Automated Data Warehousing @ KKStream

...

黃仲安 許凱嵐

About

- KDDI & VideoPass
- KKStream & Operations Center
- Amazon Web Services (AWS)

KDDI & VideoPass

- Second largest telecommunication operator in Japan
- OTT service provider
- 3+ million subscribers with VideoPass

ビデオパス

約 10,000 本の
ドラマ・アニメ・映画が見放題！

ビデオパス なら
こんなに便利！

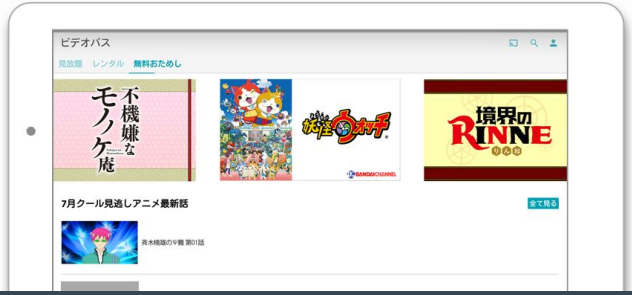
- 返却不要
- 貸出中なし
- いつでも見られる
- au だから安心
- 登録なしで
  ライブ配信が楽しめる
- ダウンロード視聴可能

マルチデバイス対応

だから、24 時間
いつでもどこでも楽しめる！

iPhone、iPad、テレビ、パソコンにも対応

7月クールアニメの最新話を無料で視聴！



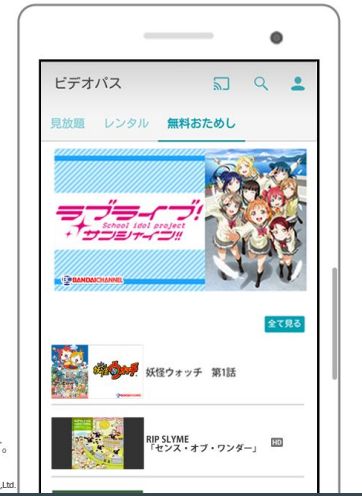
Chromecastで映画をテレビで視聴

chromecast

ビデオパスで音楽ライブを視聴！

アニメ、韓流音楽ライブなど見放題！

au 公式 ビデオパス



※画像はイメージです。

KKStream & Operations Center

- KKBOX Group
- video streaming solutions provider
- operations center supports 24/7 operations


KKStream


KKbox
Like
HAMI+
FREE
PAN
CHU~
KKbox
2015


KKbox
stay hungry, stay

Amazon Web Services (AWS)

- S3 (Storage)
- Lambda (Compute)
- Athena (Database)
- Glue (ETL)
- CloudFront (CDN)
- CloudWatch (Cron Job)
- SES (Mailing Service)

The Challenge

- by Numbers
- Time Limit
- Available Resources
- Exisiting Solution

by Numbers

- 3 distributions
- ~4000 log files
- 300,000,000+ records
- 31 columns
- 100GB of size per day (extracted)
- 3-4TB per month
- 3 times data scanned daily
- scale to 20+ distrubutions & 20+ log events

Time Limit

- 3 days / week * (2 weeks / Jan. + 2 weeks / Feb.) = 12 days = 96 hours
- 2 days / week * (4 weeks / Mar. + 4 weeks / Apr.) = 16 days = 128 hours

Available Resources

- AWS products (with limited permissions)
- AWS documentations (blog + tutorials + Google)
- OC Team
- Service Team
- < 24 hours writing Python
- 0% knowledge in automation
- 0% knowledge in CI/CD
- 0% knowledge in data warehousing
- 0% knowledge in data processing
- 100% knowledge in reading files line by line

Existing Solution

- manual screenshots + downloads + e-mails
- time-consuming
- difficult to manage
- lack of accessibility
- not cost-efficient

Distributions
What's New *

Reports & Analytics
Cache Statistics
Monitoring and Alarms
Popular Objects
Top Referrers
Usage
Viewers

Security
Origin Access Identity
Public key
Field-level encryption

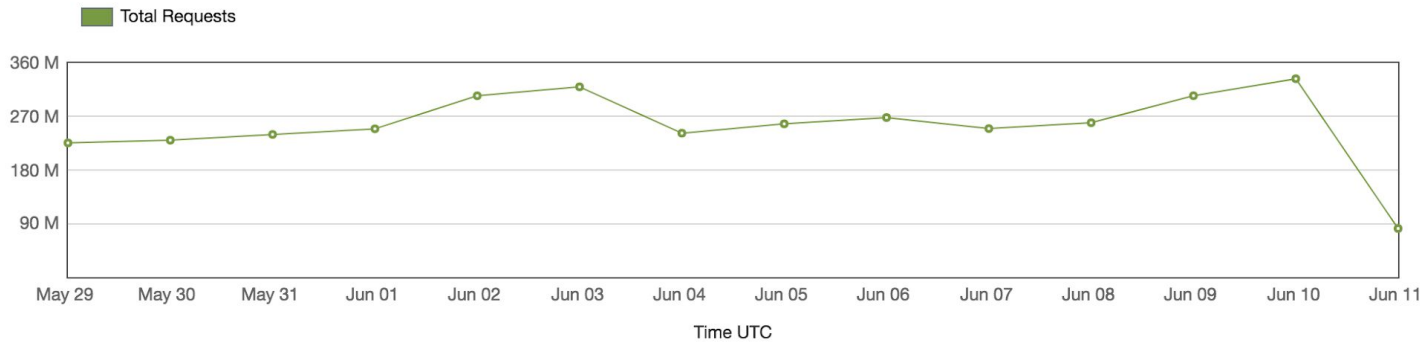
CloudFront Cache Statistics Reports

Start Date 2018-05-29 Granularity Daily (any period in previous 60 Web Distribution All Web Distributions

End Date 2018-06-11 Viewer Location All Locations Update CSV

The following charts show selected values from CloudFront access logs. In the access logs for a distribution, each row corresponds with one viewer request. If you choose All Web Distributions in the Distribution list, the charts include totals for all of your web distributions that had activity during the specified period and that have not been deleted. Data for deleted distributions and RTMP distributions is not available. Note: Cache Statistics Reports on this page are based on the location of the viewers making the request while Usage Reports are based on CloudFront Billing Region.

Total Requests (Millions | Thousands | Not Scaled) Show Details



Average: 254.7672 M **Total:** 3,566.7411 M **Maximum:** 333.4928 M **Minimum:** 81.7656 M

aws-lambda APP 9:10 AM

production-theater daily reort

CDN Cache Report 2018-06-10

Peak
298.96 GB (39.86Gbps) at 13:37

Average
188.28 GB (25.10Gbps)

June bandwidth usage

Yesterday
0.26 PB

To-Date
2.17 PB

Estimate
6.51 PB

CacheStatistics-2018-01-28 Inbox x

andy hsu <andyhsu@kkstream.com.tw> Jan 29

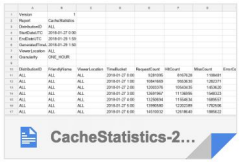
to videopass-oc, vp-service, 黃振修, kks24, me

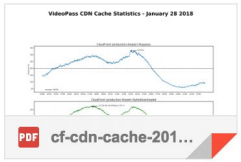
お世話になっております。

早速ですが、日次の通信データ容量を添付いたしました。
資料により最大ピークは271.83 GB (36.24Gbps) at 07:17です、ご確認お願いします。

以上、よろしくお願いいたします。

3 Attachments





Design and Build

- ETL Pipeline
- Progress Tracker
- Error Handling

ETL Pipeline

1) [Lambda] preprocess CloudFront log at 8:30 UTC+0
2) [Lambda] dispatch Lambda consolidation worker
3) [Lambda] run an algorithm to merge files
4) [Lambda] check total size before compression
5) [Glue] map, concat, and compress distribution 1 of 3
6) [Glue] map, concat, and compress distribution 2 of 3
7) [Glue] map, concat, and compress distribution 3 of 3
8) [Glue] update Athena table via Glue Crawler
9) [Lambda] query data from Athena and send to s3
10) [Lambda] download data from s3, generate report, and send to s3
11) [Lambda] download files from s3 generate e-mail body, and dispatch e-mail

