Updated:
April 28, 2020

야놀자 BigQuery 사용사례

Google Cloud

Cloud Innovator MEGAZONE

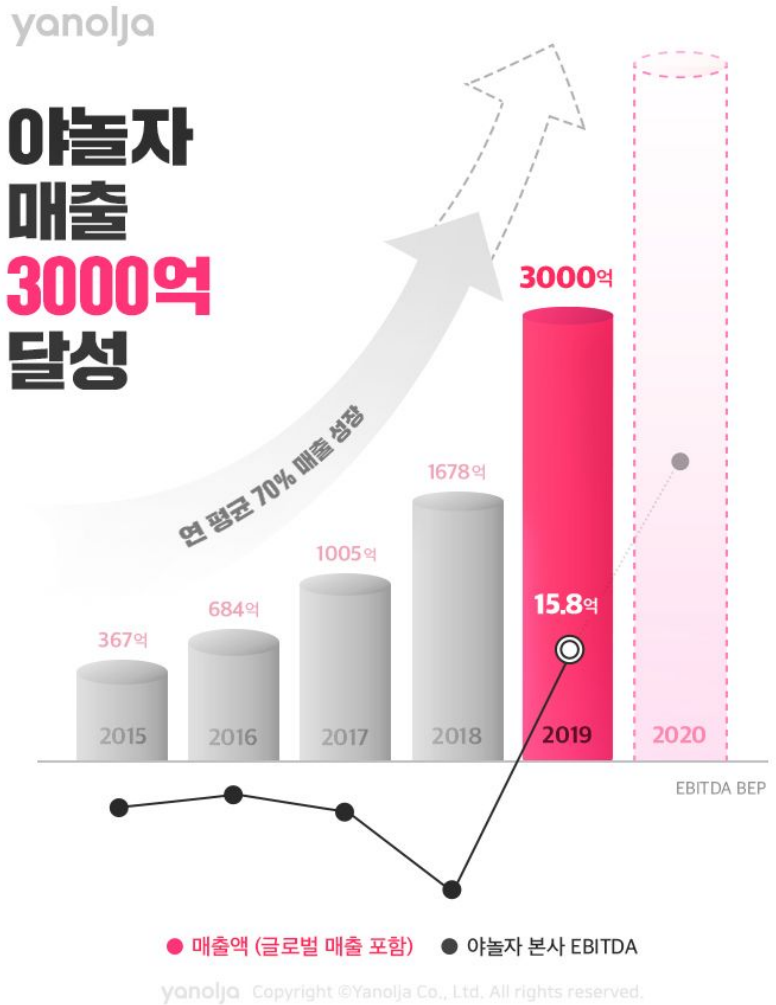
Contents

- yanolja is evolving
- Requirements
- Log parsing and collecting
- Metric analytics
- Reporting
- BigQuery evolves
- Next steps

yanolja is evolving

Global R.E.S.T Platform

- 재충전(Rest)과 재미(Entertain)를 극대화하는 좋은 숙박(Stay)을 기반으로, 액티비티와 문화까지, 사용자에게 여행(Travel)의 새로운 경험과 가치를 선사합니다.



Google Cloud

Requirements

01

Stackdriver Logging

보안 로그를 수집하고,
분석하고, Report 를
생성하는데 있어
많은 운영 resource 를 들이지
않고 구축하고 싶다

02

Fluentd

WAF Appliance 에서 발생하는
delimiter 기반 비표준 로그를
직접
정형화 하여(JSON)
저장하고 싶다

03

Data Studio

기간 별 Report 를
자동으로 생성하고 싶다

04

BigQuery

기존의 SIEM 의
제한적인 분석보다는
다양한 metric 을
직접 수정하며
빠르게 분석하고 싶다

Log parsing and collecting

```
DETECT|WAF|2019-12-02
16:26:09|10.1.**.10|10.0.**.126|KR|54.36.***.197|FR|10.2.**.34|443|Web|Scanner/Proxy/Spambot|www.yanolja.com|/|
GET / HTTP/1.1
X-Forwarded-For: 54.36.***.197
X-Forwarded-Proto: https
X-Forwarded-Port: 443
Host: www.yanolja.com
X-Amzn-Trace-Id: Root=1-5d***c92-54f766******e2873a9fec74
User-Agent: Mozilla/5.0 (compatible)
Accept: */*
Accept-Encoding: deflate, gzip
```

