WHITE PAPER

Media and Communications
Content Creation and Distribution

Rethinking Visual Cloud Workload Distribution

Creating a New Model

With visual computing workloads growing at an accelerating pace, cloud service providers (CSPs), communications service providers (CoSPs), and enterprises are rethinking the physical and virtual distribution of compute resources to more effectively balance cost and deployment efficiency while achieving exceptional performance.

Visual cloud deployments accommodate a diverse range of streaming workloads, encompassing media processing and delivery, cloud graphics, cloud gaming, media analytics, and immersive media.

Contending with the onslaught of new visual workloads will require more nimble, scalable, virtualized infrastructures; the capability of shifting workloads to the network edge when appropriate; and a collection of tools, software, and hardware components to support individual use cases fluidly. Advanced network technologies and cloud architectures are essential for agile distribution of visual cloud workloads.

A 2017 report, *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021*, projected strong growth in all Internet and managed IP video-related sectors. Compound annual growth rate (CAGR) figures during this time span, calculated in petabytes per month, included these predictions:

- Content delivery network (CDN) traffic: 44 percent increase globally
- Consumer-managed IP video traffic: 19,619 petabytes per month (14 percent increase) by 2021
- Consumer Internet video: 27 percent increase for fixed, 55 percent increase for mobile

The impact of this media growth on cloud-based data centers will produce a burden on those CSPs, CoSPs, and enterprises that are not equipped to deal with large-scale media workloads dynamically. Solutions to this challenge include:

- **Increasing flexibility and optimizing processing**: Virtualization and software-defined infrastructure (SDI) make it easier to balance workloads on available resources. Open platforms and widely available acceleration technologies give CSPs and CoSPs a flexible way to optimize processing of diverse workloads. Learn more in the section titled *Balancing Dynamic Workloads in a Cloud-Based Data Center*.
- **Scaling compute, storage, and network resources**: Dynamic elasticity is a major advantage when contending with visual cloud workloads. Adopting a modern cloud infrastructure powered by a new generation of scalable processors increases resource availability substantially. Pervasive high-bandwidth and low-latency networks will also play a vital role for delivering rich user experiences and ensuring business model scalability. Learn more in the section titled *TCO Considerations*.
- **Enhancing development processes**: Software development tools, media-specific toolkits, and highly optimized software codecs support rapid application development and efficient media delivery. Learn more in the section titled *Open Source Software for the Visual Cloud*.

Table of Contents

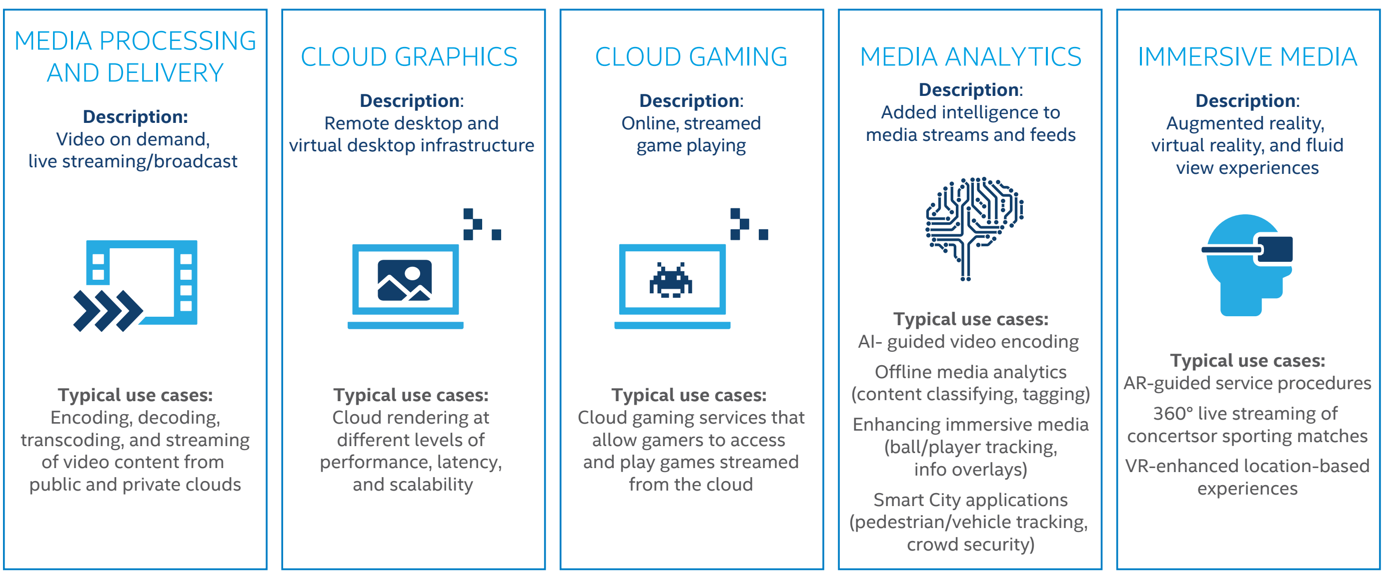
- **Deploying purpose-built solutions for select use cases and edge computing**: Specialized hardware—including discrete graphics processing units (GPUs), integrated GPUs, field programmable gate arrays (FPGAs), video processing units (VPUs), and image-processing units (IPUs)—can boost performance for select applications in which a single, targeted workload must be handled on a large-scale basis. Learn more in the section titled *Open Source Software for the Visual Cloud*.
- **Implementing modern cloud architectures**: For performing operations on large, complex data files and delivering processing power to efficiently handle inference for AI solutions (as well as many other intensive computing operations), modern cloud architectures hold the key to effective workload distribution.

