

Project Goals

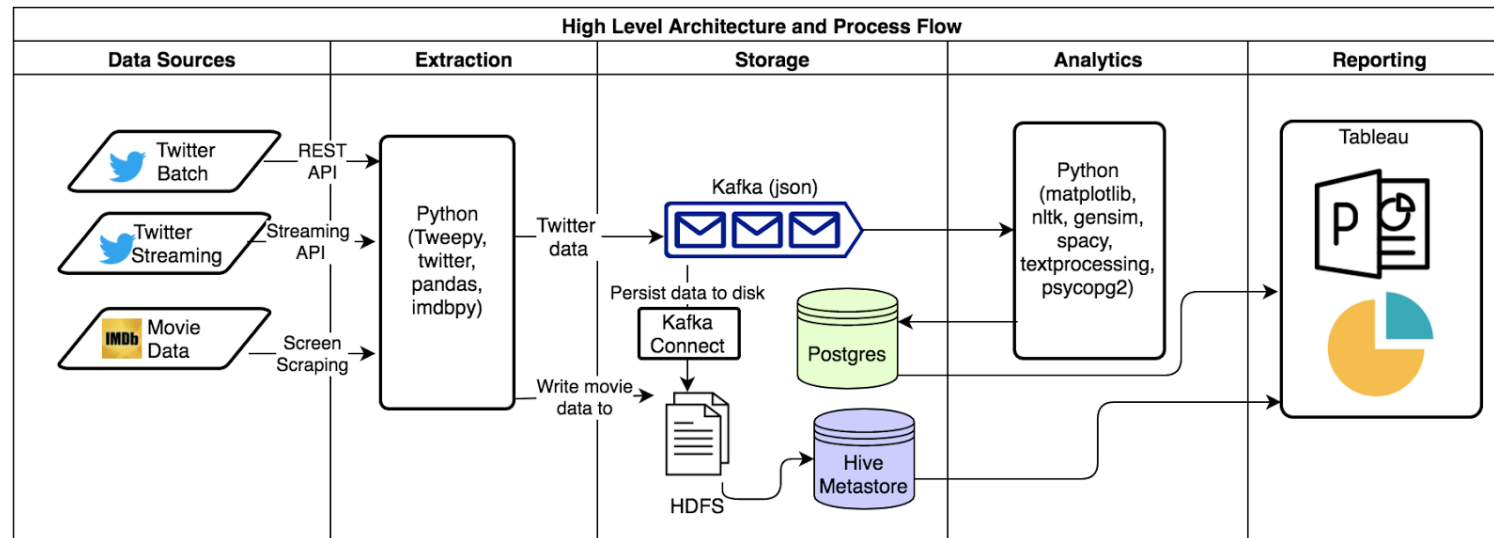
Build a big data storage, retrieval and analytic pipeline with Twitter data

- End to end Twitter data storage and retrieval process
- Demo various data sources: online restful, batch, and streaming
- Apply basic sentiment analysis on the data
- Scraping online data to enrich analytics and data serving
- Use tableau to visualized analytics



Architecture

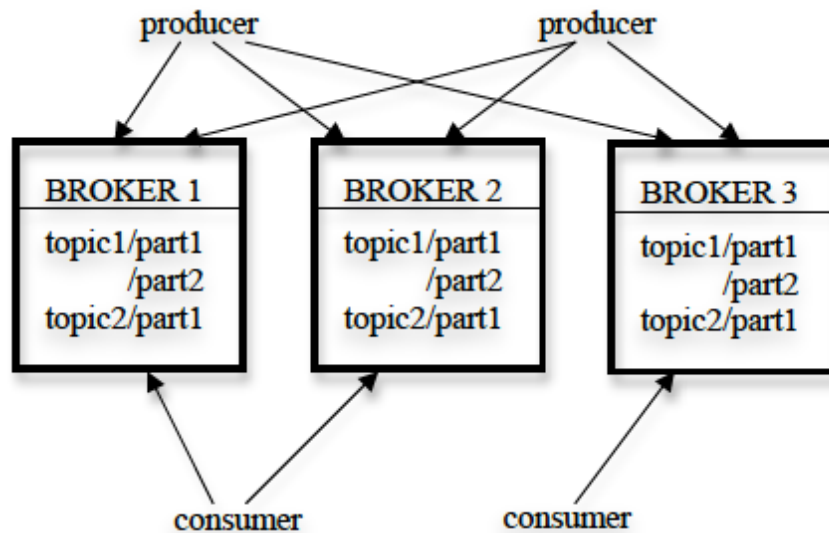
Use Kafka as ingest layer to interface Twitter restful API and streaming API; Store persist data to Hadoop HDFS. Use python to analysis streaming data and store analytic results in postgres for Tableau to consume and visualize



Implementation Notes

Why Kafka?

