

DADS6005 Data Streaming and Real-Time Analytics

มูลค่ารวม 1.46 แสนล้านดอลลาร์สหรัฐ คิดเป็นกว่า 3 เท่าของมูลค่าบีโอกซ์ซอฟฟิศทั่วโลก ในปี 2019

จำนวนผู้เล่นเกมในไทย ปี 2020

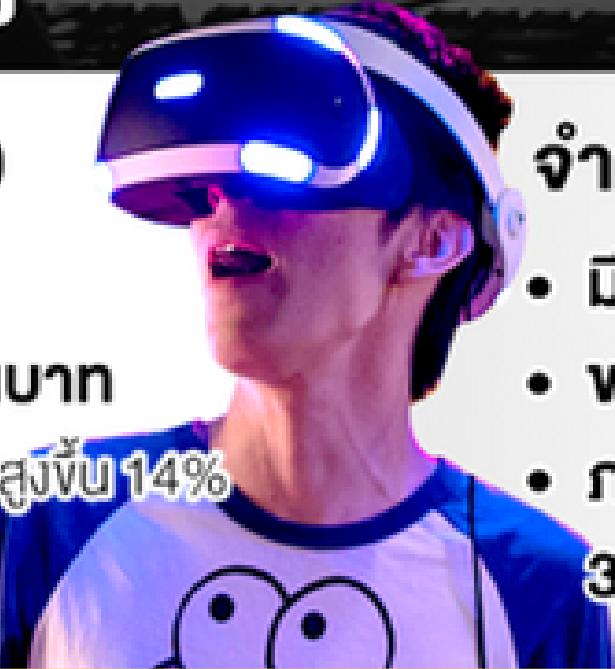
- สูงถึง 27 ล้านคน
 - คาดปีนี้เงินสะพัดรวม 33,000 ล้านบาท
- ปี 2019 อุตสาหกรรมเกมมีอัตราการขยายตัวสูงขึ้น 14%
คิดเป็นมูลค่า 29,000 ล้านบาท)

จำนวนผู้เล่นเกมทั่วโลก ในปี 2020

- มีอัตราการเติบโต 20% YoY
- ขยายตัวเป็น 27,000 ล้านคน
- ภายใน 3 ปี คาดว่าจะขยายได้ถึง 31,000 ล้านคน (ในปี 2023)



ปี 2023 คาดการณ์ว่าจะมีจำนวนผู้เล่นเกมบนสมาร์ตโฟน
สูงถึง 25,000 ล้านคน หรือราว 1 ใน 3 ของประชากรโลก



Data source: <https://workpointtoday.com/game-industry/>

PROFESSIONAL ESPORT GAMERS



DOTA2 TOURNAMENTS

streaming in 2 years difference





GROUP MEMBER

Metpiya Lertakkakorn

6410422005

Khodchapan Vitheethum

6410422015

Peerat Pookpanich

6410422017

Anyamanee Pornpanvattana

6410422031

Patcharapruet Watanangura

6410412007

AGENDA



- 1** Diagram
- 2** Data Source
- 3** Data Processing
- 4** Distribute Data Tool
- 5** Data Storage
- 6** Data Visualization
- 7** Insight and Conclusion





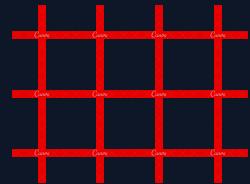
EXPECTATION DESIGN

1. DESIGN AND IMPLEMENT DATA STREAMING AND REAL-TIME ANALYTICS SYSTEM (DRAW THE DESIGN DIAGRAM)
2. THE SYSTEM CONSISTS OF 4 COMPONENTS
 - AT LEAST TWO DATA SOURCES
 - USE DISTRIBUTE DATA TOOL (EG. KAFKA, GOOGLE PUB/SUB)
 - REAL-TIME PROCESSING AND ANALYTICS COMPONENT
 - DISPLAY COMPONENT: MONITORING AND SHOWING ANALYTICS RESULTS

DIAGRAM



DOTA The International:
FINAL DAY



2021

Data
Publisher 1



Cloud Pub/Sub

subscriber 1



Table 1



Google
BigQuery



Grafana



2022

Data
Publisher 2



Cloud Pub/Sub

subscriber 2



Table 2

DATA SOURCE

เราดึงข้อมูลคอมเมนต์จาก YouTube live จากการ live เกม
การแข่งขัน DOTA2 ในปี 2021 และ ปี 2022





