



# 2017 Vacation Trends Analysis

## #vacation

Tracking top trending vacation destinations on Twitter using stream processing

Chandan Gope  
Daniel Lee  
KC Tobin



---

# Introduction

Idea: Analyze vacation related tweets to better understand and predict trends in travel destinations.

Approach: Utilize stream processing technologies to collect and analyze tweets in real-time.



---

# Problem statement

- Travel industry becoming increasingly competitive
- Need to be nimble
- Rapidly changing preferences



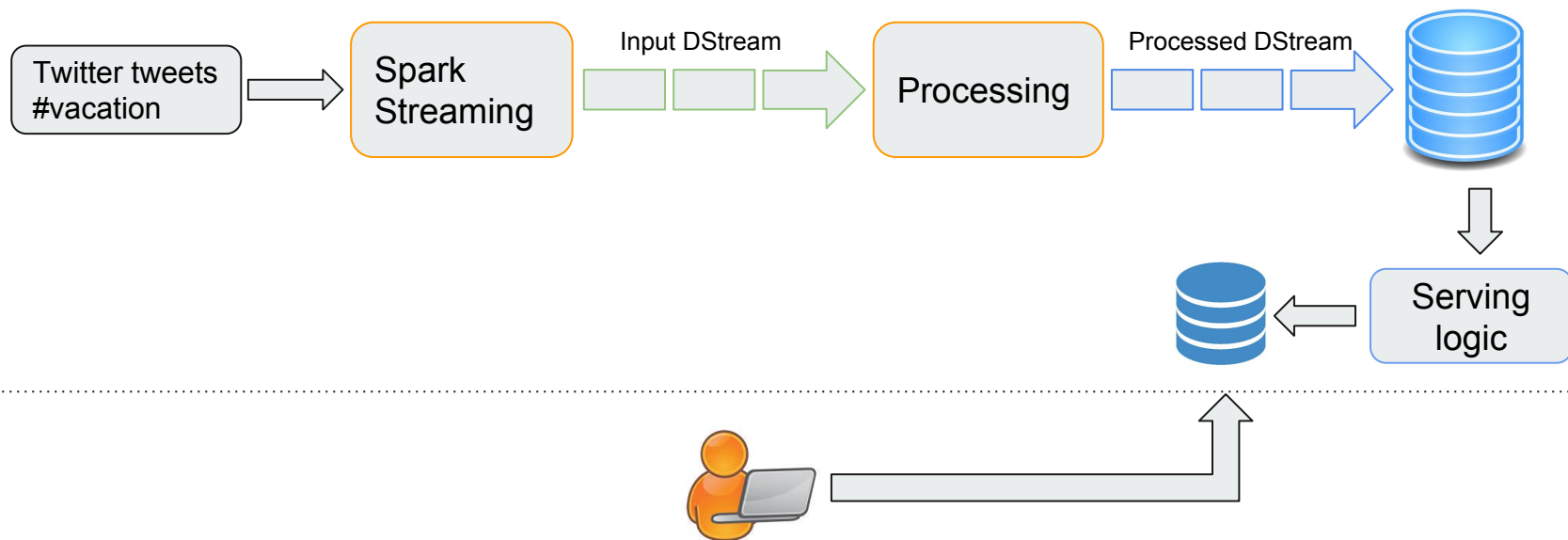


## Steel thread

- Gather vacation reviews/information
- Identify locations
- Store information
- Process information
- Visualize results of top destinations

