

# 솔루션 소개

가속 컴퓨팅 시스템 및 그래픽  
미디어 처리 및 전송



## 미디어 처리 및 전송을 위한 인텔® 데이터센터 GPU Flex 시리즈



미디어 처리 및 전송 제공업체는 인텔® 데이터센터 GPU Flex 시리즈로 컴퓨팅 밀도와 대역폭 소비를 모두 최적화하여 네트워크를 효율적인 비용으로 미래에 대비할 수 있도록 지원합니다.



시스코(Cisco)는 비즈니스와 소비자 동영상이 전 세계 인터넷 트래픽의 80% 이상을 차지한다고 보고하였습니다.<sup>1</sup> 실제로 동영상은 사용자가 만들어낸 풍부한 콘텐츠가 커뮤니케이션과 표현의 주요 수단이 되어서 엄청난 사회적 중요성을 가지게 되었습니다.

이전보다 정교한 콘텐츠에 대한 요구는 처리 인프라와 전송 대역폭 모두에 대해 제공업체의 비용 상승 압력을 야기합니다. 가입자들은 증가하는 다양한 장치에서 연속적으로 더 높은 해상도를 포함하여 인터넷 전송에서 높은 방송 품질을 기대합니다. 동시에 5G 네트워크의 구축은 모바일 가입자들이 사용할 수 있는 처리량을 엄청나게 증가시키고 있습니다.

사용자당 평균 매출(ARPU)이 더 이상 올라가지 않는 상황에서, 높은 품질에 대한 가입자 기대치를 충족시키는 것은 공급자에게 상당한 효율성 문제를 야기합니다. 높은 처리 성능은 서버당 스트림 밀도를 높이고 AV1을 포함한 고급 스트리밍 코덱에 대한 최적의 지원을 제공함으로써 비용 요구 사항을 충족하는 데 중추적인 역할을 합니다. 이러한 전환은 더 높은 해상도와 프레임 레이트의 고급 비디오 콘텐츠를 공급하기 위해 제공업체에 매우 중요합니다.

인텔® 제온® 스케일러블 프로세서는 오늘날 미디어를 처리하고 전송하는 데 있어 최고의 표준 솔루션입니다. 이 프로세서는 미디어 제공자가 폐쇄적이고 독점적인 환경의 제약을 극복하기 위해 매우 원하는 공통 오픈 소스 도구와 호환됩니다. 그러나 이와 동시에 기기종 하드웨어 및 소프트웨어 환경은 고유한 워크로드에 대해 지속되므로 워크로드가 복잡해지는 가운데 고밀도의 데이터 센터에 효율성을 제공하는 동시에 새로운 요구사항을 충족할 수 있는 유연한 솔루션이 필요합니다.

인텔® 데이터센터 GPU Flex 시리즈는 서버급 안정성, 가용성 및 확장성을 갖춘 미디어 스트림 밀도 및 품질에 최적화된 범용 데이터센터 그래픽 프로세서입니다. 개별적으로 또는 인텔® 제온® 프로세서와 함께 사용하여 유연하고 개방적인 미디어 솔루션을 제공하여 오늘날의 변화가 심하고 도전적인 미디어 전달 요구 사항을 충족할 수 있습니다.

