

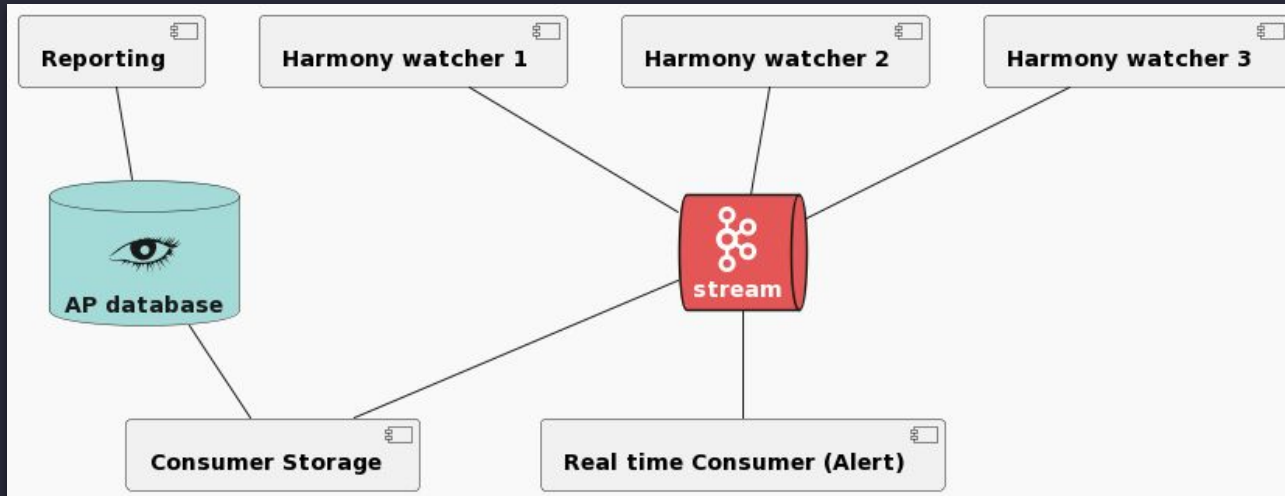
Harmony State Project : Part II

Proof Of Concept

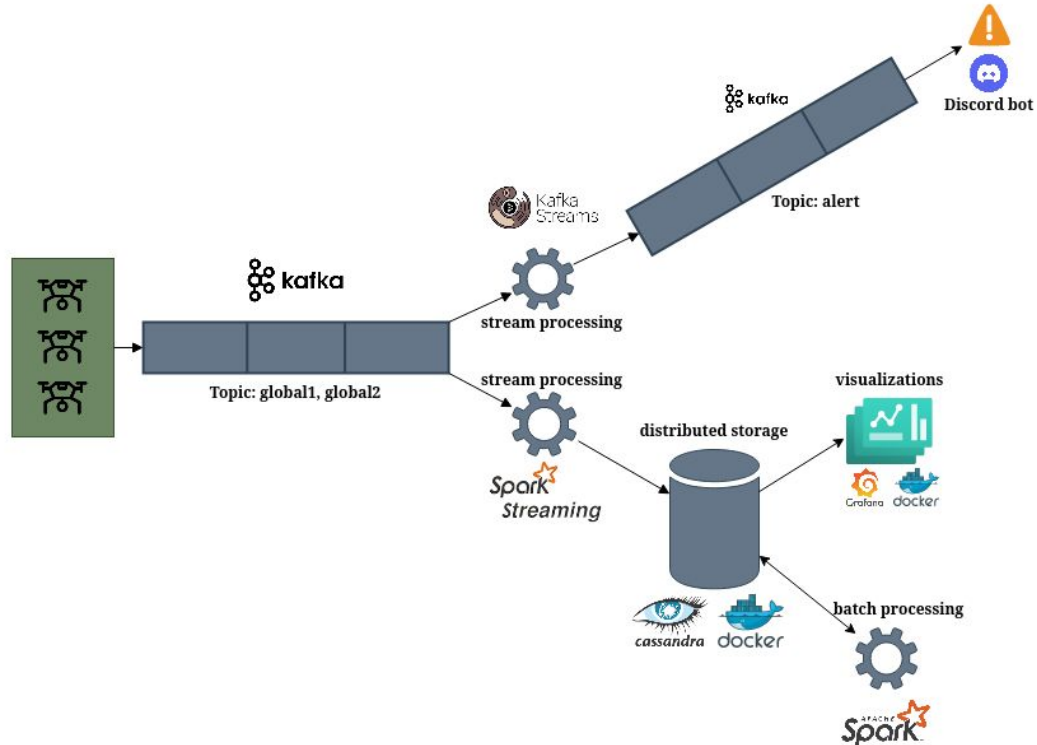
Bastien Pouëssel, Antoine Boutin, Tom Genlis