What is the Visual Cloud?

Visual cloud computing consists of a set of capabilities for remotely consuming content and services that center around efficient delivery of visual experiences from the cloud—both live and file-based. Additionally, visual cloud includes media analytics applications that add intelligence to video content. As shown in the figure below, the visual cloud has five major workloads, each related to a set of media use cases.

Visual Cloud Workloads

All workloads require high performance, high scalability, and full hardware virtualization



Adapting to the Evolving Media Landscape

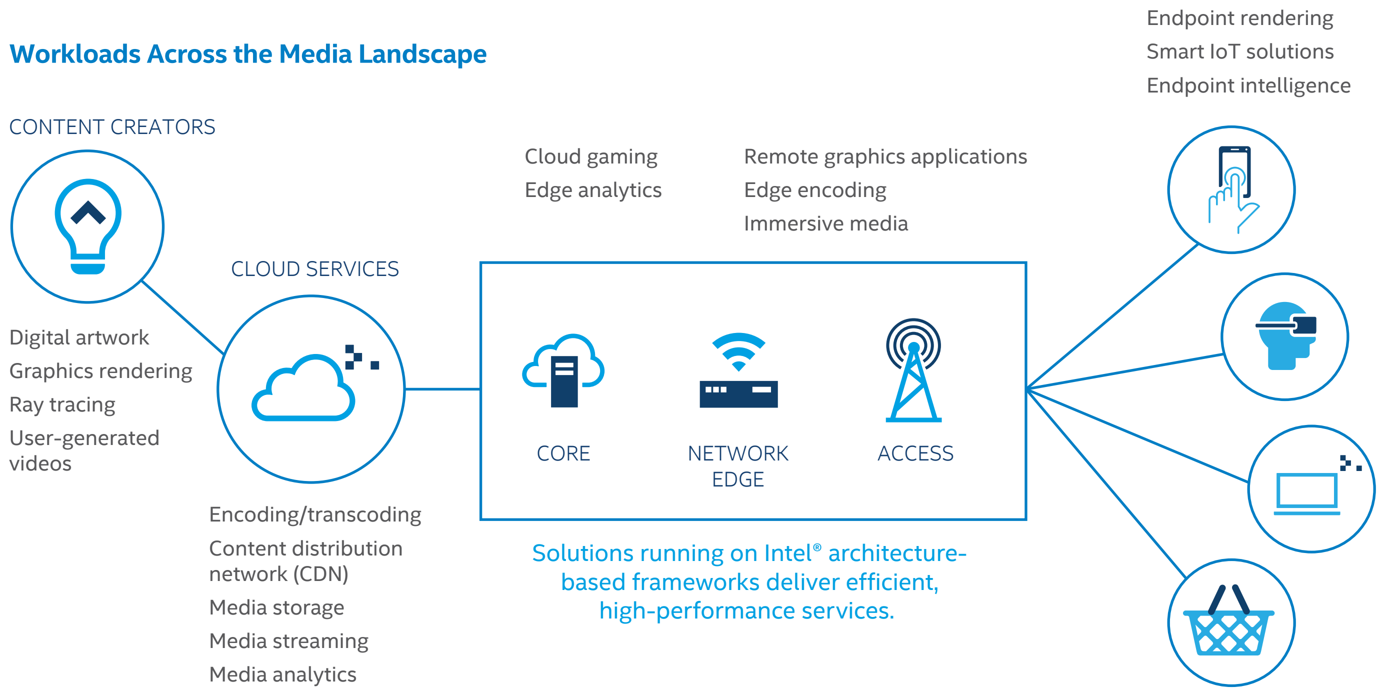
The new media space goes beyond basic consumption models. It includes immersive experiences, in which users may engage in a 360-degree virtual reality (VR) experience. It includes delivering more intelligence with the video that is being distributed, incorporating analytics as a part of the media workflow. Analytics, performed on the contents of video sent to the end user, makes it possible to deliver additional services or value. Media content as a part of Internet of Things (IoT) solutions is growing in importance. Sometimes visual workloads will need to be handled at the network edge, to reduce latency and provide near real-time responsiveness. Other times dense visual workloads will be best processed in the centralized data center using the most advanced processors and other scalable technologies.