### 솔루션 관련 성능 수치

**5X** 경쟁사의 절반 수준의 전력으로  
미디어 트랜스코드 처리량 제공  
인텔 Flex 시리즈140 GPU와  
NVIDIA A10 비교

HEVC 1080p60<sup>3</sup>

**2X** 경쟁사의 절반 수준의 전력으로  
디코딩 처리량 제공  
인텔 Flex 시리즈140 GPU와  
NVIDIA A10 비교

across HEVC, AV1, AVC, VP9<sup>3</sup>



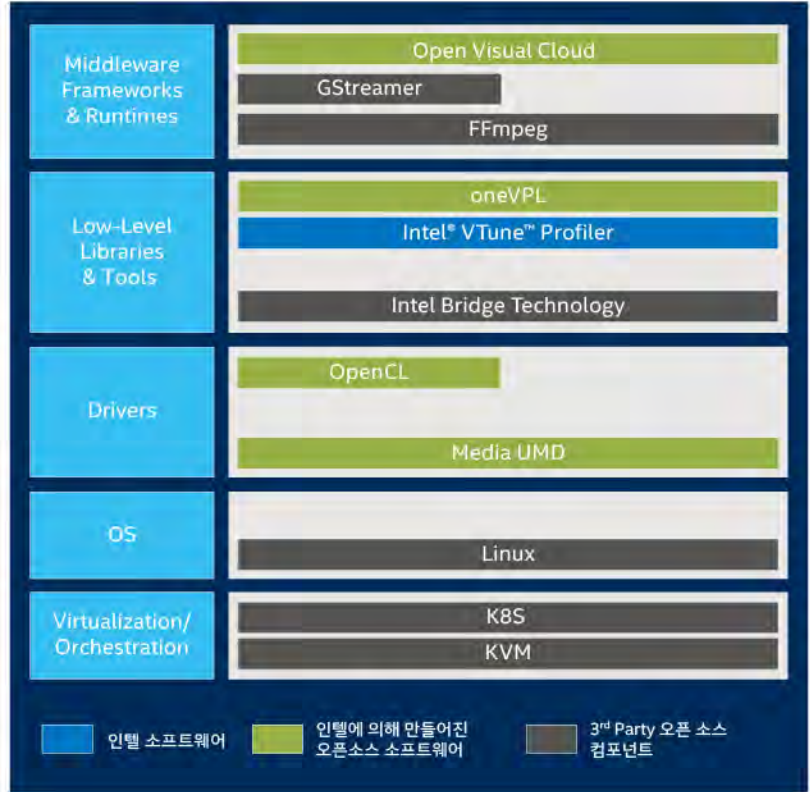
## 개방형 표준 아키텍처

CUDA와 같은 독점 프로그래밍 모델에서 GPU용으로 개발된 코드는 다른 하드웨어로의 이식성이 부족하여 조직을 폐쇄된 생태계에 가두는 고립화된 개발 관행을 만듭니다. 이와 대조적으로 인텔® 데이터센터 Flex 시리즈 GPU는 oneAPI 아키텍처 간 프로그래밍과 함께 개방형 표준 기반 소프트웨어 스택을 지원하므로 개발자는 인텔 CPU 및 GPU에서 원활하게 실행되는 고성능 미디어 애플리케이션 및 솔루션을 구축할 수 있습니다.

oneAPI를 기반으로 하는 개방형 표준 코드 개발은 오픈 소스 도구, API 및 드라이버를 포함하는 대규모 개방형 에코시스템의 장점을 제공합니다. 이러한 유연성은 조직으로 하여금 복잡성, 비용 및 시간에 대한 요구 사항을 해결하도록 도와주어서 엔지니어와 프로그래머가 코드를 유지 관리하는 대신에 새로운 서비스와 솔루션을 시장에 출시하고 새로운 아키텍처의 채택을 간편하게 하고 혁신할 수 있도록 지원합니다.

## 산업 에코시스템

표준 기반 개방형 아키텍처와 최적화된 소프트웨어 스택의 이점을 확장하는 인텔® 데이터센터 GPU Flex 시리즈는 서비스 제공업체, ISVs(Independent Software Vendors), OEM(Original Equipment Manufacturer) 및 기타 업체로 구성된 광범위한 에코시스템을 활용하여 광범위한 사용 사례를 지원합니다.



이러한 회사는 게시된 oneAPI oneVPL 사양을 활용하여 GPU를 미디어 처리 및 전달을 위한 기술과 통합합니다. 프로그래밍 모델의 개방성과 투명성은 또한 오픈 소스 커뮤니티의 채택을 장려하여 소프트웨어 스택을 더욱 향상시키는 선순환을 만듭니다.

