RAHUL PATIL

Email ID: rahul.basix@gmail.com

Mobile No: 8530387596
LinkedIn Profile:

Data Engineer with 3.7 years of experience in building Data Pipelines, Data-Wearhouse Solutions, ETL, Big Data and managing operations of multi-Terabyte systems for Edtech clients.

PROFESSIONAL SUMMARY

- I have experience working in Big Data and Big Data related technologies.
- Proven ability to design, develop and implement big data solutions that meet business requirements using Agile Methodology.
- Hands-on work experience on Big Data Core Components and Eco-System including Data Ingestion, Data Processing and Data Loading (PySpark, Python, Hive, HDFS and AWS).
- Skilled in programming languages such as Python and SQL to manipulate and analyze large datasets.
- Uploaded and processed terabytes of data from various structured and semi-structured sources into HDFS.
- Very Good understanding of Partitioning, Bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance.
- Have experience making AWS Glue Jobs for data flow, data pipeline and workflow management.

TECHNICAL SKILLS

- **Language:** Python, SQL
- **AWS:** S3, Glue, EC2
- **Big Data & ETL:** PySpark, HDFS, Spark SQL, HIVE, YARN
- **Project Management:** JIRA
- **Orchestration:** CA Autosys
- **Version Control:** GIT

WORK EXPERIENCE

Feb 2021 - Present

Organization: Unacademy (Edtech)   D*esignation:* Data Engineer

Project Name: Tracking and Predicting Users' Behaviour using Big Data

Technologies Used: PySpark, AWS S3 and AWS Glue

Description: This project deals with monitoring, analyzing and predicting the users' behavior with the help of big data produced after their engagement on Android/iOS Applications and Web pages.

Key Result Areas:

- Ingestion of big data generated by users' interaction into the S3 bucket.
- Development of end-to-end data transformation framework using PySpark on AWS.
- Creating Glue crawler for importing the database & tables and creation of data catalogs.

- Creation of AWS Data Lake for the purpose of analytics projects and Machine Learning.
- Involved in Data pre-processing data operations like data validation, data cleaning, schema evolution, etc. by using AWS Glue.
- Using SKlearn library, performed Scaling on data.
- Created glue workflow to run the ETL Jobs and store the data in S3 location.

Aug 2019 – Feb 2021	*Organization*: OmvSab IT Solutions	*Designation*: Big Data (Associate)

Project Domain: Banking

Technologies Used: MySQL, HDFS, HIVE, HBase, Spark

- Involved in gathering the requirements, designing, developing, and writing the script files for processing data and loading to HDFS.
- Loading files to HDFS and writing hive queries to process required data.
- Creating Hive tables with Partitions to store the processed data in tables.
- Set up Hive with MySQL as a Remote Meta store.
- Responsible to manage data coming from different sources and loaded into HDFS.

EDUCATION

- **Graduation:** B.E in Electronics and Telecommunication, 2014
- **HSC:** With Mathematics, Compute Science from Maharashtra Board.

PERSONAL DETAILS

- **Languages:** English, Hindi and Marathi
- **Date of Birth**: 11th August 1991
- **Marital Status:** Single
- **Current Address**: Domlure, Bengaluru-560071

Declaration: I do hereby declare that the above information is true and complete to the Best of my knowledge and belief.