

SHRIDHAR KHADSE

Address: Yavatmal India | Mobile: 8983547297

Email: shridharkhadse07@gmail.com

Linkedin: www.linkedin.com/in/shridhar-khadse

SUMMARY

- Having **3+ years** of experience in IT industry including technical proficiency in big data environment.
- Hands on experience with Big Data core components and Hadoop Eco System including Data ingestion and Data processing.
- Good understanding of Hadoop Architecture and its major core components like Hive, HDFS, Pyspark, Yarn.
- Good understanding of Spark Architecture including Spark Core, SparkSQL, DataFrame.
- Have experience writing data extraction logic in **SQL**.
- Capable of processing large set of Structure and semi-structure data.
- Perform task such as writing scripts, SQL queries, Python cases.
- Experience in writing **Pyspark** code to migrate data from **RDBMS** to **HDFS** and **Hive** tables using the **Pyspark** jobs.
- Experience to in writing spark program to perform data cleansing, Data Ingestion using Pyspark.
- Solved performance issues in spark with understanding of Joins, group and aggregations.
- Good understanding of **Partitioning**, **Bucketing** and Loading in **Hive**.
- Strong problem solving & Technical skills coupled with clear decision making.

PROFESSIONAL EXPERIENCE

Zielotech Infosolutions India LLP:

Web Telecom Unit Analysis
Data Engineer

Pune India

April 2021 - present

Description:

The purpose of Web Telecom Unit Analysis project is the rapid rise in use of smartphone and other connected mobile device has a spurt in the volume of data flowing through the network of telecom operators. It is necessary that the operators process, store and extract insights operators from the available data. The data analysis can help them to increase profitability by helping optimize network usage and service and enhance the customers experience and improves the security.

Role & Responsibility:

- Importing Historical data from **RDBMS** (PostgreSQL) into the **Nosql** database (Hbase) using pyspark scripts.
- Understand the client requirement for the developing project.
- Write Hive DDL to create **Hive External** table & **Hive manage** table for optimize query Performance.
- Writing pyspark jobs for importing the data into Hbase and from Hbase to Hive.
- Implement the data validation, quality check, profiling, de-duplication & deals with nulls.
- Worked on Optimization of Spark jobs.
- Involved in creating some UDF when needed.
- Involved in importing data from Hive Managed to External Hive table and defined the **static** and **dynamic partitions** as per requirement for optimize the performance on production datasets.
- Write Hive query for data analysis to meet the business requirements.
- Creating pyspark jobs for data transformation and aggregation.

Environment:

Pyspark, Hadoop, Hive, Hbase, Phoenix, AWS EMR cluster, AWS S3, RDS

SHRIDHAR KHADSE

Address: Yavatmal India | Mobile: 8983547297

Email: shridharkhadse07@gmail.com

Linkedin: www.linkedin.com/in/shridhar-khadse

Zielotech Infosolutions India LLP:
Emporium Data Analytics

Pune India
Apr 2020-Feb 2021

Description:

Emporium Data Analytics is an analysis for everything in your business, from your sales and inventory to your customer data. It gives you the ability to effectively track customer actions like their purchases and foot traffic in your store and to increase their overall revenue of our client.

Role & Responsibilities:

- Daily task checking on sprint board.
- Developed Pyspark program to extract the data from S3 Bucket.
- Writing a pyspark script for data validations, data cleaning.
- Developed pyspark program for data transformations in Databricks environment and load that data into RDS.
- Involved in working on spark SQL code as an alternative approach for faster data processing and better performance.
- Monitoring Jira tickets and taking care of them.

Environment:

Databricks, SQL, Pyspark, AWS S3, RDS(MySQL)

EDUCATION

Jagadambha College of Engineering and Technology

Bachelor of electrical Engineering

- CGPA 7.9/10

Yavatmal, India

Aug 2013 - July 2017

TECHNICAL SKILLS

Hadoop Ecosystem: HDFS, MapReduce, Hive, Sqoop, Spark, Hbase, SparkSQL, Pyspark

Programming Language: Python

Cloud Services: AWS (S3, EMR, RDS, EC2, Redshift, Athena, Glue, IAM role)

Relational Databases: MySQL, Oracle, PostgreSQL

Operating System: Unix, Windows

Cloud base Platform: Databricks

Version Control Tool: Git, Github

DECLARATION:

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

Date: / /2022

Shridhar Khadse