Reflection Paper on Capstone Project and Big Data Architecture Course

**Introduction**

I have chance to use knowledge from Big Data Architecture course in real-world project, especially in capstone project about traffic data analysis. My work on this project very similar to what we learn in class, like making data processing pipelines and building apps that can handle large scale. This paper talk about how my project and course goals are same, especially in making data pipelines, working with different types of data, and understanding cloud computing and scalable systems.

**Data Processing Pipelines and Integration of Data**

In the course, we learn how to make data processing pipelines, and I use this skill in my project. For the project, I create many scripts to read log files, calculate time differences, and process image and audio files. These scripts like different stages of data pipeline, changing raw traffic data into structured data for analysis and visualization.

For example, first script read log files and calculate time differences. This data is in JSON format, which is a type of semi-structured data. I process this data to get important metrics like time differences between events and traffic direction (forward or backward). These steps similar to what we learn in course about handling different kinds of data.

**Scalable Application Deployment and Cloud Computing**

Another important part of course is learning about scalable apps and cloud computing. In capstone project, I use this knowledge by using SSH and SFTP to connect with remote server. I download large sets of image and audio files from server for processing on my local PC.

This experience very similar to what we learn about making apps that can handle big data, especially in cloud environment. While project didn’t involve directly deploying apps in cloud, using remote servers and processing large data files is very related to cloud computing principles we study in course.

**Practical Application of Big Data Architecture Principles**

In course, we focus on big data architecture from pipeline perspective. This show in my project work. The scripts I develop for traffic data processing need to think about scalability and efficiency, which are important in big data systems. Even though project didn’t use Hadoop, ideas of processing large datasets and distributed processing still applied.

For example, I group and count log entries by 15-minute intervals, which need efficient data processing techniques when handling large traffic data. The ability to script over big data and access it in structured way, like we learn in course, is very important for success of this project.

**Visualization and Plots**

The course focus on scripting over big data and real-time data access is very relevant when I process log entries and create visualizations. The need to handle and show data efficiently, while making sure the scripts can scale with more data, is very similar to course objectives.

**Conclusion**

Big Data Architecture course give me strong foundation in big data systems, which I apply directly in capstone project. By creating data processing pipelines and handling different types of data, I learn more about big data architecture and how to use it in real-world. I gain important skills needed to do well in big data analytics.