Digital Business Group 23

邱彥慈 簡妤璇 王鏡媛 許志瑄

Outline

- Overview
- Workflow
- Architecture
- Implement
- Insight

Overview

Github link

Webpage link

Website ● - - - ⭐ 01

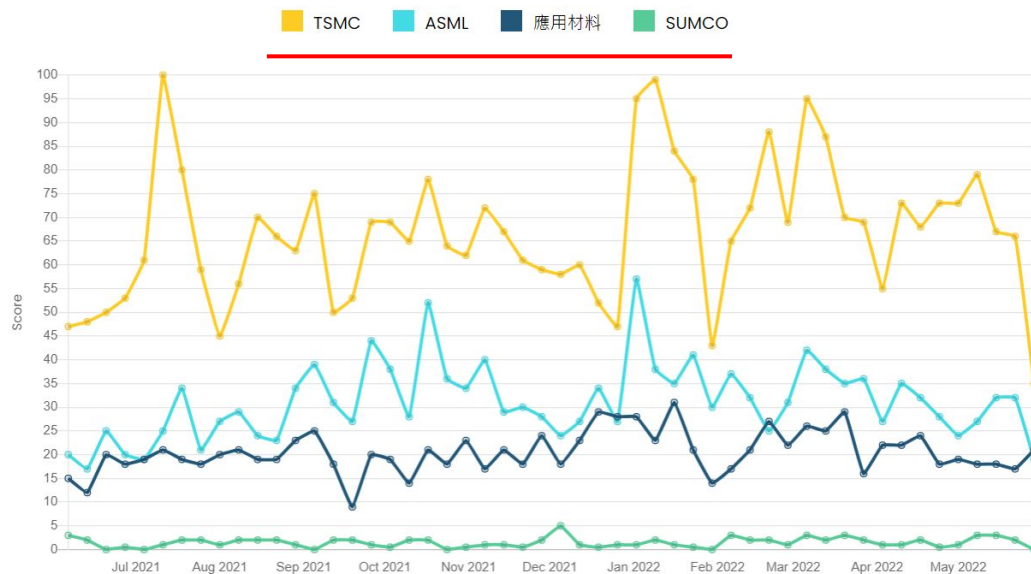Workflow ● - - - ⭐ 02

Architecture ● - - - ⭐ 03

Website – All

Website – Specific Company

Website – Detail

Website – Show Detail

Workflow

Crawler

Crawl google trend once per day using cron as scheduler

Backend

Read data got by crawler and return to frontend

Frontend

Build website base on data returned from backend
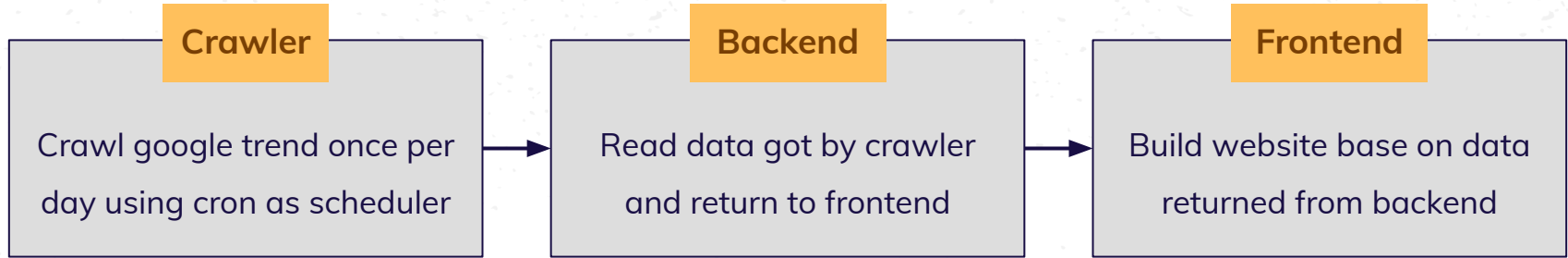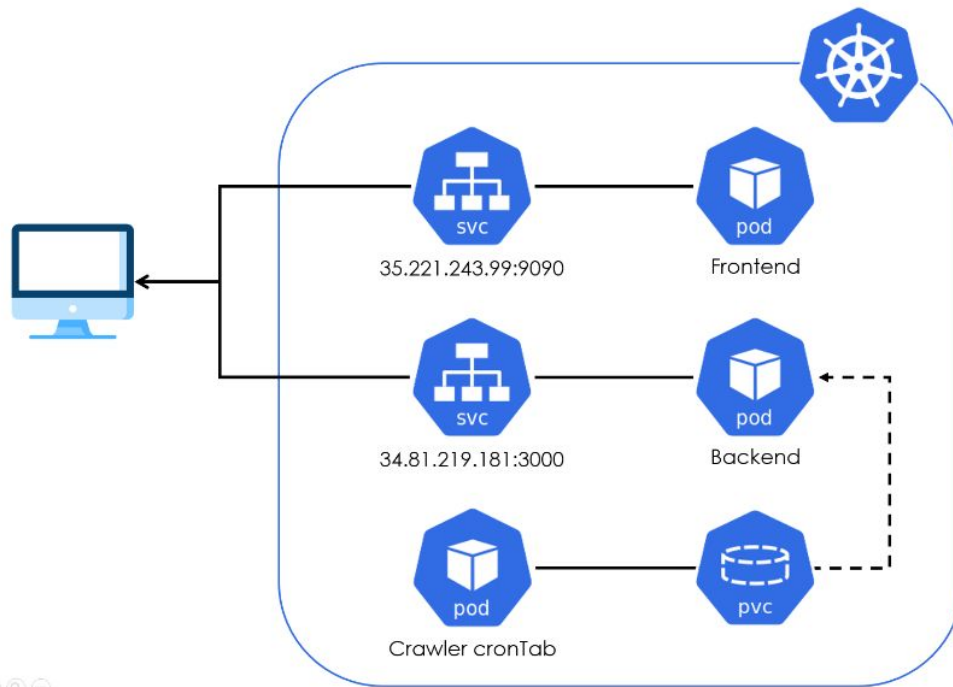
Architecture

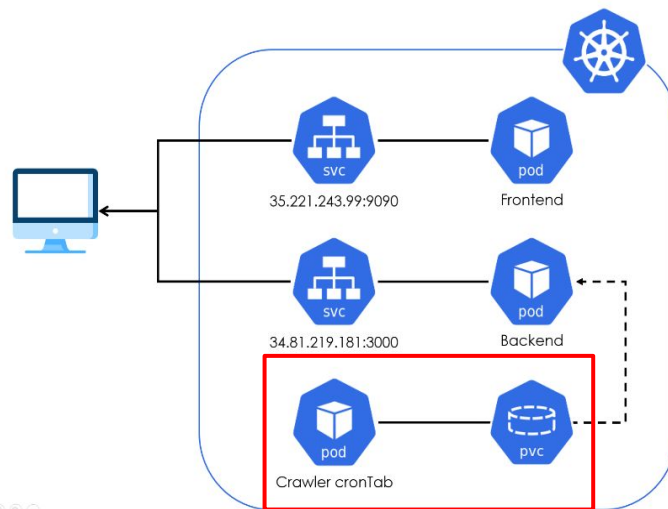Implement

Storage/ Cron

Application

- ○ Crawl data on single pod QQ
- ○ Save data to where PVC is mounted
- ○ Schedule task by cron

```
NAME                              READY   STATUS    RESTARTS
pod/crawler-pod                   1/1     Running   0
pod/flask-app-d5c747f94-84tnk     1/1     Running   0
pod/react-app-787ddfdb99-rf8tx    1/1     Running   0
```
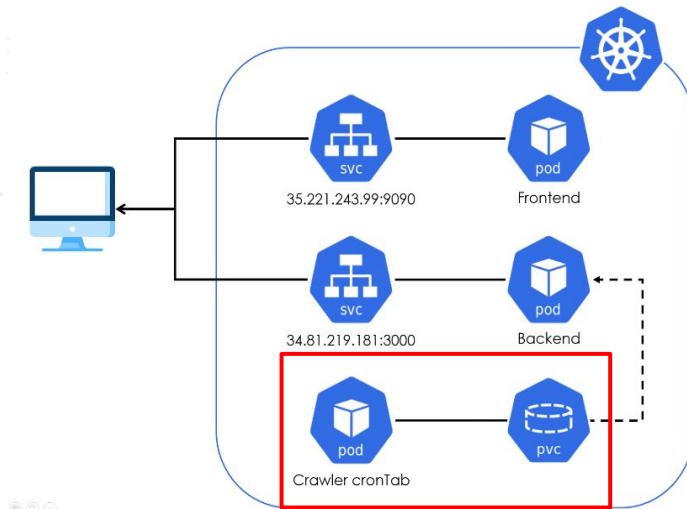
```
pvc-demo    Bound    pvc-98abbbc0-2401-4872-9d38-606fa86f0a52    30Gi    RWO    standard-rwo
```

Storage/ Cron

Problem : Fail to setup read-write-many PVC

- ○ GCEPersistentDisk is rwo / rox
- ○ In autopilot mode :
  - ■ Pods are not assigned on same node
    - ✖ NodeSelector
    - ✖ Inter-pod affinity
  - ■ Rwx storage relies on pre-provisioned such

    as filestore instance
    - ✖ Filestore
    - ✖ NFS tutorial

Storage/ Cron

Problem : Fail to setup read-write-many PVC

- ○ GCEPersistentDisk is rwo / rox

- ○ In standard mode :

  - ■ Assign pods on same node
    - ✔ NodeSelector
    - ✔ Inter-pod affinity
  - ■ Rwx storage relies on pre-provision
    as filestore instance
    - ✘ Filestore
    - ✔ NFS tutorial

```
irriiissss26268989@cloudshell:~ (nycu-lab-23)$ kubectl get pod
W0614 12:47:02.441476    4646 gcp.go:120] WARNING: the gcp auth plugin is depr
To learn more, consult https://cloud.google.com/blog/products/containers-kuber
NAME                                    READY   STATUS      RESTARTS   AGE
crawler-pod                             1/1     Running     0          28m
cronjob-crawler-27586840-rl7qm          0/1     Completed   0          7m2s
flask-app-5f495c9df8-dhzvj              1/1     Running     0          13m
nfs-client-provisioner-7c975854b8-d6fqp 1/1     Running     0          16m
nfs-pod                                 1/1     Running     0          15m
nfs-server-5555b46767-x5btr             1/1     Running     0          16m
react-app-84689b69b5-9thrb              1/1     Running     0          12m
irriiissss26268989@cloudshell:~ (nycu-lab-23)$ kubectl exec -it nfs-pod bash
kubectl exec [POD] [COMMAND] is DEPRECATED and will be removed in a future ver
W0614 12:47:09.455240    4655 gcp.go:120] WARNING: the gcp auth plugin is depr
To learn more, consult https://cloud.google.com/blog/products/containers-kuber
root@nfs-pod:/sample_crawler# cd history
root@nfs-pod:/sample_crawler/history# ls -l
total 32
-rw-r--r-- 1 root root   577 Jun 14 12:40 crawler_data.csv
-rw-r--r-- 1 root root    46 Jun 14 12:40 crawler_result.csv
-rw-r--r-- 1 root root   577 Jun 14 12:47 data.csv
-rw-r--r-- 1 root root    16 Jun 14 12:30 index.html
drwx------ 2 root root 16384 Jun 14 12:30 lost+found
```

Crawler

Cron

- Execute trend.py once per day on crawler-pod
- Simple script to cp files QQ

Cronjob

- Complete nltk.download() in build-up image stage

```
src="nfs-pod"
for pod in $(kubectl get pod -o name)
do
    if [[ $pod == *"flask-app"* ]]; then
        stripped_pod="${pod:4}"
        kubectl cp $src:/sample_crawler/history/data.csv ./data.csv
        kubectl cp ./data.csv $stripped_pod:/server/data.csv
    fi
done
```

```
NAME                              READY   STATUS      RESTARTS   AGE
crawler-pod                       1/1     Running     0          20m
cronjob-crawler-27572633-2mxxp    0/1     Completed   0          2m27s
cronjob-crawler-27572634-2dt56    0/1     Completed   0          87s
cronjob-crawler-27572635-p6dwm    0/1     Completed   0          27s
flask-app-d5c747f94-84tnk         1/1     Running     0          20h
react-app-787ddfdb99-rf8tx        1/1     Running     0          20h
```

CI

trigger docker building Build up docker file #1

⌂ Summary

Jobs

docker ⟳

docker
succeeded 5 days ago in 1m 15s

> ✓ Set up job
> ✓ Checkout
> ✓ Set up Docker Buildx
> ✓ Login to DockerHub
> ✓ Build and push
> ✓ Post Build and push
> ✓ Post Login to DockerHub
> ✓ Post Set up Docker Buildx
> ✓ Post Checkout
> ✓ Complete job

```
name: Build up docker file
on:
  workflow_dispatch:
  push:
    paths:
      - 'cron/Dockerfile'
```

trigger pytest Tests #2

⌂ Summary

Jobs

✓ **build**

build
succeeded 5 days ago in 19s

🔍 Search logs

> ✓ Set up job
> ✓ Initialize containers
> ✓ Run actions/checkout@v2
∨ ✓ Test with pytest

```
1  ▶ Run pytest -v cron/test/trend_test.py
4  =========================== test session starts ===========================
5  platform linux -- Python 3.7.13, pytest-7.1.2, pluggy-1.0.0 -- /usr/local/bin/pyt
6  cachedir: .pytest_cache
7  rootdir: /__w/TSMC_NYCU_FINAL_23/TSMC_NYCU_FINAL_23
8  plugins: cov-3.0.0
9  collecting ... collected 3 items
10
11  cron/test/trend_test.py::test_get_trend_cols PASSED                    [ 33%]
1  cron/test/trend_test.py::test_get_trend_lens PASSED                    [ 66%]
1  cron/test/trend_test.py::test_get_trend_type PASSED                    [100%]
```
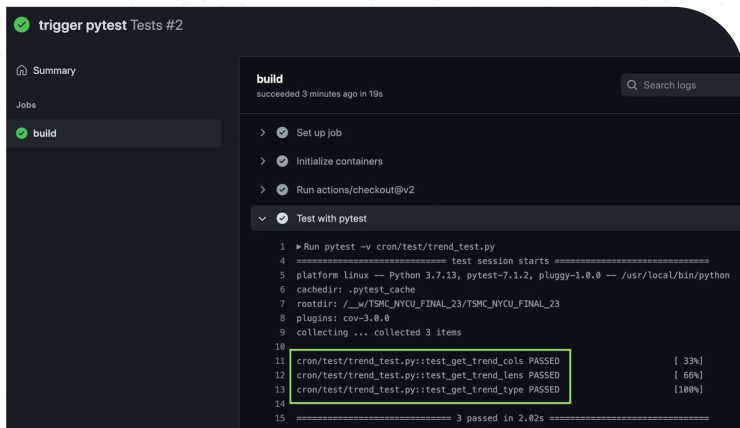
```
name: Tests
on:
  workflow_dispatch:
  push:
    paths:
      - 'cron/**.py'
```

Testing

Coverage for **trend.py**: 82%

22 statements | 18 run | 4 missing | 0 excluded
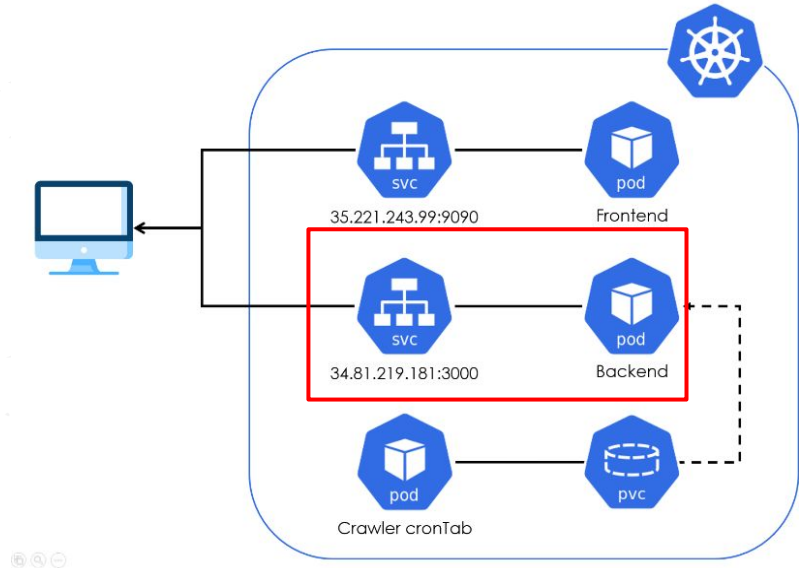

✓ Assert columns equalTo keywords

✓ Consistent row length

✓ Assert all (is_numeric_dtype)

Backend

- Use flask framework of python
- Read the data gotten by crawler
- Build API for the communication of backend server and frontend
- Two types of API
  - Get the information of all companies
  - Get the information of specific company

Frontend

- Use React.js framework
- Get information using API and draw the charts according to the information
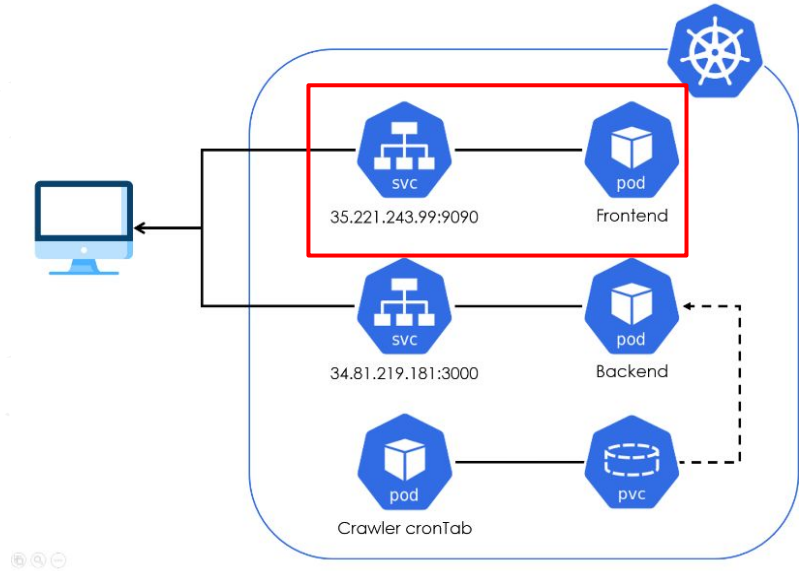
Deployment

Backend

- Use a service to connect the backend pod
- Set the service type to LoadBalancer to get external IP

Frontend

- Use a service to connect the frontend pod
- Reach the backend server through the external IP of backend service
- Set the service type to LoadBalancer to get external IP
- Reach the website through the external IP of frontend service
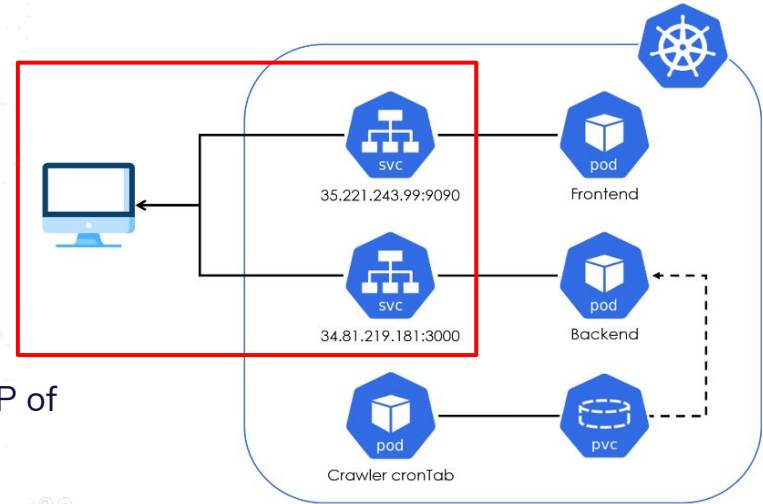
Insight

數據分析

- 整體聲量趨勢：

  ○ TSMC > ASML >
    應用材料 > SUMCO

- 分析造成聲量高峰的
  可能原因：

  ○ 疫情相關新聞
  ○ 法說會
  ○ 股價波動
  ○ 其他重大事件
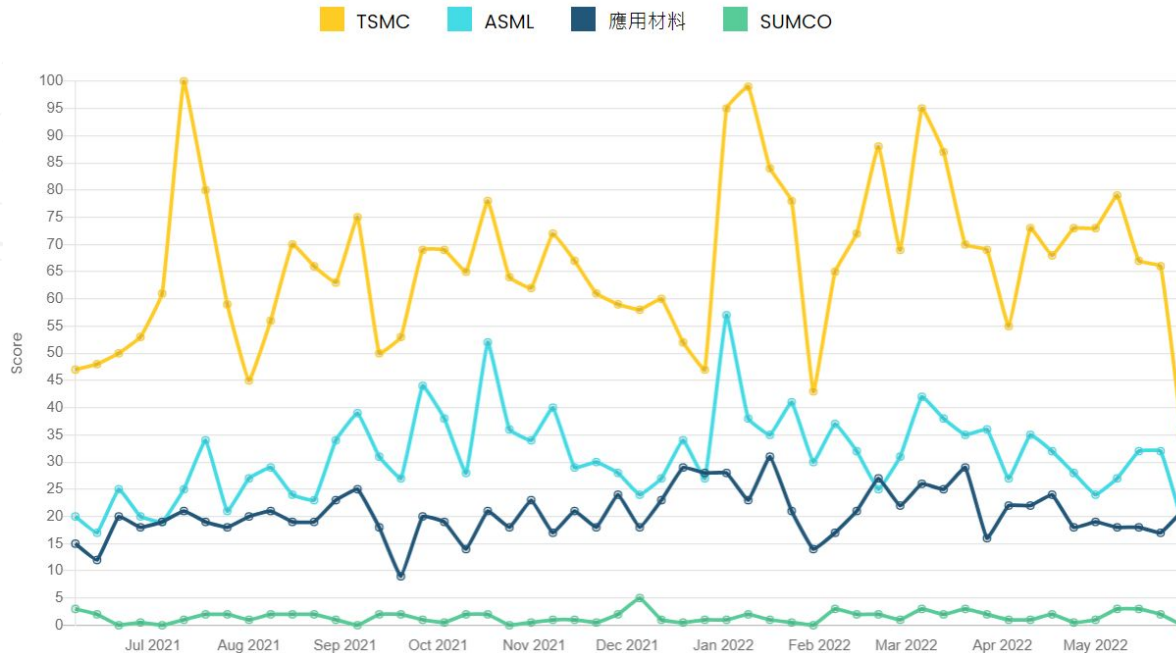
疫情相關新聞聲量

- 2021/07/11 TSMC 有一個明顯的聲量高峰，發現可能是因為有疫苗相關的新聞，在疫情嚴峻的情勢下，因此達到聲量高峰。

FT Financial Times

TSMC and Foxconn join forces to secure vaccines for Taiwan

TSMC and Foxconn join forces to secure vaccines for Taiwan. Tech champions step in after sovereignty prevented Taipei from buying BioNTech Covid jabs.
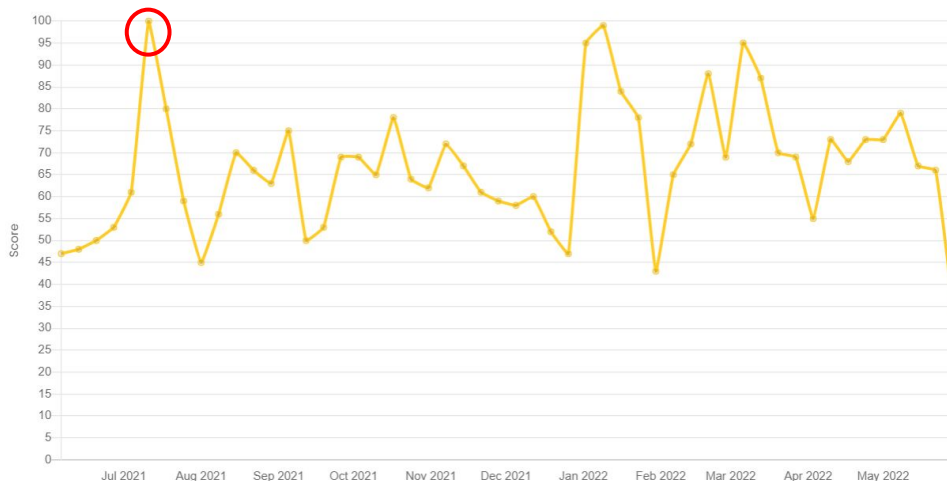
2021年7月11日
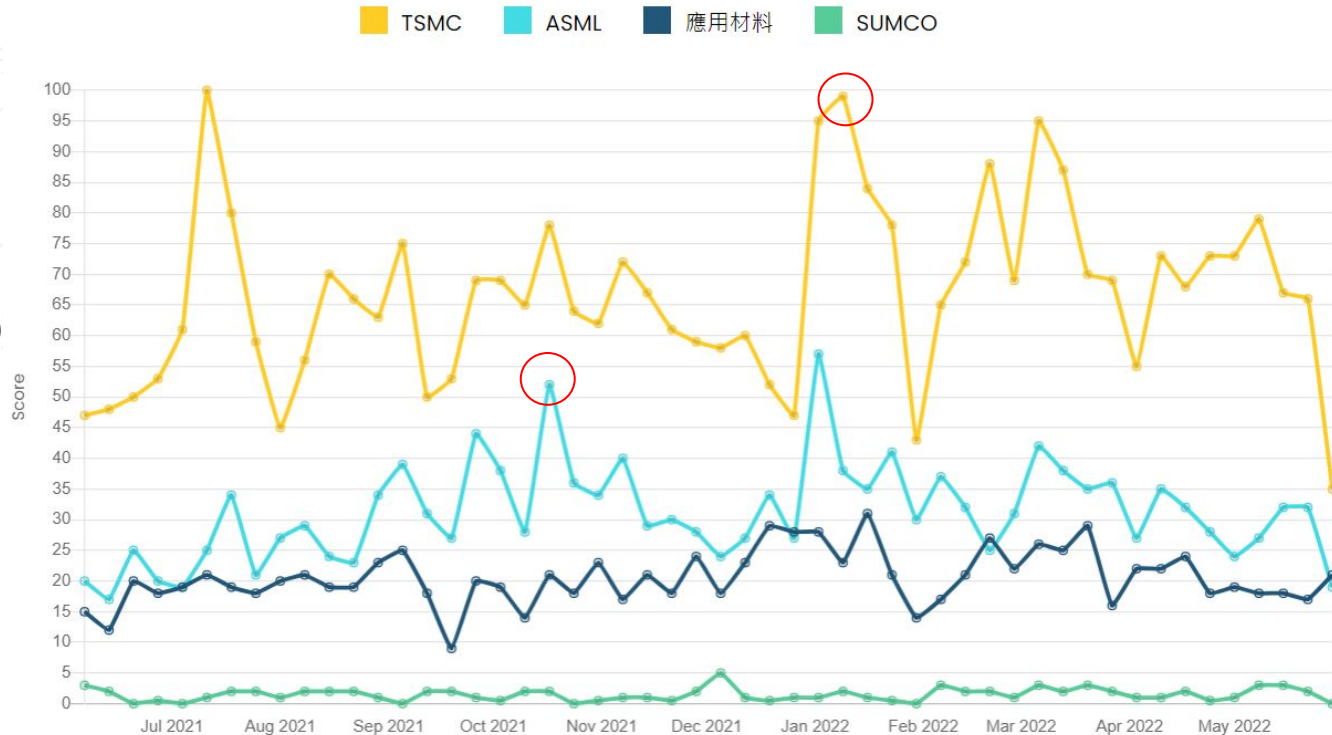
聯合報

上海復星宣布：向台積電、鴻海簽協議 銷售1000萬劑疫苗

本公司控股子公司復星實業分別與（1）台積電（即買方一）、（2）鴻海、永齡基金會（即買方二）以及裕利醫藥（有進口疫苗資質的醫藥公司）簽訂《銷售協議》，復星實業將向...

2021年7月11日

法說會聲量

- 台積電與艾司摩爾的聲量，在發布財報及法說會前後各有一個高點。

- ASML 2021/10/20發布2021年第3季財報

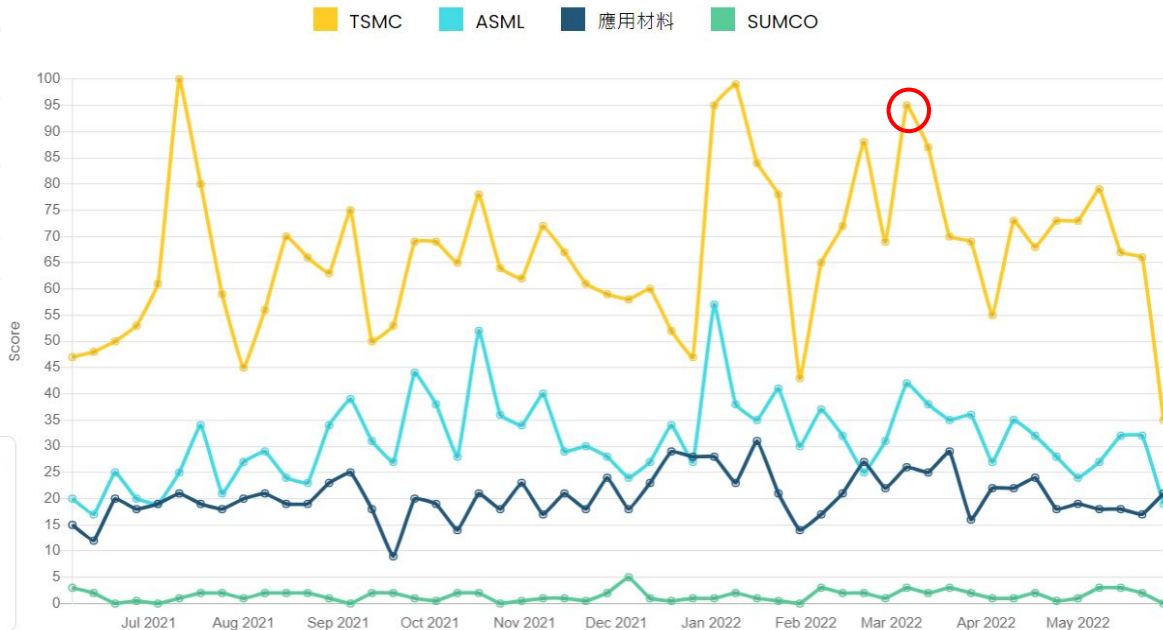- TSMC 2022/01/13線上舉行第四季法人說明會

股價影響聲量

- 當股價有特別波動時，可能會影響聲量

- TSMC 2022/03/06

鉅亨

台積電跳水重挫逾3% 市值跌破15兆元

烏俄戰火持續延燒，傳美、歐擬禁運俄國石油，美股期貨電子盤重挫，台股今(7) 日股匯雙殺，晶圓代工龍頭台積電(2330-TW)(TSM-US) 持續下殺，股價跳水重挫超過3%，最低下...

2022年3月6日

其他重大事件造成聲量

- 艾司摩爾柏林廠於 2022/01/03 發生火災

鉅亨

半導體設備商ASML德國廠傳祝融之災損失仍在評估中| Anue 鉅 ...

ASML 是晶片製造商關鍵設備的主要供應商，由於全球半導體短缺，目前盡全力生產，任何出貨延遲都有可能影響到想擴增產能的客戶。 ASML 傳出火災的工廠前身是Berliner Glas，...
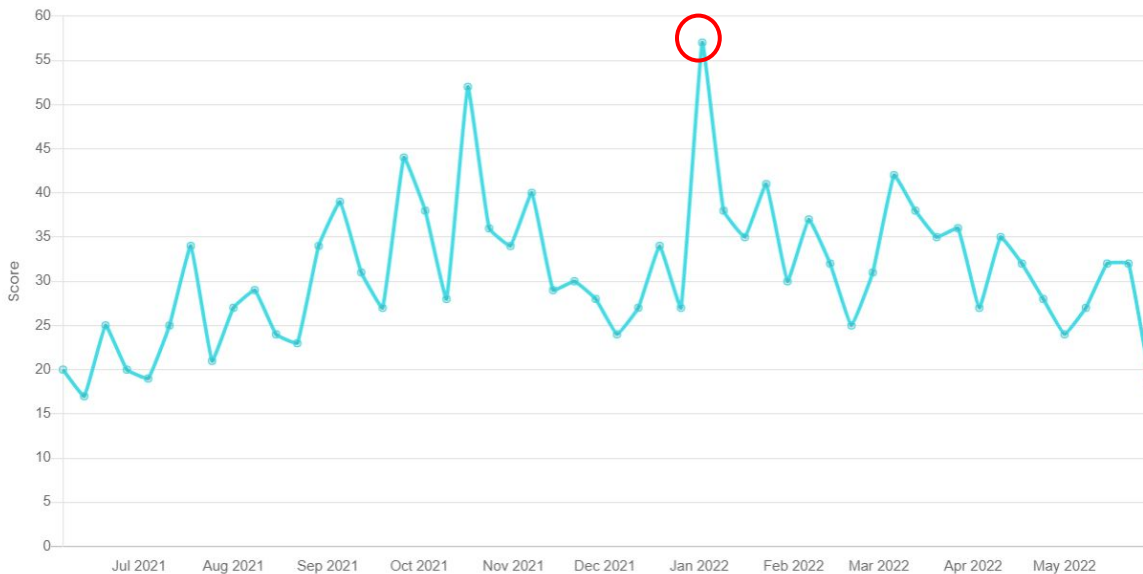
2022年1月3日

Reuters

ASML reports fire at its Berlin factory

ASML, a key equipment supplier to all major computer chip manufacturers, is currently operating at full capacity due to the global semiconductor shortage. Any...

2022年1月3日

分工

邱彥慈
- Code：Crawler、CI、Testing
- Report：投影片製作 (Code)、報告

簡妤璇
- Code：Frontend、Backend、Deployment
- Report：投影片製作 (Code)、打講稿 (Code)

王鏡媛
- Report：投影片製作 (Insight)、報告

許志瑄
- Report：投影片製作 (Insight)

Thank You