|  |  |
| --- | --- |
| 1. 1) Migrating Large Deep Learning Models to Serverless Architecture | |
| **저자** | Dheeraj Chahal, Ravi Ojha, Manju Ramesh, Rekha Singhal |
| **학회** | 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW) |
| **한줄 요약** | |
| 내용  : 대규모 딥러닝 모델을 서버리스 아키텍처로 효율적으로 이주하는 방법과 성능 평가에 대해 다루고 있음. | |
| **요약** | |
| 내용  : 대규모 딥러닝 모델을 서버리스 아키텍처로 이주하는 방법과 이에 따른 이점 및 도전 과제에 대해 다루고 있음. 또한 사전 훈련된 대규모 머신러닝 및 딥러닝 모델 FaaS(FaaS)와 같은 서버리스 컴퓨팅 플랫폼에 배포하는 방법론적 제안과 구현 전략을 제시하며, 성능 및 비용 평가를 제공함. | |
| **강점/약점** | |
| 내용   * 강점 * 인프라 관리의 단순화: 서버리스 모델은 서버 프로비저닝, 패치 작업 및 유지 관리와 같은 인프라 관리 작업을 단순화 함. * 자동 확장: 서버리스 아키텍처는 수요에 따라 자동으로 확장될 수 있음. 리소스는 동적으로 프로비저닝되어, 대규모 데이터 분석과 같은 복잡한 엔터프라이즈 클라우드 생태계에서도 강력한 기반을 제공함. * 약점 * 콜드 스타트 문제: 서버리스 아키텍처에서는 콜드 스타트 문제가 발생할 수 있으며, 이는 응답 시간에 영향을 줌. * 메모리 및 실행 시간 제한: 서버리스 환경은 메모리 및 실행 시간에 제한이 있어, 이는 대규모 딥러닝 모델을 실행하는 데 있어 제약이 될 수 있음. 또한, 이러한 제한은 모델의 성능과 정확도에 영향을 줄 수 있음. | |
| **Minor Comments** | |
| 오타 / 논리 부족 / 용어 미흡  : 논문을 검토하는 과정에서 위와 같은 부분은 없었음. | |