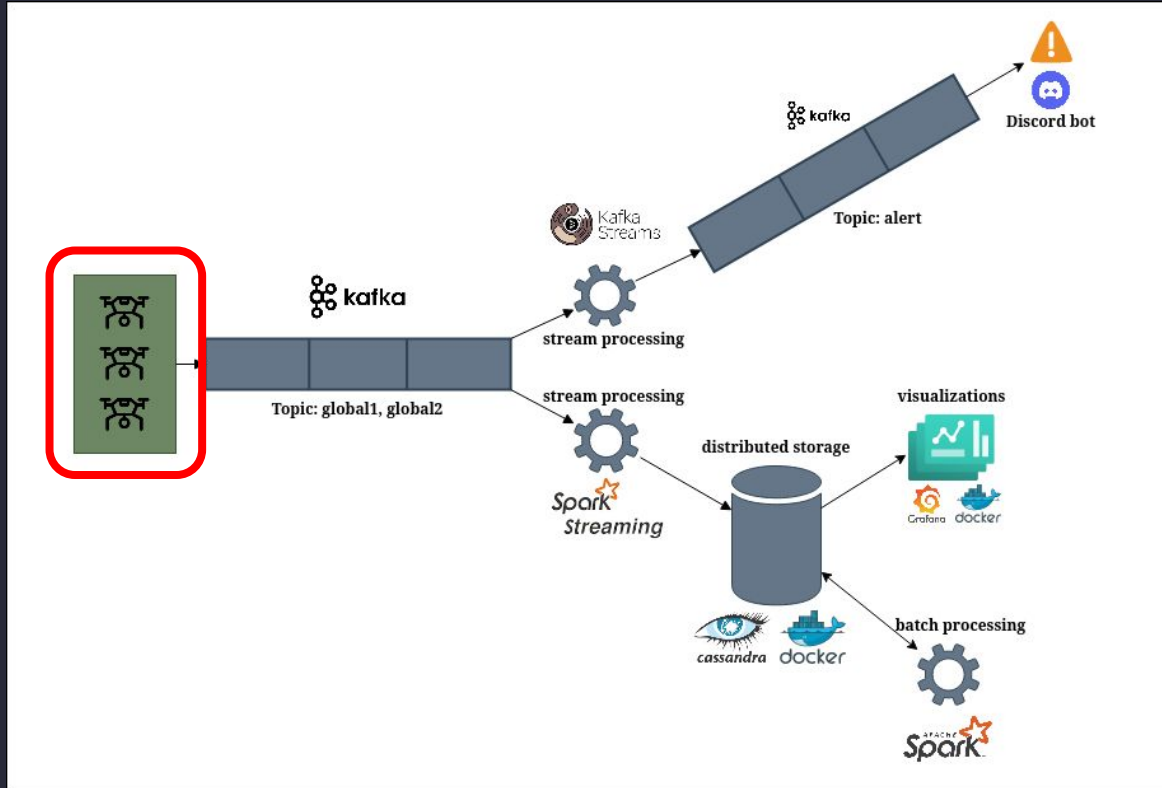
First Architecture



Architecture

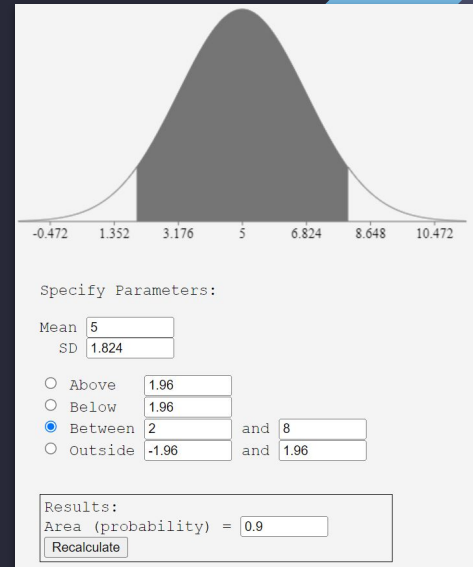


Architecture Focus : Producers



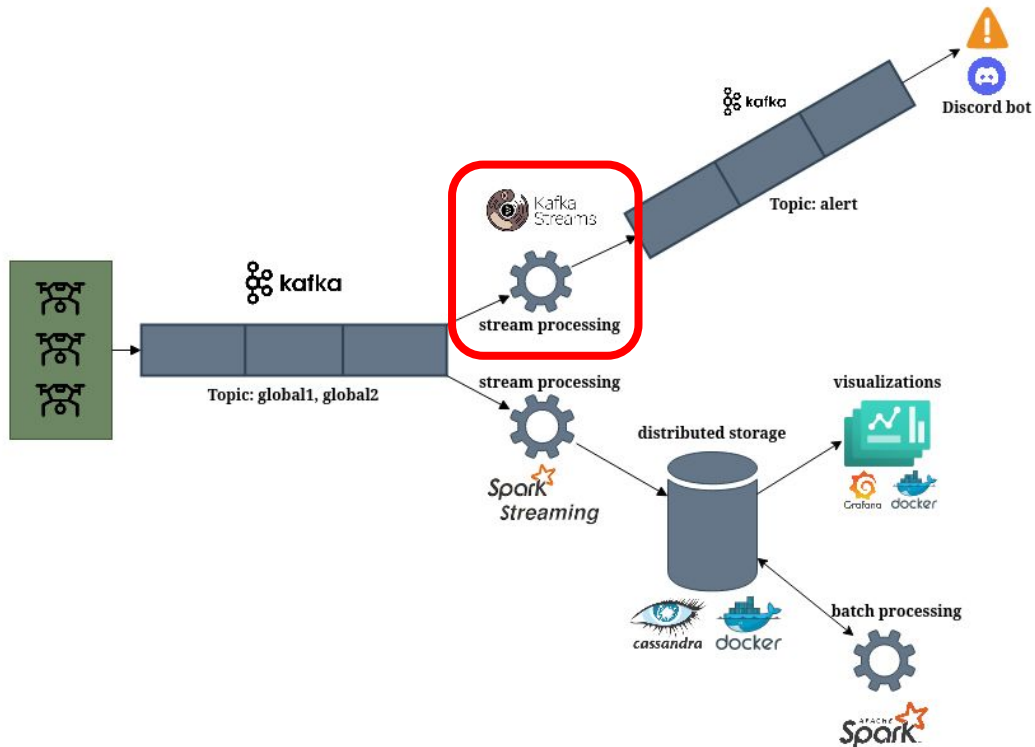
Drone Producer

- Produces one report per drone per minute (**100 / min**)
- **Reports** contain :
 - Drone ID
 - Drone Coordinates
 - Persons
 - Firstname, Name
 - From a database
 - Harmony Scores (0 - 10)
 - Words Used
 - From a database of sentences
 - Date
- **5%** of reports contain **Alerts**



	A	B
1		text happiness
2	I love my community and feel a sense of belonging.	10
3	The government is doing a great job.	9
4	I feel so happy today.	10
5	Life has been tough lately.	5
6	I am angry with the recent policies.	2
7	I am upset and thinking of protesting.	1
8	I have never been happier.	10
9	I am indifferent to the current state of affairs.	5
10	I feel a sense of hopelessness.	1
11	The recent events have left me feeling joyful.	8
12	I am concerned about the direction of our society.	3
13	The government needs to be held accountable.	0
14	I am dissatisfied with the status quo.	2
15	I believe our leaders are corrupt.	0
16	I am thinking of joining the rebellion.	0
17	Life has never been better.	10
18	I have complete trust in our government.	9
19	I am feeling unhappy with the current leadership.	2
20	I will not stand for this injustice.	0

