Rakesh Gupta

Data Engineer

Highly motivated and results-oriented Big Data Engineer with 3+ years of experience in designing, developing, and maintaining scalable data solutions using Hadoop, Hive-SQL, Data Warehouse, PySpark, Azure Databricks, Teradata, Snowflake, Python, and Kafka. Possesses a strong understanding of data structures, SQL concepts, and complex SQL scripting with expertise in building and optimizing data pipelines for high-volume data processing. Proven ability to collaborate effectively with cross-functional teams to deliver innovative solutions and ensure high customer satisfaction.

✉ om.rakesh27@gmail.com 📱 9137280180 📍 Pune, India in linkedin.com/in/om-rakesh1999

EDUCATION

• Bachelor of Engineering in Electronics and Telecommunication
Thakur College of Engineering and Technology, Mumbai
08/2017 - 06/2021   *7.78 CGPA*

WORK EXPERIENCE

Associate Consultant
Capgemini (Morgan Stanley client)
06/2021 - Present   *Pune, Maharashtra*
Role and Responsibilities:

- Designed and implemented scalable data storage solutions using Hadoop and Hive-SQL databases.
- Developed and maintained big data processing pipelines using Hadoop, hive, Azure Databricks and Apache PySpark.
- Distributed data downstream using platforms such as Hadoop, Snowflake, and Teradata.
- Wrote and tested data processing scripts using Python.
- Utilized automation tools like Autosys, TWS (job scheduler), TCM, and batch monitoring tools for more than 50 jobs.
- Used UNIX shell scripts for validations and generic file-watcher scripts. Resolved bugs in the testing environment, considering QA and PROD batch stability.
- Collaborated with business stakeholders to understand data requirements and identify opportunities for data-driven decision making to ensure the security and privacy of sensitive data.
- I have migrated tables from the hive to Snowflake at table level, base objects containing data in TBs of data from Hadoop Data Lake to on-premise, Azure ADLS and Teradata.
- Monitored performance and optimized big data systems by staying up-to-date with emerging technologies and trends in big data and data engineering to ensure they meet performance and availability requirements.
- Documented and communicated technical designs, solutions, and best practices and mentored junior team members by providing technical guidance as needed.

ORGANIZATIONS

Capgemini (06/2021 - Present)
2+ years' experience in ETL development utilizing Big Data (Hive, Teradata), & multiple relational databases (SQL Server, MySQL, Databricks,Snowflake). Involved in all phases of the SDLC.

SKILLS

SQL(MyQSL, PostgreSQL, MS SQL Server)   Pyspark

Python   Hadoop   Hive   Azure Databricks

Unix and Shell scripting   Teradata   GIT   Snowflake

Data lake & Data warehousing concepts   ETL tools

Agile methodology   HDFS   Sqoop   Kafka

Problem Solving   Scala

ACHIEVEMENTS AND PROJECT

• Star of the Month Award from CG Gems for quality of work, dedication, and commitment (twice).

• Received compliments and appreciation from higher management for enhanced control validation.

Real-time social media analytics pipeline (01/2024 - Present)
- Develop a real-time pipeline that ingests social media data through Kafka. Use PySpark to process the data, extracting relevant information like sentiment, entities, and keywords. Store the processed data in a NoSQL database like MongoDB for further analysis or visualization.

CERTIFICATES

• Coursera Certified Data Analysis Using PySpark

• Coursera Certified Big Data with Spark and Hadoop

• Hackerrank Certified in SQL & Python

INTERESTS

Drawing   Travelling   Music

Providing mentorship to younger teammates