Cloud Native MCP AIOps

2025-1 AJOU SOFTCON

팀 명 여기서 잘해서 AWS 가겠습니다

팀 원 이민서, 김종민, 신명권, 황신철

지도교수 손태식 교수님

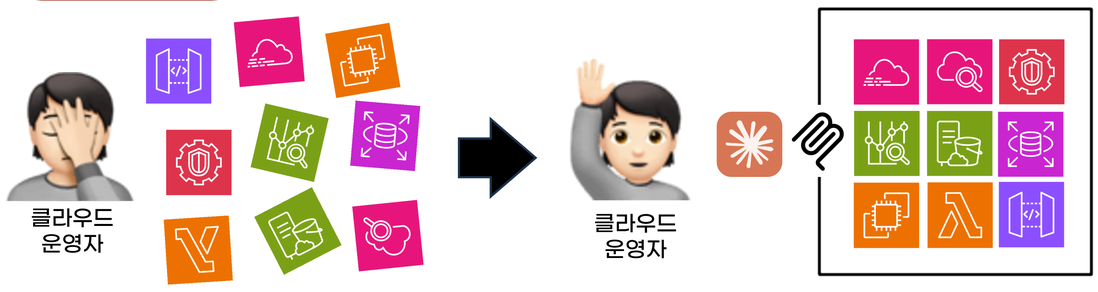
멘 토 인정식 (삼성전자MX 클라우드)

개발동기 및 목적

Target 클라우드 운영 및 관리 기업

Problem MSA 운영의 어려움, 전문 인력의 부족

Solution AIOps 챗봇으로 손쉬운 클라우드 운영 정보 분석



쿼리 및 콘솔을 통해
파편화된 MSA 운영 관리

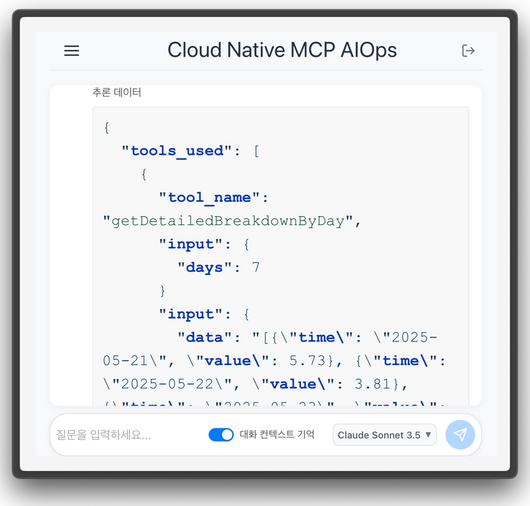
AIOps 챗봇을 통한
손쉬운 통합 운영 관리

개발 내용

Feature 1 웹/슬랙(메신저) 기반 자연어 운영 정보 질답

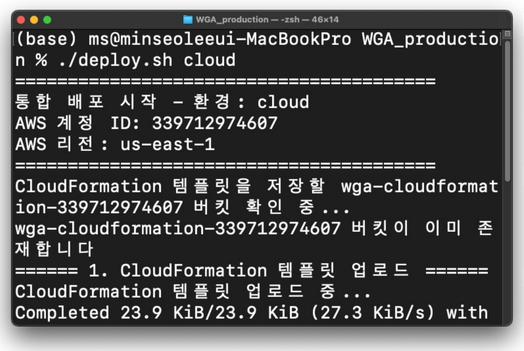
- 운영 관련 질문을 자연어로 질의, LLM이 **핵심 정보 요약 및 응답**
- **Slack 메신저 연동**을 통해 사용중인 채널에서도 바로 활용 가능
- **다중 모델 선택** 및 **대화 컨텍스트 기억 기능**을 통해 사용자 맞춤형 응답 품질 조절과 자연스러운 대화 흐름 유지 가능
- 복잡한 쿼리, 명령어, 콘솔 조작이 필요 없는 **사용자 친화적 채팅 인터페이스** 제공

Feature 2 MCP기반 데이터 접근 및 분석



- **LLM 추론 과정**과 **쿼리 실행 흐름**을 상세하게 확인 가능
- **MCP를 능통해 LLM이 직접** AWS 접근, 로그 및 메타 데이터 분석
- 단순 응답 생성에 그치지 않고, AWS SDK 등 외부 도구 호출을 통해 **실시간 데이터 분석 및 실행형 응답** 제공
- **비용, 이상징후 탐지 및 시각화**를 위한 다양한 **MCP 도구** 제공, 추가 연동 가능
- 입/출력 **비용 최적화** (1회당 약 200원)

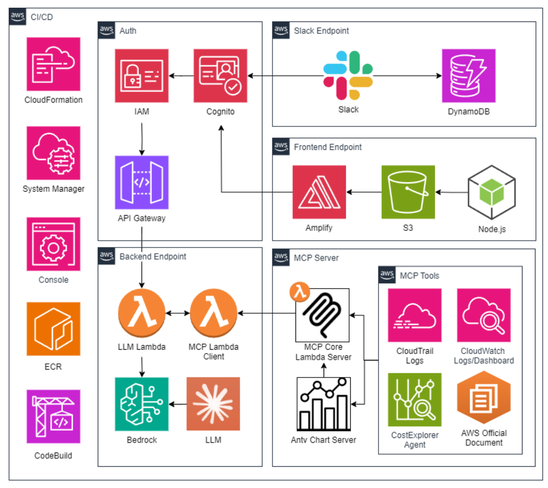
Feature 3 AWS CLI기반 CI/CD 원클릭 설치 및 배포



- AWS CDK와 CodeBuild를 활용, MSA 환경을 **IaC 코드** 기반으로 모든 환경에서 **고객의 인프라 내에 자동 배포**
- CloudFormation 템플릿 구조로, **확장성 및 유지보수 용이**
- 본 SW의 구축, 실행, 정보 접근, 요금 청구가 모두 **고객의 인프라 내에서 발생**

