

AWS환경에서 Serverless로

- 실시간 데이터 분석
플랫폼 만들기





하수용(syha)

MEGAZONECLOUD | AWSKRUG

Solutions Architect*(잡부) | syha@megazone.com

Interested in...

IaC(Infrastructure as Code)

Amazon Web Services

DevOps

AWS환경에서 Serverless로
실시간 데이터 분석 플랫폼 만들기





as Code)
ervices



건전한 토론 환영합니다.

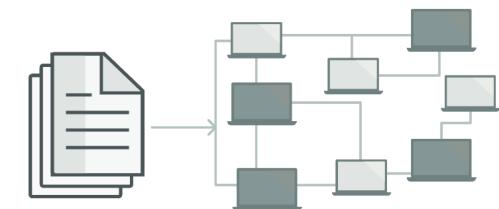
AWS 환경에서의 Serverless는?

AWS 환경에서의 Serverless를 말하기 앞서 '기존 환경은'?

기존 환경은



Server
provisioning/managing



그게 IaC로 배포된
VM/Cloud 환경이라도...

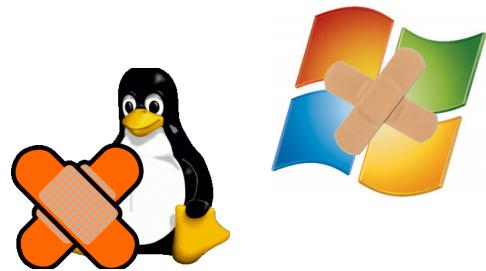


용량 확장 계획
수립

기존 환경은



취약점이 발견되면



Patching

그래서,



전지전능 해야 했거나,

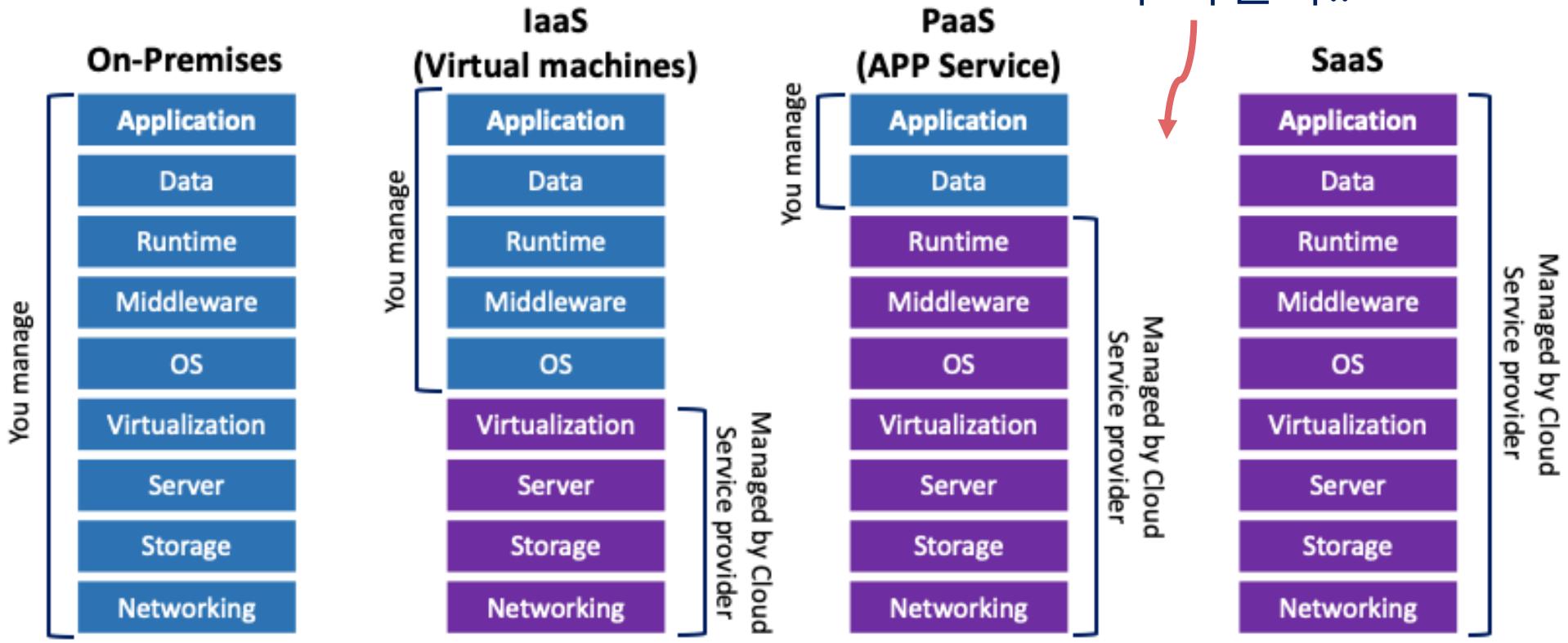
그래서,



작업은 분업했는데.. 어..음...

AWS 환경에서의 Serverless는?

요기 어딘가..



AWS 환경에서의 Serverless는?

서버 환경
구성/관리/운영
취약점 패치



좀 더 빠르게 개발,
좀 더 빠르게 배포

상대적으로 신경을 덜 쓰고

AWS 환경에서의 Serverless는?

Computing



AWS Lambda
Lambda Edge



AWS Fargate

Storage



Amazon S3



Amazon EFS

Data Store



Amazon DynamoDB



Amazon Aurora
Serverless

API Proxy



Amazon API G/W

Analytics



Amazon Kinesis



Amazon Athena

AWS 환경에서의 Serverless는?

App integration



Amazon SQS



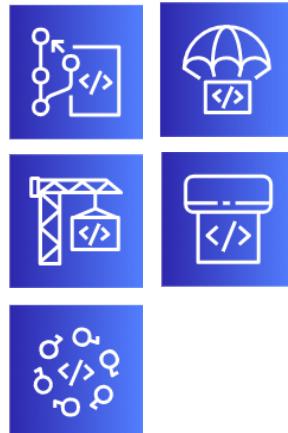
Amazon SNS

Orchestration



AWS Step Functions

Developer tool



AWS Code Series

ETC...

Amazon RDS Proxy
AWS AppSync
Amazon EventBridge
Amazon QuickSight
AWS KMS
Amazon Cognito
Amazon CloudWatch
.....