Architecture Focus : Kafka Streams



Kafka Stream

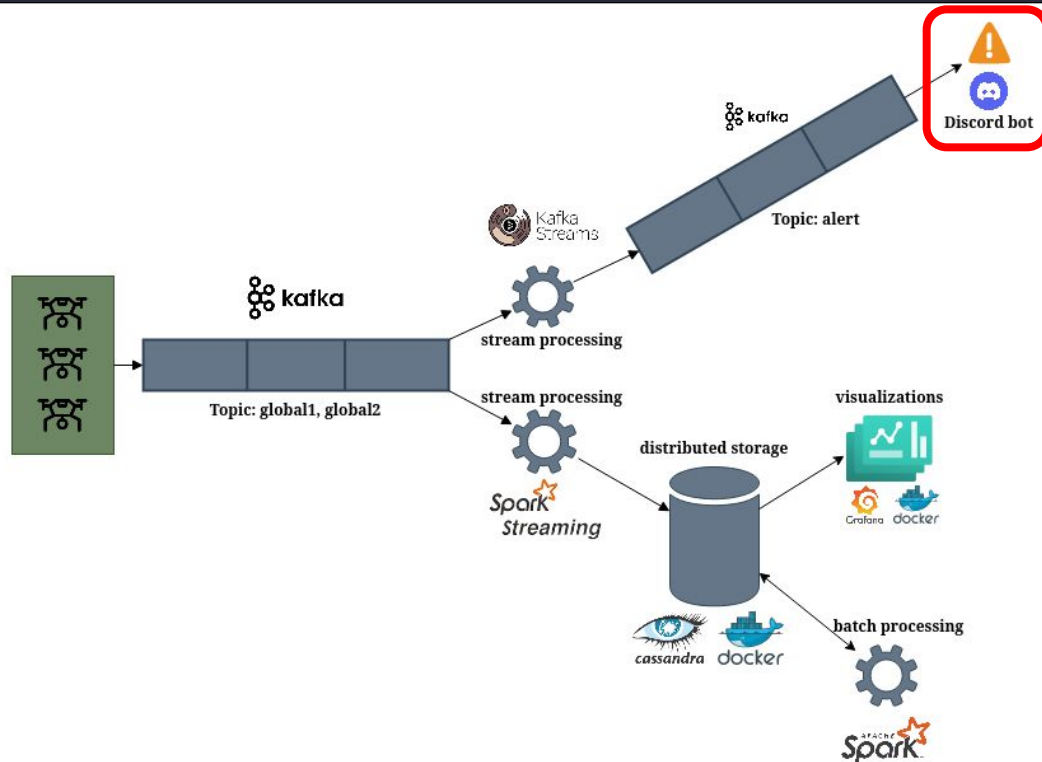
- Stream processing between to topics
- Data transformation → Filter harmony score < 2
- Data goes from topics global1 → alert

Why Kafka streams ?

- Input and output data are stored in Kafka clusters
- Ease of use
- High Level API





Architecture Focus : Alert Consumer




Alert Message


- **Functional Scala** only **Discord**  Bot using **AckCord** 
- Consumes **Alert Kafka Topic**, parses **JSON** data and sends **messages** on **Discord**

 **Harmony Bot** BOT Aujourd'hui à 16:37


 Online

 **Alert !**

- ◆ Drone : 7
- ◆ Position : -47.61786910283755, 90.12430854095561
- ◆ Sad persons : Gwendolyn Doyle (0)
- ◆ In contact with : Grayson Schneider (5)
- ◆ Key words : losing, direction, hurdles
- ◆ Date : Thu Jun 29 16:38:17 CEST 2023