Certain kinds of services—especially those subject to sudden surges in traffic—create challenges in building a supporting infrastructure that can adapt to quickly changing conditions. For example, NFL game views can skyrocket from hundreds to thousands to millions in a matter of several minutes—fueled by social media—and drop back down again just as quickly. A virtualized, software-based infrastructure provides many benefits for service providers delivering new media content in many kinds of use cases growing in popularity, such as:

- User-generated live streaming (for example, Facebook Live*).
- Over-the-top linear streaming (Yahoo or Twitter* streaming of NFL games).
- Cloud-gaming applications (such as Shadow* or LiquidSky*), which can create substantial compute demands when large numbers of users are participating simultaneously.
- Rising video content generated through social media applications (for example, Facebook and Snapchat* typically handle eight billion average daily video views).

All of this is part of new media. The added value for service providers is the intelligence that can be delivered with the video content. It is more about the context of use, the type of content, and the additional value that can come from the consumer, as well as the service provider looking to monetize and deliver new services.

Workloads Across the Media Landscape



Balancing Dynamic Workloads in a Cloud-Based Data Center

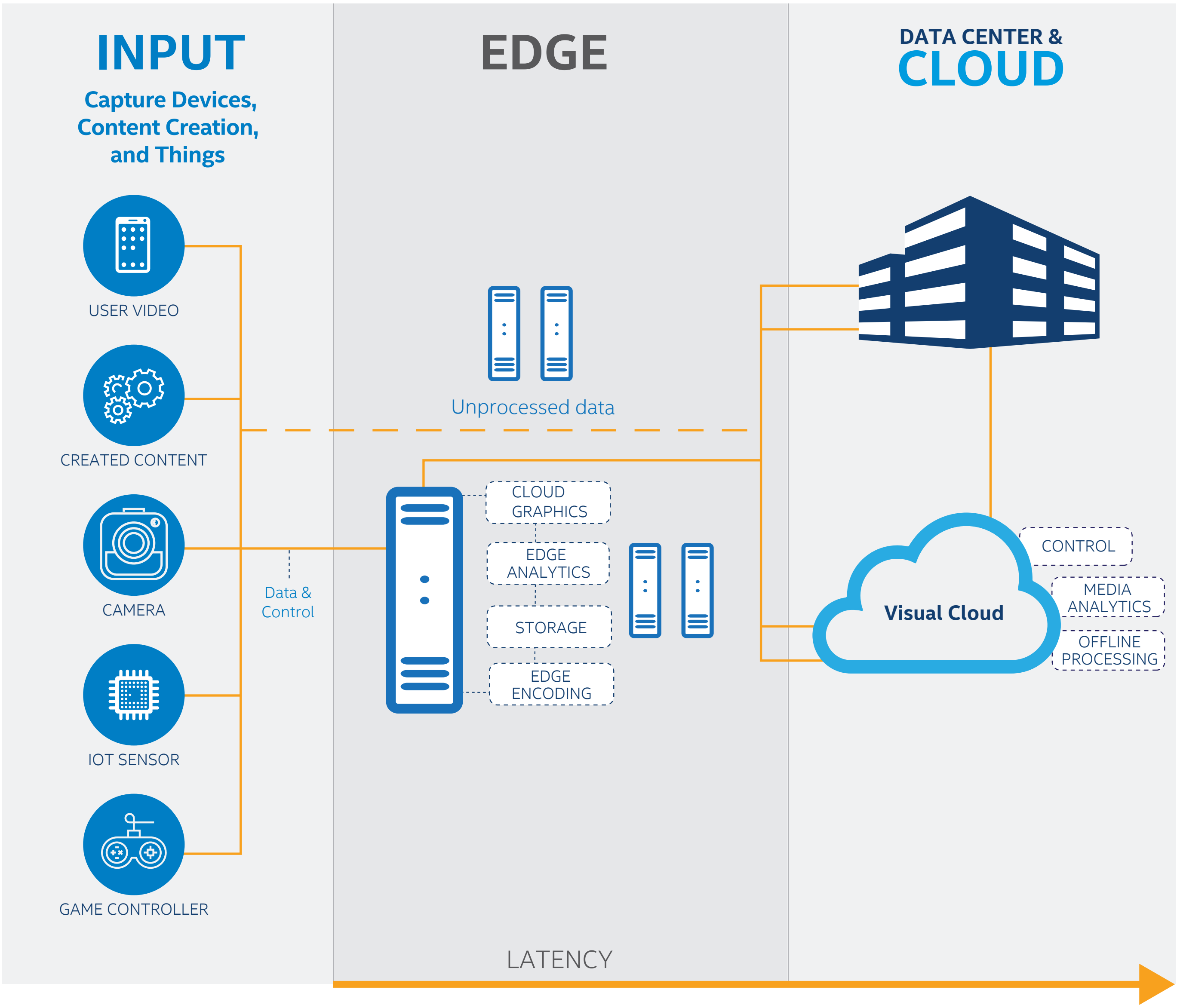
Significant increases in new media processing and delivery workloads in the data center call for rethinking the approaches and technologies for handling current and future processing requirements.

