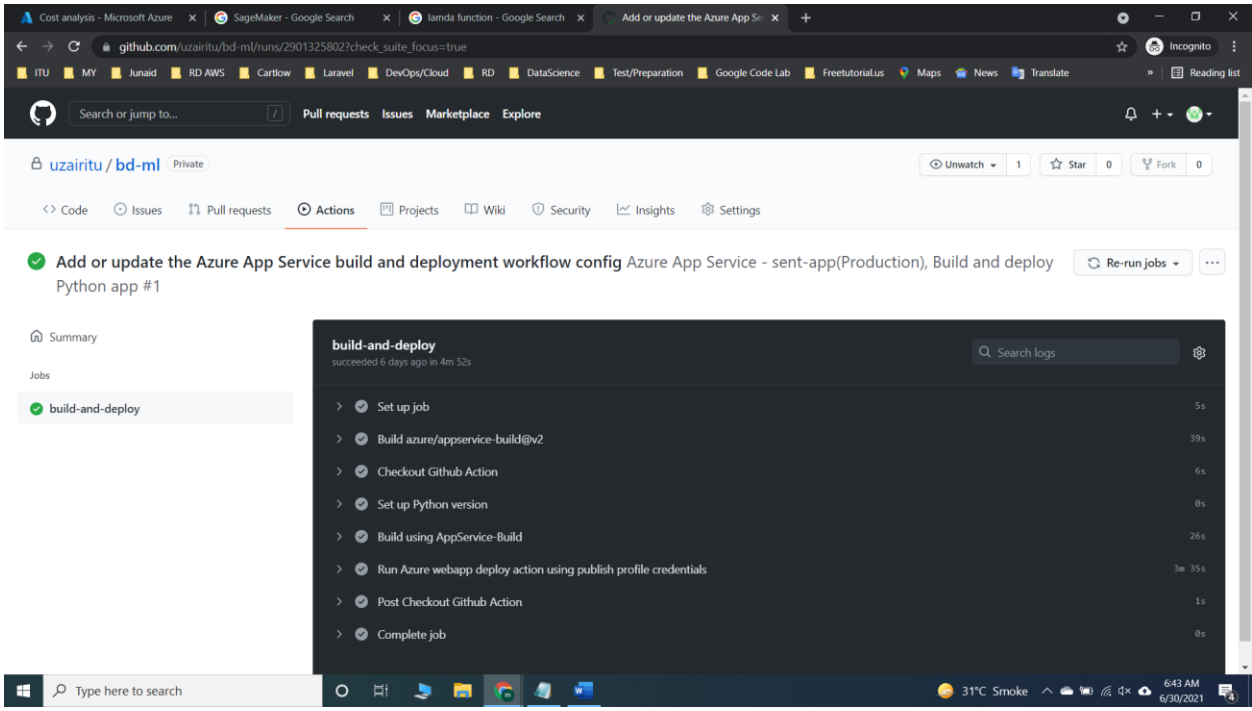
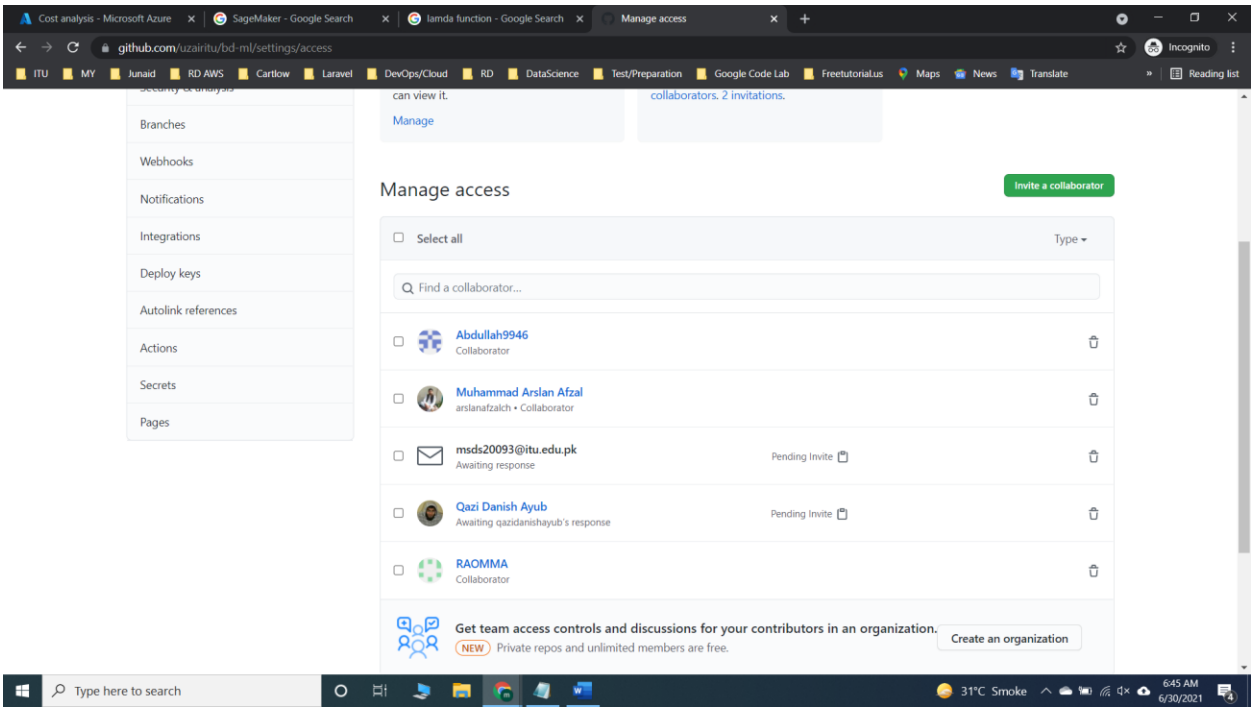
Arsalan , Abdullah, Ramesha ,Mubashir (Group 1)

