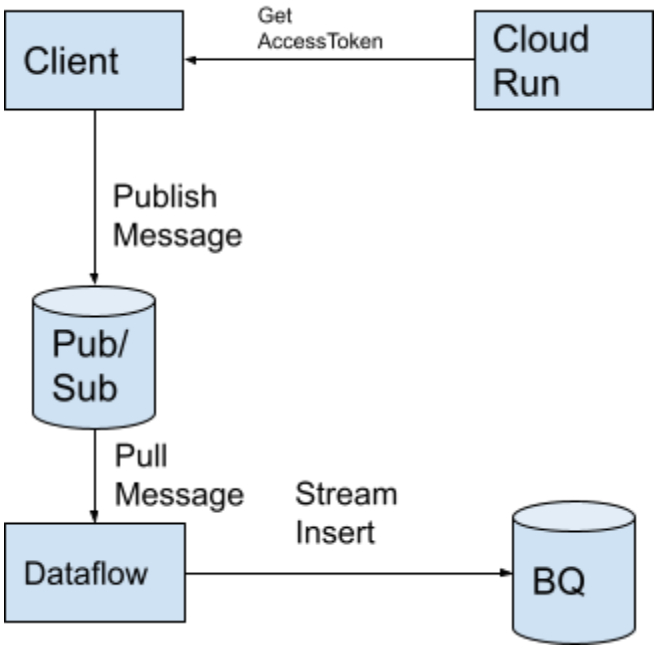
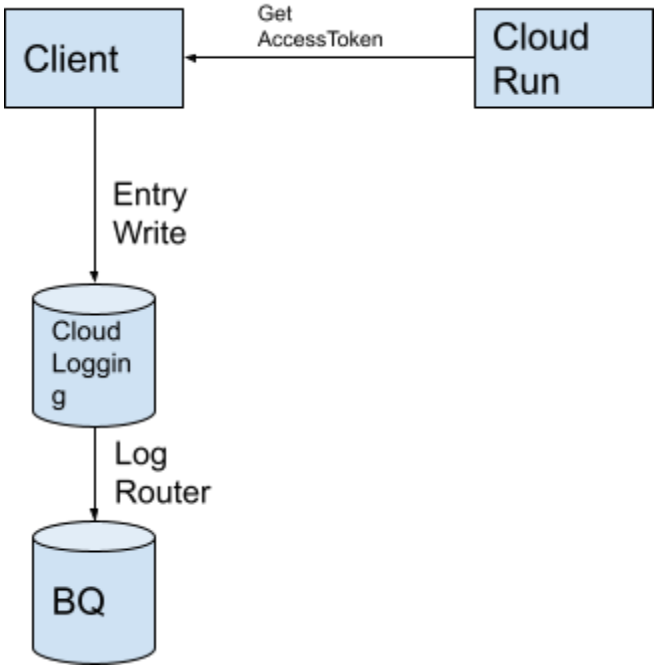
1.架构图

- Cloud run部署根据关联的sa生成access token的服务
- 客户端获取access token并保存
- 客户端根据jwt向pubsub 发布日志消息或者向cloud logging发日志消息
- dataflow(或者客户自己的消息处理程序消费消息), 流式插入bq
- 整个CD过程通过Cloud Build来实现

1.1方法1


Client
Get
AccessToken
Cloud
Run
Publish
Message
Pub/
Sub
Pull
Message
Stream
Insert
Dataflow
BQ

1.2方法2


Client
Get
AccessToken
Cloud
Run
Entry
Write
Cloud
Loggin
g
Log
Router
BQ

2.可扩展性

2.1.Pub/Sub 可扩展性

https://cloud.google.com/pubsub/quotas#throughput_quota_units

Pubsub Quota:

> Regional quotas are divided into 2 types:
>
> - Large regions: `europe-west1`, `us-central1`, `us-east1`
> - Small regions: all other regions

X客户, 在印度和新加坡： 120GB/Min

Quota	Default quota limit	Description
Publisher throughput per region	• 12,000,000 kB per minute (200 MB/s) in large regions • 3,000,000 kB per minute (50 MB/s) in small regions	**`pubsub.googleapis.com/regionalpublisher`** Quota usage is based on the size of the published **`PubsubMessage`**s: • PubsubMessage (REST) • PubsubMessage (RPC) Note that multiple messages can be included in a single publish request, and there is no additional quota charge per message. If messages have ordering keys, the maximum publisher throughput is 1 MB/s per ordering key.