> The trends and projections are bullish as new media use cases are developed and deployed. The Cisco Visual Networking Index (VNI) report for 2017 forecasts that, by 2021, Ultra HD will represent 20.7 percent of IP video traffic. The same report projects a seven times increase in Internet video surveillance traffic between 2016 and 2021. According to the World Economic Forum, immersive media, including augmented reality (AR) and VR, will grow to a USD 95 billion market by 2025, spurred by live events, gaming, and video entertainment.

The capabilities of virtualization allow diverse workloads to access and share a common collection of compute, storage, and network resources. In environments that are demand driven and that frequently experience processor-intensive workloads, virtualized hardware can scale freely and relocate workloads as required across infrastructures. This has spawned the massive increases of efficiency that customers have found in the cloud.

Cloud-based data centers built using modern cloud infrastructures can capitalize on network function virtualization (NFV), software-defined networking (SDN) and SDI to support functions historically delivered by dedicated network appliances. These new technologies adapt well to the demands of shifting media workloads. As visual computing use cases evolve and mature, enterprises, CSPs, and CoSPs—relying on Intel® architecture-based infrastructures—will have a wide range of choices as to where and how to process workloads. This may involve using the network edge, the cloud, or the core data center, depending on the scenario, data sizes involved, security issues, and workload distribution. The broad choice of tools and technologies available from Intel to efficiently handle the workloads—from pure software solutions to those based on hardware acceleration—also give solution architects the freedom to design, develop, and deploy high-performance visual solutions in the most cost-efficient and scalable way.

Span of The Visual Cloud



TCO Considerations

Throughout the ecosystem, running visual compute workloads on pure software architectures gives customers the flexibility of making improvements to media processing and delivery as well as customizations to the visual quality.

To maximize the business value of infrastructure investments, CoSPs, CSPs, and enterprises need servers flexible enough to accommodate a diverse range of both compute and visual workloads.

Several key factors affect the total cost of ownership (TCO) of a media-processing system. Factors to consider include:

- Scalability of server infrastructure to address all workloads for maximizing IT investments.
- Optimized software features that enhance the value of data center and cloud resources.
- Hardware platform characteristics with respect to cost, cooling, overall utility, deployment complexity, and maintenance.
- GPU license fees by other vendors have an annual cost of licensing.
- Server density considerations and overall resource utilization in the data center, factoring in virtualization capabilities and workload balancing tools.
- The degree of flexibility and agility for managing the balance between video quality and the streaming bit rate, as well as availability of customizations and ongoing performance enhancements.

- Tradeoffs between investing in an open standards hardware platform with widely available open source software components and a proprietary hardware platform subject to vendor lock-in and a greatly reduced range of supported applications.
- Depth and breadth of the ecosystem supporting the hardware/software platform and whether it is adequate to provide reasonable choices and fulfill business objectives.

There is tremendous potential to leverage existing infrastructures, and the majority of the world's leading data centers are populated with Intel® Xeon® processor family servers. Infrastructure managers can take advantage of low-demand use periods to handle media-processing applications that have been optimized and tuned for these servers. This—in addition to leveraging open source and off-the-shelf packaged open source-based projects—helps minimize the need to acquire additional hardware.

Meeting Service-Level Expectations

accomplished by increasing the workload density within the data center. This, in turn, can help lower the TCO for the CSP. By upgrading to Intel® Xeon® Scalable processors, a CSP can substantially increase the density of virtual machines in the data center, letting them support more customers per rack.[1]

Intel Xeon Scalable processors—available with as many as 28 cores—deliver exceptional performance and provide fluid scalability for dense media workloads. To improve media operations, Intel® Advanced Vector Extensions 512 (Intel® AVX-512) feature double the number of floating point operations per second (FLOPS) per clock cycle, a significant leap forward compared to the past-generation capability. Demanding tasks that are vital to a new media environment gain performance boosts, including media analytics, video data encoding and decoding, digital content creation, 3D modeling and simulation, and visualization.

For cost-effective processing, products such as the Intel® Xeon® D processor series benefit from software optimizations that bring new performance capabilities for handling media streams. For example, the Intel Xeon D processor now supports transcoding of a single full 4Kp60 10-bit HDR stream or three 1080p60 10-bit HDR streams in a single socket, thanks to optimized software. Wide selections available in the Intel Xeon processor family can help resolve the challenge of balancing performance and cost for any given deployment.

