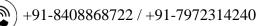
Vidya Magar

Data Engineer



magarvidya1@gmail.com

Chikhli, Dist. Buldana, Maharashtra

Profile

Looking for a long term career in big data Developer, to obtain a responsible position in the Software Industry, where my capabilities can be fully utilized.

Career Summary

- Total 3.2 years of experience in different domains in IT industry out of which having 2+ years of relevant experience in big data which includes Hadoop Ecosystem, Spark API's, Cloud Services (AWS).
- In-depth knowledge of Hadoop architecture and its components like YARN, HDFS, Name Node, Data Node, Job Tracker, Application Master, Resource Manager, Task Tracker and Map Reduce.
- > Hands on experience in various big data application phases like data processing, analytics.
- > Have experience in building optimized big data pipeline using S3, PySpark on AWS cloud.
- > In-depth understanding of Spark Architecture including SparkCore, SparkSql, and Data frames.
- > Have experience writing data extraction logic in Pyspark.
- Experience in working on file compression techniques as well as various file formats like ORC, Parquet, AVRO.
- Hands on experience on the transformation glue job to move data from the S3 cleansed to the redshift using lambda and cloudwatch.
- > Good understanding of AWS VPC, Athena, EMR, EC2, IAM.
- Very good understanding of Partitions, bucketing concepts in Hive and both Managed and External tables in Hive to optimize performance.
- > Understanding of transferring data from RDBMS to HDFS and Hive tables using Sqoop.
- > Hands on development on several tools like Putty, WinSCP etc.
- > Well aware of various essential commands in linux and shell scripting.
- > Experience on working on an Agile/Scrum Methodology.
- > Strong problem solving & technical skills coupled with clear decision-making

Technical Skill Set

- Big Data Ecosystems: Pyspark, Apache Hadoop, Apache Hive, Apache Spark, Sqoop, HDFS, YARN.
- Cloud Technologies: AWS (S3, IAM, Lambda, Step functions, ECR, Sagemaker, Cloudwatch, Glue, Redshift)
- Language: Python, SQL, C, C++
- > Web Development: HTML, XML, CSS
- **Database**: MySQL, Oracle
- > Operating Systems: Windows 10/11, Linux Ubuntu/CentOS
- **IDE**: PyCharm, Jupyter notebook

Educational Qualification

- > Pass out in 2018, Master of Science in **Computer Science** from Amravati University.
- > Pass out in 2015, Bachler of Science in **Computer Science** from Amravati University.

Work Experience

Project #1

Organization: SYS ADMIN SOLUTIONS (INDIA) PVT LTD, BENGALORE

Project Name: Bank Email Data Analysis

Duration: Sept. 2021 – Current

Role: Data Engineer

Description:

Build automated data pipeline on cloud for data scientist to save their time, in this project, we worked with a dataset with feedback collected for a business. our task was to analysis which feedback is positive and which one is negative.

Roles and Responsibilities:

- ▶ Working in Data Analysis team for client in banking domain.
- > Responsible for creating Lambda function, Glue jobs.
- > Responsible for automating Machine Learning Models using Amazon Sagemaker.
- Responsible for developing and deploying various applications, optimized big data pipeline using different AWS services.
- > To create ECR image and to get output using AWS Sagemaker and Step functions.
- Responsible for applying data cleaning techniques to correct and remove in-accurate and corrupted data, to improve data quality which significantly improves productivity.
- > Applying various **PySpark transformation** to cleaned and validate data in glue jobs.
- ▶ Writing desired output in **S3** bucket.
- Building Docker image in Sagemaker for required and requested libraries for data scientist based upon code.

Environment: AWS S3, Lambda, Step functions, Sagemaker, ECR, Docker image, Cloudwatch.

Project #2

Organization: SYS ADMIN SOLUTIONS (INDIA) PVT LTD, BENGALORE

Project Name: Data Warehouse Design for an E-commerce Platform

Duration: Sept. 2020 to Aug. 2021

Role: Data Engineer

Description:

The client uses on premises OMS (order management system). It leads to facing the slow performance of large analytical data, so to avoid this we have to prepare a data lake and data warehouse for unifying the data and to get meaningful information from it.

Roles and Responsibilities:

- > Attaining a daily scrum call with team members and discussing the task on a particular sprint.
- ▶ In every sprint we have assigned tasks, that we are completing as per timeline.
- > Weekly call with client for their requirements and further enhancement.
- ➢ Job monitoring.
- Work on the transformation glue job to move data from the s3 cleansed to the redshift dimensional and fact tables.
- > Optimize code using pyspark for better performance.
- > Creating a DDL script for creating tables in redshift.
- Work with the teammates to understand and support them in importing data.

Environment: Pyspark, SQL, Python, AWS Glue, AWS Lambda, Amazon Redshift, AWS Cloudwatch.

Personal Information

Name: Vidya Nandkishor Magar

Languages Known: English, Hindi, Marathi

Address: Mahavir Nagar, A/P Chikhli, Tq. Chikhli, Dist. Buldana, Maharashtra

Declaration:

I hereby declare that all the information mentioned above is true to the best of my knowledge.

Place:

Vidya Nandkishor Magar