Representor
Arsalan / Abdullah

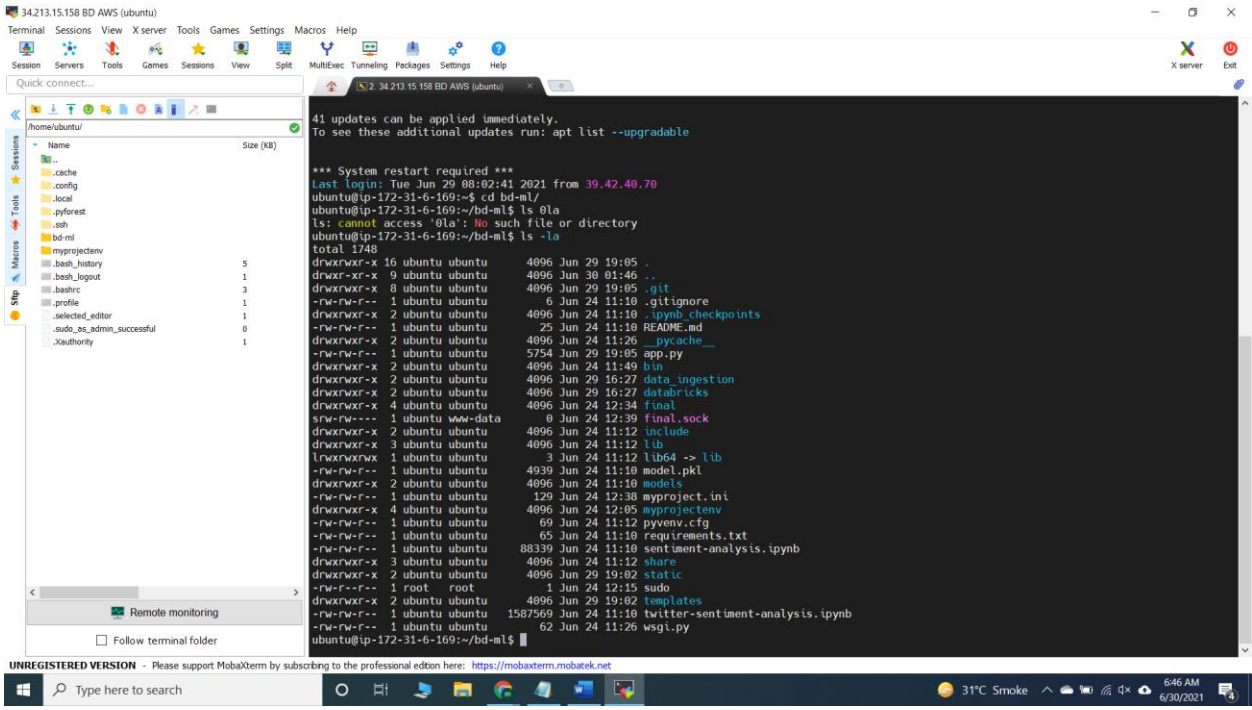
- How we trained the sentiment analysis model on data set of tweets
- Which techniques we used need 1 slide on results and data processing
- Why we not deploy on Azure Machince Learning?
    - we try to use that due to in experience we lost 1 account balance of our team member
    - we also want some server where we continue capture data from tweeter and pass forward to our pipeline. Azure is so expensive to use as Azure Web Solution
    - We use CI/CD deployment tools of Azure during experiments.