3rd Party WAF 어플라이언스에서보내주는 탐지 로그 샘플

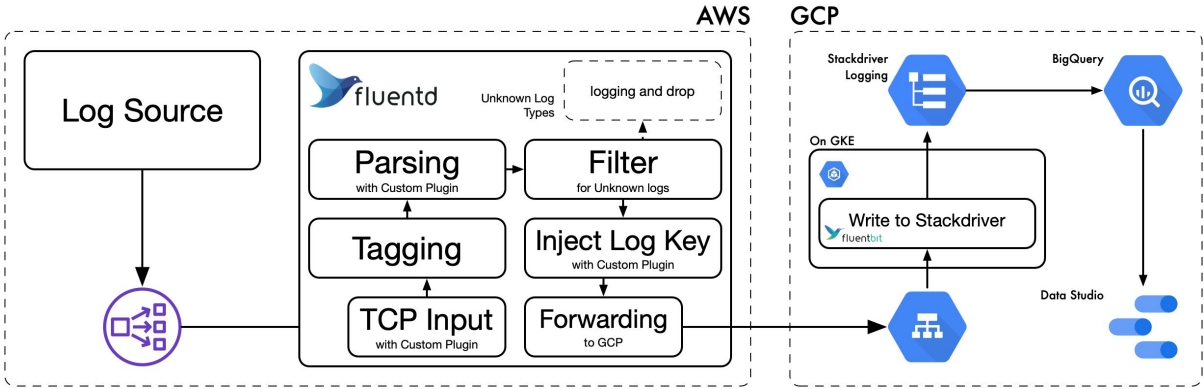
Issues

- delimiter 로 '|' 를 사용
- WAF 특성 상 http request 내용이 payload 에 포함 (carriage return 값에 대한 concatenation 필요)
- DETECT 외 불필요 로그 항목들

Log parsing and collecting

Fluentd 에서 전달된 로그를 파싱

- **Tagging** : 용도별 로그에 태깅
- **Parsing** : delimiter (|) 및 payload 중간의 HTTP Request 부분을 구분
- **Filter** : 태깅된 DETECT 로그만 필터
- **Forwarding** : GCP 인프라 내 Stackdriver 로 전달

