Swaroop Pagonda

Actively looking for a part time role to share professional knowledge on Data engineer stack. Having an experience of 7.5 years in the Software Industry. Worked in all levels from design, development, bug fixes, performance tuning, and production deployment activities. Proficient in data processing, data mining, and data pipeline management using Python and R.

Education

Jan 2022 - Dec 2022
University of Essex
Post Graduation in Applied Data Science

Sept 2010 - Jun 2014
RV College of Engineering
Bachelors of Engineering, 8.13 CGPA

Contact

Email
swaroop516@gmail.com

Contact Number
07776589326

Address
23 Lucas Garden, East Finchley
N2 8HJ

LinkedIn
https://www.linkedin.com/in/swaroop-pagonda

Skills

Data Science
Machine learning   Deep learning
Statistical analysis   Pattern recognition

Machine learning
Regression Classification Clustering

Programming
Python   SQL PLSQL R

Databases
Teradata Redshift   Postgress
Scylla   Oracle Cassandra

ETL
Spark Pyspark   Talend Airflow
Argo

Cloud
AWS: S3 EC2 RDS CloudWatch
Azure: Blob Containers DataBricks

Framework
Docker Kubernetes Helm Terraform

Miscellaneous
Unix Shell Scripting FSLDM Tableau
Agile Kanban JIRA Graphana

Achievements

Certification
Teradata 14 Certified Professional.

Awards
Employee of Quarter Q3 2018 at
Teradata India Pvt Ltd

Experiences

Dec'22 to Present
Data Engineer at Extreme reach
- Involved in migration of legacy pipelines.
- Working on design to migrate Pentaho data pipeline to Python based data pipelines.

Oct'19 to Jan'22
Sr Data Engineer at Zebra Technologies
- Designed data ingestion & data validation in AWS S3 data lake.
- Developed pipelines in Apache Airflow for around 11 data process flows for the supply chain customer.
- Refactored spark code for optimization improving the performance by 35%.
- Upgraded data bricks to utilize the new features and handled 60% of the compatibility errors occured.
- Designed & developed unit testing framework for CI/CD Pipeline
- Designed denormalized data model for wide column storage database Scylla DB to reduce around 85% of multiple API calls.
- Migrated 100% of code in R to spark.
- Designed code to suit disaster recovery.

Sr. Data Engineer at Zebra Technologies
- Designed the data lake in Azure for sell _in B2B data from scratch which will be the source for Data science team.
- Modularized code to handle billions of batch data.
- Designed and developed validation report to check the file count w.r.t to the last run reducing the manual effort by 90%.
- Involved in migration of SAS to Antuit Demand Forecasting Product Engine.
- Handled 60 million to 1 billion of data..

July'17 to Oct'19
Sr. Technical Associate at Teradata India Pvt. Ltd.
- Developed Text based search engine for onboarding in Banking.
- Have written python code to calculate string matching percentage for searching in backend DB.
- Used string matching algorithms like Levenstein distance, Jarowinkler and cosine similarity.
- Developed Python flask APIs to interact with UI.
- Used Redis cache for fast processing improving the performance by 58%.
- Based on the previous search patterns, used ranking based sort order.

Sr Technical Associate at Teradata India Pvt. Ltd.
- Involved in the design and development phase to extract 300 tables in the data lake for the retail customer.
- Automated the data extraction using TPT script. And TPT script generation is automated using shell script and saved 30 hours of effort
- Dumping the data into GCP cloud which was later used for data science projects in P&G.

July'14 to July'17
Associate Consultant at Capgemini India Pvt. Ltd.
- This project involved building a new semantic layer for capital, market, and risk reporting to the treasury of the bank.
- Involved in the architecture, design, and data modelling of semantic layer in Data warehouse.
- Daily interactions with business and stakeholders on requirements and issues.
- Wrote Transformation views and stored procedures in Teradata SQL.
- Generated reconciliation reports for Regulatory reporting.