ANURAG THAKUR

DATA ENGINEER | MICROSOFT AZURE CLOUD

PROFILE SUMMARY

- Total of 2 years of working experience as Data Engineer and Cloud platforms Azure & AWS.
- Experience with ETL using Python, PySpark, Kafka, Big Data, Java, SQL, AWS Quick Sight etc.
- Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) & Azure SQL DB).
- Well-acquainted with data processing and data engineering pipelines.
- Experienced on using Cloud services like Azure Data Factory (ADF), Databricks & Compute Clusters
- Experienced with SQL.
- Worked in project with Agile methodology.
- Quick understanding of business problem and have good analytical skills.
- Experienced in working with various Python IDE like PyCharm & Jupyter Notebook, etc.
- Excellent at report generation, following process-flow, project documentation.
- Good communication skills.

EDUCATIONAL CLARIFICATIONS

Bachelor of Science in Data Science & Big Data Analytics   MIT World Peace University, Pune, India.

WORK EXPERIENCE

3 years of experience as Data Engineer and Cloud platforms Azure and AWS.

TECHNICAL SKILLS AND TOOLS

Cloud Computing:	Validation activity in Azure Data Factory and Synapse Analytics pipelines
	Have good experience working with Azure BLOB and Data Lake storage and loading data into Azure SQL Synapse analytics (DW).
	Microsoft Azure - Azure Synapse Analytics (DW), Cloud Services (PaaS & IaaS), Active Directory, Application Insights, Azure Monitoring, Azure Search, Data Factory, Azure Analysis services, Azure Data Lake.
Python Libraries	NumPy, Pandas, PySpark, Matplotlib, Seaborn, Scikit-learn, OS etc.

Cloud Platforms	Microsoft Azure & AWS.
Tools	pgAdmin4, RapidMiner, Kedro Framework, etc.

PROJECTS

Project - 1 Designing and implementing Sales Prediction Model architecture on Azure Cloud Services.

Role	Data Engineer
Description	Designing and implementing the Sales Prediction Model Architecture on Azure Data Bricks.
Responsibilities	- Created various ETL pipelines using ADF. - Deploying scripts on Azure Databricks. - Connecting to Azure SQL Server and CRM. - Securing credentials through Azure Key-vault. - Triggering Python notebook and scripts from Azure Data Factory by creating proper pipelines. - Scheduling Jobs.

Project - 2 Data Validation & Migration to Salesforce using Python, Azure Data Factory and Databricks

Role	Data Engineer	Azure
Description	Design, develop and deployment of an on-demand data loader module to validate and migrate data.
Responsibilities	- Performed ETL on data. - Created data validation rules (Config files). - Implemented python validation for each attribute including Picklist, Multipicklist, Email, Contact Number etc. - Migrated data to Azure Cloud Storage.

Project 3: Thyroid Disease Prediction System Using Machine Learning Technique

Roles	Data Engineer

Description	Developing a model for early detection of Thyroid based on the stages.
Responsibilities	- Pre-processed huge dataset consisting of 40 clinical variables by Data Cleaning & handling missingness, Handling Data Imbalance and Dividing the data. - Annotated each image with 2 classifiers and trained with 10000 steps with batch size 2. - The information is separated into three sections: - Equipment for training (70 percent) - Validation checklist (15 percent) - Make a test set (15 percent) - Created different data engineering pipeline for DS team.

Project 4: Sentiment analysis on Women Clothing Review

Roles	AWS Data Engineer| AWS Cloud Storage
Description	Sentiment analysis provides a way to various businesses (here, e-commerce) to views of customers to their product or services to understand the correlation of various variables in customer reviews of women's clothing and their classification into positive, negative or neutral.
Responsibilities	- Implemented data Pre-processing - Performed Exploratory Data Analysis and removed Outliers. - Exploratory Data Analysis and Feature Analysis. - Wrote a python script to extract files Amazon S3 Bucket; Created NLP based preprocessing techniques like tokenization, lemmatization, and vectorization