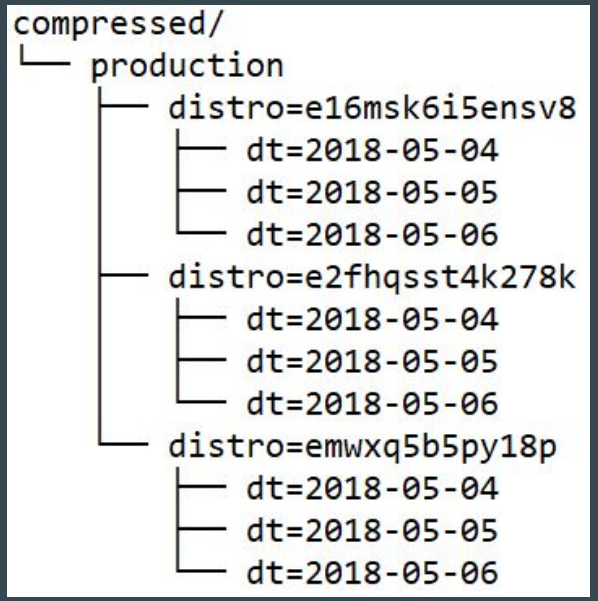
CloudWatch (8:30 UTC+8)

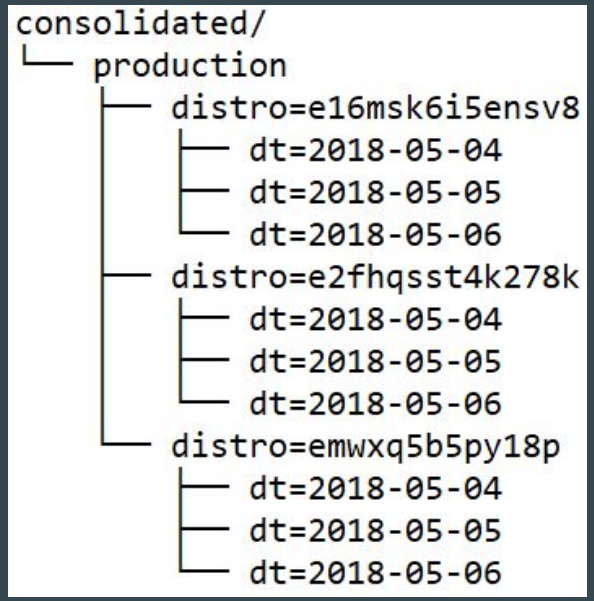
1	2	3	4	5	CW	6	7	8
合併	檢查	壓縮1	壓縮2	壓縮3		產CSV	產PDF	寄信
Lambda	Lambda	Glue	Glue	Glue		Lambda	Lambda	Lambda
S3	S3	S3	S3	S3		S3	S3	S3
				Athena				SES

