

Kiran S Vyavahare

Mob No: +91 8149805279

Email: kiranyyavahare1999@gmail.com

CAREER OBJECTIVE

To be associated with a progressive organization that provides an opportunity for a challenging and rewarding career by applying my knowledge, skills, and potential in the profession. I would also like to make a positive contribution to your organization by promoting team spirit.

CAREER SUMMARY

- **Overall, 5 years of experience** in different domains in the IT industry. 2 years experience in Testing and another 3+ years experience in **Big Data** like Hadoop Ecosystem Development and PySpark and Cloud service (Amazon WebServices).
- Currently Working as **Data Engineer** in Data Quarter Technologies Pvt Ltd since 3+ years.
- Hands on experience on **SQL, HADOOP** and **Spark Architecture** including Spark SQL, Data Frames, Hands-on experience in SQL query analysis.
- Having good knowledge in Hadoop and Spark ecosystems include **Hive, Shell Script, Spark SQL, Spark Streaming and MapReduce** along with various distributions such as **Cloudera**.
- Have experience writing data extraction logic in pyspark.
- Having experience in all stages of the project including requirements gathering, designing & documenting architecture, development, performance optimization, data extraction, cleaning and reporting
- Experience processing large amounts of structured including Integrating data from multiple sources.
- Experience in writing spark programs to perform data cleansing, Data Ingestion.
- Develop batch processing , integration solutions and process structured Data.
- Hands on experience on Major components of Hadoop Ecosystem like HDFS, SPARK, HIVE, SGOOP, YARN.
- Solved performance issues in Spark and Hive scripts with understanding of Joins, Group and aggregation.
- Very Good understanding of **Partitioning And Bucketing** concepts in Hive and designed both **Internal and External tables** in Hive to optimize performance.
- Have a hands-on experience on Amazon web services mainly **S3 , EC2 , Glue, EMR , Redshift , IAM Role , RDS, LAMBDA**.
- Experience on using tools such as **PyCharm, winscp, Databricks, Jupyter Notebook**.
- Experience in launching EMR cluster, Redshift cluster, EC2 instances, Amazon Data Pipeline, Simple Work flow Services.
- Experience in creating and loading data from Database to S3 Buckets.

- Experience in using Hadoop distribution like Cloudera.
- Experience in working on file compression techniques as well as various file formats like **ORC, Parquet, CSV, RC, JSON, AVRO**.
- Experience in transferring data from RDBMS to HDFS and Hive tables using Sqoop and Spark API's.

TECHNOLOGY PROFICIENCY

- **Platforms** : Windows, Linux.
- **Programming language** : Python basic, SQL
- **Scripting Language** : Shell Scripting.
- **Big Data Ecosystem** : HDFS, Sqoop ,Hive ,Pyspark, Pyspark-SQL
- **Distribution Package** : AWS
- **IDE** : PyCharm, Jupiter Notebook, Google-Colab,
- **Cloud Technologies** : AWS (S3, EC2 & EMR Cluster, Redshift, RDS, IAM Role, Glue, Lambda), Azure(Databricks)
- **Relational database** : Oracle, MYSQL, SQL developer
- **Another tools and technology** : GitHub

ORGANIZATIONAL SCAN

- **Organization:** Data Quarter Technologies Pvt Ltd
- **Designation:** Data Engineer

WORK EXPERIENCE

Project #1 : Health and Insurance Data Analysis
Environment: Mysql,Oracle,Hadoop,Hive,HDFS

Role: Data Engineer

The purpose of the project is to do the ingestion of the medical store data that is present in Mainframe system to data Warehouse. It's Mainframe to **HDFS** ingestion project in Real Time. Data migration utility to migrate existing system into **Oracle**

Description & Responsibilities:

- Import Data into the **Hive** from various Relational Databases (**Oracle, MySQL**) using **Sqoop**.
- Write **Hive** DDL to Create **Hive** Table for Optimize Query Performance.
- Work with Various File Formats in HDFS **Flat Files** like Delimited, ORC Files, Sequence Files, Fixed Length, etc. into **Hive Warehouse**.
- Implement Data Validation, Quality Checks.
- Involved in Collecting Business Requirements from Business Users, Translate into Technical Design (Data Pipelines and **ETL** workflows).
- Involved in import data from various **RDBMS** into **HDFS** using **Sqoop** which includes Incremental Load to populate **Hive External Table** and vice-versa.
- Involved in importing data from various **RDBMS** into **Hive** Tables which includes Queries using **Sqoop**.
- Designed both **Managed and External Hive Tables and Defined static and dynamic partitions** as per requirement for optimized performance on production datasets.
- Written Hive Queries for Data Analysis to meet the business requirements

Project #2: Data Pipe Line (customer relational Data services)

Environment: -MySQL, Pyspark, AWS (S3, IAM, Glue, Lambda, Redshift)

Role: Data Engineer

Provided **Data Warehouse** solution to client using **Data Pipeline** using **AWS** services like **S3, IAM, Glue, Lambda, RedShift** etc. Client had data on various databases and we need collect it at one place for giving data warehouse solution.

Description & Responsibilities:

- Providing warehousing solution to client with their problems related to business growth, fraud analysis etc.
- Import data from **MySQL** database using **PySpark/Python** using **MySQL** Connector.
- Creating landing **S3** bucket in **AWS S3**. Using Pyspark as a ETL tool we load the data into **AWS S3**.
- For migration purpose creating **GLUE** job.
- Create **LAMBDA** function for triggering purpose .
- Upload file in landing Bucket and check Lambda job got automatically invoked or not and check job got triggered or not. Check and Validate data got copied or not to final bucket.
- To Create **Data Warehousing** solution uses **AWS REDSHIFT** . Copy data from **S3** final bucket to **REDSHIFT** using copy command.

EDUCATIONAL QUALIFICATION

- Done B.Tech in July 2015.
- Done HSC (Science) from Maharashtra State Board in July 2011 .
- Done SSC from Maharashtra State Board in July 2009.

PERSONAL DETAILS

Date of Birth : October 1st 1993
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.

Kiran Vyavahare