WEEK 4 REPORT

Topic: Live Data – APIs, CRM, Zendesk, WebSockets, Web Scraping & ETL Pipelines

Intern Name: Harika Vavilapalli
Internship Domain: Data Analytics / Data Science
Tool/Concepts Covered: Live Data, APIs, CRM, Web Scraping, ETL Pipelines
Focus: Understanding real-time data flow and how data is collected, processed, and used in real applications

1. Introduction

In the fourth week of my internship, I focused on understanding the concept of **Live Data** and how real-world applications continuously collect, update, and manage data. Compared to static datasets (which we download once and analyze), live data keeps changing every minute or every second depending on the system.

This week was mainly about learning how live data is handled in real-time projects using different methods such as **APIs, CRM tools, Zendesk ticketing systems, WebSockets, Web Scraping, and ETL pipelines**. I also learned why automation becomes important when the data is continuously updated.

Overall, Week 4 helped me understand how companies manage data in real environments and how data moves from the source to the final application or dashboard.

2. Objectives of Week 4

The main objectives of Week 4 were:

- To understand what **live data** is and how it differs from static data
- To learn how data is collected from real-time sources using **APIs**
- To understand the role of **CRM systems** in handling customer and business data
- To explore **Zendesk** and understand how support tickets work in real-time
- To learn the basics of **WebSockets** for real-time communication
- To understand **Web Scraping** as an alternative method for collecting data
- To learn the concept of **ETL pipelines** for continuous data processing
- To connect all these concepts with real-world data workflows

3. Work Done / Tasks Completed

3.1 Understanding Power BI Basics

This week was not directly focused on Power BI, but I understood how live data concepts can later connect with Power BI dashboards. Power BI becomes more useful when it receives updated data regularly, which helps in making dashboards dynamic and up-to-date.

This made me realize that once live data is collected and processed properly, it can be used for reporting and monitoring systems in real-time.

3.2 Importing Data into Power BI

I did not import data into Power BI during this week, but I learned an important concept: **live data can also be connected to dashboards** through automated updates or pipelines. Instead of manually importing files again and again, data can be refreshed automatically when new records arrive.

This helped me understand that live systems reduce manual work and improve accuracy.

3.3 Data Cleaning Using Power Query

This week, the focus was more on **continuous data flow**, but I understood that live data still requires cleaning. The difference is that instead of cleaning once, we need to clean continuously using automation.

So the concept of cleaning is still the same:

- handle missing values
- remove duplicates
- correct formats, but the approach becomes automated through scripts and pipelines.

Live Data Concepts Covered (Week 4)

1) Live Data

I learned that live data means data that keeps updating continuously from a source such as:

- websites
- sensors
- customer interactions
- support tickets
- online transactions

Live data is important because it helps companies make decisions based on what is happening right now instead of old data.

2) APIs (Application Programming Interfaces)

I learned that APIs are one of the most common ways to collect live data. APIs allow applications to communicate with each other and share information securely.

For example:

- weather apps collect data using APIs
- stock market apps get updated prices using APIs
- disaster and emergency platforms can use APIs for updates

This concept helped me understand how real-time data can be fetched and updated automatically without manual downloading.

3) CRM (Customer Relationship Management)

I learned that CRM platforms are used by companies to manage customer-related data such as:

- customer details
- interaction history
- issues/requests
- service updates

CRMs help businesses keep track of customer communication and maintain proper support and service management. This made me understand how important organized data is in real-world systems.

4) Zendesk

I explored **Zendesk** and understood how it works as a customer support platform. The main learning was how Zendesk converts customer messages into tickets automatically.

Key things I understood:

- When a customer sends an email, a **ticket gets created**
- Support teams can reply through Zendesk and the response goes back to the customer's mail
- Zendesk also provides AI features that help in improving the reply tone and formatting

This gave me a real-world understanding of how live customer support systems work.

5) WebSockets

I learned that WebSockets are used for **real-time communication**, where data updates instantly without refreshing the page.

Unlike APIs (which are usually request/response based), WebSockets allow continuous connection between client and server.

Examples of WebSocket use cases:

- live chat applications
- live sports score updates
- live notifications
- real-time tracking systems

This helped me understand how real-time data can be delivered instantly to users.

6) Web Scraping

I also learned about **web scraping**, which is a method of collecting data from websites when APIs are not available.

This includes:

- extracting information from web pages
- collecting updated content automatically
- storing scraped data for analysis

I understood that web scraping can be useful, but it must be done carefully and ethically based on website rules and permissions.

7) ETL Pipelines

One of the most important concepts I learned this week was **ETL Pipelines**, which stands for:

Extract → Transform → Load

- **Extract:** collect data from sources like APIs, files, websites, CRMs
- **Transform:** clean, filter, format, remove duplicates, handle missing values
- **Load:** store the cleaned data into a database, dashboard system, or storage

This concept helped me understand how live data is continuously processed and updated in real-world applications.

4. Key Learnings

By the end of Week 4, I learned:

- Live data is continuously updated and needs automation
- APIs are the most common way to fetch real-time data
- CRM platforms help in managing customer and business data efficiently
- Zendesk works using ticket-based live support systems
- WebSockets allow instant updates through continuous connection
- Web scraping is useful when APIs are not available
- ETL pipelines are essential for handling live data in a structured way
- Automation is necessary for real-time workflows

5. Challenges Faced

Some challenges I faced during this week were:

- Understanding the difference between APIs and WebSockets clearly
- Learning how live data systems manage constant updates
- Connecting all the concepts together into one workflow
- Understanding how data is cleaned continuously in real-time projects

But step by step, the concepts became clearer and more practical.

6. Outcome of Week 4

At the end of Week 4, I was able to:

✅ Understand what live data is and why it is important
✅ Learn how APIs are used to collect real-time data
✅ Understand CRM systems and their role in business data handling
✅ Explore Zendesk and understand ticketing + AI support features
✅ Learn the basics of WebSockets for real-time updates
✅ Understand web scraping as an alternate data collection method
✅ Learn the working of ETL pipelines for continuous processing

7. Plan for Next Week

In the upcoming week, I plan to:

- Apply live data concepts to a mini project or dataset
- Explore API integration practically
- Work on storing and updating data continuously
- Strengthen understanding of automation pipelines and real-world workflows

Conclusion

Week 4 was focused on understanding **live data systems and real-world data flow**. I learned how data is collected using APIs and web scraping, how it is managed using CRM platforms, how Zendesk works in real-time customer support, and how WebSockets provide instant updates. I also understood the importance of ETL pipelines for extracting, transforming, and loading continuously updated data. This week gave me a strong understanding of how modern applications handle real-time data efficiently.