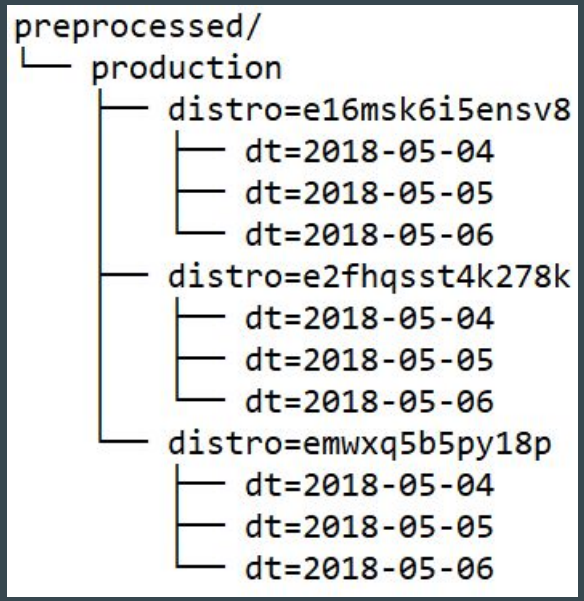
EMWXQ5B5PY18P.2018-06-10-00.017639fa.gz	Jun 10, 2018 8:03:51 AM GMT+0800	4.2 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.02cca497.gz	Jun 10, 2018 8:48:17 AM GMT+0800	7.2 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.07021ce4.gz	Jun 10, 2018 8:41:10 AM GMT+0800	7.2 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.0724ce6a.gz	Jun 10, 2018 9:04:36 AM GMT+0800	972.8 KB	Standard
EMWXQ5B5PY18P.2018-06-10-00.07787196.gz	Jun 10, 2018 8:44:06 AM GMT+0800	7.1 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.08b01dff.gz	Jun 10, 2018 8:58:19 AM GMT+0800	7.4 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.0bc38d4d.gz	Jun 10, 2018 8:14:07 AM GMT+0800	6.4 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.0bf909c6.gz	Jun 10, 2018 8:03:14 AM GMT+0800	3.5 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.0f4fe408.gz	Jun 10, 2018 8:36:00 AM GMT+0800	6.5 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.0f937a76.gz	Jun 10, 2018 8:33:17 AM GMT+0800	6.8 MB	Standard
EMWXQ5B5PY18P.2018-06-10-00.10a708a3.gz	Jun 10, 2018 8:34:38 AM GMT+0800	7.0 MB	Standard

```
#Version: 1.0
#Fields: date time x-edge-location sc-bytes c-ip cs-method cs(Host) cs-uri-stem sc-status cs(Referer) cs(User-Agent) cs-uri-query cs(Cookie) x-edge-result-type x-edge-requ
2018-06-15      00:50:25        NRT57-C2        179747  219.109.241.228 GET     d1fh4xdv91g8dx.cloudfront.net   /32/509532_fea64354b641a61bcc16008ddbe33f96/1518082389_dash
2018-06-15      00:50:26        NRT57-C2        380408  61.23.131.108   GET     d1fh4xdv91g8dx.cloudfront.net   /11/627811_c5a006c82fbbe4eb7b5753ee577fccbb/1504206276_thum
2018-06-15      00:50:15        NRT51   2273451 153.179.203.250 GET     d1fh4xdv91g8dx.cloudfront.net   /43/678143_67baac740b4939786d0cf17efe377d21/1522198888_hls_fps/720p
2018-06-15      00:50:15        NRT51   437917  126.209.235.32  GET     d1fh4xdv91g8dx.cloudfront.net   /24/391324_318b65e2d5e50abcb55c1f9739e551b9/1463986943_dash/480p_7C
2018-06-15      00:50:15        NRT51   480410  133.155.170.10  GET     d1fh4xdv91g8dx.cloudfront.net   /76/682876_b64d465b8bd5ca3ca60c29bd9ac94a88/1528949897_dash/360p_7C
2018-06-15      00:50:15        NRT51   45823   60.236.132.116  GET     d1fh4xdv91g8dx.cloudfront.net   /33/355033_b5c675f5c1928d3407e058cc998f24d7/1424858939_dash/64k/117
2018-06-15      00:50:16        NRT51   31003   223.216.28.211  GET     d1fh4xdv91g8dx.cloudfront.net   /29/665229_9a9c57c1701c39fdd1576b7ca4a39768/1527534241_dash/32k/196
2018-06-15      00:50:17        NRT51   132394  60.236.132.116  GET     d1fh4xdv91g8dx.cloudfront.net   /33/355033_b5c675f5c1928d3407e058cc998f24d7/1424858939_dash/240p_35
2018-06-15      00:50:17        NRT51   65670   126.209.235.32  GET     d1fh4xdv91g8dx.cloudfront.net   /24/391324_318b65e2d5e50abcb55c1f9739e551b9/1463986943_dash/96k/264
2018-06-15      00:50:18        NRT51   4531570 114.174.108.111 GET     d1fh4xdv91g8dx.cloudfront.net   /18/665918_af7ae8d63f3bb68d6993b3c6387264ca/1517857463_dash/720p_7C
2018-06-15      00:50:19        NRT51   155528  223.216.28.211  GET     d1fh4xdv91g8dx.cloudfront.net   /29/665229_9a9c57c1701c39fdd1576b7ca4a39768/1527534241_dash/240p_24
2018-06-15      00:50:19        NRT51   30883   126.242.167.94  GET     d1fh4xdv91g8dx.cloudfront.net   /76/682776_7fd2fe455c1506e9ecb257e2eb732331/1527843479_dash/32k/8.m
2018-06-15      00:50:20        NRT51   45786   60.236.132.116  GET     d1fh4xdv91g8dx.cloudfront.net   /33/355033_b5c675f5c1928d3407e058cc998f24d7/1424858939_dash/64k/118
2018-06-15      00:50:20        NRT51   504334  126.209.235.32  GET     d1fh4xdv91g8dx.cloudfront.net   /24/391324_318b65e2d5e50abcb55c1f9739e551b9/1463986943_dash/480p_7C
2018-06-15      00:50:21        NRT51   78771   133.155.170.10  GET     d1fh4xdv91g8dx.cloudfront.net   /76/682876_b64d465b8bd5ca3ca60c29bd9ac94a88/1528949897_dash/96k/862
2018-06-15      00:50:21        NRT51   4402822 153.191.148.190 GET     d1fh4xdv91g8dx.cloudfront.net   /60/675260_0ded6d706bebaebafe246d4c4756e576/1524721272_dash/720p_7C
2018-06-15      00:50:21        NRT51   1509388 153.180.29.187  GET     d1fh4xdv91g8dx.cloudfront.net   /29/682829_280ec0eebc8bee1b4356847b0e7fbba8/1528789907_hls_fps/720p
2018-06-15      00:50:22        NRT51   83425   60.236.132.116  GET     d1fh4xdv91g8dx.cloudfront.net   /33/355033_b5c675f5c1928d3407e058cc998f24d7/1424858939_dash/240p_35
2018-06-15      00:50:22        NRT51   511405  133.155.170.10  GET     d1fh4xdv91g8dx.cloudfront.net   /76/682876_b64d465b8bd5ca3ca60c29bd9ac94a88/1528949897_dash/360p_7C
2018-06-15      00:50:22        NRT51   65812   126.209.235.32  GET     d1fh4xdv91g8dx.cloudfront.net   /24/391324_318b65e2d5e50abcb55c1f9739e551b9/1463986943_dash/96k/265
2018-06-15      00:50:22        NRT51   31031   223.216.28.211  GET     d1fh4xdv91g8dx.cloudfront.net   /29/665229_9a9c57c1701c39fdd1576b7ca4a39768/1527534241_dash/32k/197
2018-06-15      00:50:23        NRT51   4688353 114.174.108.111 GET     d1fh4xdv91g8dx.cloudfront.net   /18/665918_af7ae8d63f3bb68d6993b3c6387264ca/1517857463_dash/720p_7C
2018-06-15      00:50:24        NRT51   4355851 153.191.148.190 GET     d1fh4xdv91g8dx.cloudfront.net   /60/675260_0ded6d706bebaebafe246d4c4756e576/1524721272_dash/720p_7C
2018-06-15      00:50:24        NRT51   2878324 153.179.203.250 GET     d1fh4xdv91g8dx.cloudfront.net   /43/678143_67baac740b4939786d0cf17efe377d21/1522198888_hls_fps/720p
2018-06-15      00:50:24        NRT51   1934057 153.179.203.250 GET     d1fh4xdv91g8dx.cloudfront.net   /43/678143_67baac740b4939786d0cf17efe377d21/1522198888_hls_fps/720p
2018-06-15      00:50:25        NRT51   45739   60.236.132.116  GET     d1fh4xdv91g8dx.cloudfront.net   /33/355033_b5c675f5c1928d3407e058cc998f24d7/1424858939_dash/64k/119
2018-06-15      00:50:25        NRT51   405903  126.209.235.32  GET     d1fh4xdv91g8dx.cloudfront.net   /24/391324_318b65e2d5e50abcb55c1f9739e551b9/1463986943_dash/480p_7C
2018-06-15      00:50:25        NRT51   31107   126.242.167.94  GET     d1fh4xdv91g8dx.cloudfront.net   /76/682776_7fd2fe455c1506e9ecb257e2eb732331/1527843479_dash/32k/9.m
2018-06-15      00:50:25        NRT51   237148  223.216.28.211  GET     d1fh4xdv91g8dx.cloudfront.net   /29/665229_9a9c57c1701c39fdd1576b7ca4a39768/1527534241_dash/240p_24
2018-06-15      00:50:26        NRT51   78778   133.155.170.10  GET     d1fh4xdv91g8dx.cloudfront.net   /76/682876_b64d465b8bd5ca3ca60c29bd9ac94a88/1528949897_dash/96k/863
2018-06-15      00:50:12        NRT53   1491134 117.74.29.29    GET     d1fh4xdv91g8dx.cloudfront.net   /75/371775_fc7dc38c7f607c3609180edb946eb84a/1518168416_hls_fps/480p
2018-06-15      00:50:13        NRT53   30626   115.36.134.232  GET     d1fh4xdv91g8dx.cloudfront.net   /46/678346_957bca15cd38876cac780a04c1fae075/1528793072_dash/32k/75.
2018-06-15      00:50:14        NRT53   315235  115.36.134.232  GET     d1fh4xdv91g8dx.cloudfront.net   /46/678346_957bca15cd38876cac780a04c1fae075/1528793072_dash/360p_42
2018-06-15      00:50:16        NRT53   429489  180.196.242.238 GET     d1fh4xdv91g8dx.cloudfront.net   /33/497533_a08cf841e4a26d5797faf188e7232eaa/1428747245_dash/480p_7C
2018-06-15      00:50:18        NRT53   4426727 122.103.126.224 GET     d1fh4xdv91g8dx.cloudfront.net   /15/302315_771ae88a83a99850246810de192a82ba/1508353459_dash/720p_4C
2018-06-15      00:50:18        NRT53   65959   180.196.242.238 GET     d1fh4xdv91g8dx.cloudfront.net   /33/497533_a08cf841e4a26d5797faf188e7232eaa/1428747245_dash/96k/211
2018-06-15      00:50:19        NRT53   1415023 122.103.126.224 GET     d1fh4xdv91g8dx.cloudfront.net   /15/302315_771ae88a83a99850246810de192a82ba/1508353459_dash/720p_4C
2018-06-15      00:50:19        NRT53   30884   115.36.134.232  GET     d1fh4xdv91g8dx.cloudfront.net   /46/678346_957bca15cd38876cac780a04c1fae075/1528793072_dash/32k/76.
2018-06-15      00:50:20        NRT53   215306  115.36.134.232  GET     d1fh4xdv91g8dx.cloudfront.net   /46/678346_957bca15cd38876cac780a04c1fae075/1528793072_dash/360p_42
2018-06-15      00:50:21        NRT53   65869   114.69.23.37    GET     d1fh4xdv91g8dx.cloudfront.net   /36/506936_6482108ee7c799acd3c219f1a530bb1a/1518080324_dash/96k/221
2018-06-15      00:50:21        NRT53   2664237 111.104.219.59  GET     d1fh4xdv91g8dx.cloudfront.net   /19/677619_75bb3bfbdc2a5b2039d1e2ad272ae38c/1525848835_dash/1080p_3
2018-06-15      00:50:21        NRT53   453388  180.196.242.238 GET     d1fh4xdv91g8dx.cloudfront.net   /33/497533_a08cf841e4a26d5797faf188e7232eaa/1428747245_dash/480p_7C
```

☐	compressed	--	--	--	3
☐	consolidated	--	--	--	2
☐	oc-query-data	--	--	--	4
☐	oc-query-result	--	--	--	5
☐	preprocessed	--	--	--	1


compressed/
production
distro=e16msk6i5ensv8
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06
distro=e2fhqsst4k278k
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06
distro=emwxq5b5py18p
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06


consolidated/
production
distro=e16msk6i5ensv8
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06
distro=e2fhqsst4k278k
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06
distro=emwxq5b5py18p
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06


preprocessed/
production
distro=e16msk6i5ensv8
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06
distro=e2fhqsst4k278k
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06
distro=emwxq5b5py18p
dt=2018-05-04
dt=2018-05-05
dt=2018-05-06

Amazon S3 > videopass-processing-log / preprocessed / production / distro=emwxq5b5py18p / dt=2018-06-10 / tm=00

Overview

Type a prefix and press Enter to search. Press ESC to clear.

