## OUTPUT STEPS

## We have started with setting up Azure Databricks

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1092).png

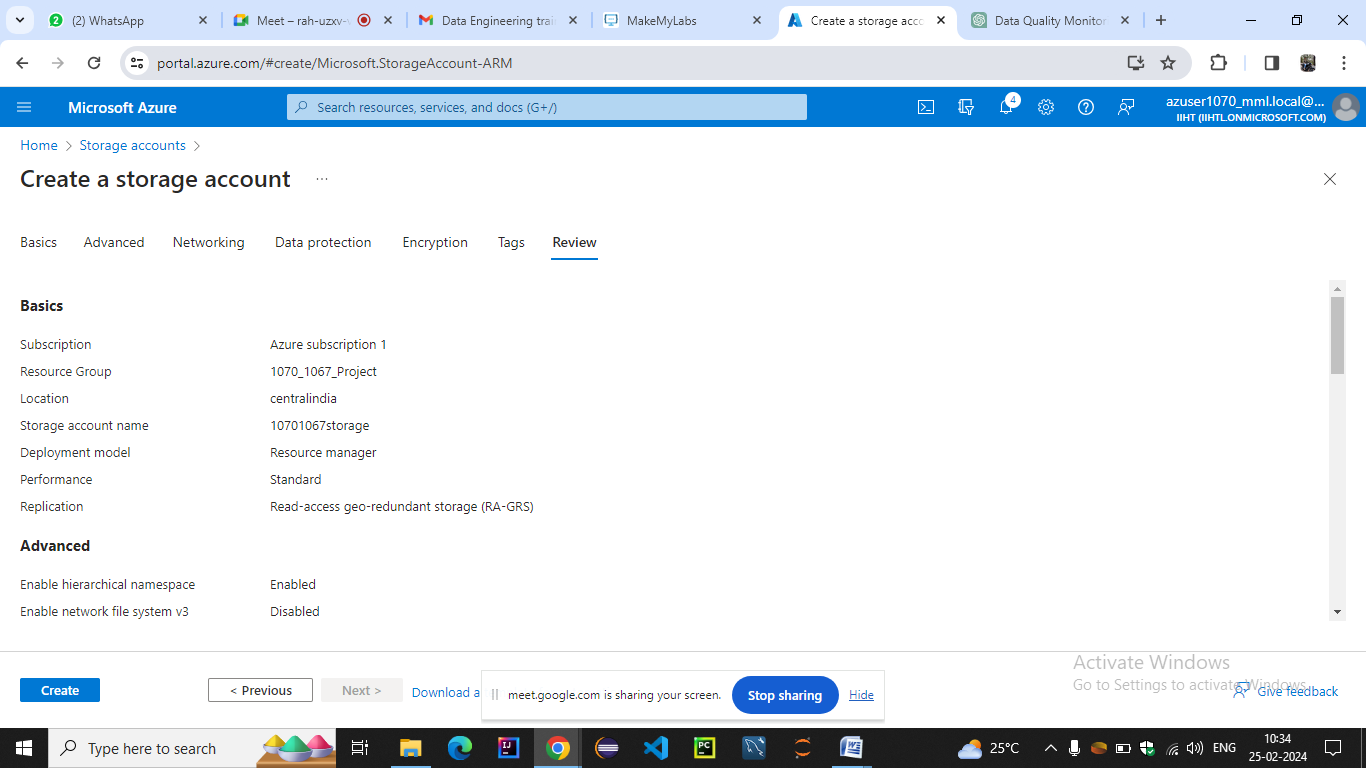
## After Launching Workspace we have created cluster

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1096).png

## We have taken our data source from kaggle and working on titanic boat dataset

## C:\Users\welcome_\Downloads\WhatsApp Image 2024-02-25 at 10.44.18 AM.jpeg

## We have Created Storage Account (ADLS GEN 2) for storing file



## We have launched account and created container and uploaded file in it .And we can also do it with azure storage explorer

