

경영진을 위한 SLO 가이드

새로운 고객 경험 패러다임을 위해 경영진이
서비스 수준 목표(SLO)를 고려해야 하는 이유



요약

현대의 애플리케이션은 그 어느 때보다 복잡하고 분산되어 있으며, 애플리케이션 성능은 그 어느 때보다 기업에 중요하게 작용합니다. 따라서 경영진은 조직이 비즈니스에 미치는 영향력을 입증해야 하는 부담을 느끼고 있습니다. 다행인 것은 애플리케이션을 현대화하고 관리하는 데 사용되는 것과 동일한 기술로 IT가 비즈니스에 미치는 영향을 직접적으로 연계할 수 있다는 사실입니다. IT 팀과 플랫폼 팀은 환경의 상태를 측정하고, 애플리케이션에 대한 기대치를 설정하며, 이러한 노력을 비즈니스 컨텍스트, 즉 서비스 레벨 목표(SLO)에 적용하는 새로운 방법을 발견했습니다. 오늘날 현대적 애플리케이션이 비즈니스에 차지하는 중요도를 고려할 때, 고위 경영진은 비즈니스 영향과 고객 경험 측면에서 SLO를 정의하는 조직의 사고방식을 가이드할 수 있어야 합니다.

본 문서는 SLO를 정의하고 구현하는 데 있어 일반적인 접근 방식과 빠지기 쉬운 함정을 검토하고, SLO를 활용하여 더 나은 비즈니스 경영 성과를 도출하는 모범 사례에 대한 명확한 지침을 제공합니다.

맥락

오늘날 “성공적인 디지털 비즈니스 운영을 위해서는 애플리케이션 성능과 안정성에 대한 최종 사용자의 만족이 매우 중요합니다.”¹ 하지만 최신 애플리케이션과 이를 기반으로 실행되는 플랫폼과 인프라가 그 어느 때보다 복잡하고 분산되어 있기 때문에 뛰어난 최종 사용자 환경을 제공하는 것은 결코 쉽지 않습니다. 실제로 2021년 멀티클라우드 상태 보고서에서 알 수 있듯이, 복잡성은 비즈니스 목표를 달성하기 위한 가장 중요한 숙제 중 하나로 꼽혔습니다.²

IT 팀은 현대적인 애플리케이션의 개발과 함께 “하이브리드 IT, 멀티클라우드, 컨테이너를 포함한 신속한 변화를 위한 인프라로의 전환” 과정에서 발생하는 엄청난 양의 데이터를 파악하는 데 어려움을 겪고 있기 때문에, “기존 인프라 모니터링 도구의 실행 가능성에 도전하고 있습니다.”³

IT 팀과 플랫폼 팀은 이러한 어려움에 대응하기 위해 환경의 상태를 측정하고, 애플리케이션에 대한 기대치를 설정하며, 그러한 노력을 비즈니스에 미치는 영향과 연결하는 새로운 방법을 찾았습니다. 이와 함께 최종 사용자 기대치가 지속적으로 증가하고 있으므로, 경영진이 의미 있는 측정 기준을 가지고 애플리케이션 성능과 비즈니스에 미치는 영향을 평가하도록 보장하는 것이 중요합니다.

1 출처: 2020년 전 세계 애플리케이션 성능 관리 소프트웨어 시장 점유율 | 2021년 6월, IDC #US47989021

2 출처: 2021 Turbonomic 멀티클라우드 상태 보고서, 2020년 CNCF 설문조사

3 출처: 현대 서비스 및 인프라의 모니터링과 관측성 | 2020년 6월, Gartner G00720854

현재 접근 방식: 노동 집약적, 반응형

IT 팀과 플랫폼 팀이 환경의 상태를 측정하기 위해 취하는 일반적인 방식은 서비스 수준 지표(SLI), 서비스 수준 계약(SLA), 서비스 수준 목표(SLO)를 식별하고 구성하는 것입니다.

- 서비스 수준 지표(SLI)는 팀이 목표와 관련하여 서비스를 수행하는 방법을 측정하는 데 사용되는 메트릭입니다.
- 서비스 수준 계약(SLA)은 IT 팀과 플랫폼 팀이 고객 또는 최종 사용자에게 제시하는 가용성에 대한 약속입니다.
- 서비스 수준 목표(SLO)팀이 달성하기 위해 노력하는 서비스 수준 목표 또는 목표를 SLI로 측정하는 것입니다.

