Datadog Korea User Group

Datadog 한국 사용자 모임

Datadog과 함께 성장한 여정, 꿀팁과 노하우

인프랩

DevOps 정인호

안녕하세요, 정인호입니다.

정인호 (조슈아)

DevOps 엔지니어

인프랩

2019년 입사, 백엔드 + 인프라 업무 담당
2021년 데브옵스

지난 3년 동안 경험한 Datadog의 좋았던 점,
그리고 조심해야하는 점을 이야기하고자 합니다.

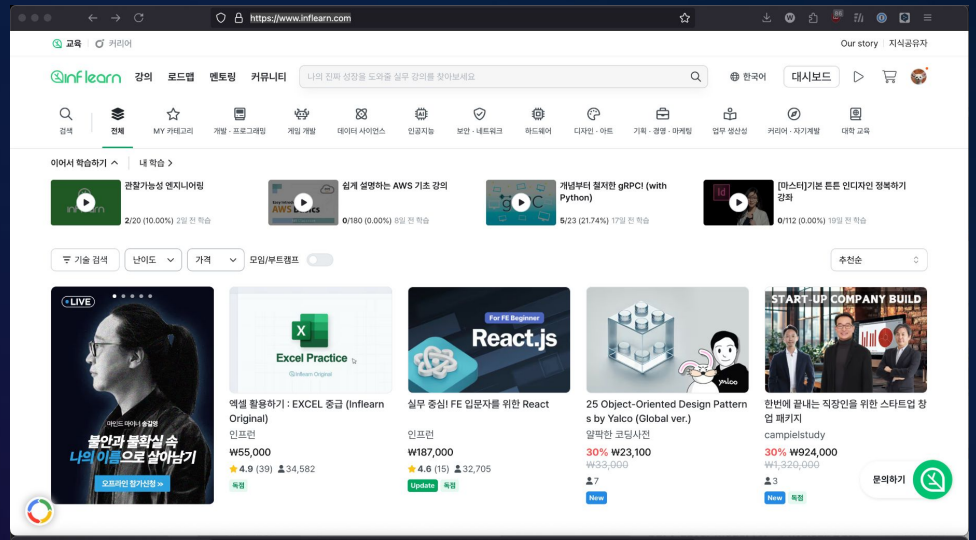
순서

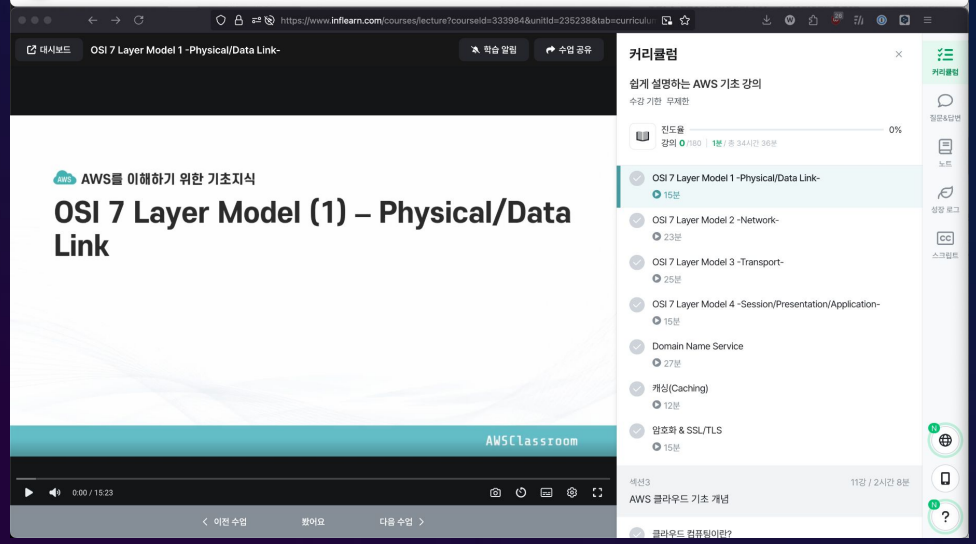
- 인프런(Inflearn)과 랠릿(Rallit)
- Datadog 첫 시작
- 잘 사용하고 있는 Datadog 제품 활용 사례
- Datadog 자체를 잘 써보기
- 마무리

인프런 (Inflearn) 과 랠릿 (Rallit)

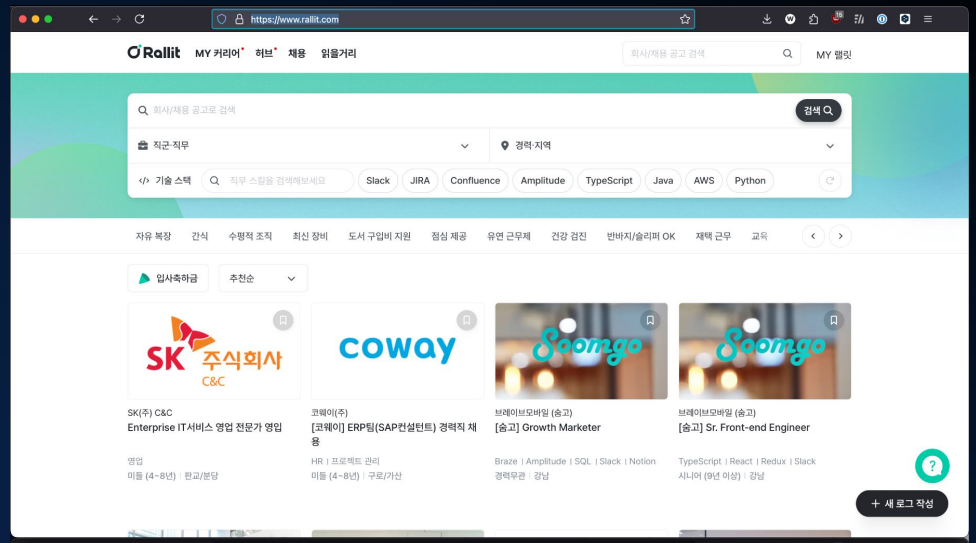
인프런과 랠릿

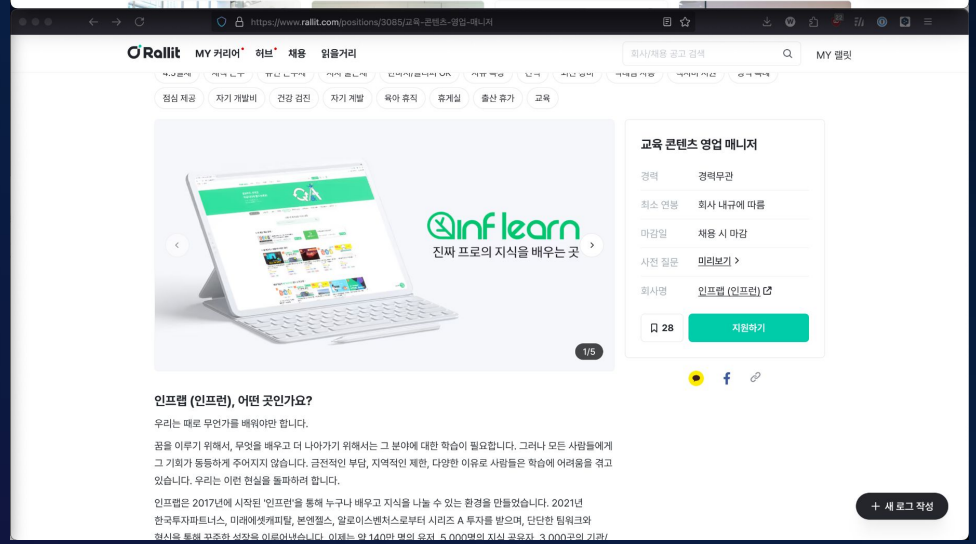
인프런: 라이프타임 커리어 플랫폼





랠릿: IT 인재 채용 플랫폼





인프런과 랠릿

데브옵스 관점

인프런: 온라인 동영상 학습 플랫폼

- Node.js, Kotlin 기반
- API 호출 대량 발생
- 동영상 인코딩 및 암호화, 그리고 자막과 더빙을 위한 배치성 파이프라인 작업
- 동영상 탈취 시도 방어

랠릿: 취업 채용 연결 플랫폼

- Node.js 기반
- 구직자 정보 암호화 및 접근제어 필수

- 분산형 서비스 구조로 서비스간 호출 빈번

인프런과 랠릿

Datadog 관점

인프런: 온라인 동영상 학습 플랫폼

- Application Performance Monitoring (APM)
- Database Monitoring (DBM)
- Datastream Monitoring (DSM)
- CI Visibility
- Log
- ASM

랠릿: 취업 채용 연결 플랫폼

- APM
- Cloud SIEM

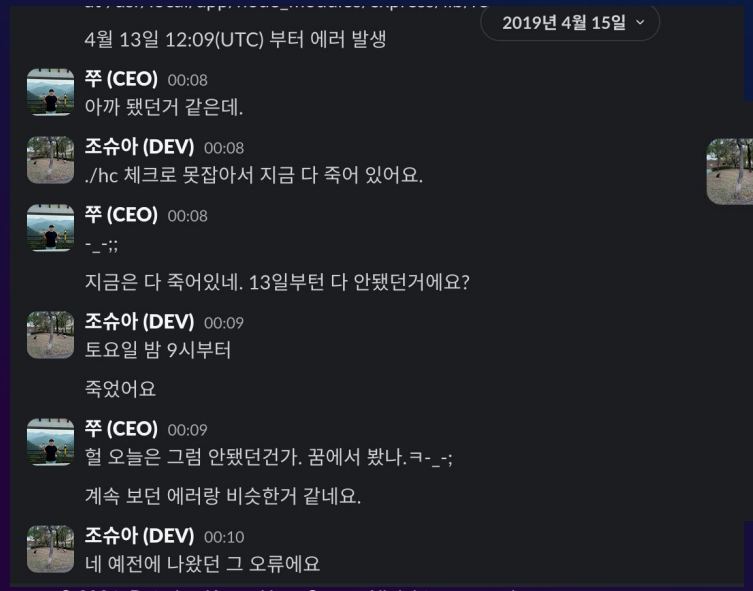
- Service Catalog, Service Map
- Monitoring

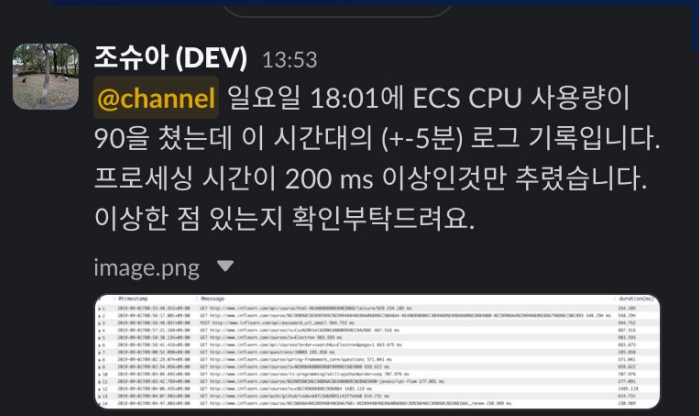
Datadog 첫 시작

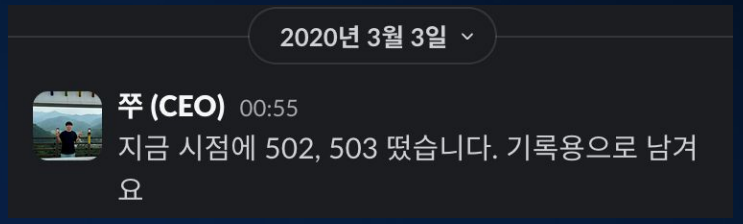
APM 도입의 필요성

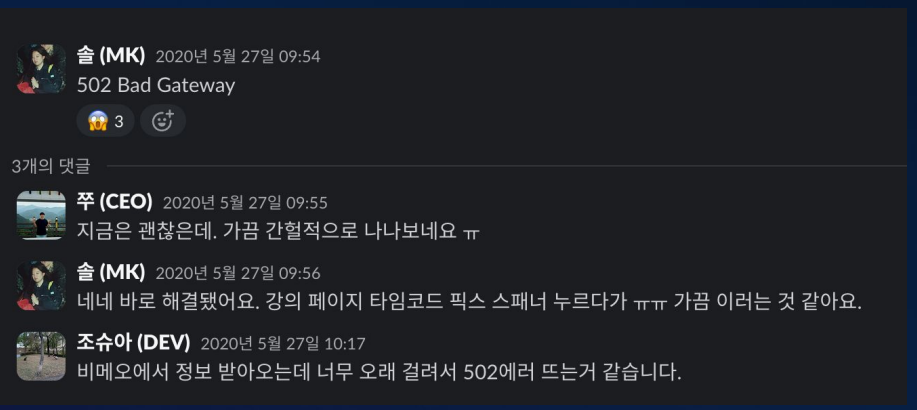
장애가 발생했는데 원인을 모르겠어요.

- 서비스 성장에 따라 같이 빈번해지는 장애와 오류
- 장애 원인을 분석하기 위한 분석 및 모니터링 도구의 부재
- AWS CloudWatch Log와 Alarm만으로는 다양한 장애의 원인을 제때 파악하고 대응하기 어려움





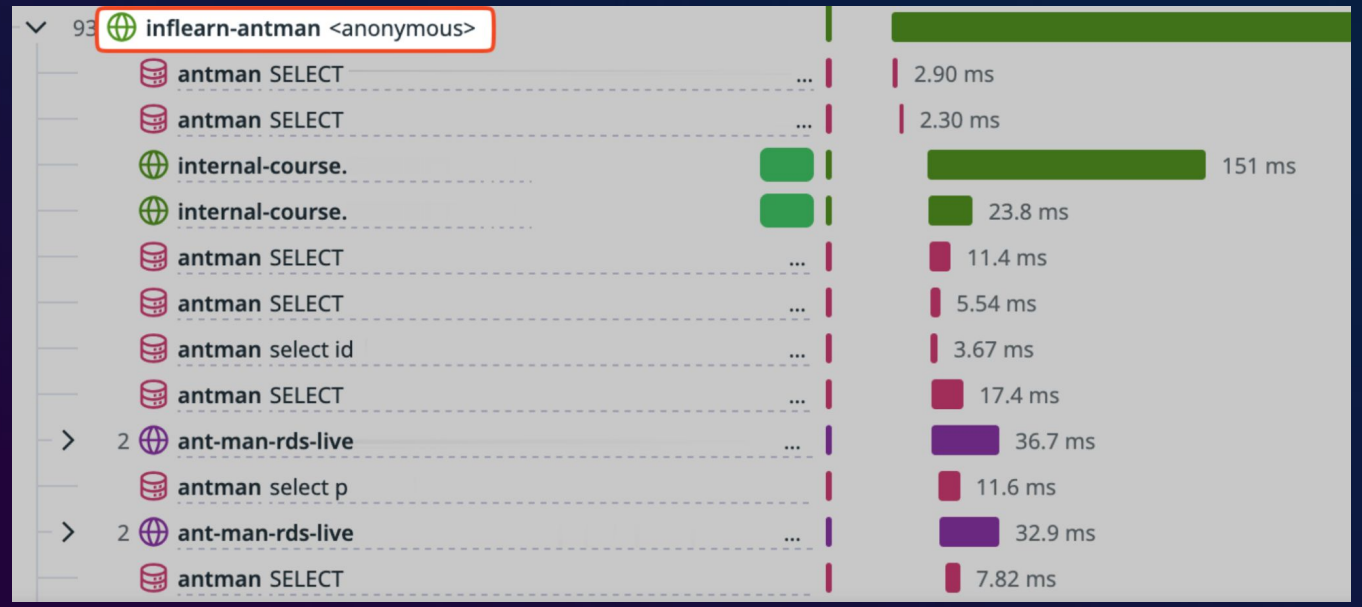




휴먼 경보 운영하던 시절

Datadog을 선택한 이유

- 당시 Node.js 를 가장 잘 지원
  - 익명(anonymous) 함수에 대해 자동 span 유일하게 지원



Datadog 첫 날

2021년 7월 22일

조슈아 (DEV) 15:26
APM 트레이스 배포하고 정상 동작 중입니다.
메트릭이랑 로그 통합하기 전에
개발 파트 가이드용으로 트레이스 사용 문서 작성하고
안내할까요? (+ 개발파트원 데이터독 회원 가입)