A Software-Based Approach to Media Processing

A software-based approach to media processing offers flexibility and the opportunity for new revenue channels. Without having to rework the physical architecture or update the hardware, a software-based model—running code on a CPU—provides these advantages:

- Supporting new codecs that emerge, through simple software updates rather than expensive hardware changes.
- Adding intelligence to video segments through analytics software that can be rapidly iterated on with new algorithms.
- Adding efficiency to encoding operations through artificial intelligence, enabled by software.
- Advancing video compression to new levels, through software enhancements.
- Balancing video quality with bandwidth, mediated by software.

Ninety-five percent of the server processors powering the cloud are based on Intel® technology.[2] The vast supporting ecosystem and deep expertise surrounding open platforms based on Intel architecture ensures a reliable, long-term foundation for launching media applications and services. It also provides a framework in which software updates are delivered frequently, through open source contributions as well as media toolkits that are regularly updated, adding to the long-term business value.

Open Source Software for the Visual Cloud

CSPs, CoSPs, and enterprises increasingly rely on open source solutions and components to drive innovation, reduce time-to-market, lower costs, and minimize obstacles when commercializing products.

To help strengthen the ecosystem and provide ready access to the tools and technologies for building effective solutions, Intel is working within the open source community to standardize and extend capabilities. This includes releasing many tools, libraries, and previously commercial products to help build visual cloud solutions using the open source ecosystem. The focus is on accelerating availability of open source, validated building blocks—across encode, decode, inference, and rendering core building blocks—that support visual cloud workloads. The goal is to minimize barriers and provide clear channels for creating, storing, distributing, and monetizing new media solutions deployed through cloud-based servers and edge servers.

Among the top contributions to the visual cloud ecosystem, Intel plans to release two new codec cores to open source: Scalable Video Technology for HEVC (SVT-HEVC), available Q3'2018, and Scalable Video Technology for AOMedia Video 1 (SVT-AV1), available 1H'2019.

- **SVT-HEVC**: An extremely scalable codec core with great visual quality/density tradeoffs and released to open source, this advanced, ultra-high-definition encoder is being provided under permissive licensing.
- **SVT-AV1**: This codec core is an alternative implementation of the open source AV1 standard, providing interoperability with the latest Intel processors and based on the same architecture as SVT-HEVC, with similar scalability and attention to visual quality/density tradeoffs. The royalty-free AV1 codec delivers 4K ultra-high-definition video with provisions for supporting even higher definition video content as new formats emerge.

In support of the visual cloud, Intel helps ensure that the software tools are optimized on an ongoing basis for the latest hardware and processor advances. Working collaboratively with industry leaders in the visual cloud sector, Intel is committed to helping build a strong, thriving ecosystem supported by enterprise-ready open source software solutions and advanced tools to enhance development processes.

Intel frameworks, including the Intel® Media SDK and the OpenVINO™ toolkit extend capabilities for developers working on accelerating media and media analytics applications.

- The Intel® Media SDK for Linux* delivers tools for developing media and video applications, including an application programming interface (API) for accessing Intel® Quick Sync Video capabilities. This includes hardware-accelerated encoding, decoding, and pre-processing operations. Used in combination with Intel® Media Server Studio, developers can balance visual quality against performance to optimize streaming content for targeted applications. Support is included for H.265 (HEVC), H.264 (AVC), MPEG-2, VP9, VC-1, and MJPEG. Built-in support for a number of video operations includes resize, scale, deinterlace, color conversion, composition, denoise, sharpen, and more. This comprehensive media tool suite streamlines the process of creating innovative, optimized media processing and delivery solutions, cloud graphics applications, and other types of enterprise-caliber video solutions.
- The OpenVINO toolkit provides deep learning and traditional computer vision analytics capabilities. Components of the OpenVINO toolkit - Intel® Deep Learning Deployment Toolkit support for Intel CPU and GPU (Intel® Processor Graphics) and heterogeneous plugins, and Intel's Open Model Zoo pre-trained models and samples - will be available in open source Q3'18. The toolkit is optimized for Intel architecture-based platforms, and enables developers to extract intelligence from video streams and create new use cases. To empower the visual cloud community and help accelerate solution development, Intel will continue to deliver reference code, developer training, and community support. The OpenVINO toolkit enhances the development of media analytics applications, using a write once, deploy everywhere model aligned with Intel architecture-based platforms, including pure software and several hardware acceleration options. The acronym stands for Open Visual Inferencing and Neural Network Optimization, and the toolkit helps transform vision data into business insights using analytics.

Additional Development Tools for Building Visual Cloud Solutions

Familiar Intel® development tools include the Intel® C++ Compiler, Intel® VTune™ Amplifier, and Intel® Code Builder for OpenCL™ API, and provide the tuning and optimization capabilities for ensuring efficient performance of media processing solutions on Intel architecture-based platforms.

In addition to the development tools mentioned previously, these additional tools help build visual cloud solutions:

- Intel® Rendering Framework–A collection of open source ray tracing and rasterization libraries that includes:
  - Embree
  - OSPRray
  - OpenSWR
  - Open Image Denoise
- Intel® Collaboration Suite for WebRTC–Hardware acceleration of real-time communications, including video conferencing, scheduled for release to open source.