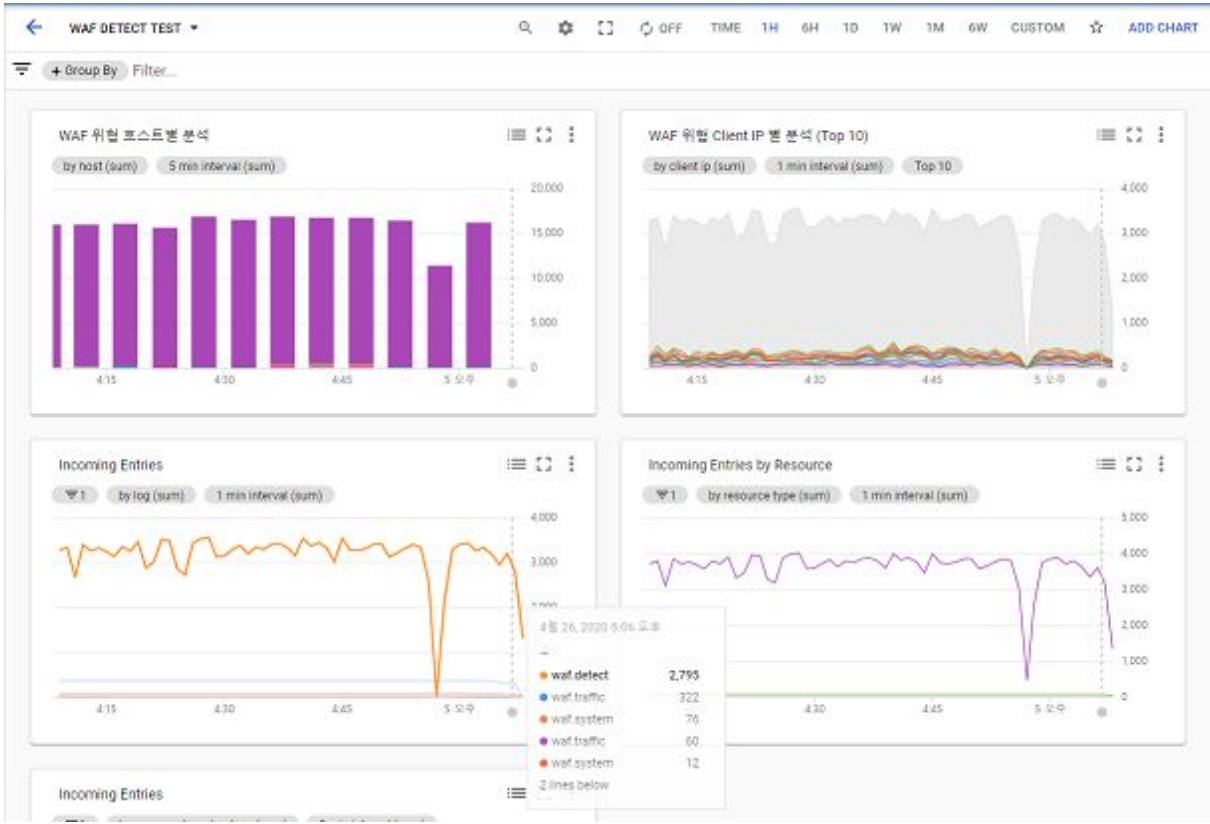


Metric analytics

Stackdriver 모니터링 > 대시보드 이용

- 👍 로그 기반 실시간 모니터링 가능
- 👎 필요한 metric을 사용자 정의 측정 항목으로 미리 지정해야 함
- 👎 측정 항목으로 지정하기 이전 데이터에 대한 분석은 어려움
- 👎 X축은 시계열로 고정됨