Upload + Create folder More

Asia Pacific (Tokyo)

Viewing 1 to 144

Name	Last modified	Size	Storage class
017639fa.gz	Jun 10, 2018 8:06:08 AM GMT+0800	4.2 MB	Standard
02cca497.gz	Jun 10, 2018 8:48:46 AM GMT+0800	7.2 MB	Standard
07021ce4.gz	Jun 10, 2018 8:41:39 AM GMT+0800	7.2 MB	Standard
0724ce6a.gz	Jun 10, 2018 9:04:42 AM GMT+0800	971.2 KB	Standard
07787196.gz	Jun 10, 2018 8:44:34 AM GMT+0800	7.0 MB	Standard
08b01dff.gz	Jun 10, 2018 8:58:47 AM GMT+0800	7.4 MB	Standard
0bc38d4d.gz	Jun 10, 2018 8:14:30 AM GMT+0800	6.3 MB	Standard
0bf909c6.gz	Jun 10, 2018 8:03:28 AM GMT+0800	3.5 MB	Standard

Amazon S3 > videopass-processing-log / consolidated / production / distro=emwxq5b5py18p / dt=2018-06-10

Overview

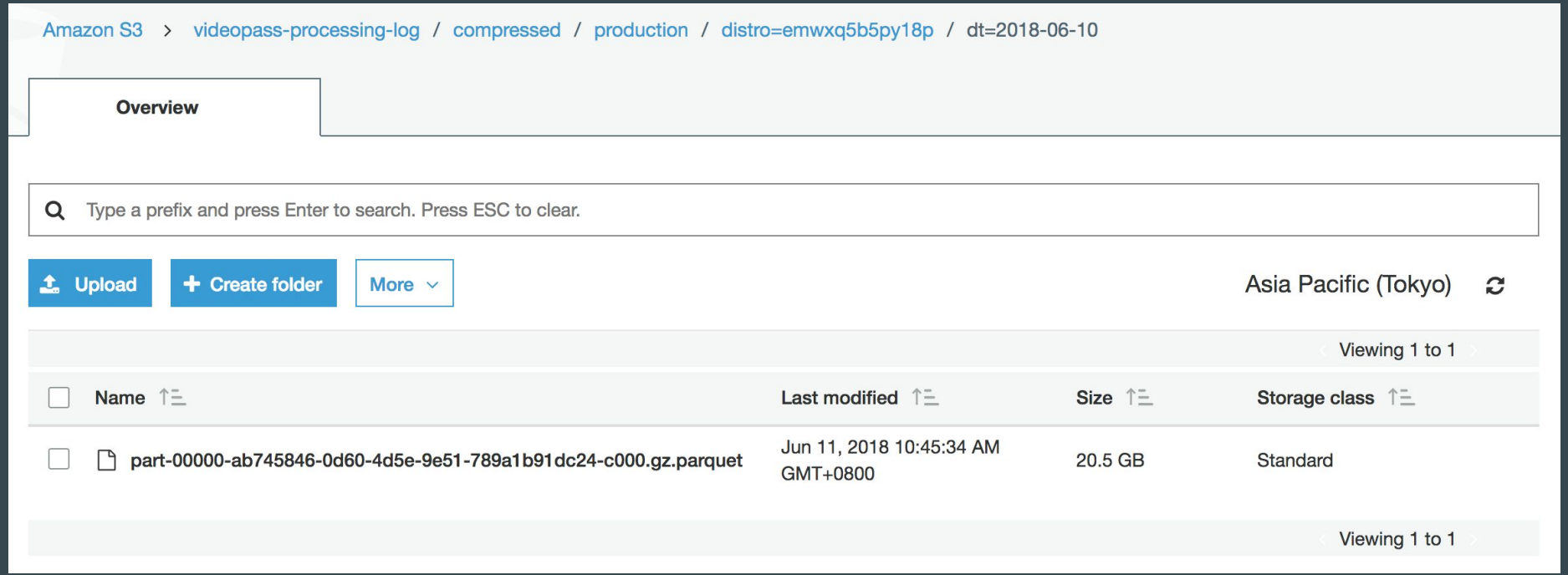
Type a prefix and press Enter to search. Press ESC to clear.

Upload | + Create folder | More

Asia Pacific (Tokyo)

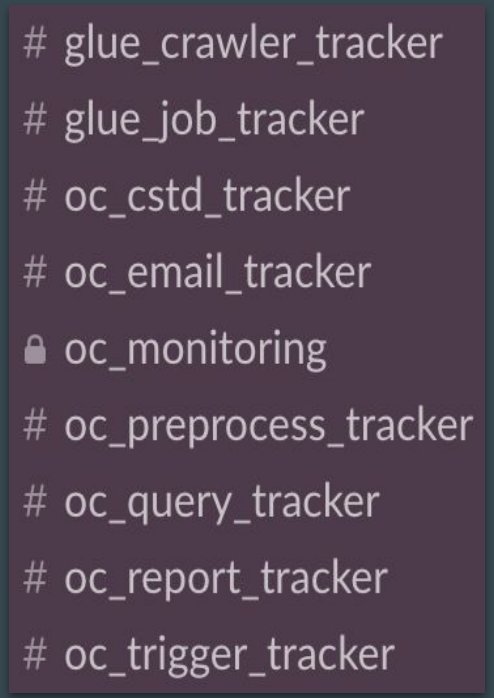
Viewing 1 to 300 >

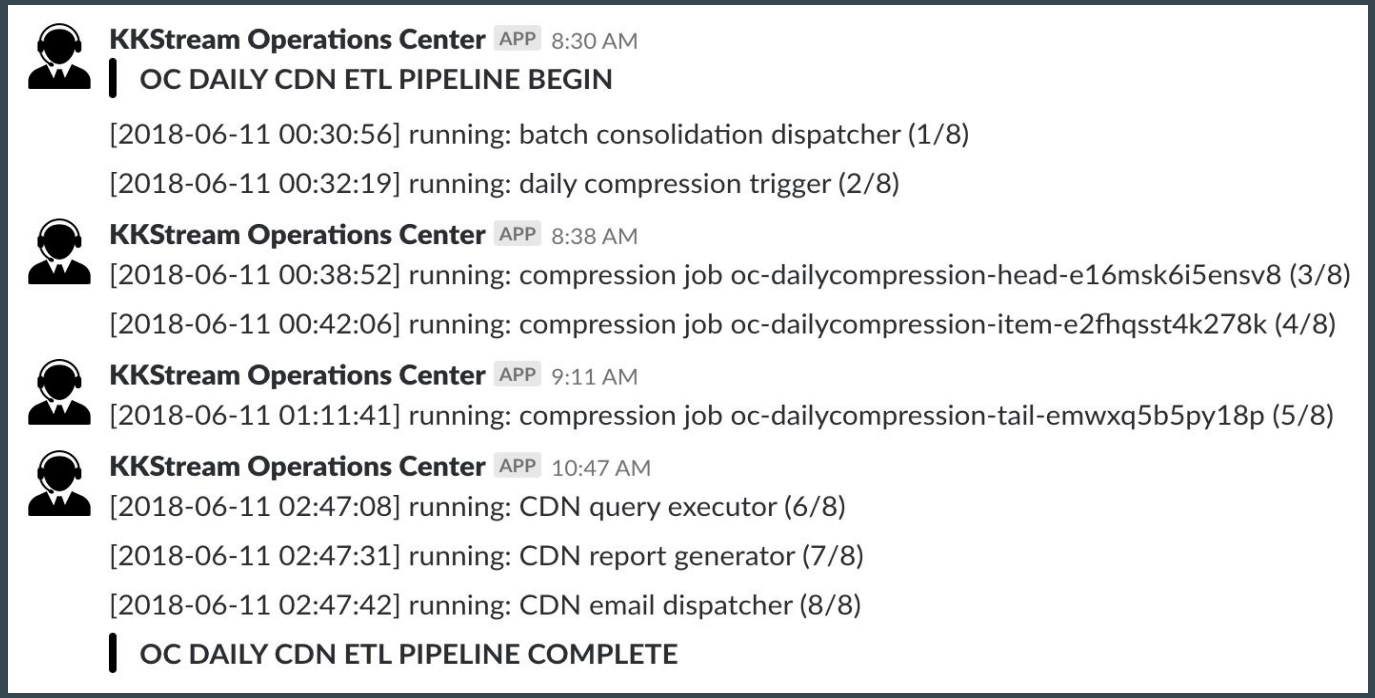
Name	Last modified	Size	Storage class
00-cstd1.gz	Jun 11, 2018 8:31:22 AM GMT+0800	57.0 MB	Standard
00-cstd10.gz	Jun 11, 2018 8:31:53 AM GMT+0800	57.7 MB	Standard
00-cstd11.gz	Jun 11, 2018 8:31:56 AM GMT+0800	57.0 MB	Standard
00-cstd12.gz	Jun 11, 2018 8:31:59 AM GMT+0800	59.7 MB	Standard
00-cstd13.gz	Jun 11, 2018 8:32:03 AM GMT+0800	57.5 MB	Standard
00-cstd14.gz	Jun 11, 2018 8:32:06 AM GMT+0800	60.9 MB	Standard
00-cstd15.gz	Jun 11, 2018 8:32:09 AM GMT+0800	35.5 MB	Standard
00-cstd2.gz	Jun 11, 2018 8:31:25 AM GMT+0800	57.2 MB	Standard


