

Niharini Nandella

DATA ENGINEER

Stillwater, OK | [7208823437](tel:7208823437) | niharininandella@gmail.com

SUMMARY

- Data Engineer with Around 4 years of experience in Python, PySpark, and R Data Extraction, Data Modelling, Data Mining, and Data Visualization.
- Proficient in data acquisition, validation, predictive modeling, and visualization using tools such as NumPy, Pandas, Matplotlib, SciPy, TensorFlow, and ggplot2.
- Expertise in SQL, database design, and ETL pipeline implementation, coupled with hands-on experience in AWS (EC2, Lambda, S3, Athena, Redshift, Appflow, Glue, Quicksight), GCP, and Azure (BlobStore, Data Factory, EventHub), showcasing expertise in leveraging diverse cloud services.
- Leveraging Tableau and Power BI to create impactful visualizations and reports demonstrating expertise in utilizing metrics, attributes, charts, filters, hierarchies, trendlines, sets, grouping, data blending, parameters, and complex calculations.
- Extensive experience in building Data Warehouses/Data Marts using ETL tools like Informatica and SSIS.

EDUCATION

Master of Science in Information Management | Oklahoma State University, OK
Bachelor of Civil Engineering | Maturi Venkata Subba Rao, Nadargul India

May 2024
Sep 2020

EXPERIENCE

JPMorgan Chase & Co., IL | Data Engineer

Dec 2023 – Present

- Optimized ETL data migration pipeline employing AWS Python Boto3, facilitating the transfer of 2TB of data from Salesforce to S3 buckets in Parquet.
- Formulated AWS Glue jobs by parsing JSON definition and used AWS Athena for SQL queries and **reducing service creation time from 5 hours to 8 mins**.
- Constructed MapReduce-based ETL pipeline to extract, transform, and load 7TB of customer data to a data warehouse, **reducing data ingestion time by 30%**.
- Leveraged **PySpark in Databricks with 12B** records in parallel to execute ETL tasks on sensor-generated time-series data.
- Built 10+ automated workflows using Databricks Jobs and notebooks, streamlining data pipelines and **reducing manual intervention by 28%**.
- Conducted in-depth analysis of customer complaints data using BERT and Google T5 sst2 deep learning models to identify negative sentiments on various product parts and recognize growth opportunities that improved yearly **Net Promoter Score by 13%**.
- Formulated **SQL queries for reports** to engage with stakeholders, designed data models, and presented data across **12+ Tableau dashboards**, delivering valuable global and regional market share insights and conducting competitor analysis.
- Improved data accuracy and **reduced data processing time by 38%** by leveraging PostgreSQL to create and manage extensive databases.
- Designed and implemented Snowflake-based scalable data warehousing, **reducing storage costs by 40%** and improving data accessibility for business users.
- Automated Alteryx workflows for computing warranty claim metrics over 14 M sales records to create a product insights dashboard in Power BI.

R-Financial, India | Data Analytics Engineer

Nov 2020 – Aug 2022

- Enhanced Opex and Capex datasets of 8M+ rows by automating EDA using Pandas along with a cross-functional team **reducing Excel processing time by 25%**.
- Implemented **Spark** for processing large-scale datasets, improved processing times by 35% compared to traditional frameworks.
- Collaborated with **15+ interactive Tableau reports**, reducing data analysis time by **20% and saving \$500K in marketing quarterly**.
- Optimized SQL for monthly report generation enhancing **Tableau dashboard** performance by **30%**, reducing report creation time from **4 hours to 12 min**.
- Integrated and centralized data, reducing redundancy in data by 45% and streamlining maintenance procedures, **achieving a 25% increase in team efficiency**.

XLogic Technologies, India | Data Engineer

Jun 2019 – Oct 2020

- Conducted ad-hoc data analysis from design to production, on AWS S3 Data Lake using SQL in Apache Hive environment to **increase FTE production by 10%**.
- Optimized PySpark code for performance and efficiency, and achieved an **18% reduction** in the processing time of the project.
- Implemented data modeling using SQL and Excel improving data accuracy and **reducing processing time by 40%** to facilitate data cleaning and preparation.
- Utilized Spark Structured Streaming for high-throughput processing of streaming data with **15% reduced latency** compared to traditional methods.
- Executed Alteryx ETL process, enabling extraction from diverse sources (SQL Server, Excel, CSV) and scheduled workflows for efficient data management.
- Implemented dynamic scheduling and dependencies between **Airflow** tasks for efficient workflow execution.

KEY PROJECTS

Sentiment Analysis on Gun Violence in the U.S. Using Twitter

Jan 2023 – Mar 2023

- This project's main objective has been to analyse people's sentiments and emotions towards Gun Violence based on their tweets on Twitter using hashtags and managing database integrity, data generation, loading, query writing to fetch data, and performance tuning using indexes.
- Employed Python, and NoSQL (MongoDB) to scrutinize Twitter datasets of over 540M rows by conducting analysis.

ETL and Data Management for Retail Store Analytic

Oct 2022 – Nov 2022

- Engineered comprehensive ETL pipeline, managing the extraction, transformation, and loading of 9M rows of data, ensuring accuracy and quality in data migration and warehousing.
- Deployed a resilient ETL data pipeline to migrate a sizable dataset to AWS S3, resulting in a notable 40% decrease in data transfer latency.

SKILLS

Programming Language: Python, R, SQL, PySpark

Big Data Ecosystem: Hadoop, MapReduce, Pig, DynamoDB, HDFS, Spark

Databases: MS SQL Server, PostgreSQL, Azure Databricks, MongoDB

Visualization Tools: Tableau, Power BI, Excel, Looker

Packages: NumPy, Pandas, Matplotlib, SciPy, ggplot2, TensorFlow, Seaborn, Scikit-learn

Cloud: AWS (EC2, Lambda, S3, Athena, Redshift, Appflow, Glue, Quicksight), GCP, Azure (BlobStore, Data Factory, EventHub)

Tool: Google Analytics, Airflow, Git, Snowflake, SSIS, One Trust, Service Now, Power Automate

ETL Skills: Data Cleaning, Data warehousing, Data Wrangling, Predictive analysis