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1095).png

## We have launched workspace and created new notebook

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1097).png

## We have ingested the data from adls gen2 into workspace

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1098).png

## 8.PERFORMING DATA PROFILING

## Print the schema

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1113).png

## Display the data

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1114).png

## Summary statistics (count,mean,stddev,min,max)

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1115).png

## Count null values for each column

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1116).png

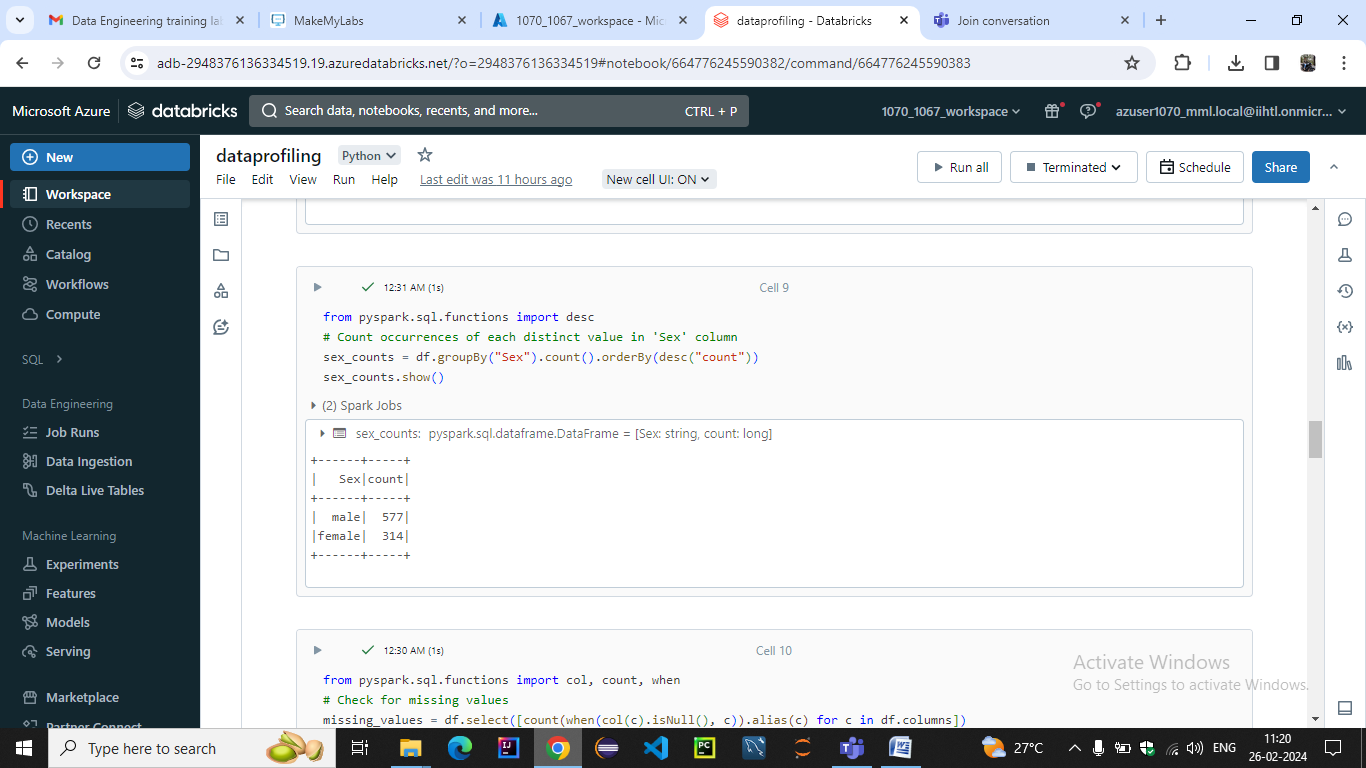
## DISTINCT VALUES

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1117).png

## COLUMN PROFILING (DISTINCTVALUES COUNT,AVG VALUE LENGTH)

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1118).png

## COUNT OCCURENCES OF EACH DISTINCT COLUMN LIKE(SEX )



## CHECK FOR MISSING VALUES

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1120).png

## COUNT OF DUPLICATE VALUES WITH CONDITION

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1121).png

## FREQUENCY COUNT

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1122).png

## OUTLIER DETECTION

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1123).png

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1124).png

## DATA HISTOGRAMS

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1125).png

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1126).png

## QUALITY

## COMPLETENESS SCORE

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1127).png

## COMPLETENESS

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1128).png

## ACCURACY

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1129).png

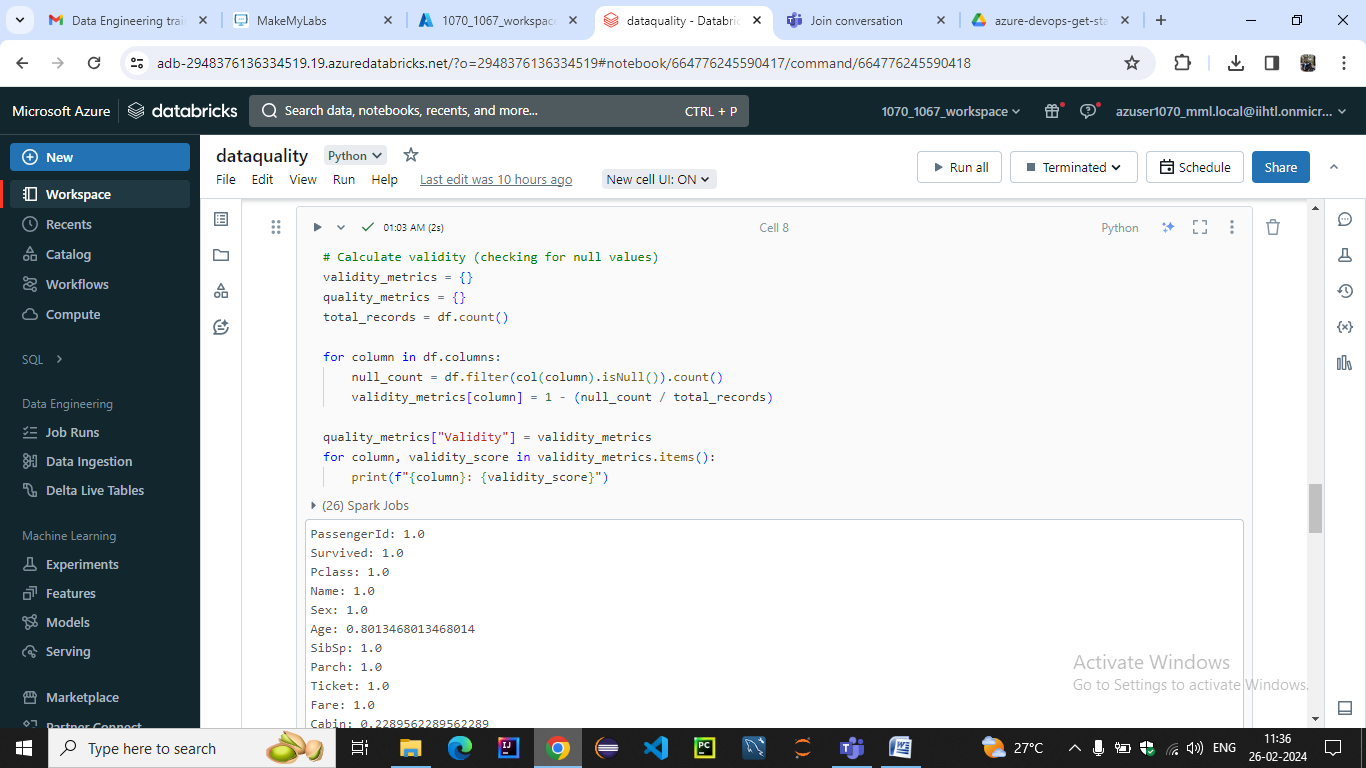
## CONSISTENCY

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1130).png

## UNIQUENESS

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1131).png

## VALIDITY



## RELIABILITY

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1133).png

## INTEGRITY

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1134).png

## PRECISION

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1135).png

## RELEVANCE

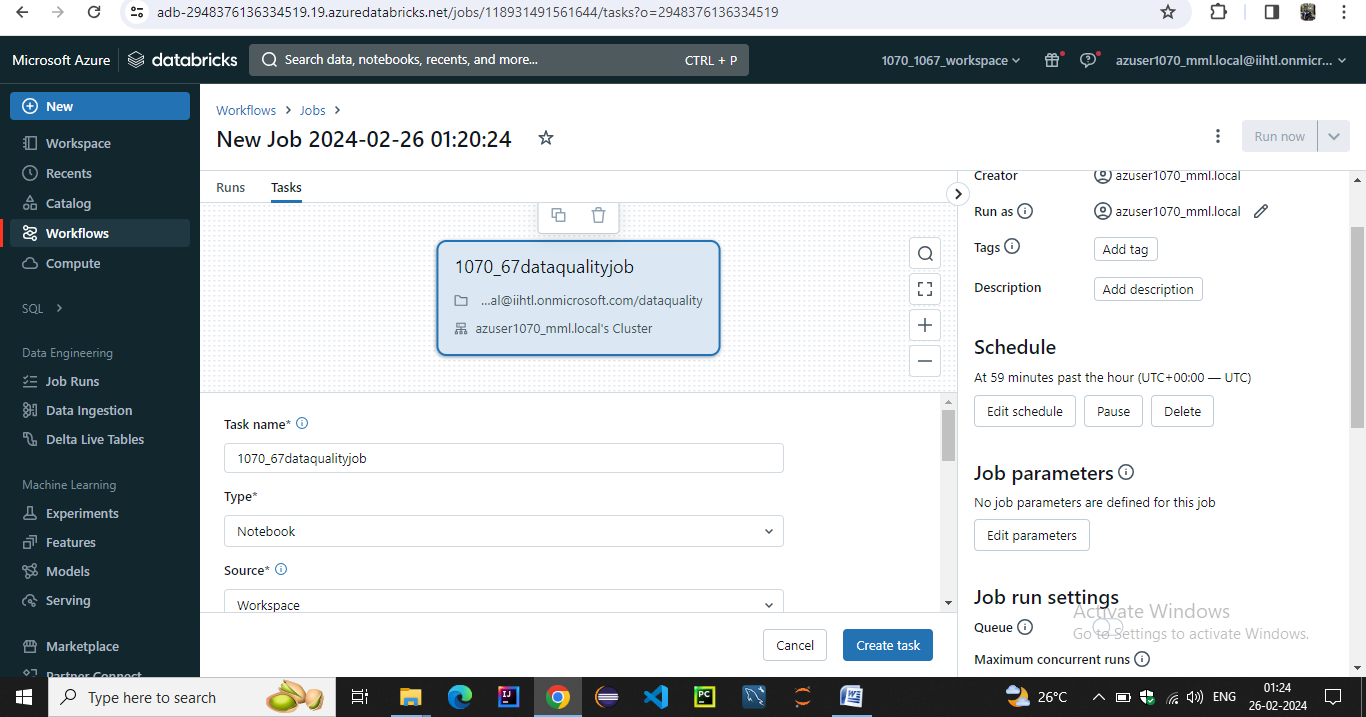
## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1136).png

