



AVINASH KUMAR

DATA ENGINEER 📍 SIWAN, INDIA 📞 7888480687,9779938468

◦ DETAILS ◦

Siwan, India
7888480687,9779938468
kavinash267482@gmail.com

◦ LINKS ◦

[Linkedin](#)

◦ SKILLS ◦

Git
Data Analysis
ETL
SQL
Python
AWS
Airflow
Pandas
Numpy
PostgreSQL
PySpark
Tableau
Redshift
S3

◦ LANGUAGES ◦

Hindi

English

Bhojpuri

Punjabi

◦ HOBBIES ◦

Movie& Music, Travelling, Politics,

👤 PROFILE

As a skilled Python AWS Developer has a skilled Python AWS Developer with 5.1 years of practical experience, I am dedicated to utilizing my expertise in cloud computing and programming to make meaningful contributions to innovative projects. I am focused on using my Python development and AWS services skills to design strong solutions and propel business growth in a dynamic and challenging setting. With a strong passion for adopting emerging technologies, I am committed to continually improving my skills while delivering impact solutions that surpass client expectations.

📁 EMPLOYMENT HISTORY

Python Developer at Wipro Limited, Pune

June 2019 — May 2022

Developed automation scripts in Python and SQL to automate tasks and build an auto-ticket-solving system, reducing job workload by 30-40% and saving significant time. Conducted qualitative and quantitative analysis to identify data patterns and derive meaningful insights. Utilized Python for data cleaning, model application, identifying data patterns, and generating data visualizations.

Created over 20 Tableau reports and dashboards, leading to a 25% increase in operational efficiency. Proficient in understanding incidents and problems, analyzing tasks, and ensuring timely resolution to prevent SLA breaches.

Data engineer at Tata consultancy services, Pune

May 2022 — July 2024

Design, Development, and Integration: Designed and maintained ETL processes to extract, transform, and load data from various sources, integrating data from relational databases, flat files, and cloud storage into a unified repository.

Optimization, Quality, and Automation: Optimized ETL performance through query tuning and partitioning, ensured data accuracy with validation checks and cleansing routines, and automated workflows using tools like Apache Airflow.

Collaboration, Documentation, and Troubleshooting: Collaborated with stakeholders to gather requirements, created comprehensive documentation for ETL processes, and diagnosed and resolved workflow issues to ensure timely data availability.

Restart-ability Code in Python: Developed Python code by using Pandas to ensure restart-ability in case of failures. If the code encounters issues, it automatically retries upon fixing the problem. In case of failure, the code skips the problematic line, logs the error, and executes the remaining code.

SQL Query Generation and Execution: Created a module to generate SQL queries based on requests from various databases. Automated the execution of SQL queries within the Python environment, enhancing efficiency and reducing manual intervention.

Data Cleaning and Type Transformation: By Using **Pandas** Implemented data cleaning procedures to ensure data integrity and quality. Dynamically changed data types as per specific requirements, facilitating downstream processes. Utilized

NumPy for efficient numerical computations, enabling the manipulation and analysis of large datasets.

Data Processing in PySpark: Utilized PySpark for various data processing tasks. Conducted data ingestion, identification, and cleaning operations to prepare data for analysis or further processing. Constructed custom code modules tailored to specific project requirements. Generated JSON files and dispatched AWS SQS messages seamlessly within the PySpark environment.

Parquet File Processing in PySpark: Developed PySpark code to read Parquet files stored in AWS S3 buckets. Applied customized logic as per project specifications to process data or perform data analysis effectively.



EDUCATION

B.Tech, Shaheed Bhagat Singh State Technical Campus ,

July 2015 — May 2019

7.5



COURSES

Master's Program - Data Scientist , Simplilearn

December 2020 — September 2021

AngularJS & NodeJS Certification, Naresh IT

May 2018 — August 2018

Web Development Certification, Ducat

June 2017 — August 2017



Personal Project

Movie Lens Case Study (02/2021 - 02/2021) : Built predictive model by analyzing historic data, and using classification ML algorithms.

Customer Service Request Analysis (04/2021 - 05/2021): Performed basic EDA and hypothesis testing on historic dataset.

Mercedes-Benz Greener Manufacturing (06/2021 - 07/2021) : Performed basic EDA, dimensionality reduction and built predictive model using xgboost.

Sales Performance Analysis (08/2021 - 09/2021) : Used sample superstore dataset. Created bullet chart to measure sales with respect to target. Built a dashboard with this view by using Tableau.



EXTRA-CURRICULAR ACTIVITIES

National Service Scheme (NSS) Volunteer

October 2015 — March 2019

Community Engagement and Social Responsibility

- Actively participated in various community service activities, including organizing blood donation camps, cleanliness drives, and awareness programs on health and education.
- Collaborated with local NGOs and government bodies to address community issues, contributing to sustainable development and societal welfare.

Leadership and Teamwork

- Led a team of volunteers in planning and executing events, ensuring effective coordination and successful implementation of initiatives.
- Developed strong teamwork and communication skills by working closely with diverse groups, fostering a spirit of cooperation and mutual support.