Neha Nagle

Data Engineer

☐ (+91) 7768826331 | ✉ nehanagle45@gmail.com | ⌨ neha | in nehanagle | **Location:** Pune, Maharashtra, India

Experience [3.9 years]

Michelin   *Domain: Product Manufacturing*

Data Engineer (Big-Data & Azure)   *Nov'2021 - Present*

- Assessed data closely with product managers and developed **data pipeline** along with **dashboard** to help business decisions.
- Automated **Web Scraping** from multiple websites to analyze Rate of change**(ROC)** for supply & demand of the product month on month basis.
- Designed batch-pipeline with the help of **Azure Data factory (ADF)** and performed ETL within **databricks** by using **Pyspark**,**SQL**,**Python** and sent notifications with appropriate metrics leveraging **logicApp**. Implemented the same with multiple data sources on multiple projects.
- Streamlined various data sources and **designed schema** alongwith data dictionary to efficiently capture and integrate the data accurately.

Tata Consultancy Services   *Domain: Healthcare, Logistics, BFSI*

Data Engineer (Analytics)   *Jan'2019 - Oct'2021*

- Developed pipeline for converting healthcare sales data into actionable insights with the help of **python** scripts and **SQL**.
- Spearheaded in-depth **analysis** of sales details to avoid discrepancies in operation and maintain data integrity across system.
- Performed **Exploratory Data Analysis** using **python** to interpret data sets, paying particular attention to trends and patterns
- Implemented a **POC for de-mystification of the documents** cluster containing medical supply information as well as patient information by using Entity recognition **NLP** model as well as **object detection** algorithm to capture required information for business analysis
- Automate the message Queue Over 70 hours per week were saved by using Python scripting for queue monitoring
- Responsible for **Root Cause analysis** of production-related failures by analyzing logs with linux command
- Administered and **mentored** new associates for domain knowledge and Technology.

Skills

Programming: Python, Pandas, Numpy, Pyspark, SQL,Scala

Big Data: Spark, Hadoop, Hive, ETL, Sqoop

Cloud: Azure Data Factory, Azure Data Lake, Azure Blob Storage,Azure Synapse Analytic, GCP, AWS

Database: Azure SQL Database, MySQL, DynamoDB, MongoDB

Other Tools: Jupyter, Tableau, Kafka, Rest-API, Excel,Git, Microservices,Databricks

Soft Skills: Teamwork, Communication, Problem-Solving, Flexibility,Leadership

Personal Projects

Youtube Data Analytics   *Big Data, AWS*

Tech Used: Python, PySpark, SQL ,AWS   *August 2021*

- Implemented a data intensive platform where youTube user interaction data can acquire and visualise factors that determine the popularity of videos on YouTube.
- Built a datawarehouse by moving the data to Redshift to use it for further for dashboarding purposes.

Real Time Fraud Detection   *Big Data*

Tech Used: PySpark, Hadoop, Hive, Kafka, Airflow, SQL   *July 2020*

- Implemented to verify several business rules to determine whether the transaction took place was fraudulent or genuine

Certifications & Courses

2022	**DP - 203 Data Engineering on Microsoft Azure**, Learning: Azure Services	*WizLab*
2022	**Big Data Masters**, Learning: Hadoop,Spark,Hive, Kafka	*Trendytech*
2022	**Microsoft Azure Fundamentals DP-900**, Learning: Services provided by Microsoft	*Udemy*
2022	**Azure Databricks Fundamentals**, Learning: Delta Table,Data Warehouse	*Databricks*
2021	**Python for Data Science and ML Bootcamp**, Learning: Data Science and Analysis	*Udemy*
2021	**Python and SQL**, Certification on Problem Solving	*HackerRank*

Education

RGCER   *Nagpur, Maharashtra, India*

B.E.(Hnours) in Electrical Engineering, [CGPA: 8.21/10]   *Jul'2013 - Jul'2017*