- Kafka is designed for scale-out distributed messaging system for log
- Kafka cluster can store large dataset for longer period
- Allow clients to read in different orders
- Internal zookeeper allow producers and consumers to work at different speeds
- Handle data bursts and fault tolerance



How is Kafka different from other real time data tools

- Zookeeper: order messages within a partition
- Scale-out: for distributed across machines
- Persistence: hold a lot of data
- Schema: do not apply schema, trade-off complexity with speed

Kafak Producer Example: A listener to tweet streams

Language Python

```
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
from kafka import KafkaProducer
from kafka.client import SimpleClient
from kafka.consumer import SimpleConsumer
from kafka.producer import SimpleProducer

client = SimpleClient("localhost:9092")
producer = SimpleProducer(client)
consumer_key = "cW3RTKoG5kiNkzfdbSb8aBMyY"
consumer_secret = "iwj5uOncngUMYk2BMNkF3WwL9VR7FXCPvXJVYwbDNDmuy8yRkH"
access_token = "4175914697-j5Ghb209PGZ0kobm0cnh9nZ2zMwrrigfVGczYbA"
access_token_secret = "0lK2XYfh1oksmpymqgTRBLrhR5nGLMUr84N0yhicWuUq2"

class StdOutListener(StreamListener):
    """ A listener handles tweets that are received from the stream.
    This is a basic listener that just prints received tweets to stdout.
    """
    def on_data(self, data):
        producer.send_messages('movies', str(data))
        print(data)
        return True

    def on_error(self, status):
        print(status)

if __name__ == '__main__':
    l = StdOutListener()
    auth = OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_token_secret)
    stream = Stream(auth, l)
```

```
stream.filter(track=['#moana'],languages=["en"])
```

Kafak Consumer Example: A client consumes records from the kafka cluster

Language Python

```
from kafka import KafkaConsumer
import pandas as pd
import json
import pandas as pd
import nltk
import spacy
import textprocessing
import re
from nltk.sentiment.vader import SentimentIntensityAnalyzer as Vader

def main():
    '''
    main function initiates a kafka consumer, initialize the tweetd
    Consumer consumes tweets from producer extracts features, clean
    calculates sentiments and loads the data into postgres database
    '''
    # set-up a Kafka consumer
    consumer = KafkaConsumer('movies')
    tweets,conn, dbcur = initialize(db_name = "tweetdata")
    for msg in consumer:
        output = []
        output.append(json.loads(msg.value))
        print output
```

```
print '\n'

# Function to extract features from tweets
extracttweetfeatures(tweets, output)

# Text cleaning
cleantweettext(tweets)

# Calculate sentiment using nltk vader sentiment library
calculatesentiments(tweets)

# Create data frame
cleanse_dataframe_and_load(tweets, conn, dbcur)

if __name__ == "__main__":
    main()
```

Twitter DataFrame for Analytics

In [96]: tweets

Out[96]:

