

Tanisha Negi

Application Development Analyst

Email Id: tanishanegi9899@gmail.com

Location: Gurugram, India

Mobile: +91-7906685089

Professional Summary

An enthusiast with 2 years of experience in Data Engineering. Working with large data sets to meet the business requirements along with the exposure to cloud platform in Ecommerce domain to increase sales and productivity. Along with this, having basic knowledge of ML concepts.

Skills

- **Programming Language:** Python, SQL
- **Framework:** Apache Spark (PySpark, Spark SQL)
- **Python Libraries:** Numpy, Pandas, Matplotlib, Seaborn, NLTK
- **Machine Learning (Basic):** Linear Regression, Decision Tree, Random Forest, Collaborative filtering
- **Familiar With:** AWS (Redshift, S3 and EMR), Kubernetes, Unix Shell scripting, Unit Testing

Professional Experience

Accenture

July 2019 – present

- Understanding the business requirement and utilizing python in spark and SQL operations to get right set of data from the large data sets.
- Strong experience in building SQL queries using views, aggregations and windows function and retrieve data to find anomalies.
- Use **PySpark** and **Spark SQL** to create complex mapping using different transformations like filter, joiner, expressions and aggregators to pipeline data to Redshift.
- Build data pipeline to load data from **S3 bucket** into **Redshift database** tables and perform SQL operations.
- Change data Capture (CDC) and slowly changing Dimensions (SCD II) to improve the performance of system.
- Involved in design and maintenance of dimensional models like Star Schema and Snowflake Schema.
- Schedule production jobs for batch and streaming processes.
- Currently involved in utilizing unit testing on Spark code to test and validate the results for variety of usecases.
- Implementation of Python libraries on a large set of transformed data to get the graphical representation for further analysis.
- **Segregation of text data and perform topic modeling-** worked on text data and performed basic NLP preprocessing-cleaning, transformation, tokenization, stemming and lemmatization to abstract topics from a dataset.

Education

- **Bachelor of Technology** in Electronics and Instrumentation from Banasthali Vidyapith, Banasthali 2019
- **Higher Secondary** from NDS, Rishikesh 2015

Certification

- Analyze Data with SQL ([certificate](#))
- Taming Big Data with Apache Spark and Python- Hands On! ([certificate](#))
- Kubernetes Certified Application Developer (CKAD) with Tests ([certificate](#))