SLO는 기준을 정의하고 조직의 애플리케이션 성능에 대한 명확한 기대치를 설정하기 때문에 특히 중요합니다.

오늘날 대부분의 조직은 시간과 노력이 많이 소요되는 수동 SLO 구성 프로세스를 사용하고 있습니다. IT 팀은 먼저 비즈니스와 최종 사용자의 경험에 직접적인 영향을 미치는 애플리케이션의 서비스를 결정해야 합니다. 다음으로, 서비스 수준 지표(SLI)로 사용할 메트릭을 식별해야 합니다. 일반적인 SLI 메트릭은 가용성, 지연 시간, 트랜잭션 처리량입니다. 적절한 SLI가 선택되면 IT 팀은 해당 메트릭과 특정 측정 기간에 대한 SLO 목표를 결정해야 합니다. IT 팀은 이 프로세스를 완료한 후 각 SLO에 대한 오류 예산을 작성하고 임계값 기반 경고 시스템에 연결합니다. 많은 조직이 이러한 임계값 기반 접근 방식을 채택하고 있는 이유는 인력이 애플리케이션의 성능을 연중무휴 24시간 동안 모니터링하는 것이 불가능하기 때문입니다.

안타깝게도 SLO 임계값을 설정해도 성능은 해결되지 않습니다. 이러한 IT 전략은 멀티클라우드 환경이나 컨테이너형 환경에서 실행되는 애플리케이션을 갖춘 오늘날의 민첩한 인프라에 너무 민감하게 반응하기 때문에 비효율적입니다. 서비스에 문제가 발생하고 임계값 또는 경고가 트리거되면 해당 서비스의 성능이 순식간에 저하되므로 최종 사용자 환경이 열악해집니다.

IT 팀은 컨테이너형 환경에서 실행되는 애플리케이션을 위해 HPA(Horizontal Pod Autoscaling)를 구현하여 임계값/경고 시스템을 개선하려고 노력했습니다. 그러나 HPA 역시 우수한 최종 사용자 환경을 보장하지 못하고 성능 저하도 피할 수 없습니다. SLO 구성 프로세스와 마찬가지로 리소스 수요를 충족하도록 수평 자동 확장을 설정하려면 IT 팀이 리소스 요구 사항을 가장 잘 나타내는 메트릭을 식별하고 목표를 구성한 후 임계값을 설정하고 테스트해야 합니다. 이러한 프로세스는 애플리케이션의 모든 서비스마다 반복되어야 합니다. 일부 애플리케이션은 수백 가지의 다양한 서비스를 제공하므로 HPA를 대규모로 구현하는 것은 상당히 어렵습니다. 또한 서로 다른 HPA 정책은 여전히 임계값에 의존하고 서로 상관되거나 따르지 않으므로 한 서비스를 확장할 경우 다른 서비스에 부정적인 영향을 미칠 수 있습니다. 마지막으로, 이것은 일회성 연습이 아니며 HPA 확장 정책이 효과적이기 위해서는 지속적인 재구성과 모니터링이 필요합니다.



SLO에 대해 고려할 점

고객 경험 측면에서 SLI, SLO를 정의

업계 전문가들은 “빠른 성능과 100% 가동 시간이 디지털 비즈니스 성공의 관건이므로 최종 사용자 환경을 최적화하기 위해서는 **시스템과 애플리케이션 관리에 집중하는 것이 중요하다**”고 말합니다.⁴ 이제는 빠른 성능과 가동 시간이 중요한 요소이기 때문에, 조직은 애플리케이션 성능을 평가할 때 최종 사용자 경험을 직접적으로 나타내지 않는 데이터를 수집하는 데 시간을 낭비해서는 안 됩니다.

올바른 데이터를 수집하는 것은 쉬운 일이 아닙니다. 예를 들어 가용성은 일반적으로 사용되는 메트릭이지만, 성능이 직접적으로 확장되는 것은 아닙니다. 이유는 애플리케이션을 사용할 수는 있다 해도 여전히 리소스 병목 현상과 성능 저하로 인한 문제가 일어나기 때문입니다. 조직에서 성능을 측정하는 데 사용할 수 있는 다양한 측정 기준이 있으므로 IT 팀과 LOB가 어떤 데이터를 수집하고 보고할 지에 대해 합의하는 것이 중요합니다. Gartner는 다음과 같이 말합니다. “대표적이고 의미 있는 SLI를 선택하는 것은 매우 중요합니다. 대부분의 경우 인프라 기반 메트릭(‘사용 가능한 메모리’ 또는 ‘%의 사용 가능한 작업자 노드’)은 서비스 사용자가 별로 신경 쓰지 않기 때문에 의미가 없습니다. 사용자의 서비스 경험을 직접 측정하는 SLI를 선택해야 합니다.”⁵