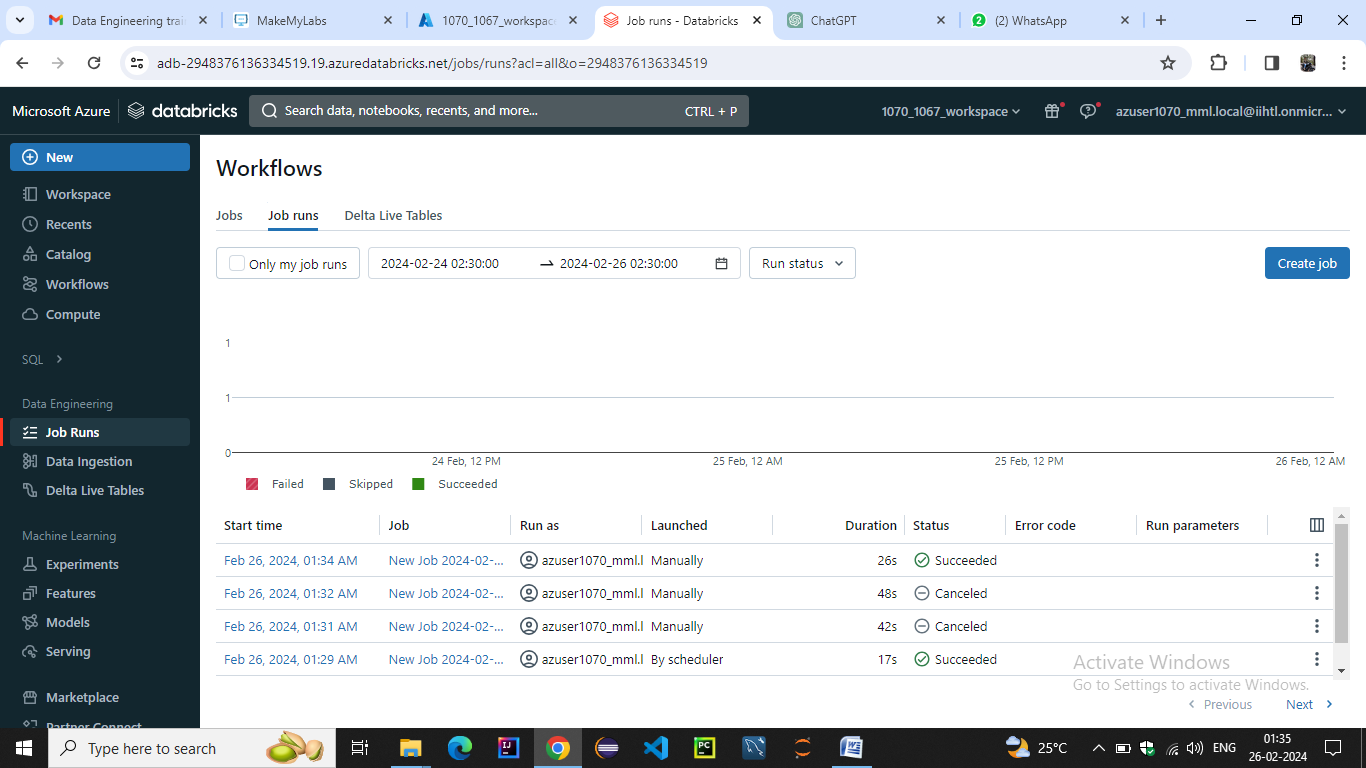
## CONFORMITY

## C:\Users\welcome_\Pictures\Screenshots\Screenshot (1137).png

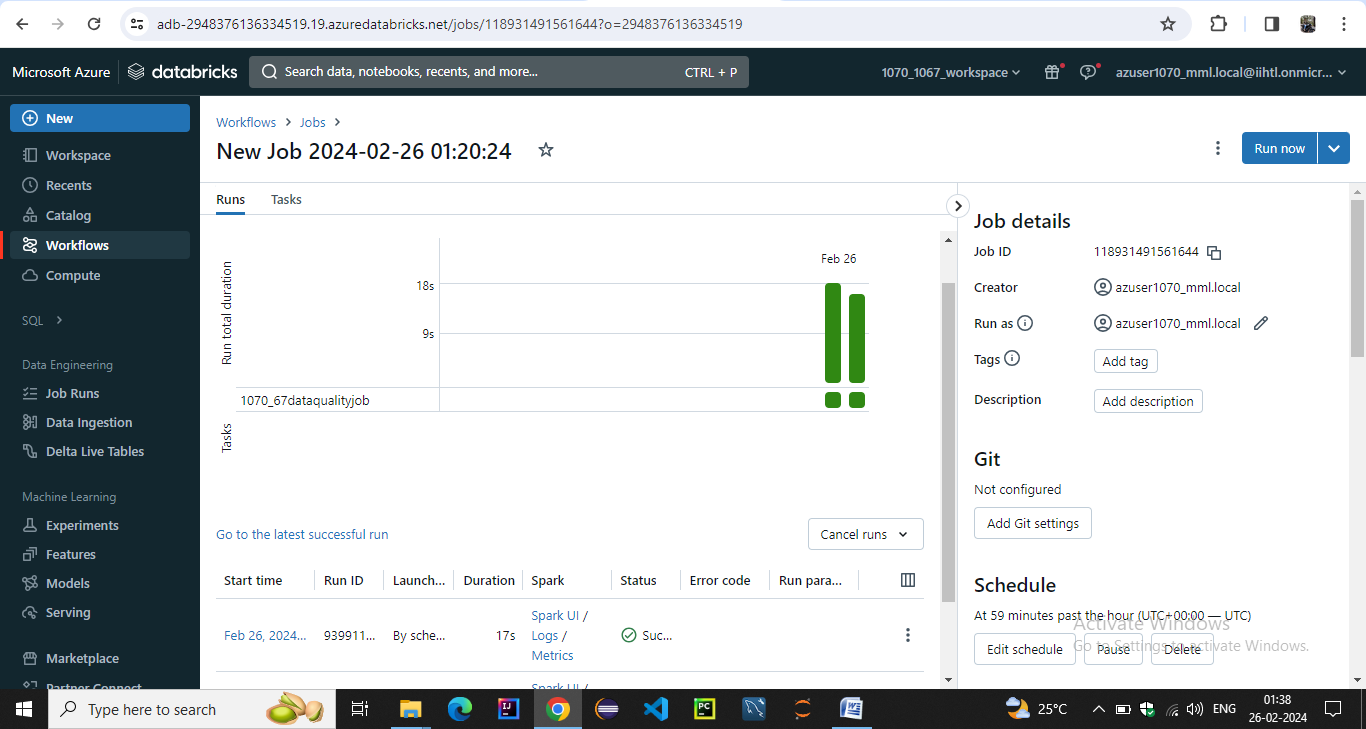
**9.PERFORMING CONTINUOUS MONITORING**

**We have scheduled job for continuous monitoring of data quality and data profile notebooks**

****

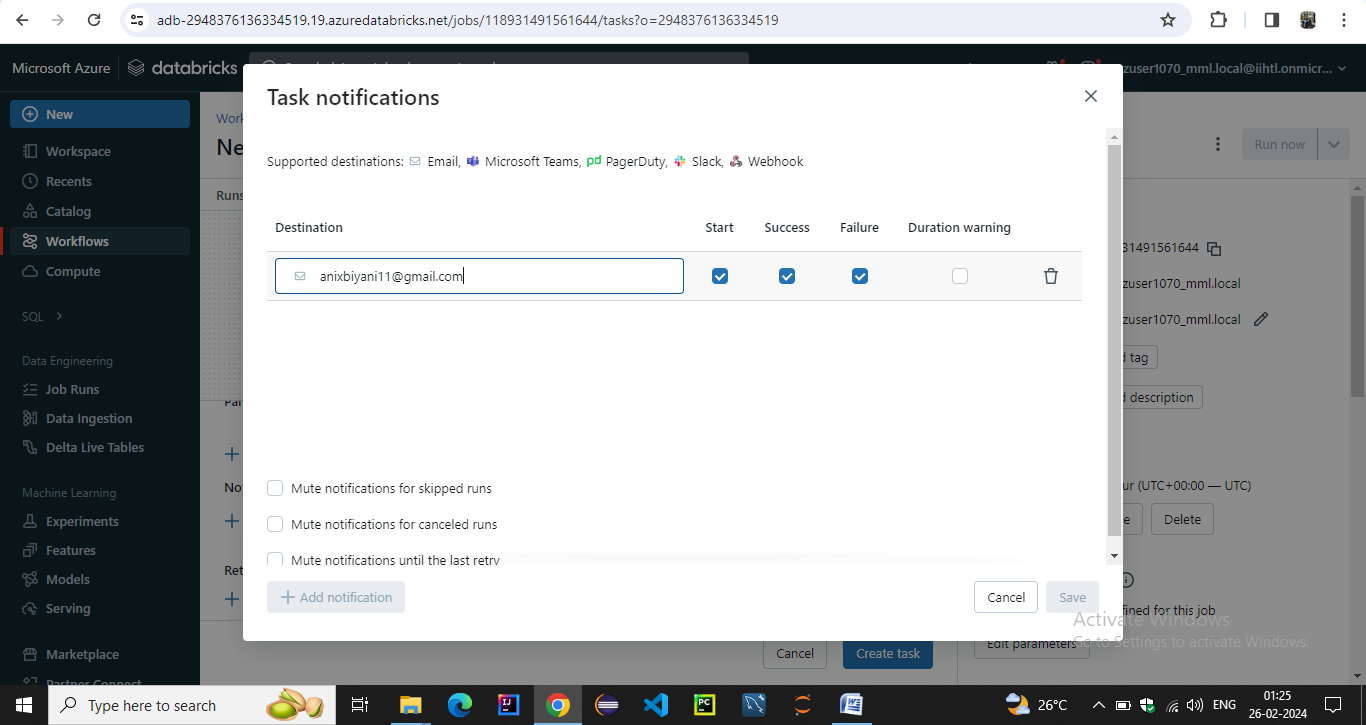
**Workflow of the jobs after scheduled interval **