Amazon S3 > videopass-processing-log / compressed / production / distro=emwxq5b5py18p / dt=2018-06-10
Overview
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
More
Asia Pacific (Tokyo)
Viewing 1 to 1
Name
Last modified
Size
Storage class
part-00000-ab745846-0d60-4d5e-9e51-789a1b91dc24-c000.gz.parquet
Jun 11, 2018 10:45:34 AM GMT+0800
20.5 GB
Standard
Viewing 1 to 1

Progress Tracker

- 9 channels
- channels with prefix oc_ for OC-wide monitoring
- channels with prefix glue_ for Glue-wide monitoring


glue_crawler_tracker
glue_job_tracker
oc_cstd_tracker
oc_email_tracker
oc_monitoring
oc_preprocess_tracker
oc_query_tracker
oc_report_tracker
oc_trigger_tracker


KKStream Operations Center APP 8:30 AM
OC DAILY CDN ETL PIPELINE BEGIN
[2018-06-11 00:30:56] running: batch consolidation dispatcher (1/8)
[2018-06-11 00:32:19] running: daily compression trigger (2/8)
KKStream Operations Center APP 8:38 AM
[2018-06-11 00:38:52] running: compression job oc-dailycompression-head-e16msk6i5ensv8 (3/8)
[2018-06-11 00:42:06] running: compression job oc-dailycompression-item-e2fhqsst4k278k (4/8)
KKStream Operations Center APP 9:11 AM
[2018-06-11 01:11:41] running: compression job oc-dailycompression-tail-emwxq5b5py18p (5/8)
KKStream Operations Center APP 10:47 AM
[2018-06-11 02:47:08] running: CDN query executor (6/8)
[2018-06-11 02:47:31] running: CDN report generator (7/8)
[2018-06-11 02:47:42] running: CDN email dispatcher (8/8)
OC DAILY CDN ETL PIPELINE COMPLETE

Outcomes

- Optimization
- Stability
- Scalability
- Accessibility

Optimization

- remove unnecessary data at arrival (1-2% data saved)
- cosolidate small files (40-50% time saved during compression)
- compress using gzip + parquet (99% data saved on single column queries)
- partition using date (30-40% time saved during queries)
- repartition to 1 file for each distribution (70-80% time saved during queries)
- saved ~$100-150NTD per day / ~$50,000NTD per year

Stability

- column number check at arrival (Slack + SMS)
- data size check before compression (Slack)
- compression state tracking (Slack + SMS)
- conditional sequential compression
- crawler state tracking (Slack + SMS)
- conditional query execution
- progress tracking at each stage (Slack)
- overall progress tracking (Slack)
- sum by column for statistics overview and correctness (e-mail)
- 97% up time (down instance: Spark's memory exceeded, divide by zero exception)

Scalibility

- worker Lambda to prevent timeout and improves division of labour
- capable of handling 20+ distributions
- 10 times user traffic? partition by hour/minute/second
- 100 times? process every hour/minute/second
- 10,000 times? real-time processing
- 100,000,000 times? build your own data center

Accessibility

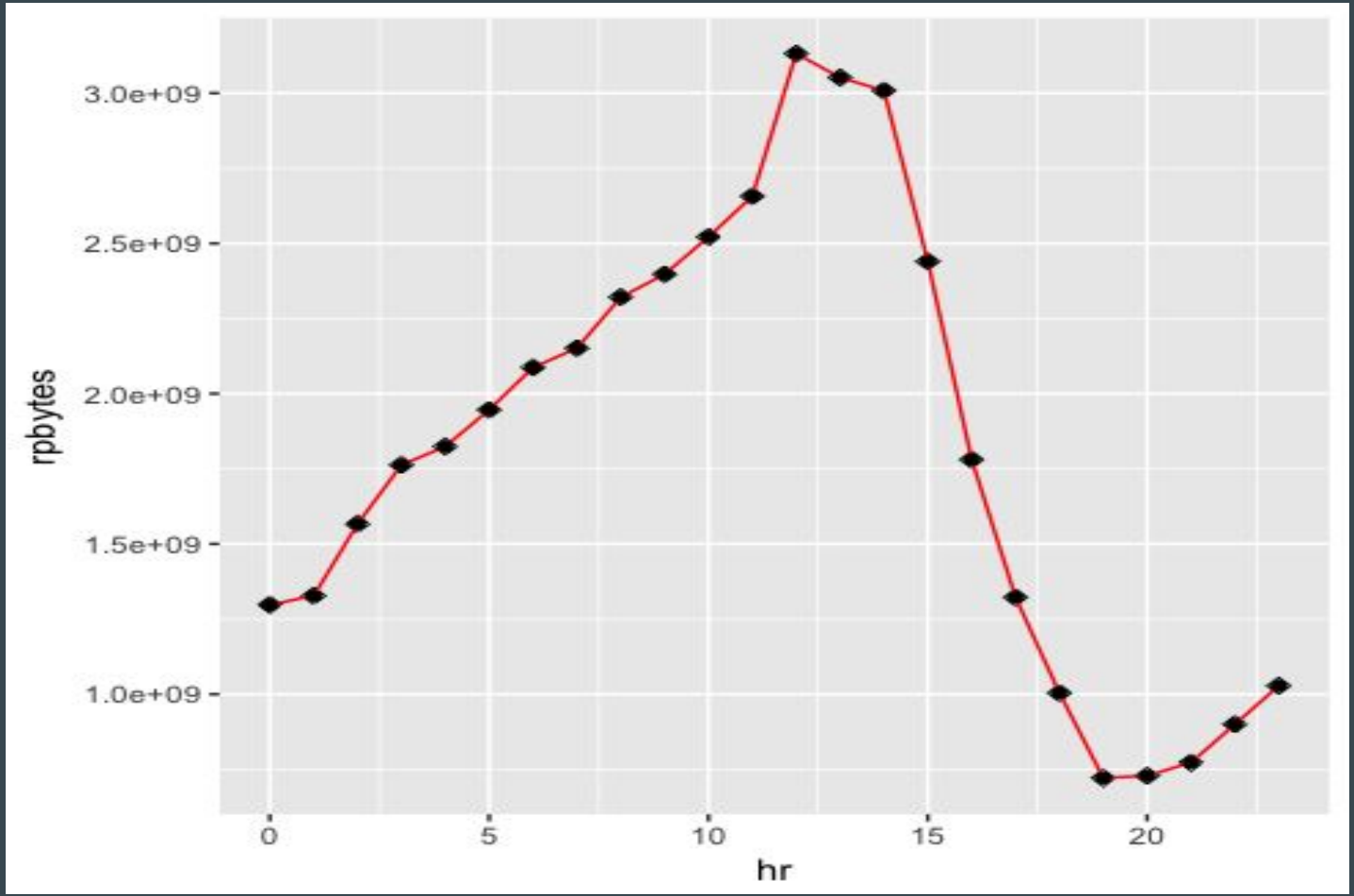
- CloudFront statistics/CloudFront report/CloudWatch report attached to e-mails
- direct report download from s3
- fast and optimized query-ready Athena tables (SRE)
- report generation via Slack commands (OC)
- direct gzipped parquet download from s3 for further analytics/migration
- derived columns for immediate device/os/browser-specific investigations

