Harinath Reddy Mandha

harinathreddy0035@gmail.com | https://www.linkedin.com/in/harinath-mandha/ | 571-634-2774 | Faifax, VA. 22030.

- Data enthusiast and a results-driven data science professional with Masters in Data Analytical Engineering
- Proficient in Python, SQL, R and SAS for data cleaning, data wrangling, statistical analytics and visualizations.
- Hands-on experience in building scalable data pipelines using Pyspark, DBT, Airflow, and Apache Kafka.
- Possess extensive hands-on experience in managing and constructing robust infrastructure, as well as deploying comprehensive storage solutions across leading cloud services, including GCP, Microsoft Azure, and AWS.
- Expertise in performance tuning SQL queries by analyzing code, optimizing indexes, and collecting statistics.
- Well-versed in AWS technologies, including EC2, S3, Redshift, Lambda, Athena, RDS, and DynamoDB.
- Integrated PySpark with DBT to build robust data pipelines, streamlining data ingestion, transformation, and loading processes in a cloud environment. Implemented data lineage tracking and documentation using DBT.
- Designed and built scalable ETL pipelines and data workflows to migrate the enterprise data to cloud data warehouse solutions like Snowflake in order to reduce the redundancy and to improve query performance.
- Conducted performance tuning and optimization of PySpark jobs, reducing resource consumption and improving overall data pipeline efficiency, resulting in improved throughput and latency.
- Demonstrated ability to integrate and optimize data pipelines across multiple systems and platforms, ensuring data consistency, integrity, and security. Holds data governance best practices that meet industry standards.
- Skilled in integrating and extracting NoSQL databases with data pipelines and (ETL) Extract, Transform and Load processes to efficiently manage and analyze unstructured and semi-structured data in data lakes.
- Extensive experience in Production support and On-call for troubleshooting and monitoring data pipelines. Holds practical experience in Kubernetes, Docker Containers, Jenkins CI/CD and other devops principles.
- Possess strong skills in statistical modeling, dimensionality reduction, feature engineering, and machine learning operations and exploratory data analysis, adept at addressing big data and machine learning challenges.
- Extensive experience in big data processing and analytics using AWS tools like Glue, EMR and Kinesis.
- Extensive experience in building advanced machine learning models and deep neural networks for Computer Vision, Recommendation systems, Segmentation, Time-series modeling and Natural Language processing NLP.
- Proficient in utilizing advanced features of visualization tools and libraries such as Tableau, PowerBI, Matplotlib, Plotly & Shiny to translate complex datasets into visually compelling stories and actionable insights.
- Expert in utilizing tools such as JIRA and confluence for project management and Git for version control to ensure seamless team coordination and efficient workflow. Leveraged markdown files to create documentation.
- Experience with both Agile and Waterfall methodologies in the Software Development Life Cycle (SDLC).
- Possess strong interpersonal skills and ability to effectively handle projects involving cross-functional teams. Demonstrated excellence in communication, problem-solving, and critical thinking to solve business problems.
- Passionate about Data Science, driven by continuous learning, problem-solving, and adding value to businesses with attention to detail. Strong Visual storytelling and Project Management skills.

TECHNICAL-SKILLS

PROGRAMMING LANGUAGES: Python, R, SQL, and SAS.
CLOUD: AWS, AZURE, GCP, DATABRICKS, SNOWFLAKE AND OCI.
VISUALIZATION TOOLS: TABLEAU, POWER BI, R-SHINY.
LIBRARIES: TensorFlow, PyTorch, Matplotlib, Scipy, Pyspark, Plotly, NumPy, Pandas, and scikit-learn.
ETL TOOLS: Dataflow, AWS Glue, Synapse Analytics and DBT.
BIG DATA TOOLS: Spark, kafka and Airflow
ML/DL APPLICATIONS: Customer Segmentation, Recommendation systems, Predictive Analytics, Time-series Forecasting, Computer Vision, NLP and Optimization.

TOOLS LEARNING: LLMs, Retrieval Augmented Generation (RAG), Langchain, knowledge Graphs and building hugging face portfolio.

EXPERIENCE

Data Engineer,
U.S. Bank. **May 2022 - April 2024.**

Collaboratively handled an end-to-end cloud data migration project, from deploying storage solutions, and designing data warehouses to ensuring optimal storage conditions and data integrity in a cloud environment.

- Actively contributed to developing a CRM tool by utilizing SQL extensively for database querying, data modeling, and dimensional modeling, enhancing data accuracy and reporting capabilities by **18%.**
- Contributed to developing a predictive model for incident analysis within ITSM ServiceNow using Python and PySpark, resulting in a notable **20%** decrease in SLA breaches by estimating the incident age.
- Acquired hands-on expertise in ETL using Spark, DBT and Airflow. Actively contributed to the application development, honing skills in building, crafting, testing and debugging big data pipelines.
- Demonstrated proficiency in cloud data engineering, with hands-on experience in building Cloud Data Warehouses using technologies such as Snowflake, Redshift, and BigQuery.
- Gained experience in big data ecosystems Apache Spark, Apache Kafka and Map reduce for data processing.
- Contributed to employing Amazon Athena and PySpark to perform ad-hoc SQL queries on large datasets stored in Amazon S3, reducing time-to-insight from hours to minutes.
- Implemented CI/CD pipelines using Terraform, Docker, and Kubernetes, ensuring seamless deployment and scalability of data solutions in a cloud environment.
- Configured Amazon RDS to support transactional databases with automated backups, monitoring, and scaling, enhancing database reliability and performance.
- Deployed Amazon Elasticsearch Service to analyze and visualize log data, improving system monitoring and troubleshooting capabilities with PySpark for data pre-processing.
- Utilized Splunk and ELK to create dashboards and visualizations, transforming transactional data into actionable key performance indicators (KPIs). This enabled real-time monitoring and comprehensive analysis of crucial business metrics.