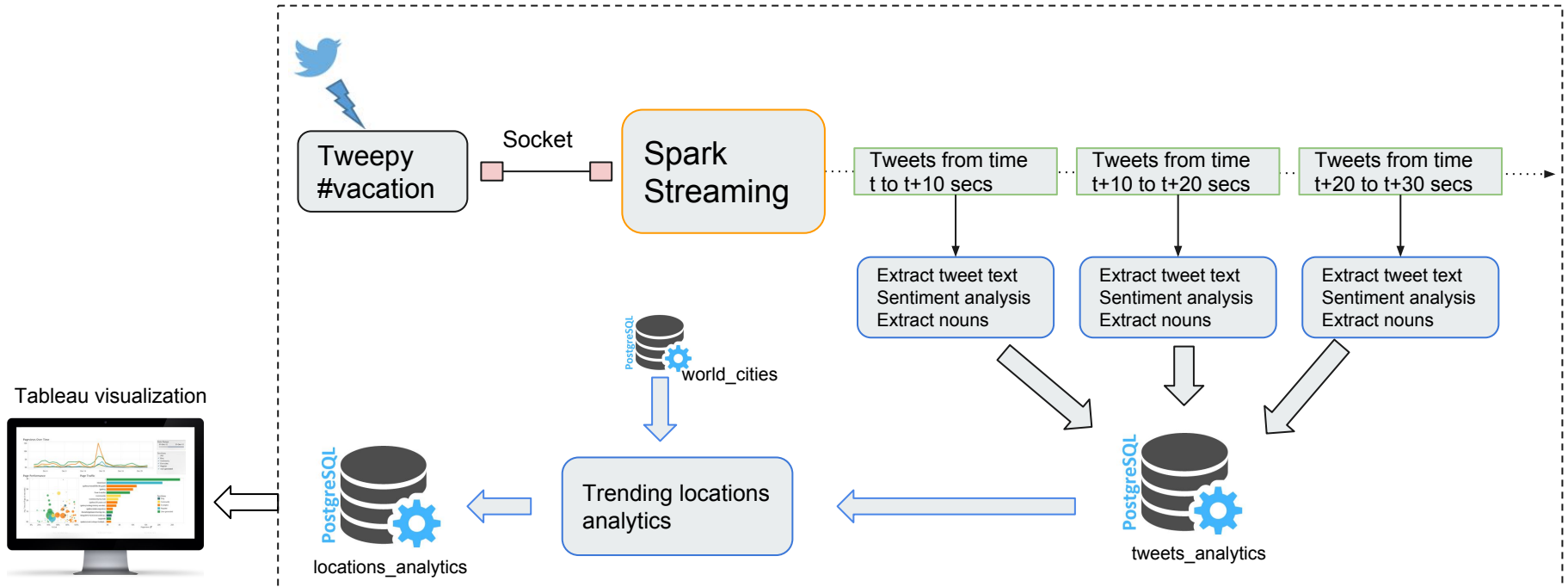
# System design

Cloud



# Software architecture

AWS Cloud





# Implementation

- Created an AWS *t2.medium* (2 vCPU, 4 GB RAM) instance
- Installed software - Spark, Postgres, Tweepy, psycopg2, nltk, git
- All code in python
- System running live non-stop for 7+ days
- Collected ~170K tweets (**V**olume)
- Tweets coming at the rate of ~ 3 tweets/10 secs (**V**elocity)



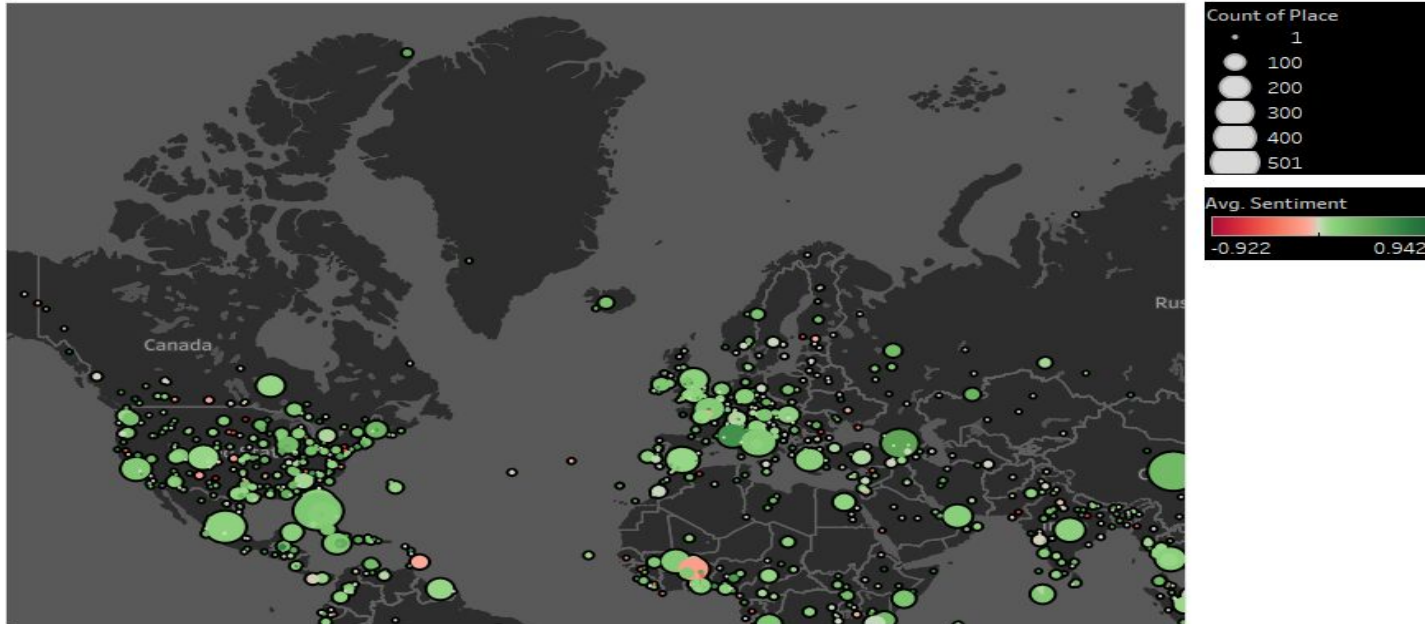
# Scale-out

- Switch from storing data in Postgres to Hive or Red Shift
- Incorporate other data sources, (e.g. TripAdvisor, other social media sites, etc.)
- Move the location extraction processing from batch processing to stream-processing
  - Takes about 0.5 seconds to find words matching locations against a table of 7300 world cities
  - Currently done serially, after items are added to the postgres table