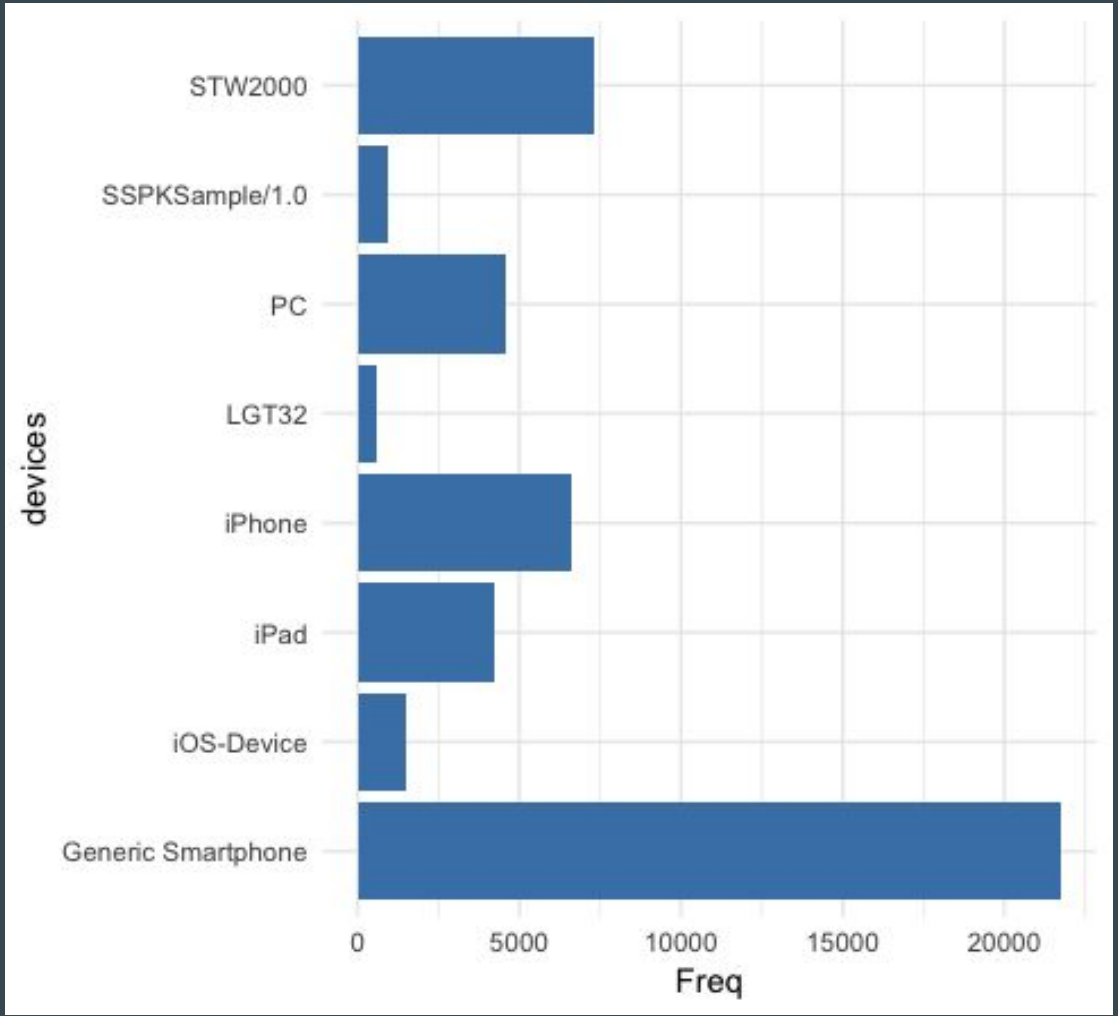
Analysis & Visualization

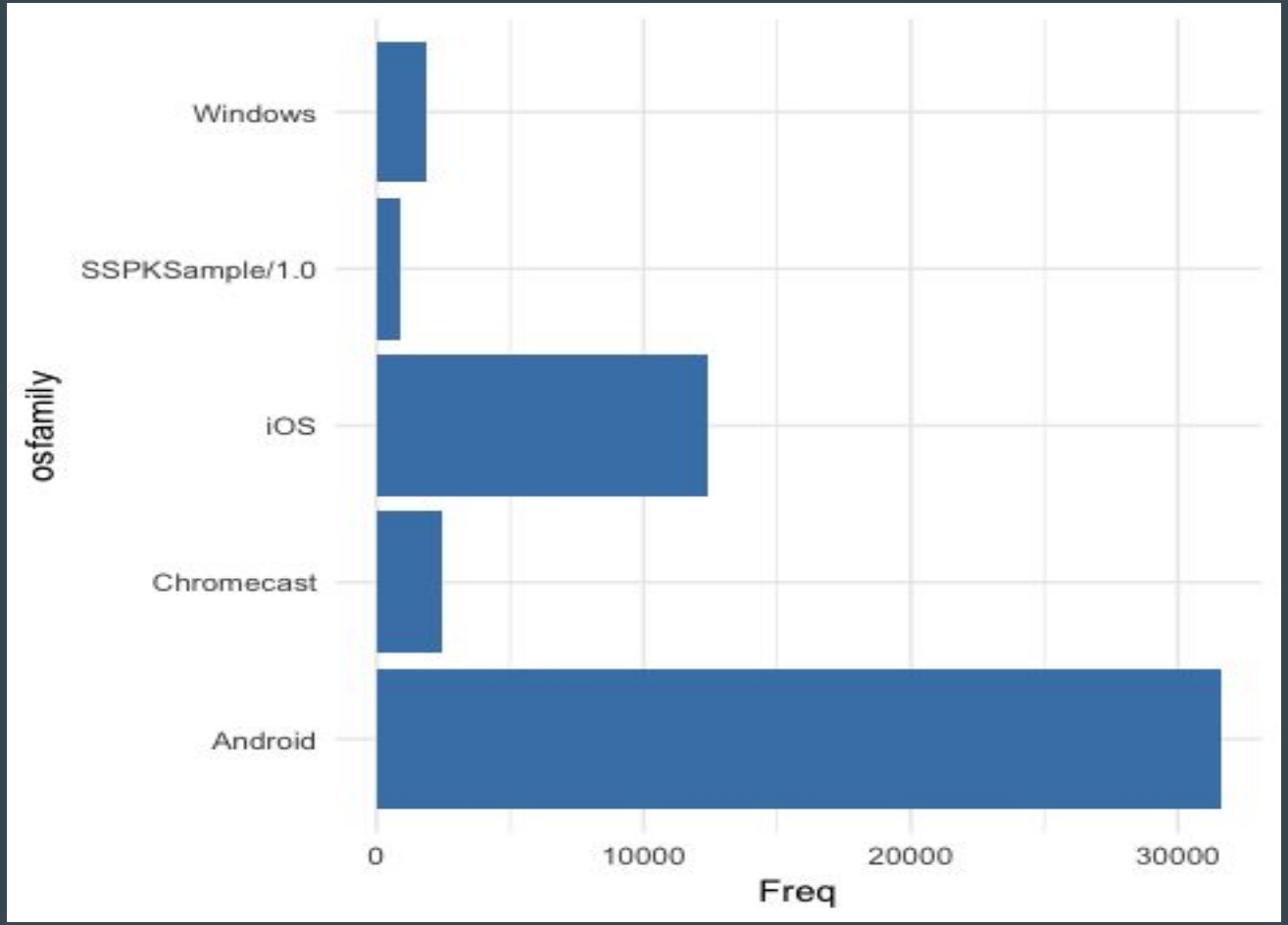
- EDA
- SQL using Amazon Athena
- CDN Statistics
- CDN Report (via raw log)
- CDN Report (via API)
- Statistics Insights
- E-mail

EDA

- sample data of size .0002
- detailed insights for clients
- determine major causes for HTTP miss records


3.0e+09
2.5e+09
2.0e+09
1.5e+09
1.0e+09
rpbytes
0
5
10
15
20
hr


STW2000
SSPKSample/1.0
PC
LGT32
iPhone
iPad
iOS-Device
Generic Smartphone
devices
0
5000
10000
15000
20000
Freq


Windows
SSPKSample/1.0
iOS
Chromecast
Android
osfamily
0
10000
20000
30000
Freq

lhs	rhs	support	confidence	lift	count
{locationstr=NRT57-C1, hr=12,device=PC}	{result0=Miss}	0.000906266	0.6	2.230042505	42
{locationstr=NRT57-C1, hr=10,device=PC}	{result0=Miss}	0.000755222	0.636363636	2.365196597	35
{locationstr=NRT57-C1, hr=13,device=PC}	{result0=Miss}	0.000733644	0.618181818	2.297619551	34
{locationstr=NRT57-C1, hr=14,device=PC}	{result0=Miss}	0.000712066	0.647058824	2.4049478	33
{locationstr=NRT57-C1, hr=11,device=PC}	{result0=Miss}	0.000690489	0.603773585	2.24406793	32
{locationstr=NRT57-C2, hr=11,device=Generic Smartphone}	{result0=Miss}	0.000668911	0.596153846	2.215747361	31