Pubsub Limits:

Publish request	10MB (total size) 1,000 messages

```
"messages": [
            {
                "data": encoded_element,
                "ordering_key": "first order",
                "attributes": {"somekey": "somevalue"}
```

```
            }
        ]
```

2.2.Cloud logging可扩展性

https://cloud.google.com/logging/quotas#api-limits

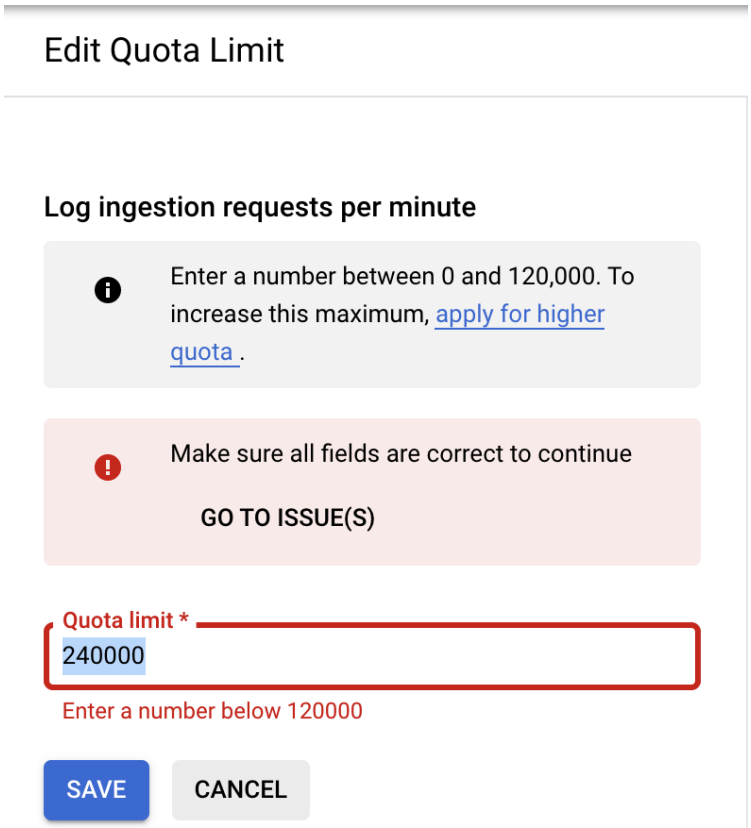
https://cloud.google.com/logging/docs/reference/v2/rest/v2/entries/write

备注: QPM比较小, 120k,但是可以考虑write entry的时候采用batch的方式, 降 低实时性; 但是整体来说不太适合客户端日志直接打入

Logging API quotas and limits

Category	Maximum value
API usage	To view your quotas, go to the API dashboard. Click an API and select **Quotas**.
Lifetime of API page tokens	24 hours
Number of open live-tailing sessions per Cloud project	10
Number of live-tailing entries returned per minute	60,000
Size of an entries.write request	10 MB
Number of entries.write API calls	120,000 per minute, per Cloud project[1]
Number of entries.list API calls	1 per second, per Cloud project
Number of Cloud projects or other resource names in a single entries.list API call	100

[1] Using log exclusions doesn't reduce this number because logs are excluded after the entries.write call is made.



2.3.Bigquery Stream Insert 可扩展性

https://cloud.google.com/bigquery/quotas#streaming_inserts

- 如果没有insertID 字段, stream insert只会限制吞吐量 1GB/s

Streaming inserts

The following limits apply for streaming data into BigQuery.

If you do not populate the `insertId` field when you insert rows, the following quotas apply. For more information, see Disabling best effort de-duplication. This is the recommended way to use BigQuery in order to get higher streaming ingest quota limits.

Note: Currently, these quotas apply to the **us** and **eu** multi-regions. However, these quotas are *not* applicable when streaming to template tables.

