Project Proposal – Group 8

Broadband: Data usage, Data Quality and Latency

About Connecting Windsor-Essex (CW-E):

- Connecting Windsor-Essex began as an information technology project, focused primarily on building an extensive high-speed fibre optic network that would bring world-class connectivity to several of the region leading institutions. CW-E now expanded to provide services for the regions business and especially in under-serviced rural areas, residents who wanted access to high-speed internet.
- Connecting Windsor-Essex is a non-profit organisation comprised of Board of Directors, Executives and consortium of over 40 member companies, associations and other organisations in Windsor-Essex.
- CW-E is funded by stakeholder groups including The City of Windsor, The County of Essex, Essex County Library, UoW, St. Clair College and other organisations.

CW-E Projects:

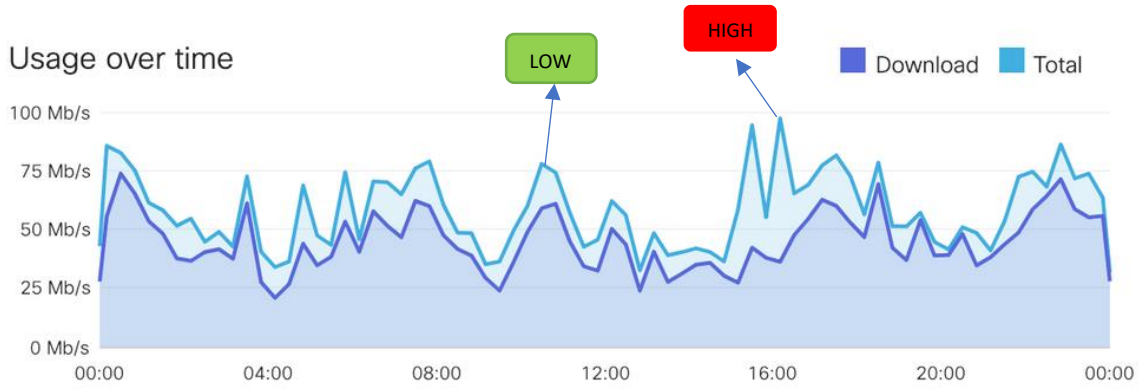
- **SWIFT (South Western Integrated Fibre Technology):** SWIFT is a non-profit corporation and was initiated by the Western Ontario Wardens' Caucus to address regions connectivity challenge. The project is delivered in partnership with member municipalities as well as with government of Ontario and Canada.

- **Technovation Windsor-Essex:** Technovation is a global challenge started in California in 2009 and has since grown to encourage 10,000 girls across 78 countries to pursue technology courses and careers. Girls work in teams of 3-5 students and are challenged to identify a problem in their community, research and create a mobile app that solves the issue. Student teams work closely with mentors to build the idea into a working app with a business plan as guided through the curriculum.

- **Internet Exchange Point:** An Internet exchange point (IX or IXP) is a physical infrastructure through which multiple networks can connect and share resources. CWE will be host to WEDIX, which serves to improve internet services by keeping local traffic local, in turn improving security and reducing latency.

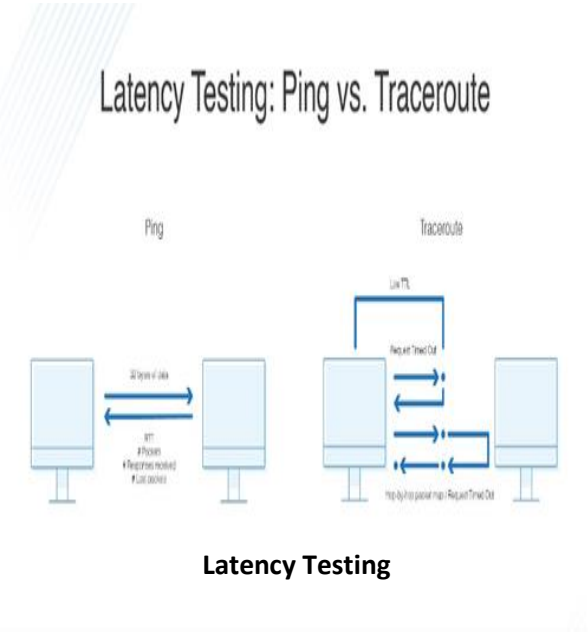
Problem Statement:

It has been a concern for local independent carriers in providing connectivity to few referred locations as per CW-E. High consumption of data volume during the day and further in the evening when students return from schools are the main reasons for the concern. Therefore, Carriers are managing high data usage throughout the entire day. In the past data usage looked like a sine wave i.e., high data usage in the evening and low for the day.