**Job details**

****

**We have setup job task notifications on email for start success failure**

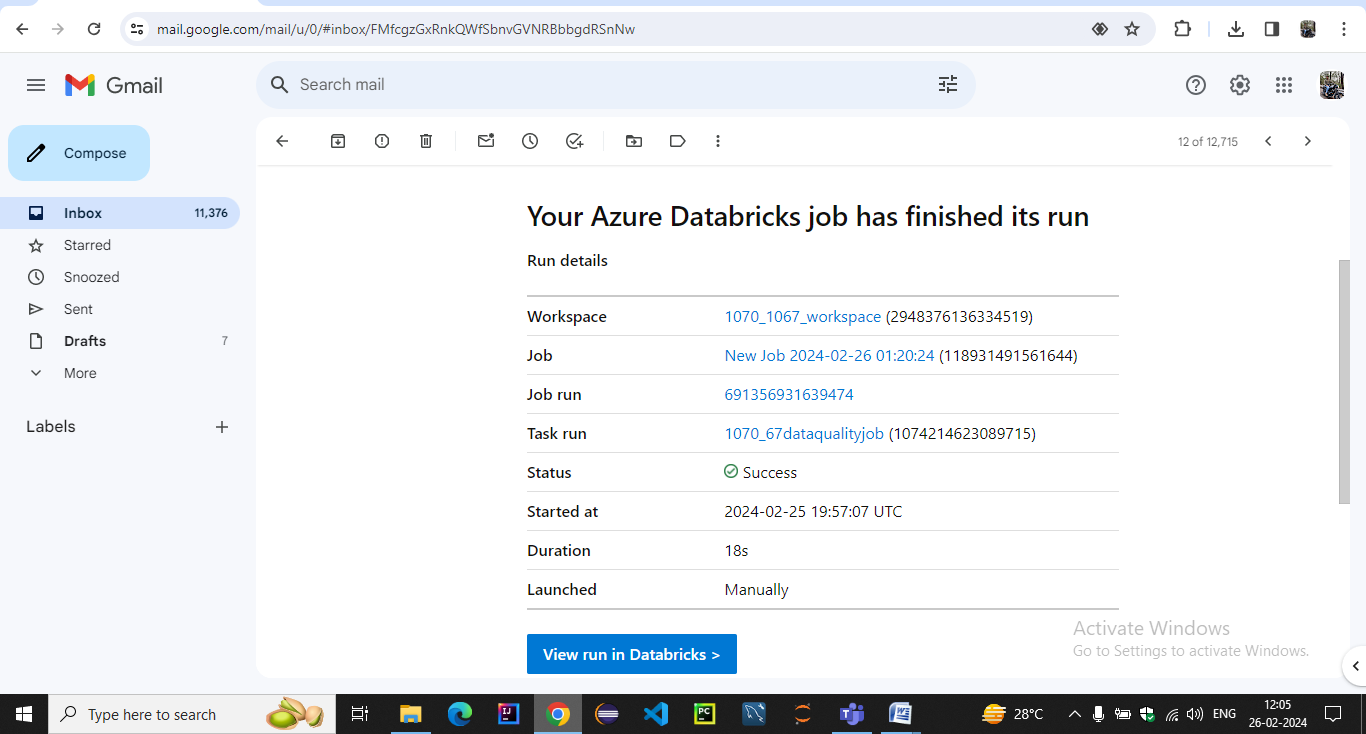
**After each interval**

****

**We have alerted tasks notifications for success start and failure notifications on the email which we provided**

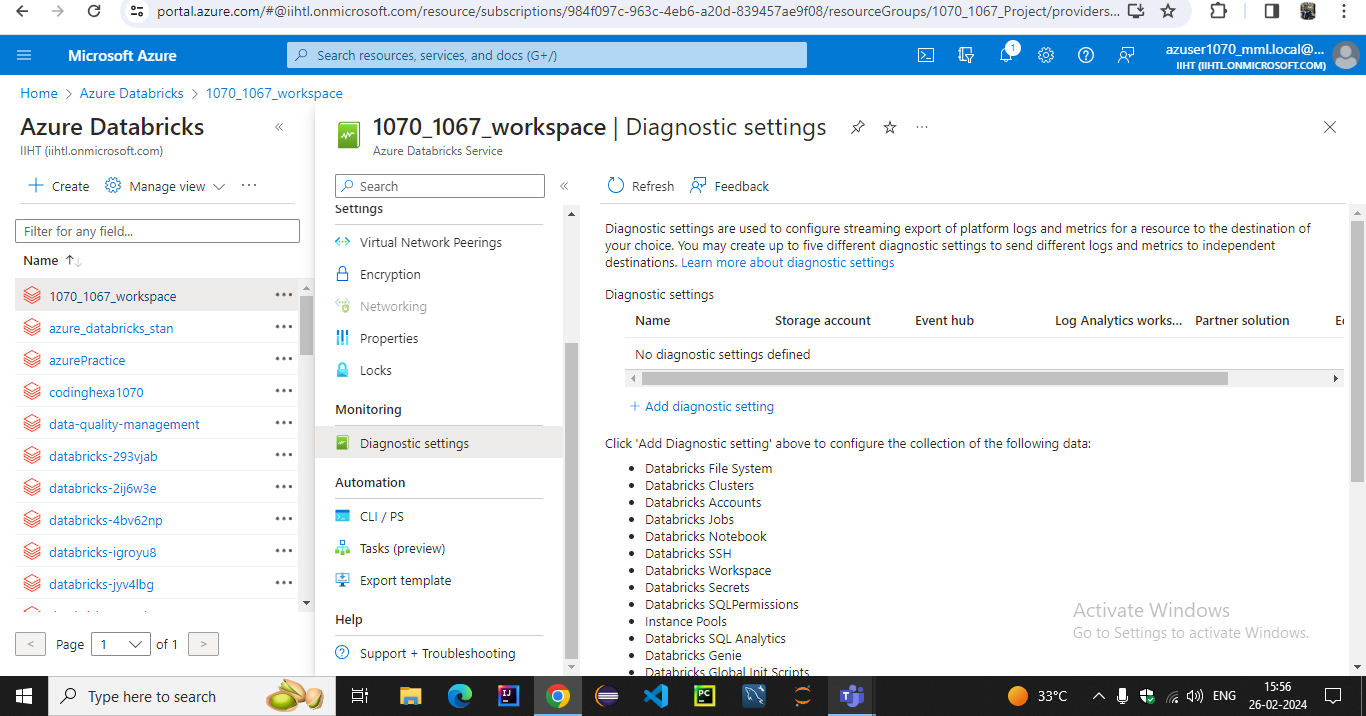
****

**After running at scheduled interval it also reverted success back**

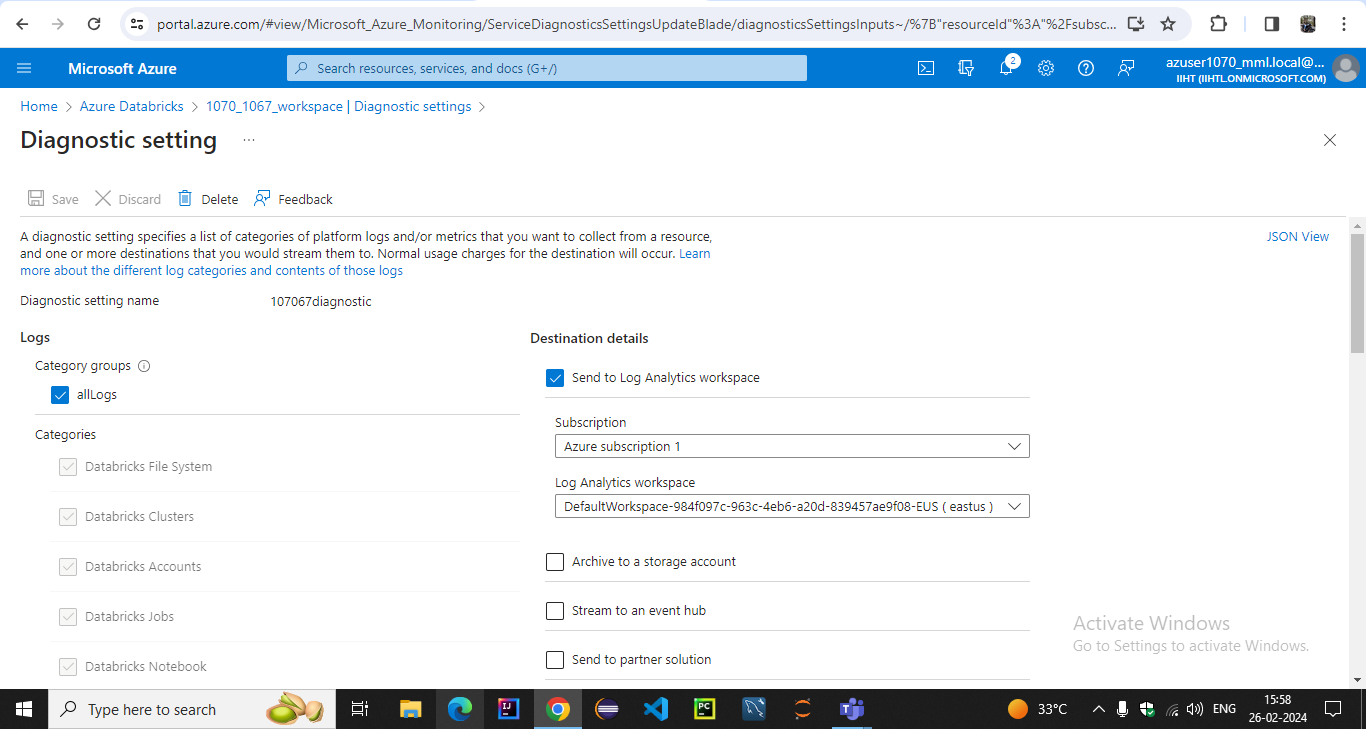
****

**10.Generating Alerts**

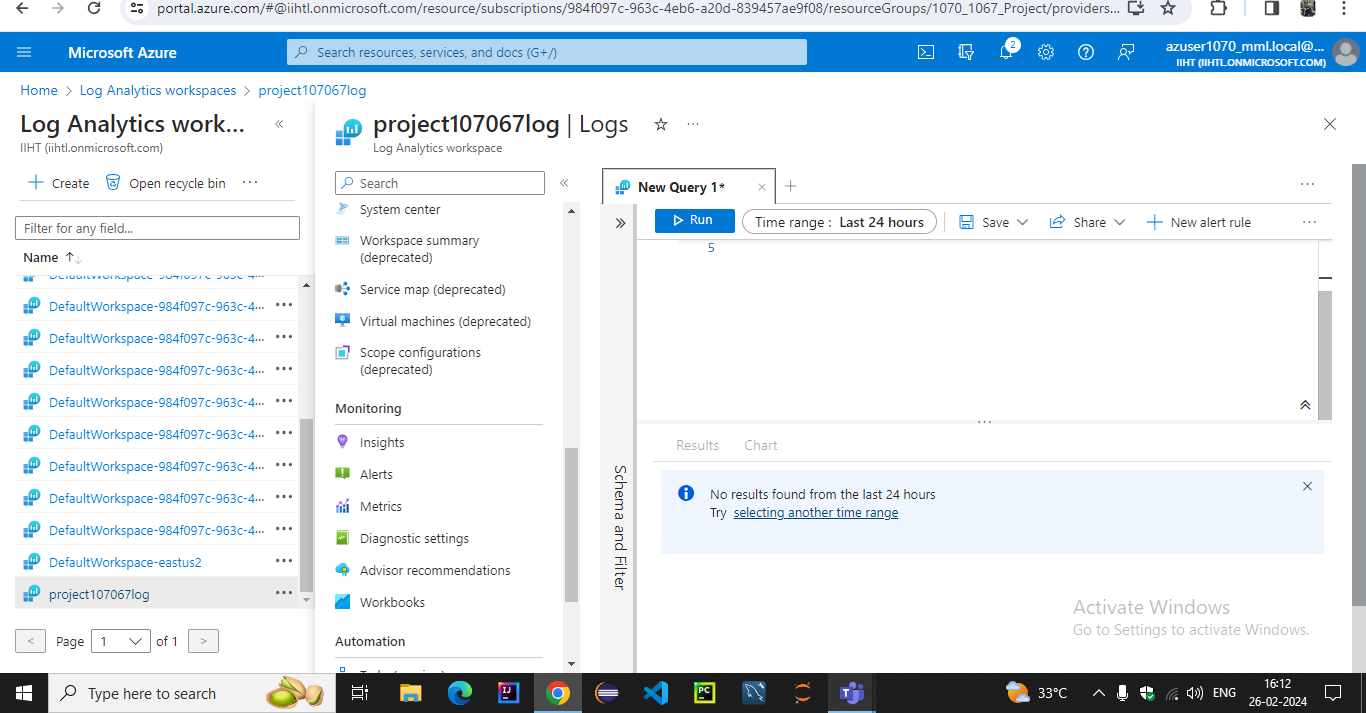
**For creating alerts we have to integrate azure databricks workspace to log analytics workspace**