DOTA TOURNAMENTS

DIAGRAM



DOTA The International:
FINAL DAY

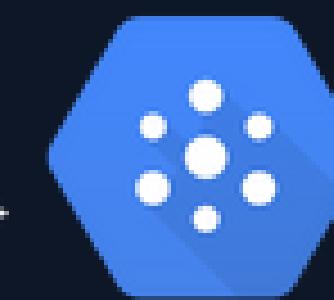


2021



2022

Data
Publisher 1



Cloud Pub/Sub

subscriber 1

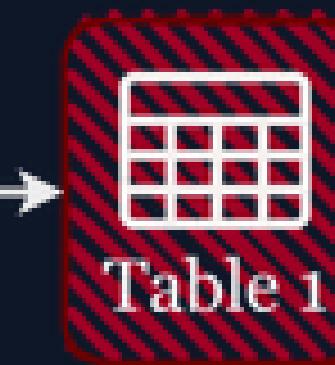


Table 1

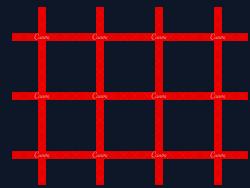
subscriber 2



Table 2



Google
BigQuery



Grafana

DATA PROCESSING

ในส่วนของ Data Processing เรานำ data ในแต่ละ source เก็บข้อมูลในรูปแบบ Table โดยมีข้อมูลคือ date time, author name และ comment และนำ comment ที่ได้มาในแต่ละช่วงเวลา count ตามเวลา comment จริงและแสดงในรูปแบบของกราฟ ดังนี้

- quantity
- cumulative



```

1 !pip install pytchat
2 !pip install google-cloud-pubsub
3 # Import library
4 import time
5 import pytchat
6 import json
7 from google.cloud import pubsub_v1
8 from time import sleep
9 import os
10
11 # Connect with google Pub/Sub
12 credential_path = "/content/substantial-art-374705-bb95d1f58737.json"
13 os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = credential_path
14
15 # Streaming comment data on YouTube platform
16 chat = pytchat.create(video_id="UhQmAfzaw7c") # DOTA 2 game
17
18 # Inject data to google cloud
19 project_id = "substantial-art-374705"
20 topic_id = "projects/substantial-art-374705/topics/dota2_pub_2"
21
22 publisher = pubsub_v1.PublisherClient()
  