Other components for enabling visual cloud solutions include:

- Special-purpose SDKs, such as the Intel® Network Edge Virtualization SDK that provides an NFV platform for mobile-edge computing applications and services, streamline solution development.
- Commercial and open source hypervisors, container technologies, and the related SDN orchestration (including OpenStack* and Kubernetes*) and performance optimizations (Open vSwitch*) provide diverse, efficient virtualization and deployment options.

Through visual cloud software running on an x86-based infrastructure, the development community has a platform from which to launch comprehensive end-to-end media processing solutions, with powerfully optimized building blocks and the flexibility of open source enhancements to support the latest technology advances.

Media Processing and Delivery Use Case

Cloud-based media processing and delivery demands in response to escalating video traffic across wired and wireless channels have become a vitally important service capability for CSPs and CoSPs.

The predominant use cases involving media delivery and processing, encompassing both content creation and streaming, fit into two basic categories:

- **Broadcast media**—Communications media providers, such as Comcast and DirecTV (AT&T), produce and deliver both live and streaming video content to audiences—at least partially using their own CoSP network infrastructure.
- **Over-the-top (OTT) media**—OTT content is captured, produced, and delivered over the Internet, either through mobile services or wired connections typically provided by a CoSP. This category includes video on demand content as well as live streaming video from companies such as Apple TV, Amazon, and Netflix, and web entities such as Facebook Live and YouTube*.

Media Workloads

BROADCASTING - LIVE AND VOD	OVER THE TOP - VOD	OVER THE TOP - LIVE
Communications media providers (Comcast, DirectTV, AT&T)	Cloud media providers (Amazon, Netflix, Hulu)	Live streaming video (Twitch*, YouTube*, Facebook Live*)
Live TV streaming	Video on Demand content	User-generated content
Video on Demand content		

WORKFLOW SEQUENCE: CONTENT CAPTURE AND PRODUCTION → CONTENT STORAGE AND DISTRIBUTION → CONTENT CONSUMPTION

Enabling Hardware and Software

The latest lineup of advanced Intel Xeon processor family and Intel® graphics products provide an agile hardware foundation for delivering visual cloud services. In combination with the enabling visual cloud software, providers can deliver media services efficiently and handle a diverse range of media content.

The release of SVT-HEVC to open source, the first in a line of standards-based encoder libraries designed to scale for demanding media environments, empowers the developer community, CSPs, and CoSPs with a powerful means for processing and delivering rich media.

Also on a near-term release track, SVT-AV1—adapted from AOMedia Video 1 to handle high-quality video transmissions over the Internet, including 4K video streams and higher—offers an open-standards, royalty-free video coding format for the visual cloud community. This codec will help serve increasing user demand for high-quality video delivery, as high-resolution mobile devices and VR-enabled devices gain wider acceptance.

Examples of Media Processing and Delivery Use Cases

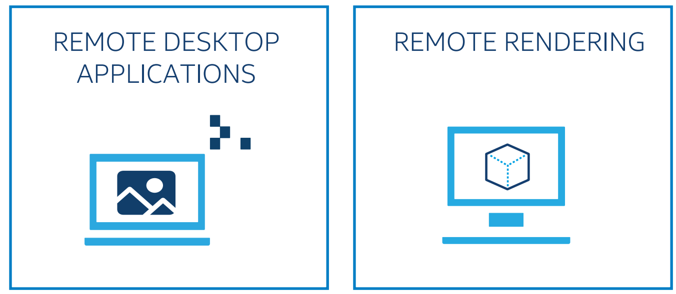
Among the most noteworthy use cases involving media processing and delivery are:

- **Transcoding**—A large part of handling media workloads involves converting video formats, adjusting bit rates, and controlling compression intelligently, both in real time and offline. Intel AVX-512, part of the Intel Xeon Scalable processor platform, makes it possible to accelerate media workloads, supporting ultra-wide 512-bit vector operations and handling 32 double-precision and 64 single-precision floating point operations per second, per clock cycle.
- **Distribution of workloads**—The nature of CPU-centric infrastructures makes it easy to perform cloud load balancing of media workloads. The additional flexibility of Intel based infrastructures employing virtualization technology, based on open standards and representing the clear majority of cloud service installations, gives CSPs and CoSPs more precise management control and agility in processing workloads dynamically.

Cloud Graphics Use Case

By being able to tap into powerful cloud-based servers to perform graphics-intensive operations, organizations can selectively choose the level of performance, latency, and scalability that best fits their business needs. Using new technologies, such as cloud-hosted workstations powered by Intel® Xeon® processor E3-1500 v5 with Iris® Pro graphics, enterprises can equip technical staff members to work collaboratively and access centralized graphics applications remotely from wherever they want to work. With data and applications accessible securely from the data center, complex renderings and visualizations can be handled by high-performance servers in a virtualized environment in which workload processing can be adjusted dynamically to meet requirements. Remote rendering of 3D animations or 3D models represents another promising use of cloud graphics. Autodesk currently uses a software component, Embree, (that was developed by Intel and released to open source) as a rendering engine in several of their top graphics applications. Embree takes full advantage of all available CPU cores for interactive rendering to address large models using full system memory.

