Yashraj Nigam
Gurugram, HR | *+91 87709 00297* | *yashrajnigam97@gmail.com* | *yashrajnigam.tech* | *linkedin:yrnigam* | *github:yrnigam*

Professional Summary

Experienced Data Engineer with over 4+ years of expertise, currently employed at KPMG. Proficient in Python and skilled at analyzing structured and unstructured data. Proven track record of utilizing Data Science technologies to solve real-world problems. Committed to delivering high-quality solutions and catalyzing data-driven success for organizations.

Technical Skills

Cloud Platform : AWS, Databricks
Cloud Services : S3, EMR, Step functions, Glue, Lambda, SNS, Eventbridge, DataSync
Databases : Oracle, MS-SQL, Redshift, DynamoDB, MySQL, Redis, Kafka, Elasticsearch
ETL Tools : Talend (Open-Studio and Enterprise)
Languages : Python, SQL, PySpark
IDEs : Jupyter, VS Code, DBeaver , Google Colab
Version Control : GIT, JIRA
Operating Systems : Linux, Windows
Libraries explored : Boto3, BS4, OpenCV, Scikit-learn, Pandas, SpaCy, NLTK

Experience

Consultant - Data Engineer | *KPMG India (Gurugram, HR)* *November 2021 – Present*

- Designed and implemented a configurable ingestion job that migrated 300+ tables from Oracle to a Data Lake and Data Warehouse using Talend, Glue, and Amazon S3.
- Improved data accuracy and accessibility by designing and implementing a data pipeline using AWS Glue and PySpark.
- Reduced data processing time from 1 hour to 20 minutes, achieving significant performance and cost improvements in data pipelines and queries.
- Seamlessly integrated Talend with Python to enhance data ingestion efficiency for networked shared drives.
- Proficiently curated data in data lake's curated layer using AWS Glue, DynamoDB, Step functions, and Lambda functions, demonstrating strong AWS cloud expertise.
- Developed ETL job from scratch for data replication from Oracle DB to Oracle DB using Talend, implementing conditional checking and stored procedure calls.

Project Engineer - Data Science | *C-DAC (Hyderabad, TS)* *September 2020 – October 2021*

- Worked on a data analysis tool focused on both structured and unstructured data, encompassing over 20 modules.
- Improved code performance using in-process parallelism and multiprocessing.
- Enhanced language processing by implementing text transliteration, Named Entity Recognition (NER), and an offline language translation Python API using an open-source model.
- Utilized image and video analytics techniques, including face detection, cropping, and face matching for multimedia analysis.
- Dockerized Python code for automation and performance enhancements, streamlining deployment and scalability.

Computer Vision Engineer | *Tech Driven Basic (Indore, MP)* *February 2020 – August 2020*

- Trained deep learning model to automate the manual data entry work of invoices information.
- Created a system for multi-object detection in images using SSD and Faster R-CNN.
- Extracted invoice information through OCR, storing data in CSV files and databases.
- Utilized LabelImg for image annotation, Pytesseract for OCR, and Regular Expressions for data processing.
- Leveraged the power of Google Colab, along with OpenCV and TensorFlow, to accurately detect multiple objects in images.

Education

DAVV, Indore — **MBA in Business Analytics and Data Science** | GPA: 8.2/10.0 — *2018 – 2020*
RGPV, Bhopal — **BE in Computer Science Engineering** | GPA: 7.3/10.0 — *2014 – 2018*

Awards and Certifications

Generative AI Fundamentals accreditation | *Databricks* *March 2024*
Super Employee Award | *KPMG, India* *February 2023*
Rising Star Award | *KPMG, India* *April 2022*
Excellence Award | *CDAC Hyderabad* *February 2021*
Arctic Code Vault Contributor badge | *GitHub* *April 2020*