```



DOTA TOURNAMENTS



LIVE
DOTA The International:
FINAL DAY

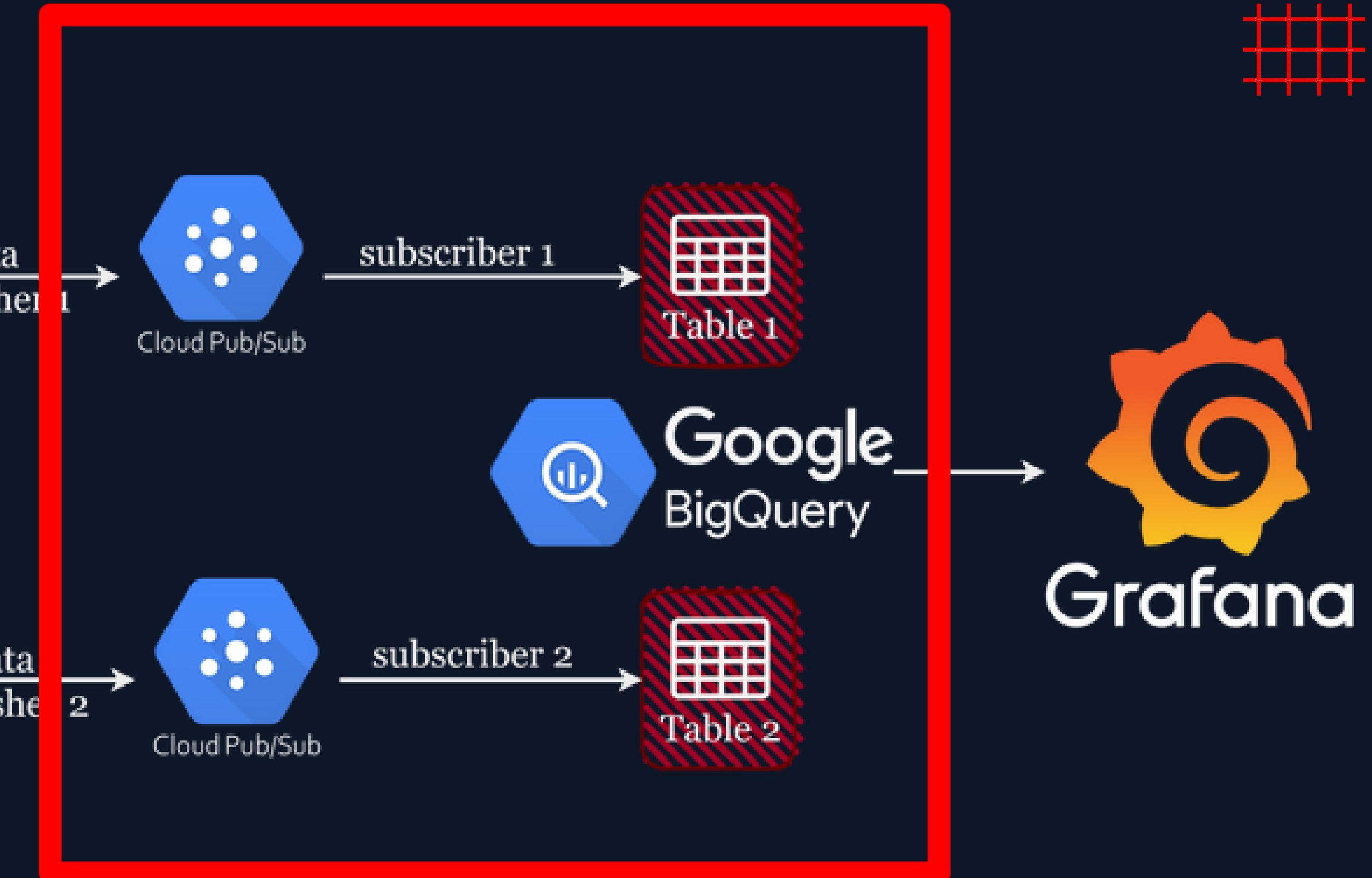


2021



2022

DIAGRAM





DISTRIBUTE DATA TOOL

Cloud Pub/Sub

Create a topic

A topic forwards messages from publishers to subscribers.

Topic ID * ?

! ID is required

Add a default subscription ?

Use a schema ?

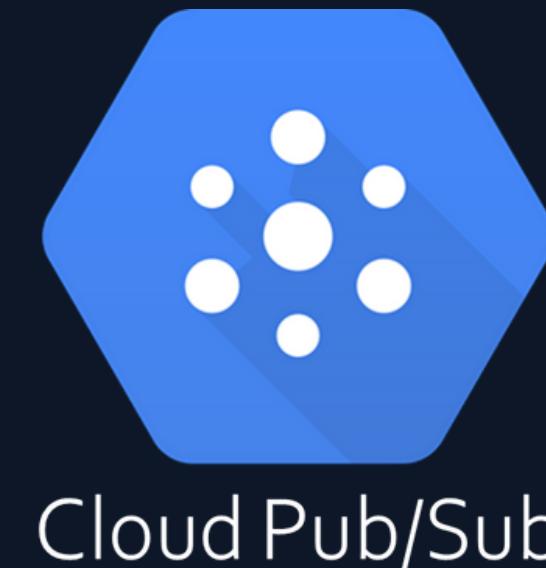
Set message retention duration (not free) ?

Encryption

Google-managed encryption key
No configuration required

Customer-managed encryption key (CMEK)
Manage via [Google Cloud Key Management Service](#)

[CANCEL](#) [CREATE TOPIC](#)



Cloud Pub/Sub





DISTRIBUTE DATA TOOL

Cloud Pub/Sub

A subscription directs messages on a topic to subscribers. Messages can be pushed to subscribers immediately, or subscribers can pull messages as needed.

Subscription ID * ?

Subscription name: projects/substantial-art-374705/subscriptions/dota2_live_sub

Select a Cloud Pub/Sub topic * ▾

Delivery type ?

Pull

Push

Write to BigQuery

A variant of the push operation. Select this option if you want Pub/Sub to deliver messages directly to an existing BigQuery table. [Learn more](#)

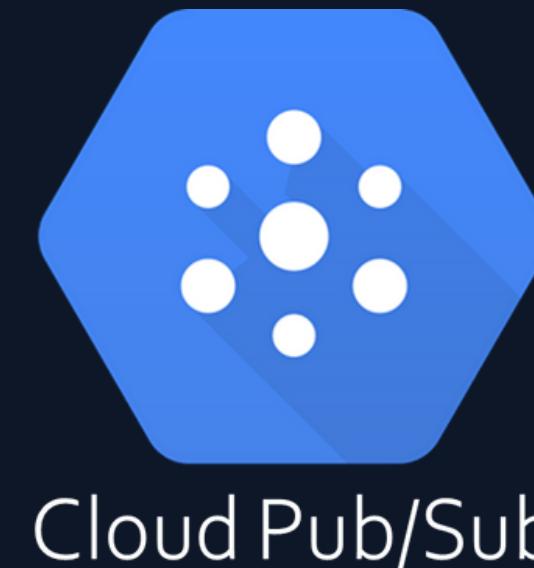
Project * [BROWSE](#)

Dataset *

Table *

Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

- Use topic schema
When enabled, the topic schema will be used when writing to BigQuery. Else, Pub/Sub writes the message bytes to a column called *data* in BigQuery.
- Write metadata
When enabled, the metadata of each message is written to additional columns in the BigQuery table. Else, the metadata is not written to the BigQuery table.
[Learn more](#)
- Drop unknown fields
When enabled along with the **Use topic schema** option, any field that is present in the topic schema but not in the BigQuery schema will be dropped. Else, messages with extra fields are not written and remain in the subscription backlog.