SQL using Amazon Athena

- query engine for S3
- serverless engine that automatically scales
- priced based on data scanned

```
select
    dt as DATE,
    hr as TIME,
    count(*) as REQUEST_COUNT,
    count(case when result0 = 'Hit' then result0 end) as HIT_COUNT,
    count(case when result0 = 'Miss' then result0 end) as MISS_COUNT,
    count(case when result0 in ('LimitExceeded', 'CapacityExceeded', 'Error') then result0 end) as
ERROR_COUNT,
    count(case when status/100 = 2 and result1= 'Error' then status end) as INCOMPLETE_COUNT,
    count(case when status/100 = 2 then status end) as HTTP2XX_COUNT,
    count(case when status/100 = 3 then status end) as HTTP3XX_COUNT,
    count(case when status/100 = 4 then status end) as HTTP4XX_COUNT,
    count(case when status/100 = 5 then status end) as HTTP5XX_COUNT,
    sum(rpbytes) as BYTES_REQUESTED,
    sum(case when result0 = 'Miss' then rpbytes end) as BYTES_MISSED
    from cf_compressed_production where dt between '2018-06-09' and '2018-06-11' and distro in
('e16msk6i5ensv8', 'e2fhqsst4k278k', 'emwxq5b5py18p')
    group by dt, hr
    order by dt, hr asc
```

Run query | Save as | Create view from query | (Run time: 8.2 seconds, Data scanned: 2.98GB) | Format query | Clear

Use Ctrl + Enter to run query, Ctrl + Space to autocomplete

Results

	DATE	TIME	REQUEST_COUNT	HIT_COUNT	MISS_COUNT	ERROR_COUNT	INCOMPLETE_COUNT	HTTP2XX_COUNT	HTTP3
1	2018-06-09	00	9826190	7550222	2272501	3152	31813	9502627	310056
2	2018-06-09	01	10658743	8090744	2564054	3587	31918	10310057	340761
3	2018-06-09	02	11394144	8603274	2786668	3982	33740	11033475	352670
4	2018-06-09	03	12543512	9488518	3050366	4312	39648	12127299	406182
5	2018-06-09	04	13139083	9803853	3329887	5012	41554	12716056	412776
6	2018-06-09	05	13800980	10311916	3483866	4923	42960	13385134	402000

CDN Statistics

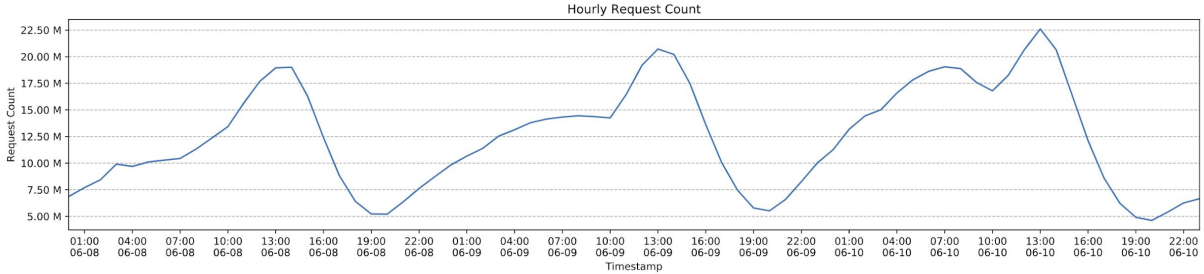
- aggregated by hour
- contains data within last 72 hours (3 days)
- 13 columns
- automatically attached to e-mail

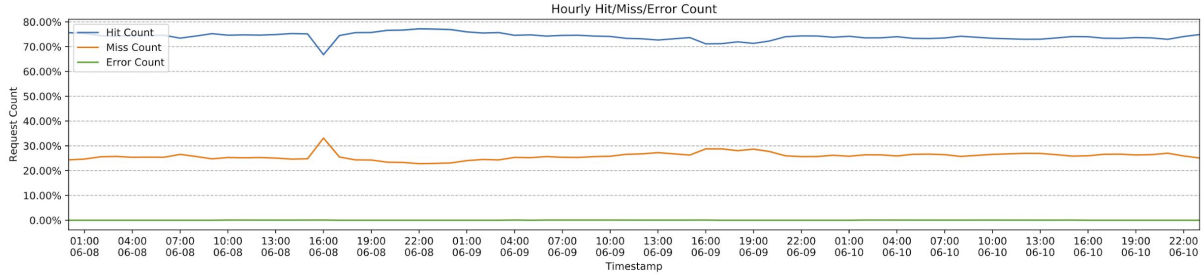
DATE	TIME	REQUEST_COUNT	HIT_COUNT	MISS_COUNT	ERROR_COUNT	INCOMPLETE_COUNT	HTTP2XX_COUNT	HTTP3XX_COUNT	HTTP4XX_COUNT	HTTP5XX_COUNT	BYTES_REQUESTED	BYTES_MISSED
2018/6/8	0	6855133	5187941	1665132	1956	19897	6618125	231834	53	0	5.88426E+12	1.69332E+12
2018/6/8	1	7701624	5800475	1898701	2353	23824	7436867	258593	79	0	6.92179E+12	1.99574E+12
2018/6/8	2	8440530	6275475	2161986	2897	24551	8169888	259521	237	0	7.64499E+12	2.21914E+12
2018/6/8	3	9907842	7356793	2547298	3587	33475	9591708	306640	15	0	8.62388E+12	2.49422E+12
2018/6/8	4	9692986	7228630	2461151	3028	28146	9392213	293240	47	0	8.89562E+12	2.53427E+12
2018/6/8	5	10107640	7534995	2569575	2936	30938	9815117	284006	25	0	9.21635E+12	2.65671E+12
2018/6/8	6	10277146	7661053	2613001	2994	31305	9967166	301100	53	0	9.3653E+12	2.71782E+12
2018/6/8	7	10435368	7658824	2773025	3414	27580	10115658	309998	49	0	9.3943E+12	2.80184E+12
2018/6/8	8	11316945	8405396	2907467	3911	34779	10962683	344501	108	0	1.00062E+13	2.81531E+12
2018/6/8	9	12364231	9299349	3060051	4607	42998	11960139	392066	76	340	1.07032E+13	2.98011E+12
2018/6/8	10	13427469	10026713	3395404	5100	52717	12974939	439503	136	0	1.11943E+13	3.15046E+12
2018/6/8	11	15648850	11697917	3943818	6806	62270	15124917	504843	76	356	1.28983E+13	3.63883E+12
2018/6/8	12	17678916	13201181	4470167	7257	85840	17114684	543727	81	320	1.47027E+13	4.13072E+12
2018/6/8	13	18945169	14188598	4749197	7001	76941	18393088	532982	130	2	1.59763E+13	4.39704E+12
2018/6/8	14	19013997	14320330	4686541	6711	76147	18499582	492980	118	0	1.63098E+13	4.36758E+12
2018/6/8	15	16297794	12247607	4043642	5630	62625	15927314	353828	121	0	1.42221E+13	3.80762E+12
2018/6/8	16	12398680	8282254	4110732	5419	45665	12113614	272661	62	72	1.10174E+13	4.04817E+12
2018/6/8	17	8823037	6569687	2250086	2954	23414	8617933	198642	40	0	7.88436E+12	2.23429E+12
2018/6/8	18	6410913	4845422	1563082	1990	14522	6235577	171414	50	0	5.73226E+12	1.56707E+12
2018/6/8	19	5229728	3956318	1271863	1446	10516	5064325	162942	31	0	4.60534E+12	1.28664E+12
2018/6/8	20	5213583	3991866	1219975	1605	13666	5020792	188248	44	0	4.40122E+12	1.17554E+12
2018/6/8	21	6358049	4872845	1483432	1666	17080	6133861	220681	30	0	5.2457E+12	1.39163E+12
2018/6/8	22	7624492	5885771	1736303	2165	25605	7359219	258527	41	0	6.32448E+12	1.67928E+12
2018/6/8	23	8736914	6728834	2004760	2815	29170	8444409	284855	105	0	7.62917E+12	1.99721E+12
2018/6/9	0	9826190	7550222	2272501	3152	31813	9502627	310056	33	0	8.7461E+12	2.30246E+12
2018/6/9	1	10658743	8090744	2564054	3587	31918	10310057	340761	108	0	9.69001E+12	2.68222E+12
2018/6/9	2	11394144	8603274	2786668	3982	33740	11033475	352670	114	0	1.04573E+13	2.94564E+12
2018/6/9	3	12543512	9488518	3050366	4312	39648	12127299	406182	48	0	1.12281E+13	3.14259E+12
2018/6/9	4	13139083	9803853	3329887	5012	41554	12716056	412776	81	0	1.20279E+13	3.47376E+12
2018/6/9	5	13800980	10311916	3483866	4923	42960	13385134	402000	31	3	1.28262E+13	3.68415E+12
2018/6/9	6	14142825	10503457	3633567	5621	49661	13731994	396435	158	1	1.32625E+13	3.82216E+12
2018/6/9	7	14333205	10686023	3641116	5898	55255	13916318	401378	107	306	1.32723E+13	3.7363E+12
2018/6/9	8	14446577	10779925	3660605	5908	49751	14011879	420717	55	9	1.3433E+13	3.78079E+12
2018/6/9	9	14374493	10677988	3690453	5706	51973	13909148	452626	80	101	1.29878E+13	3.69129E+12
2018/6/9	10	14247987	10565812	3676513	5374	51317	13783521	450443	47	15	1.25091E+13	3.54066E+12
2018/6/9	11	16441845	12060679	4373725	7055	59421	15912929	514487	100	287	1.41086E+13	4.1562E+12
2018/6/9	12	19196868	14045325	5143302	7862	80253	18613668	563286	116	0	1.62505E+13	4.79259E+12
2018/6/9	13	20718952	15060697	5649731	8130	78286	20150111	549532	109	0	1.76679E+13	5.27402E+12
2018/6/9	14	20207454	14786561	5413020	7474	77460	19673881	512348	61	0	1.72599E+13	5.03411E+12
2018/6/9	15	17497237	12885190	4604197	6342	64513	17087279	395389	97	0	1.53759E+13	4.40729E+12
2018/6/9	16	13618811	9692840	3919631	5100	42360	13299950	308993	89	0	1.22048E+13	3.61778E+12
2018/6/9	17	10045094	7154862	2886023	4055	27805	9812238	225823	91	0	9.12818E+12	2.76635E+12
2018/6/9	18	7453498	5357458	2093437	2454	19422	7271617	176529	80	0	6.76571E+12	2.05063E+12
2018/6/9	19	5788541	4131328	1655329	1812	12515	5620656	164871	21	0	5.18238E+12	1.62079E+12

