

# Distributed system project



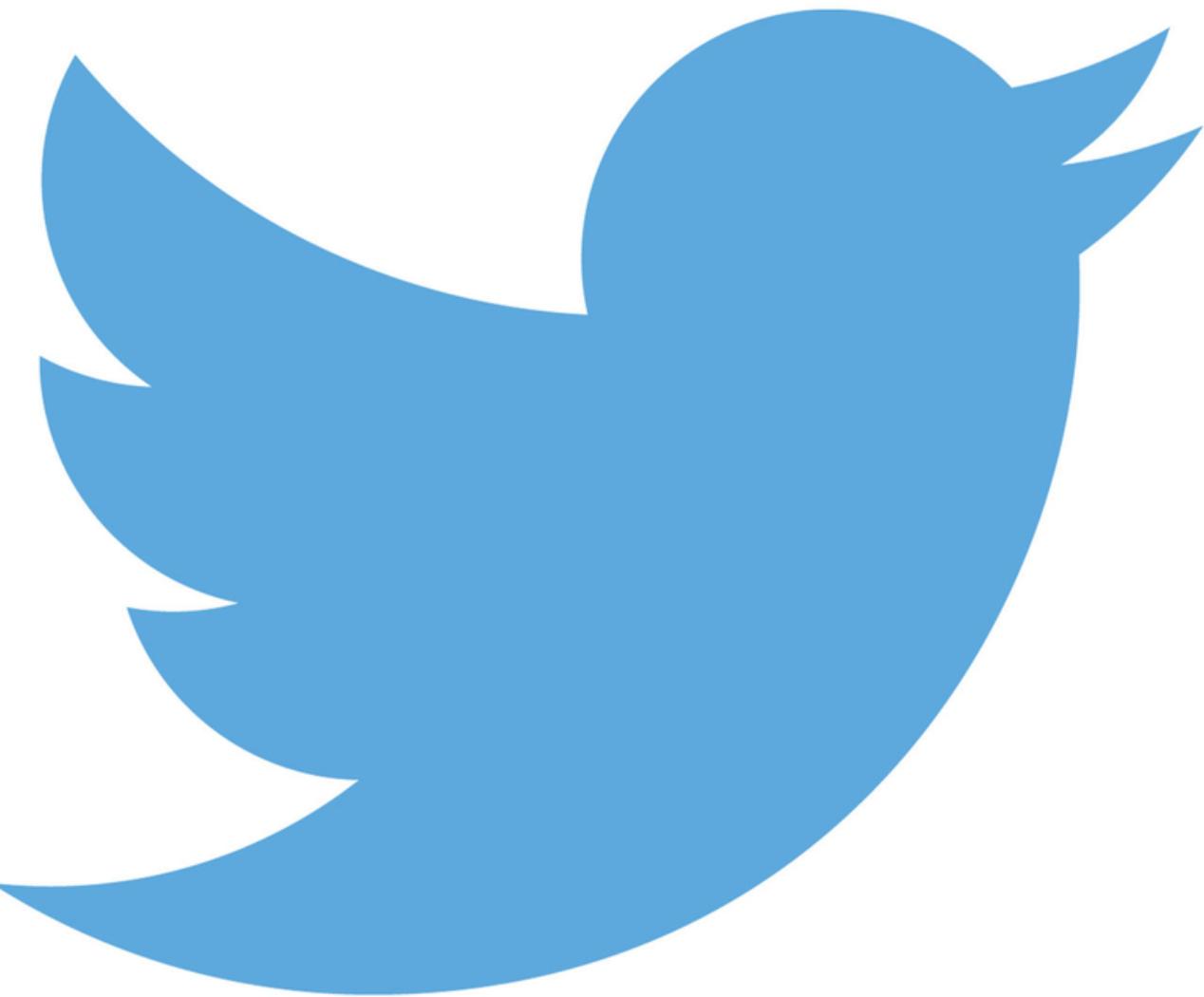
Satisfaction research from World Cup 2022 in Qatar



# Data collection

**Get our data from Twitter**

Fetch data from Tweets linked to a specific topic. Here, the topic will be the World Cup 2022. Once set, we'll analyse those data in the next process.



# Satisfaction study

## Analyze Tweets

The goal is to define the sentiment behind a tweet. Whether it is positive, negative or neutral. In the end we want to define the public's average opinion about the World Cup 2022 in Qatar.



# Results

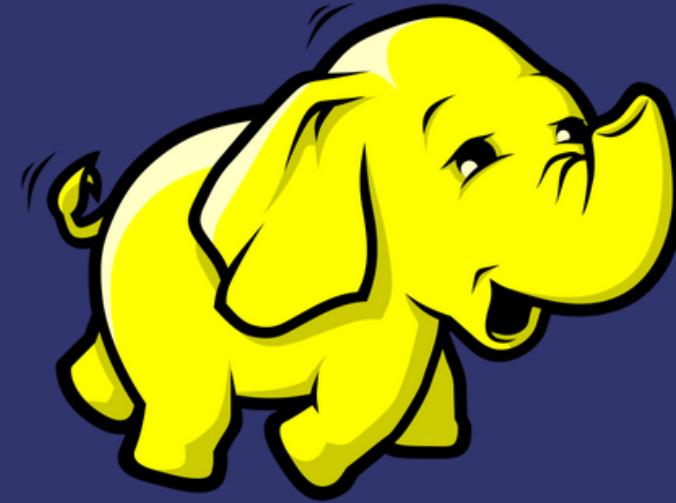
Overall view on satisfaction

## World Heatmap

With the data collected we will assert a heatmap representing every country's average opinion on the World Cup 2022 in Qatar.



# Technologies (Data)



**Hadoop**

Opensource  
framework

*Manage large amount  
of data on distributed  
systems*



**Flume**

Opensource  
framework

*Analyze data in real-  
time process*



**Oozie**

*Tool for Hadoop  
Job scheduler for  
Hadoop tasks*



**Hive**

*Large amount data  
storage*

*Store data in SQL*

# Technologies (Analysis)



**Java**  
Programming  
language



**Microsoft Azure's**  
**Cognitive Service**  
**for Language**

*A collection of features to  
analyze text within  
documents / tweets*

# Exemples

Twitter Sentiment Analysis in real-Time (monkeylearn.com)

Test with your own text

Qatar don't sell beer anymore for World Cup 2022. It's a shame...



Classify Text

Results

TAG	CONFIDENCE
Negative	82.0%

Test with your own text

Qatar invested tons of money in World Cup 2022. They are so rich!



Classify Text

Results

TAG	CONFIDENCE
Positive	73.6%

# Thank you

Questions?