

Mansimran Kaur

Data Engineer



<https://www.linkedin.com/in/mansimran-kaur-a97636190/>



Delhi, India

SUMMARY

- 4+ years of experience in Data engineering and Analytics through development of systems and processes to ingest, transform and analyze large scale data for external clients.
- Designing and building scalable data pipelines, real-time data processing solutions, and data applications.
- Good working knowledge with different modules of Hadoop Ecosystem(SQOOP,Hive,HDFS,Spark)
- Experience in building and optimizing Hive queries to improve the performance
- Good knowledge of HDFS architecture and MapReduce programming model

WORK EXPERIENCE

Consultant

Jun-2021-Present

Deloitte USI

Project 1 : Data Migration

- Involved in designing,developing and optimizing data pipeline for Integrating vast amount of structured and unstructured using Hadoop and spark framework which supports the company ability to complete analysis on multiple datasets quickly.
- Worked on a live 30 nodes Hadoop cluster.
- Worked with highly structured and semi structured data of 90 TB in size (270 TB with replication factor of 3).
- Built Spark Scripts for Cleaning and Processing data by utilizing Scala depending on the requirement.
- Worked on data ingestion with Sqoop to load data in HDFS.
- Created and worked on Sqoop jobs with incremental load to populate Hive External tables.
- Worked on performance tuning in Spark Code and it's configuration

Project 2 : Data Migration

- Worked on creating pipeline objects, reading data into the pipeline, and applying transformations to it using dataflow runner for execution.
- Extracting complete data/delta records from different source systems like MD Staff, Salesforce, Symedical via API call or using SQL queries.
- Reading and writing data to Pub/Sub and BigQuery.
- Applying transformations like standardization, canonical conversion, data quality checks.
- Loading transformed data into Target system via API call .
- Unit testing the functionality
- Worked on creating architectures and flow charts for various source systems

Assistant System Engineer

Tata Consulting Services

Jun-2019 to May-2021

Project 1 : ELT workflow

- Building a pipeline to extract data from multiple sources from Azure storage explorer and applying transformations to get insights from the customers data, calculating credit score and customers loan availing patterns .
- Scheduling the pipeline to run every day.
Tools: Azure Data Bricks,Azure Storage,PySpark,Hive

Project 2 : Data Analysis

- Created End to end pipeline for extracting the data from more than 50 sources and applying various complex transformations (to more than 900 target fields) using dynamic SQL queries.
- Created JSONS for more than 5 million cases based on the nested path of these target field and loaded it in S3 bucket.
Worked on performance tuning. Improved performance by more than 40%
- Created audit log table for logging success/failure of each case in transformation layer as well as while creating JSONS.

TECHNICAL SKILLS

- **Hadoop Tools/framework :**
Sqoop,Hive,HBase,Spark (Batch, SparkSQL, Spark Streaming)
- **Programming language :**
Java ,Scala
- **Database :**
Hive,Oracle, MySQL
- **Cloud :**
GCP(Dataflow,BigQuery,Pub/Sub,Cloud Storage), Azure(Azure Data Bricks)

CERTIFICATION

- Databricks – Data Engineer Associate
- Google Cloud Certified Professional Data Engineer
- Data Science and machine learning with Python

PERSONAL PROJECTS

- Building a pipeline to extract data from Google cloud storage and apply transformations to it and load the transformed data into BigQuery table.
Scheduling the pipeline to run every day.
Tools: Google cloud storage,Spark with Scala,BigQuery
- HR Analytics (Exploratory & Predictive Analytics) Description: To build a model and this allows organizations to Identify high-risk employees who are going to attrite, build profiles of those most likely to leave or stay and understand how risk is distributed throughout the organization.
Analytics Tools: Excel, Python
Analytics Technique: Classification (Logistic Regression)

ACHIEVEMENTS

- **Outstanding Performance Award(Dec 2023)**
Appreciated for delivering the project successfully
- **Outstanding Performance Award(Nov 2022)**
Appreciated for solving complex design issues
- **Applause Award (Jul 2022)**
Recognized for my work in the area of performance improvement and code designing
- **Applause Award(Nov 2021)**
Appreciated for picking up Java on ad-hoc demand from client and helping development team in achieving the sprint deliverables

EDUCATION

- **Bachelor of Technology in IT**
GGSIPU Per : 81.4%
2015-2019
- **Class 12th**
Holy Child Auxilium Per : 83.4%
2015
- **Class 10th**
Holy Child Auxilium Per : 90%
2013