향로 (CTO) 15:37
네 가이드 문서가 있어야할것같아요.
다들 APM이 처음이셔서

조슈아 (DEV) 15:38
옙

1

넵! 2

2021년 7월 22일

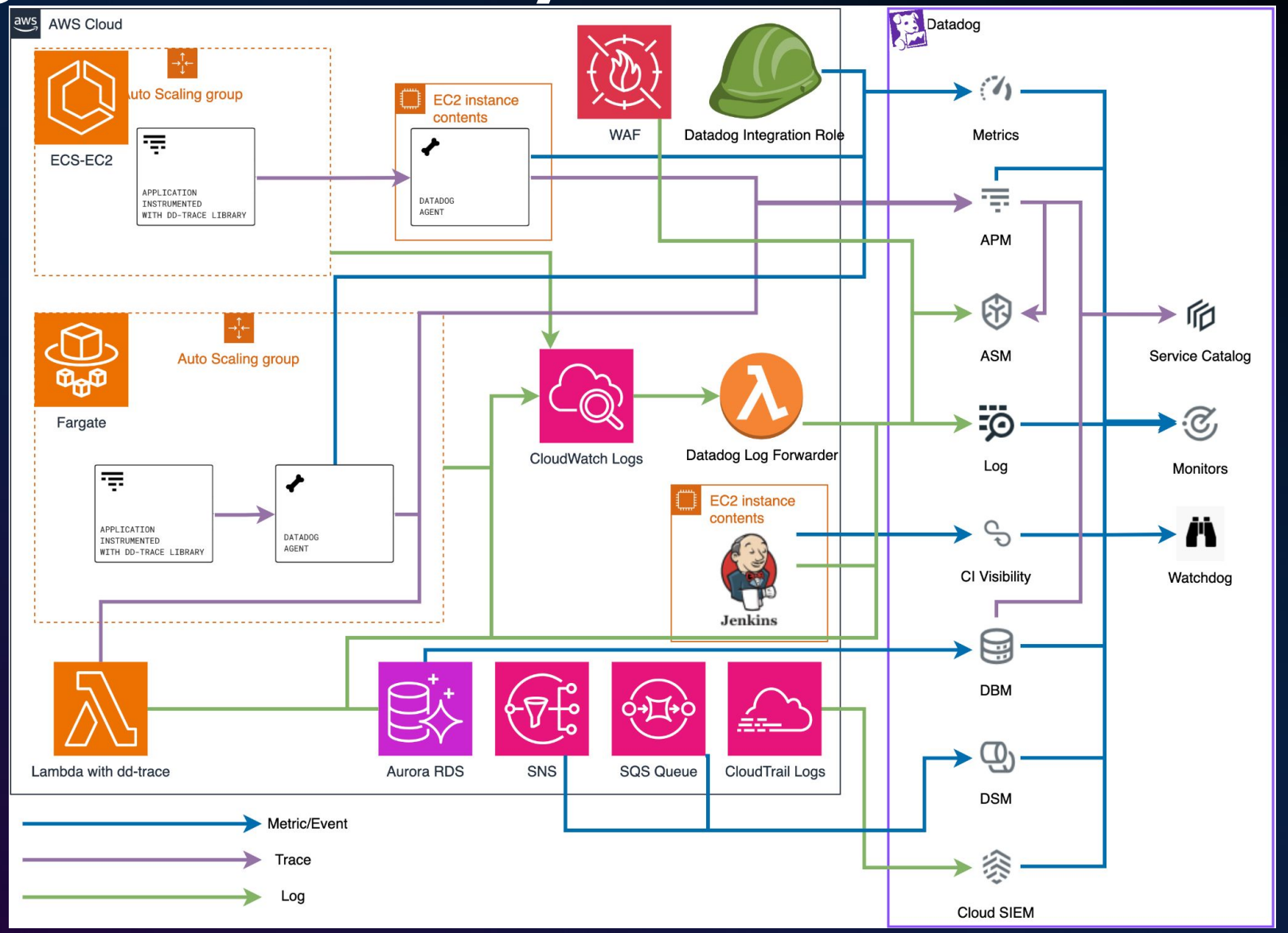
조슈아 (DEV) 15:54
@p-dev Datadog 서비스에 초대드릴 예정입니다. 이메일로 초대장이 발송될거에요.
회원 가입 후 이름을 슬랙 형식으로 입력해주세요~

넵! 1

3개의 댓글 4년 전 마지막 댓글

잘 사용하고 있는 Datadog 제품 활용 사례

Datadog Observability Architecture



APM과 Log

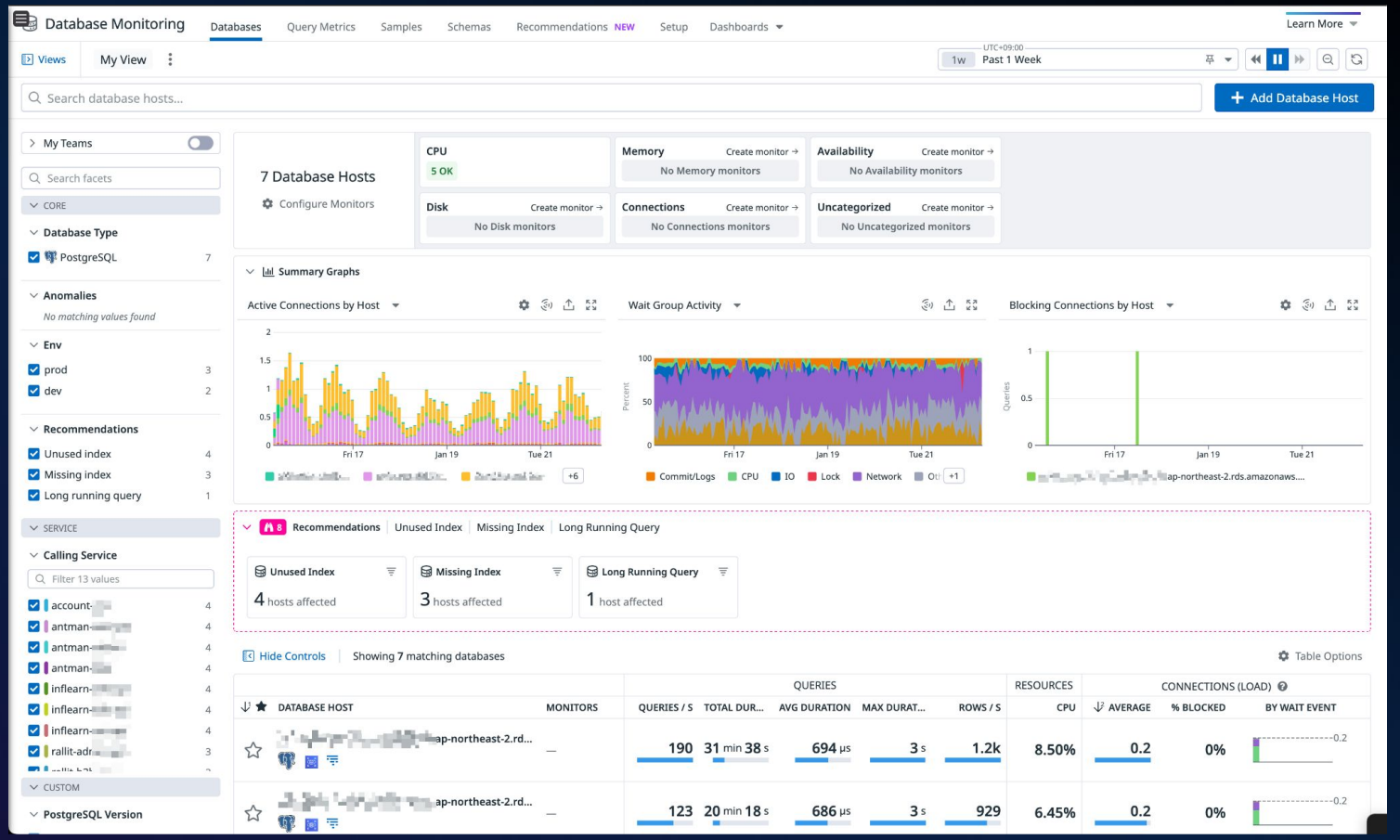
다들 너무 잘 쓰고 계실 것이라 생각합니다.
따라서...

SKIP

APM과 Log는 발표 후반부의 **Datadog 잘 써보기**에서 등장합니다.

Database Monitoring (DBM)

- 디비의 성능을 끌어올리는 것에 집중된 서비스
- 스키마 뷰어, 쿼리 메트릭, 쿼리 샘플링, index 생성 추천 등 다양한 기능 제공
- APM과 연동

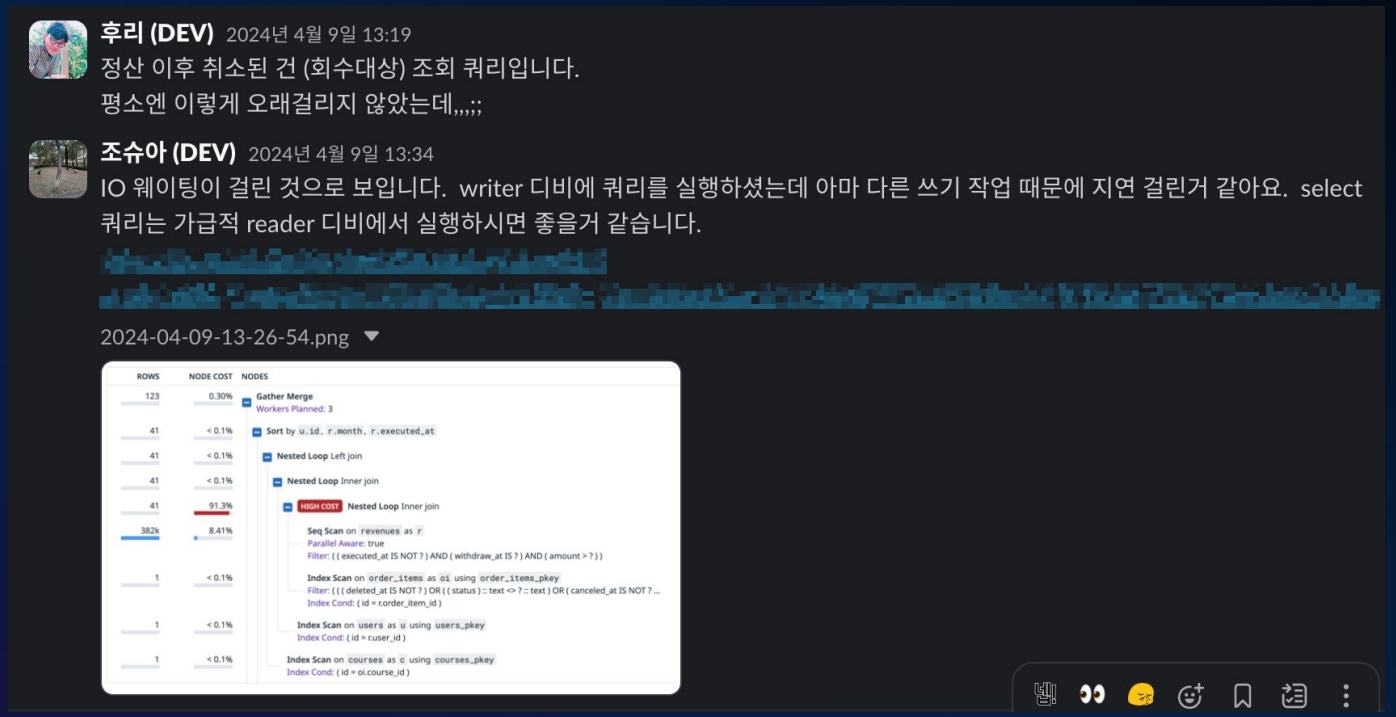


Database Monitoring (DBM)

활용 사례 1 - Query Explain Plan

- 실제 실행된 쿼리 대상으로 explain plan을 제공하여 쿼리가 느린 이유를 식별

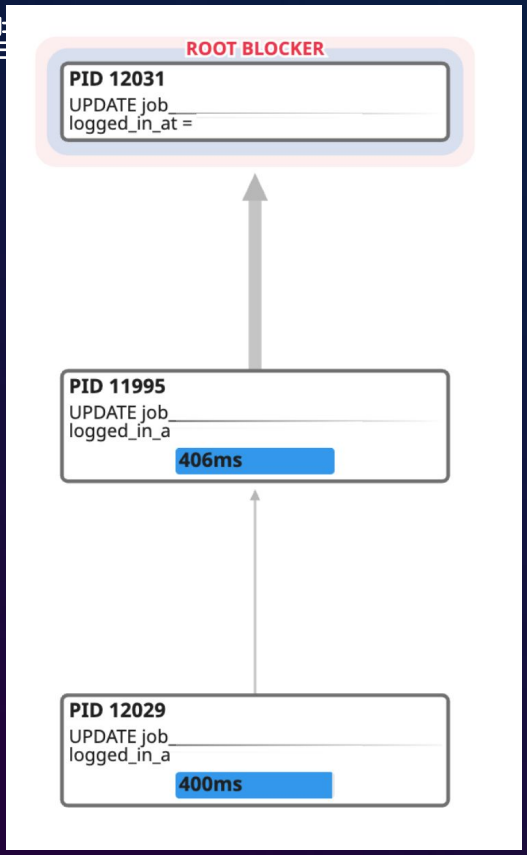
ROWS	NODE COST	NODES
123	0.30%	**Gather Merge** Workers Planned: 3
41	< 0.1%	**Sort** by `u.id`, `r.month`, `r.executed_at`
41	< 0.1%	**Nested Loop** Left join
41	< 0.1%	**Nested Loop** Inner join
41	91.3%	HIGH COST **Nested Loop** Inner join
382k	8.41%	**Seq Scan** on `revenues` as `r` Parallel Aware: true Filter: ((executed_at IS NOT ?) AND (withdraw_at IS ?) AND (amount > ?))
1	< 0.1%	**Index Scan** on `order_items` as `oi` using `order_items_pkey` Filter: (((deleted_at IS NOT ?) OR ((status) :: text <> ? :: text) OR (canceled_at IS NOT ? ... Index Cond: (id = r.order_item_id)
1	< 0.1%	**Index Scan** on `users` as `u` using `users_pkey` Index Cond: (id = r.user_id)
1	< 0.1%	**Index Scan** on `courses` as `c` using `courses_pkey` Index Cond: (id = oi.course_id)



Database Monitoring (DBM)

활용 사례 2 - Lock race condition 식별

- DBM의 Blocking Activity 기능으로 락 경쟁 상태를 식별



향로 (CTO) VIP 2024년 9월 9일 14:48

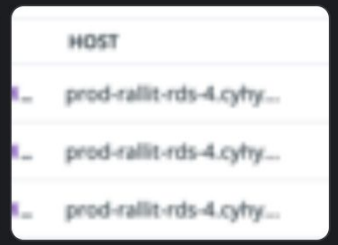
@c-rallit-be 나중에 시간되실때 왜 commit이 한번씩 1초이상걸리는지 한번 찾아봐주시면 좋을것같아요.
1초가 긴 시간은 아닌데 빈도수가 꽤 높은것 같아서요.
DB 스펙도 이젠 T 시리즈가 아닌데도...

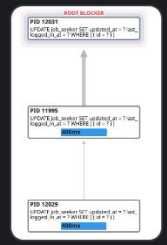
우드 (DEV) 2024년 9월 9일 14:53

추측하기로는 jobSeeker를 업데이트하는 세션 API가 많이 호출되기도 하고,
jobSeeker 업데이트 자체가 워낙 많은 곳에서 이루어지다보니까 요거랑 비슷하게 락이 잡혀서 그런 것 같습니다.
간헐적으로 발생하는 슬로우쿼리들도 이와 관련된 것이 아닐까 싶어요 (편집됨)

> @하루 (DEV) 이번에 발생한 슬로쿼리는 동일한 쿼리가 0.004초 간격으로 3번 발생했다는 특징이 있는데요. 1월 1일에도 비슷한 상황에서 3개의 쿼리가 기록되어 있어 확인을 해보니 락 경쟁 상황이 발생했던것 같습니다. (링크)
> - PID12031 이 Network - ClientRead로 인한 슬로우가 발생함
> - PID11995 가 PID12031이 끝날 때까지 기다림 (Wait event: transactionid)
> - PID12029 가 PID11995가 끝날 때까지 기다림 (Wait event: tuple)
>
> Network로 인한 지연은 간헐적으로 발생하는 wait event 인데, 지연이 발생하는 시점에 락이 걸려있는 쿼리들이 여러 개 있는 상황은 부자연스러워보입니다. 왜 동일한 리소스(튜플)를 거의 동시에 접근하게 됐는지 원인 파악이 필요해보입니다. (편집됨)
>
> 3개 파일

연속적 락이 걸리는 상황

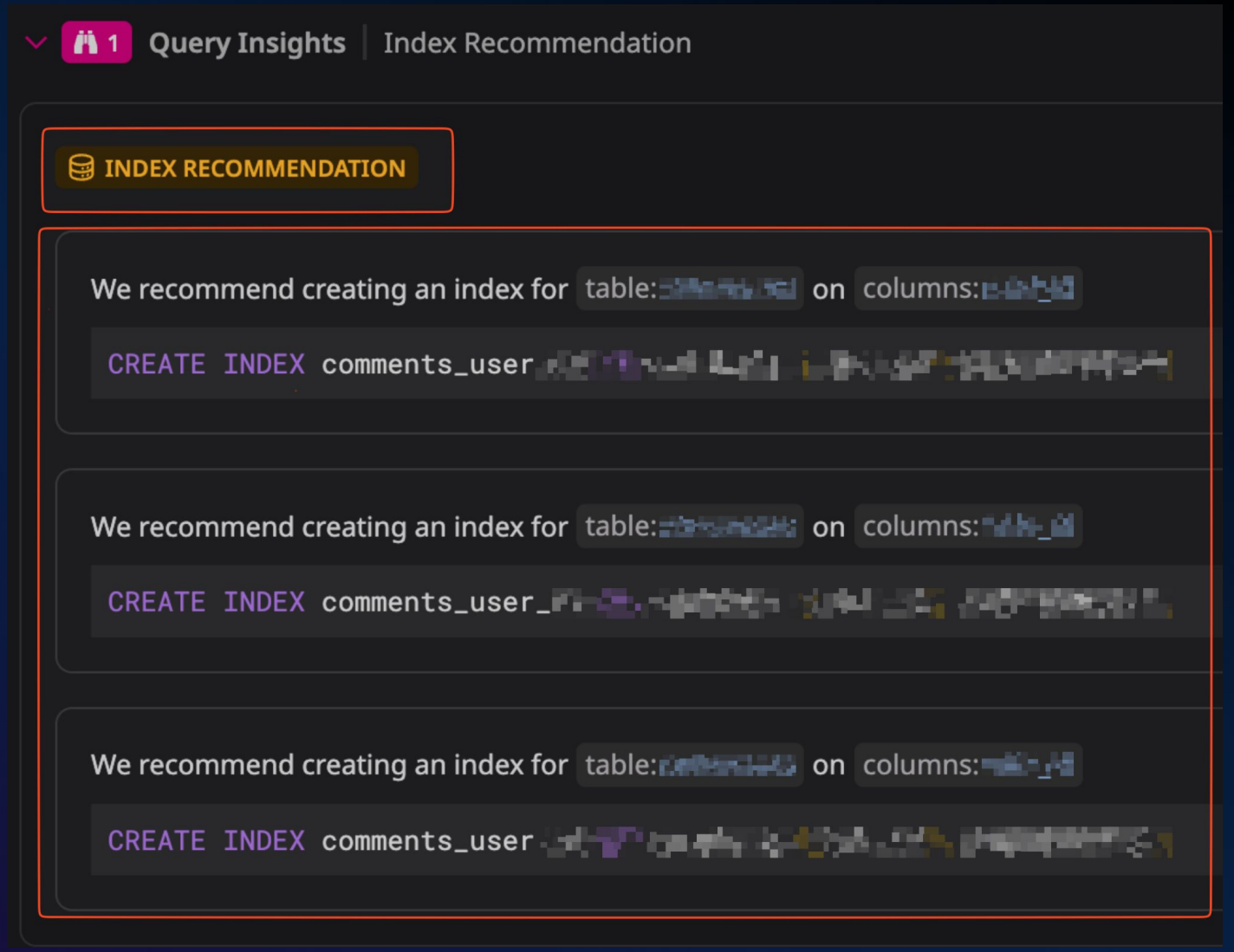




Database Monitoring (DBM)