****

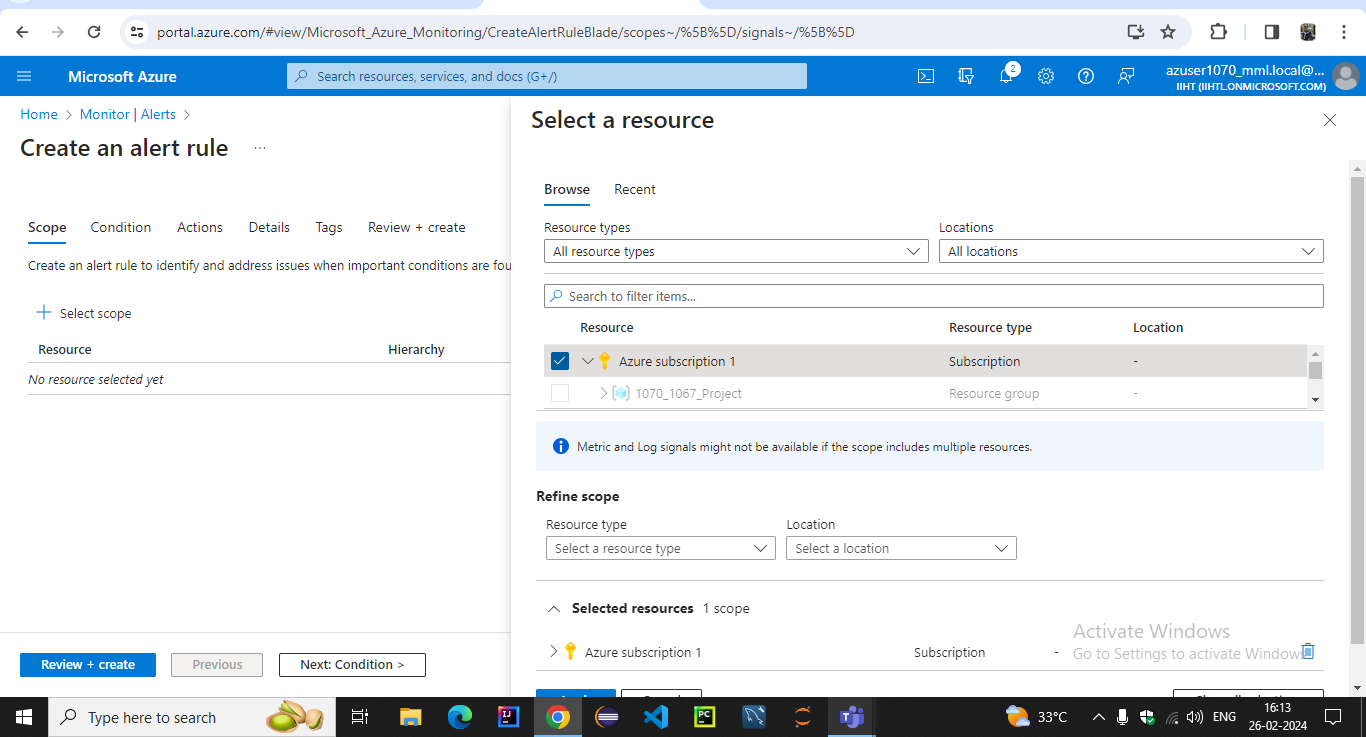
**we have diagnosed workspace to long analytics workspace**

****

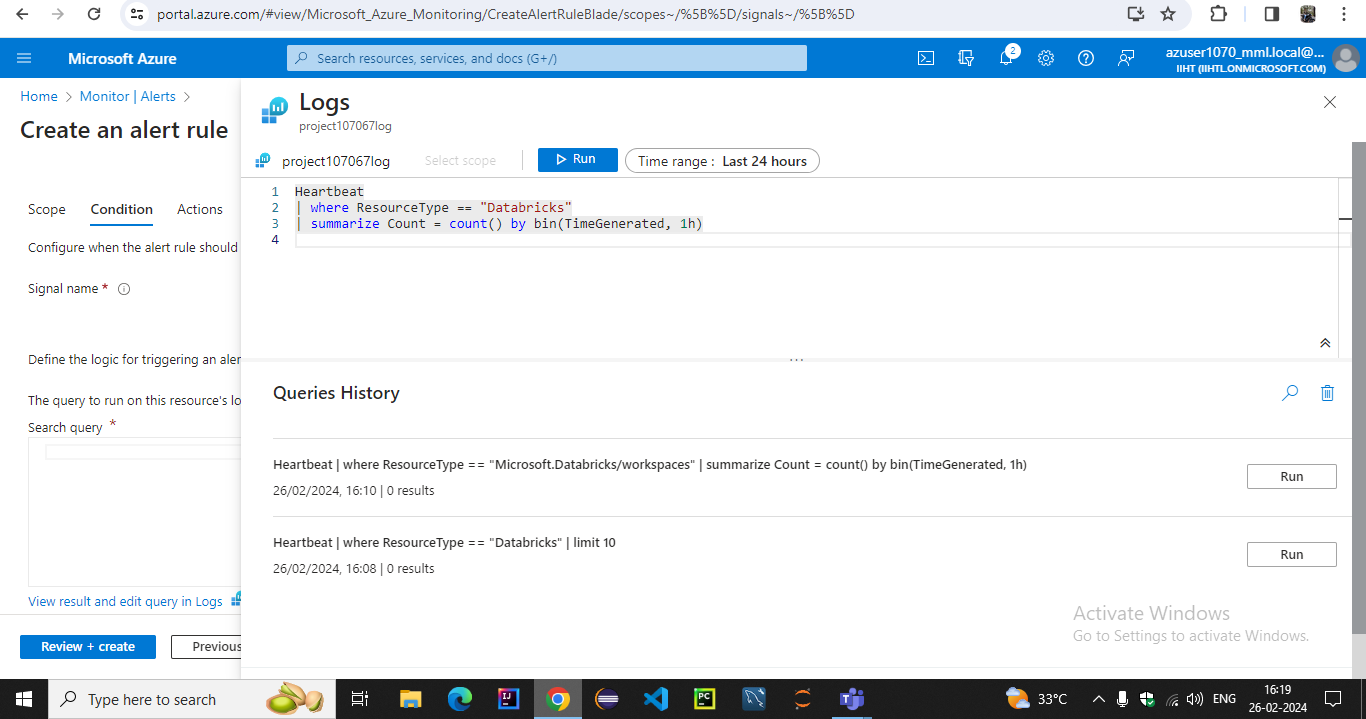
**In the logs workspace we checked our query and it runned successfully**

****

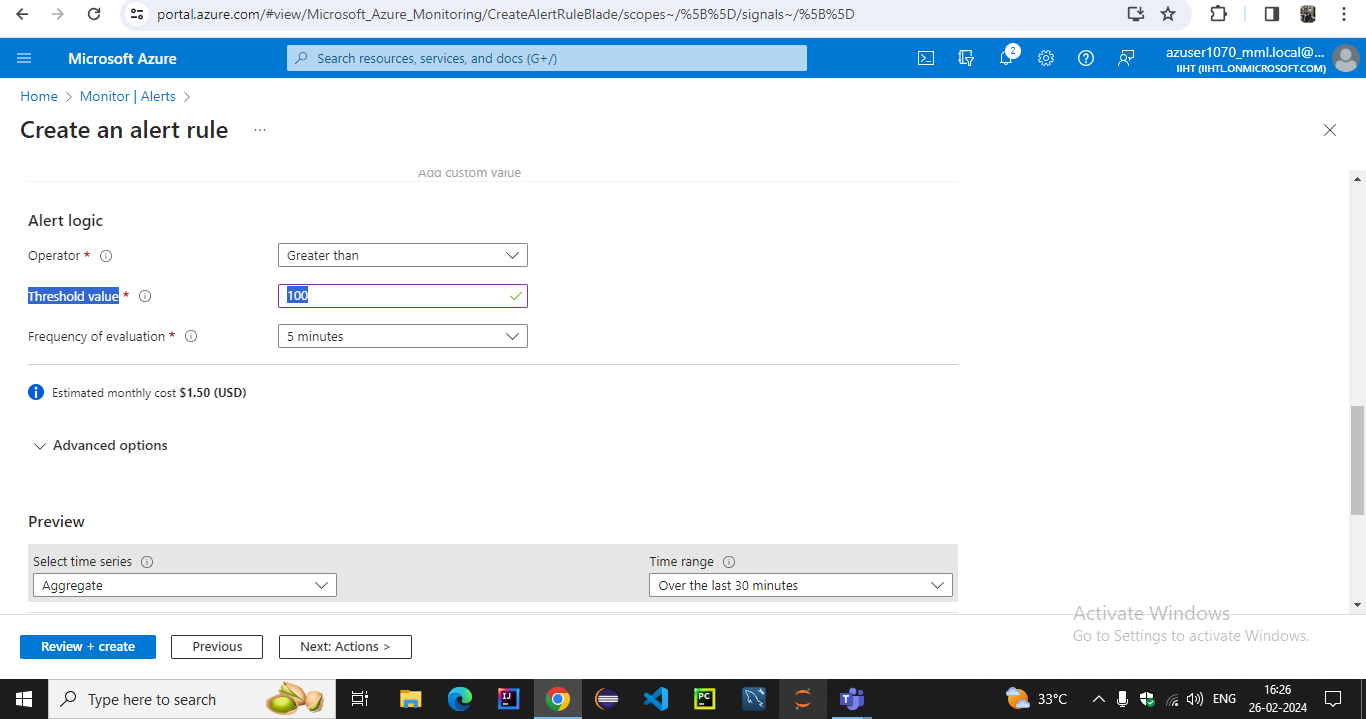
**Now going to azure monitor we have created alert with log metrics and configured scope with log analytics workspace**