	text	lang	country	user_nm	coordinates	location	retweets_count	text_clean	sentiment_score
0	RT @finaldraftinc: #TheGirlOnTheTrain #screenw...	en	None	Ian Jon McCoy	None	Atlanta, Georgia	3	rt erin cressida wilson take u behind scene hi...	0.1027
1	I don't understand the movie "The Girl on the ...	en	None	Bourlart Noemie	None	Lille, France	0	dont understand movie girl train take place ne...	0.0000
2	RT @BritishVogue: Have you seen #TheGirlOnTheT...	en	None	Aron Adimin	None	Liwutung, Sulawesi Utara	35	rt seen yet here s need know watching film	0.0000
3	RT @girlontrainfilm: #TheGirlOnTheTrain is num...	en	None	Aron Adimin	None	Liwutung, Sulawesi Utara	62	rt number film uk seen yet	0.0772
4	#thegirlonthetrain #rainydayread #thankyouomeg...	en	None	Alle Davis	None	New Plymouth District, New Zea	0	bekahs day b made	0.0000
5	Just saw #TheGirlOnTheTrain Emily Blunt's perf...	en	None	Heather K.	None	North Carolina, USA	0	saw emily blunts performance amazing	0.5859
6	RT @MissCharleyF: Wow! @PaulaHWrites' #TheGirl...	en	None	craig a ball	None		1	rt wow outstanding thriller full twist turn su...	0.8481
7	Finally saw #TheGirlOnTheTrain @TGOTTFilm @emi...	en	None	Carrie Tyrie	None	Scotland, United Kingdom	0	finally saw amazing imagined	0.5859
8	yay almost done w reading my book! took advant...	en	None	Saritah	None	AZ	0	yay almost done w reading book took advantage ...	0.6597
9	Loved, loved the book: #thegirlonthetrain Jus...	en	None	Sheila Tidwell	None	Nashville, TN	0	loved loved book saw movie and	0.8316
10	#TheGirlOnTheTrain is the best movie I've ever...	en	None	Ky Hoffman	None		0	best movie ive ever seen sure	0.7579

Tableau Analytics Dashbroad

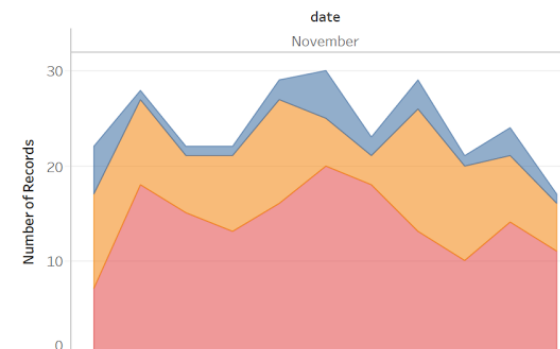
Movie
Moana

Movie description

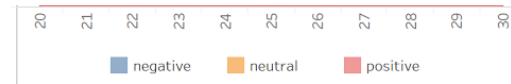
Director: Ron Clements
Cast: Auli'i Cravalho, Dwayne Johnson, Rachel House, Temuera Morrison
Genre: Animation, Adventure, Comedy, Family, Fantasy, Musical
Language: English
Summary: The main character will be Moana Waialiki, a sea voyaging enthusiast, and the only daughter of a chief in a long line of navigators. When her family needs her help, she sets off on an epic journey. The film will also include demi-gods and spirits taken from real mythology. Princess Moana is a daughter of the chief of her tribe. Coming from a long line of navigators she sets off for a fabled island with her hero, the demi God Maui. Along the voyage they battle the treacherous ocean and all which it hides, all the while learning what the power love between friends can accomplish. An adventurous teenager sails out on a daring mission to save her people. During her journey, Moana meets the once-mighty demigod Maui, who guides her in her quest to become a master way-finder. Together they sail across the open ocean on an action-packed voyage, encountering enormous monsters and impossible odds. Along the way, Moana fulfills the ancient quest of her ancestors and discovers the one thing she always sought: her own identity. A young woman

Month
November

tweets by sentiment



uses her navigational talents to set sail for a fabled island. Joining her on the adventure is her hero, the legendary demi-god Maui.



Top positive tweets

- 1: Moana is so brave, smart, strong, crafty, athletic, heroic, incredible, amazing, never been done before. I'm not exaggerating. #Moana
- 2: When Te Fiti was revealed I was like "WOW. OH WOW. OH MY GOD! WOW! OH... OH... OH MY GOD!" #Moana
- 3: RT @Simszter: Ended up watching #Moana again lol best choice! It motivated me again to be awesome. "It's who I am. Who I truly am" :)@TheRŠ—_
- 4: Just finished watching #moana, it was a beautiful movie, amazing story and just awesome. GreatŠ—_ <https://t.co/gFVmBn1oyo>
- 5: Just saw #Moana with my 3 year old sister and we loved it. Very beautiful and detailed_™... Great job and I loved @TheRock 's character!

Wordcloud



Project Repo