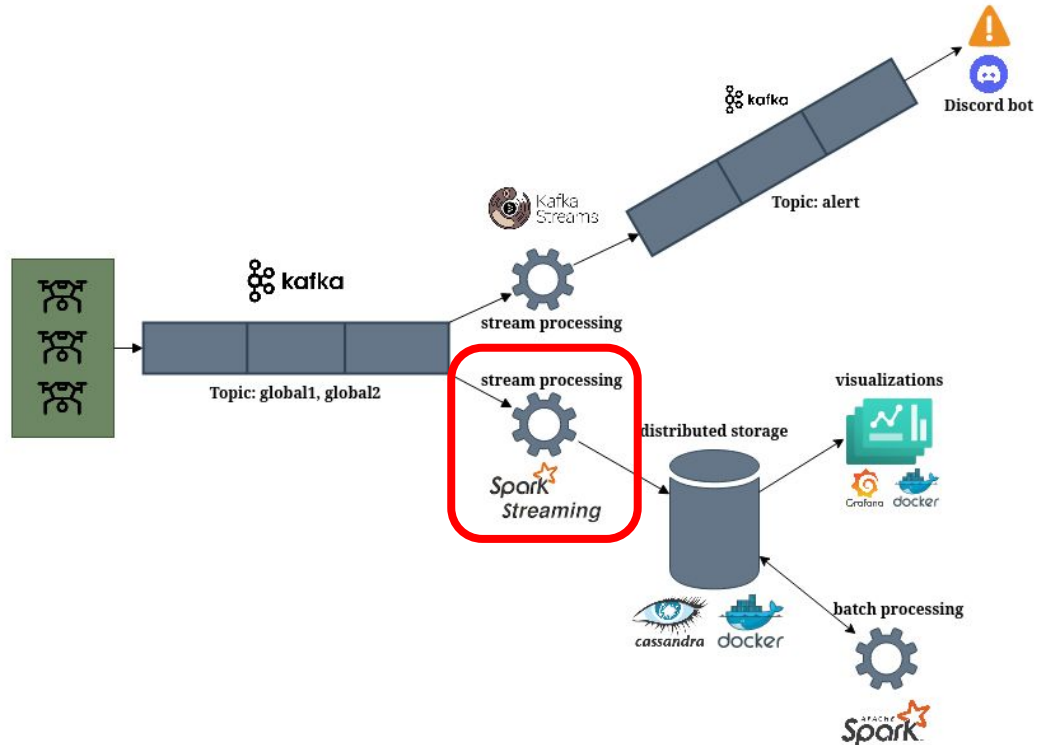
 **Alert !**

- ◆ Drone : 1
- ◆ Position : 61.29676852338048, 36.240194910746
- ◆ Sad persons : Thatcher Gonzalez (1)
- ◆ In contact with : Olivia McLaughlin (6)
- ◆ Key words : upset, hopelessness, ahead, left, job
- ◆ Date : Thu Jun 29 16:38:17 CEST 2023

 **Alert !**

- ◆ Drone : 0
- ◆ Position : 52.06283306128017, 50.372576118665165
- ◆ Sad persons : Emily Mathis (0), Ryker Reeves (1)
- ◆ In contact with : Madison Abbott (3)
- ◆ Key words : divide, feels, becoming, injustice, protest, status, recent, situation, happenings, happenings
- ◆ Date : Thu Jun 29 17:57:17 CEST 2023

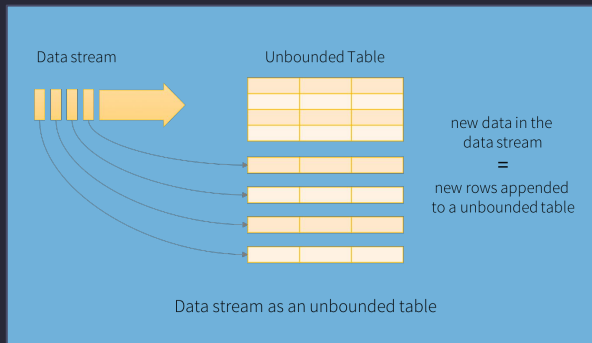
Architecture Focus : Spark Stream Processing