간단히 말해서, SLO는 애플리케이션이 비즈니스에 필요한 작업을 수행하고 있는지 여부를 조직에 알려주는 도구로 사용되어야 합니다. 애플리케이션이 성능을 발휘하지 못하는 경우, 애플리케이션의 적절한 수준에서 의미 있는 SLO를 규정해 놓은 조직은 성능을 복원하는 데 가장 큰 영향을 미치는 작업이 무엇인지 정확히 파악해야 합니다. 조직마다 LOB에 따른 구체적인 측정 메트릭이 필요합니다. 그러나 조직이 단일 애플리케이션 아키텍처에서 벗어나 보다 현대적이고 분산적인 환경으로 전환함에 따라 높은 메모리 및 CPU 사용량과 같은 기존 성능 지표는 유용성이 떨어지고 있습니다. 그 대신, 조직은 응답 시간, 트랜잭션 처리량과 같이 일반적으로 비즈니스와 관련된 메트릭 측면에서 SLI/SLO를 정의해야 합니다. 예를 들어 트랜잭션 처리량에 대한 SLO를 정의하면 IT 팀이 각 개별 팟 또는 VM에 대한 요청이 어떻게 처리되고 있는지 정확하게 파악할 수 있으므로 이러한 메트릭은 보다 직접적으로 성능을 측정할 수 있습니다. 이 메트릭은 사용량 메트릭을 기준으로 SLO를 정의한 후 응답 시간, 처리량과 같은 메트릭과 대강 연결하는 것이 아니라, 성능을 보다 직접적으로 평가합니다.

4 출처: 2020년 전 세계 애플리케이션 성능 관리 소프트웨어 시장 점유율 | 2021년 6월, IDC #US47989021

5 최신 인프라 및 애플리케이션 모니터링을 위한 솔루션 경로 | 2019년 6월, Gartner

응답 시간과 트랜잭션 처리량은 고객 경험과 애플리케이션 성능을 평가하는 가장 일반적인 방법 중 하나이지만 이러한 메트릭이 모든 조직에 적용되는 것은 아닙니다. 예를 들어, VDI 서비스 공급자인 조직에서는 SLO를 트랜잭션 처리량에 대해 정의하지 않고 지원 티켓에 대해 정의하려고 합니다. SLO는 가상 데스크톱의 성능을 직접 측정하므로 지원 티켓이 더 효과적인 메트릭입니다. 궁극적으로 SLI/SLO는 각기 다른 조직과 해당 LOB에 따라 구체적으로 정의되어야 합니다.

애플리케이션 수요 변화에 대해서도 지속적인 분석이 필요

현대의 애플리케이션과 인프라는 탄력적이며 리소스 수요는 역동적입니다. 이러한 지속적인 변화로 인해 서로 다른 소스와 데이터 유형 간의 관계를 이해하는 것은 상당히 어렵습니다. 조직은 흔히 서로 다른 도구를 사용하여 여러 부서 간에 각기 다른 스택 계층을 모니터링하는 동시에 동일한 문제의 다른 측면을 해결하려고 합니다. 이러한 조정력 부족은 문제의 근본 원인으로 잘못된 결과로 이어질 수밖에 없습니다. 또한 데이터는 지속적으로 수집되어야 합니다. 그렇지 않은 경우, IT 팀과 플랫폼 팀에서 추측한 데이터만 가지고 문제를 해결해야 하기 때문입니다. Gartner가 2020년 모니터링 및 관측성 보고서에서 언급한 바와 같이, “잠재적 원인 데이터를 수집하는 것은 반드시 지속적으로 이루어져야 합니다. 증상에 대한 대응으로만 데이터 수집을 한다면 원인을 완전히 놓치기 십상입니다.”⁶

이러한 딜레마를 피하기 위해서는 조직은 애플리케이션 스택의 모든 계층에서 데이터를 집계하고 상호 연관시킬 수 있는 중앙 저장소가 필요합니다. 이 시스템을 통해 지속적으로 수집된 데이터를 통합하여 SLO에 반영할 수 있습니다. 이러한 지속적인 분석을 통해 조직은 모든 것을 맥락화하고 성공적인 애플리케이션 성능을 위해 어떤 목표에 도달해야 하는지 알 수 있습니다. 또한 목표를 달성하지 못할 경우 애플리케이션 스택의 기본 계층에서 수정해야 할 문제는 무엇일까요? 궁극적으로는 지속적인 분석을 수행할 수 있는 시스템을 구현하는 것이 최신 애플리케이션과 인프라의 동적 리소스 요구를 충족시킬 유일한 방법입니다.