****

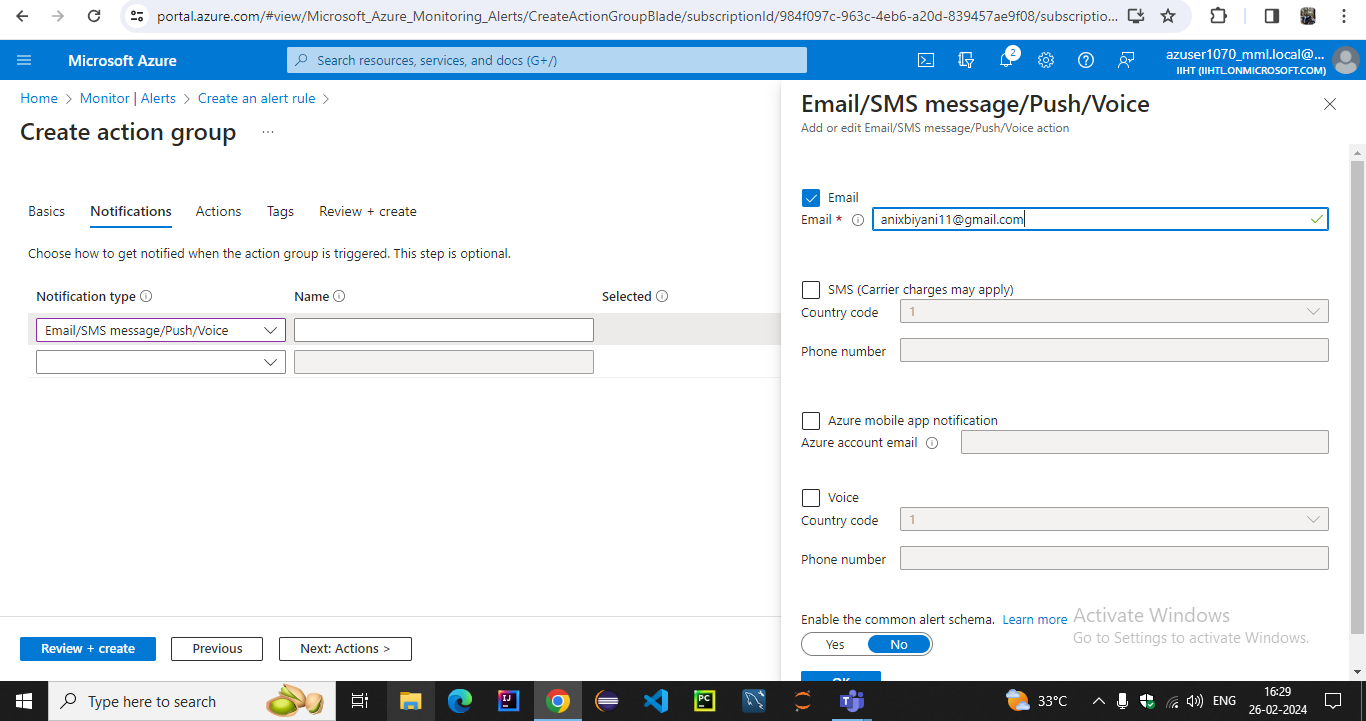
**We have specified the condition in the log query space with signal name as log analytics**

****

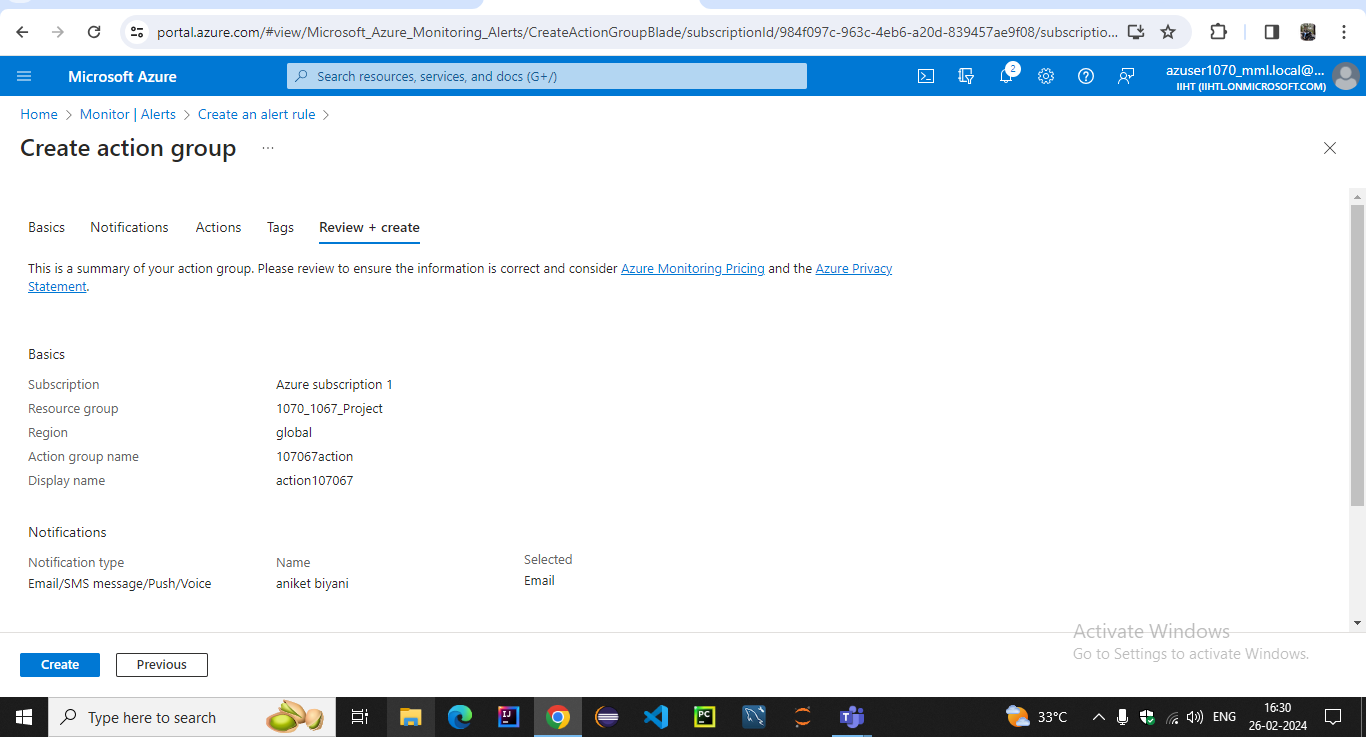
**In the alert logic we specified the threshold value operator and frequency of evaluation**

****

**We have set alert notifications email in the action group**



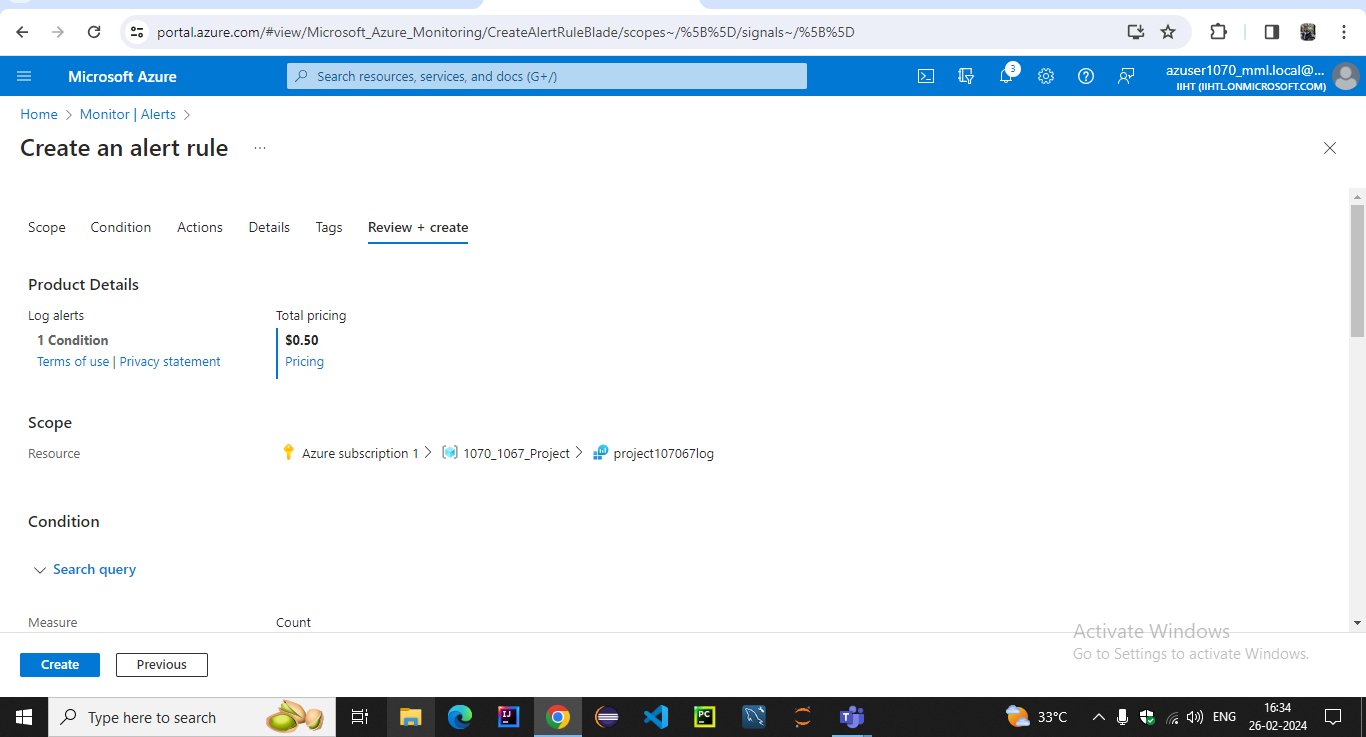
**We have created action group in the alerts**