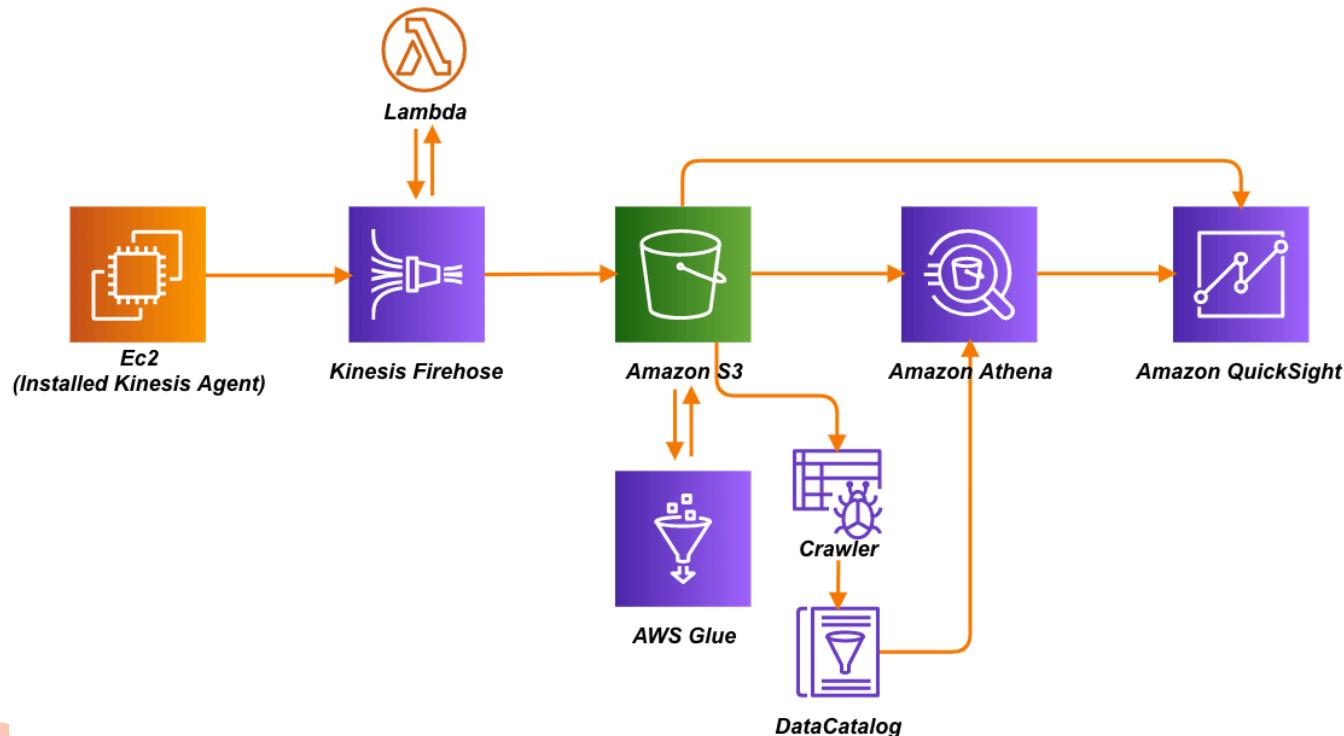




그래서 오늘의 이야기는

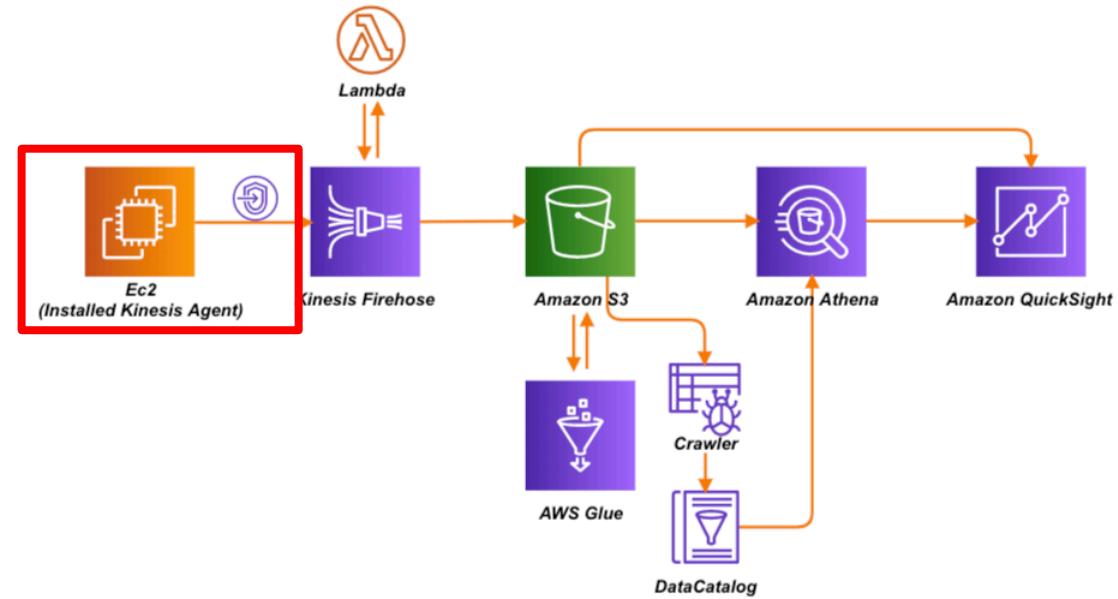
AWS환경에서 Serverless로 실시간 데이터 분석 플랫폼 만들기

Architecture