활용 사례 3 - Index 생성 추천

- Index를 타지 않는 쿼리가 지속적으로 발생하는 경우 해당 쿼리를 분석하여 Index 생성을 추천



Database Monitoring (DBM)

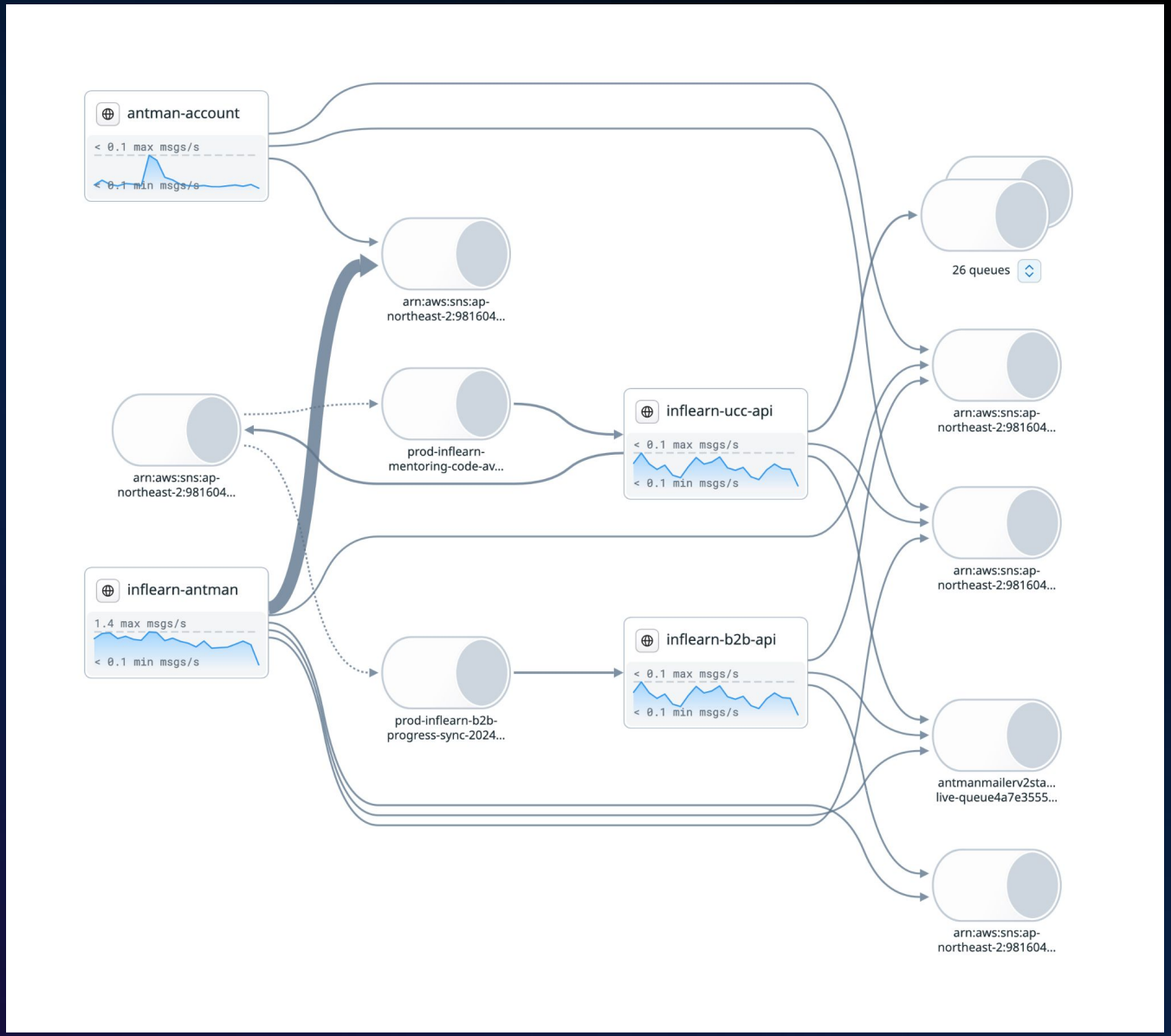
주의할 점

- 실행 중 오류가 발생한 쿼리는 수집되지 않음
  - 오류가 발생한 쿼리는 APM 단에서 확인 필요
- 100% 모든 쿼리가 수집되지는 않음
  - 다만 대표성이 있는 쿼리는 99% 수집된다고 함
- DBM 수집 에이전트는 디비 성능에 영향을 줌
  - CPU: ~1% of the CPU used on average
  - Memory: ~300 MiB of RAM used (RSS memory)
  - Network bandwidth: ~30 KB/s ▼ | 30 KB/s ▲

https://docs.datadoghq.com/database_monitoring/agent_integration_overhead/?tab=postgres

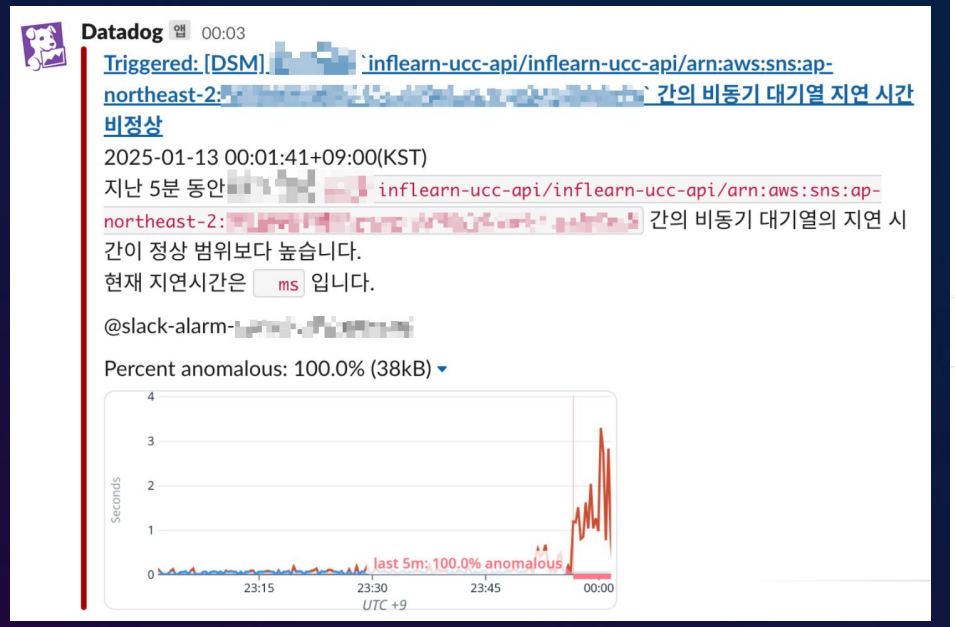
Data Streams Monitoring (DSM)

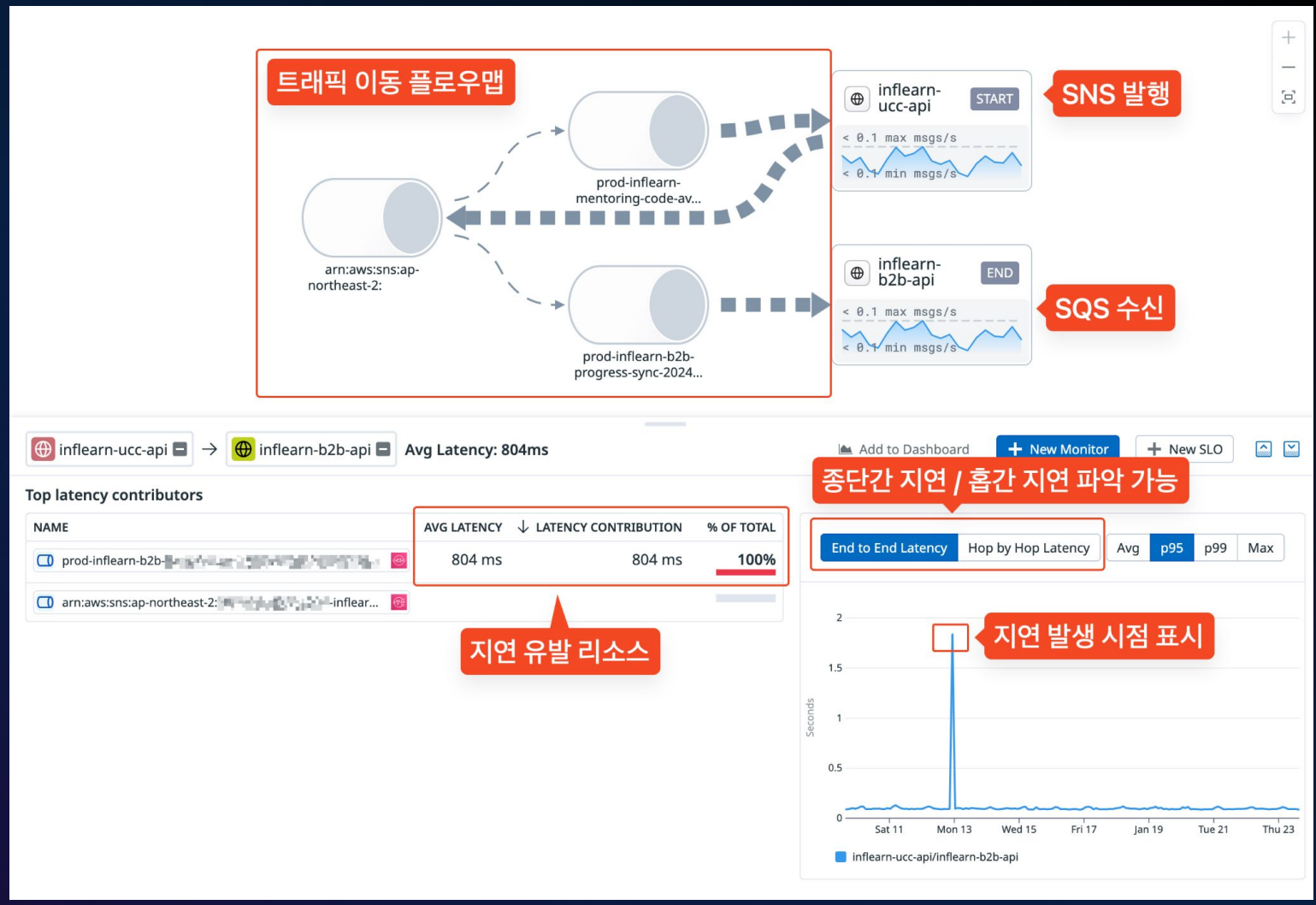
- AWS SQS, SNS, Kafka 등 비동기 대기열을 모니터링
- APM으로 확인할 수 없는 비동기 영역에 대한 추적 제공



Data Streams Monitoring (DSM)

