Kunj Hiteshkumar Pathak

Location: Halifax, NS, Canada

LinkedIn | GitHub | HackerRank   pathakkunj12@gmail.com | Mobile: +1 7828828522

Education

Dalhousie University Halifax, NS, Canada
Master of Applied Computer Science *Jan 2024 – May 2025*

Gujarat Technological University Gujarat, India
Bachelor of Engineering in Computer Science and Engineering *June 2018 – June 2022*
C.G.P.A : **8.46/10.0**

Experience

Data Engineer Intern Dec 2022 – March 2023
Diggibyte Technologies Private Limited *Bangalore, India*

- Learned concepts of data transformation with Spark/PySpark framework. Furthermore, worked with dataframes, exported the data using various write methods into cloud storage.
- Worked mainly with Azure blob storage and Azure data lake storage to store data and write the transformed data.
- Got hands-on experience with technologies like Azure Data Factory, and Azure Synapses to make pipelines, monitor them, and manage data.
- Used Databricks majorly for transformation and writing the data by creating the storage layers in the storage and achieved the knowledge of ETL process.
- Successfully Performed the unit-testing of the developed spark applications which consisted of the data-transformation code.
- Worked with Apache airflow to monitor the dag for the spark jobs created. Wrote dags with various operators such as Python and Bash to schedule jobs. Also was able to use Cron syntax for scheduling the tasks. Further used xcoms to establish communication between two functions.
- Also got a chance to work with the leadership group of the firm for the project to create a P.O.C for the D.B.T tool used in E.T.L process to supplant the Databricks as it was required by the client. Further created staging layers and applied basic transformation from SQL and worked to containerize the whole project with Docker.

React Development Internship Jan 2022 – Apr 2022
Imbuesoft L.L.P *Rajkot, India (Remote)*

- Completed the development of the Landing page and Authentication pages in ReactJS for the Vistaderm Project.
- Worked with React route to link authentication pages and also, used sass as a styling element for the entire project.
- Gained knowledge of sass scripting for the CSS part and had a basic understanding of the working of react framework.

Technical Skills

Languages	:	Python, Java, SQL, C, HTML, CSS, Javascript
Frameworks	:	Spark, PySpark, React
Databases	:	MySQL, OracleDB
Cloud Terms	:	Azure, AWS Basics
Platforms	:	Azure Data Factory, Apache AirFlow, Terraform, Docker, Jenkins, Azure DevOps, Linux, SDLC, DBT(Basic), Raspberry PI
Libraries	:	Pandas, NumPy, Unittest, React-Route, Plotly, Matplotlib
Version Control	:	Git, Github

Projects

Caption Generation WebApp *Transformers, NLP, PyTorch, Encoder-Decoder-Model, CV* Source Code

- Web-App built with Django which houses the logic of deeplearning for generating the captions from the user uploaded image. Using a pre-trained from transformers provided by the huggingface community more specifically a VisionEncoderDecoderModel that combines a Vision Transformer (ViT) and a gpt2 languagemodel
- The image's characteristics are extracted using the ViTImageProcessor, and the captions are tokenized using the AutoTokenizer
- The mkOutput function takes the path of the uploaded image as an input, open the image, changes it to RGB mode, uses feature-extractor to extract the pixel values or more specifically extract the feature- tensor , and then uses the pre-trained model to output the predicted caption
- The pre-trained vision encoder-decoder model is loaded and used by the VisionEncoderDecoderModel class. Device management and tensor operations are done with the pytorch module. Image processing is done using the pillow (PIL) module.

Spark Application *Python, PySpark, unittest, Spark* Source Code

- Created three spark applications for data transformation as an assignment task during the data engineering internship. The First uses a CSV file and the other file consists of nested log file data to be transformed.
- Lastly, the requirement demanded creating a dataframe and applying the transformation. Hence successfully performed the mentioned actions.
- Further performed the unit testing of all three applications and uploaded to github.

ETL process with Databricks *Python, PySpark, Azure Databricks, Azure Data Factory, Microsoft Azure* Source Code

- Created staging Layers in Azure storage account Taken one nested json file copied using Azure Data Factory Copy Tool into Bronze Layer.
- Flatten it by applying transformations using pyspark and data bricks.Stored that transformed data into Silver layer of staging
- Further performed mergeColumn function joined two different tables as per business requirements and Wrote that data into gold layer.

COVID-19-ANALYSIS-VISUALIZATION-AND-PREDICTION *Python, Plotly, Regression, Pandas* Source Code

- Worked in Data Science and Machine Learning domain, analyzed and visualized the existing data for covid-19.
- Predicted the (future) for the Positive cases, death toll, and recovered cases.

Certifications

- Databricks Lakehouse Fundamentals, Academy Accreditation
- Python 3, from scratch to pro