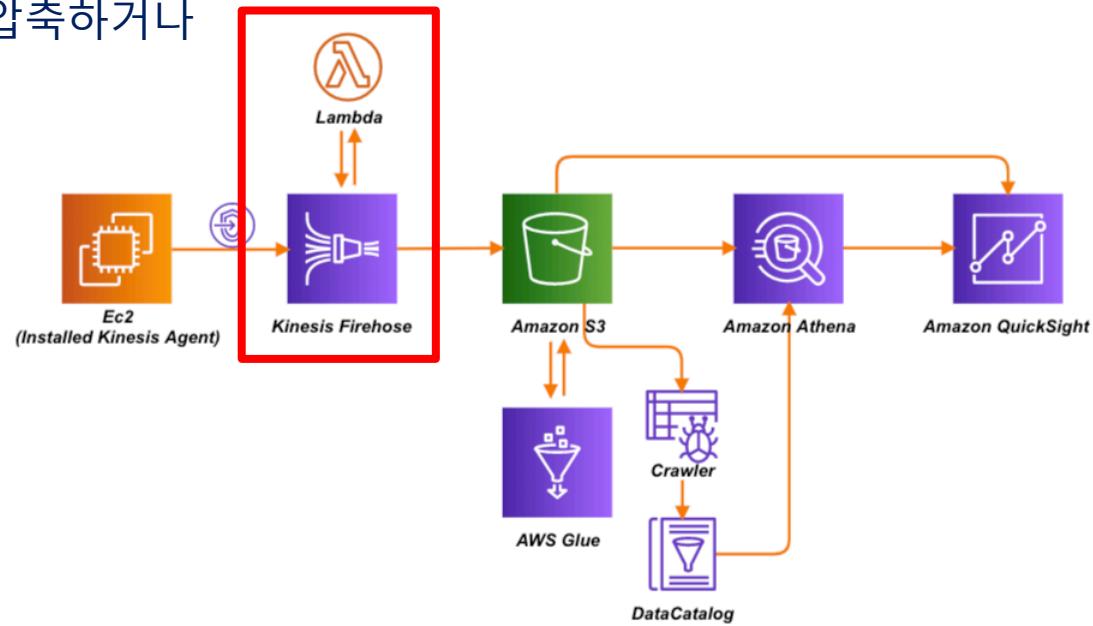
각 서비스 별 역할

- Kinesis Agent를 통해 로그 데이터를 Kinesis로 실시간 전송
- VPC endpoint를 통해 Private 통신/트래픽 비용 최소화



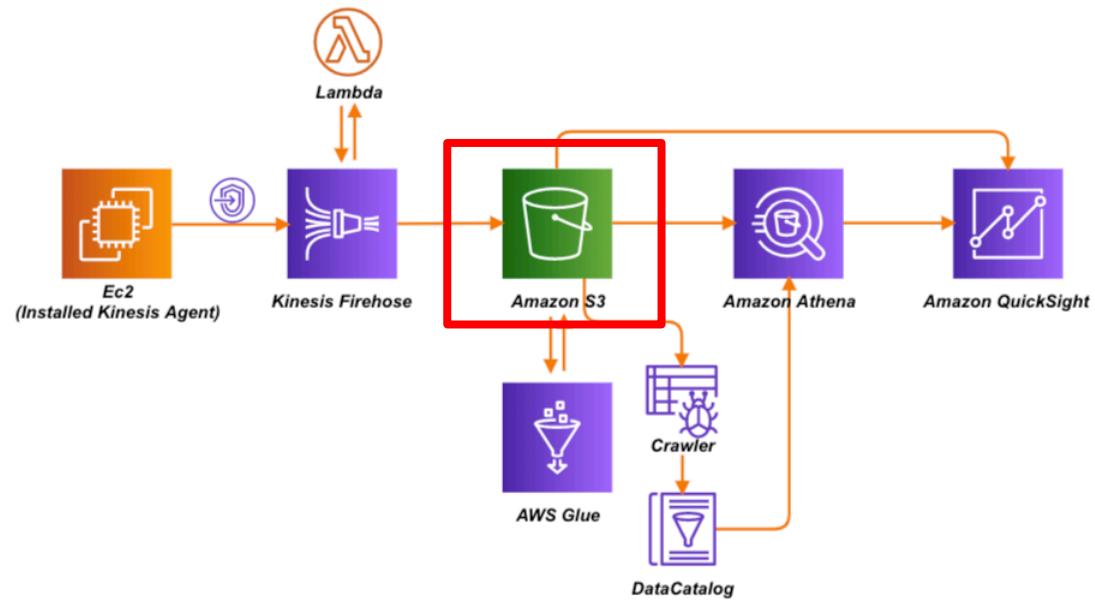
각 서비스 별 역할

- Agent로부터 수집된 데이터를 lambda를 통해 transform
- 정제된 데이터는 Data store인 S3로 데이터 저장
- 버퍼 크기 단위로 gz, zip, snappy로 압축하거나 Parquet, ORC 변환 가능



