

Deepak Sharma

Data Engineer | HCL Technologies Limited

Noida

+91-8447221657

deepakdsharma03@gmail.com

www.linkedin.com/in/deepak-sharma03

Summary

Approx. **Three years** of experience in Information Technology (IT) industry in **ETL, Data Warehousing, AWS Cloud, Business Intelligence** and **analytics**.

- Working on a Pharma project to perform **ETL**, data **ingestion**, data **transformation** to push clinical data into Enterprise Clinical DataMart.
- Hands on experience on creating ETL process using **Python, PySpark** and **Spark-SQL** on **AWS EMR and Databricks** while loading data from **AWS S3** buckets to create DataMart on **Redshift** and **ATHENA**.
- Worked briefly in **AWS Athena & Redshift** to run queries and perform other table operations to **load data** from S3 Buckets.
- Responsible for creating **S3 Data Lake** to load files from server.
- Worked on basic SQL queries for testing and manipulating data.
- Hands on with basic Linux, Docker commands.
- Hands on experience with **Git** to push, pull, edit and run code.
- Team player with excellent communication and problem-solving skills.

Technical Skills

- Python, Basic Pandas, PySpark, Spark SQL, Databricks
- AWS-EMR, S3, ATHENA, RedShift, GLUE
- GitLab
- Linux

Experience – HCL, Noida

AUG 2018 - Present:

Project: **Enterprise Clinical Trials Systems – Data Mart**

- Working in a US based pharmaceutical client project towards design & creation of Clinical Trials Data Mart in AWS for the reporting needs of the project while analyzing and integrating the data from clinical data sources.
- Performed data ingestion, data integration, ETL and data validation using Python, Pyspark, **AWS S3, Athena, EMR** and **GLUE**.
- Developing Python and Pyspark codes from various clinical trials datasets to create an ETL having transformations including filter conditions, missing values, dropping duplicates, validating conditions using python & Spark functions, Data frames and Spark SQL.
- Hands on with SQL to write Scripts/queries for commercial claims, Attribution and Clinical Data Transformation.
- Creating buckets and loading data into **Amazon S3 data lake via AWS CLI**.
- Creating tables and querying data in **Amazon Athena and transforming data via EMR - spark**.
- Using **Pyspark transformations** and **Spark SQL** to create ETL jobs.
- Have been a part of Client Calls and have presented multiple times.

Academics

YEAR	DEGREE	BRANCH	UNIVERSITY	PERCENTAGE
2016	B.C.A	Computer Science	SIKKIM MANIPAL UNIVERSITY	62%

Personal Details

- DOB- 03-02-1996
- Languages Known- English, Hindi
- Native Place- Ranikhet, Uttarakhand