

Yogita Bhakad

Data Engineer

+91 9764244905

bhakad.yogita11@gmail.com

[Linkedin](#) | [Github](#)

Career Objective

To work in an organization where I can contribute and enhance my knowledge and growth and also to achieve high carrier growth through continuous learning process and keep myself dynamic, visionary and competitive with changing scenario of the world.

Career Summary

- Overall **4+ years** of experience in IT industry.
- Over **1.8 years** of relevant experience in Big Data technology.
- Hands on experience with Big Data core components and Eco System including Data Ingestion and Data Processing (**Spark, Hive, Sqoop, HDFS and, AWS cloud**)
- Involving in Importing/Exporting data to/from RDBMS using Sqoop.
- In deep understanding of Spark Architecture including **Spark-core, Spark-SQL, DataFrame.**
- Very Good understanding of **Partitioning, Bucketing** concepts in Hive and designed both **Internal and External tables** in Hive to optimize performance.
- Uploaded and processed of data from various structured and semi-structured sources into HDFS.
- Hands-on experience with **AWS EC2, S3, RDS, EMR, Glue.**
- Perform tasks such as writing Scripts, write SQL query, etc.

Projects

1. Product Analysis and Marketing (PAAM)

Role: Data Engineer

In this project we get data from number of source data of our client into s3 bucket, our client is having various number of data sources and datasets. We take that upstream data with the help of AWS Glue through the various stage we process that data and then send to the RDS.

Roles and Responsibility:

- Understand the nature and characteristics of data and check how data flow.
- Extract source file from Source (S3).
- Involved in import data from one S3 bucket to another S3 bucket.
- Transform and load to RDS location using Glue jobs.
- Data Transforming with the help of Glue jobs and monitoring the process. •

Utilized :- AWS(S3, Glue, RDS), Spark, Python.

2. Data Magretion and Data Lake

Role: Data Engineer

The purpose of the project is to do the ingestion of the data that is present in source to Data lake. Source to Hadoop migration project. Data migration utility to migrate existing system into Hadoop (Big Data) from source system and perform aggregation operation on Hadoop Big Data platform.

Roles and Responsibility:

- Understand data migration requirements and needs.
- Involved in extracting data from sources. Analyze and organize raw data.
- Efficiently loading data into a spark data frame. Perform Spark Transformation.
- Data cleaning is to save the cleaned S3 bucket.
- Load data from S3 to Redshift.

Utilized : AWS(S3,Redshift), Spark

Education

Deogiri College, Aurangabad

2016 - 2018

Master of Computer Science (MCS)

Skills

Big data

Data Management: Yarn

Data Processing: Hadoop(MR,Hive),Spark

Data Ingestion: Sqoop

Workflow Scheduler: Oozie

Cloud Platform

AWS: EC2, S3, RDS, EMR, Redshift, Glue

Cloud Database: Snowflake

Programming Language

Language: Python,SQL

IDE: PyCharm, Google Colab, Jupyter.

Other Skill

Operating System: Linux, Windows

Database: MySQL, Oracle, PostgreSQL

Emulator: Putty

Bug Tracking Tools: Agile Methodology

Bug Tracking Tools: Jira,Git,Github.

Experience

Advent Informatics PVT. Ltd.

Pune— Data Engineer

April 2021 – PRESENT

EC Mobility PVT. Ltd.

Aurangabad— Data Engineer

OCT 2018 - March 2021

Personal Information

Date of birth : 18 June 1995

Marital status : Married

Nationality : Indian

Language : English, Marathi, Hindi