Cloud Pub/Sub



DISTRIBUTE DATA TOOL

Cloud DataFlow

[← Create job from template](#)

Job name * Must be unique among running jobs

Regional endpoint *

Choose a Dataflow regional endpoint to deploy worker instances and store job metadata. You can optionally deploy worker instances to any available Google Cloud region or zone by using the worker region or worker zone parameters. Job metadata is always stored in the Dataflow regional endpoint. [Learn more](#)

Dataflow template *

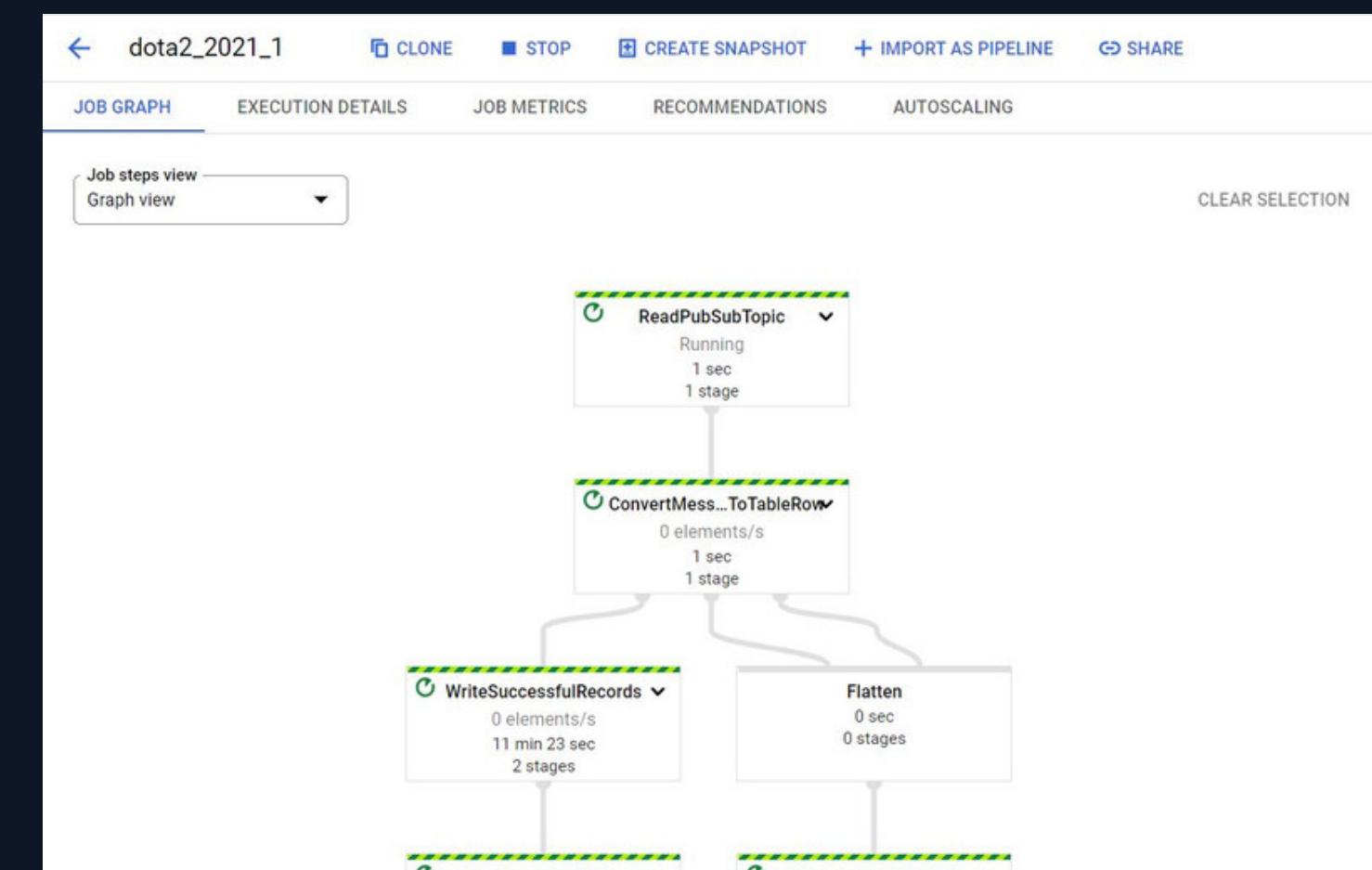
Streaming pipeline. Ingests JSON-encoded messages from a Pub/Sub topic, transforms them using a JavaScript user-defined function (UDF), and writes them to a pre-existing BigQuery table as BigQuery elements. [OPEN TUTORIAL](#)

