GOURI BHOSALE

Data Engineer/Analyst

Pune, Maharashtra 411057

Phone: - +91 8999844316 /+91 8806817947

Email: - gouribhosale1996@gmail.com

LinkedIn:- https://www.linkedin.com/in/gouri-bhosale-8458a0243

Website: http://gouribhosale.github.io/gportfolio/

A Skilled Data Analyst/Engineer consultant with expertise in Amazon Web Servicess, SQL, Pyspark, CICD Tools. Experienced in analyzing large datasets and deriving meaningful insights to drive business decisions. Ability to handle complex data projects from inception to implementation. Having overall 2 years of experience as a data analyst and as a devops consultant

SKILLS

- AWS Services: AWS S3, Redshift, RDS, IAM, EC2, VPC.
- Python, Linux, Databricks, HTML.CSS3.
- Jupyter notebook
- Redwood RMJ, Brickftp
- Terraform
- Jenkins, CICD
- Github
- Technical Design Document, SOP Documents

WORK HISTORY/PROJECTS

2022 current
Data Engineer
Capgemini, Pune

Project1:

Client: Sage Therapeutics
Title: Commercial and medical datahub/Market Access datahub
Purpose of the project
The objective was to define workflow of Commercial Data Hub & Reporting. Here ingestion of data is being carried through Databricks in automated scheduled way Via

Redwood RMJ (Automation tool) into the Snowflake data warehouse through below layers:

Layers:
Source to Core: -
Data are pushed from the below type of sources as is to the snowflake core layer.
Flat files (.xlsx,.csv,.txt, .zip)
Snowflakes tables and views
Core to Conformed: -
Business rules are applied on core data and pushed to conformed layer mostly adhering SCD type 2. Audit columns and table are populated accordingly.

Conformed to Consumption/ARD: -
Business/ aggregation rules are applied on conformed data to populate analytics ready layer(ARD) for reporting and further analytics.

Technology used:
- Pyspark, Python
- Aws S3, Brickftp, Databricks, Snowflake, SQL. Redwood RMJ (Automation Tool for Orchiestration), GitHub

Roles and responsibilities
- Analyzation of large data set to identify trends, patterns, and correlation for business insights Gathered requirements from stakeholders to design data models and reports.
- Conducted data analysis to prepare forecast and identify trends.
- Created /reviewed /maintained technical design document and support documents.
- Used AWS, SQL, and pyspark to analyze bigdata.
- Collaborated with teams to develop code and ensure smooth operations.
- Provided support for a Devops project, using aws, pyspark, databricks, snowflake sql, python, Redwood RMJ
- Built and managed a central data repository for Market Access and Medical Datahub Analysis
- Check data on frequency based whether data got loaded.
- Implementation of python scripts based on requirement.
- Validation data / Testing while implementation

Project 2:
Title: Development of web application for Career framework.

Purpose of the project
The objective of this project was to collect all the Employee Information, like their current details, experience, qualification details also their aspirations.

Technology used:
Backend: Python,SQL lite, postgre sql
Front-End: HTML5, CSS3, python, JS

Roles and Responsibilities:

- Design the Front-end pages.
- Connection between frontend and backend with SQLite
- Gathering requirement, Analyzation, Preparing projects documents
- Design Data Models
- Design ERD, DFD
- Creating forms like, for login registration, user profile form, ▯ Testing, validation, verification.

Project 3:

Purpose of the Project:

Using the Redshift data warehouse as the source and s3 as the Target. Uploading the data (CSV files) into S3 bucket and loading the data into redshift data warehouse using s3 as a data source. By using the data available in the redshift warehouse, we're establishing a connection between Pyspark and redshift then storing the data in dataframe then performing the Pyspark dataframe queries based on the required reports.

Technology used:

- AWS S3 for source
- AWS redshift Database
- Jupyter Notebook
- PySpark

Roles and Responsibilities:

- Made used of AWS S3 to store files.
- Loading the data into redshift via s3 and making connection betweens3 and jupyter notebook to perform operations.

EDUCATION

2018-2021	**Master of Computer Application (MCA)** Chhatrapati Shahu Institute of Business Education and Research, Kolhapur. (CSIBER)
2015 – 2018	**Bachelor of Computer Application (BCA)** V.P. Institute of Management Studies & Research (VPIMSR)
2014	HSC Completed in Miraj Mahavidyalaya Miraj (MMM), Miraj
2012	SSC Completed in Cambridge school Miraj.

ADDRESS

Address: Menlo Homes Hinjewadi phase 2 Plot no 912, B wing
City: Pune

DESCLAIMER

I, Gouri Bhosale, hereby declare that the information contained herein is true and correct to the best of my knowledge and belief.