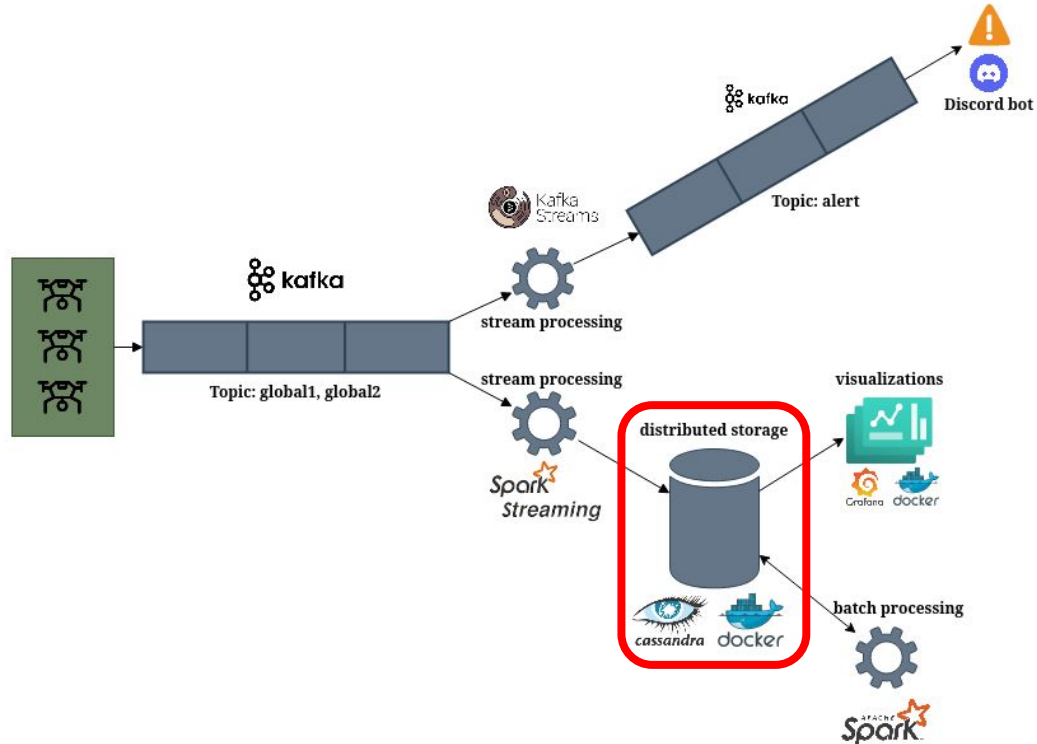
Message Storage



- Stream processing between Kafka global2 topic and Cassandra
- Usage of Spark cassandra connector (Datastax)
- Micro-batch processing
- Structured Streaming
- Dataframe Spark API ← Manipulation of Json
- Possibility of data transformation (aggregation)



Architecture Focus : Database



Message Storage

Apache Cassandra: Distributed Data Storage



- **Distributed Data Model:** Horizontal data distribution across a cluster of nodes.
- **Columnar Model:** Data organization in grouped columns, allowing schema flexibility.
- **High Availability and Optimized Performance:** Data replication for continuous availability and optimized write operations.

Message Storage

Tables :

- Report

```
droneid | time | latitude | longitude | persons | words
-----+-----+-----+-----+-----+-----
```

(0 rows)

- HarmonyScores

```
time | harmonyscore | id
-----+-----+-----
```

(0 rows)

- Persons

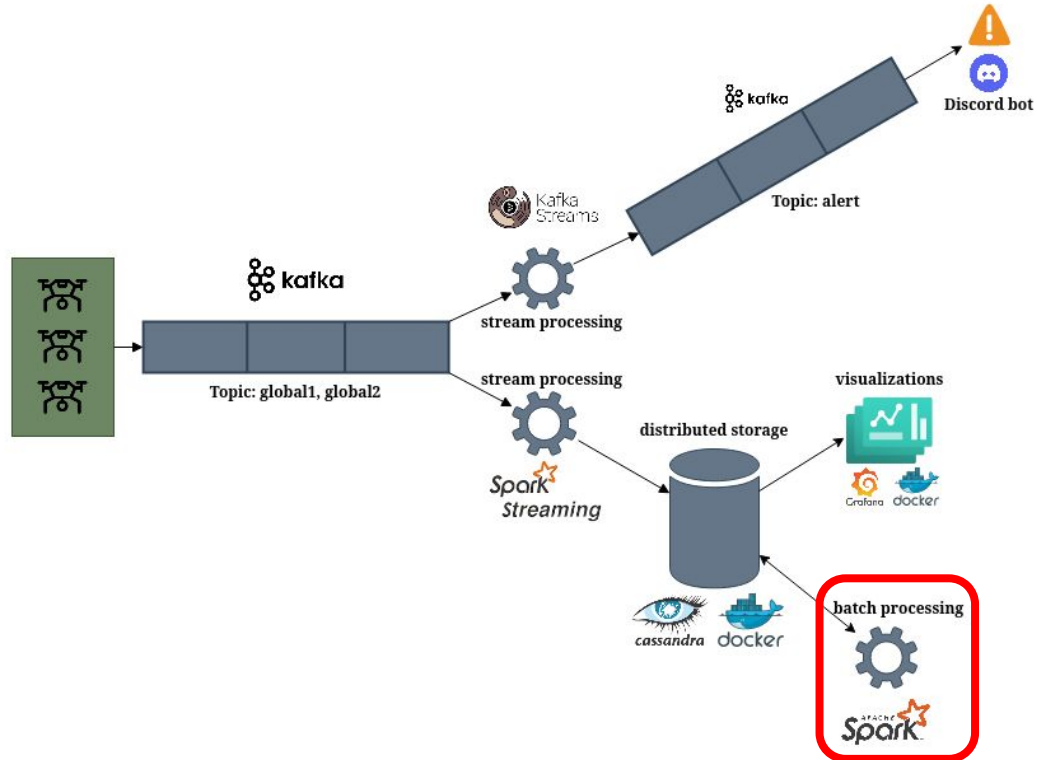
```
time | count | id
-----+-----+-----
```