Required parameters

Input Pub/Sub topic *

BigQuery output table * The location of the BigQuery table to write the output to. If you reuse an existing table, it will be overwritten. The table's schema must match the input JSON objects. Ex: your-project:your-dataset.your-table

Temporary location * BROWSE



DATA STORAGE

Google BigQuery

เราเลือกใช้ Google BigQuery เนื่องจากมีความสะดวกในการ setup และใช้งานได้ง่าย นอกจากนี้ตัว BigQuery ยังมี API ที่ทำให้สามารถ Connect กับ Gafana ได้โดยตรง ทำให้สะดวกในการทำ data มาแสดงผล



Google
BigQuery

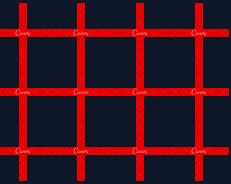
Query results

SAVE

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	source	time		author		message
1	Dota2_2021	2021-10-17T05:15:26		Just Do It		laper gengs
2	Dota2_2021	2021-10-17T05:16:04		KSY		wao guys wao
3	Dota2_2021	2021-10-17T05:16:12		KSY		do u know wao
4	Dota2_2021	2021-10-17T05:19:10		AI M		:bat: CHINA
5	Dota2_2021	2021-10-17T05:19:31		AI M		bat :bat:
6	Dota2_2021	2021-10-17T05:19:34		AI M		bat :bat:
7	Dota2_2021	2021-10-17T05:19:49		Miguel		Marci <3
8	Dota2_2021	2021-10-17T05:26:03		abhi 6ix		let's goooo spirit
9	Dota2_2021	2021-10-17T05:27:04		Sam K		Is Dota 1 still around?
10	Dota2_2021	2021-10-17T05:33:00		nayo1505		I think final gonna be lgd vs tsp...
11	Dota2_2021	2021-10-17T05:34:43		Prima Danny		Stay Down
12	Dota2_2021	2021-10-17T05:35:06		Prima Danny		rememmrber
13	Dota2_2021	2021-10-17T05:35:15		Prima Danny		Stay Down