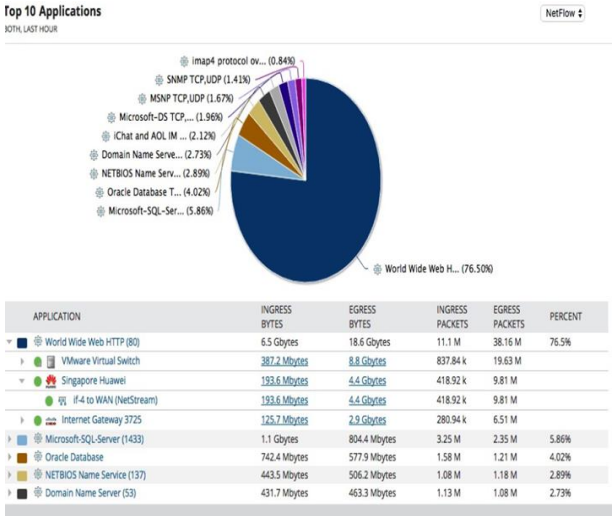


Challenges:

1. Tracking data usage and quality of connection.
2. Monitoring data consumption per school.
3. Identifying peak and low times of the day.
4. Identifying applications that consume high amount of data.
5. Identifying the most used application.
6. Latency analysis.
7. Comparing quality of the two carries being used.
8. Reporting cycle and dashboard to communicate the insights.



Latency Testing



Top used applications

Obstacles:

- **Data Extraction:**
  The school board does not allow anyone to log into their facilities to obtain data. Identify the tools that can be used to extract and allocate the data to a secure location.

Motivation:

We being a group of 25 year old techie's, our journey in this world of internet has been very exciting. From the phase of enormous number of wires and machines to compact routers and contact-less satellite services, internet has made a remarkable technological leap. Our lives have become more convenient as compared to the times without the access to internet. It has also become such an important part of our lives that we cannot live without it. Also, the Internet is the invention of cutting-edge science and modern technology.

As students at St. Clair College and being active residents of Windsor-Essex region we feel highly responsible and grateful to align our data analytics skills and support the progress of this region in the field of broadband services. After initial discussion with Dan Circelli, CEO & President, CW-E, We learnt the importance of quality broadband services with respect to schools and major organisations in this region.

In regard to the problem statement, this is an interesting interdisciplinary research and insights problem involving analytics and networking. With previous experience in regard of broadband and its metrics, We believe our knowledge with respect to the domain and application of our analytical skills to this area networking can be a bridge for bringing in ideas, analysing various factors and finding solutions related to the same. Considering all the above, this project has motivated us.

Evaluation with respect to challenges (changes with the progress in project):

1. **Data extraction** with support of **CW-E** and appropriate people at the **school board**.

2. **Loading data** securely into SQL database by integrating **SSMS** (SQL Server Management Studio) with **Docker containers** or **Azure SQL Server**.

3. **ETL** operations using **SSIS** (SQL Server Integration Services) or PySpark (depends on the size of data)

4. **Data Cleaning** and **Transformation** using **MS Excel** and **T-SQL querying**.

5. **Analysing** broadband metrics such as data consumption.

6. Latency and quality using python analysis which include **EDA** and **regression and tests**.

7. Building **ML Prediction Model** for data consumption using **scikit-learn**.

8. Building **reports** using **SSRS** (SQL Server Reporting Services).

9. Building **dashboards** for broadband metrics using **Tableau** and **Azure Synapse Analytics.**

By performing and working on all above procedures and meeting the requirements proposed by stakeholders, we consider the project to be successful.

Estimated Impact on stake holders

High data consumption can lead to higher broadband costs affecting the stakeholders directly. Restricting applications that consume more data can be more economical to the stakeholders. Stakeholders will obtain clear information on the amount of data being consumed with respect to time of the day and applications being used allowing them and further allowing them to make decisions as per their requirement.

Group Members:

Sai Abhiram GP- 0779155
Bramha Teja Dasarraju Venkata- 0787625
Pavan Kumar Morasa- 0783875
Yamini Bhupendrakumar Prajapati- 0793972
Shweta Sharma- 0795669

References:

- About company, projects and domain : Connecting Windsor-Essex (cw-e.ca)
- Data Latency and Consumption: What is data latency and how to measure it | Snowplow
- Data consumption patterns impact our behaviour (kpmg.com)