이러한 미디어 스트림을 제공할 때 기본 하드웨어의 기능을 효과적으로 실현하기 위해 인텔은 소프트웨어 에코시스템이 이를 활용할 수 있도록 지원합니다. 이 작업은 소프트웨어 표준, 프레임워크 및 FFmpeg, GStreamer 및 Handbrake와 같은 오픈 소스 기술과 이를 사용하는 고객이 일반적으로 CPU에서만 가능했던 GPU 성능을 달성할 수 있도록 합니다. 이 목표를 달성하기 위해 인텔은 CPU 및 GPU 아키텍처 전반에 걸쳐 미디어 처리 및 전달을 위한 프로그래밍 기능을 활성화하기 위해 상당한 투자를 했습니다.

## oneAPI 동영상 프로세싱 라이브러리

인텔® oneAPI 비디오 처리 라이브러리(oneVPL)는 통합 및 개별 GPU에서 최적화된 미디어 트랜스코딩 성능을 제공합니다. oneVPL은 미디어 처리 및 전송, 방송, 스트리밍, 주문형 비디오(VoD), 클라우드 게임 및 원격 데스크톱 솔루션에 걸친 애플리케이션에서 비디오 디코딩, 인코딩 및 처리를 위한 비디오 중심 API를 제공합니다.

oneVPL에서 제공하는 로우레벨 인코더 및 속도 제어를 통해 개발자는 FFmpeg 및 GStreamer와 같은 표준 프레임워크를 넘어 사용 사례별로 고유한 밀도/성능 균형을 위해 인코더 구성을 미세 조정할 수 있습니다. 또한 자체 속도 제어를 구현하여 고객 도메인 전문 지식을 혁신적인 인텔 하드웨어와 결합할 수 있습니다. oneVPL은 인텔® 미디어 SDK 코어 API와 역호환(backward-compatible)됩니다.<sup>3</sup>

oneVPL은 개별적으로 무료로 다운로드할 수 있습니다. 또한 인텔 CPU 및 GPU에서 고성능 데이터 중심 응용 프로그램을 개발하기 위한 핵심 도구 및 라이브러리 세트인 인텔® oneAPI Base Toolkit에 포함되어 있습니다.

## 인텔® VTune™ 프로파일러로 응용 프로그램 성능 분석

GPU 코드에서 가장 시간이 많이 소요되는 부분을 식별하고 SYCL, OpenCL 코드, Microsoft DirectX 또는 OpenMP 로드 코드에 대한 데이터 전송 및 GPU 오프로드 스키마를 최적화하여 애플리케이션 컴퓨팅 집약적 작업을 가속화합니다. 마이크로 아키텍처 제약 또는 비효율적인 커널 알고리즘으로 인해 발생하는 성능 병목 현상에 대해 GPU 기반 코드를 분석합니다.

지원하는 운영 체제에 대해 자세히 알고 싶으면 [시스템 요구 사항](#)을 참조하십시오.



## 고효율 코덱

대규모 데이터 스토리지의 가격이 점차 낮아짐에도 불구하고 해당 데이터에 액세스하기 위한 대역폭 사용 비용은 여전히 비쌉니다. 향상된 압축을 통해 미디어 처리 및 전송 공급자는 이러한 대역폭 요구 사항을 줄여 운영 비용을 절감할 수 있습니다.

아마존, 시스코, 구글, 인텔, 마이크로소프트, 모질라, 넷플릭스가 설립한 산업 간 컨소시엄인 오픈 미디어를 위한 연합(Alliance for Open Media)는 2018년에 오픈 소스 AV1 코덱을 도입했습니다. 인텔 Flex 시리즈 GPU에 내장된 이 차세대 코덱은 최고의 성능을 제공합니다. 모든 대역폭에서 모든 최신 장치로 확장 가능한 고품질 실시간 비디오. 인터넷 스트리밍에 최적화된 낮은 계산 공간으로 상업용 또는 비상업용 사용자 생성 콘텐츠를 전달할 수 있습니다. 스트리밍 품질 저하 없이 30% 더 나은 압축으로 이 모든 작업을 수행하여 스트림당 비용을 줄입니다.