DIAGRAM

LIVE

DOTA The International:
FINAL DAY

2021

Data
Publisher 1

Cloud Pub/Sub

subscriber 1



Table 1



2022

Data
Publisher 2

Cloud Pub/Sub

subscriber 2



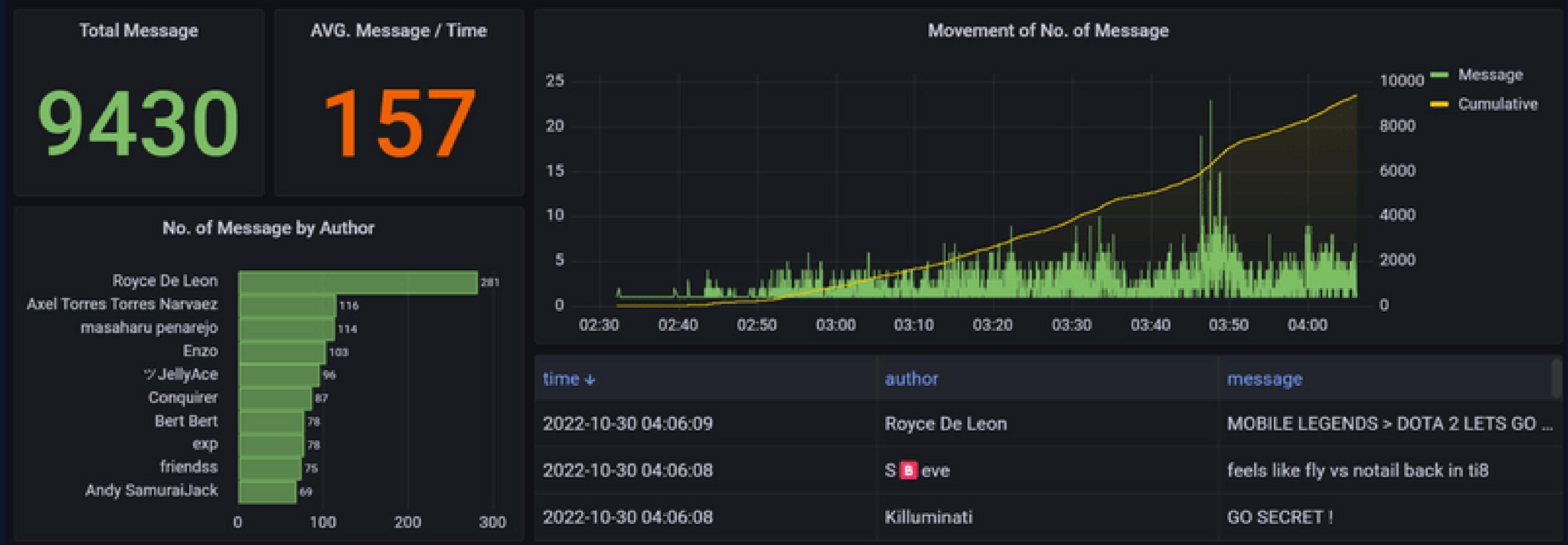
Table 2

Google
BigQuery



DATA VISUALIZATION

DOTA2 TOURNAMENT STREAMING IN YEAR 2022 as of 2022-10-30 02:32:07 2022-10-30 04:06:08





DATA VISUALIZATION

DOTA2 TOURNAMENT STREAMING IN YEAR 2021 as of 2021-10-17 05:15:26 2021-10-17 06:51:30

