Tarun Rajvanshi
Data Engineer

TarunRajvanshi2017ei@gmail.com
+91-6396505925

ABOUT

Seasoned Data Engineer with 8 years of experience in IT, currently leveraging expertise as a Financial Systems Analyst II at American Express. Adept at building and optimizing data pipelines, automating processes, and migrating data across platforms using Python and Pyspark. Recognized for a keen eye for detail, a strong adaptability to diverse technologies, and a continuous drive for learning and improvement. Demonstrated success in enhancing process performance and delivering actionable insights to stakeholders. Proven ability to teach and mentor others in data engineering and analytics, with a track record of impactful training sessions and innovative solutions

EXPERIENCE

American Express, *Financial Systems Analyst II*

Aug 2021 - PRESENT

Primary responsibilities have been to,

- Build the data pipelines for data warehousing with scalable architecture.
- Automate data reconciliation process for various data sources and generated aggregate reports.
- Migrate SAS code to Big Data Platform using Python and Pyspark.
- Generate Monthly KPIs and share it with Senior leadership and other stakeholders
- Improved the Performance of various processes by reducing the time consumption 70%.
- Collaborate with various teams for improvising data governance rules.

Xebia, *Consultant -- Big Data*

May 2021 - Aug 2021

Primary responsibilities have been to,

- Web scrape websites and Parse the magazine's PDF for Information collection.
- Transform the Unstructured data into Structure data using NLP and Regex.
- Ingest the data from the AWS S3 bucket to a No-SQL database.

TCS, *ML Engineer and Python Developer*

Mar 2016 - May 2021

Primary role have been as,

- Developed and maintained scalable data pipelines, ensuring efficient data processing and integration.
- Implemented machine learning models and algorithms to enhance data-driven decision-making.
- Designed and executed Python scripts for data extraction, transformation, and loading (ETL) processes.
- Built and optimized SQL queries for data manipulation and retrieval, supporting business intelligence initiatives.

DataMantar,Freelance Trainer

Mar 2019 - Present

Primary role have been

- Delivered training to over 4,500 professionals globally as a freelance Trainer for DataMantra.
- Specialized in SQL, Python,Excel , Pyspark ,Linux
- Achieved a 4.8 CSAT rating

PROJECTS

Integrated Delivery Model

- Role : Data Engineer and Python Developer
  - Build the pipeline for aggregated data containing very large dimensions and standardized the data for BI layer.
  - Setting up the process for manual adjustment without the need of re-generation of the whole data.
  - Automate the data feed from HDFS to Jethro and then to BI Layer.
- Technologies : Pyspark, Python, Pandas, Tableau

Setup Monthly KPI reporting Data Pipeline

- Role : Data Engineer and Python Developer
  - Build the data pipeline to ingest the monthly extracts and generate the KPIs for Financial reporting purpose.
  - Build the control checks for data ingestion failure and logging for auditing purposes.
  - Build the pipeline for Tableau Dashboard
- Technologies : Pyspark, Python, Pandas, Tableau

Automated Data Reconciliation

- Role : Data Engineer and Python Developer
  - Setup the process for performing data reconciliation among the data sources and generate the report.
- Technologies : Pyspark, Python, Pandas

Financial Report Generation and Migration

- Role : Data Engineer and Python Developer
  - Various reports were SAS based or were being prepared Manually.
  - All the reports were automated using Python/Pyspark jobs and results were published in the form of summarized reports.
- Technologies : Pyspark, Python, Pandas

EDUCATION

MIET, Meerut— B.Tech
- Aug 2012 - Jul 2016 (A..K.T.U.)
- Electronics and Instrumentation Engineering (69.6%)

AWARDS

Hackathon 2020: Won 1st prize for Solution of Onboard Guest Location Tracking.

TCS Awarded "On the spot Award": For fixing the bug in application.
Amex awarded "Innovation hub 2023" for Integrated delivery solution proposed.

TECH SKILLS

- Python (Data Engineering & Data Analytics)
- Pyspark
- Linux
- Sql & Hive
- Excel