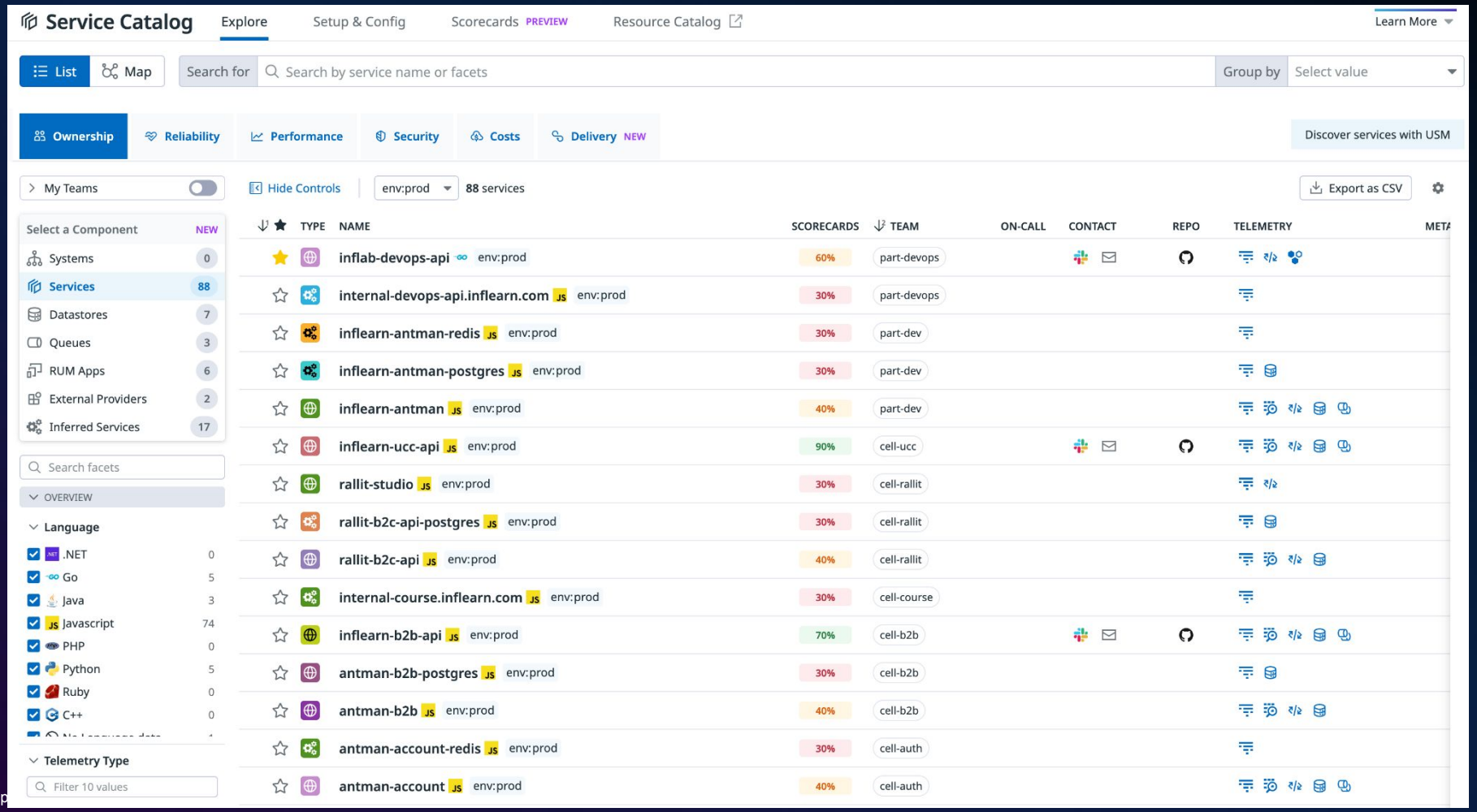
활용 사례 - End to end 지연 식별





Service Catalog

- 기본적으로 아무 설정 없어도 사용 가능
- Datadog에 연동한 모든 서비스가 표시됨

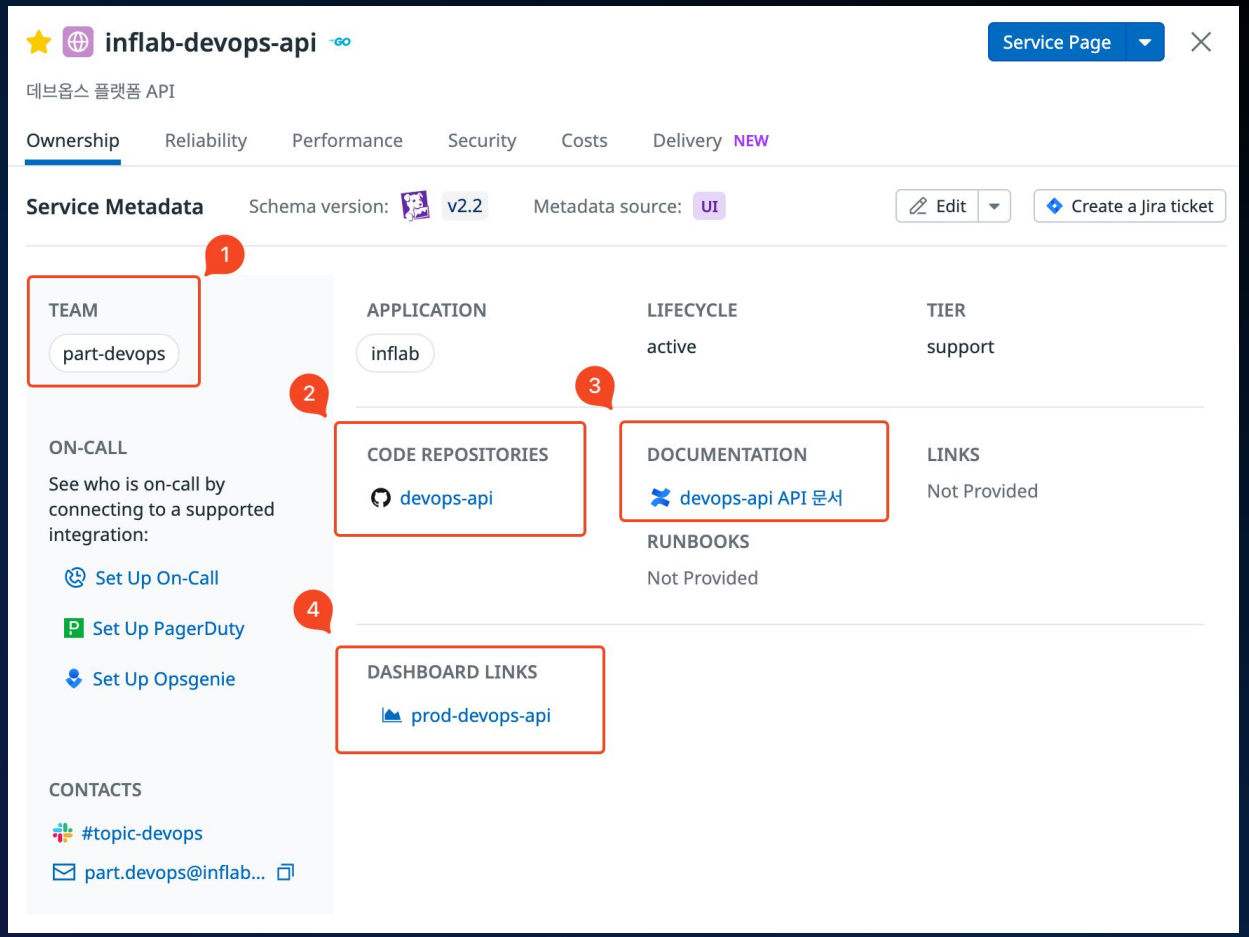


Service Catalog

소유권 (Ownership)

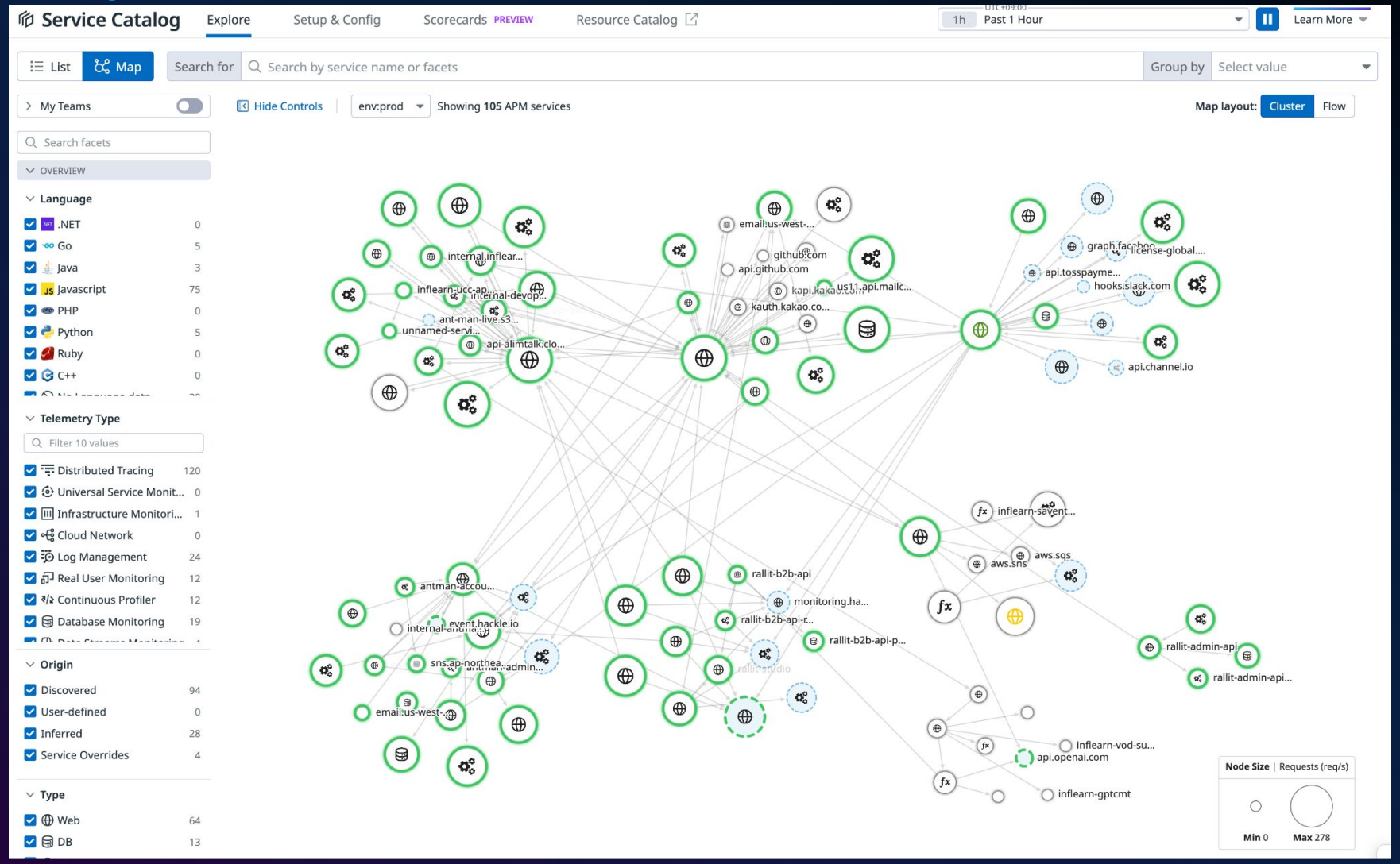
소유권탭에서 서비스의 책임자 정보를 확인 가능

1. 담당팀
2. 코드 저장소
3. 서비스 관련 문서
4. 서비스 대시보드



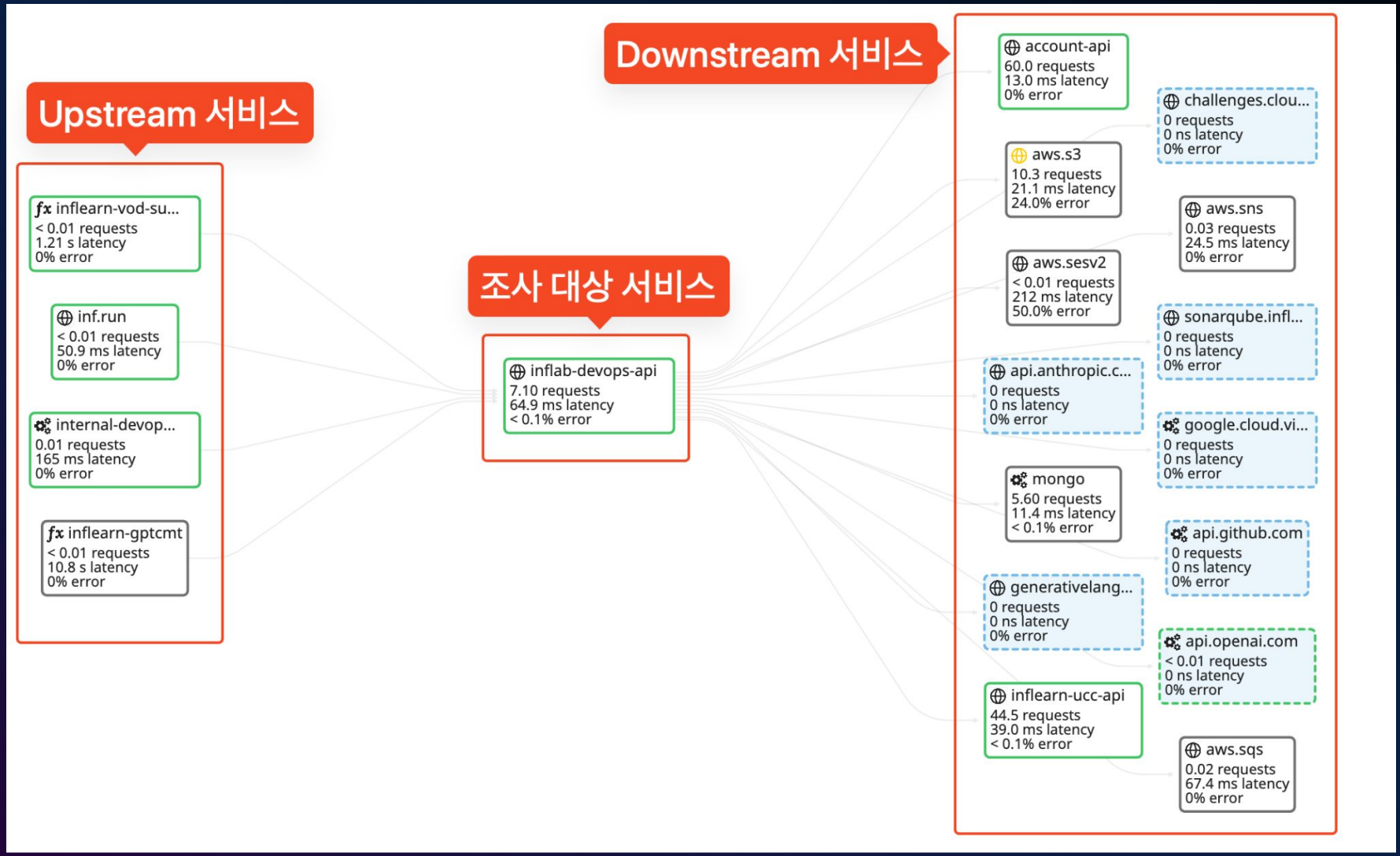
Service Catalog

서비스 맵(Service Map) - Cluster



Service Catalog

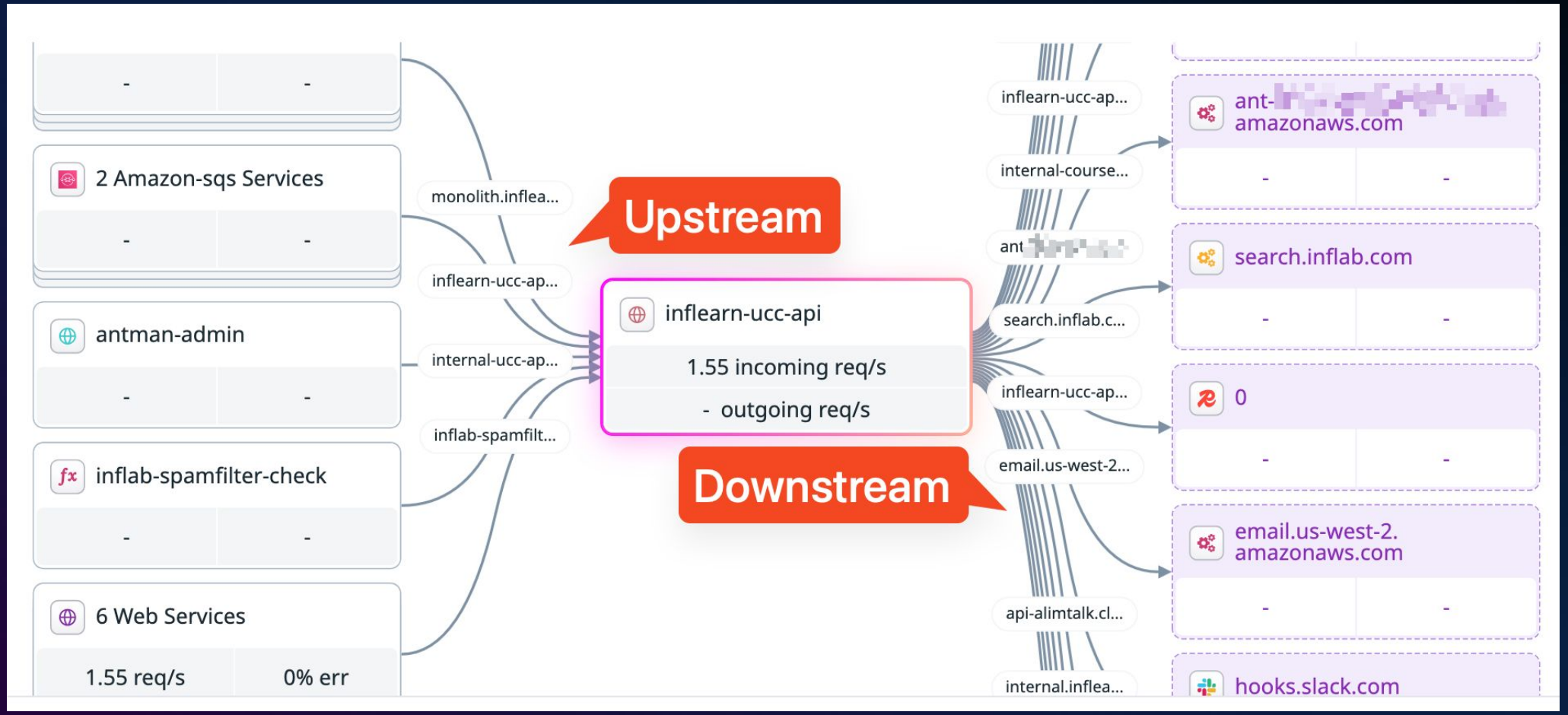
서비스 맵(Service Map) - Flow



Service Catalog

활용 사례 - 서비스 종속성 맵으로 운영 서비스에 대한 큰 그림 그리기

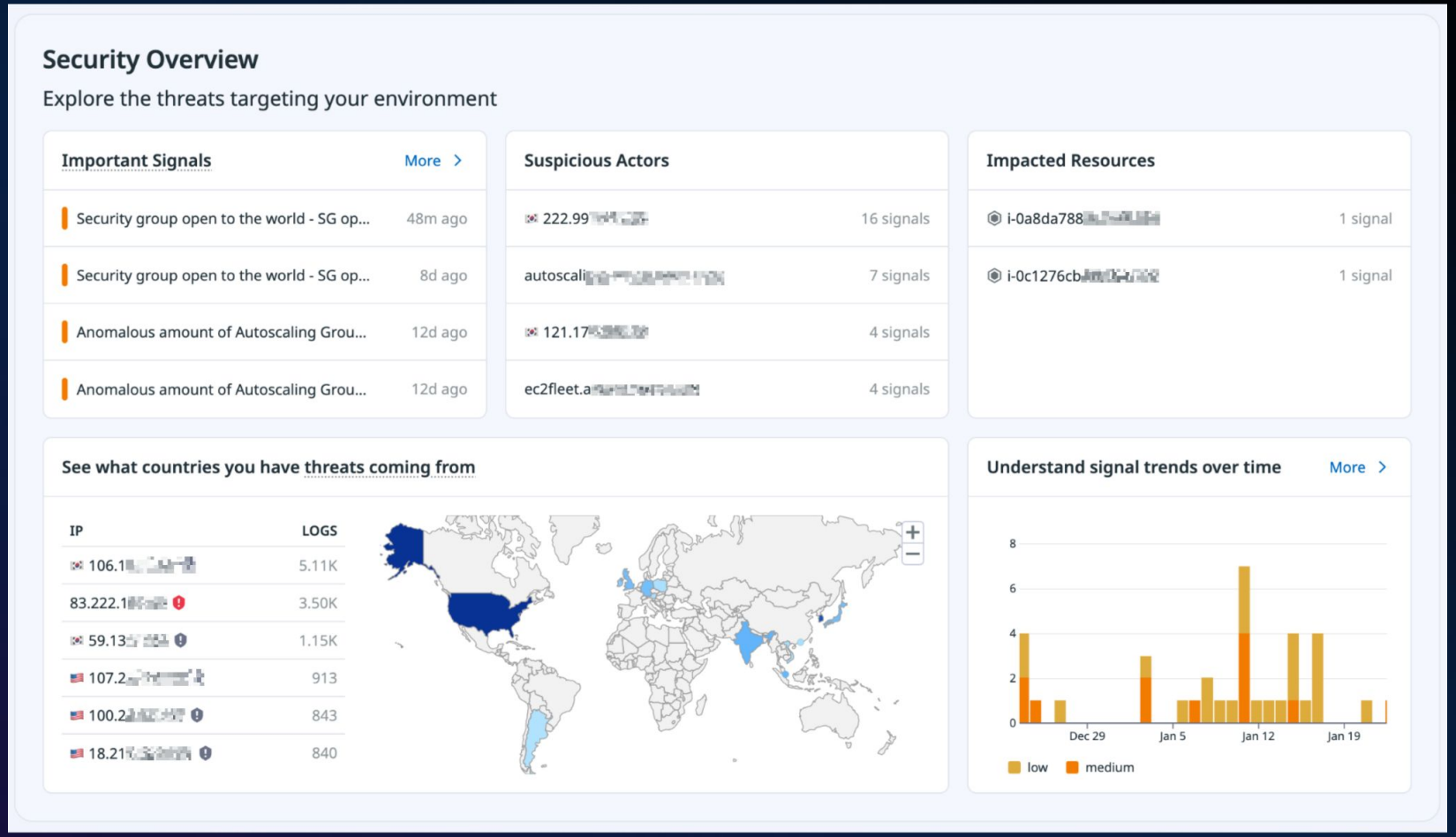
- 서비스의 상하 의존구조를 한 눈에 파악
- 장애 추적시 서비스 종속성 맵을 활용해 신속하게 원인 pin-down



