

**Name** : Ms. Malvika Sunil Deokar

**Mob No** : +91 8805141405

**Email** : [malvikadeokar1611@gmail.com](mailto:malvikadeokar1611@gmail.com)

**Address** : Trimurti Vihar , Near Shiv Palace 1, Bakori Road, Wagholi, Pune - 412207

### CAREER OBJECTIVE

Being highly qualified Data Engineer with experience in the industry, I enjoy creative problem solving and getting exposure on multiple projects, and I would excel in the collaborative environment on which your company prides itself.

### CAREER SUMMARY

- **Overall, 3+ years** of experience in different domains in the IT industry and mainly in Hadoop Ecosystem Development and **PySpark** and Cloud service (Amazon Web Services).
- Have experience writing data extraction logic in **PySpark**.
- Have Experience building optimized big data pipeline using HDFS, Hive, **PySpark** on Cloud Premises
- Have a hands-on experience on **Amazon web services** mainly S3, EC2, EMR, RDS, IAM, Glue and Lambda.
- Experience processing large amounts of structured and unstructured data including integrating data from multiple sources
- Perform tasks such as writing scripts, calling APIs, write SQL queries.
- Develop batch processing, integration solutions and process structured and Unstructured Data.
- Good experience on Spark Architecture including Spark-Core, Spark-Sql.
- Experience in using Hadoop distribution like AWS.
- Experience in working on file compression techniques as well as various file formats like ORC, Parquet and AVRO.
- Experience in transferring data from RDBMS to HDFS and Hive tables using Sqoop and Spark API's.
- Implemented proof of concept in Glue.
- Experience in creating tables, partitioning, bucketing and loading in Hive.
- Gather and process raw data at scale and analyze processed data.
- Analytical mind and ability to learn , relearn concepts.

## TECHNOLOGY PROFICIENCY

- **Platforms:** Windows
- **Programming language:** Python, SQL
- **Big Data Ecosystem:** HDFS, Sqoop, Hive, **PySpark** , PySpark-SQL, ETL
- **Distribution Package:** AWS
- **IDE:** PyCharm, Jupyter Notebook
- **Cloud Technologies:** AWS (S3, EC2 ,EMR Cluster, RDS, IAM, Glue)
- **Relational database:** MySQL, Oracle
- **Another tools and technology:** JIRA Tool, Github

## ORGANIZATIONAL SCAN

- **Organization:** Infister Technology Pvt Ltd
- **Designation:** Data Engineer
- **Location:** Pune
- **Duration:** From June 2020 to Till Date.

## WORK EXPERIENCE

### **Project #1: DATA ANALYTICS FOR RETAIL DOMAIN**

**Role:** Data Engineer

#### **Outline:**

The purpose of the project is to migrate the client's data from the source (AWS S3), Perform ETL and Data-Processing on the Databricks and push the data to storage (AWS S3). The data was used to feed different applications in the analytic division.

**Environment:** Pyspark, spark-SQL , Pycharm, Databricks, Jupyter Notebook , S3

#### **Roles & Responsibilities:**

- Daily stand up meetings for tasks status discussions.
- Apply the technical knowledge in PySpark to solve the daily problem.
- Creating dataframes from different file formats.
- Design Complex Schemas for Data Frames.
- Gathering data from the entire source for a client then ETL the data using Spark, Pandas, and Spark Sql.
- Perform many transformations to clean the data as per the business requirement.
- Analyzed the SQL scripts and designed it by using spark SQL for Faster performance.
- Involved in working on Spark Sql Code for Faster Data Processing and Better Performance.
- Installing required libraries in the cluster based on task.
- Optimizing code performance by using various optimizations as required.
- Writing / storing the processed data to required destinations in various file formats.

## **Project #2 : AWS DATA MIGRATION**

**Role: Data Engineer**

### **Outline:**

The purpose of the project is to build the data warehouse using Amazon Redshift which is cloud based data storage.

**Environment:** AWS S3, DMS, AWS Glue, Redshift, PySpark, SQL, Python.

### **Roles & Responsibilities:**

- Attend daily Scrum calls and team meetings.
- Weekly call with client for their requirements and further enhancement.
- Managed Redshift data warehouse project overseeing data extraction, migration and transformation.
- Extract data from sql using AWS DMS, migrated to redshift and optimizes storage in s3.
- Design glue PySpark script for data transformation including joins and aggregations.
- Created cloud architecture using AWS services (DMS, S3, Glue, Redshift) for seamless data flow.
- Build a proper pipeline for the data .
- Athena integration: Integrated Athena for processed data, enabling smooth s3 to redshift transfer.
- Data Quality Analysis : Ensure data Accuracy and resolved anomalies during extraction and transformation.
- Documentation Coordinator: Produced user-friendly documentation outlining architecture, processes and scripts.

### **EDUCATIONAL QUALIFICATION**

- Done BE (CSE) from University of BAMU in MAY-JUNE 2014.
- Done HSC from Maharashtra State Board, Aurangabad in MAY- 2010.
- Done SSC from Maharashtra State Board, Aurangabad in JUNE- 2008.

### **PERSONAL DETAILS**

Date of Birth : November 16<sup>th</sup> 1992

Languages known : English, Hindi and Marathi

I hereby declare that the above Information given is true and correct to the best of my knowledge and belief.

Date:

Place: Pune

(Ms. Malvika Sunil Deokar)