Meetup Data Analysis and Reporting

Balaji Rajaram
Data Science Graduate Student
Indiana University
Bloomington
Tel. +1-602-330-1316
brajaram@umail.iu.edu

Malabika Biswas
Data Science Graduate Student
Indiana University
Bloomington
Tel. +1-602-459-5752
mbiswas@umail.iu.edu

ABSTRACT

Meetup [4] is a social networking platform that enables people to create groups of common interest and organize offline events for effective communication and sharing of ideas. The offline events happen in various localities in the world and there are about 24 categories in the meetup such as Learning, Photography, Tech, etc. Meetup has exposed certain data using RESTful APIs [5]. We are going to make use of these APIs for our analysis and visualization of meetup data.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

General Terms

Documentation

Keywords

Meetup data analysis, Big Data project, Apache Kafka [2], Apache Spark [3], Apache HBase [1], python [7], MySQL [6], Tableau [8], text analysis.

1. INTRODUCTION

The objective of the project is to analyze and visualize the meetup data such as most active cities in the meetup, to find out the level of interest that people are showing in groups that are related to entrepreneurs, and to find out trending events and topics in technology specifically related to Big Data, Machine Learning, and Artificial Intelligence. These analyses may help us to get more information about the technologies that people are interested in and also it can help to find out the locations where people are more interested in startups.

2. DATA ANALYSIS

This section details about various analysis and visualization that will be performed on the meetup data. The first section discusses the types of analyses that will be performed on meetup data (§2.1), the implementation details (§2.2), and the artifacts to be created (§2.3). Finally, we discuss about the timelines of the project (§2.4).

2.1 Types of analyses

We are focusing on three kinds of analysis to understand the trend of popular meetups in various locations. Temporal analysis to understand the type of meetups which are popular at what time of the year, text analysis to analyze the title and description of the event to infer the core topics and finally as part of geolocation analysis, we want to cover the most popular events by cities based on the RSVP data, number of technical meetups in the city, and number of startup meetups by cities. The plan is to determine popularity or demand of an event based on the capacity of the event and number of waitlists. We also want to analyze no-show data as well. Using text analytics, we want to do a deep dive into few categories of technical meetups. We may also want to add additional metrics based on the data/time availability.

2.2 Implementation details

The end-to-end data pipeline will be built using Apache Kafka, Apache Spark, Apache HBase, MySQL, Python, and Tableau. The meetup data will be consumed via its RESTful API. Table 1 lists the APIs that will be used for this project. The host for the API call is api.meetup.com and it will be a GET method. The response from the API will be in JSON format, which will be processed in python.

API	Description
/2/cities	Returns Meetup cities. This method supports search by latitude/longitude/radius, by country/state, by query term/zip, or a combination of all of these.
/2/categories	Returns a list of Meetup group categories
/:urlname/events	Gets a listing of all Meetup Events hosted by a target group
/find/locations	Provides a query interface for finding known locations
/find/groups	Text, location, category, and friend-based group searches
/:urlname/events /:event_id/rsvps	Retrieves list of event RSVPs

Table 1: Meetup API List

A topic will be created in Apache Kafka to which the data retrieved from meetup will be published. The data from Kafka will be consumed via Spark streaming and pushed to Apache HBase. The data will then be analyzed and aggregated using Spark MLib, and SparkSQL. The aggregated data will then be pushed to MySQL.

Tableau will be used to visualize of the aggregated data. We will be using python for most of the programming needs and shell scripts will be used to automate certain processes. Figure 1 shows the high-level design of this project.
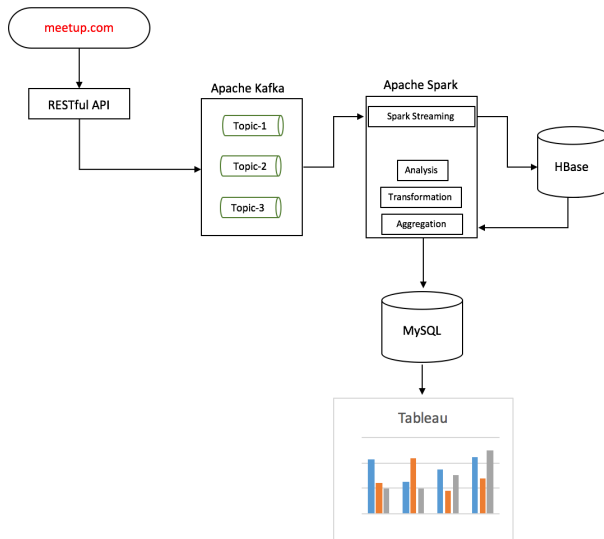
Figure 1: High-Level Design

2.3   Artifacts
As part of this project, we will be creating the following artifacts

1. Project proposal
2. Python scripts
3. Shell scripts
4. MySQL load scripts
5. Tableau reports
6. Project report

2.4   Team
We will be working as a 2-member team and the below are the project members

1. F16-DG-4058, brajaram, Rajaram, Balaji
2. F16-DG-4011, malabikab16, Biswas, Malabika

2.5   Timelines
We are planning to complete and deliver the project by November 30, 2016.

3.   REFERENCES

[1] Apache HBase. http://hbase.apache.org

[2] Apache Kafka. http://kafka.apache.org

[3] Apache Spark. http://spark.apache.org

[4] Meetup. https://www.meetup.com

[5] Meetup API. https://www.meetup.com/meetupapi

[6] MySQL. https://www.mysql.com

[7] Python. https://www.python.org

[8] Tableau. http://www.tableau.com