Vijay Deshmukh

Associate Consultant

✉ deshmukh.vijay0395@gmail.com

📍 Aurangabad(MH),India

📱 **+91 8668586378**

🔗 linkedin.com/in/vijay-deshmukh-180b57240

Professional Summary

- Over all 3+ years of experience in IT industry and over 2+ years of relevant experience in BIG DATA technologies including Hadoop Ecosystem, Pyspark, Cloud Services (AWS), Oracle database, NoSQL database such as HBase, Sqoop, Linux shell script .

- Having experience in all stages of the project including requirements gathering, designing & documenting, development, performance optimization, data extraction, cleaning and reporting.

- Exposure to AWS cloud technologies such as EC2, S3, EMR, RDS, IAM, Redshift, Athena etc.

- Knowledge of programming language Python and its libraries related to bigdata.

- Knowledge of Spark using python, spark core, spark architecture, and Spark-SQL for faster testing and processing of data.

- Worked on structured, semi-structured data frame.

- Good understanding of Partitioning, bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance.

- Worked in following domain such as Telecom, Retail industry.

- A good team player, good communicator.

- Comprehensive knowledge Agile methodology, coupled with excellent communication skills.

- Experienced in branching, tagging and maintaining the version across the Environments using tools like Git, GitHub.

Professional Experience

O2 Soft Solutions

**Designation: Associate
Consultant**   Jan 2022 to till date

Project # 1:

Enterprise Data Warehouse Analysis

The purpose of The Project is to do the ingestion of the Telecom data that is present on cloud storage as well as Clients Database to Data pipeline which was modelled as per the client.it is cloud to HBase ingestion Project. Data Migration Utility to migrate existing system into Data hub which was built on top of Hadoop and uses spark for operations.

Roles and Responsibilities:

- Understand the upstream source nature and work with business cases.
- Involved in Collecting Business Requirements from Business Users, Translate into Technical Design (Data Pipelines and ETL workflows).
- Written spark jobs(python) to extract records from Upstream source.
- Work closely with the business and analytics team in gathering the system requirements.
- Import and export batch data and delta into Phoenix (HBase) and Hive using Pyspark.
- Involve in creating Hive tables, loading with data and writing Hive queries.
- Raise JIRA tickets for Infrastructure, Platform issues.
- Resolve JIRA tickets raised by scrum master.
- Comes up with missing scenarios and get them Clearfield.
- Reconcile the imported or transformed data as per business requirement.
- Responsible for branching, tagging and maintaining the version across the Environments using SCM tools like SCM as Git, GitHub.
- We have to filter data and perform multiple operations in order to get meaningful insights that will help the client in decision making and analysis.
- Implement Data Validation, Quality Checks, Profiling.

**Environment: Spark, Python, Phoenix, HBase, Hive, Hadoop, GitHub,
   AWS (EMR, RDS, S3) Zookeeper.**

O2 Soft Solutions
Designation: Associate Consultant

Project # 2:   Aug 2020 to Nov 2021

Retail Store Data Pipeline

      The purpose of this project is to do the ingestion of the super market that is present in mainframe system to data warehouse. Its mainframe to HDFS ingestion project in Realtime. Data migrate into MySQL. Later the extraction would be done in PowerBI utility.

Roles & Responsibilities:

- Importing and exporting data into HDFS and HIVE using big data ETL tool (Sqoop).
- Involved in creating tables, loading data into hive tables and process it as per the business logic using hive queries.
- Write Hive DDL to Create Table for Optimize Query Performance
- Writing Hive queries for data analysis.
- Ingest Flat Files like Delimited, Fixed Length, etc. into Data Warehouse.
- Worked with various HDFS file formats like ORC, RC, Avro, Csv.
- Load processed final data to AWS S3.
- Data Reconciliation to ensure good Data Quality.
- Implement Data Validation, Quality Checks.
- Resolve JIRA tickets raised by scrum master.

     Environment: Hive, Sqoop, Hadoop, GitHub, AWS (EMR, RDS, S3)

Professional Qualification

Bachelor of Engineering
Deogiri Institute of Engineering and Management Studies, Aurangabad,
Dr. Babasaheb Ambedkar Marathwada University
Percentage: 66.00% [2017]

Higher Secondary Education (HSC)
Swami Vivekanand Junior College,
Aurangabad.
 Score:72.00% [2013]

School Secondary Education (SSC)
Shahid Bhagat sing High school, Bajaj
Nagar, Aurangabad.
Score:86.55% [2011]

Personal Skills

- Eager to learn new things as well as good learner.
- Strong analytical, design skills and problems solving abilities.

Personal Dossier

Full Name	: Vijay Rustumrao Deshmukh
Date of Birth	: 03rd October 1995
Languages	: English, Hindi, Marathi.
Known Address	: Mhada Housing Society Flat no 60, (M4 wing), Near Alphonsa High School, Bajaj Nagar, Aurangabad Maharashtra 431136

Declaration

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

Vijay Deshmukh