동적 리소스 자동화

현대 애플리케이션의 동적인 특성과 이를 기반으로 실행되는 인프라에 지속적인 분석이 요구되는 것처럼, 경영진과 조직 또한 SLO를 완전히 활용하고 애플리케이션 관리를 위한 예방 시스템을 구축하기 위해서는 지속적인 자동화를 구현해야 합니다. IDC의 최근 분석 보고서에서는 조직이 미래 경쟁력을 유지하려면 “제품 기능의 자동화의 역할을 고려해야 한다”고 밝혔습니다.⁷

6 출처: 현대 서비스 및 인프라의 모니터링과 관측성 | 2020년 6월, Gartner G00720854

7 출처: 2020년 전 세계 애플리케이션 성능 관리 소프트웨어 시장 점유율 | 2021년 6월, IDC #US47989021

현대 인프라의 발전은 미래의 애플리케이션 복원력과 탄력성을 약속했지만, 오늘날 대다수 조직은 애플리케이션의 성능을 유지하기 위해 고군분투하고 있습니다. 복원력과 탄력성에서 비롯되는 이점을 얻기 위해서는 최신 애플리케이션 관리에 자동화를 구현하는 것이 필수입니다. 올바른 SLO를 구성하고 지속적으로 데이터를 수집한다 해도 자동화 없이는 그 해결책이 효과적일 수 없습니다. IT 팀과 플랫폼 팀은 동적 리소스 자동화 없이 애플리케이션 성능을 보장할 수 없습니다. 경고가 트리거되고 리소스 결정이 수동으로 처리된다는 것은 이미 성능이 저하되었다는 의미입니다.

궁극적으로, 이 접근 방식의 문제는 이미 문제가 발생했다는 사실을 전제로 한다는 점입니다.

자동화는 임계값에 의해 트리거된 이벤트에 대한 단순한 반응 이상의 기능을 제공합니다. 지속적인 데이터 수집을 활용하고 LOB에 적합한 주요 SLI, SLO를 식별함으로써 조직은 자동으로 실행 가능한 결정을 생성하는 소프트웨어를 보유할 수 있습니다. 조직에서 필요한 단계를 수행하면서 자동화에 전념할 경우, 애플리케이션의 리소스 할당 방식을 사전 예방적으로 관리해주고 지속적인 성능을 보장하는 진정으로 탄력적인 환경을 구축할 수 있습니다. 이러한 유형의 자동화를 위해서는 동적으로 변화하는 환경을 분석하고, 필요한 일련의 의사 결정을 자동화하여 성능 저하가 발생하기 전에 문제를 해결하는 지능형 시스템이 필요합니다. 스프레드시트와 경고 프로세스를 통해서 이러한 성능과 탄력성을 달성하는 것이 불가능합니다.

사후 대응적인 접근 방식에서 벗어나려면 애플리케이션을 관리하는 다양한 팀과 인프라를 관리하는 팀의 이해 관계자들이 완전히 동의해야 합니다. 애플리케이션 및 제품 소유자는 흔히 애플리케이션에 대한 제어를 자동화하는 것을 주저합니다. 이는 자동화에 대한 신뢰가 부족하기 때문이지만 극복될 수 있습니다. 자동화를 위해서는 IT 조직의 문화적 변화가 필요합니다. 진정한 탄력성과 복원력을 얻으려면 조직이 자동화 조치를 신뢰할 수 있어야 합니다. 플랫폼과 인프라의 동적 리소스와 직접 연결된 의미 있는 비즈니스 및 애플리케이션 중심 SLO가 갖춰진다면 애플리케이션 및 제품 소유자는 자동화를 더욱 쉽게 신뢰할 수 있을 뿐 아니라, 비즈니스 워크플로우에 자동화가 완벽하게 구현된다고 확신할 수 있습니다.