- **Maximum bytes per second: 1 GB** — If you don't populate `insertId` field for each row inserted, you are limited to 1 GB per second, per project. This limit applies at the project level. It does not apply to individual tables. Exceeding this amount will cause `quotaExceeded` errors.

- 如果有insertID 字段, stream insert限制会更多

If you populate the `insertId` field when you insert rows, the following quotas apply.

- **Maximum rows per second per project in the `us` and `eu` multi-regions: 500,000**

  If you populate the `insertId` field for each row inserted, you are limited to 500,000 rows per second in the `us` and `eu` multi-regions, per project. This quota is cumulative within a given multi-region. In other words, the sum of rows per second streamed to all tables for a given project within a multi-region is limited to 500,000. Each table is additionally limited to 100,000 rows per second.

  Exceeding either the per-project limit or the per-table limit will cause `quotaExceeded` errors.

- **Maximum rows per second per project in all other locations: 100,000**

  If you populate the `insertId` field for each row inserted, you are limited to 100,000 rows per second in all locations except the `us` and `eu` multi-regions, per project or table. This quota is cumulative within a given region. In other words, the sum of rows per second streamed to all tables for a given project within a region is limited to 100,000.

  Exceeding this amount will cause `quotaExceeded` errors.

- **Maximum rows per second per table: 100,000**

  If you populate the `insertId` field for each row inserted, you are limited to 100,000 rows per second per table.

  Exceeding this amount will cause `quotaExceeded` errors.

- **Maximum bytes per second: 100 MB**

  If you populate the `insertId` field for each row inserted, you are limited to 100 MB per second, per table.

  Exceeding this amount will cause `quotaExceeded` errors.

https://developers.google.com/identity/protocols/oauth2/service-account#jwt-auth

3.价格体系

3.1pub/sub 价格体系

https://cloud.google.com/pubsub/pricing#pubsub-pricing

Message ingestion and delivery

Message delivery pricing

Message ingestion and delivery are priced per volume of data transmitted in a calendar month. The first 10 gigabytes of usage are free. After that, the price for ingestion or delivery of messages is $40 per TiB.

The data volume of a message is the sum of the following:

- The number of bytes in the encoded message body string
- For each attribute, the size of the key and its value
- 20 bytes for the timestamp
- The size of the `message_id` string
- Additional optional fields, such as those associated with early access and other restricted access APIs.

3.2 cloud run价格体系

https://cloud.google.com/run/pricing

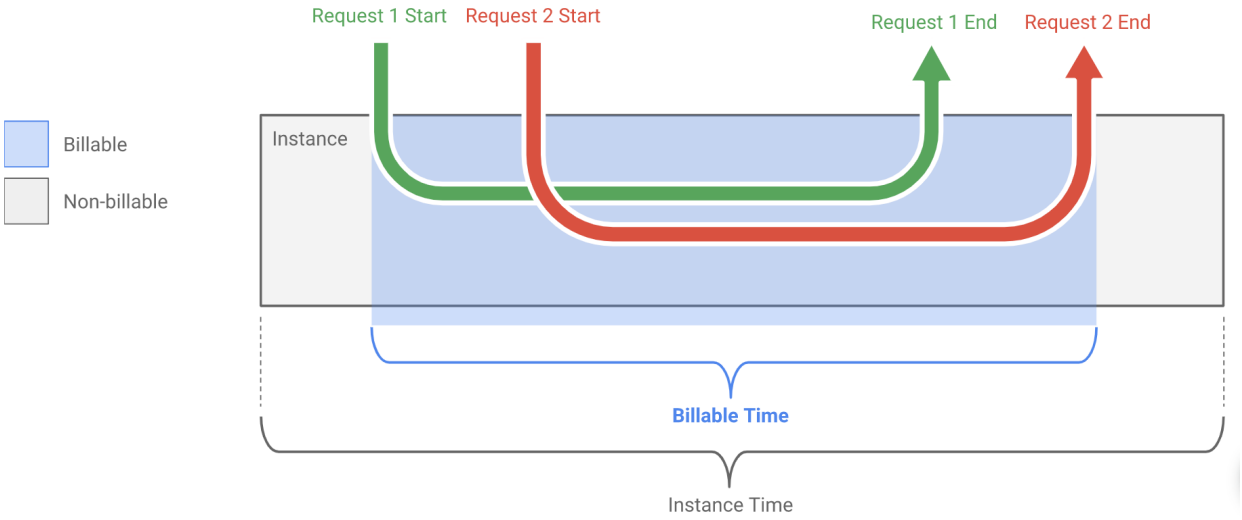
Tier	CPU	Memory	Requests	Networking
Free	First 180,000 vCPU-seconds free per month	First 360,000 GiB-seconds free per month	2 million requests free per month	1 GiB free egress within North America per month
1	$0.00002400 / vCPU-second beyond free tier CUD[1]: $0.00001992 If idle[2]: $0.00000250	$0.00000250 / GiB-second beyond free tier CUD[1]: $0.000002075 If idle[2]: $0.00000250	$0.40 / million requests beyond free tier CUD[1]: $0.332	Google Cloud Network Premium tier pricing beyond free tier.