Data Analyst,
Reliance Digital retail LTD. **January 2019 - July2021.**

- Conducted sensitivity analyses to assess the impact of changes in key parameters (such as demand variability, holding costs, and ordering costs) on the inventory optimization model reducing combined costs by **$150K.**
- Contributed to the development of a recommendation engine using collaborative filtering techniques in PySpark, leading to a **20%** increase in cross-selling revenue. Actively monitored KPIs and effectively communicated strategic business decisions to stakeholders
- Designed and deployed a real-time data ingestion pipeline using AWS Kinesis and AWS Lambda to monitor and adjust key inventory parameters dynamically.
- Utilized AWS Glue and Amazon Redshift Spectrum in conjunction with PySpark for advanced data cleansing and pre-processing, improving data transformation run time by **12%.**
- Designed and implemented a normalized database schema for an e-commerce application using MySQL, improving data integrity and reducing redundancy by **6%.**
- Utilized Python to conduct statistical analysis for consumer behavior insights and to optimize marketing strategies. Enhanced decision-making contributing to overall effectiveness, increased sales by **8%.**
- Collaboratively executed advanced data cleansing and pre-processing techniques using Python and DBT to transform raw data into a structured format, improving transformation run time by **12%.**
- Developed advanced data visualizations and interactive dashboards using Azure Synapse Analytics and Azure Databricks, delivering near real-time analytics for enhanced decision-making and actionable business insights.
- Leveraged Snowflake for high-performance data warehousing, supporting business intelligence, reporting, and analytics requirements, ensuring scalability and reliability of data storage and processing.

EDUCATION AND CERTIFICATIONS

MS in Management, St.Francis College. NY **Finance, Business Analysis, Project Management.**
Masters in Data Analytics Engineering, George Mason University. VA. **Statistics, Machine Learning, Probability.**
Bachelor in Computer Science, GITAM University. **Discrete Math, System Analysis, Databases, Python.**
Google certified Database Engineer **2024**
Applied Data Science Professional (MIT) **2023**
Graph Data Science Certification (Neo4J) **2023**

PROJECTS

Project 1: Malaria Detection with CNNs @MIT Professional Education.
This robust malaria detection model is built using Convolutional Neural Networks (CNN). Leveraging powerful Python libraries such as TensorFlow, PIL, and scikit-learn, I've developed a model capable of accurately identifying parasites (infections) in blood smear images. This comprehensive approach includes preprocessing with NumPy and pandas, data visualization with Matplotlib and Seaborn, and efficient model training utilizing advanced CNN layers like Conv2D, MaxPooling2D, and Dense layers.
Skills: Deep Learning, Computer Vision, Neural-Networks, Image Processing, Data Augmentation, Python.

Project 2: Covid-19 Vaccination Data Visualization associated with Prof: Lily Wang at GMU.
Developed a robust Shiny app featuring interactive choropleth maps, time-series plots, and other visualizations. Leveraged CDC-health data to perform data wrangling, visualization, and exploratory data analysis (EDA) on vaccine distribution across the United States.
Skills: Statistical Modeling, R programming Language, Dashboards, Data Analytics, Visualizations.

Project 3: Case Study on Electrify America's Charging stations
This case study utilizes Python, K-means clustering, Geocoding API and NaturalEarth Shape file to identify and display optimal locations for future Electrify America charging stations in Virginia. I've integrated Google Cloud Platform's GeoCoding API for precise location display on my plots. The approach offers strategic insights to support Electrify America's expansion efforts, promoting EV adoption in Virginia.
Skills: **Python, Machine Learning, Clustering Analysis, Data Visualizations**

Project 4: Data-Driven Excellence in Food Aggregation
As a data analyst, I conducted a comprehensive Exploratory Data Analysis (EDA) on the dataset from a prominent food aggregator company. The dataset encapsulates a myriad of orders made through their online portal, providing valuable insights into the dynamics of customer preferences and restaurant performance.
Skills: **Exploratory Data Analytics, Python, Data Visualization, SQL.**

Project 5: Predictive Analytics for Patient Length of Stay Optimization
Spearheaded a comprehensive Predictive Analytics initiative focused on deciphering the nuanced factors influencing the Length of Stay (LOS) for patients within a hospital setting. Leveraging advanced machine learning techniques, I engineered a robust predictive model capable of accurately forecasting the LOS based on admission data and key medical-test outcomes.
Skills: Statistical Modeling, Machine Learning, Python, Data Visualization.