DEEPAK ASHOK LAGAS

✉ deepaklagas1007@gmail.com   📞 +91 855-290-5325   📍 Aurangabad-431001, MH, INDIA

☞ I am looking forward to work in an organization which provides growth & knowledge. I have an urge and inclination for learning new technologies and I want to prove my abilities by accepting challenges. IT professional having **3+ years** of experience have knowledge of **PySpark, Python, AWS, Hadoop and SQL**.

EXPERIENCE SUMMARY:

☞ Working closely with business and engineering teams to encourage statistical best practices with respect to experimental design, data capture and data analysis.

☞ **3+ years** of experience in **data pipeline, data extraction, data processing, data preparation/cleaning, data ingestion, data validation, python programming, data visualization etc**.

☞ Closely collaborate with data engineers, software development and product teams to ensure accuracy and reliability of tools and data.

☞ Experience on using tools such as **PyCharm, WinSCP, Putty**.

☞ Knowledge of **Python and Shell Scripting.**

☞ Good knowledge of **pandas, NumPy and matplotlib** to work with data.

☞ Create, design, develop, test data-pipeline to ingest data from multiple source systems.

☞ Good experience on **Spark Architecture including Spark-Core, Spark-Sql**.

☞ Have a hands-on experience on Amazon web services mainly **S3 ,EC2 , Aws EMR , AWS Redshift , IAM Role , RDS**.

☞ Strong analytical and problem-solving skills.

☞ Experienced Scrum Process and **Agile Methodologies**.

☞ Good collaboration with proficiency at grasping new technical concepts quickly and utilize the same in a productive manner.

TECHNICAL SKILLS:

☞ Programming Languages   :   SQL, Python

☞ Libraries   :   Pandas, NumPy

☞ Data Analysis Tool   :   Jupyter Notebook

☞ Cloud Technologies   :   AWS (S3, EMR, RDS, IAM ,EC2, Redshift, Glue, Lambda)

☞ Database Skills   :   Oracle, SQL Server

☞ Big data   :   PySpark, Spark SQL, Hadoop

☞ Versioning Tool   :   GIT

☞ Development Tools   :   PyCharm, WinSCP, Putty

☞ Operating System   :   Window, Linux

☞ Ticketing Tool   :   JIRA

EDUCATIONAL QUALIFICATION:

☞ Bachelors of Engineering in Mechanical Engineering from Savitribai Phule Pune University (Maharashtra) with **First Class** in 2019.

WORK EXPERIENCE:

☞ Total **3+ Years** of experience in IT Firm with **BITWISE GLOBAL, PUNE as Data Engineer** from July-2019 to Till Date.

PROJECT HANDLE:

Project #1(Current)
Project Title: Financial Intelligence Unit
Duration: Nov-2020 to Till Date
Domain: Banking
Technologies: Spark-SQL, HBASE, Phoenix, HIVE, PySpark, AWS(EMR, S3, RDS),GitHub

Description:

☞ The purpose of the project is to do the ingestion of the credit card data that is present in RDS to EMR cluster. Making data pipeline with the help of PySpark to get some business insights for various use cases and to generate a report out of it and storing it into AWS S3.

Roles and Responsibilities:

☞ Understand the upstream source nature and work with business cases.
☞ Engage with BA to understand the requirements clearly.
☞ Import and export batch data and delta into Phoenix (HBase)and Hive using PySpark.
☞ Responsible for applying data cleaning techniques to correct and remove in-accurate and corrupted data, to improve data quality which significantly improves productivity.
☞ Filtered data and performed multiple operations in-order to get meaningful insides that help our client in decision making and analysis.
☞ Load, transform large sets of structure data.
☞ Involve in loading data from AWS RDS (PostgreSQL) to HBASE And Hive Environment i.e., HDFS using Apache Spark.
☞ Creating UDF's.
☞ Use Python function in Pyspark.
☞ Involved in processing data and do operations on it e.g., doing validation of data, deduplication and other.
☞ Operations as per the client requirement.
☞ Data Reconciliation to ensure good Data Quality.
☞ Resolve JIRA tickets raised by Client.
☞ Involved in client interaction to sort out requirement issue.
☞ Participate in daily scrum calls.
☞ Create the reporting and share with top management.

Project #2
Project Title: Data Migration into Data Lake
Duration: Dec-2019 to Sep-2020(**10 Months)**
Domain: Retail Store
Technologies: PySpark, AWS(S3, Redshift)

Description:

☞ The purpose of project is to create data pipeline for migrating the unstructured data or uncleaned data which is comes in client-side AWS S3 bucket. And then our object is to clean, filter the unstructured format data into structured format by using PySpark and stored into cleaned AWS S3 bucket. After that check the count of records before and after processing the data by using PySpark and stored to Amazon Redshift.

Roles and Responsibilities:

☞ Understand data migration requirements and needs.

☞ Engaged with BA to understand the requirements clearly.

☞ Efficiently loading data into a spark dataframe.

☞ After processing data using Spark-SQL we export that data back into client's data lake(Amazon Redshift).

☞ Responsible for daily data validation.

☞ Responsible to write PySpark code for extraction.

☞ Cleansing and normalizing the data and schema.

☞ Performed PySpark transformations.

☞ Responsible to write efficient code as per required and requested.

☞ Using Git, Jira and coordinating with team.

DECLARATION:

I hereby acknowledge that the information furnished above is correct to the best of my knowledge.

Deepak Ashok Lagas