Billable time

You are billed only for the CPU and memory allocated during billable time, rounded up to the nearest 100 milliseconds.

For a given container instance, *active* billable time occurs when:

- The container instance is starting
- The container instance is shut down and the `SIGTERM` signal is handled.
- At least one request is being processed by the container instance. Active billable time begins with the start of the first request and ends at the end of the last request, as shown in the following diagram:



3.3 Dataflow价格体系

https://cloud.google.com/dataflow/pricing#pricing_overview

Worker resources pricing

Iowa (us-central1)

Dataflow Worker Type	vCPU (per hour)	Memory (per GB per hour)	Data Processed[4,5] (per GB)
Batch [1]	$0.056	$0.003557	$0.011
FlexRS [2]	$0.0336	$0.0021342	$0.011
Streaming [3]	$0.069	$0.003557	$0.018

3.4 BQ价格体系

https://cloud.google.com/bigquery/pricing#storage

US (multi-region) | Monthly

Operation	Pricing	Details
Active storage	$0.020 per GB	The first 10 GB is free each month.
Long-term storage	$0.010 per GB	The first 10 GB is free each month.

US (multi-region) | Monthly

Operation	Pricing	Details
Batch loading	Free using the shared slot pool.	Customers can choose flat-rate pricing for guaranteed capacity. Once the data is loaded into BigQuery, you are charged for storage.
Streaming inserts (`tabledata.insertAll`)	$0.010 per 200 MB	You are charged for rows that are successfully inserted. Individual rows are calculated using a 1 KB minimum size.

4.方法1: CloudRun-Pubsub-Dataflow-Demo

demo步骤请参考:

1. Cloud run上部署一个service,用来获取access token
2. 客户端通过get cloud run service,获取token后, 通过rest请求往自定义的 topic上publish log message
3. 创建bigquery table(schema = log schema)
4. 创建dataflow: 读取pubsub subscription,stream insert到bigquery.

参考文档:
Service identity | Cloud Run Documentation
Google-provided streaming templates | Cloud Dataflow
Using OAuth 2.0 for Server to Server Applications | Google Identity

具体请参考:

https://github.com/liuchenggang2014/clientLog

5.方法2: CloudRun-Logging-BQ-Demo

demo步骤请参考:

1. Cloud run上部署一个service,用来获取logging access token
2. 客户端通过get cloud run service,获取token后, 通过rest请求向cloud logging上write log entry
3. 根据需要创建logging router(sink), 自动同步到bigquery

具体请参考:

https://github.com/liuchenggang2014/getLoggingToken

参考文档:

https://cloud.google.com/logging/docs/api/v2/resource-list
https://cloud.google.com/logging/docs/reference/v2/rest/v2/entries/write
https://cloud.google.com/logging/docs/reference/v2/rest/v2/LogEntry
https://developers.google.com/identity/protocols/oauth2/scopes
https://cloud.google.com/blog/products/operations/psyonix-on-cloud-logging-and-monitoring

注意事项:

Request parameters

No method-level parameters

Show standard parameters

Request body


```
{
  "logName": "projects/cliu101/logs/clientlog",
  "entries": [
    {
      "jsonPayload": {
        "id": "1",
        "name": "1"
      }
    },
    {
      "jsonPayload": {
        "id": "2",
        "name": "2"
      }
    }
  ],
  "resource": {
    "type": "generic_node",
    "labels": {
      "locations": "us-central1-a",
      "project_id": "cliu101"
    }
  }
}
```