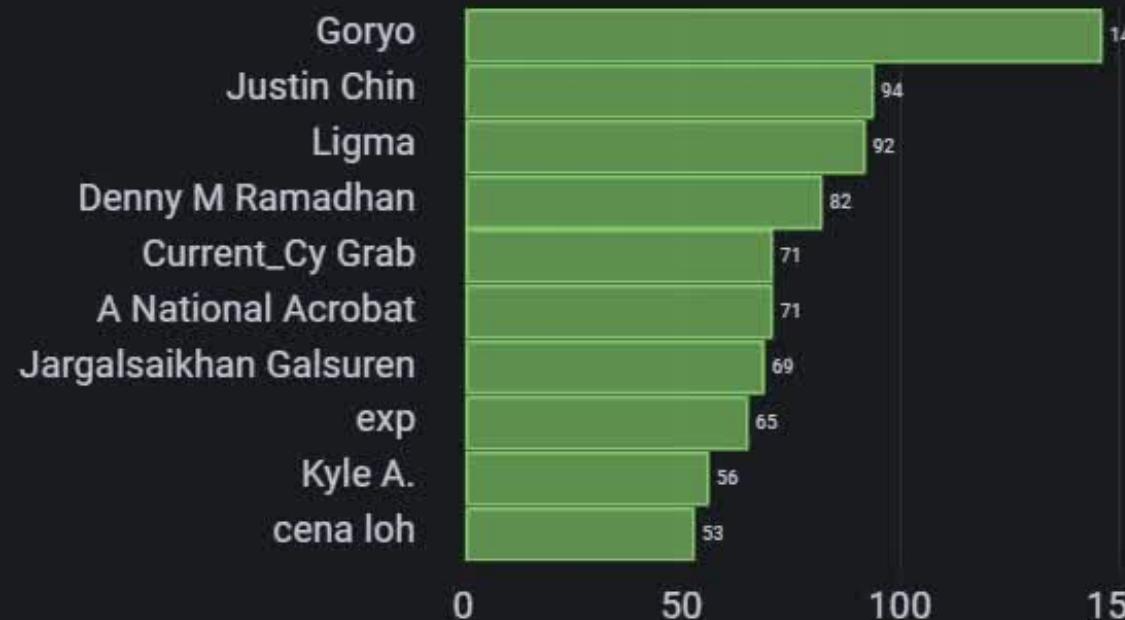
Total Message

5988

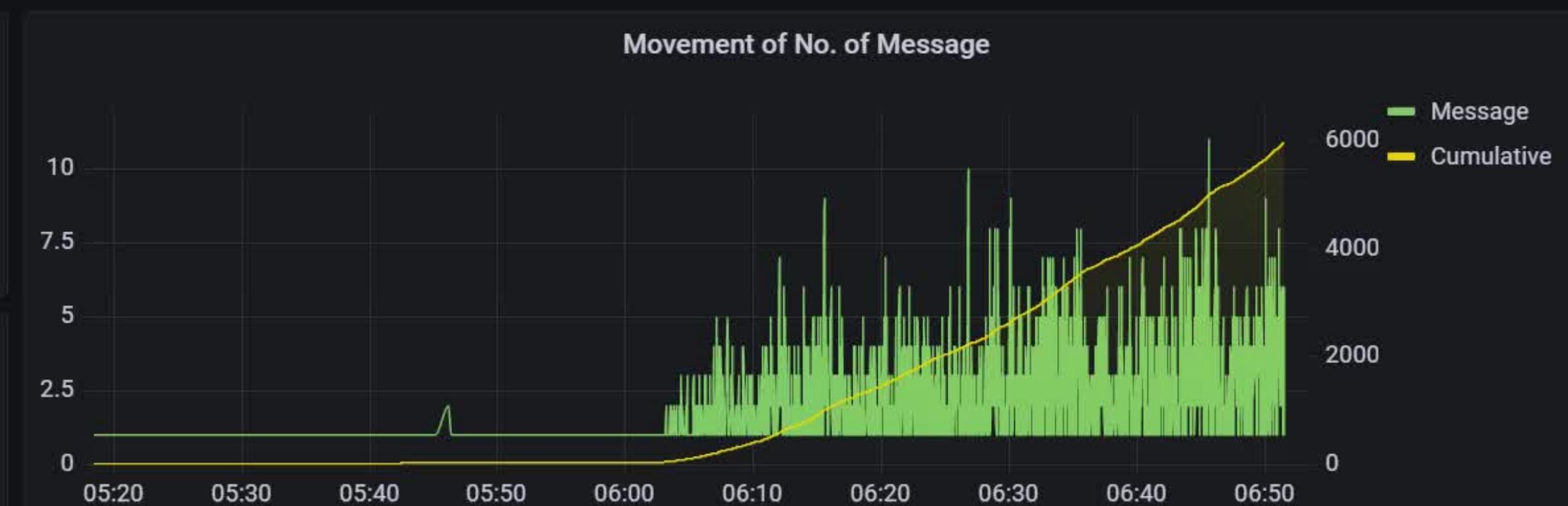
AVG. Message / Time

99.8

No. of Message by Author



Movement of No. of Message



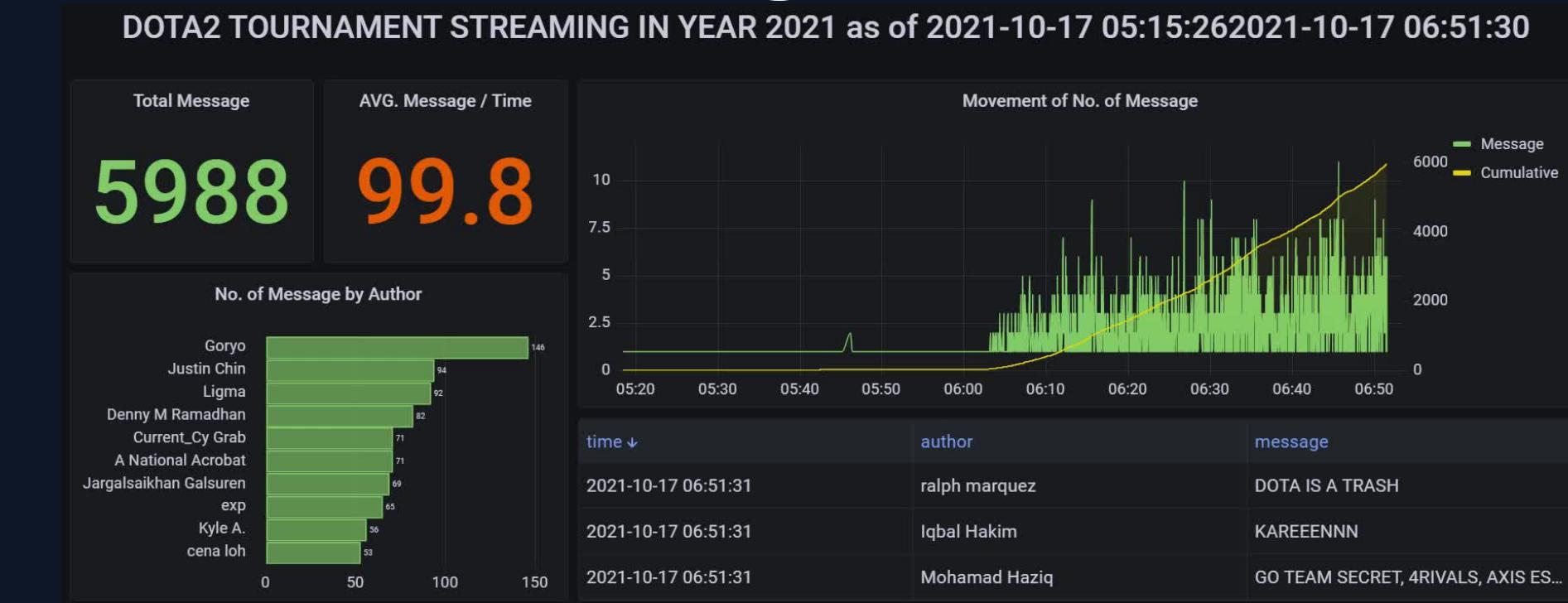
time ↓	author	message
2021-10-17 06:51:31	ralph marquez	DOTA IS A TRASH
2021-10-17 06:51:31	Iqbal Hakim	KAREENNN
2021-10-17 06:51:31	Mohamad Haziq	GO TEAM SECRET, 4RIVALS, AXIS ES...

