

Building a Real Time Analytic Dashboard using BigData and Hadoop using Azure services

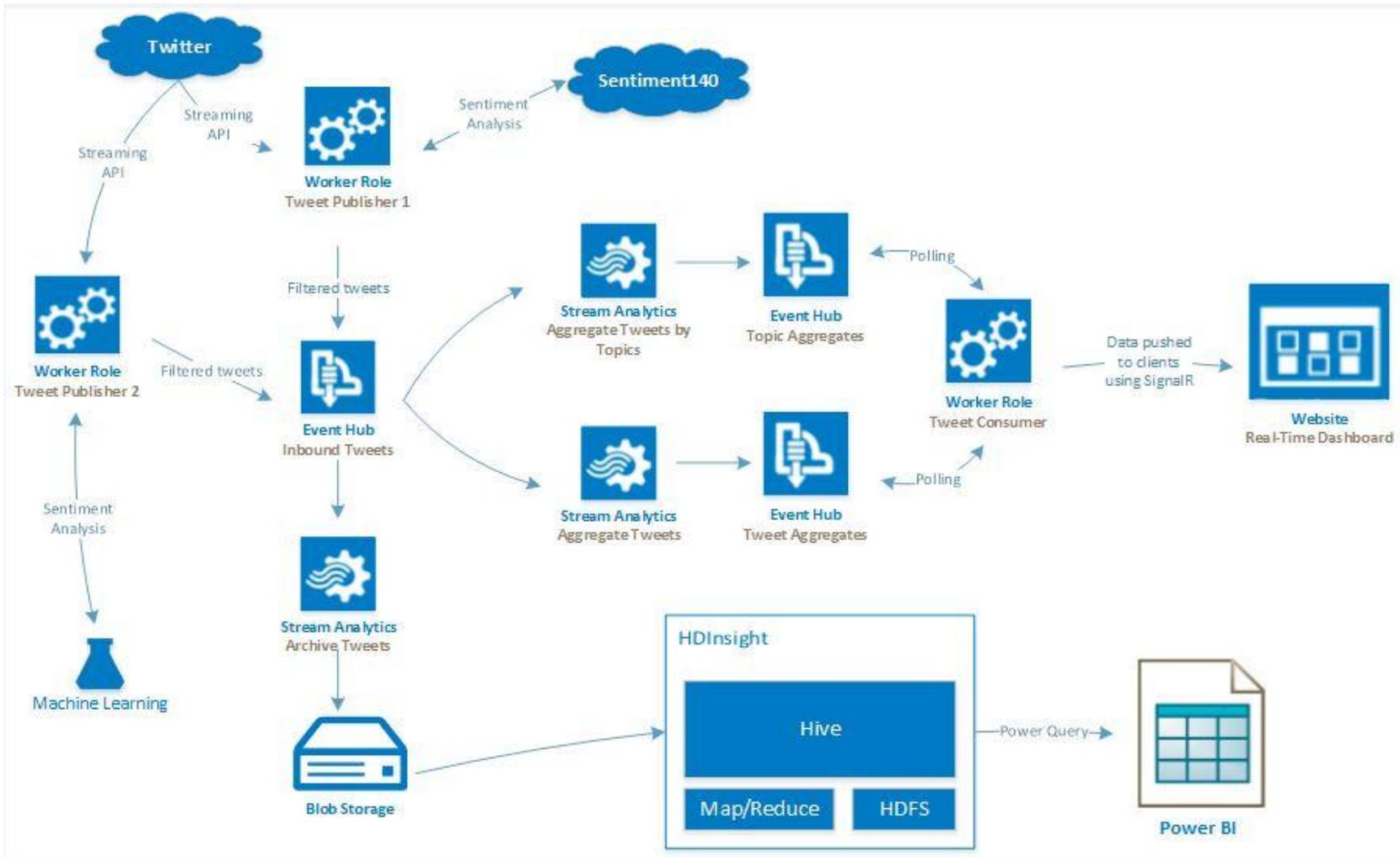


Debarchan Sarkar [@debarchans](https://twitter.com/debarchans)

Support Escalation Engineer, Microsoft



Logical Architecture



Analysis on Data at motion

- **Azure Worker Role**, *Tweet Publisher* connects to Twitter Streaming API with its own Twitter Application ID and a list of keywords to track.
- The worker role then calls an **Azure Machine Learning** web service to perform sentiment analysis on each tweet that it receives and assigns a sentiment score. It also does simple topic detection and then send the tweet to an **Azure Event Hub** called *tweetsin*

Analysis on Data at motion.....continued

There are three **Azure Stream Analytics** jobs monitoring the *tweetsin* event hub:

- The first job, *TweetsArchive* archives all the tweets to **Azure Blob Storage**.
- The second job, *TweetsByTopic* counts the number of tweets along with the average sentiment for *each topic* within the last 5 seconds. The aggregated results are saved to another **Azure Event Hub** called *TweetsOut2*
- The second job, *TweetsSummary* counts the number of tweets along with the average sentiment for all tweets within the last 5 seconds. It sends the result to an **Azure Event Hub**, call *TweetsOut1*.

Analysis on Data at motion.....continued

- There is another **Azure Worker Role**, *Tweet Consumer* that constantly polls the two event hubs, *TweetsOut1* and *TweetsOut2*, for any data. If there's anything in these event hubs, it picks it up, then send it on to the real-time dashboard using ****SignalR**
- The real-time dashboard is an **Azure Web Site** written in simple HTML5 page that connects to the SignalR host, *TweetConsumer* Worker Role each time it is loaded in a browser

****ASP.NET SignalR is a new library for ASP.NET developers that makes it incredibly simple to add real-time web functionality to your applications.**

Analysis on Data at rest

- On a daily basis, you can use an **Azure Automation Runbook** to spin up a new **HDInsight** cluster, execute a **Hive** script, then shut down the cluster
- There's an Excel workbook that uses **Power Query** to connect to **HDInsight**, bring the Hive tables into **Power Pivot**, and then visualize using **Power View**
- Finally, the Excel workbook is uploaded to **Power BI Preview** (on Azure) and then configured for data refresh. It can use various PowerBI services like **Q&A** to provide insights on the data with a few clicks and key presses

**Thank
You!**

saltmarch
MEDIA

GREAT INDIAN
**DEVELOPER
SUMMIT**

