# Final Project – Technical contribution

**PROG8451: BIG DATA INTEGRTION & STORAGE**

Aakash Baskaran | 8937042

Harshitha Anishetty |

Big Data Solutions Architecture, Conestoga College

PROG8451 – Big Data Integration & Storage

Professor Ms. Shanti Couvrette

Technical contribution in ETL report of news API data analysis

In our final ETL project, each of the two team members played a vital role in the technical execution. Below is a summary of the individual technical contributions made by each member:

## Team Members Name (Student Id) & Contributed Topic:

### **Aakash Baskaran (8937042)**

* **Creating Producer, Consumer Script**

I (Aakash) started the project by working on producer.py file, where I initialize Kafka producer and topic and getting data from news\_api and finally sending the extracted data to kafka producer. And for consumer.py file is where I handle receiving data from kafka producer and saving it in local file system.

* **Initialize Spark Streaming**

I (Aakash) did initialize the Spark Streaming, created schema, initialize readStream, added watermark and defined writeStream

### **Harshitha Anishetty ()**

* **Analyzation of stream data in spark**

I (Harshita) worked on analyzing the streaming data from spark and wrote queries using aggregate functions to get valuable insights and store the result in output directories.

* **Added additional pipeline component**

I created an additional component to the pipeline by creating hive tables to handle the streamed real time data and ran queries to get more insights.