

# Amar Biradar

## Professional Synopsis

✉ biradaramarsk@gmail.com

📞 +91 9359459620

### CAREER OBJECTIVE

Accomplish 3.0+ years of experience in Hadoop ecosystem for designing and executing solutions of complex business problems. Having capability to explore, learn and understand newer business domain and technology. Problem solving capability, peered with good communication skills.

### Experience Summary

- **3+ years of Experience** in different domains in IT industry including Hadoop Ecosystem, Pyspark API, Databricks, Cloud Services (AWS)
- Hands on experience with Databricks Platform, AWS services and Ecosystem including Data Ingestion and Data Processing (Pyspark, Hive, SQOOP, HDFS, Aws services)
- Have hands on experience Databricks platform, Spark SQL, Dataframe, Aggregation, Datetimes, Spark Architecture, SparkCore, AWS Services S3, IAM, EMR, AWS Glue
- Capable of processing large sets of structured and unstructured data.
- Very Good understanding of Partitioning, bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance.
- Solved performance issues in Hive scripts with understanding of window functions, Joins, Group and Aggregation.
- Developed jobs for ingesting tables in Hive and from Hive to Big Query.
- Strong problem solving & technical skills coupled with clear decision-making.
- Received appreciation and recognition from client several times.
- Hands-on Experience in mainframes and Hadoop eco-system. Good exposures in Requirement Analysis, Customization and user acceptance.

### Skill Set

- Operating System: Windows, Linux
- Hadoop Ecosystem: HDFS, YARN, Apache Hive, Sqoop, Apache Spark.
- Big Data Processing Platform: Databricks, AWS Glue
- Cloud service: AWS S3, IAM, EMR, AWS Glue
- Programming Language: Python, SQL
- Ticketing Tool: JIRA
- Database: MySQL, Oracle

## Work experience

- **Organization:** - Advent Informatics

Role: DATA ENGINEER

Duration: - 2019-Present

### **Project-1 Overview:**

#### **Data Analytics on Retail Domain**

**Duration: Jun/2021-Present (1.1 Years)**

- **Technologies:** Databricks (Tool), Pyspark, Aws Cloud service AWS S3, Apache Hive

- **Description:**

We are receiving feeds from different categories of product like Accessories, Electronic devices, supermarket chain product also the feeds like inventory and price with respect to main items from source/upstream. We used to get this file in s3 location. Then we clean and validate the data. This processed data will be used by PowerBi team to do the meaningful insights this is accomplished through interactive and user-friendly dashboards and reports. PowerBi team uses this for front and end Purpose.

- **Roles & Responsibilities:**

- Attending daily scrum call with team member and discussing the task on the particular sprint.
- Responsible for applying data cleaning techniques to correct and remove inaccurate and corrupted data using PYSPARK, and performing various pyspark transformation based on the requirement.
- To run Query Optimization techniques in pyspark while doing processing.
- Filtered data and performed multiple Transformations in-order to get like Snappy filtered data and performed multiple Transformations in-order to get Leaning Model and Analysis.
- Implement Data Validation, Quality Checks.
- Responsible in case of failure of code in Pre-Prod.
- Support PowerBi Team in developing meaningful dashboard as per requirement.
- Tools and Technology used in this project S3, Databricks, Pyspark, Hive.

### **Project-2 Overview:**

#### **ENABLING E-COMMERCE**

**Duration: Jul/2020 - mar/2021(0.8 years)**

- Technologies: AWS S3, RDS, Glue, Pyspark

- **Description:**

E-com receives feeds from different categories of product like tablets, liquid/syrup, capsule, spray, drop, topical medicine etc, Also the feeds like inventory, price with respect to main items from source/upstream. We use to get these files into s3 location. In load glue job we clean and validate the data. For storing purpose, we are using AWS RDS(ORACLE) For each feed which we use to load will have its respective extract glue job, job will get the data from RDS and as per extract logic will dump CSV to the target S3 location. This data is used by SAP/downstream. SAP uses this for front end purpose.

- **Roles & 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.
- Analyzing PROD tickets – Issues can be data related or code related.
- Job monitoring.
- Import data from S3 raw bucket and do the cleaning and masking and again move to s3 clean bucket by using glue.
- Worked on the transformation glue job to move data from the s3 cleanse to the RDS dimensional and fact tables.
- Creating a DDL script for creating tables
- Validating the records once the transformation glue job completed the loading of data into EDS.
- Work with the teammates to understand and support them in importing data from source to S3 by using AWS DMS Environment: S3, AWS GLUE, AWS RDS, MySQL, PowerBI

## **Educational qualification:**

- (2014-2018) Bachelor of Engineering from CSMSS College of Engineering, Aurangabad with 81.00%.

## **Achivements:**

- Academy Accreditation- Databricks Lakehouse Fundamentals (09/2022 – 09/2023)

## **Personal Information:**

- **Date of Birth:** 12/02/1996
- **Languages Known:** English, Hindi, Marathi, kannada
- **Permanent Address:** Aurangabad, Maharashtra, INDIA

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

**Date: 14/10/2022**

**Place: Aurangabad**