When thinking about the smart cities and the evolution of the systems that we are going to be creating there are lots of problems that will start to arise from all the data being created that we attempt to use for analytics and tracking. One significant issue in such a system could revolve around the sheer Volume of data generated from various sensors, devices, and endpoints spread across the city. This influx of data, encompassing information on traffic patterns, environmental conditions, energy consumption, and more, can quickly escalate, overwhelming the system's capacity to process and derive actionable insights efficiently. Also considering the Velocity at which data is produced in real-time, coupled with the Variety of data formats from diverse sources, poses a formidable challenge for data aggregation and analysis. To address the 5 Vs of Big Data which are, Volume, Velocity, Variety, Veracity, and Value. They need to implement robust data storage solutions, efficient data processing algorithms, and scalable infrastructure to handle the data deluge effectively. (Wang, 2022) There are some solutions being offered by Microsoft which allow for the data pipelines to be created and stored effectively. However, I think that there should be some proper planning done when deciding how to use and make these devices. I have had some experience working on ETL pipelines and developing a way for data to be input into servers, I have come across the problem of, “is the data you’re storing useful?” Considering that there are two kinds of data, which I have found to be a common theme when making Neural networks, there is Junk data, and good data. I think that there needs to be some thought on what you need the model to do, and what inputs have a high correlation to be used in a model. (Dan Adams, Senior Vice President, Data Strategy & Operations, Precisely, 2022) Or better yet, which data are you targeting, and which data has a correlation to the output. On the next topic, a traffic flow sensor might need an extremely low latency connection because traffic jams can happen in a mere instance but on the other hand, a rainfall sensor might be okay with higher latency. When also considering the other factors such as data storage, I think that there are many reasons for the frequency of the update of the database, as I think that a scalable solution may be in order. Working with databases, I have come across so many ‘Junk’ tables, that have no actual meaning to our operation, but they are being updated constantly. I think that if the solution you are making has junk tables in it, in the case of storage and value, these will drive up the value cost of the data storage. Which is something to consider if the network of devices you are creating will be producing a ton of data. I think that there needs to be ton of planning when developing these smart cities, and the actual project planning might take years before the first round of devices are actually developed and implemented.

References

Dan Adams, Senior Vice President, Data Strategy & Operations, Precisely. (2022, March 21). *Moving from “Junk” data to data integrity*. Database Trends and Applications. https://www.dbta.com/Editorial/News-Flashes/Moving-From-Junk-Data-to-Data-Integrity-151759.aspx

Emb, T. (2024, February 6). *Smart Cities: The revolution of urban planning through IoT*. EMB Blogs. https://blog.emb.global/urban-planning-through-iot/

Wang, R. (2022, June 14). Data Storage and the Internet of Things - Phison blog. *Phison Blog*. https://phisonblog.com/data-storage-and-the-internet-of-things-2/