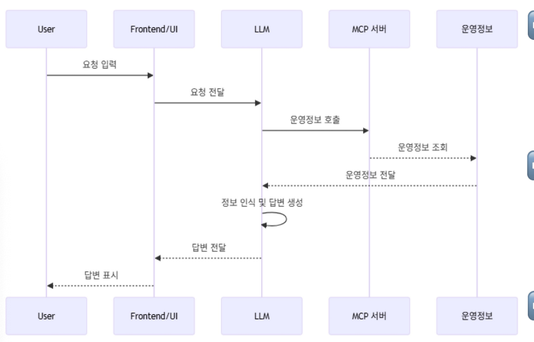
주요기술

Architecture Cloud Native with Serverless



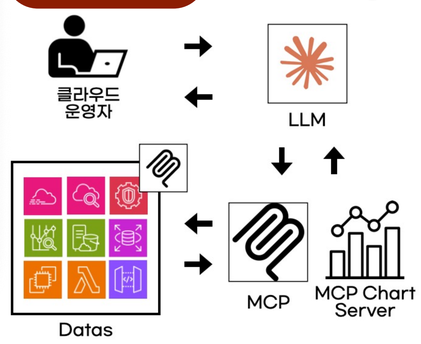
- CloudFormation, Codebuild, ECR, CLI, SystemManager 기반 **원클릭 완전 자동 배포** 구현
- Cognito OAuth 기반 인프라 **IAM 계정 및 권한 Authentication** 구현
- SSM Parameter Store, KMS 기반 **키, 환경변수 암호화 및 관리**
- Lambda 기반 **자동관리 Serverless Backend** 구현, 사용한 만큼만 비용을 지불하는 **Pay to Use 구조**
- S3, Amplify 기반 **Frontend CI/CD**
- DynamoDB 기반 **Stateful 세션 관리**로 ms 단위의 낮은 레이턴시와 빠른 처리 지원

Logic MCP (Model Context Protocol)



- Streamable-HTTP 기반 MCP 서버와 JSON-RPC 2.0 프로토콜을 통한 **실시간 도구 연동 및 다중세션 처리** 최적화
- 최대 15단계 MCP-Chaining을 통해 답변자가 검토 후, 다양한 **MCP 도구를 스스로 선택하여 연계**
- Presto SQL 엔진 활용 **CloudWatch 대용량 로그 병렬 처리**

Flow Working Flow



1. 클라우드 운영자가 LLM에 **질문**
2. LLM이 질문 판단 후 **MCP 클라이언트 생성**
3. MCP 클라이언트가 **MCP 서버 및 도구 호출**
4. MCP 도구 스스로 클라우드 운영 **데이터 탐색**
5. 데이터 기반 **답변 생성**
6. 검토 후 필요한 경우 다른 **도구 재요청**
7. 최종 답변 생성 후 **반환**

결과 및 분석

Evaluation LLM-MCP 성능 평가표 (자체 지표 및 실험)

	input 토큰 평균 (단위 : 개)	output 토큰 평균 (단위 : 개)	평균 가격 (Sonnet 3.7) (단위 : 원)	Fallback 비율 (단위 : %)	정확도 (단위 : %)	평균 답변 시간 (단위 : 초)	본 서비스로 인한 평균 답변 지연 (단위 : 초)
성공 가능성 높은 질문	45224.16	701.63	184.91	0	100	20.88	11.08
일반적이지만 다양한 표현을 가진 질문	45251.27	700.36	184.62	0.5	92.5	20.79	10.97
정보 부족하거나 모호한 질문	66725.35	688.11	272.23	2	95	23.51	13.88
오류를 유도하는 질문	32005.28	666.68	130.58	0	75.5	23.29	13.61
전혀 관련 없는 질문	4325.2	241	17.64	0	80	5.15	1.78

Review 클라우드 종사자의 서비스 평가

- 대기업 K사 Cloud DevOps 김**: 대박이다. **사내에 초청해서 PT 요청** 드리고 싶다.
- AWS 파트너 E사 Cloud DevOps 박** : MCP 구축이 정말 어려운데, **완성도가 높다.**
- AWS 파트너 N사 Cloud Consultant 이** : **실제 시장에서 좋은 평가**를 받을 것 같다.
- 중견기업 M사 Architect 민** : 유사 서비스를 사내 구축하여 사용 중. **상용화 기대한다.**

활용방안 및 기대효과

복잡한 AWS 콘솔 조작, 쿼리문 조회 대신 **자연어로 운영 정보 조회**
실시간 모니터링 및 장애 대응 업무 효율성 향상
모듈화된 구성으로 다양한 MCP 도구를 유연하게 **추가/확장 가능**

오픈소스 URL

https://github.com/WeGoAWS

아주대학교 | SW융합교육원