      Insights of CI/CD tool during learning

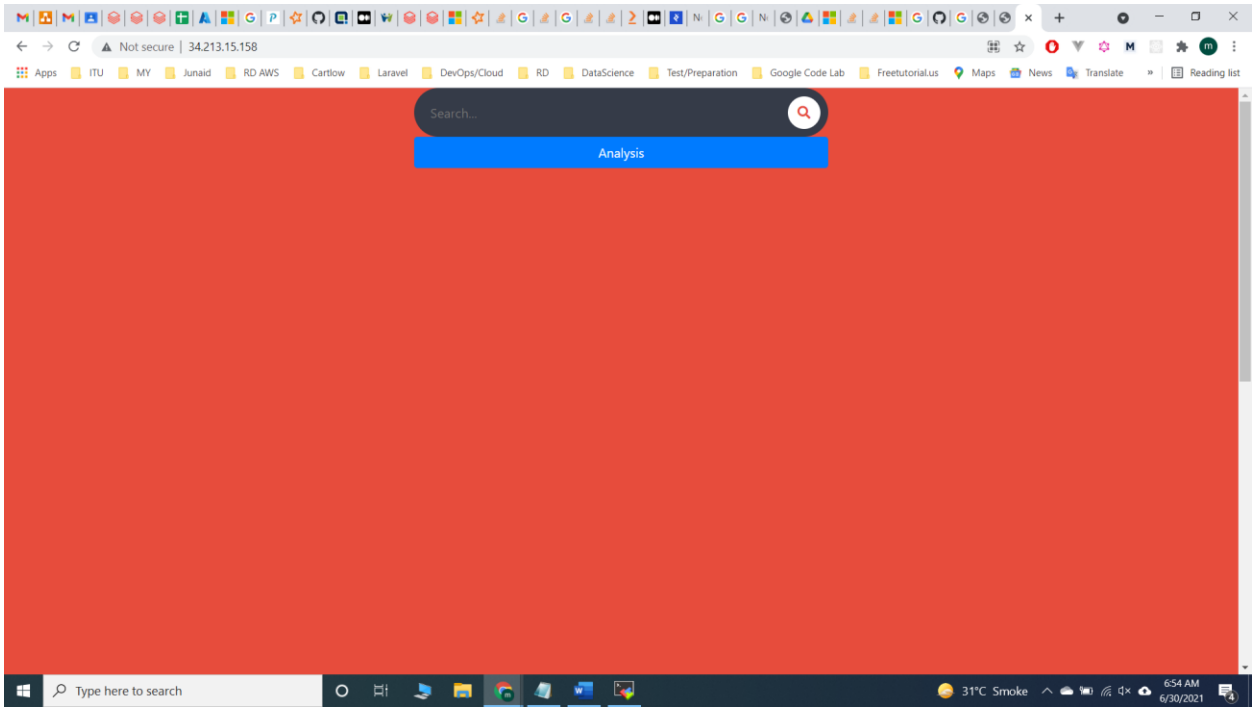




- Why we use AWS EC2 instead of Sage Maker Machine learning deployment?
  - We do not want additional resources to use with Sage Maker like S3, Lambda functions.
  - Secondly, we also want some server where we continue capture data from tweeter and pass forward to our pipeline and deployment of our model.
  - Main reason cost cutting & hands on E2c as well
  - We use Direct SSH client of for code deployment to Ec2 server.



- URL of server :
  - http://34.213.15.158/
  - Api Writing using flask framework to serve ML model in general and for data bricks.



