

SCALA Project

Dorian Mignon
Christopher Diamana Lutete
Alexis Morton
Jose A. Henriquez Roa

May 22, 2021

Contents

1	Preliminary questions	1
1.1	Question 1	1
1.2	Question 2	1
1.3	Question 3	2
1.4	Question 4	2
2	Project	2
2.1	Architecture Components	2
2.2	Architecture Diagram	2

1 Preliminary questions

1.1 Question 1

Question What technical/business constraints should the data storage component of the program architecture meet to fulfill the requirement described by the customer in paragraph “Statistics”?

As described by the “Statistics” section of the subject the storage component will be required to hold peacemakers’ reports as well as the peacewatchers’ temporal series. We assume the former to be some type of unstructure data (word documents or PDFs) while the latter will most like be represented in a semi-structured data format (XML, JSON, etc.). Additionally, it is specified that the estimate of a day’s worth of peacewatchers’ reports will occupy around 200Gb, which on it’s own greatly influences the choice of type storage component. Finally with regards to the CAP theoreme, It is also safe to assume that since this is a police force related task the storage component will most likely rather benefit from being “available” and “partition tolerant” and trade-off consistency by having data be propagated at night or some other time when there is less activity.

Thus, we have the following requirement:

1. AP type
2. able to natively store different data types
3. able to manage store large amounts of data

Question So what kind of component(s) (listed in the lecture) will the architecture need?

Out of the ones presented during the lectures, the storage component type that is fits all of the previous points is the *Data Lake*.

1.2 Question 2

Question What business constraint should the architecture meet to fulfill the requirement describe in the paragraph “Alert”?

The “Alert” section mostly describes constraints on the type of distributed processing. This section brings forth the necessity of an “fast” alert system.

The two distributed processing types presented during the lecture were Batch processing and Stream processing. As implied by the name batch processing treats the data only on a given schedule or when a given quota is met. Stream processing on the other hand treats the data in a continous matter. It also allows for multiple data buses for when the data needs to be redirected to multiple (stream) consumers.

Seen as peacewatcher’s are deployable field units meant for gathering data it is best to offload any type of alert directed data-processing to a separate dedicated and more secure unit. The task of this new unit would then consist of processing incomming data from the peacewatchers and forward alerts to the peacemakers when need be. This new unit essentially defines another data (stream) consumer in addition to our previous data storage component. Finally, we note from the peacewatchers’ description that the generated minute reports are mostly consisting of a small number of integers and strings making the entire reports relatively small in size, and thus a good fit for Stream processing.

In summary, we note the new requirement for *Stream processing* and a dedicated *Alert Processing Unit*.

1.3 Question 3

Question What mistake(s) from Peaceland can explain the failed attempt?

The sentence: “*not been able to set up a scalable program that can handle the load*” leads to believe that a mistake was made in the choice of the storage component, where one not natively fit for large amounts of data was chosen. The mistake could also have been in the choice of the distributed processing type, where the team could have chosen batch processing, which would inevitably result in a large accumulation of data if processing was setup to be done on an ill-adapted schedule.

1.4 Question 4

Peaceland has likely forgotten some technical information in the report sent by the drone. In the future this information could help Peaceland make its peacewatchers much more efficient. Which information?

From what is gathered by the “Drone Description” section, regarding the recorded words heard in the vicinity of the peacewatcher, being able to discern from which citizen the recorded words originated from will most likely be used for updating their designated “peace score”. Seen as the drone is already able to detect the name of the citizen through facial recognition and hear the words spoken in its surrounding, this was most likely a forgotten feature.

2 Project

2.1 Architecture Components

TODO

2.2 Architecture Diagram

TODO