(0 rows)

- BadPersons

```
time | count | id
-----+-----+-----
```

Architecture Focus : Analysis



Analyse stored message

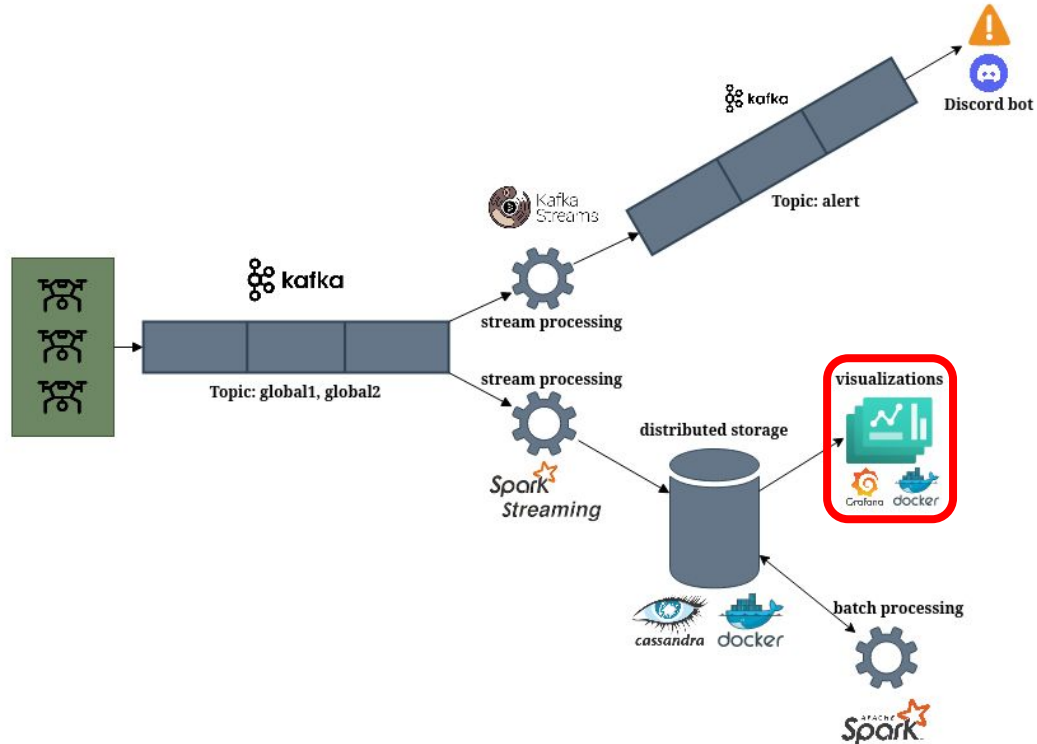
- Batch processing
- DataFrame API ← consistency with Spark Streaming
- Transformation of large amount of Data
- Statistics stored in Cassandra
- Trigger by action show or save into Cassandra

Why Spark batch processing ?

- Scalability and Distributed (Handling large amount of Data)
- Speed due to parallelization



Architecture Focus : Data Visualization



Data Visualization

- Which 10 people have the most frequently low harmony scores?
- How does the average harmony score evolve over time?
- How does the average number of supervised people evolve over time?
- How does the average number of people with a harmony score < 2 evolve over time?

Grafana time series visualization





That's all folks !