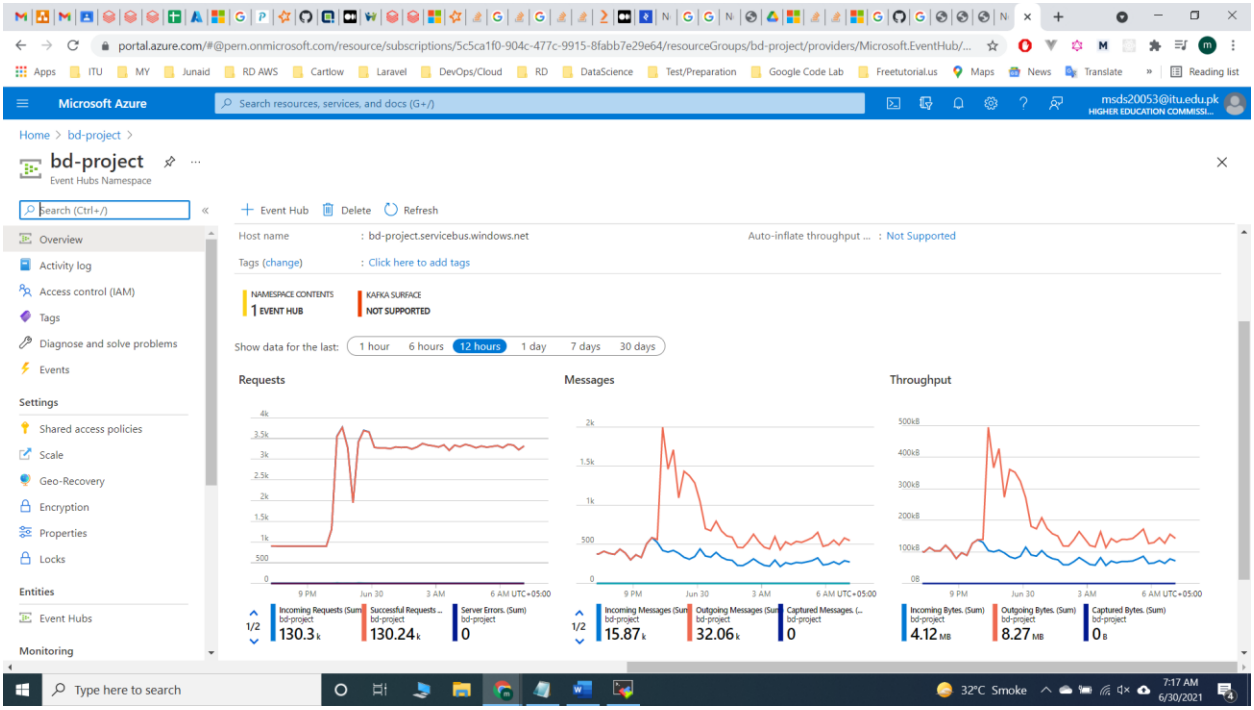
Qazi, Khuzama , Ahmad (Group 2)

Representor
Qazi & Uzair

- Qazi will represent data ingestion related part
  - We use **2 different tweets** sources from twitter using tweepy package of python,
  - One source we pass to **Azure Events Hub for** direct stream to **Azure Databricks (qazi did this work)**
  - Second source we pass to **Azure Cosmos DB for batch processing at Azure Databricks (khuzama did this work)**
  - Both sources are different, have no resembles of tweets as well (we filter them on different topics)
  - Tweepy able to continuous run , we deploy both scripts on EC2 server which is our sentiment analysis server as well. To make sure tweepy script continuous runs we reschedule using crontab of Operating system after 10 sec's **(ahmad did this work)**

