

Team Number:

8

Title:

Identification and Analysis of popular trends around the world.

Description:

The objective of this project is to develop a SaaS Web Application/API Framework which provides real-time analytics of current events using Twitter, Reddit*, and Meetup streaming data. This tool can be used to visually identify active social gatherings and help organize targeted activities with high audience participation.

Members:

Aditya Bhardwaj (ABHARDW2)
Gautam Verma (GVERMA)
Samir Jha (SJHA4)

Deliverables:

1. SaaS Web Application that displays current information on events like meetups/trends/news/forecast in real-time.
2. Application offers insights (Batch Analytics) of the data with respect to different time frames such as hour/day/month.
3. Application offers slice-dice operation as per the demographics specified by user.

Status

- One Amazon Kinesis Stream has been created to support three streaming APIs' viz. Reddit, Twitter and Meetups.
- Producer scripts to fetch and send data from all three streams have been completed with success.
- Architectural Model which includes the flow of data and business logic of the project has been created.
- Mock Scripts to store streaming data into AWS DynamoDB have been developed.

Issues

1. Kinesis Stream breaks frequently when the load of streaming data suddenly spikes. This helped in determining number of shards for the stream.
2. Restriction in calling in Twitter API calls per hour. How to compensate for this?
3. Which other databases can perform better in the project over default DynamoDB?

Note: * Reddit was not included in the initial claims.