# Live visualization

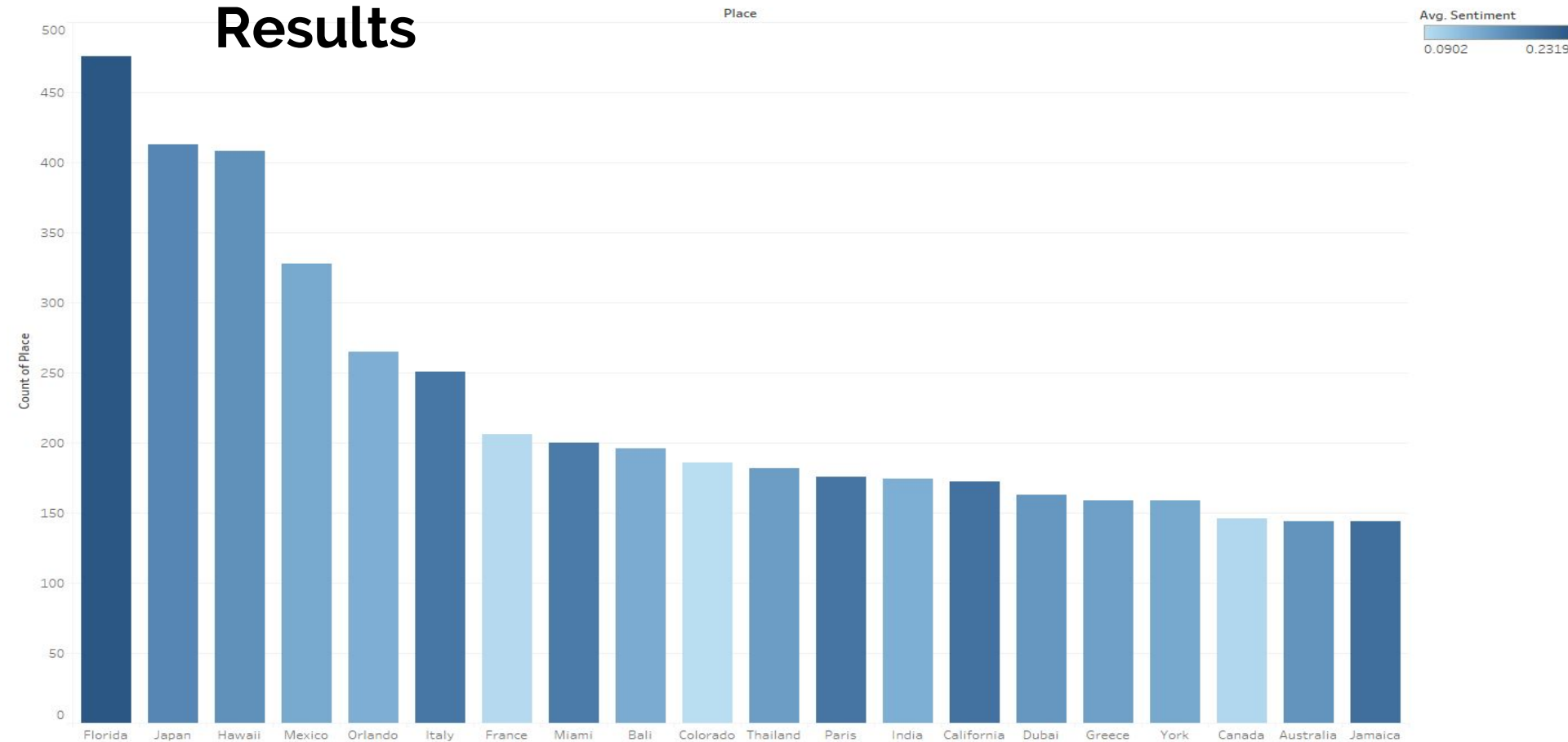
Symbol\_Map



Map based on Longitude and Latitude. Color shows average of Sentiment. Size shows count of Place. Details are shown for Place. The view is filtered on Exclusions (Latitude,Longitude,Place), which keeps 1,023 members.

Top20Count

# Results



Count of Place for each Place. Color shows average of Sentiment. The view is filtered on Place, which keeps 20 of 1,023 members.



# Future Development

- Personalized recommendations, creating user profiles, preferences, mentions, nearest neighbor analysis based off prior travel mentions
- Incorporate additional data sources, TripAdvisor, other social media sites
- Incorporate trending metrics over time
- More robust extraction of locations from tweets
  - “Nice vacation” => Not talking about the place Nice
- Really understand what the tweet means
  - “Back to Miami after a great vacation” => Does not imply Miami is a great vacation spot!
  - Use advanced NLP techniques



# Thank you.

Questions?