For suggestions, press control+space or click one of the

cURL HTTP JAVASCRIPT

```
curl --request POST \
  'https://logging.googleapis.com/v2/entries:write?key=[YOUR_API_KEY]' \
  --header 'Authorization: Bearer [YOUR_ACCESS_TOKEN]' \
  --header 'Accept: application/json' \
  --header 'Content-Type: application/json' \
  --data '{"logName":"projects/cliu101/logs/clientlog","entries":[{"jsonPayload":{"id":"1",
  --compressed
```

application/json Raw HTTP Response

200

```
{}
```

Query results

Jump to now Actions Configure

SEVERITY TIMESTAMP HKT SUMMARY

Showing logs for last 1 hour from 6/22/21, 5:20 PM to 6/22/21, 6:20 PM. Extend time by: 1 hour Edit time

2021-06-22 18:19:35.011 HKT { "id": "1", "name": "1" }

Hide log summary Expand nested fields Copy to clipboard Copy link


```
{
  insertId: "yyxszwfq839et"
  jsonPayload: {
    id: "1"
    name: "1"
  }
  logName: "projects/cliu201/logs/clientlog"
  receiveTimestamp: "2021-06-22T10:19:35.011990456Z"
  resource: {2}
  timestamp: "2021-06-22T10:19:35.011990456Z"
}
```

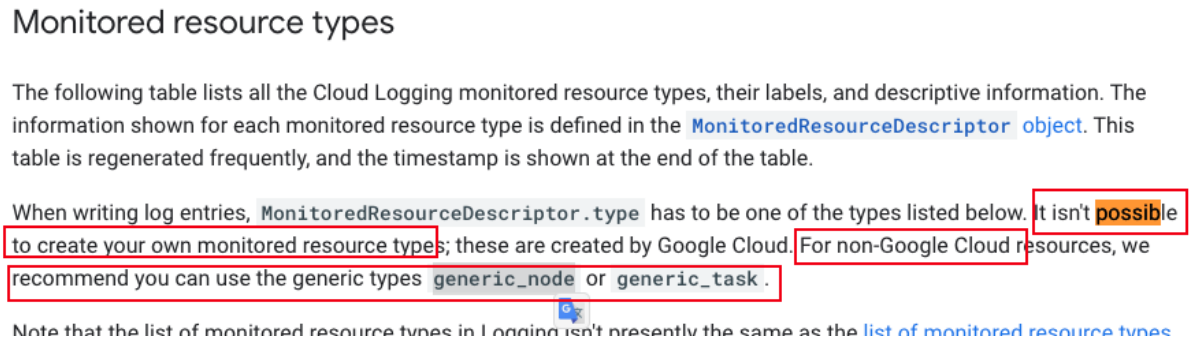

2021-06-22 18:19:35.011 HKT { "id": "2", "name": "2" }

Hide log summary Expand nested fields Copy to clipboard Copy link


```
{
  insertId: "yyxszwfq839ew"
  jsonPayload: {2}
  logName: "projects/cliu201/logs/clientlog"
  receiveTimestamp: "2021-06-22T10:19:35.011990456Z"
  resource: {2}
  timestamp: "2021-06-22T10:19:35.011990456Z"
}
```


cloud logging收到的timestamp

客户端调用的write的timestamp



Monitored resource types

The following table lists all the Cloud Logging monitored resource types, their labels, and descriptive information. The information shown for each monitored resource type is defined in the `MonitoredResourceDescriptor` object. This table is regenerated frequently, and the timestamp is shown at the end of the table.

When writing log entries, `MonitoredResourceDescriptor.type` has to be one of the types listed below. It isn't possible to create your own monitored resource types; these are created by Google Cloud. For non-Google Cloud resources, we recommend you can use the generic types `generic_node` or `generic_task`.

Note that the list of monitored resource types in Logging isn't presently the same as the list of monitored resource types

Network diagnose develop in android and ios:

https://github.com/qiniu/iOS-netdiag
https://github.com/qiniu/android-netdiag
https://github.com/ucloud/netanalysis-sdk-android
https://github.com/ucloud/netanalysis-sdk-ios