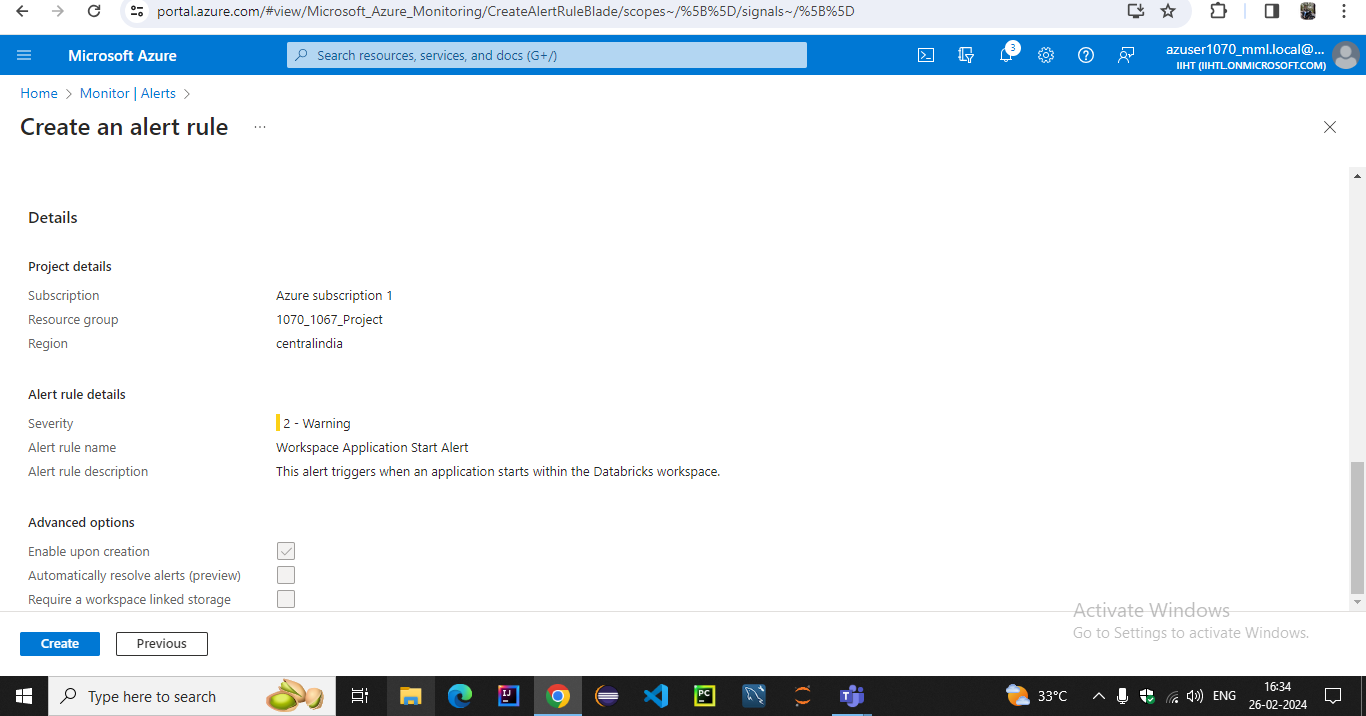
**We have created alert rule by providing severity name description**



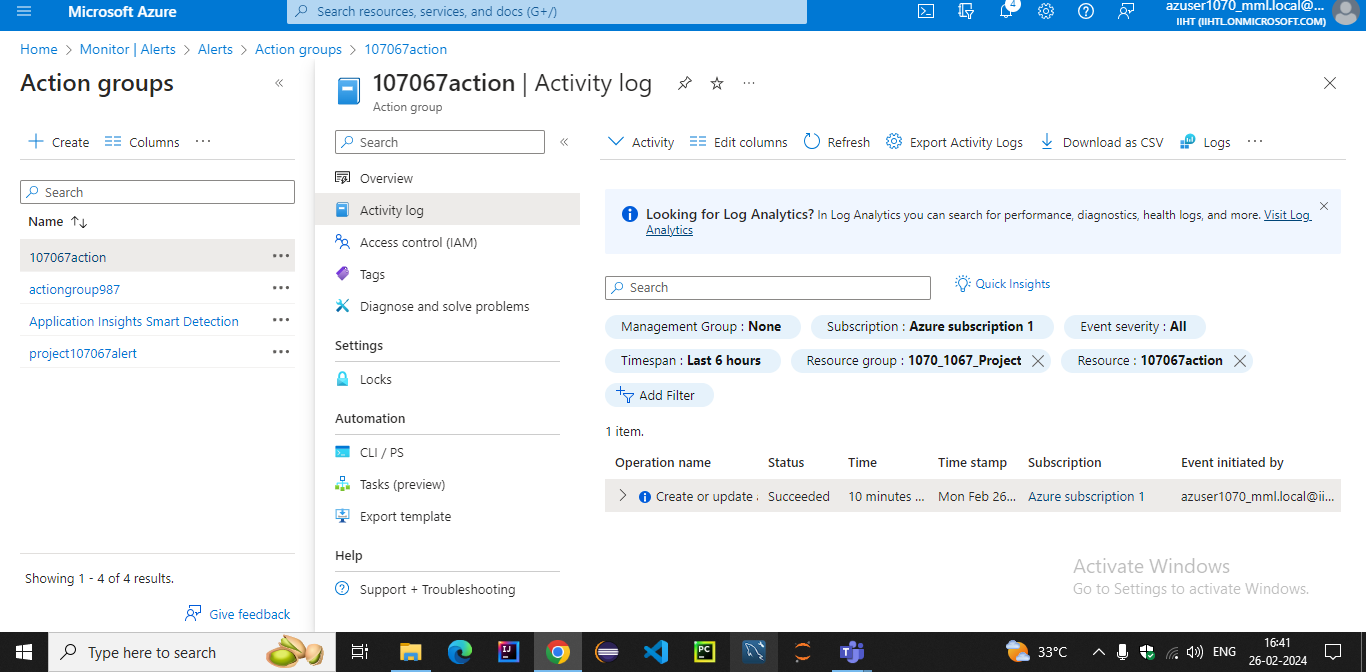
**Now reviving and creating with pricing of 0.5 dollar with scope and condition**

****

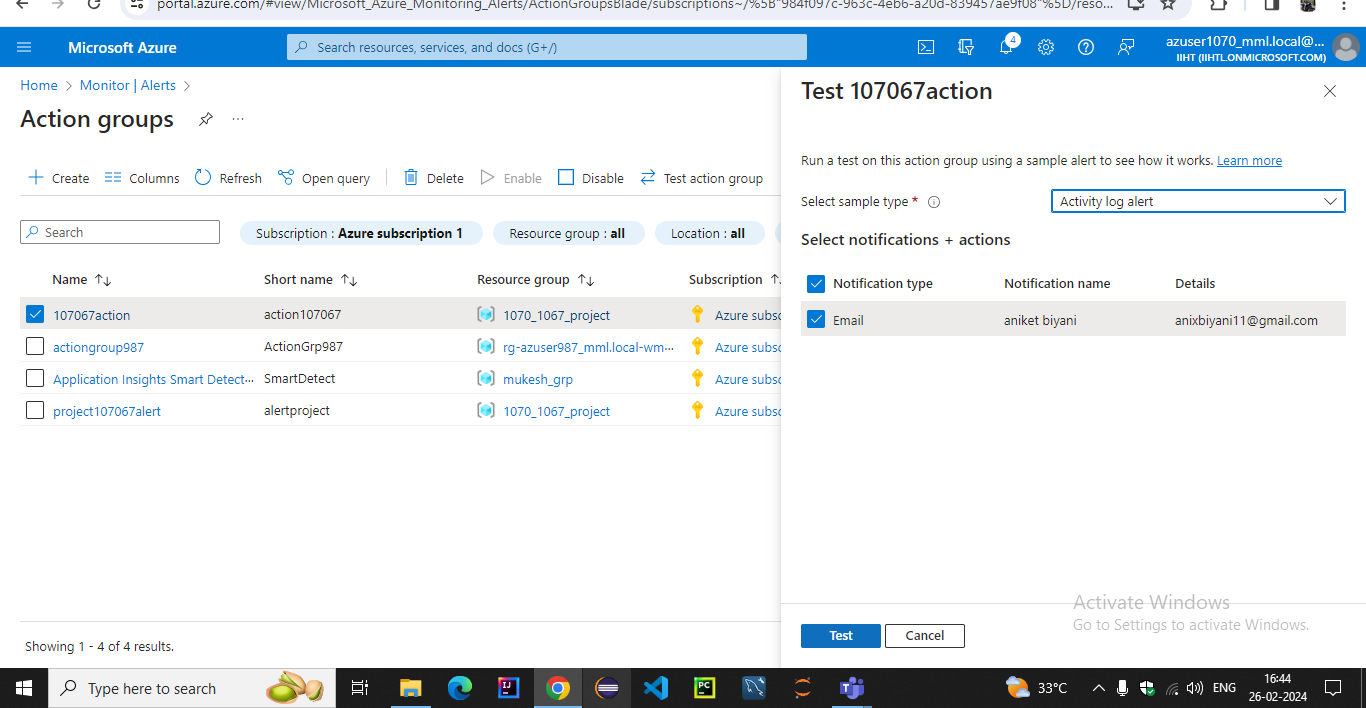
**Finally we created alert rule successfully**