INSIGHT & CONCLUSION

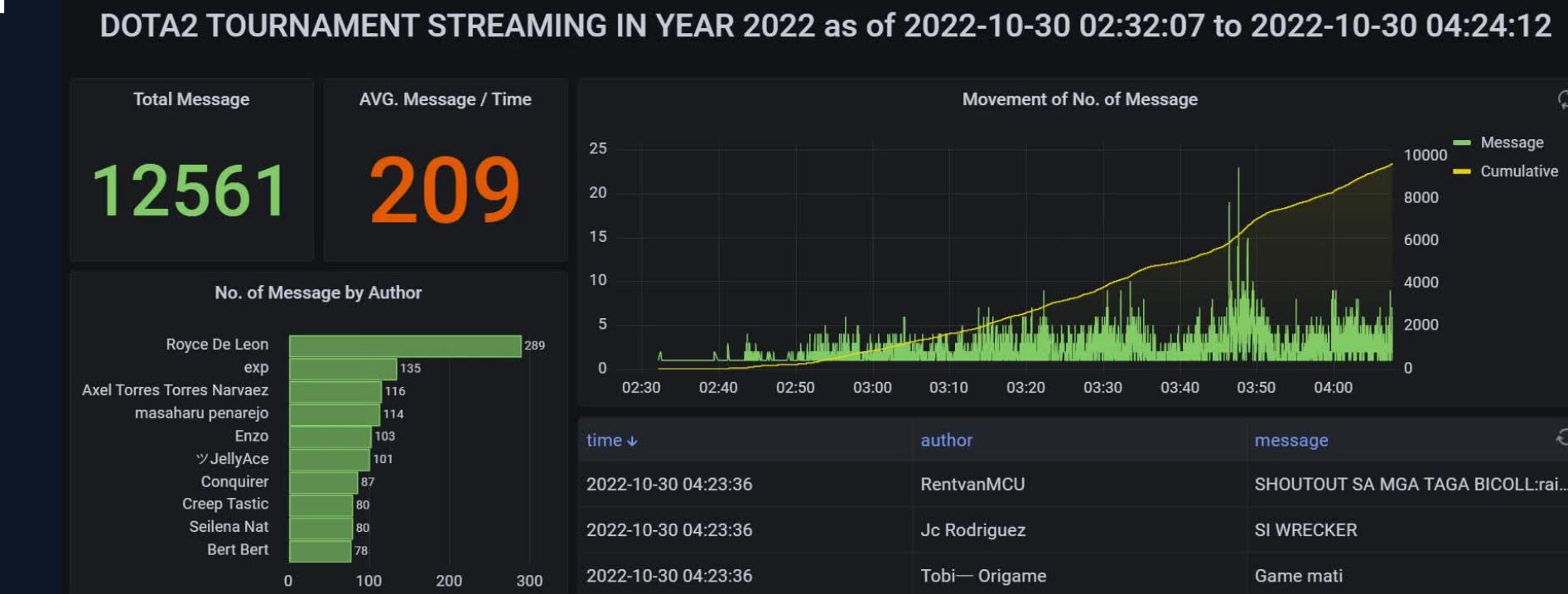
- ในช่วงเวลา **1.5 ชั่วโมง** เท่ากัน ในปี 2022 มีคอมเม้นมากกว่าในปี 2021 โดยเพิ่มขึ้นกว่า 56%
 - ในปี 2021 มีคอมเม้นกั้งหนด 6,049
 - ในปี 2022 มีคอมเม้น 9,430
- ในปี 2022 ผู้รับชมการแข่งขันออนไลน์เริ่มมาคอมเม้นใน VDO การแข่งขันตั้งแต่ 10 นาทีแรกของการเริ่มการแข่งขัน ในขณะที่ปี 2021 เริ่มมีการคอมเม้นต์เมื่อเวลาผ่านกว่า 40 นาที
- Average comment อัตราการเพิ่มขึ้นของปี 2021 ไปปี 2022 อยู่ที่ 40%
 - ในปี 2021 อยู่ที่ประมาณ 101 คอมเม้นต์/นาที
 - และในปี 2022 อยู่ที่ประมาณ 157 คอมเม้นต์/นาที



2021



2022



PROFESSIONAL ESPORT GAMERS



THANK YOU