LY Jean
MARECHAL Théo

Data engineering I

Preliminary questions

1. The statistic section of the document highlights the fact that we will need to handle failure in a way that guarantees that we don't lose any data but also that can handle a constant stream of new information as every peacemaker drone gives a rapport every minute.

   As such our failure handling policy will be an "at least once", and we want to use an AP system as we need the system to always be available to receive reports but don't have a high need for consistency. Finally, we want it to be optimized for Writing as we don't need to regularly read the data inside.

   So, we will use a NoSQL key-oriented database (like Riak) to store the old reports and one stream per drone to handle the constant reports.

2. In order to keep the harmony of Peaceland, it's important that the system is able to notice, process, and transmit alter as fast as possible. To that end, we will need a way to mark reports as containing alerts so that they can be treated before anything else. To accomplish that we will double the system, each drone communicating with 2 streams: 1 for general reports as described above and 1 for alert reports which immediately processes and send the relevant data to the peacemakers' office.
3. Peaceland's previous failure could be explained by the fact that the team was made up of Data Scientists when setting up systems isn't their core skill.
4. Peaceland's current Peacewatchers only send the position of the drone not of the citizens and while the 2 are likely near each other, it could still lead to delay in the peacemakers' intervention, especially in a dense urban environment. For the same reason, it seems like sending the altitude of the target could also be useful in case the target is inside a multi-floors building.

Graph of the system: