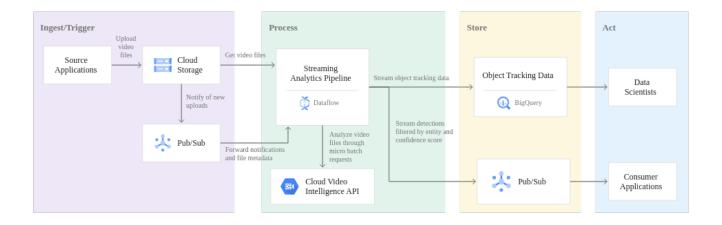
Creando u PIPELINE para Analitica de Video con : <a href="https://cloud.google.com/architecture/building-a-streaming-video-analytics-pipeline">https://cloud.google.com/architecture/building-a-streaming-video-analytics-pipeline</a>

https://cloud.google.com/video-intelligence/docs/people-detection#video\_detect\_person\_gcs-python

https://cloud.google.com/video-intelligence/docs/feature-person-detection

- BigQuery
- Cloud Storage
- Dataflow
- Video Intelligence API
- Pub/Sub
- Cloud Build

## Arquitectura de la solucion



#### Donde:

- Las aplicaciones cliente captura o suben los archivos de video de origen a un bucket de Cloud Storage.
- Para cada carga de archivos, el sistema notifica de forma automática al cliente publicando un mensaje en Pub/Sub. (Es como Kafka / RabbitMQ)

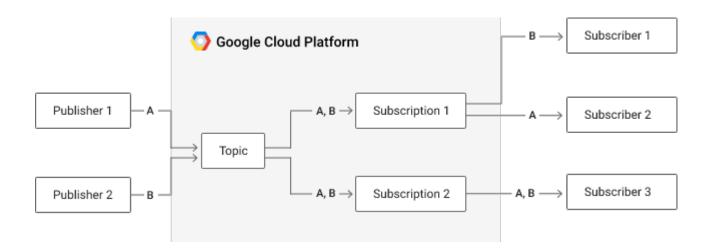
## •Concepto:

**Pub/Sub** se usa para las canalizaciones de integración de datos y estadísticas de transmisión a fin de transferir y distribuir datos. Es igual de efectivo como un middleware orientado a la mensajería para la integración de servicios o como una cola a fin de paralelizar las tareas.

Pub/Sub te permite crear sistemas de <u>productores</u> y <u>consumidores</u> de eventos, llamados editores y suscriptores. Los publicadores se comunican con los suscriptores de forma asíncrona mediante la transmisión de eventos, en lugar de llamadas de procedimiento remoto (RPC) síncronas

#### donde usarlo:

- · Interacción del usuario y eventos del servidor de transferencia.
- Distribución de eventos en tiempo real
- Replicar datos entre bases de datos
- Procesamiento y flujos de trabajo paralelos
- Bus de eventos empresariales
- Transmisión de datos desde aplicaciones, servicios o dispositivos de IoT.
- Actualización de cachés distribuidas
- Balanceo de cargas para la confiabilidad

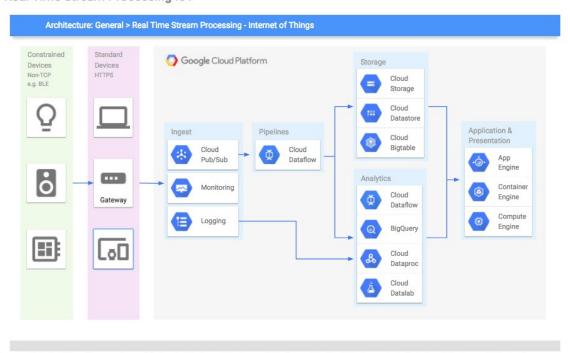


## Que es DATAFLOW ? (es como Spark)

Procesamiento de datos de streaming y por lotes unificado, rápido, rentable y sin servidor

- Servicio totalmente gestionado de procesamiento de datos
- Aprovisionamiento y gestión automáticos de recursos de procesamiento
- Autoescalado horizontal de recursos de trabajadores para sacarles el máximo partido
- Innovación en software libre motivada por la comunidad mediante el SDK de Apache Beam
- Procesamiento exacto, fiable y uniforme
- analíticas de datos de streaming con rapidez. Dataflow permite desarrollar flujos de procesamiento de datos de streaming con menos latencia de datos de manera rápida y más sencilla.

General > Real Time Stream Processing IoT



- Para cada notificación nueva, la canalización de **Dataflow** realiza las siguientes acciones:
- Lee los metadatos del archivo del mensaje de Pub/Sub.
- Carga el archivo en la memoria y lo divide en segmentos de 5 segundos.
- Envía cada segmento a la API de Video Intelligence, que luego muestra anotaciones.
- Almacena todas las anotaciones en una tabla de **BigQuery**.
- Envía mensajes de salida a Pub/Sub para los segmentos de video que contienen anotaciones significativas (por ejemplo, todos los segmentos que contienen personas), como se especifica en los parámetros personalizados de la canalización.
- Una aplicación de consumidor lee los mensajes de salida de Pub/Sub para mostrar las anotaciones filtradas.

#### Primero Habilitamos los siguientes servicios :

```
gcloud services enable dataflow.googleapis.com \
containerregistry.googleapis.com \
videointelligence.googleapis.com
```

## Listado de Variable de trabajo

gcloud auth list
gcloud config list project

```
spideroriol@cloudshell:~ (my-project-91798-video-ai)$ gcloud auth list
Credentialed Accounts

ACTIVE: *
ACCOUNT: spideroriol@gmail.com

To set the active account, run:
    $ gcloud config set account `ACCOUNT`

spideroriol@cloudshell:~ (my-project-91798-video-ai)$ [
```

```
spideroriol@cloudshell:~ (my-project-91798-video-ai)$ gcloud config list project
[core]
project = my-project-91798-video-ai

Your active configuration is: [cloudshell-19199]
spideroriol@cloudshell:~ (my-project-91798-video-ai)$
```

## Definiendo de Variable de trabajo

```
export REGION=us-central1
export PROJECT=$(gcloud config get-value project)
export GCS_NOTIFICATION_TOPIC="gcs-notification-topic"
export GCS_NOTIFICATION_SUBSCRIPTION="gcs-notification-subscription"
export VIDEO_CLIPS_BUCKET=${PROJECT}_videos
export BIGQUERY_DATASET="video_analytics"
export OBJECT_DETECTION_TOPIC="object-detection-topic"
export OBJECT_DETECTION_SUBSCRIPTION="object-detection-subscription"
export ERROR_TOPIC="error-topic"
export ERROR_SUBSCRIPTION="error-subscription"
export DATAFLOW_TEMPLATE_BUCKET=${PROJECT}_dataflow_template_config
```

## Clonamos un repositorio de Google con los templates

drwxr-xr-x 3 spideroriol spideroriol 4096 Mar 20 18:53 src

```
git clone https://github.com/GoogleCloudPlatform/dataflow-video-analytics.git

spideroriol@cloudshell:~/dataflow-video-analytics (my-project-91798-video-ai)$ ls -l
total 76
-rw-r--r-- 1 spideroriol spideroriol 27 Mar 20 18:53 build.gradle
-rw-r--r-- 1 spideroriol spideroriol 1100 Mar 20 18:53 _config.yml
-rw-r--r-- 1 spideroriol spideroriol 1100 Mar 20 18:53 CONTRIBUTING.md

drwxr-xr-x 2 spideroriol spideroriol 4096 Mar 20 18:53 diagram

drwxr-xr-x 3 spideroriol spideroriol 4096 Mar 20 18:53 gradle
-rwxr-xr-x 1 spideroriol spideroriol 5764 Mar 20 18:53 gradlew
-rw--r---- 1 spideroriol spideroriol 3056 Mar 20 18:53 gradlew.bat
-rw-r----- 1 spideroriol spideroriol 11307 Mar 20 18:53 LICENSE
-rw-r----- 1 spideroriol spideroriol 10038 Mar 20 18:53 README.md

drwxr-xr-x 2 spideroriol spideroriol 4096 Mar 20 18:53 scripts
-rw-r----- 1 spideroriol spideroriol 373 Mar 20 18:53 settings.gradle
```

#### **I PARTE**

#### **PUB/SUB**

Crea una notificación de Pub/

gcloud pubsub topics create \${GCS\_NOTIFICATION\_TOPIC}

Crea una suscripcion de Pub la el TOPICO creado

gcloud pubsub subscriptions create \${GCS\_NOTIFICATION\_SUBSCRIPTION} \
--topic=\${GCS\_NOTIFICATION\_TOPIC}

Se crea un bucket para almacenar los clips de video de entrada

gsutil mb -c standard -l \${REGION} gs://\${VIDEO\_CLIPS\_BUCKET}

spideroriol@cloudshell:~/dataflow-video-analytics (my-project-91798-video-ai)\$ gsutil mb -c standard -l \${REGION} gs://\${VIDEO\_CLIPS\_BUCKET} Creating gs://my-project-91798-video-ai\_videos/... spideroriol@cloudshell:~/dataflow-video-analytics (my-project-91798-video-ai)\$ gsutil ls gs://my-project-91798-videoai\_videos/ gs://storage\_video01/ spideroriol@cloudshell:~/dataflow-video-analytics (my-project-91798-video-ai)\$ [

## Cree una notificación de Pub/Sub para el bucket

gsutil notification create -t \${GCS\_NOTIFICATION\_TOPIC} \
-f json gs://\${VIDEO\_CLIPS\_BUCKET}

#### **BIGQUERY**

Crea un conjunto de datos y una tabla de BigQuery

En Cloud Shell, crea un conjunto de datos de BigQuery:

bq mk -d --location=US \${BIGQUERY\_DATASET}

Crea una tabla de BigQuery:

```
bq mk -t \
--schema src/main/resources/table_schema.json \
--description "object_tracking_data" \
${PROJECT}:${BIGQUERY_DATASET}.object_tracking_analysis
```

```
spideroriol@cloudshell:~/dataflow-video-analytics/src/main/resources (my-project-91798-video-ai)$ cat
table_schema.json
[
  "mode": "REQUIRED",
  "name": "file_name",
  "type": "STRING"
 },
  "mode": "REQUIRED",
  "name": "entity",
  "type": "STRING"
 },
  "fields": [
    "name": "processing_timestamp",
    "type": "STRING"
   },
    "name": "timeOffset",
    "type": "STRING"
   },
    "name": "confidence",
    "type": "FLOAT"
   },
   {
    "name": "left",
    "type": "FLOAT"
   },
    "name": "top",
    "type": "FLOAT"
   },
    "name": "right",
    "type": "FLOAT"
   },
    "name": "bottom",
    "type": "FLOAT"
   }
  ],
  "mode": "REPEATED",
  "name": "frame_data",
  "type": "RECORD"
}
]
```

#### **II PARTE**

Crea un tema y una suscripción de Pub/Sub para los resultados filtrados gcloud pubsub topics create \${OBJECT\_DETECTION\_TOPIC}

Crea una suscripción a Pub/Sub:

gcloud pubsub subscriptions create \${OBJECT\_DETECTION\_SUBSCRIPTION} \
--topic=\${OBJECT\_DETECTION\_TOPIC}

Crea un tema y una suscripción de Pub/Sub para detectar errores gcloud pubsub topics create \${ERROR\_TOPIC}

## Crea una suscripción a Pub/Sub:

gcloud pubsub subscriptions create \${ERROR\_SUBSCRIPTION} \
--topic=\${ERROR\_TOPIC}

#### III PARTE

Ejecuta el trabajo de canalización de Dataflow

En Cloud Shell, compila el código de la canalización de Apache Beam gradle build

Crea una imagen de Docker para la plantilla de Dataflow Flex

```
gcloud auth configure-docker
gradle jib \
    --image=gcr.io/${PROJECT}/dataflow-video-analytics:latest
```

```
Built and pushed image as gcr.io/my-project-91798-video-ai/dataflow-video-analytics

Executing tasks:

[==========] 100.0% complete

Deprecated Gradle features were used in this build, making it incompatible with Gradle 7.0.
Use '--warning-mode all' to show the individual deprecation warnings.
See https://docs.gradle.org/6.0/userguide/command_line_interface.html#sec:command_line_warnings

BUILD SUCCESSFUL in 1m 19s
3 actionable tasks: 3 executed
spideroriol@cloudshell:~/dataflow-video-analytics (my-project-91798-video-ai)$
```

Crea un bucket de Cloud Storage para almacenar la plantilla de Dataflow Flex:

```
gsutil mb -c standard -l ${REGION} \
gs://${DATAFLOW_TEMPLATE_BUCKET}
```

Sube el archivo de configuración JSON de la plantilla al bucket:

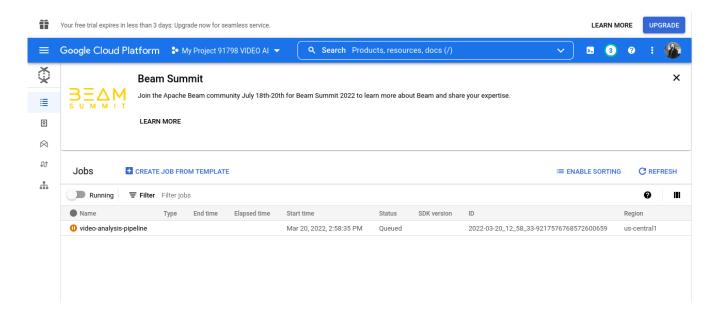
```
cat << EOF | gsutil cp -
gs://${DATAFLOW_TEMPLATE_BUCKET}/dynamic_template_video_analytics.json
{
    "image":
    "gcr.io/${PROJECT}/dataflow-video-analytics:latest",
    "sdk_info": {"language": "JAVA"}
}
EOF</pre>
```

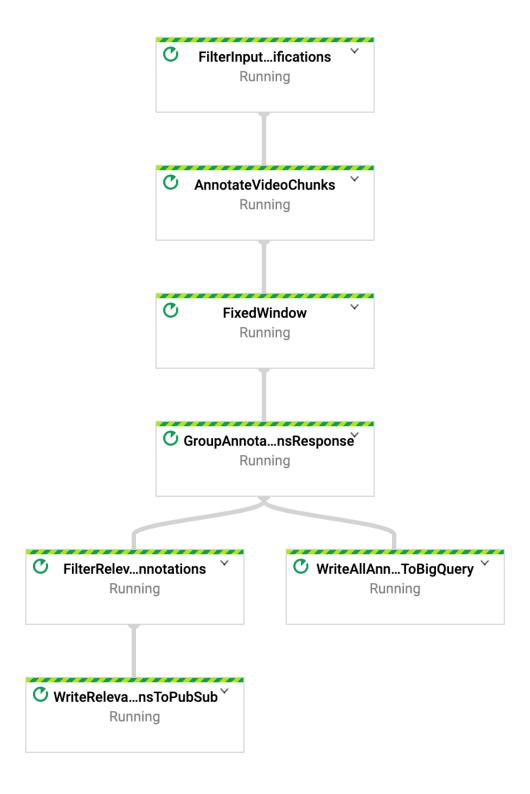
## Ejecuta la canalización de Dataflow

```
gcloud beta dataflow flex-template run "video-analysis-pipeline" \
--project=${PROJECT} \
--region=${REGION} \
--template-file-gcs-location=gs://${DATAFLOW_TEMPLATE_BUCKET}/
dynamic_template_video_analytics.json \
--
parameters=^-^autoscalingAlgorithm="THROUGHPUT_BASED"~numWorkers=5~maxNumWorkers=5~workerMachineType=n1-highmem-4\
~inputNotificationSubscription=projects/${PROJECT}/subscriptions/$
{GCS_NOTIFICATION_SUBSCRIPTION}\
~outputTopic=projects/${PROJECT}/topics/${BJECT_DETECTION_TOPIC}\
~errorTopic=projects/${PROJECT}/topics/${ERROR_TOPIC}\
~features=OBJECT_TRACKING~entities=window, person~confidenceThreshold=0.9~windowInterval=1\
~tableReference=${PROJECT}:${BIGQUERY_DATASET}.object_tracking_analysis\
~streaming=true
```

```
~streaming=true
job:
    createTime: '2022-03-20T19:58:35.283518Z'
    currentStateTime: '1970-01-01T00:00:00Z'
    id: 2022-03-20_12_58_33-9217576768572600659
    location: us-central1
    name: video-analysis-pipeline
    projectId: my-project-91798-video-ai
    startTime: '2022-03-20T19:58:35.283518Z'
spideroriol@cloudshell:~/dataflow-video-analytics (my-project-91798-video-ai)$
```

# Nos Valos a Dataflow - JOBS





La canalización se está ejecutando y esperando recibir las notificaciones de entrada de Pub/Sub. Aún no se procesan datos.

#### IV PARTE Sube archivos de video de muestra

1. En Cloud Shell, instala ffmpeg:

```
sudo apt update
sudo apt install -y ffmpeg
```

2. Descarga los archivos de video originales:

```
gsutil -m cp gs://dataflow-video-analytics-test-clips/*.mp4 .
```

3. Divide los archivos originales en segmentos de 5 segundos:

```
for file in *.mp4
do
ffmpeg -i "$file" -codec:a aac -ac 2 -ar 48k -c copy \
-movflags faststart -f segment -segment_format mpegts \
-segment_time 5 "${file%.*}~"%1d.mp4
done
```

```
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~2.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~3.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~5.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~5.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~5.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~6.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~7.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur~8.mp4' for writing
[segment @ 0x55ca9957c780] Opening 'gbikes_dinosaur-8.mp4' for writing
[segment @ 0x5ca9957c780] Opening 'gbikes_dinosaur-8.mp4' for
```

# Sube los segmentos de video a Cloud Storage:

gsutil -m cp \*~\*.mp4 gs://\${VIDEO\_CLIPS\_BUCKET}/

#### **NOTA**

Cuando gsutil sube cada segmento de video, Cloud Storage:

- envía automáticamente una notificación a **Pub/Sub**.
- Pub/Sub reenvía la notificación al trabajo de **Dataflow**.
- En la página web del trabajo de **Dataflow**, puedes ver que el trabajo procesó los archivos de video y recopiló las anotaciones de la API de Video Intelligence para esos archivos.

# **V PARTE - BIGQUERY**

SELECT file\_name, entity
FROM `video\_analytics.object\_tracking\_analysis`
WHERE entity like 'bicycle%'
OR entity like 'person%'
OR entity like 'cat%'
GROUP by file\_name, entity

#### file name entity

- 1 gbikes\_dinosaur.mp4 bicycle
- 2 cat.mp4 cat
- 3 gbikes\_dinosaur.mp4 bicycle wheel
- 4 gbikes\_dinosaur.mp4 person

3. Ejecuta la siguiente consulta para mostrar todos los objetos detectados en los clips de video y los niveles de confianza asociados con la detección:

SELECT file\_name, entity, max(fd.confidence) AS max\_confidence FROM `video\_analytics.object\_tracking\_analysis`, UNNEST(frame\_data) AS fd GROUP by file\_name, entity
ORDER by file\_name, entity, max\_confidence DESC

ow	file_name	entity	max_confidence			
1	cat.mp4	cat	0.86749231815338135			
2	gbikes_dinosaur.mp4	bench	0.58585751056671143			
3	gbikes_dinosaur.mp4	bicycle	0.73626726865768433			
4	gbikes_dinosaur.mp4	bicycle wheel	0.93204206228256226			
5	gbikes_dinosaur.mp4	building	0.57778167724609375			
6	gbikes_dinosaur.mp4	chair	0.71668219566345215			
7	gbikes_dinosaur.mp4	dinosaur	0.85336136817932129			
8	gbikes_dinosaur.mp4	footwear	0.53572624921798706			
9	gbikes_dinosaur.mp4	house	0.50764185190200806			
10	gbikes_dinosaur.mp4	person	0.91313296556472778			
11	gbikes_dinosaur.mp4	sculpture	0.80446499586105347			
12	gbikes_dinosaur.mp4	shorts	0.5866851806640625			
13	gbikes_dinosaur.mp4	tire	0.66517066955566406			
14	gbikes_dinosaur.mp4	train	0.63094544410705566			

Ejecuta la siguiente consulta para obtener las compensaciones de tiempo, las coordenadas del cuadro de límite y los objetos detectados en un archivo de video:

SELECT entity, fd.\*
FROM `video\_analytics.object\_tracking\_analysis`,
UNNEST(frame\_data) AS fd
WHERE file\_name LIKE 'gbike%.mp4'
ORDER BY timeOffset

w	entity	processing_timestamp	timeOffset	confidence	left	top	right		bottom
1	bicycle	2022-03-20 20:09:19.561000	0.0	0.59062814712524414	0.4034018	0.15256836	0.69659823	0.7057067	
2	bicycle wheel	2022-03-20 20:09:19.556000	0.0	0.7534714937210083	0.41102248	0.4479253	0.56480426	0.75166166	
3	bicycle wheel	2022-03-20 20:09:19.559000	0.0	0.71192306280136108	0.54492337	0.3912022	0.68621135	0.71723974	
4	tire	2022-03-20 20:09:19.560000	0.0	0.65913408994674683	0.16820216	0.54425687	0.32418936	0.9403182	
5	tire	2022-03-20 20:09:19.560000	0.0	0.62130814790725708	0.003933981	0.3694937	0.117960714	0.76988363	
6	bicycle	2022-03-20 20:09:19.579000	0.1	0.59062814712524414	0.4031254	0.15000394	0.6960945	0.7027823	
7	bicycle wheel	2022-03-20 20:09:19.577000	0.1	0.7534714937210083	0.41015756	0.4458365	0.56328356	0.7486158	
8	bicycle wheel	2022-03-20 20:09:19.578000	0.1	0.71192306280136108	0.5437518	0.38750565	0.6843778	0.71250725	
9	tire	2022-03-20 20:09:19.578000	0.1	0.65913408994674683	0.16952908	0.5527643	0.31797844	0.9291841	
10	tire	2022-03-20 20:09:19.579000	0.1	0.62130814790725708	0.0007861784	0.35973758	0.11952875	0.77915204	
11	bicycle	2022-03-20 20:09:19.603000	0.2	0.59062814712524414	0.4080667	0.15741487	0.68814486	0.6860029	
12	bicycle wheel	2022-03-20 20:09:19.600000	0.2	0.7534714937210083	0.41015452	0.44048452	0.5632796	0.7432624	
13	bicycle wheel	2022-03-20 20:09:19.601000	0.2	0.71192306280136108	0.54224455	0.38348514	0.6828697	0.70848525	
14	tire	2022-03-20 20:09:19.602000	0.2	0.65913408994674683	0.16877338	0.54875207	0.31721187	0.9251435	
15	tire	2022-03-20 20:09:19.602000	0.2	0.62130814790725708	3.1638006e-05	0.34774378	0.11878101	0.7671857	
16	bicycle	2022-03-20 20:09:19.624000	0.3	0.59062814712524414	0.40832192	0.15405133	0.68732506	0.6806221	
17	bicycle wheel	2022-03-20 20:09:19.622000	0.3	0.7534714937210083	0.40886548	0.43597716	0.5619905	0.7387549	
18	bicycle wheel	2022-03-20 20:09:19.622000	0.3	0.71192306280136108	0.5406548	0.37924585	0.6812798	0.7042458	
19	tire	2022-03-20 20:09:19.623000	0.3	0.65913408994674683	0.1660389	0.54221874	0.3144765	0.91860783	
20	tire	2022-03-20 20:09:19.623000	0.3	0.62130814790725708	0.00013423576	0.3530335	0.11387438	0.75466514	
21	bicycle	2022-03-20 20:09:19.661000	0.4	0.59062814712524414	0.40775508	0.14991154	0.68666947	0.67631567	
22	bicycle wheel	2022-03-20 20:09:19.659000	0.4	0.75223219394683838	0.41248083	0.44228667	0.5556548	0.7304086	
23	bicycle wheel	2022-03-20 20:09:19.660000	0.4	0.71852326393127441	0.5390674	0.37905973	0.6837034	0.6889242	
24	tire	2022-03-20 20:09:19.660000	0.4	0.65392041206359863	0.14796299	0.4766562	0.3144405	0.90441537	
25	tire	2022-03-20 20:09:19.660000	0.4	0.62284284830093384	0.015212158	0.36214855	0.11572887	0.7261003	
26	tire	2022-03-20 20:09:19.661000	0.4	0.59460932016372681	0.38246545	0.30786604	0.4869209	0.5010258	
27	bicycle	2022-03-20 20:09:19.701000	0.5	0.59062814712524414	0.4080338	0.14574242	0.6869407	0.67213255	
28	bicycle wheel	2022-03-20 20:09:19.700000	0.5	0.75140595436096191	0.41226032	0.4410909	0.55452925	0.727368	
29	bicycle wheel	2022-03-20 20:09:19.700000	0.5	0.72292345762252808	0.5398735	0.37460747	0.6844134	0.68306726	
30	tire	2022-03-20 20:09:19.700000	0.5	0.6504446268081665	0.14392571	0.45547235	0.31176454	0.8870521	
31	tire	2022-03-20 20:09:19.701000	0.5	0.62386602163314819	0.02243729	0.36699522	0.11745444	0.7117471	
32	tire	2022-03-20 20:09:19.701000	0.5	0.59460932016372681	0.38281196	0.3055591	0.48671907	0.49861482	
33	bicycle	2022-03-20 20:09:19.745000	0.6	0.59062814712524414	0.40823737	0.14239833	0.6871437	0.66878736	
34	bicycle wheel	2022-03-20 20:09:19.743000	0.6	0.75081580877304077	0.4115864	0.43707752	0.55378073	0.7232026	
35	bicycle wheel	2022-03-20 20:09:19.744000	0.6	0.72606641054153442	0.54028434	0.37013727	0.6848163	0.6784812	
36	tire	2022-03-20 20:09:19.744000	0.6	0.64796191453933716	0.13811554	0.43544927	0.31394485	0.8877176	
37	tire	2022-03-20 20:09:19.744000	0.6	0.62459683418273926	0.021817965	0.35423782	0.120689884	0.71272427	
38	tire	2022-03-20 20:09:19.745000	0.6	0.59460932016372681	0.38281256	0.30154416	0.4867189	0.49459973	
39	bicycle	2022-03-20 20:09:19.785000	0.7	0.59062814712524414	0.40839088	0.13949041	0.6872971	0.6658793	
40	9	2022-03-20 20:09:19.783000	0.7	0.7503732442855835	0.4117425	0.43356672	0.5539306	0.719679	
41	9	2022-03-20 20:09:19.784000	0.7	0.7284235954284668	0.5405212	0.36750323	0.6850525	0.67583746	
42	tire	2022-03-20 20:09:19.784000	0.7	0.64609992504119873	0.13807198	0.42914525	0.31457207	0.88315034	
43	tire	2022-03-20 20:09:19.785000	0.7	0.62514495849609375	0.025124053	0.35718518	0.12002423	0.7015523	
44	tire	2022-03-20 20:09:19.785000	0.7	0.59460932016372681	0.38345718	0.29844964	0.48736346	0.4915052	
45		2022-03-20 20:09:19.848000	0.8	0.7503732442855835	0.41128904	0.4302369	0.5534771	0.7163492	
46	•	2022-03-20 20:09:19.848000	0.8	0.74527448415756226	0.54500616	0.36141402	0.68185484	0.67292786	
47	•	2022-03-20 20:09:19.849000	0.8	0.55888181924819946	0.015016886	0.8832664	0.22182398	0.99439746	
48	tire	2022-03-20 20:09:19.848000	0.8	0.64246129989624023	0.14971313	0.43874398	0.3169887	0.8814454	
49	tire	2022-03-20 20:09:19.849000	0.8	0.62298280000686646	0.009338304	0.33131048	0.12854198	0.7313317	
50	tire	2022-03-20 20:09:19.849000	0.8	0.59460932016372681	0.38422015	0.29560477	0.4881264	0.48866034	