Vivek Upadhyay
(LWD– 19th Dec 2022)

Email: - up.vivek.up@gmail.com
Contact: +91-9748274803

Profile Summary:

Experienced Big Data/Cloud Data Engineer and Analyst with 4+ years of overall experience in the Data Analytics field.

Looking for challenging opportunities. Would like to leverage Big Data tools, Statistics and Machine Learning to come up with Analytics solutions that help businesses take data driven decisions.

Experience:

Organization	Designation	Period
Optum (UHG)	Associate Data Analyst	20th Sep 2021- Till Date
Capgemini	Associate Consultant	13th Dec 2017-15th Sep 2021

Technical Skills:

ETL Tools	SSIS
Database	SQL Server, Oracle, My SQL
Scripting	Python, UNIX Shell Scripting
Big Data Frameworks	HDFS, HIVE, Spark (PySpark, SparkSQL)
Cloud Services	Azure Data Factory, Azure Databricks, Azure DataLake, Azure Key Vault
Version Control	GITHUB, TFS

Work Summary:

I have been involved with the following activities till date-

- Dimensional and Fact Modelling for Data-Marts.
- Developing and designing ETL/ELT pipelines for Analytics workloads using SQL, PySpark with Azure Databricks and Python.
- Scraping data from frontend systems using Python and Selenium.
- Worked with various type of source and target systems including AWS S3, Google Big Query, Azure Data Lake, Azure Blob Storage and on-premises Hadoop cluster.
- Gathering requirements from stakeholders and ensuring timely deliverables.
- Development of KPIs relevant to business reporting and building data pipelines that would contribute to those KPIs.

Projects:

1. **Working to automate reporting and drive business growth for United Health Care Contact Center Operations Unit.**

a. Gathered business requirements from BU and worked towards building an Analytics Solution for the client.

b. Built a data pipeline with AWS S3, Azure Blob, remote NAS drives and various third party applications such as Nexidia and Qualtrics as source systems and Azure Data Lake as target system. Used Azure Databricks for data transformation and Azure Data Factory for orchestration.

c. Used Python, SQL and PySpark with Azure Databricks for data cleaning and transformation.

d. **The proposed solution helped in gaining more customers for UHC and saving 40+ hrs of manual effort monthly to analyse data and build a manual report.**

2. Working on a Supply Chain Analytics project for efficient procurement of healthcare and medical equipments by leading hospitals in the US.

- Gathered business requirements from chain of hospitals (CDOs) and worked with a large amount of unstructured data from sources like AWS S3, Google Big Query, remote NAS drives.

- Used Hive and PySpark for data cleaning, transformation and loading data into an on premises Hadoop Cluster (UHG Data Lake).

- **The proposed solution has saved 1.1 Million USD in two quarters of the financial year enabling hospitals in acquiring of medical equipment at competitive prices from various distributors.**

3. Working in an ETL development project for a major insurance company named Willis Towers Watson.

a. Gathered business requirements from client and worked towards building a Business Intelligence Solution for the client.

b. Maintained and developed Big Data pipelines using SQL Server, Python, Hive and SSIS.

Educational History:

Bachelor of Engineering (Electronics & Communication) - July 2013 – July 2017
University: West Bengal University of Technology, Kolkata
College: MCKV Institute of Engineering, Liluah
Aggregate CGPA: 7.91

XII, CBSE - Apr 2012 - April 2013
School: Kendriya Vidyalaya Fort-William
Percentage : 65