- Tweepy script have 10 sec ‘s run time. **(ahmad did this work)**

Event Hub insight

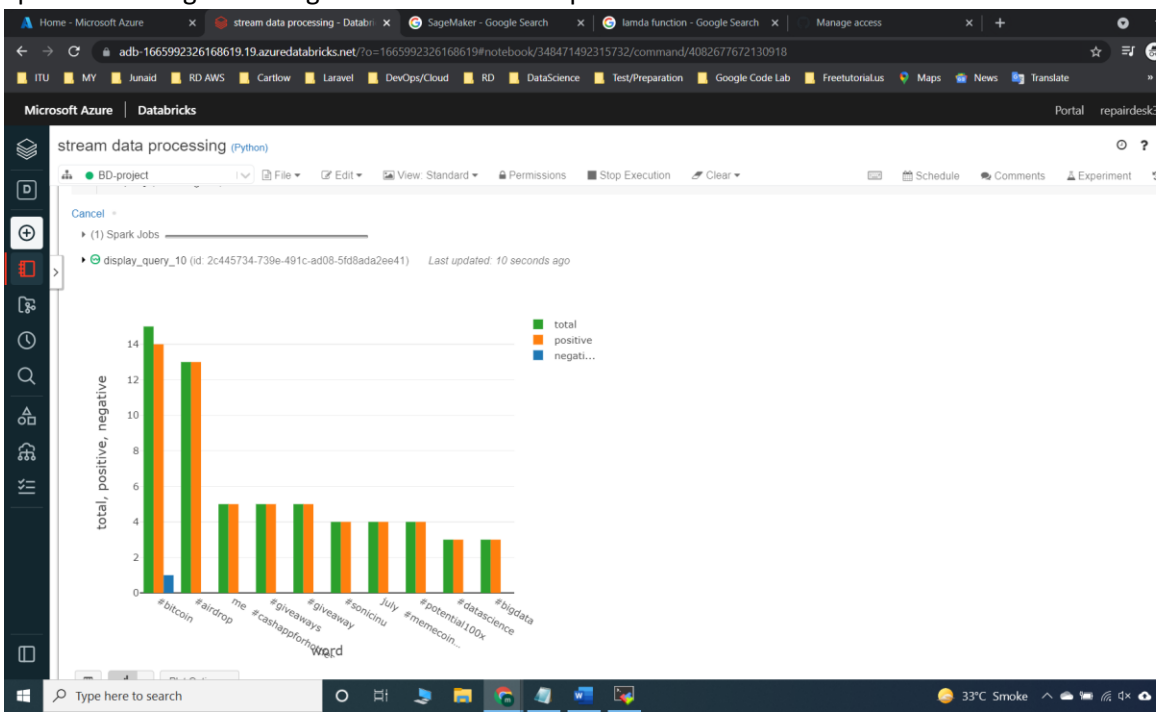


- Data ingest to Cosmos DB

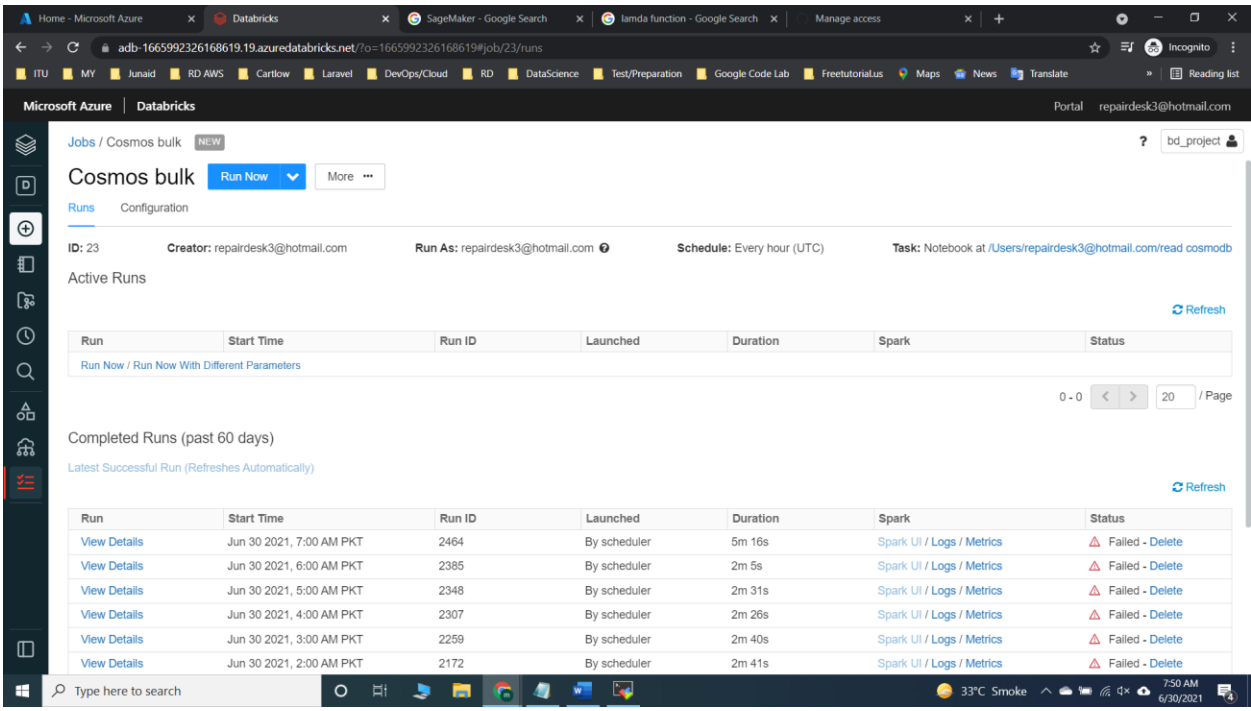

```
{
    "user": "Brandy Flora",
    "created_at": "2021-06-28 16:50:46",
    "location": "San Antonio, TX",
    "friends_count": 10,
    "followers_count": 16,
    "tweet": " L*SPACE Wendy Dress Wrap Ruffle Floral Medium NWT.… https",
    "id": "44176dae-2d78-4c07-9598-faad7e07b27a",
    "_rid": "2iMbAJwZjBgDAAAAAAAAAA==",
    "_self": "dbs/2iMbAA==/colls/2iMbAJwZjBg=/docs/2iMbAJwZjBgDAAAAAAAAAA==/",
    "_etag": "\"22005a5b-0000-2000-0000-60d9fded0000\"",
    "_attachments": "attachments/",
    "_ts": 1624899053
}
```