⇒ 로그 라우터를 이용하여
수집한 Stackdriver Logging 데이터를
BigQuery 로 매우 쉽게 전달



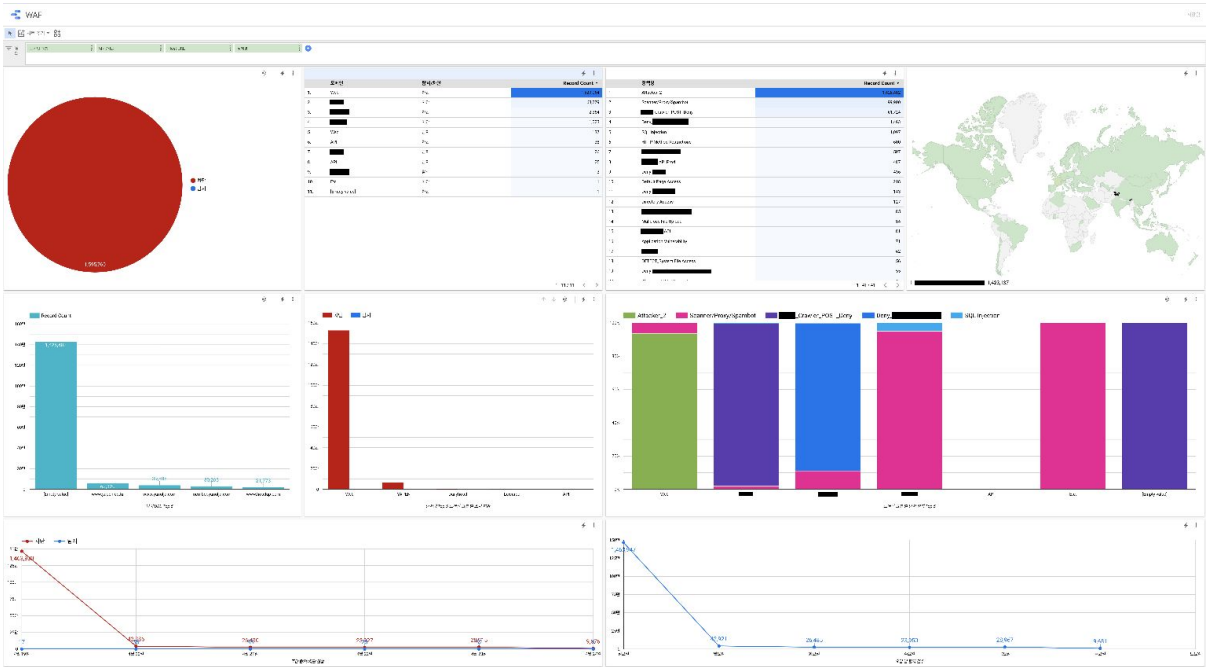
Google Cloud

Cloud Innovator MEGAZONE

Reporting

BigQuery 에 적재한 테이블을
데이터 원본으로 설정,
Data Studio 에서 보고서 생성

- 👍 BigQuery 테이블에서
  Metric 을 뽑아 다양한
  시각화 구성을 편리하게 설정
- 👍 특정 항목 클릭하여
  다른 지표들을 같이 필터링해
  연관된 내용을
  빠르게 확인 가능



Google Cloud

Cloud Innovator MEGAZONE

BigQuery evolves

로그 정보가 BigQuery 테이블로 생성되면서 아래와 같은 작업도 가능해짐

- 여러 WAF에서 수집된 로그를 한 곳에서 쉽게 Query 가능
- 수집된 Field 별로 추가적으로 조건을 줄 수 있어, SIEM에서 조회 가능했던 기능들을 동일하게 구현 가능, 좀 더 빠른 쿼리 결과
    - 예: 지난 1주일 간 대량 탐지된 Client IP 를 탐지된 수로 정렬
    - 예: 대량 탐지된 Client IP 를 C Class 기준으로 정렬
- 수행한 내용을 파일 등으로 저장할 수 있어 2차 가공도 가능

Google Cloud

Cloud Innovator MEGAZONE

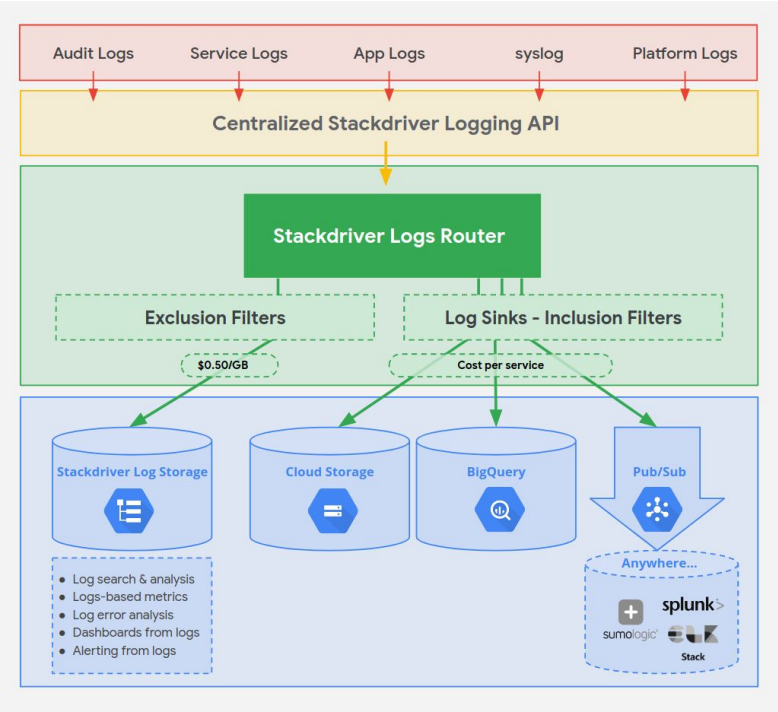
BigQuery evolves

기존에 로그 데이터를 수집하고, 시각화하고, 분석하기 위해서는 여러 시스템들을 구축하고 운영해야 했으나 BigQuery 로는 별도의 Cluster 를 운영하지 않고도 같은 효과 가능

- 사용중인 데이터 수집, 분석 관련 시스템
  - Elastic Search, EMR, Zeppelin, Re:dash ...

Log Router 를 이용하여 필요한 로그만 선별적으로 저장 가능

- 로그 수집 인터페이스 통합
- 로그 저장 비용 절감
- 대상 로그를 줄여서 저장함으로써 좀 더 빠른 분석에 도움



Next steps

1. WAF Log 수집/분석 PoC
2. Data Studio 를 이용한 주기별 분석 보고서
3. Application Logging PoC
4. Metric 모니터링 Dashboard PoC

Thank you.