Cloud Graphics Use Cases



Enabling Hardware and Software

Several Intel technologies contribute to efficient cloud graphics implementations. Intel® Processor Graphics technology accelerates video processing and remote graphics workloads. Intel Quick Sync Video, a dedicated hardware-based accelerator, streamlines video transcoding operations for moving media assets through the workflow. Intel AVX-512, part of the Intel Xeon Scalable platform, makes it possible to accelerate vector- and integer-based imaging, video editing, simulations, and modeling operations. Intel® Graphics Virtualization Technology gives virtual machines access to available GPUs and integrated video transcode accelerators, enabling workstation remoting and media streaming.

Examples of Cloud Graphics Use Cases

Important cloud graphics use cases include:

- **Remote desktop applications**—Desktop and workstation remoting solutions, powered by the Intel Xeon processor E3-1500 v5 family with Iris Pro graphics, deliver professional graphics capabilities to geographically diverse technical teams, adding to productivity, improving collaboration, and minimizing the risk of lost or stolen intellectual property.
- **Remote rendering**—The intensive compute cycles required to process complex animations, scientific data visualizations, ray-traced content creation, and 3D models lends itself well to remote rendering in the cloud where virtualization can be used to dynamically provision compute, storage, and network resources to handle demanding media creations.

Cloud Gaming Use Case

In much the same way as movie distribution evolved from DVDs to online streaming as Internet connectivity speeds increased, online game services have begun to take advantage of faster connection rates and improved compression techniques. Video streaming lets users stream games in real time from remote servers without requiring that the game be downloaded first. A significant benefit of cloud gaming is that it typically employs an any pane of glass model, making it possible for gamers to play their favorite games on whatever device they prefer (laptop, desktop, tablet, or smartphone) with the performance they expect.

According to Knowledge Sourcing Intelligence, the cloud gaming market is projected to grow at a CAGR of 26.12 percent between 2018 and 2023, reaching a total market size of USD 4,284.002 million by 2023 (up from USD 1,064.500 million in 2017).[3]

Growth is being accelerated by technologies that make cloud gaming possible, including:

- High-speed, low-latency networks, including 5G and fiber-to-the-home (FTTH).
- Cloud gaming software with AAA licensed content (such as delivered by Gamestream, PlayGiga and other services).
- High performance, cost-effective graphics processors, such as the 8th Gen Intel® Core™ processors with Radeon* RX Vega M Graphics.
- Versatile and dense compute and media platforms for building cost-effective cloud gaming solutions, including the Artesyn MaxCore* platform.
- Edge computing solutions that bring games closer to the user, reducing latency.

Enabling Hardware and Software

Hardware components to support cloud gaming services extend from the client system through to the data center, encompassing edge devices, base stations, network core components, and other equipment, all of which can be based on Intel compute, network, and storage products.

Intel combined a compute processor and discrete graphics processor in a single package—the 8th Generation Intel Core processor with Radeon RX Vega M Graphics—to deliver exceptionally smooth frame rates for high-end games. Through Intel® Dynamic Tuning, power is dynamically allocated between the compute and discrete graphics processor to support enthusiast gaming, VR, and AR applications. Performance can be boosted further with selective overclocking of the CPU, GPU, and high-bandwidth memory gen 2 (HBM2) and Intel® Turbo Boost Technology 2.0. Two levels of performance allow users to switch between advanced content creation and entry-level gaming to high-performance gaming in high-definition resolutions.

The design integrates the Embedded Multi-die Interconnect Bridge (EMIB), which serves as an intelligent information bridge connecting the discrete graphics processor and high-bandwidth memory. The result is a hardware platform that can run up to six 720p games at 30 frames per second when set at a medium level. The interconnection accelerates graphics operations, including rich content creation, 4K video editing, immersive VR experiences, and fluid, high-resolution game play.

Advanced encoders, including SVT-HEVC and SVT-AV1, make it possible to stream high-definition and ultra high-definition video files efficiently to support the latest game applications and deliver rewarding user experiences. Latency—the lag time between an action and response in multiplayer gaming—generally needs to be under 100 milliseconds for a satisfying gamer experience. With the latest encoding, compute, networking, and graphics processing capabilities at the edge, latency will be reduced even further for an even better user experience.