Cloud SIEM

감사로그 중앙집중식 관리 및 위협 탐지

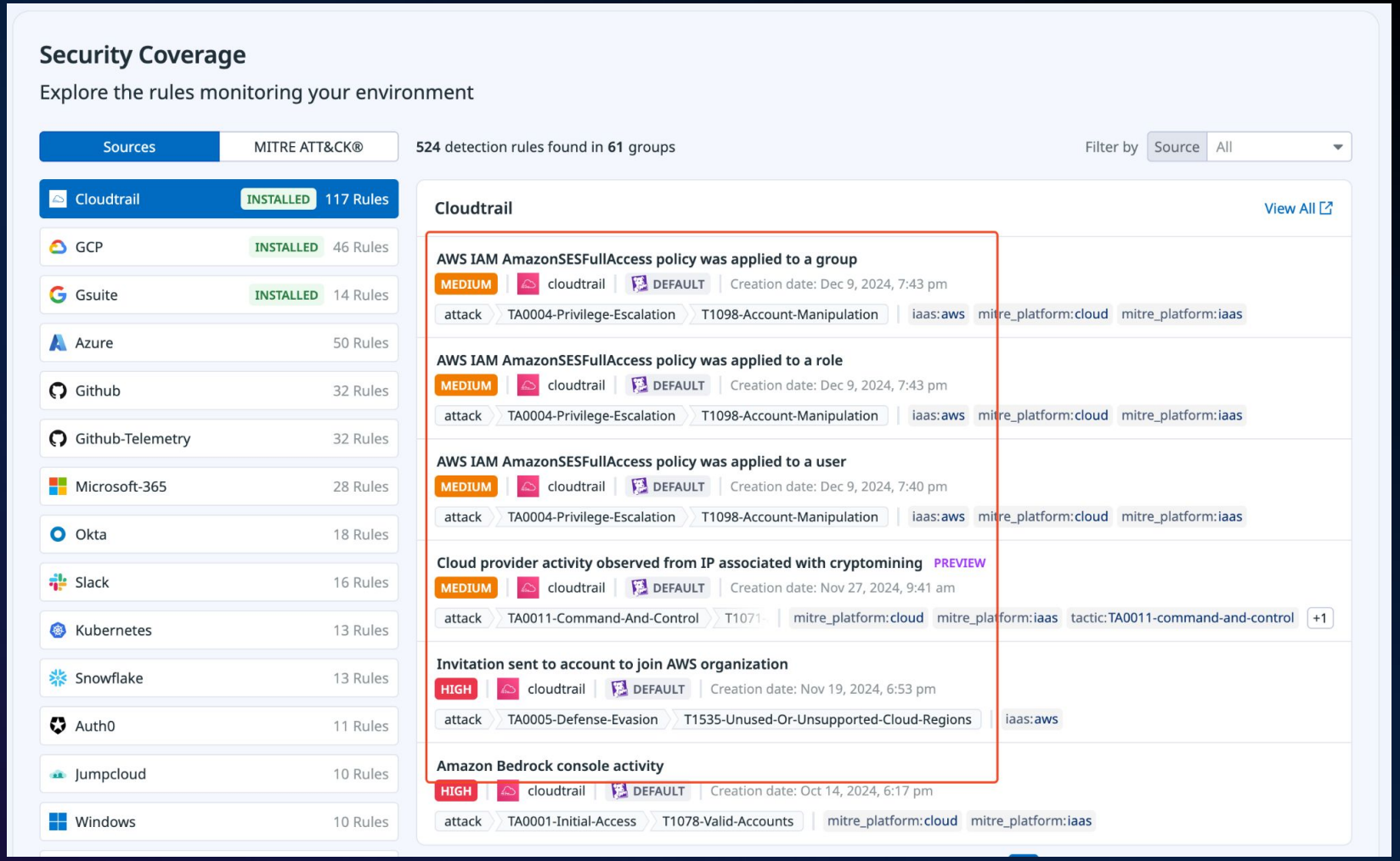
- 감사로그(Audit log)를 실시간으로 분석하여 보안 취약점 및 악의적 공격을 탐지하는 도구



Cloud SIEM

감사로그 중앙집중식 관리 및 위협 탐지

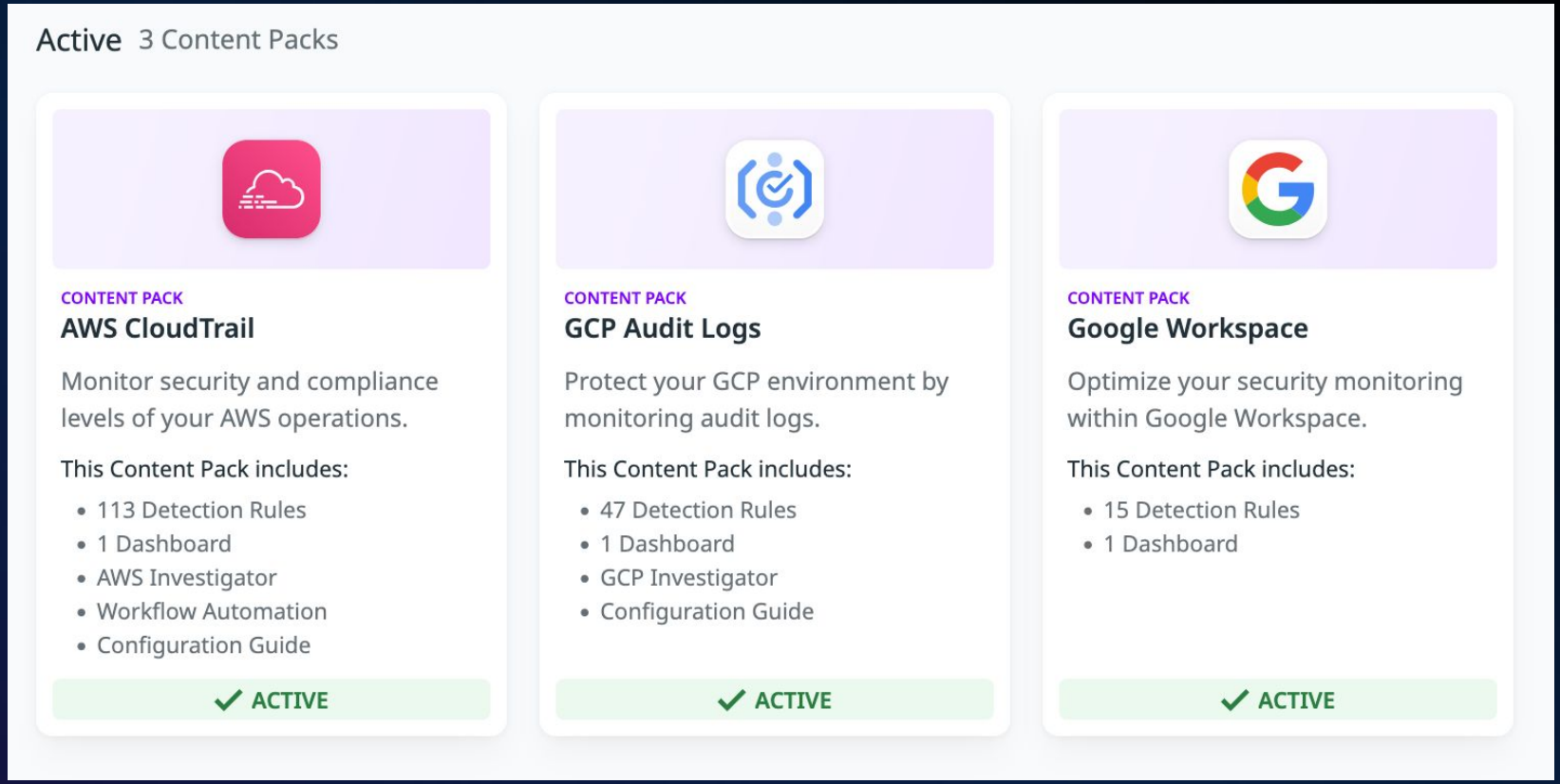
- 감사로그를 지원하는 다양한 소스를 지원
- 소스별 특화된 위협 탐지 경보 자동 설정



Cloud SIEM

SaaS 감사로그 연동

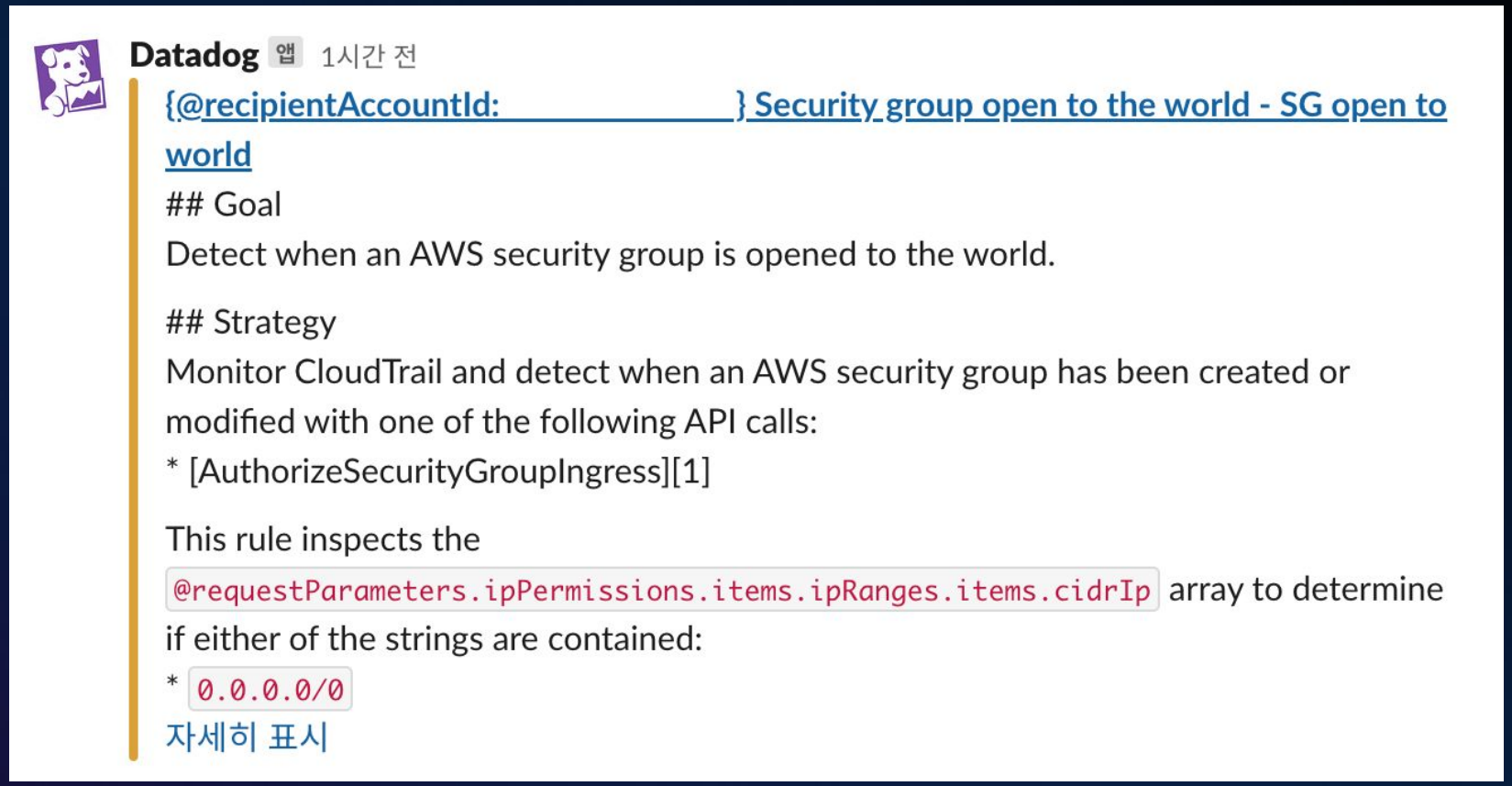
- AWS CloudTrail
- GCP
- GWS



Cloud SIEM

활용 사례 1 - 0.0.0.0/32로 오픈된 AWS EC2 보안그룹 탐지

- 실수 또는 임시조치로 인해 발생한 보안 위협에 대해 신속하게 알림



Cloud SIEM

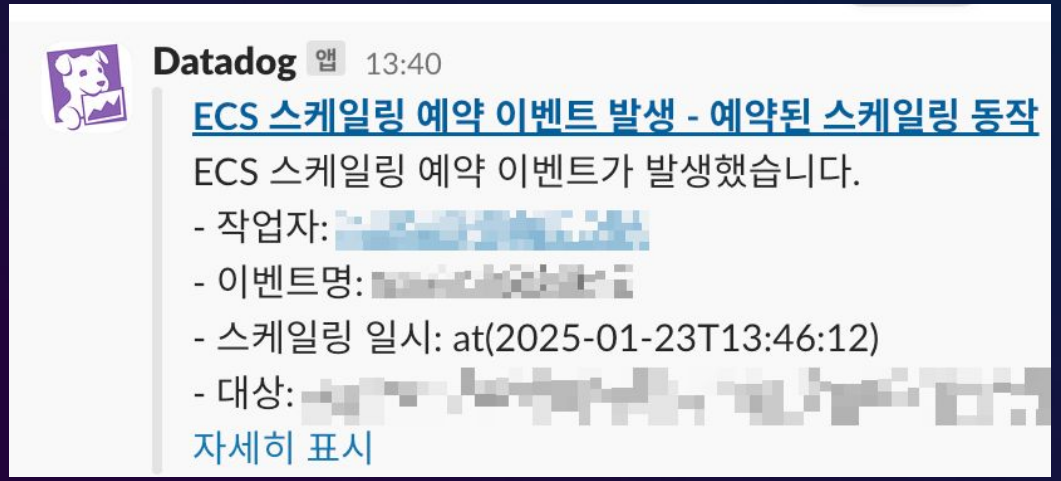
활용 사례 2 - 디비 튜플 암호화에 사용된 KMS 키에 대한 비정상적인 접근 탐지

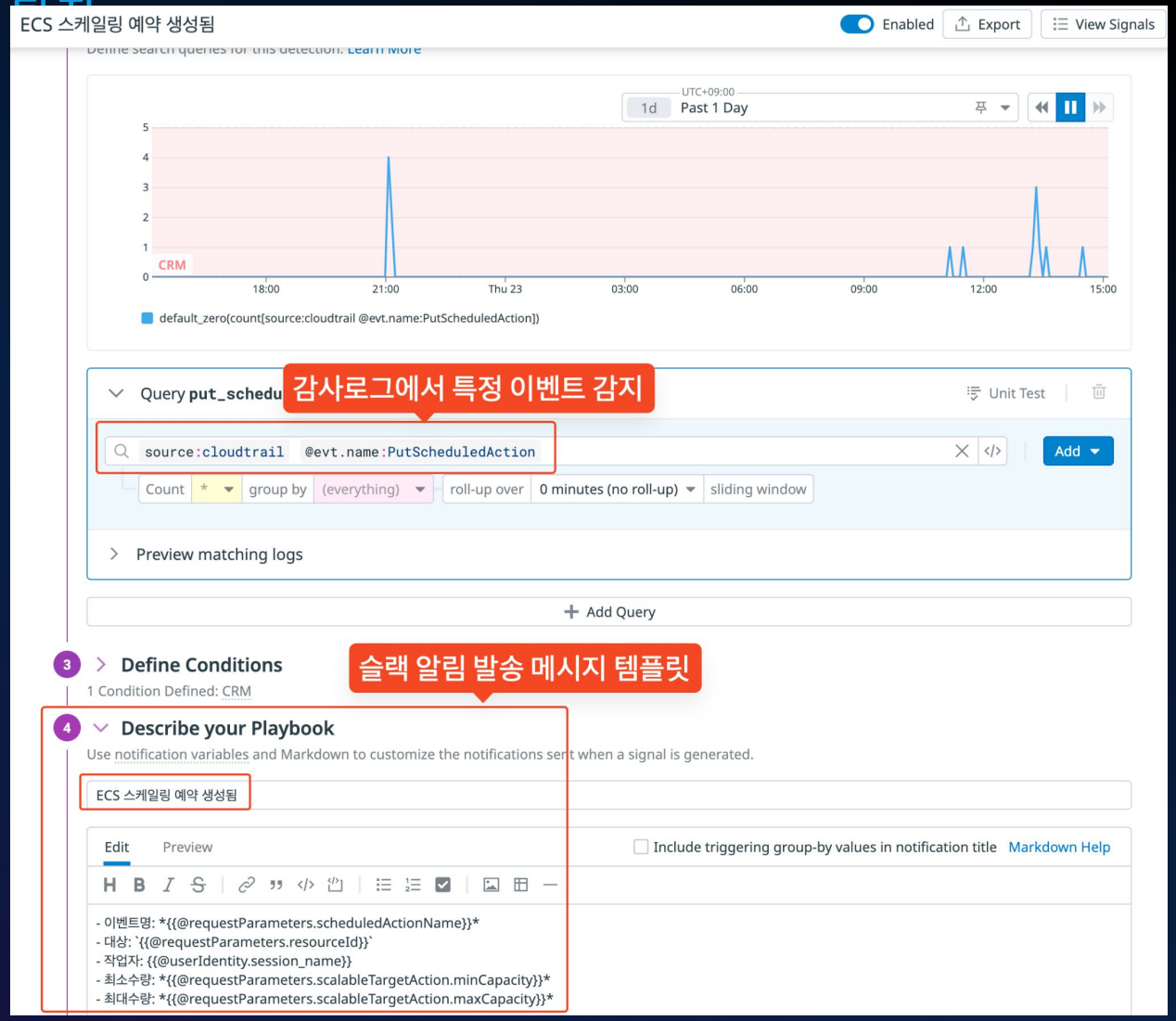
- 승인되지 않은 AWS 리소스 접근에 대해 즉시 알림
- 빠른 문제 식별 및 대처 가능

