

Manish Kumar

Experience: 1.8 years

Summary

I am a self-motivated Individual with a background in Information Technology. I have excellent interpersonal, communication and analytical skills and has overall 1.8 years of experience in Big data Hadoop and Apache Spark ecosystem experience ingestion, transformation querying, processing and analysis.

Skills

- Having 1.8+ years of experience in design and implementation of highly efficient and optimised ETL pipelines.
- Having good interpersonal skills, interaction with team, ability to work independently.
- Develops innovative solutions to Big Data issues and challenges within the team
- Working with the end-user to make sure the analytics transform data in meaningful ways.
- Design and build ETL pipelines to automate ingestion of structured and unstructured data.
- Experience in building data ingestion and data processing ETL jobs in snowflake cloud Data Warehouse.
- Ability to quickly learn new technology, business domains and processes.

Big Data Technologies	Description
Open Source Frameworks	Apache Spark, Apache Hadoop, Apache Airflow, Apache Drill, Apache Kafka
Cloud Technologies	Snowflake, AWS Glue, AWS S3, AWS Ec2
Programming Languages	Pyspark, Python, SQL, SparkSql, SnowSql

Work History

Data Engineer, Nov 19 to Current

Company-Team Computers, Gurugram

Project Details:

1. Automobile Company – Data Lake & Data Warehouse

Description: Designed and implemented the Data Lake in Hadoop ecosystem defining 3 layers(Staging, Validated, Refined) and Data Warehouse in PostgreSql. Therefore Preparing Dashboard in Qlik Sense.

Some of the multiple aspects of projects, which are:

- Implementation of data lake using HDFS.
- Developed data integration pipelines for Full-Load from source(Postgres, Flat files) to storage(HDFS) using Apache Spark and did Data Compression upto 70% (from csv to Parquet).
- Developed ETL pipelines for data validation and refined layers using pyspark ETL and query the results using Apache Drill.
- Design and Developed Data-Governance table in PostgresDb storing meta-data Information of all 3 layers.

Tools and Technologies:

Hadoop, Apache Spark, PySpark, Python, Apache Drill, SQL, PostgresDb

2. Snowflake Data Lake & Data Warehouse Reporting (POC)

Description: Design and Implemented Batch and Real-time data Ingestion Pipeline using Qlik Replicate and Snowflake Streams for CDC implementation which is further used for preparing real-time Dashboard reporting with Qlik Sense.

Involved in multiple aspects of projects, which are:

- Design and implementation of Full-Load and Streaming data from source using MQTT. Data ingestion to data lake using Qlik Replicate tool, data transformation using snowflake.
- Data flow Management using Apache Airflow.

Tools and Technologies:

Mosquitto Broker, Snowflake, Snowsql, Apache Airflow

3. Insurance Company - Data Lake & Data Warehouse

Description: Developed and Implemented ETL jobs for data ingestion and transformation. Designed and implemented the Data Lake in in HDFS using Spark and Hadoop environment.

Involved in multiple aspects of projects, which are:

- Implemented data lake in Hdfs using Spark Sql for ETL jobs
- Understanding business requirements and preparing data and preparing data pipeline
- Worked on the optimization of existing scripts written in Qlikview environment
- Built integration services in python

Tools and Technologies:

Apache Spark, Apache Hadoop, Python, Postgresql, Apache Drill, PySpark

Qualifications

Degree	Institute	Year	Percentage
Bachelor of Technology	Panipat Institute of Engineering & Technology, Kurukshetra University	2019	74%
12th	Modal School, Chandigarh	2014	67%
10th	Modal School, Chandigarh	2012	7 C.G.P.A