Angel Figueroa Cruz Email: angel.figueroa.cruz@gmail.com

Project: Twitter Stream Collector and Reporting System (TWCR)

GIT :
TW_Collector : https://github.com/afigueroa14/tw_collector.git
TW_REST : https://github.com/afigueroa14/tw_rest.git

Summary

The TWCR System is designed base on distributed architecture (Figure 1). There are two applications that compose the TWCR system. The first application (Twitter Collector) collects the Twitter Data, and second application (Twitter Rest) allows the user to query for the information. Both applications are connected using the Apache Ignite, which allows data to pass between different nodes on the Application Cluster. The Twitter Collector uses the SPARK Big Data Engine to manage the high volume of Twitter Stream.

Application Architecture

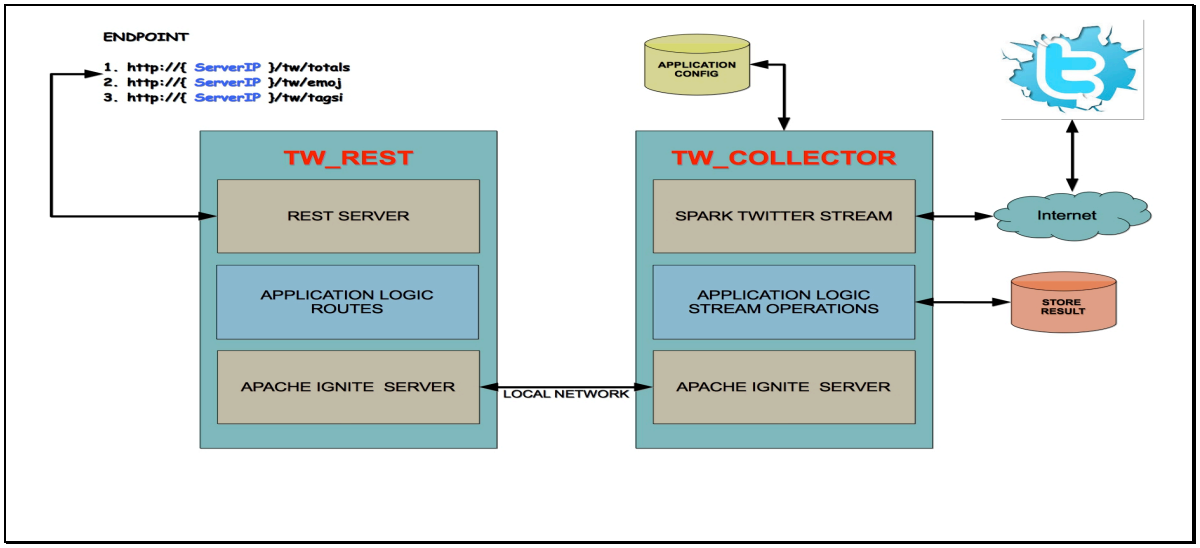

Figure 1. Twitter Architecture

TW-REST EndPoints that allows the user accesses the Information.

Requirements	Rest EndPoint
Version	http://{ ServerIP }/tw/version
Total Number of Tweets Received	http://{ ServerIP }/tw/totals
Average Tweets per hour/minute/second	http://{ ServerIP }/tw/totals
Top emojis in Tweets	http://{ ServerIP }/tw/emoji
Percent of Tweets that Contains Emojis	http://{ ServerIP }/tw/totals
Top HashTags	http://{ ServerIP }/tw/hashtags
Percent of Tweets that Contain a url	http://{ ServerIP }/tw/totals
Percent of Tweets that contain a photo url	http://{ ServerIP }/tw/photo
Top Domains of Url in Tweets	http://{ ServerIP }/tw/domaintop
URL on Tweets	http://{ ServerIP }/tw/urltop