Cloud SIEM

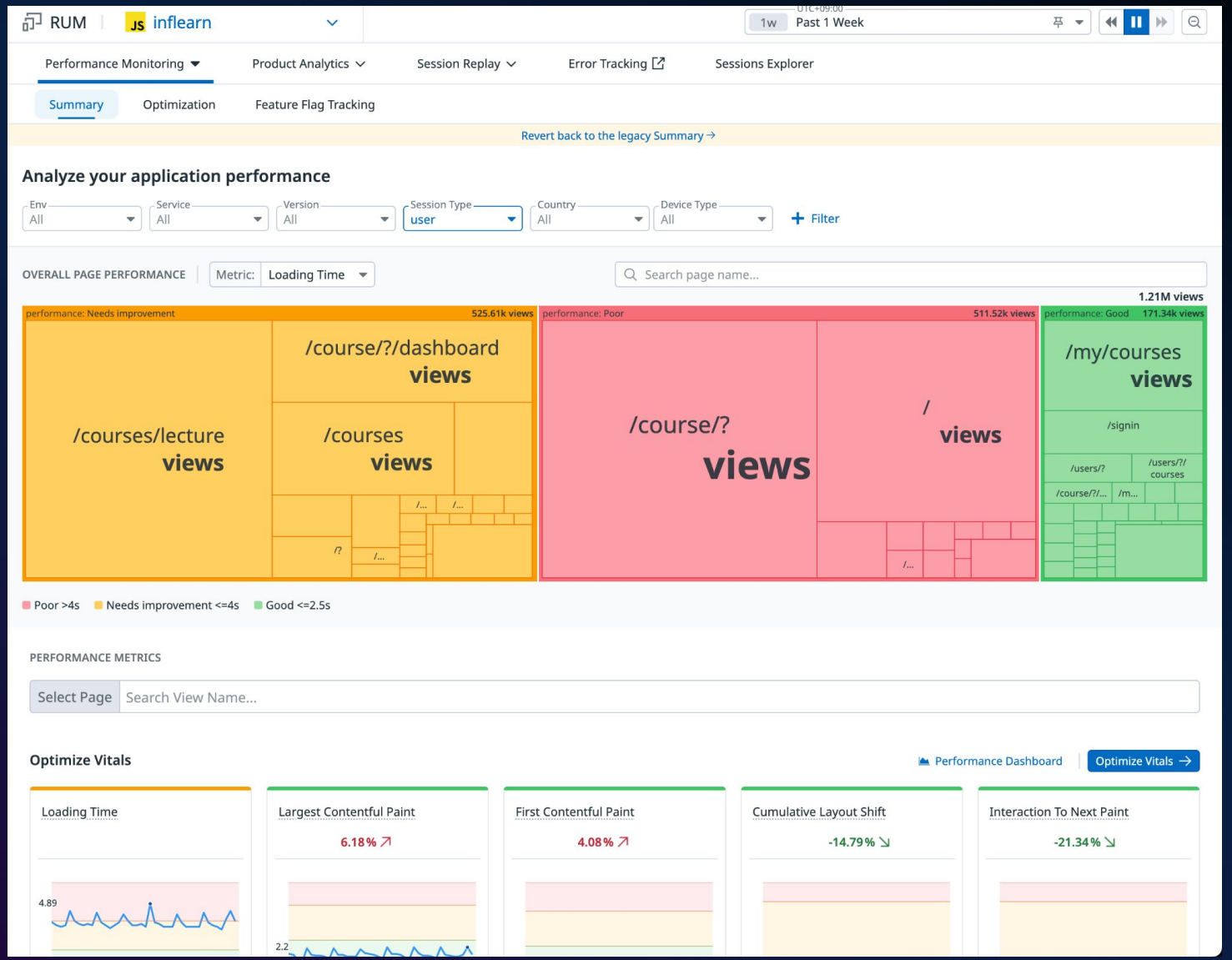
활용 사례 3 - CRM 대비를 위한 스케일링 예약 이벤트 탐지

- 스케일링 예약 이벤트 감지시 실시간 알림 발송
- 기존의 수기 공유를 대체





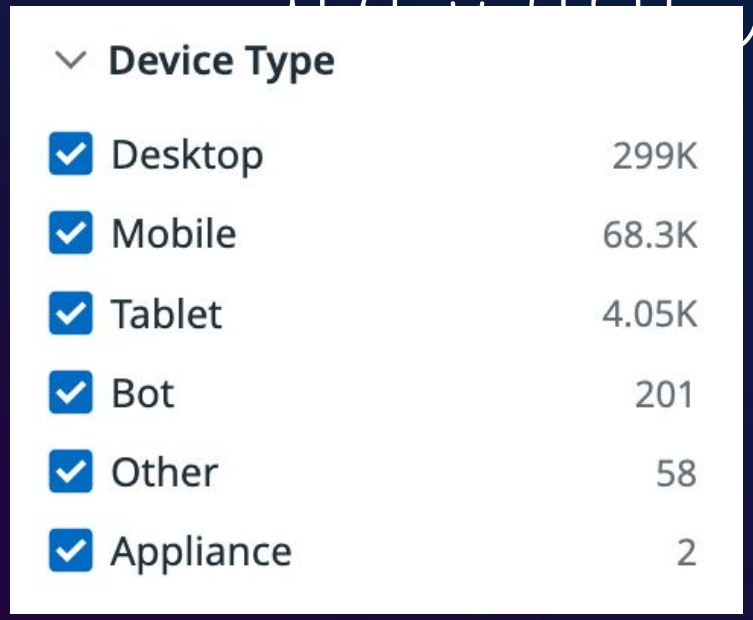
RUM (Real User Monitoring)

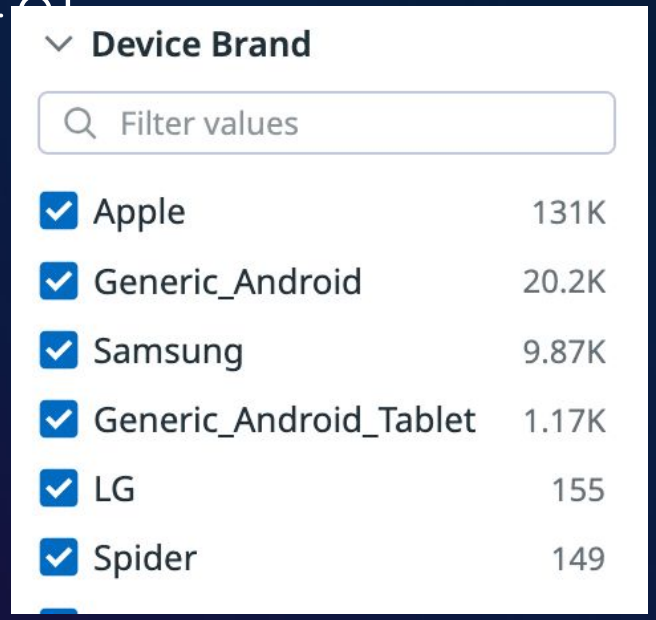


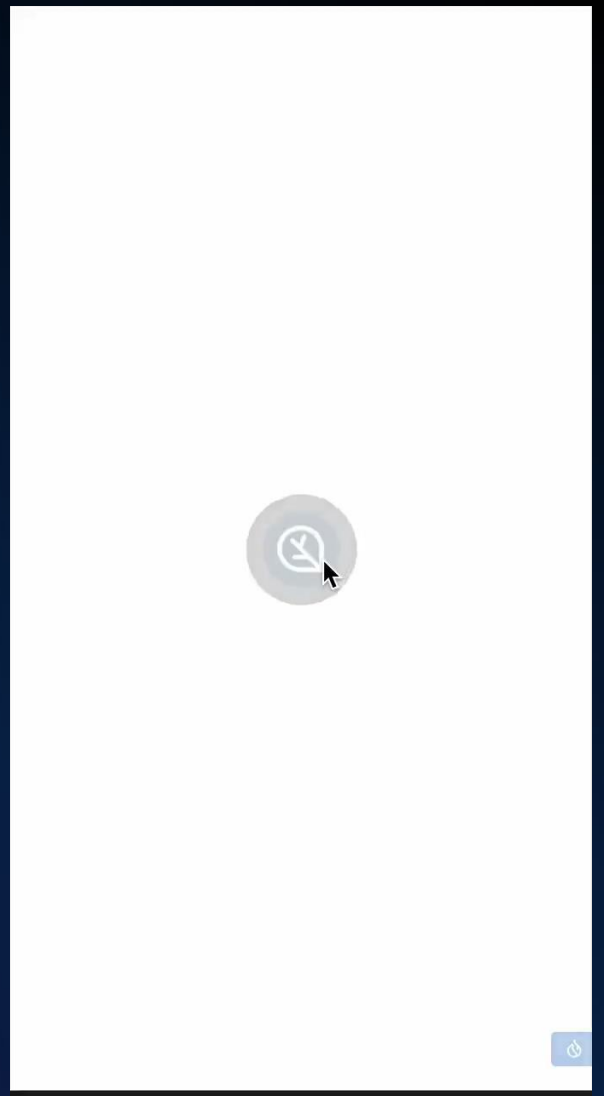
RUM (Real User Monitoring)

활용 사례 - 유저 액션 추적

- UI/UX 에서 발생하는 버그 재현에 활용
- 사용자단의 단말마다 UI가 실제로 정상 표시되는지 확인







Datadog을 잘 써보기

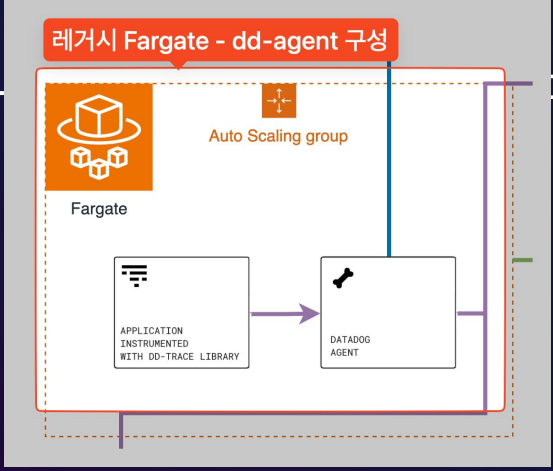
Datadog을 잘 써보기

- 적은 비용으로 최대한의 효율 뽑아내기
  - Datadog agent host 수량 최적화
  - Log 비용 최적화
- 자격증명 자동화
  - Google Workspace SAML 연동

Datadog agent host 수량 최적화

- Datadog Agent Host 월 비용: $31~45
- Datadog Fargate* 컨테이너 월 비용:

  $2.9~3.7
- Agent H PM 10대



*)Fargate: AWS의 Full Managed 컨테이너 서비스

APM Host 월 비용

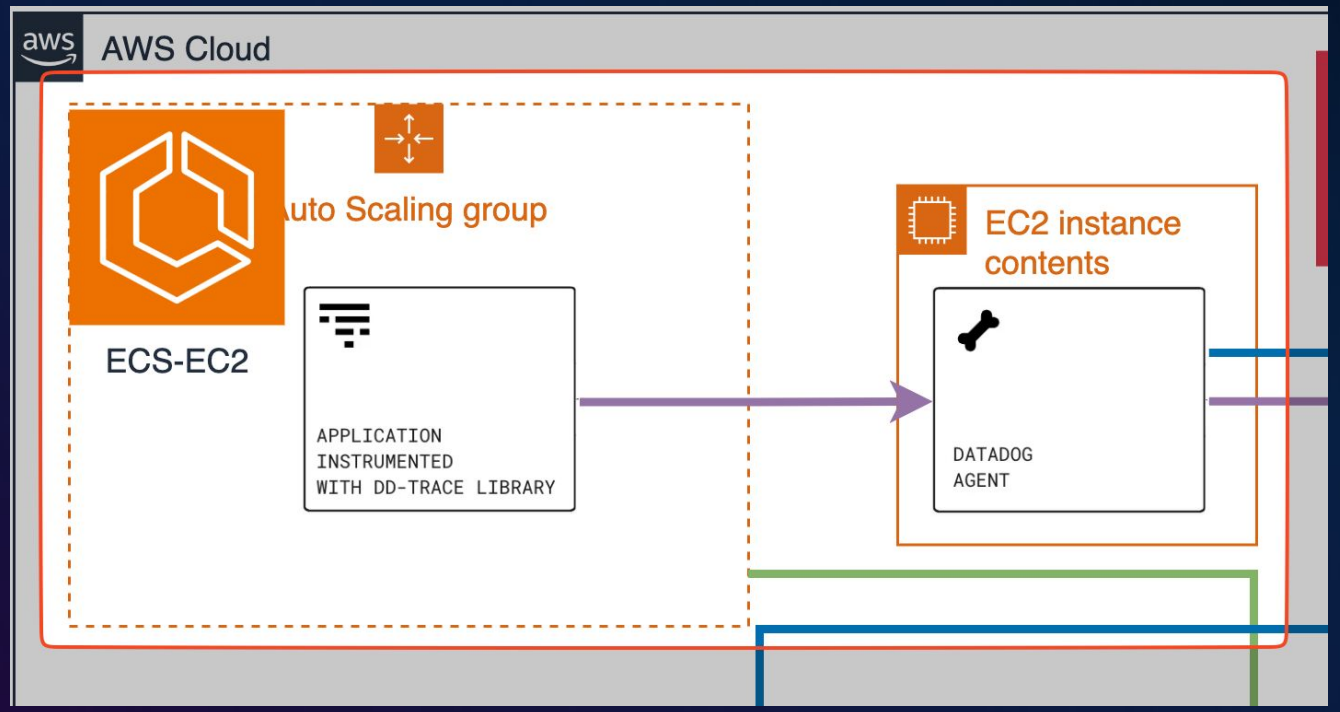
APM	APM Pro	APM Enterprise
STARTING AT $ 31 Per host, per month*	STARTING AT $ 35 Per host, per month*	STARTING AT $ 40 Per host, per month*
APM DevSecOps	**APM DevSecOps Pro**	**APM DevSecOps Enterprise**
STARTING AT $ 36 Per host, per month*	STARTING AT $ 40 Per host, per month*	STARTING AT $ 45 Per host, per month*

APM Fargate 월 비용

AWS Fargate	$2 per task ($2.90 on-demand)	$2 per task ($2.90 on-demand) 10GB span ingestion and 65,000 span retention, averaged across all your APM hosts	$2.60 per task ($3.70 on-demand) 10GB span ingestion and 65,000 span retention, averaged across all your APM hosts, 1 profiled host

