





SUBHADEEP SEN

DATA ENTHUSIAST | SENIOR CONSULTANT

EVERBODY LOVES A NICE STORY. I TELL MINE WITH DATA!!

CONTACT

-  (+91)-908-353-1573
-  subhadeep.sen@outlook.in
-  Kolkata, West Bengal, India
-  <https://www.linkedin.com/in/subhadeepsen94/>

SKILLS

- PySpark
- Python
- Databricks
- Big Query
- Cloud Function
- Dataflow, Pub-Sub
- Azure DevOps
- Amazon Redshift
- AWS Athena, AWS Glue
- AWS Lambda, EC2
- Apache Airflow
- Docker
- SQL

CERTIFICATIONS

- Databricks Lake House Fundamentals
- Databricks AI Security Fundamentals
- EY Cloud

DOMAIN

- Insurance
- TeleCommunication



ABOUT ME

I am a Data enthusiast looking forward to work in a stimulating environment, where I can implement and enhance my skills and knowledge to meet both organizational and individual goals. Team player with strong desires to face new challenges and expand my skills set.



WORK EXPERIENCE

Ernst & Young Global Delivery Service India

SENIOR CLOUD DATA ENGINEER

FEB 2024 - PRESENT

Worked as a Senior Cloud Data Engineer, managing the Transformation Team in developing Customer Data Management Solution in GCP. Design and developed Data Lake following Medallion Architecture in GCP. Involved in designing an optimized, efficient data migration framework and developing end-to-end pipelines using Dataflow, Cloud Function, Pub/Sub, Cloud Scheduler, Big Query, GCS. Designed and developed Dataflow Jobs to connect to external databases and handle CSV, Nested Complex Json data.

Experience in API integration design, creating robust and efficient data pipelines that interacts with various APIs to ingest data. Designed end-to-end solution to migrate data from Oracle, AWS S3 and GCS buckets and load them to Big Query Tables.

Responsible for developing cost optimized solution and handling PII data.

Working Experience in CICD Pipelines Using Azure DevOps. Created Technical documentation in Confluence and developer stories in JIRA

Ernst & Young Global Delivery Service India

CLOUD DATA ENGINEER

OCT 2022 - JAN 2024

Worked as a Cloud Data Engineer, leading the Transformation Team in migrating the Claims from the legacy claim system into the strategic EIS system. Worked in setting up the AWS Infrastructure for end-to-end data pipeline using AWS s3, Glue, Athena and AWS Managed Services (Apache Airflow).

Developed the customized transformation framework in Python, deployed as an AWS Glue Job. Developed Airflow DAGs for overall orchestration of the pipeline.

Created Technical documentation in Confluence and developer stories in JIRA

Tata Consultancy Services Limited

DATA ENGINEER - PYSPARK

JAN 2021 - OCT 2022

Worked as Data Engineer for migrating existing Microsoft SQL Server database and its dependencies to a centralized repository in AWS. Debugged the existing SSIS packages and re-engineered the code in Py-Spark. Developed Spark applications using Py-Spark, Spark-SQL for data extraction, for analyzing and transforming data to uncover insights into product analytics and customer usage patterns. Developed Glue Crawler and Jobs for loading the data to corresponding tables in Athena for downstream systems.

Experience in creating docker file, building docker images. Responsible for estimating the cluster size, monitoring and troubleshooting of the spark Databricks cluster.

EDUCATION

2012-2016
Bachelor of Technology
Heritage Institute Of Technology, Kolkata

2005-2012
Secondary and Higher Secondary
Ramakrishna Mission Vidyapith, Purulia

DOMAIN

- Insurance
- TeleCommunication



Tata Consultancy Services Limited

DATA ENGINEER - PYTHON

DEC 2019 - DEC 2020

Worked as Python and AWS expert for developing customized python framework to automate back-end data preparation for BI Dashboard. Developed User-Defined functions(UDF) and scripts to interact with AWS s3, uploading data to s3 bucket, unloading data from AWS s3 to Amazon Redshift tables. Developed validation and syncing scripts to auto-sync updated data to BI dashboard and refresh.

Tata Consultancy Services Limited

BACK-END DEVELOPER & ANALYST

DEC 2018 - NOV 2019

Worked as Amazon Redshift expert to design advanced Redshift Tables, developed optimized queries to manage large volume of data and generate back-end data for Analytical Reports of BI dashboard hosted in ec2 instance.

Tata Consultancy Services Limited

SOFTWARE DEVELOPER

APR 2017 - NOV 2018

Member of the development Team for developing a Data Ware House for Client using SQL



PROFESSIONAL SYNOPSIS

Overall 8+ years of IT experience with more than 6 years of experience as Data Engineer. Experienced in designing data engineering solutions, managing and leading teams Proficient in Analyzing User requirement and functional specifications. Well versed in Exploratory Data Analysis, developing Data pipeline and providing data solutions. Proficient in working with multiple GCP and AWS Cloud technologies. Team Player with excellent communication skills

PYSPARK

- Understanding of Spark framework, big data architectures and developing Data Pipeline.
- Strong Experience in data engineering and building ETL pipelines on batch and streaming data using Databricks, Spark-SQL.
- Working Experience with large volume of datasets
- Good working exposure on Cloud technologies of GCP - GCS, Cloud Functions, Big Query, Dataflow, AWS - S3, EC2, Lambda, Glue, Athena.

PYTHON

- Experienced with core Python programming, developing customized scripts for automation, developing UDF(user-defined functions), ORMs(Object-Relational Mappers)
- Used Python to develop customized framework to interact with AWS s3, Amazon Redshift for preparation of data on cloud.

BIG QUERY & REDSHIFT

- Understanding of underlying Architecture and hands on experience in developing advanced Tables and optimizing complex Queries.
- Experience in dealing with large volume of data.