주의 사항

조직이 하이브리드 IT, 컨테이너 플랫폼, 멀티클라우드와 같은 민첩한 인프라를 점점 더 많이 채택함에 따라 애플리케이션은 더욱 복잡해질 것입니다. 향후 이러한 상황이 심화됨에 따라 고객 경험과 애플리케이션 성능에 대한 높아진 기대치를 충족하는 조직은 성공을 이루게 될 것입니다. 따라야 할 모범 사례는 고객 경험 측면에서 SLO를 정의하고, 변화하는 애플리케이션 수요를 지속적으로 분석하며, 동적 리소스를 자동화하는 것입니다. 디지털 비즈니스에서 성공하고자 하는 조직은 변화하는 요구와 비즈니스 SLO를 충족하기 위해 애플리케이션을 동적으로 리소스화하는 자동화를 구현하는 데 필요한 조치를 취해야 합니다.

Turbonomic을 통해 애플리케이션 SLO를 자동으로 보장

Turbonomic은 데이터를 실행하여 애플리케이션 성능 위험을 방지하는 동시에 탄력성을 극대화하는 자동화 기능을 제공합니다.

업무상 중요한 애플리케이션 및 인프라를 현대화하는 것은 여러 가지 이점을 기대할 수 있는 투자입니다. 그러나 탄력성, 복원력, 출시 속도의 이점을 누리려면 환경을 지속적으로 분석하고 적절한 리소스 결정을 적시에 실행하여 애플리케이션 성능을 보장해 줄 수 있는 소프트웨어가 필요합니다. Turbonomic을 사용하면 애플리케이션 응답 시간, 트랜잭션 처리량 또는 기타 SLI/SLO를 동적 리소스와 연계할 수 있습니다. 수요가 변동해도 Turbonomic의 동적 리소스는 지속적인 애플리케이션 성능을 보장합니다.

HPA으로는 역부족: Turbonomic은 하향식 전체 스택 분석을 사용하여 SLO를 동적으로 보장합니다. SLO를 설정하고 AI 기반 소프트웨어를 사용하면 애플리케이션이 실행되는 곳이 어디든 관계없이 플랫폼과 기본 인프라가 SLO를 충족하는 데 필요한 리소스를 제공할 수 있습니다.

비즈니스 워크플로우에 원활하게 통합: Turbonomic과 Webhook을 통합하면 Turbonomic 작업을 애플리케이션 라이프사이클, DevOps, 인프라 파이프라인, 승인 및 감사 워크플로우, 커뮤니케이션 프로세스에 쉽게 삽입할 수 있습니다.

수동 작업 최소화: 개발자, DevOps, SRE는 임계값, 제약 조건 또는 자동 확장 정책을 설정할 필요가 없습니다. 소프트웨어는 실제로 자동화할 수 있는 작업을 제공하여 사용자에게 적합한 리소스 결정을 내립니다.

용량을 위한 과도한 비용 지출 방지: 리소스 결정을 내리는 데 있어 개발자에게 의지할 필요가 없습니다 (개발자는 안전을 위해 과도하게 프로비저닝하는 경우가 많습니다). IBM 소프트웨어는 애플리케이션 수요를 바탕으로 서비스에 필요한 리소스를 정확하게 결정합니다.

빠르고 쉬운 성장 계획 수립: IBM 소프트웨어를 사용하여 새로운 서비스의 온보딩을 시뮬레이션합니다. 그리고 새로운 성장을 지원하기 위해 무엇이 필요한지 정확히 결정합니다.

지금
Turbonomic
체험하기

turbonomic.com/try-SLO

© Copyright IBM Corporation 2022
(07326) 서울특별시 영등포구 국제금융로 10
서울국제금융센터(3IFC)

미국에서 제작됨
2022년 1월

IBM과 IBM 로고는 미국 및/또는 기타 국가에서 사용되는 International Business Machines Corporation의 상표 또는 등록 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 기타 회사의 상표일 수 있습니다. 최신 IBM 상표 목록은 다음 웹페이지를 참조하십시오. ibm.com/trademark

이 문서는 최초 발행일 기준 최신 문서로, IBM은 언제든지 해당 내용을 변경할 수 있습니다. IBM이 현재 영업 중인 모든 국가에서 모든 제품이 제공되는 것은 아닙니다.

명시된 성능 데이터 및 고객 사례는 정보 전달 목적으로만 제공됩니다. 실제 성능 결과는 특정 구성 및 작동 조건에 따라 다를 수 있습니다.

본 문서의 정보는 상품성, 특정 목적에 대한 적합성, 비침해성 보증/조건을 포함한 어떠한 명시적 또는 암시적 보증 없이 “있는 그대로” 제공됩니다. IBM 제품은 함께 제공되는 계약 동의 조건에 따라 보증됩니다.