Datadog agent host 수량 최적화

- Agent Host와 컨테이너 Host 분리
- Agent Host를 6개만 운영



Datadog agent host 수량 최적화

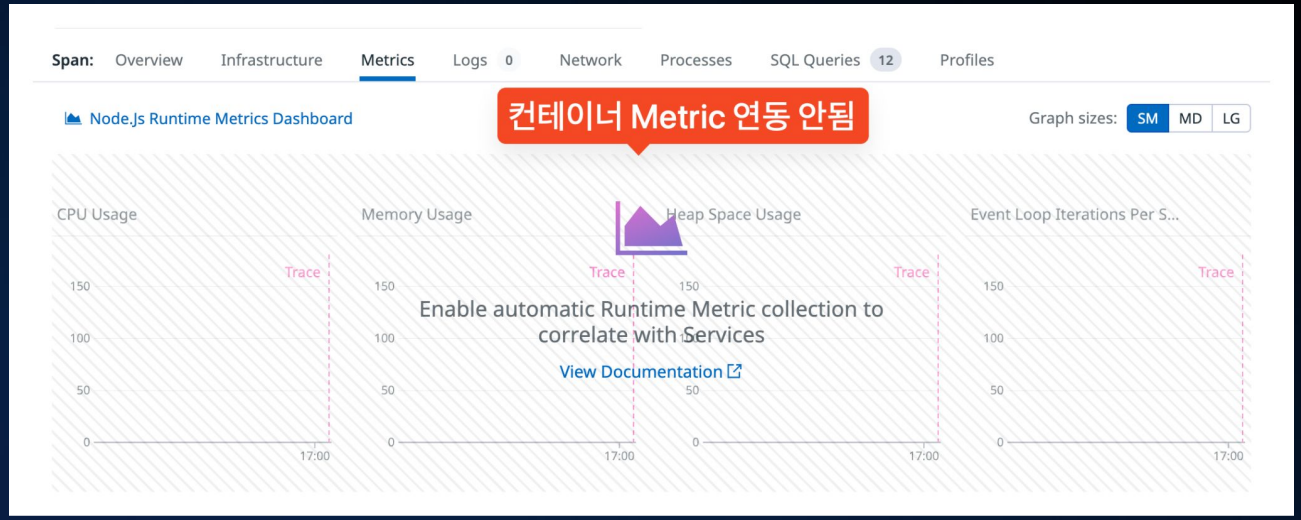
연결 방법

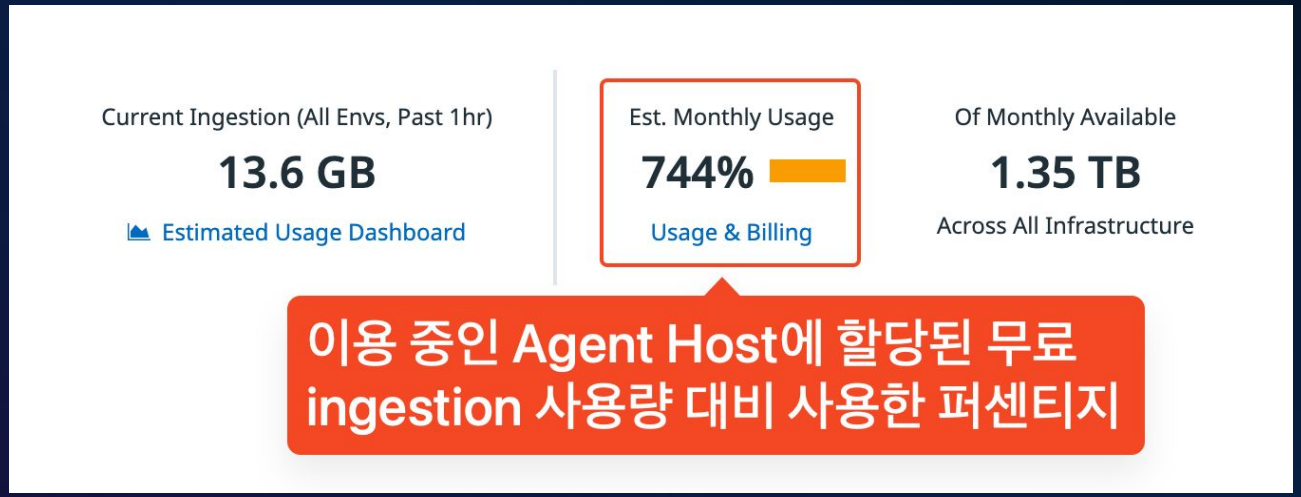
- 각 서비스 컨테이너에 **DD_AGENT_HOST**와 **DD_TRACE_REPORT_HOSTNAME** 환경 변수 정의

Datadog agent host 수량 최적화

주의 사항

- Metric 연동 불가
  - AWS 통합의 메트릭 활용 필요
- 별도의 trace ingestion 비용 발생 가능
  - 비용은 Agent Host보다 저렴

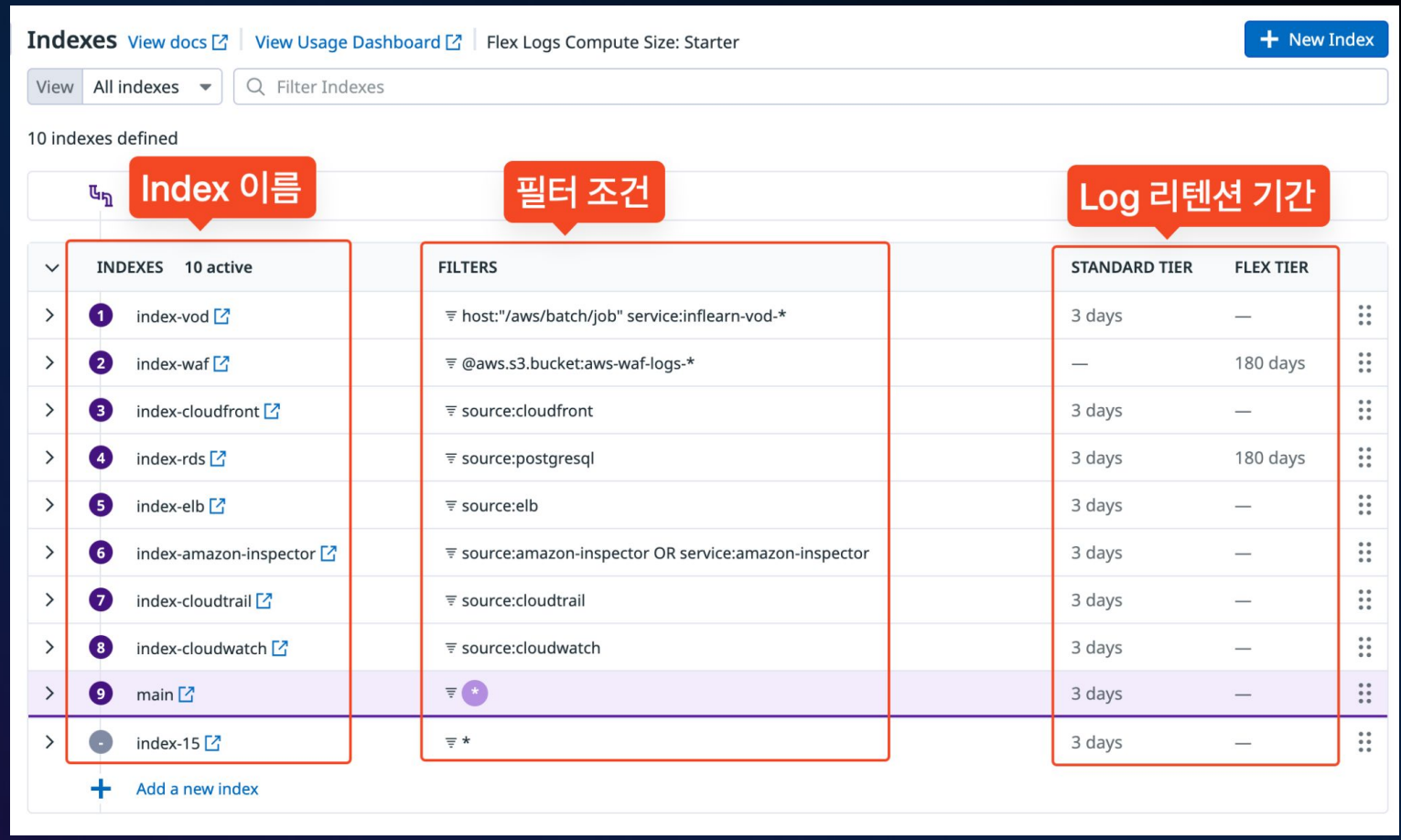




Log 비용 최적화

Index를 사용하여 유형별 Log 리텐션 기간 지정

- Retention: Log 보관 기간
  - 3일 ~ 15개월 다양
- Log 소스의 중요도에 따라 보관 기간 설정 필요



Log 비용 최적화

Index를 사용하여 유형별 Log 리텐션 기간 지정

- Index 분리시 index별 사용량 확인 가능
- 각 Log에 필요한 retention 기간 및 비용을 확인하여 개별 지정 추천

Index별 로그 사용량 확인 가능

NAME	RETENTION (D)	LIVE INDEXED LOG COUNT	REHYDRATED INDEXED LOG CO...	↓ TOTAL INDEXED LOG COUNT	CONTRIBUTION PERCENTAGE
index-elb	3	524,182,313	0	524,182,313	36.4%
index-cloudfront	3	512,580,380	0	512,580,380	35.6%
index-rds	3	247,175,511	0	247,175,511	17.2%
index-cloudtrail	3	56,996,460	0	56,996,460	3.96%
main	3	52,458,517	0	52,458,517	3.65%
index-cloudwatch	3	44,971,968	0	44,971,968	3.13%
index-amazon-inspector	3	181,452	0	181,452	< 0.1%
index-vod	3	161,086	0	161,086	< 0.1%

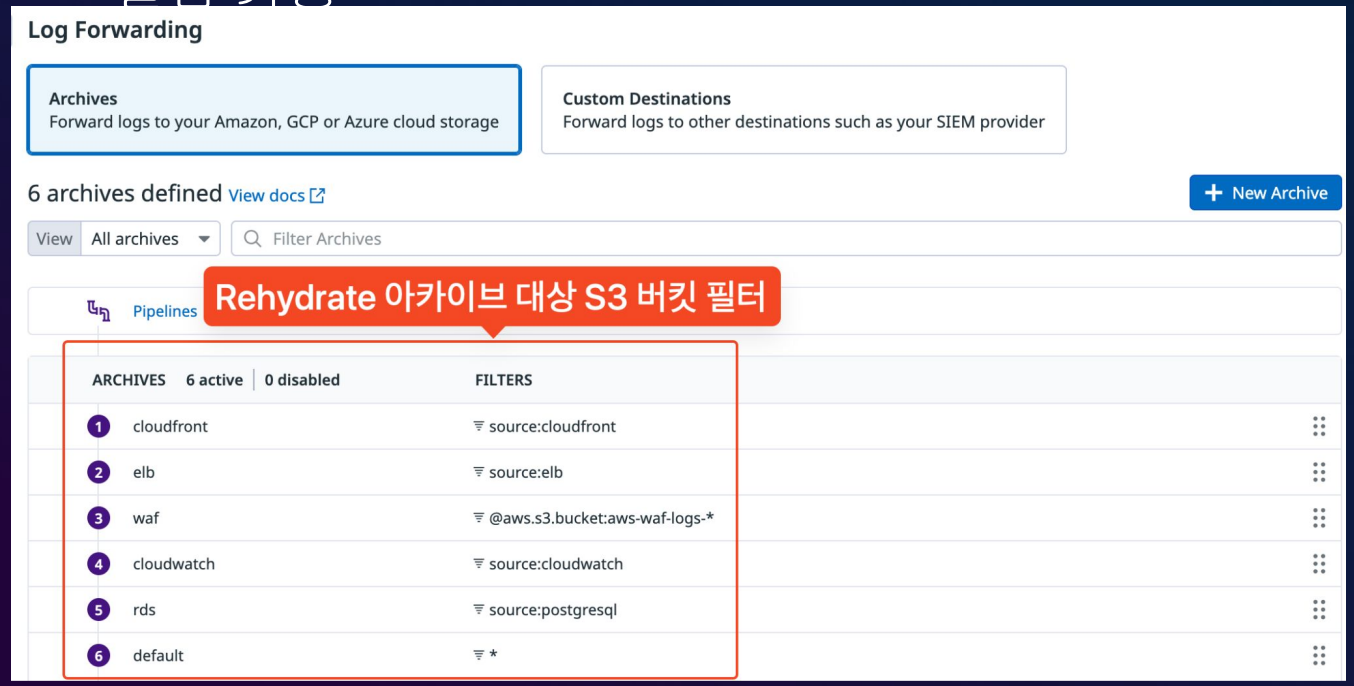
Retention 기간별 Log 요금

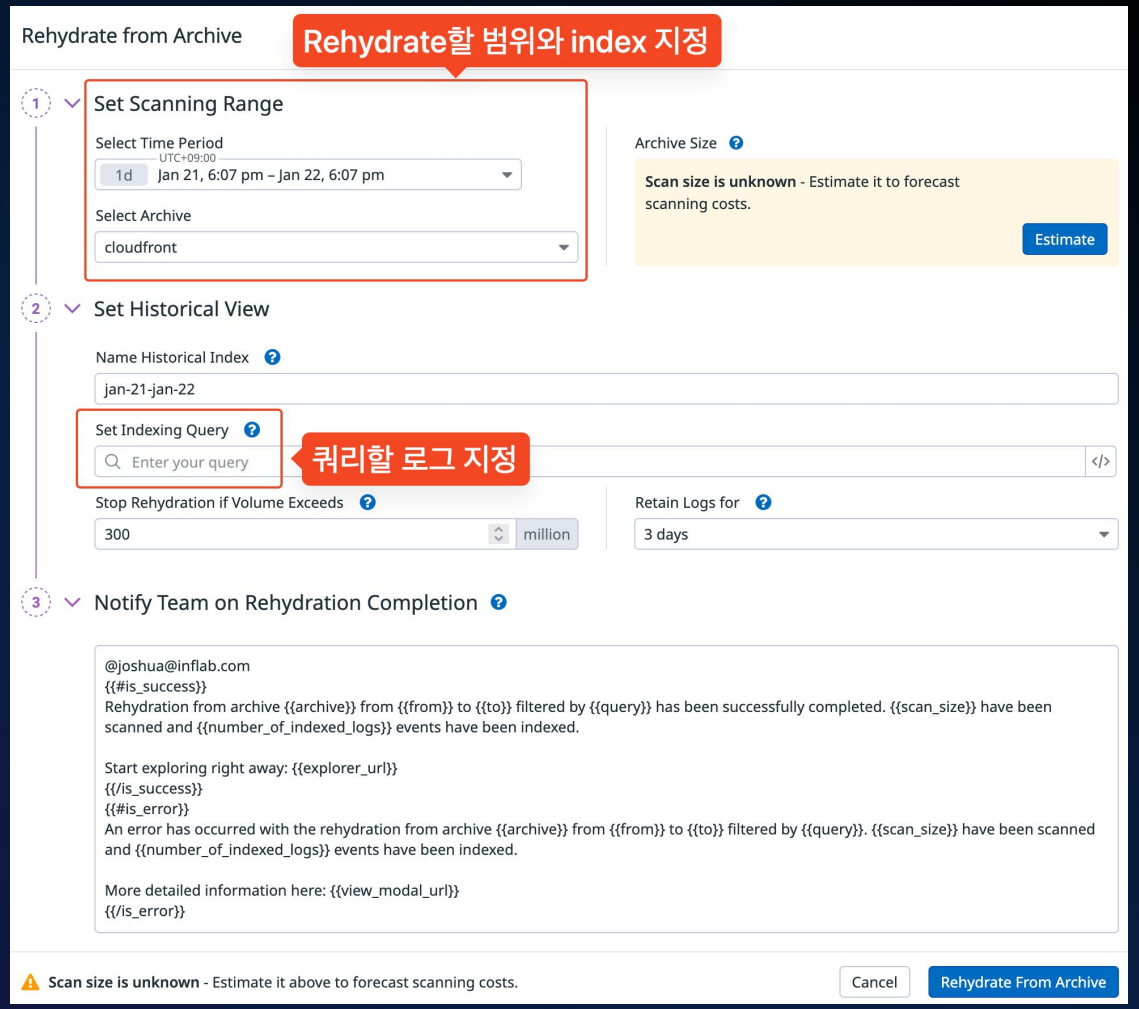
Logs - Indexed Log Events	Per 1M indexed logs (3-day retention), per month	$1.06
Logs - Indexed Log Events	Per 1M indexed logs (7-day retention), per month	$1.27
Logs - Indexed Log Events	Per 1M indexed logs (15-day retention), per month	$1.70
Logs - Indexed Log Events	Per 1M indexed logs (30-day retention), per month	$2.50

Log 비용 최적화

Rehydration 활용

- Rehydrate 기능으로 retention 기간이 지난 Log를 다시 Datadog log로 복구 가능
- 복구 범위와 쿼리 조건을 세밀하게 지정하면 비용 절감 가능

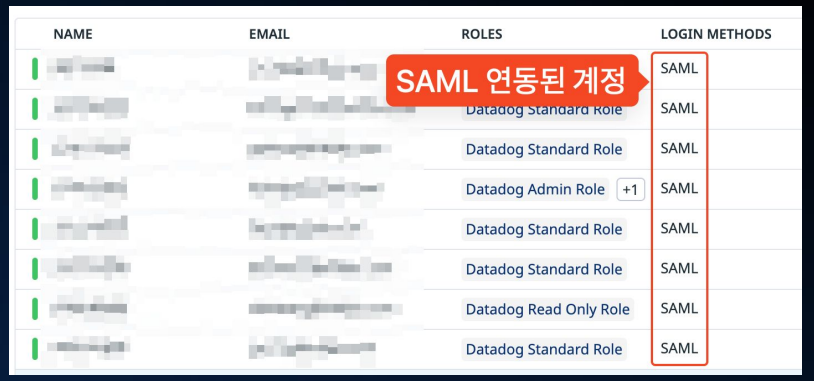


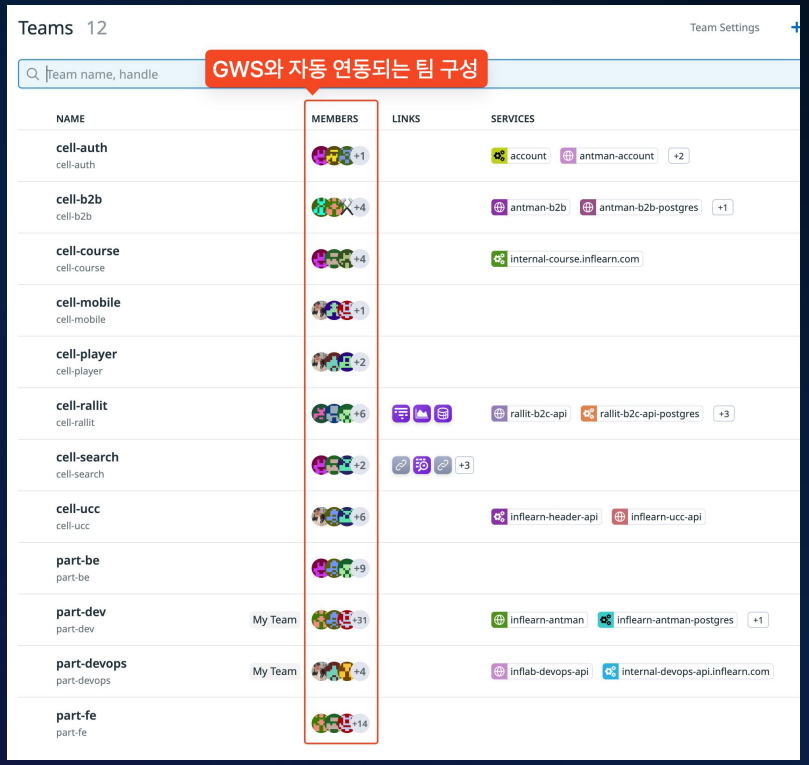


Google Workspace SAML 연동

IdP를 통한 사용자 연동

- 사용자 자동 Provision으로 효율적인 자격증명 관리
- 회사가 사용하는 IdP (자격증명제공자)로부터 사용자 연동
- 더 이상 이메일 초대는 불필요

