial SVM, to analyze and predict graduate admission probabilities.

**Car make and Model detection**

* This project employs Deep Convolutional Neural Networks to identify car makes and models. It is aimed at enhancing vehicle recognition technology using AI.