****

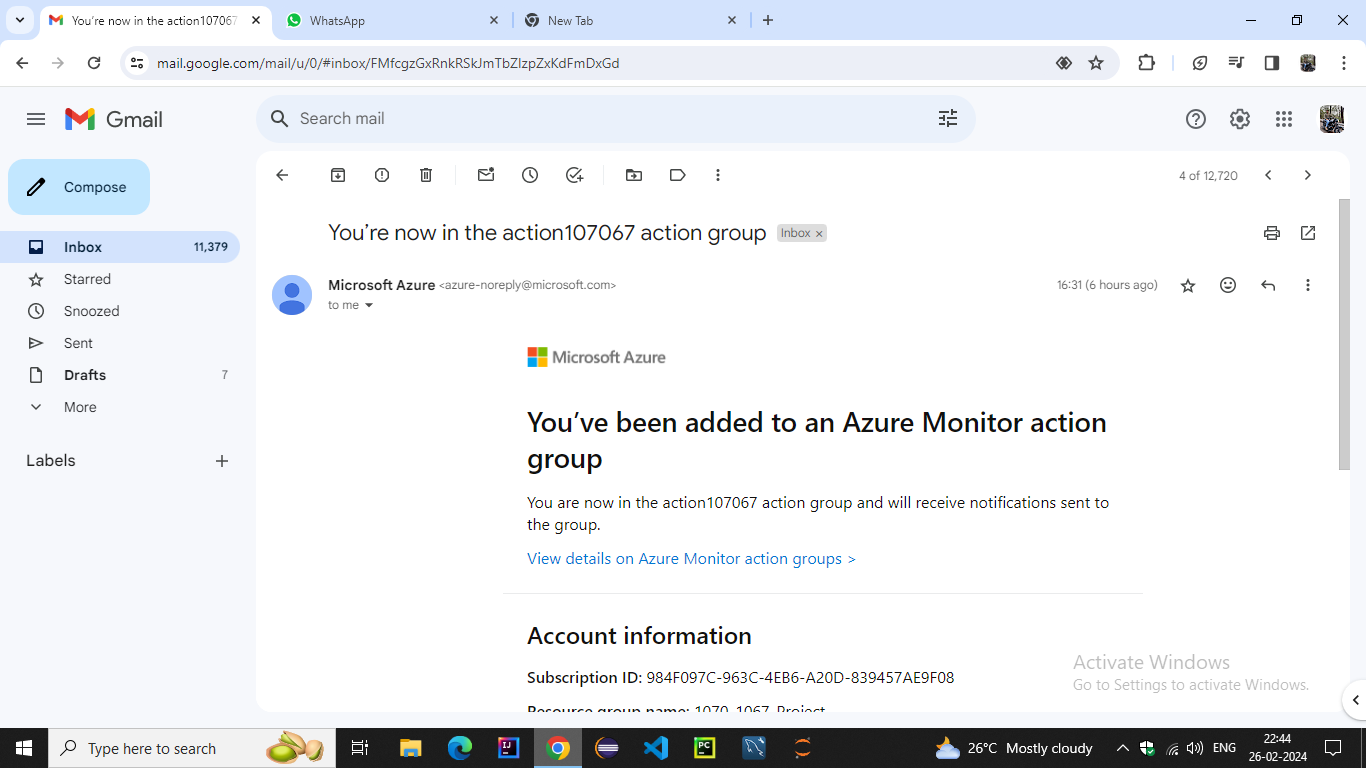
**We can see our actions in activity log**

****

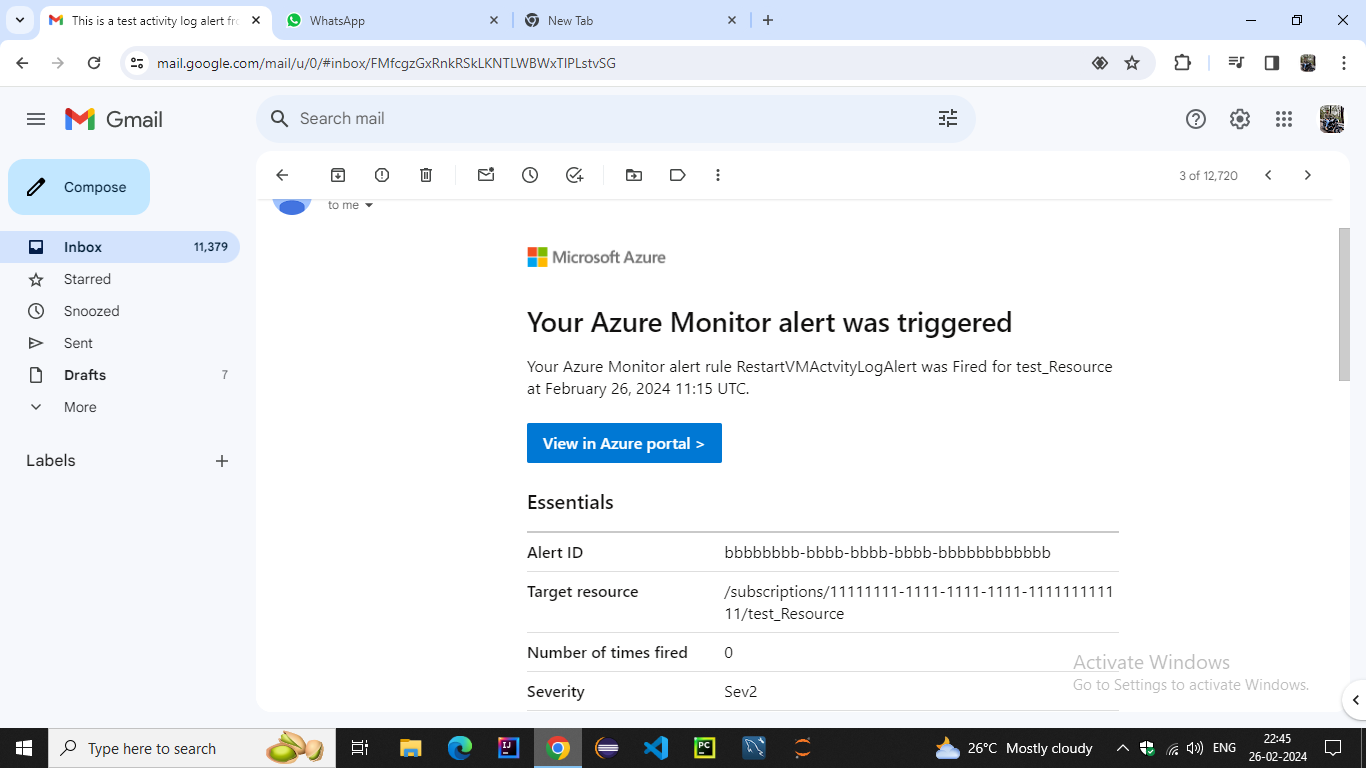
**Next step is to test our action we created for alert**

****

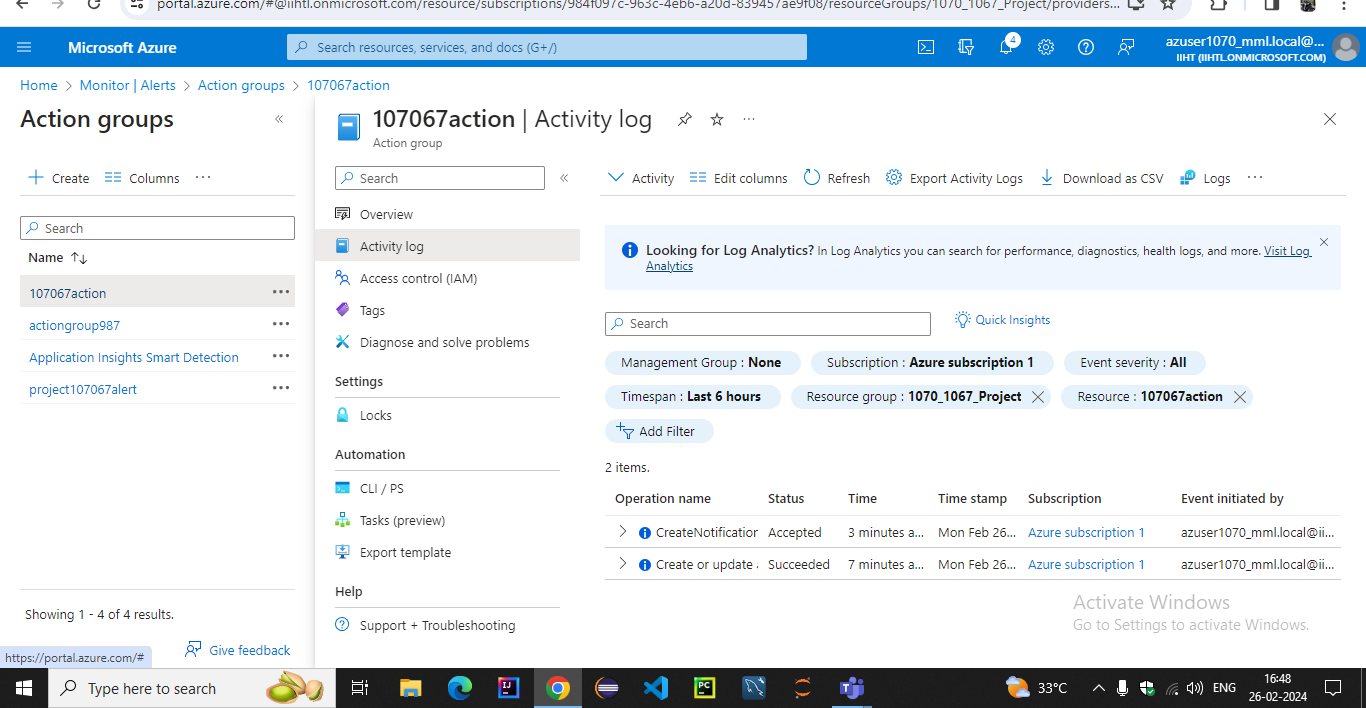
**After testing our alert has been added into azure monitor action group**

****

**Our azure monitor alert was triggered successfully**

****

**Now we can see our action triggered in the log with log status whenever alert get triggered**

****