CONTACT

✉ sharmaharsh0703@gmail.com

☐ +918791517842

📍 Bangalore, India

in https://www.linkedin.com/in /harsh-sharma-123364196/

LANGUAGES

English ● ● ● ● ●
Hindi ● ● ● ● ●
French ●

Harsh Sharma

Data Engineer

SUMMARY

2 years IT experience. Currently working as a Data Engineer. Engineering Professional with extensive experience in ETL, Data warehouse, Data integration, Data modeling and Data analysis. Demonstrated proficiency in Agile and CI/CD software development.

KEY SKILLS

ETL Tools - Apache Airflow, Informatica Powercenter 10.x, Databricks ETL
Databases - Oracle db, SQL Server, Postgre sql, Teradata, MYSQL, Snowflake, MongoDB, Cassandra
Programming Languages - SQL, Python, Unix shell scripting, Scala, Spark sql
Tools - ELK Stack, Docker, HCA WORKLOAD SCHEDULER, BITBUCKET, GIT, GITLAB, BAMBOO, Postman Api, Jenkins, Ms Office tools, Hue, Cloudera
Hadoop Ecosystem - Hive, HDFS, Yarn, Impala, MapReduce
Frameworks - Apache Spark (PySpark, Spark-Shell), Kafka(Streaming Platform), Databricks

EXPERIENCE

IQVIA

Data Engineer   Sept 2022 – Present

- Extensively working on big data technologies like Spark, Scala, Hadoop components and various other tools.
- Maintaining, Operating and Executing Airflow workflows.
- Development of Hadoop/Spark Applications in Scala for Data Acquisition, ETL workflow in data lake using Jenkins CI/CD Auto Build, Test and deployment in Cloudera Distribution.
- Using Databricks for the code changes and performing various operations in analysing the data.
- Usage of ELK Stack, for logging and visualization in Kibana.
- Responsible for platform wide support of our application, Building the reports, analysing them. Meeting other customers and road map driven requests.
- Working in setting up various projects related to healthcare and analysing the huge data for clients including **multinational pharmaceuticals.**

Carelon Global Solutions (formerly Legato)

Associate Software Engineer   June 2021 – Sept 2022

- Good understanding of Data Integration, Data Warehouse, ETL and Database development primarily with Teradata for Healthcare Industries.
- Have extensively worked on Business Requirements, Designing ETL Architecture, Data modelling, creating design document as per requirements with developing ETL program for supporting Data Extraction, transformations and loading using various ETL tool like Informatica PowerCenter.
- Writing SQL Scripts for code deployment, data extraction and data transformation.
- Writing UNIX Shell Script for automate code development, migration and scheduling ETL component.
- Preparing Data Extract Design Specification and Data mapping document by understanding the physical and logical dimensional model.

PROJECT EXPERIENCE

Oct 2022-present
Project: US Healthcare Projects Onboarding (AI/ML Platform)
Software/Tools/Technology Used: Spark, Hadoop, SQL, Apache Airflow, UNIX, Scala/Python, Kibana ELK, Jenkins, Gitlab

Description: Project involves building DAGs in Airflow and optimizing the Spark Jobs involved, for efficiently handling huge medical data. Using Hadoop with ElasticSearch to make data accessible for the dashboard.

Roles and Responsibilities:
- Analyze and transform stored data in Snowflake/Hive/HDFS by writing Spark Jobs based on business requirements.
- Optimizing Data Pipelines to reduce the time consumption of DAGs.
- Creating automation scripts to improve workflow.
- Writing UNIX shell scripts and python/scala scripts for the code changes

Mar 2022-Sept 2022
Project: Member Notification of Provider Termination
Software/Tools/Technology Used: Informatica, Teradata, SQL, UNIX

Description: Project involves creating and maintain the process where members should be notified when their provider leaves the Particular network.

Roles and Responsibilities:
- Analyzing requirements and designing data models, preparing Data Extract design specification document.
- Preparing Data Extract Design Specification and data mapping document.
- Writing UNIX shell script wherever needed in project for scheduling informatica code.

Aug-2021 – Mar 2022
Project: Payment Model for Healthcare
Software/Tools/Technology Used: Teradata, Snowflake, Informatica, SQL,Unix

Description: Project involves creating and maintaining the process to calculate the relevant saving for the HealthCare firm, after all the claims and spends processed for each term to its member.

Roles and Responsibilities:
- Business requirements gathering, Designing ETL Architecture, Data modeling, creating design document as per requirements.
- Writing UNIX shell and SQL script wherever needed in project for scheduling informatica code and data extraction.
- Project is implemented in both on-premises and off-premises.

EDUCATION

Btech Computer Science & Engineering	July 2017 – June 2021
SRM Ghaziabad, India	86%
Senior Secondary	April 2017
Dewan Public School Meerut	87%

"Think lightly of yourself and deeply of the world."
– Miyamoto Musashi