GPU는 AV1 외에도 기존 HEVC, AVC 및 VP9 코덱도 지원합니다. 공급자는 인텔이 오픈 미디어를 위한 연합과 협력하여 개발한 **SVT-AV1 소프트웨어 인코더**를 사용하여 시퀀스와 프로파일 콘텐츠가 가장 높은 채널의 품질을 극대화할 수 있습니다. 코어 라이브러리를 위한 오픈 소스 SVT(Scalable Video Technology) 프로젝트를 기반으로 하는 인코더는 인텔® 제온® 스케일러블 프로세서에 대해 고도로 최적화되어 인텔 데이터센터 GPU를 호스팅하는 동일한 서버에서 뛰어난 성능과 전력 효율성을 제공합니다.

이러한 코덱은 FFMEG 또는 Gstreamer와 같은 표준 프레임워크를 사용하거나 더 많은 제어 및 매개변수에 대한 추가 액세스를 제공하는 oneVPL을 사용하여 액세스할 수 있습니다. 하드웨어 및 소프트웨어 인코더는 모두 다양한 성능/품질 사전 설정을 제공하므로 공급자는 특정 사용 사례의 요구 사항에 따라 TCO 중심으로 조정할 수 있습니다.

### 솔루션 관련 성능 수치

**30%** 더 나은 압축을

Without degradation in streaming quality<sup>2</sup>

## 저 TCO로 고성능 구현

클라우드 게임 제공업체는 TCO를 최적화해야 하는 전략적 과제를 안고 있습니다. 인텔® 데이터센터 Flex 시리즈 GPU는 품질 저하 없이 서버당 지원할 수 있는 게임 인스턴스의 밀도를 높여 우수한 게임 플레이 경험에 대한 가입자 요구를 충족하면서 이러한 목표를 달성하도록 지원하므로, 공급자는 더 적은 수의 서버로 특정 가입자 기반을 처리할 수 있습니다. 카드당 최대 8개의 동시 4Kp60 스트림 또는 30개 이상의 1080p60 스트림을 지원합니다.<sup>2</sup> 인텔 데이터센터 Flex 시리즈 GPU는 서버당 많은 수의 스트림을 지원함으로써 공급자가 더 작은 데이터 센터 설치 공간으로 증가하는 가입자 기반을 처리할 수 있도록 하여 장비 및 시설 비용과 관련된 CapEx를 줄이는 데 도움이 됩니다. 와트당 고성능은 시스템의 이산화탄소 배출량을 줄여 기업의 친환경 이니셔티브를 지원하는 동시에 OpEx를 줄여 TCO를 더욱 낮추는 데 도움이 됩니다.

GPU는 x264 Medium에 비해 AV1을 사용하여 미디어 처리 및 전달 비용을 크게 줄일 수 있습니다. 이러한 절감 효과는 준비 및 배포 비용을 포함한 총 서비스 비용으로 확장됩니다.

서비스 제공자는 그래픽 처리 용량을 서버당 하나에서 여러 GPU로 확장하여 필요에 따라 CPU 및 GPU 리소스의 조합을 변경할 수 있습니다. 이 아키텍처의 유연성은 빠른 고품질 실시간 비디오 디코딩, 인코딩, 처리 및 미디어 형식 변환을 통해 서비스 제공업체의 진화하는 요구 사항을 충족하는 데 매우 적합합니다.

## 인텔 X<sup>e</sup> 아키텍처

인텔 X<sup>e</sup> 아키텍처를 기반으로 하는 인텔® 데이터센터 Flex 시리즈 GPU에는 최대 32개의 인텔 X<sup>e</sup> 코어 및 레이 트레이싱 장치, 최대 4개의 인텔 X<sup>e</sup> Media Engine, 인텔 X<sup>e</sup> Matrix Extensions(XMX) 및 하드웨어 기반 SR-IOV 가상화 지원 기능을 포함한 AI 가속화 기능이 들어 있습니다.