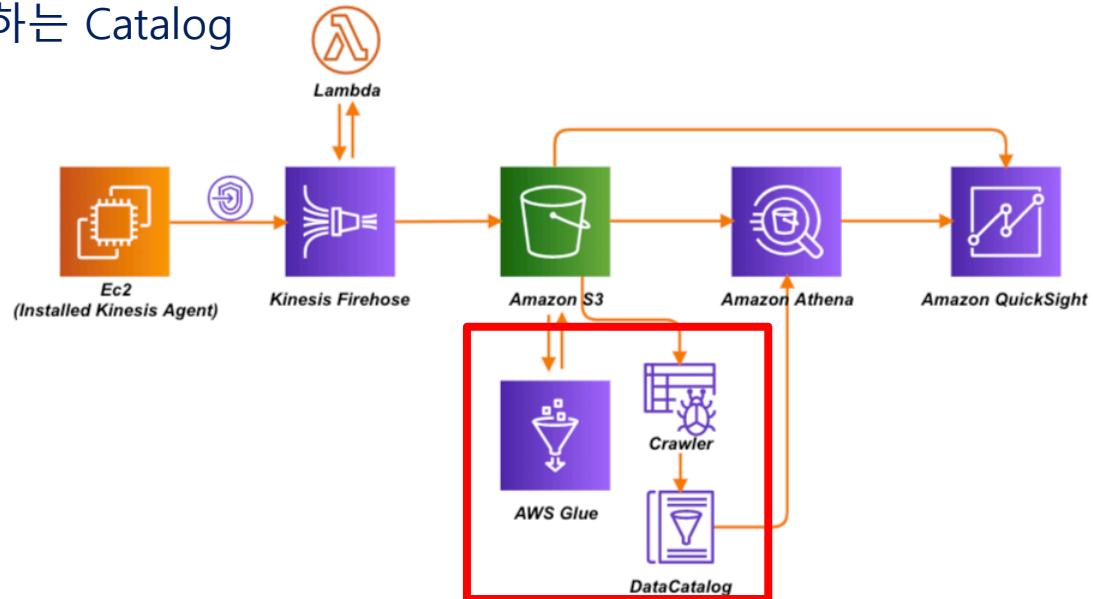
각 서비스 별 역할

- Kinesis에서 transform 된 데이터를 오브젝트로 저장
- 향후 데이터 분석을 위한 데이터 저장소로 이용



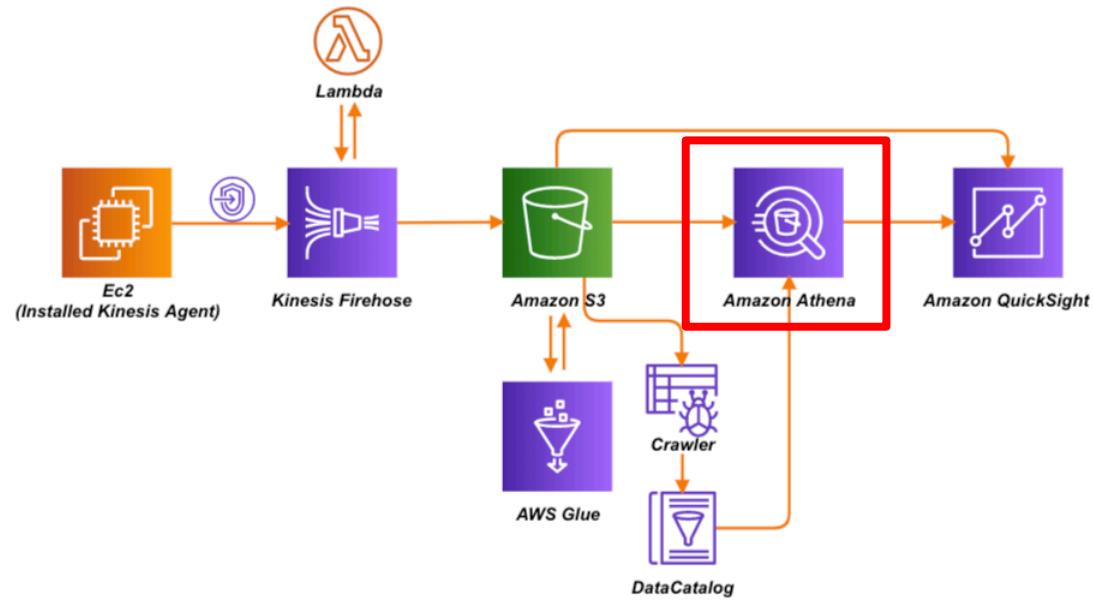
각 서비스 별 역할

- Amazon S3에 저장된 데이터를 ETL 처리
- 데이터의 스키마를 주기적으로 추적하는 Crawler
- 스키마를 저장하고, 타 서비스와 연계하는 Catalog



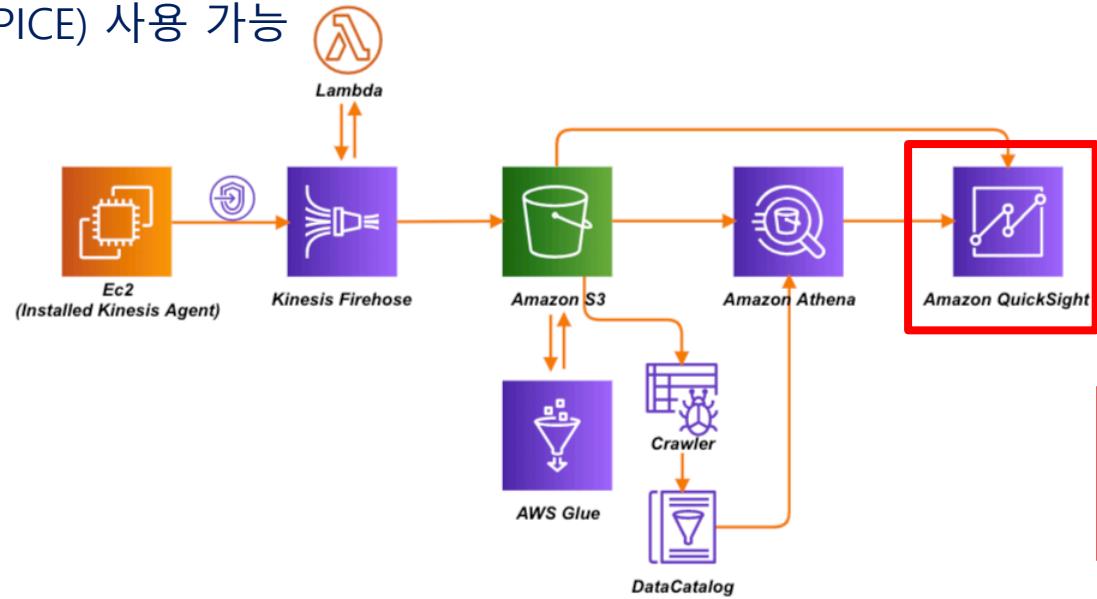
각 서비스 별 역할

- Amazon S3에 저장된 데이터를 대화식 쿼리로 분석할 수 있는 서비스
- Athena로 로드할 필요 없이 ANSI SQL을 이용한 임의 쿼리 실행
- 스캔한 데이터의 TB 당 비용 발생



각 서비스 별 역할

- AWS에서 제공하는 business intelligence tool
- 다양한 소스에서 읽은 데이터를 visualizations
- 초고속 병렬 In-memory 계산 엔진(SPICE) 사용 가능





동작 그만 데모는 안 보여주냐?



개인적인 의견으로..

Serverless가 모든 환경에서
효율적이라고는 생각하지 않습니다.



Let me know if you have any question