Gurpreet Singh Heera

SUMMARY

2+ yrs. of Experience in Data Engineering that include:

- Implementation of massive data processing using **SPARK** to Draw valuable inferences from **Bigdata** and helping **Business** to take smart & accurate decisions with **predictive analysis**.
- Establishing various data pipelines between **OLTP** source systems like **Oracle ERP** and target data warehouses like **Teradata** and **Hadoop**.
- Automating **SQL** queries using **Unix shell scripting** and **ETL** tools like **Informatica** and **NIFI**.
- **Designing data models,** flows & maintaining data received from **different source systems** by developing optimized code.
- Working on version control tools like **GitHub** and **Bitbucket.**

TECHNICAL SKILLS

- Hadoop & Teradata
- Java, Unix & Python
- SQL & Hive-QL
- Spark & Sqoop
- Oozie & NIFI
- GitHub & Bitbucket

WORK EXPERIENCE

Data Analyst (Supply Chain)
NCR Corporation Pvt. Ltd. / Jan.2019—Present

PROJECTS

Supply Line Management

- **Technologies :** Hadoop, Spark, Oozie
- **Organization :** NCR Corporation -Internal project
- **Team Size   :** Individual
- **Role   :** Developer
- **Description   :** Track all our raw material orders to see the performance of raw material providers. Data goes through various transformations & processing using HQL Scripts & Spark and then finally lands for visualization using tableau. Code is scheduled using oozie workflow.

CONTACT

gshira96@gmail.com

+917986144521

Sector-54, Gurgaon

Haryana,122002(IN)

EDUCATION

Punjabi University
Patiala, Punjab

B. Tech, ECE, GPA-6.96

Jun.2019

Mount Carmel School
Hoshiarpur, Punjab

12th, Science,75.20%

Jun.2015

10th, General, 76.33%

Jun. 2013

Build Purchase Order Repository in Hadoop

- **Technologies :** Hadoop, Hive-QL, Oozie, Unix and Informatica

- **Organization :** NCR Corporation -Internal project

- **Team Size   :** Individual

- **Role   :** Developer

- **Description   :** Integrate Purchase Order Pipeline between Oracle ERP and Hadoop. All the transactions performed on ERP tables are captured and brought into Hadoop staging tables via Informatica CDC in the form of raw data. Based on this data, records are inserted, updated or deleted from main tables to keep them in sync with ERP.

Oracle Cloud Integration

- **Technologies :** Hadoop, Hive-QL, Oozie, Unix and NIFI

- **Organization :** NCR Corporation -Internal project

- **Team Size   :** 3

- **Role   :** Lead

- **Description   :** Ingest supplies & demands data from Oracle cloud into Hadoop using NIFI. Build and schedule code for performing full refresh and incremental load. Apply Analytics on final dataset to generate reports for better inventory planning & efficient production. Here we use Hive-QL for transformations and analytics while Unix and Oozie are used for scheduling.

Order Management & Tracking

- **Technologies :** Teradata, MLOAD, SQL ,Unix

- **Organization :** NCR Corporation -Internal project

- **Team Size   :** Individual

- **Role   :** Developer

- **Description   :** Build record activation & deactivation logic using SCD 2 architecture to track and store orders on hold and released from hold. Data was ingested via files using Teradata MLOAD utility and processed using Teradata BTEQs.