### 솔루션 관련 성능 수치

**8개** 4Kp60 스트림  
동시 지원

또는

**30개 이상** 1080p60 스트림  
동시 지원  
per PCIe card<sup>2</sup>



## 미디어 처리와 전송의 미래

미디어 콘텐츠 제공업체는 수익을 정확히 파악하면서 뛰어난 품질과 고객 경험을 제공하기 위해 끊임없이 기술을 적용하고 있습니다. 독점적이고 전문화된 기술을 개방적이고 유연한 표준 기반 기술로 대체하는 산업의 전환은 혁신이 진행됨에 따라 이러한 균형을 맞추는 데 크게 기여하고 있습니다. 인텔® 데이터센터 Flex 시리즈 GPU는 매끄러운 하드웨어 및 소프트웨어 미디어 처리 및 전달 솔루션으로 이러한 전환에 기여하여 제한적인 독점 환경으로부터의 그래픽 프로그래밍 제약을 풀어줍니다.

## 인텔® 데이터센터 GPU Flex 시리즈에 대한 더 자세한 내용은 아래의 사이트에서 자세히 알아보십시오.

[https://www.intel.co.kr/content/www/kr/ko/products/docs/discrete-gpus/  
data-center-gpu/flex-series/overview.html](https://www.intel.co.kr/content/www/kr/ko/products/docs/discrete-gpus/data-center-gpu/flex-series/overview.html)



1 Cisco Systems, “VNI 완전 시장 전망 하이라이트.” [https://www.cisco.com/c/dam/m/en\\_us/solutions/service-provider/vni-forecast-highlights/pdf/Global\\_2021\\_Forecast\\_Highlights.pdf](https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2021_Forecast_Highlights.pdf).

2 성능은 사용, 구성 및 기타 요인에 따라 다릅니다. 더 자세한 내용은 성능 지수 사이트(<https://www.intel.com/PerformanceIndex>)에서 알아보십시오.

3 사소한 예외가 적용됩니다. 인텔의 “인텔® Media SDK에서 Intel® oneAPI 비디오 처리 라이브러리로 업그레이드: 전환 안내서”, <https://www.intel.com/content/www/us/en/develop/documentation/upgrading-from-msdk-to-onevpl/top/developer-details/removed-features-details-and-mitigations.html>

기반으로 하며 공개적으로 사용 가능한 모든 업데이트를 반영하지 않을 수 있습니다. 구성 세부사항은 구성 공개를 참조하십시오.

어떤 제품이나 구성 요소도 절대적으로 안전할 수 없습니다.

인텔은 제3자 데이터를 통제하거나 감사하지 않습니다. 정확성을 평가하려면 다른 출처를 참조해야 합니다.

비용과 결과는 다를 수 있습니다.

인텔 기술을 사용하려면 활성화된 하드웨어, 소프트웨어 또는 서비스 활성화가 필요할 수 있습니다.

귀하는 여기에 설명된 인텔 제품과 관련된 침해 또는 기타 법적 분석과 관련하여 이 문서를 사용하거나 사용을 촉진할 수 없습니다. 귀하는 여기에 공개된 주제를 포함하여 이후

초안이 작성된 모든 특허 청구에 대해 인텔에 비독점적 로열티 프리 라이선스를 부여하는 데 동의합니다.

여기에 공개된 주제를 포함하여 이후 작성된 모든 특허 청구에 대한 비독점적 로열티 프리 라이선스. 설명된 제품에는 제품이 게시된 사양에서 벗어나게 할 수 있는 정오표로 알려진

설계 결함 또는 오류가 포함될 수 있습니다. 현재 특정화된 정오표는 요청시 사용할 수 있습니다.

©인텔 사, 인텔, 인텔 로고 및 기타 인텔 마크는 인텔사 또는 그 자회사의 상표입니다. 다른 이름과 브랜드는 다른 사람의 자산으로 주장될 수 있습니다.

0822/MH/MESH/349354-001US