

# MANOJ KUMAR M

## BIG DATA ENGINEER

### CONTACT

+91 - 9700670090  
manojkumar.m2990@gmail.com  
Hyderabad, India

### EXPERIENCE SUMMARY

- 6+ years of IT industry experience as a Software Developer, with 3+ years in Big Data technologies.
- Extensive experience in multi-tiered architecture with Scala, Hadoop, Spark using various Data Bases Oracle and MS SQL and Hive.
- Good Experience in implementation of Data Ingestion & Transformation workflows.
- Worked with cloud services like Amazon Web Services (AWS) & Cloudera and involved in ETL, Data Integration and Migration.
- Hands-on experience in Amazon Web Services (AWS) Cloud services like EC2, S3 and EMR.
- Good Experience on Hadoop Ecosystem, HDFS, MR & YARN and distributed systems.
- Experience in Hive and UDF's.
- Experience in writing Sqoop scripts for data Exports/Imports.
- Experience in creating Oozie Jobs & Scheduler for Executing the time series Data in Hue UI.
- Good Experience in Data Management, ingestion, compute and storage with Cloud and hybrid systems.
- Good experience in implementation of Data Lake and Data Mart in Hive and spark.
- Experience in GitHub and maven.
- Good Experience in SDLC methodologies like Safe Agile, waterfall and Scrum.

### EDUCATION

#### Bachelor of Technology

RSR Engg. College, JNT University  
2008 - 2012

### TECHNICAL SKILLS

**Operating Systems:** Windows, Linux

**Languages:** Scala, SQL

**Big Data Technologies:** Hadoop, Hive, Spark, Sqoop, Oozie

**Databases:** MySQL, Oracle, Teradata

**IDE's:** Eclipse, IntelliJ

**Tools:** WinSCP, Putty, Git, Jira.

### WORK EXPERIENCE

#### Associate - Projects

**Cognizant Technologies India Private Limited**

Jan 2022 - Sep 2022

##### Cargill Walmart Validation Project - Modern Data Engineer - Jan 2022 to July 2022

The Objective of this Project is to migrate the existing TERADATA, DB2 tables into HIVE Tables.

Understanding the dataflow of source tables. Creation of mapping documents and compute business logic for derived target columns. Developing the Hive, Spark queries and YAML Scripts to load the data developing Atomic Workflows and Jobs for daily Run and Generating the Data quality Reports for the comparison of Source and Target Data.

##### Responsibilities:

- Extracted the Data from Teradata into HDFS using the Sqoop.
- Created Hive tables, dynamic partitions, buckets for sampling and working on them using Hive QL.
- Handled Importing data from various data sources (JSON, CSV, TXT)
- Performed Transformations using Hive& Spark Scala and loaded data into HDFS.
- Validating Source files with validation files using Spark Scala
- Created & Scheduled Oozie Workflows in Hue UI.
- Involved in developing UNIX Shell script to run both processed & validation files.

#### Software Engineer

**Blankbot Technologies Private Limited**

Apr 2017 - Dec 2021

##### Reliance Retail Limited - Big Data Engineer - Jun 2019 - Dec 2021

Model Service creates optimized customer offerings that reliance can use real time on the website. This project allows reliance to provide customized offerings for customers who do not normally get personal attention from Account Executives, and it is expected to increase revenue incrementally.

##### Responsibilities:

- Implemented with Data Ingestion & Transformation workflows in hive/spark with Cloudera and Create datasets in Hive using HDInsight.
- Designed the big data solution for storing and retrieving the Data models and schema using Hive and spark.
- Performed Import & Exports of remote data to AWS s3.
- Developed Spark Scala & Spark SQL code to perform transformations
- Ad-hoc analysis for Real-time log analysis using Spark.
- For bad data Processing, cleansing, and verifying the integrity of data used for analysis.
- Maintaining and administering the multi node cluster.
- Handled Optimization techniques for long running jobs in spark and hive.

##### Bharat Matrimony - ETL Developer - Apr 2017 - Mar 2019

##### Responsibilities:

- Analyzed the requirement and performed Impact Analysis based on the requirements.
- Involved in writing DML & DDL scripts to transform Data and populate in Target table.
- Involved in modifying existing procedures and ETL workflows according to the new business needs using Microsoft SQL Server Management Studio and Control M

#### ETL Tester

**PebbleBridge Technologies**

Jan 2014 - Nov 2015

##### Responsibilities:

- Record count check, before and after the transformation logic applied.
- Data flow validation from the staging area to the intermediate tables.
- Ensure the key field data is not missing or Null.
- Check if the aggregate values and calculated measures are loaded in the fact tables.
- Check modeling views based on the target tables.