Suraj Madhewad

Software Engineer

9420138473   surajmadhewad01@gmail.com   Aurangabad, IN   9420138473

CAREER SUMMARY

- Total of **3.4 years** of experience in the IT business, including technical competence in the **big data** environment and in-depth knowledge of development on the **Hadoop Ecosystem, PySpark APIs, Cloud services (AWS)**, and **NOSQL** databases like **PostgreSQL, Hbase**, etc.
- Experienced using **DataFrame APIs** and **spark-sql APIs** to write data extraction logic in **PySpark**.
- Good understanding of **Partitioning, Bucketing** concepts in **Hive** and designed both **Internal and External tables** in Hive to optimize performance.
- Worked with different RDBMS like **Oracle SQL, MySQL** and Non RDBMS like **PostgreSQL, HBase**.
- Capable of processing large sets of **structured, semi-structured** data.
- Good knowledge on file compression techniques as well as various file formats like **ORC, Parquet, JSON.**
- Extensively used **Databricks** PaaS and cluster for cleaning and processing clients source data.
- Created the dataframes, worked on **transformations** and developed **PySpark** applications.
- Hands on experience working with **SQL tables, Views, CTEs, Joins** and windowing functions.
- Experience creating an optimized big data pipeline on cloud infrastructure using **RDS, Hive, Hbase, and PySpark (AWS).**
- Learn how to use Amazon Web Services, focusing on **S3, EC2, AWS EMR, Redshift, IAM Role, and RDS.**
- Applied various **optimization techniques** in PySpark to achieve **performance tuning** of Spark Application.
- knowledge of Apache Oozie and Apache AirFlow, as well as workflow management tools and schedulers.
- Good Exposure on **Apache Hadoop, Hive,Hbase,Shell Scripting and HDFS.**
- Interacted with BA and Product Manager to devise ways to optimize the data pipeline and boost productivity.
- Strong technical, analytical and problem-solving skills. Always keen to pick up new technology learn it and apply it for optimized outcomes

PROFESSIONAL EXPERIENCE

Peaksense Technologies Pvt Ltd   December 2019 - Present

Project : Data Analytics for E-Commerce Domain
Domain: E-Commerce
Description of Project :
Aim:

- To help client **analyze visitors** to its website, **monitor the traffic** and **track users actions** to analyze product popularity and trend.
- To convert **passive visitors to active buyers** by providing them suitable offers.
- Generate weekly sales and revenue reports.

Responsibilities :

- Requirement gathering and devising ways of execution with the help of Team Lead and Scrum Master.
- Mounting S3 source data to **Databricks DBFS**.
- Worked on Data Engine part in **Databricks Pyspark.**
- Major work in **cleaning raw data, processing it, handling nulls and validation**.
- Extensive use of **DataFrame APIs** and **Spark-SQL APIs** to process data.
- Working with variety of **Datatypes** as well as manually assigning schemas to overcome datatype issues.
- Developed business logic using Spark programs and test them on sample data.
- Application of various **transformations and actions** as per business requirements.
- Handling variety of file formats including **parquet** to achieve optimized query outcomes and speedier execution.
- Application of **Optimization techniques** to fine tune Spark application job.
- Handling errors, exceptions and provide bug fixing solutions to clients.

Project : Enterprise Data Warehouse Analysis

Domain: Telecom

Description of Project :

Aim:

- To provide clients solutions to their business problems to enhance their clogged business growth,**develop subscriber revenue generation reports** and track the **subscribers complaint status**.
- To analyze customers likelihood to opt for certain plan options and check on the their trend of latest actions.

Responsibilities :

- Understand the upstream source nature and work with business cases.
- Gathered source data and push it into big data pipeline for **cleaning, processing** and business analysis.
- Use of **PostgreSQL RDS, Hbase and Hive** to build data Pipelin
- Used PySpark Engine to process the data on **EMR cluster**
- **Filtered data** and perform multiple **transformations** in order to get meaningful insights that will help the client in decision making and analysis.
- We have written Business logic in Pyspark using **AWS EMR** platform.
- Created **Hbase table using phoenix** Script.
- Worked on both **Internal and External tables of Hive** for staging and final output storage.
- Implemented **Data Validation** and unit testing on sample data.
- Analyzed and solve problems and put bug fixes on production.

Infosys BPM Pvt Ltd   September 2018 - April 2019

Process Executive

- Response to customer after invoice and order process
- Coordinated with AR operation Team for credits and debits requests.
- Queries Inter Department coordination for resolving costumer issue
- Providing support service to Cisco client via. Mails and web chat.
- Service-level agreement (SLA) committed by utilities to meet client requirements

EDUCATION

Bsc I.T (2018)

MGM G Y Pathrikar college of CS and IT

TECHNICAL SKILLS

- **Hadoop Ecosystem:** Apache Spark, PySpark, Spark-SQL,, HDFS, Apache AirFlow , Apache Kafka, Apache Hive, Apache Hadoop
- **Languages:** Python, Shell Script
- **Cloud Technologies:** AWS (S3, RDS, EC2, EMR, IAM, RedShift, Athena, Glue, Lambda)
- **Spark Platform:** DataBricks
- **Databases:** Oracle, MySQL, PostgreSQL.
- **Coordinated Service:** Zookeeper
- **Other tools:** Git and GitHub

PERSONAL DETAILS

- **Name** : Suraj Pandurang Madhewad
- **Gender** : Male
- **Marital Status** : Unmarried
- **Languages Known** : English, Hindi, Marathi
- **Date of Birth** : 09-July-1996
- **Address** : :B-14 B-WING Shri Ram Hari Complex,Near Sunny Center Pisadevi Road, Ambedkar Chowk Aurangabad 431001