Souvik Chattopadhyay

Current Employer   : PEPSICO
Experience   : 9 years
Email   : souvik.101055@gmail.com
Contact Number   : 7028929010

Technical Skills:

- SQL/NoSQL - Oracle, SQL Server , Mongo Db
- Python , PySpark
- AWS - S3, Athena, EC2, EMR, DynamoDb ,Step functions, Lambda functions, Cloudwatch
- AZURE – ADLS, ADF, Databricks
- Agile,DevOps

Experience :

- Development and testing of data pipelines in Cloud based Data Lake , using distributed computing framework (Spark) and programming tools (Python) .
- Working with structured/ unstructured datasets with different file formats (Parquet/csv/json)
- Working with SQL/NoSQL databases (MongoDb/DynamoDb/PostgreSQL) .
- Creating transformations using Spark api and spark SQL .
- Extensive knowledge in SQL .
- Working knowledge on AWS and Azure services for data pipeline .
- Experience working with geographically distributed agile teams.
- Basic knowledge in working with real time pipelines using AWS Kinesis.
- Knowledge with visualization tools like Tableau.
- Experience in client interaction and mentoring team members.

Professional Experience:
PepsiCo (Nov 2022 – Present)

- Develop Databricks notebooks for loading data to different layers of Data lake from ADLS.
- Create pipelines for data ingestion, data quality using Pyspark, ADF.
- Create DQ framework using DBT
- Manage and Review code through DevOps pipelines.

Marlabs (Sep 2020 – Present)

- Part of Agile Team and responsible for each sprint's deliverables.
- Develop components for data lake in AWS, from on-prem DWH and external files to Data Lake in S3 , through different layers , and push data to Snowflake.
- Development of batch-processing data pipeline and data transformations in PySpark SparkSQL , Python.
- Scheduling of spark jobs in EMR , debugging and optimization of spark jobs.

- Use different AWS components for orchestration and automation of data validation , data quality and monitoring activites.
- Deployment and automation using Terraform.
- Providing KT and mentor new team members.

MSH Talent Solutions (Oct 2019 – Sep 2020)

- Part of Agile Team and responsible for sprint's deliverables.
- Developing code for ETL process , to load data from MongoDb to Data Mart using PySpark, PL/SQL.
- Writing ad-hoc queries for reporting and dashboards, and creating visualizations using Tableau.
- Developing Python scripts using Pandas for data validation and data analysis from MongoDb to Oracle.

Cognizant: (Oct 2018 – Sep 2019)

- Part of ETL projects, working with development and automation testing of data load from source to target.
- Worked with creating reports based on Tableau/Oracle.

TCS (Sep 2014 – Oct 2018)

- Worked in ETL and BI projects.
- Data preparation and validation of reports from Oracle.
- Loading data from flat file and relational sources to staging database using Informatica.

Education :

	Degree/ Examination	Board/ University	From	To	% of marks	Part Time/ Full Time
Graduation	B.Tech	WBUT	2010	2014	8.70 (DGPA)	Full Time
XII	HS	WBCHSE	2008	2010	91	Full Time
X	Secondary	ICSE	2007	2008	81	Full Time

Total Experience : 7.10 years
Relevant Experience: 3 years

Personal Information:
Name   : Souvik Chattopadhyay
DOB   : 23 January 1992
Languages   : English, Hindi, Bengali
LinkedIn profile   : https://www.linkedin.com/in/souvik-chattopadhyay-4a0a0313b/