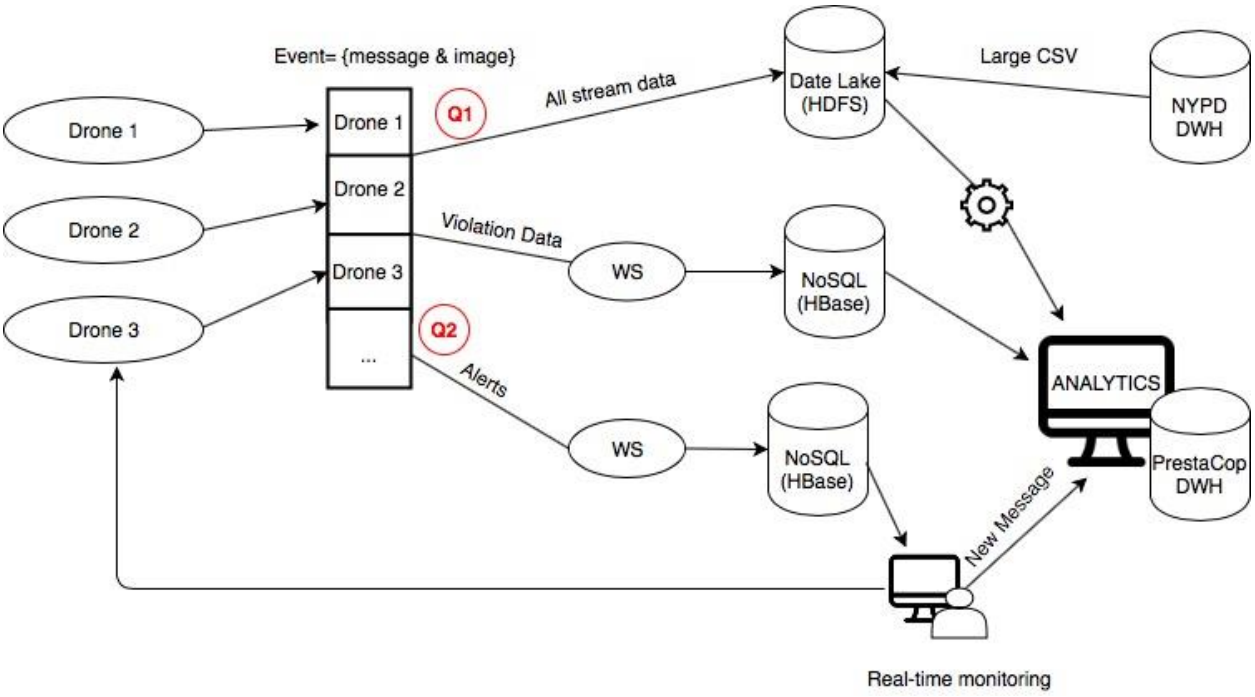
Introduction to Data Architecture Engineering



#	Case	Data	Frequency	Speed required	Priority	Database
1	Regular message	-Location -Time -Drone ID	High	Low	Variety	Stream to Data Lake
2	In case of violation	-Location -Time -Drone ID -Image -Violation code	Middle	High	CP	Stream to No SQL (by each region office) to Data Lake
3	Alert	-Location -Time -Drone ID - Image - Alert message -Violation code	Low	High	CP	Stream to NoSQL, same as #2

4	NYPD historical data (CSV)	-Historical ticket data (when, where, how, why)	Low (Batch)	Low	CP	DWH(NYPD) to Data Lake to DHW(Prestacorp)

1) What technical constraints should the architecture meet to fulfil the requirement described by the customer in paragraph «Statistics»? So what kind of component(s) (listed in the lecture) will the architecture need?

- Since they have to store all messages with no specific purpose of use, store all the message in Data Lake which allows storing a variety of data in massive amounts. A data lake will also ensure that data preparation will only be done once we read it, which in PrestaCop's case will be in the unforeseeable future. It would make sense for PrestaCop to service the data lake through the cloud because of scalability. It would be more cost-effective to scale on the cloud as PrestaCop puts more drones in circulation and also because of traffic peaks (i.e. more people park in the morning).
- By using a stream and real-time monitoring, we can quickly identify the drones that carry a violation message
- In case of violation, use stream and send the message to the closest NYPD office since it requires speed. Then store the message in Data Lake
- Regarding storage of NYPD historical data(CSV) for improving the model, use a batch system since their computer is old and not powerful

2) Same question with the paragraph «Alert»

- Again, with the stream and real-time monitoring, we can catch the alerts quickly and trace back the drone which sent the alert. Alerts would also be sent to a micro serviced database that could be monitored in real-time by an operator that would then update the message with the correct violation code.
- Use stream and send the message to the closest NYPD office since it requires speed. Then store the message in Data Lake

3) What mistake(s) from Prestacop can explain the failed attempt?

They did not process the data before sending it, thus the CSV file was too big to send.

4) Prestacop has likely forgotten some technical information in the regular message send by the drone. In the future, this information could help Prestacop make its product much more profitable. Which information?

It is suggested to add other variables when drones are collecting data, such as the peak hour, the weather conditions etc.

Moreover, after receiving regular messages, it is also recommended to store these data into a data lake, for future references and records.