Neeraj Reddy Bandi
New York | (585) 290-2136 | neerajreddy47@gmail.com| Portfolio | Linkedin

EDUCATION

Master of Science: Rochester Institute of Technology   Rochester, NY (Aug 2022 – Dec 2024)
Major in Information Technology and Analytics
Bachelor of Technology (Honors): NIT Jamshedpur   Jamshedpur, India (Aug 2014 - May 2018)
Major in Metallurgical and Materials Engineering

CERTIFICATIONS

- **Azure Data Engineer associate certification**
- **Databricks Certified Associate Developer for Apache Spark**
- **Certified Scrum Master with Proficiency in Agile Methodologies.**

SKILLS

Languages: Python, Java, SQL, Python, PySpark, JavaScript, XML, R, Selenium
Big Data Cloud Technologies: AWS (EC2, S3, EMR, Redshift), Azure (Data Factory, LakeGen2, Data Bricks, Synapse), Snowflake, Hadoop, REST APL, ETL/ ELT, Pandas, Scala, Data Warehousing, LLM, Kafka, Data Modeling, DBT
DevOps: Airflow, Azure DevOps, CI/CD, Configuration Management, Jira, Jenkins, Git, Docker, Terraform, Kubernetes
Database: SQL server, SSIS, Postgres, NoSQL (MongoDB, DynamoDB, Cassandra), Vector Database, Ne04j
Developer Tools: SAS, Tableau, PowerBI, qGIS, Plotly Dash, shiny, Streamlit, Grafana, Looker
Operating System: Windows, Linux, MAC, Unix Shell Scripting

EXPERIENCE

Rochester Institute of Technology, Graduate *Teaching Assistant*   **Aug 2023 - Present**

- Implemented Spark jobs for distributed data processing using Python – Pyspark and SQL – HiveQL in HDFS within Hadoop framework, enhancing data cleaning and data **integration efficiency by 5x.**
- Designed and implemented a big data warehouse using the **Kimball approach**, creating **STAR schema-based** dimension and fact tables to build data marts and lakes for solving **supply chain challenges.**
- Prepared **detailed documentation** of system requirements and development changes for students.

Verticle Technologies, *Data* Engineer | (Hyderabad, India)   **Aug 2019 - July 2022**

   Project: Instant Gratification for Customer Concern Streaming Architecture

- Designed ETL pipelines with **Kafka** to extract **20M** real-time e-commerce events, reducing **latency by 30%** and improving fraud decision-making by **7%**.
- Optimized **Spark** queries to process **1TB** of daily data, cutting **runtimes by 83% (30 min to <5 min).**
- Implemented ETL workflows using **SSIS** for data ingestion from various RDBMS and flat file sources, integrating data from over **15 sources** and processing **500GB** of data daily.
- Developed a CI/CD pipeline using **Jenkins and Azure DevOps** to integrate **RESTful APIs** with real-time data processing components, **achieving zero-downtime deployments**, ensuring **99.9% uptime** through automated alerts and boosting deployment efficiency by **40%.**
- Automated Linux virtual machines provision in **Azure** via **shell scripts**, reducing manual intervention **to 0%** in production servers.

   Project: Streaming Live DLT Pipelines on Databricks

- Ingesting **1M** Broker Submission Data to streaming delta live table **(DLT)** Pipelines on databricks to generate and produce quote information for underwriters on a real time lead generation on salesforce systems.
- Implemented **XML and JSON** data mapping into Delta Live Tables, integrating seamlessly with existing ML models, achieving a **15% reduction** in manual data preprocessing time, and improved pipeline **scalability** to handle **100GB** of data per day.

Volvo Trucks India, Product Data *Analyst* | (Bangalore, India)   **May 2018 - Jul 2019**

   Project: Data-Driven Marketing Optimization for VECTO India Launch

- Built Predictive ML Models using **SAS** to identify key drivers influencing vehicle purchases across three customer segments, increasing targeting precision by **20%** and improving high-potential customer engagement.
- Crafted dynamic dashboards and comprehensive reports in **Power BI**, leveraging advanced DAX formulas to empower stakeholders with actionable insights and support strategic decision-making.
- Orchestrated the successful Indian market launch of **VECTO**, partnering with a cross-functional team of **19 members** to customize, validate, and implement **16 new features**, aligning with regional requirements and customer preferences.

PROJECTS

Cold Email Generator using Vector DB and Llama (GitHub)

- Integrated **LLM-powered processing** to match job requirements with user portfolios, automating the creation of tailored cold emails with portfolio links
- Designed a vector-based portfolio storage system in **Chroma DB** to achieve **fast and efficient skill-to-job matching** with millisecond response times.

Enhanced LLM Application with Real Time Data Pipeline for recalled products (GitHub)

- Developed a scalable data pipeline, integrating **GPT-4 LangChain** with **ReACT** agent for **5x improved query generation**, and deployed via Streamlit with PostgreSQL for reliable storage.