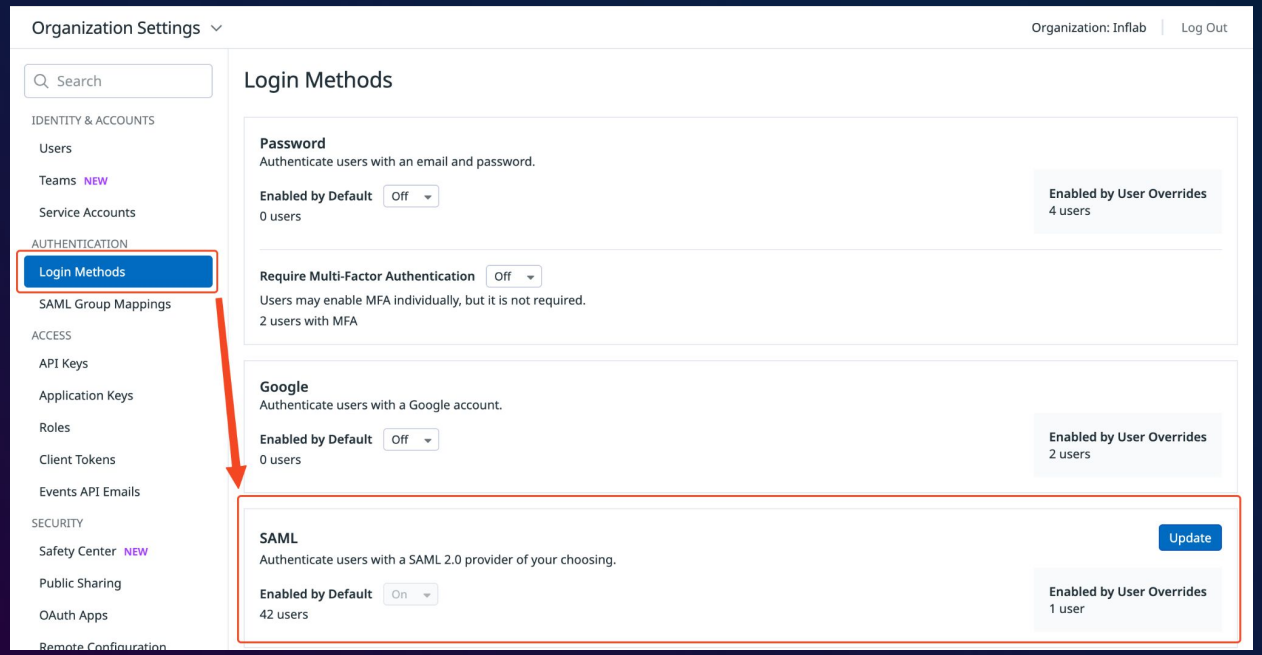


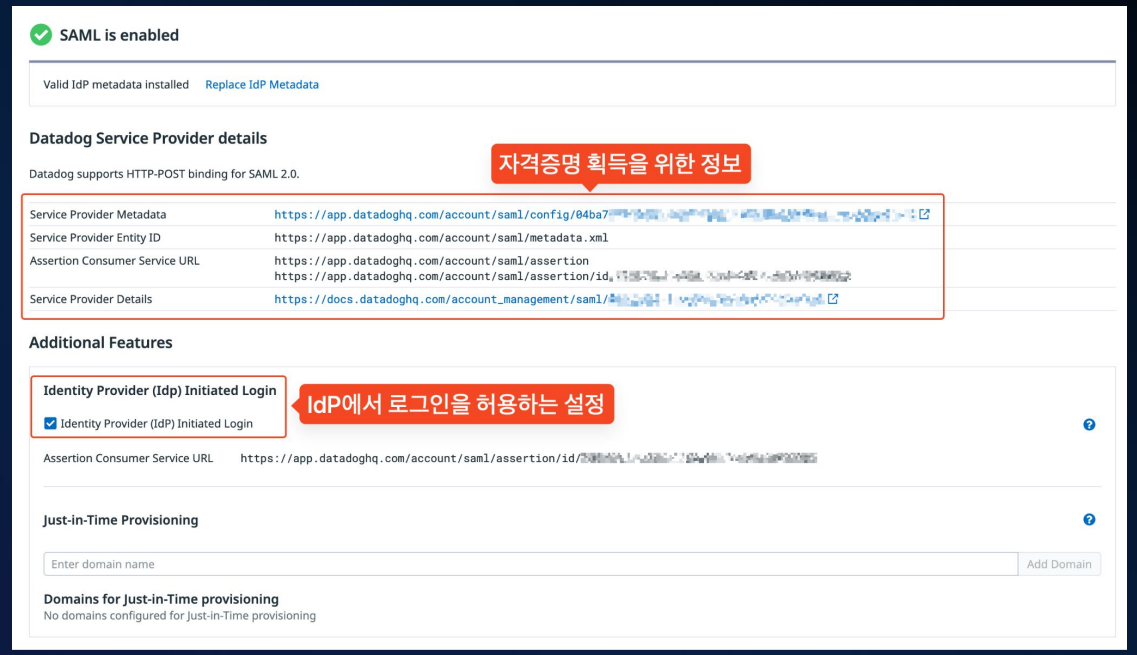


Google Workspace SAML 연동

GWS SAML 연동 방법

- 참고문서: https://support.google.com/a/answer/7553768





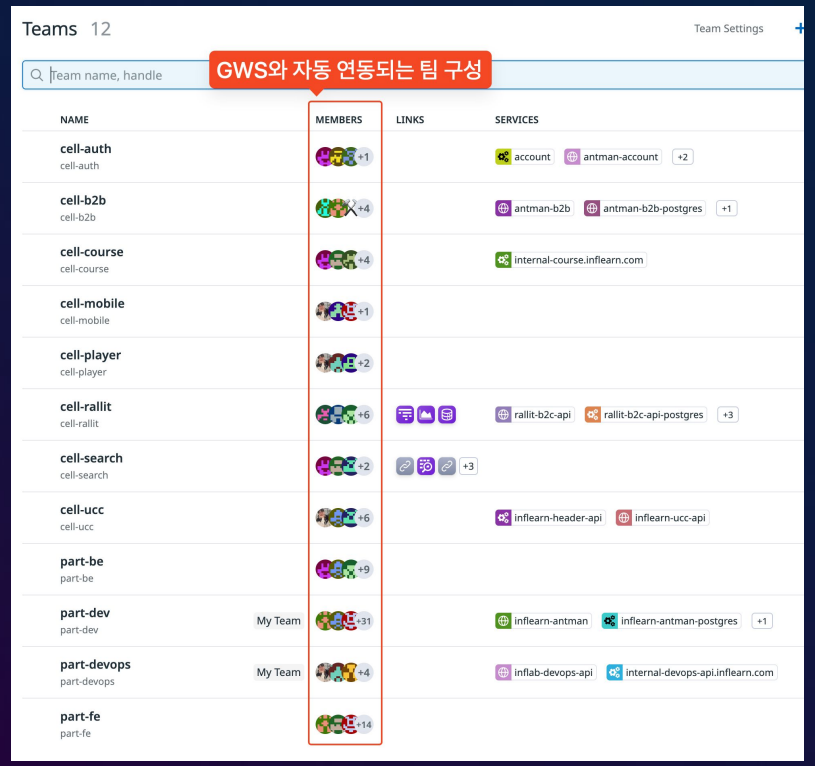
Google Workspace SAML 연동

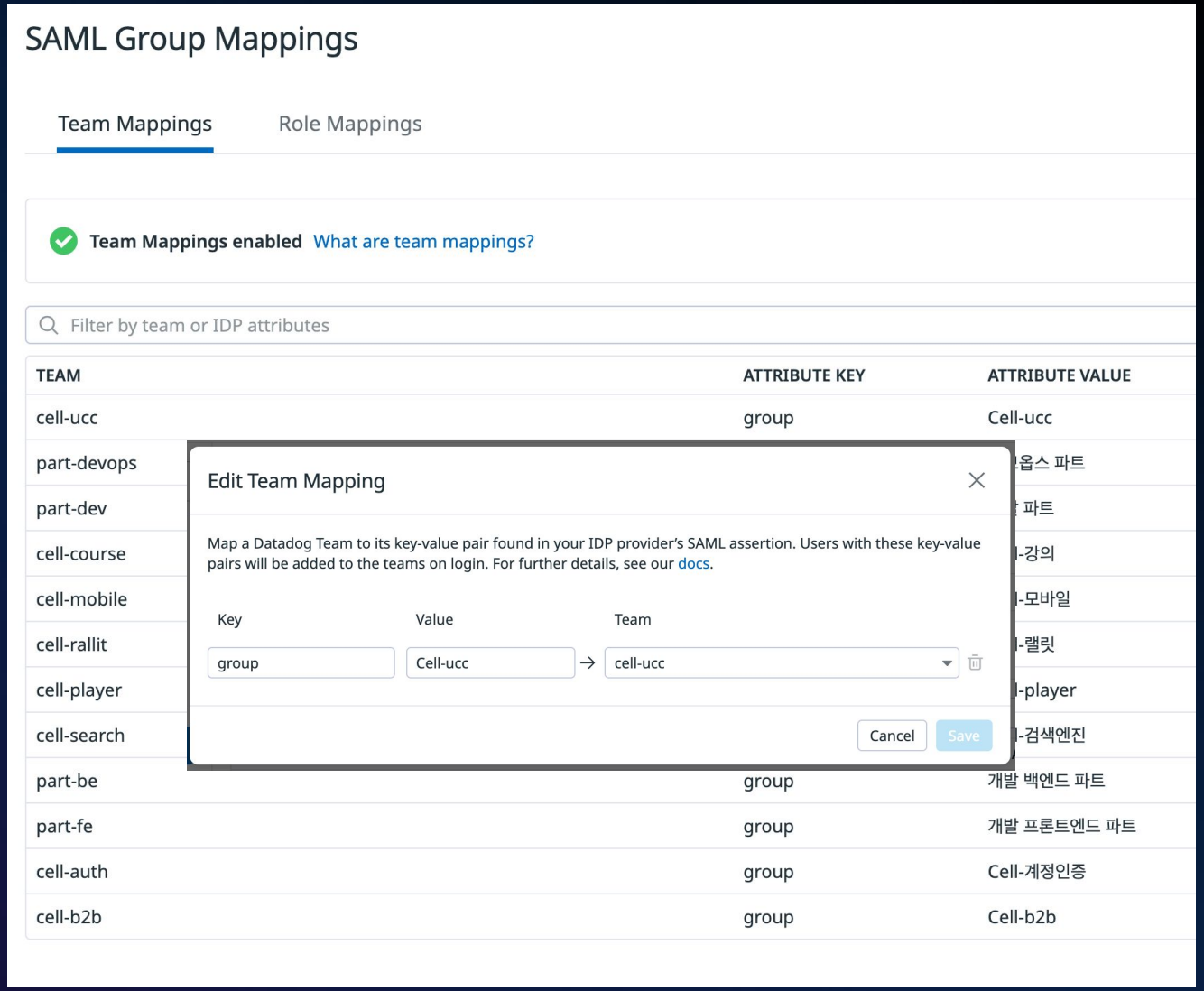
GWS 쪽 구성 방법

- 참고문서: https://support.google.com/a/answer/7553768

Google Workspace SAML 연동

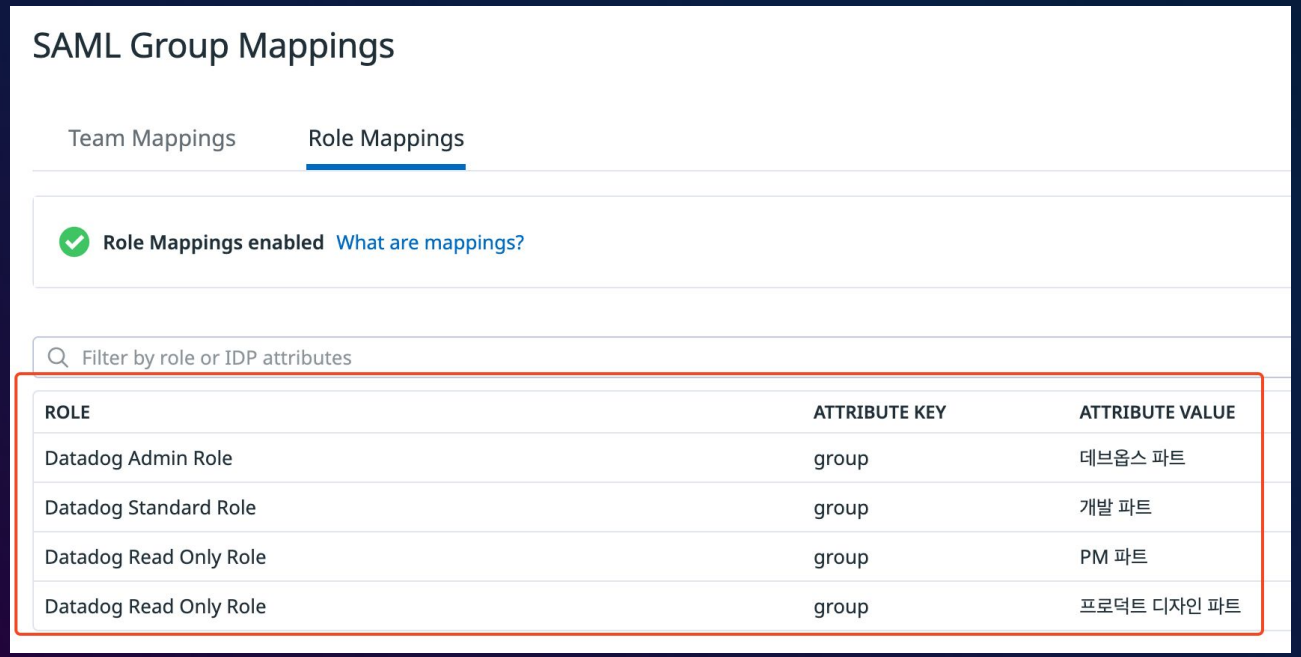
GWS 그룹 정보로 Datadog 팀 정보에 자동 연동

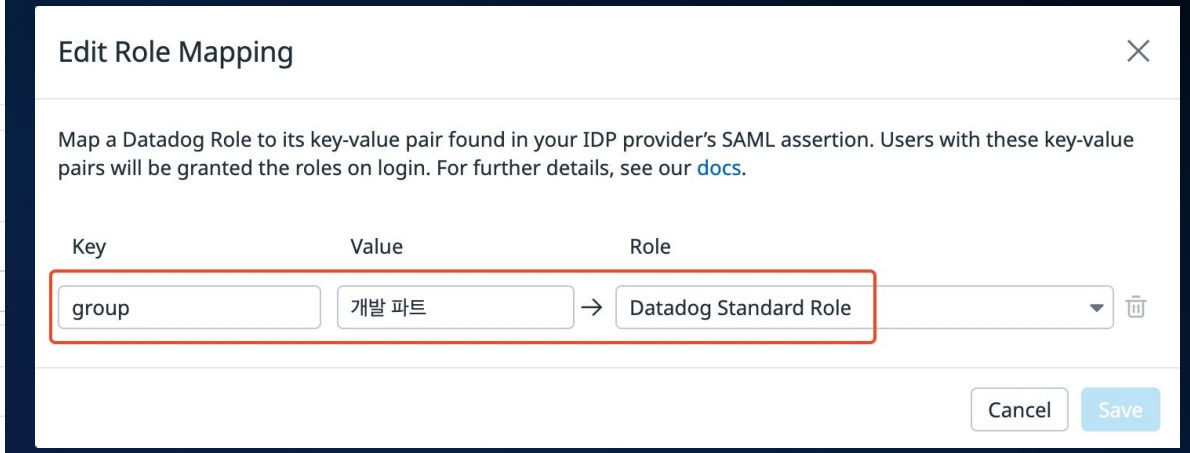




Google Workspace SAML 연동

GWS 그룹 정보로 Datadog 역할 자동 부여





Google Workspace SAML 연동

정리

- IdP (자격증명제공자)로부터 사용자 연동
- 사용자의 팀 정보 자동 연동
- 사용자의 역할 자동 부여
- **더 이상 이메일 초대 & 수동 권한 부여 불필요**

마무리

Datadog의 많고 다양한 기능 vs 사용자에게 정말 필요한 기능

Datadog 좋아요. 그리고 잘 쓰면 더 좋아요.

질문 답변

다음에 또 만나요!

Contact: joshua@inflab.com