I am a qualified professional with about 3+ years of cumulative experience in Data Engineering, Data Integration, Machine Learning, Data Warehouse, ETL, reporting and artificial intelligence.

# **Core Skills**

|  |  |
| --- | --- |
| Programming Languages | Python, C/C++, R, MATLAB. |
| Libraries & Software Environments | Pandas, NumPy, SciPy, Matplotlib, SparkMLib, NLTK, Keras, TensorFlow, Scikit-learn, SVM, XGBoost. |
| Database Management | SQL, RDBMS, NoSQL |
| BigData & Hadoop technologies | Hadoop, HIVE, HDFS, SparkSQL |
| Cloud Technologies | Azure, DataBricks, ADF, DataLake. |
| Revision Control & CI/CD | Git, BitBucket, Jenkins, Jira. |
| Reporting / Analysis Tools | Excel, Tableau, SAP BW, PowerBI |

# **Work Experience**

**Data Engineer Jan 2020 – Till Date**

***Appnovation, Toronto***

* Experience in handling Structured, Unstructured and Semi Structured data.
* Created and integrated runtime environment to extract data from **BLOB** **storage** using Azure **Data Factory**.
* Developed data processing solutions using **Azure Databricks** and wrote **Spark** transformations scripts resultant output to **SQL DW**.
* Data processing and migrated from **Oracle to Azure**. Loaded the data into **DataLake** and Build schedule **Power BI reports.**
* Experience in writing Subqueries, Stored Procedures, Triggers, Cursors, and Functions on **MySQL database.**
* Used open-source tool such as **AIRFLOW** for scheduling the **DAG jobs**
* Extensively worked on **Python** and built the custom **ingest framework** and worked on **Rest API** using Python.
* Designed and architected scalable data processing and analytics solutions, including technical feasibility, integration, development for Big Data storage, processing and consumption of Azure data, analytics, big data, business intelligence **(Reporting Services, Power BI), NoSQL, Stream Analytics and Data Factory.**
* Leveraging existing frameworks by using **Python, SparkSQL** to create new data pipelines.
* Enhance and Bug fixes on priority basis.
* Analyse the data and create new scope, attending defect triage meetings along with all stakeholders and bring the defects to closure.
* Data loaded from different source systems like **DB2** and **flat files** to **HADOOP ecosystem.**
* Developed machine learning models using **SparkMLib**.
* Assume ownership over certain components and continuously work to improve them
* Fast-moving, developer-driven organization full of brilliant and ambitious people

**Environment: Python, SQL, SparkSQL, SAS, Rest API, Hadoop, Azure, AIRFLOW, PowerBI**

**ETL developer/ Data Analyst Sep 2018 – Jan 2020**

***Angoss, Toronto***

* Enhanced and automated our existing data pipeline.
* Expertise in **ETL/ELT** creation, designing data pipelines using **SQL Server**, **Oracle, RDBMS** Platform for analysis.
* Built file format services for clients to handle multiple file formats like **CSV, FIXED, Parquetetc** using **Python, Spark, and Hadoop.**
* Developed **automation testing** framework using **PySpark** to address the data inconsistencies.
* **Extract-transform-load (ETL)** ongoing maintenance of third-party data to the **Azure** environment from **SAP HANA system**.
* **Data Integration** into **SAS** environment from different sources like **DB2, ORACLE**
* Take assignments from different stakeholders and perform SAS coding to create different database marketing lists.
* **E2E project** involves **ETL**, from **Teradata** to **Hive tables** (Raw, Transformation, and Consolidation) and validate JSON.
* Apply machine learning algorithms such as clustering, Nearest Neighbours etc., for regression and classification problems using **Python & R software.**
* Propose recommendations; give project estimation and tender inputs and proposal formats.
* Worked on inventory management.
* Developed expertise in **ETL/ELT** by creating and designing data pipelines using SQL Server and Oracle database platforms for analysis.
* Ingested large sets of data from multiple **OLTP** databases into **OLAP** data warehouse through ETL pipeline where data is queried for analytical insights using **Microsoft Power BI**.
* Unit testing in the Development box and Integration testing in the Quality box.
* Ingested large sets of data from multiple **OLTP** databases into **OLAP** data warehouse through **ETL pipeline** where data is queried for analytical insights using **Microsoft Power BI**

**Environment- Python, SAS, Azure, ETL, PowerBI**

# **Education**

**Bachelor of Engineering (Computer Science & Engineering) 2013- 2017**

Visvesvaraya Technological University, Belguam, India