Souvik Patra

Data Engineer

Souvik Patra

87/B Cossipore Road, Block- F, Room No-11, Kolkata - 700002

PH:8443003105
souvikp071@gmail.com
www.linkedin.com/in/souvik-patra-104454166

- **Proficient in Python, Pyspark, and MySQL, specializing in web scraping, data extraction via RestAPI, and overseeing end-to-end data engineering projects.**

- **Developed interpersonal skills to comprehend business requirements and spearhead research initiatives, employing a problem-solving approach by the help of various BI tools.**

- **Proven track record as a prolific and intuitive collaborator, effectively partnering with stakeholders, web developers, and database architects**

—

Skills

Python, Pandas, Pyspark, LINUX, MYSQL, AWS(S3, IAM, EC2, REDSHIFT, RDS, EMR), MYSQL, Mulesoft, Apache Beam, Java, NumPY, GCP, Databricks, Tableau

—

Experience

Royal Content research / Technical_writer
DECEMBER 2019 - March

Technical writer for Python, SQL based projects

Focus Edumatics/ E-learning Trainer (September , 2021 - January,2022)

I deliver the computer education to US students

Agilisium Consulting Pvt. Ltd / Data Engineer (March 2022 - Present)

- I specialize in powering data-driven success stories for industry leaders like UMG in the dynamic retail sector and Amgen in the cung-edge life sciences and pharmaceutical domain.Mastering the art of data alchemy, I leverage ETL tools, harness the power of web scraping, and wield deep expertise in both GCP and AWS clouds to transform raw data into strategic insights for diverse industries

Education

B.tech in Electronics and communication engineering (March 2015 - March 2019)
Bengal Institute Of Technology, Kolkata

Higher Secondary education in Science (March 2013 - February 2015)
Dum Dum Kumar Ashutosh Institution Main Boys' secondary education, Kolkata

Secondary education in Regular course for Board exam (March 2007 - March 2013)
Baranagore Ramakrishna Mission Ashrama High School, Kolkata

Awards

Solution Award: (January 2024)
Agilisium Consulting

Given a solution to the company Neocrene for end to end data processing solution by maintaining the security issue.

Certification

Entry level Python programmer
PCEP certification Institute
Databricks certified Data engineering associates(April 2023)

Core Java certification program
BAIT institute

OOPS programming concept
Great Lake Learning Platform

SQL Intermediate
Hackathon Platform

Databricks Generative AI fundamental
Databricks training platform

AWS Certified Cloud Practitioner(June 2024)
AWS

cp.certmetrics.com/amazon/en/public/veri...

Project Responsibilities

UMG CLIENT   (March 2022 - December 2022)

- Implementing a data validation process within the e-commerce department of the UMG project.
- Designing a DAG to initiate priority pipelines, facilitating the seamless transfer of necessary data from the GCP platform to BigQuery.
- Developing a backfill DAG to identify and rectify missing data by
- leveraging the capabilities of the pipeline. Utilizing API calls to fetch data from all Shopify shops and cross-referencing records for a specific date.
- Implementing a backfill process tailored for Shopify-based priority
- shops to address any data gaps. Employing GCP CLI to efficiently load data from GCP to BigQuery.
- Creating DAGs to concurrently trigger three pipelines for backfilling missing data on a specific date.

Agilisium Consulting(Internal Project)   (January 2023 - March 2023)

- Employing Spark code to load data from the source into Redshift.
- Implementing obfuscation techniques on thePython code for enhanced security.
- Leveraging the FB-prophet module to generate predictions for the source dataset spanning at least 3-4 months.
- Creating a Bash script for generating the schema and promptly inserting data into the designated database connection.
- Utilizing an EC2 cluster for executing AirflowDAGs through the server.
- Establishing a Snowflake connection using a Python script to facilitate the loading of data into an existing table.

AMGEN   March 2023

- DQ load by Mulesoft Pipeline.

ATLAS US Project(AMGEN):   March 2023

- Analyzing stored procedure for optimizing the business query.
- Migrate data from Elmac S3 to US S3 bucket.
- Using Boto configuration to upload data from local to S3 bucket.

AMGEN(GDNA)   March 2023

- Creating a back up history in Gitlab by using databricks to track the change of DDL for business related views.
- Analyzing the 'null' value and 'zero' value for columns of t2_object_configuration tables as per the affiliate requirements.
- Creating a pipeline to migrate data from EDW server to EDF server.
- Validating data across ELMAC and JAPAN countries by using apache spark application.