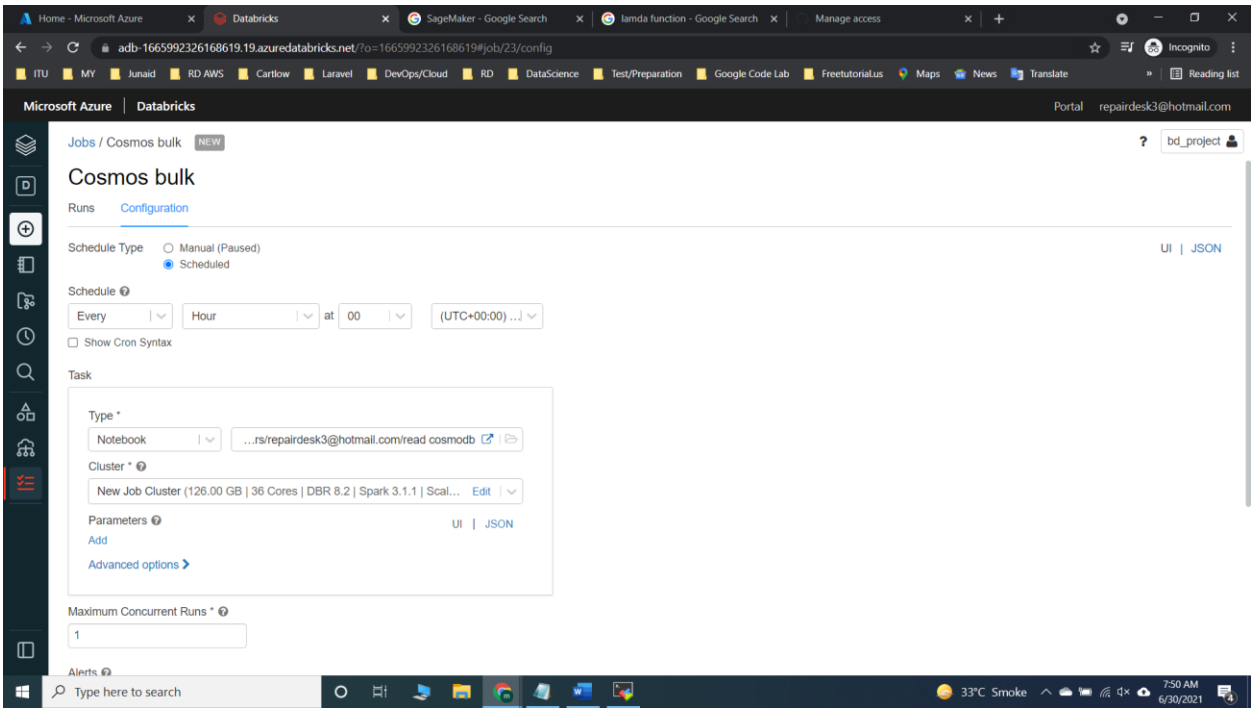
Uzair & Hira and Aiman (Group 3)

- We Consume Real time stream from event hub and pass to spark streaming where we find trends on hashtags and request for sentiment analysis of tweets to ec2 server
- We donot get any mechanism to deploy pickle or joblib related trained model to pyspark(similar solution azure have ML cogitate services)
- Spark streaming actions
    - Spark streaming data frame request for sentiment analysis
    - Spark streaming data frame split words to space and make second dataframe
    - Second dataframe aggregate on hashtag counts and positive and negative scores
    - Spark streaming sink using batch frames in complete mode to Comosdb



- Batch processing we use azure scheduler of every 1 hour to read cosmos db and dump back cosmos





- Hira Power BI
  - Hira will tell them problems of power bi with cosmosdb
  - Group by and limit problem together
  - Select clause much have group by related clause

- Power bi have no mechanism trace to ping database short interval
- Display of static data