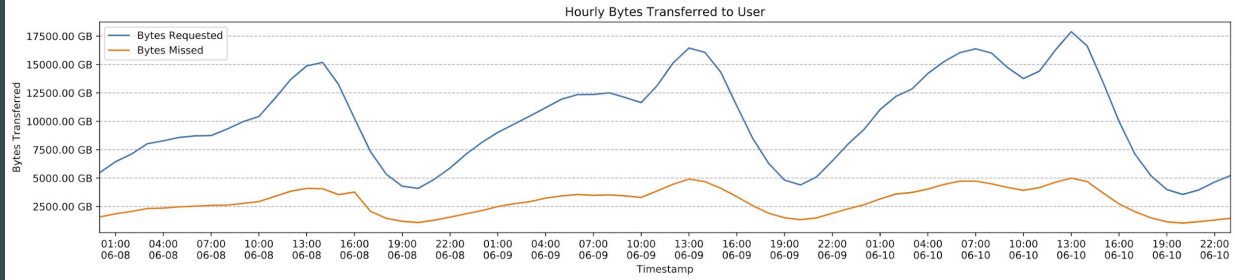
CDN Report (via raw log)

- plotted using CDN Statistics
- matches graphs using AWS console
- automatically attached to e-mail

CDN Cache Report (2018-06-08 through 2018-06-10)



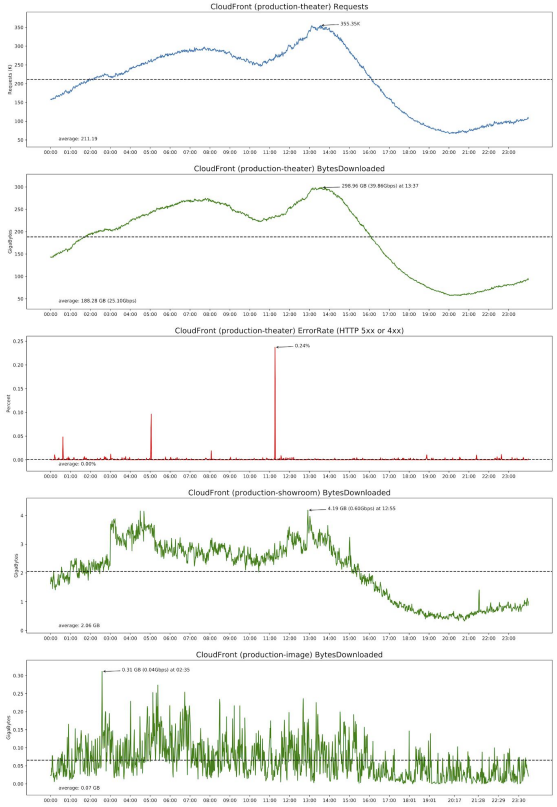




CDN Report (via API)

- plotted using CloudWatch function
- generated by daily Lambda bot
- automatically attached to e-mail

VideoPass CDN Cache Statistics - June 10 2018



Statistics Insights

- generated at e-mail creation
- Responsive Web Design (RWD) with HTML and CSS
- provides quick overview of statistics trends
- automatically generated upon e-mail creation

MEARSURES	Jun-08	Jun-09	TREND	Jun-09	Jun-10	TREND
Request Count (M)	258.92	304.25	+17.51%	304.25	332.62	+9.32%
Hit Count (M)	193.25	224.67	+16.26%	224.67	244.76	+8.94%
Miss Count (M)	65.58	79.44	+21.13%	79.44	87.70	+10.40%
Error Count (K)	90.26	113.34	+25.57%	113.34	125.33	+10.58%
Incomplete Count (K)	893.69	1024.45	+14.63%	1024.45	1166.36	+13.85%
HTTP 2XX Count (M)	251.04	295.24	+17.61%	295.24	323.04	+9.42%
HTTP 3XX Count (K)	7607.33	8725.67	+14.70%	8725.67	9253.94	+6.05%
HTTP 4XX Count	1807.00	1862.00	+3.04%	1862.00	2175.00	+16.81%
HTTP 5XX Count	1090.00	722.00	-33.76%	722.00	1283.00	+77.70%
Bytes Requested (GB)	209360.64	251622.38	+20.19%	251622.38	274131.14	+8.95%
Bytes Missed (GB)	59400.30	72729.38	+22.44%	72729.38	78325.85	+7.69%

E-mail

- Japanese body provided by OC Team
- automatically fills in peak statistics
- automatically attaches required files
- BCC for better development/production seperation

KKstream

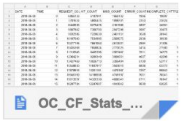
お世話になっております。

早速ですが、日次の通信データ容量を添付いたしました。
資料により最大ピークは 298.96GB (39.86Gbps) at 13:37 です、ご確認お願いします。

以上、よろしくお願いします。

MEARSURES	Jun-08	Jun-09	TREND	Jun-09	Jun-10	TREND
Request Count (M)	258.92	304.25	+17.51%	304.25	332.62	+9.32%
Hit Count (M)	193.25	224.67	+16.26%	224.67	244.76	+8.94%
Miss Count (M)	65.58	79.44	+21.13%	79.44	87.70	+10.40%
Error Count (K)	90.26	113.34	+25.57%	113.34	125.33	+10.58%
Incomplete Count (K)	893.69	1024.45	+14.63%	1024.45	1166.36	+13.85%
HTTP 2XX Count (M)	251.04	295.24	+17.61%	295.24	323.04	+9.42%
HTTP 3XX Count (K)	7607.33	8725.67	+14.70%	8725.67	9253.94	+6.05%
HTTP 4XX Count	1807.00	1862.00	+3.04%	1862.00	2175.00	+16.81%
HTTP 5XX Count	1090.00	722.00	-33.76%	722.00	1283.00	+77.70%
Bytes Requested (GB)	209360.64	251622.38	+20.19%	251622.38	274131.14	+8.95%
Bytes Missed (GB)	59400.30	72729.38	+22.44%	72729.38	78325.85	+7.69%

3 Attachments


OC_CF_Stats_...


OC_CF_Report...


cf-cdn-cache-2...

Demo

Questions?