Examples of Cloud Gaming Benefits

- **Online gaming services revenues**—The promising growth of online games, served from the cloud, represents additional revenue for CSPs and a complement to current use models for gaming. Media technology can be used to stream the game content to the end user, applying media encoding or the media pipeline processing part of the stack to deliver the game content efficiently. Visual cloud software stack components can be used for both creation and delivery of the media.
- **Any time, any place access to games**—a unique characteristic of online gaming is that the delivery model from the cloud makes it possible for gamers to access and play games from their choice of devices—from smartphones to tablets to desktop machines—wherever they happen to be. This is sometimes called the *any pane of glass* approach, signifying that because the compute operations are mainly taking place in the cloud, users can enjoy good performance on any device across the board.

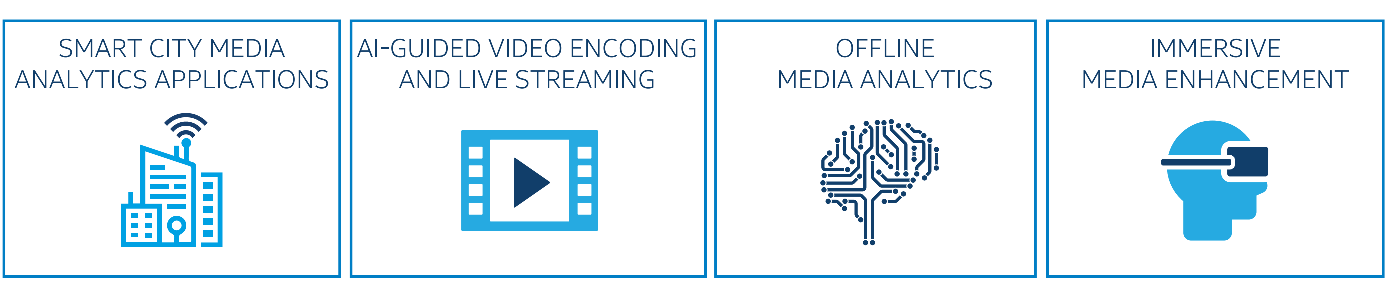
Media Analytics Use Case

Media analytics performed on live media or distributed video streams can help service providers, content aggregators, and content delivery networks better understand the nature of the visual content and derive useful intelligence from it. By incorporating technologies from the Intel® AI portfolio and leveraging IoT technologies, sophisticated media analytics applications can do everything from detecting suspicious intruders in a video surveillance feed to surveying the traffic patterns in a smart city to better control flow. With billions of pieces of visual content exchanged daily, the market opportunities are expansive for creating new, useful services and adding features and capabilities to existing services.

By 2022, MarketsandMarkets estimates that the video analytics market will reach USD 11.17 billion.[4] According to Tractica, deep learning revenues will increase from USD 655 million in 2016 to USD 35 billion by 2025.[5]

In collaboration with industry leaders and the open source community, Intel is working to develop solutions based on the Intel Xeon Scalable processor family, using the latest AI training and scoring techniques to extract understanding from media content. The latest technologies being employed in this sector are helping to generate performance boosts and simplifying deployment as enterprises and service providers become early adopters.

Media Analytics Use Cases



Enabling Hardware and Software

The OpenVINO toolkit streamlines the development of computer vision solutions that include deep learning inference capabilities. There is both an Intel version at software.intel.com/openvino-toolkit available now, and select components —Intel® Deep Learning Deployment Toolkit support for Intel CPU and GPU (Intel® Processor Graphics) and heterogeneous plugins, and Intel's Open Model Zoo pre-trained models and samples—that will be open sourced from Intel Q3'18. This toolkit gives developers a way to *develop once and deploy anywhere*, providing compatibility with a broad range of standards-based AI frameworks, including TensorFlow*, MXNet*, and Caffe*, and across a choice of Intel hardware. AI model training can be accomplished on one Intel architecture-based system and then the solution can be deployed, without further training, to a wide range of modern Intel based systems. This is a strong differentiation from certain GPU-based systems that require AI development to be conducted on the same target system to which the solution will be deployed. The OpenVINO toolkit includes components so that developers can easily deploy solutions from the network edge to the cloud, maintaining end-to-end intelligence and visibility across a diverse set of products.

Examples of Media Analytics Use Cases

Key media analytics use cases include:

- **Smart City***—Media analytics applications within the Smart City range from identifying vehicles, including their license plates, for traffic monitoring and toll collection to performing facial recognition of individuals near buildings or public spaces, for both security reasons and for crowd analysis. Often visual recognition tasks need to be performed in real time, working most efficiently in edge computing implementations.
- **AI-guided video encoding and live streaming**—By adding AI capabilities to video encoding algorithms, broadcasters and OTT producers can control quality and bandwidth parameters to meet specified goals. An AI algorithm at the heart of this process can be trained for optimal quality or bandwidth from a collection of video streams, and then automatically adjust the encoder to ensure goals are achieved.
- **Offline media analytics**—By using object and performer detection and classification, AI applications can provide a wide variety of services to help CSPs and CoSPs create new revenue channels. For example, AI techniques can identify and tag potentially violent movies, movies that include certain performers, or provide metadata to recommendation services based on movie content. These kinds of services can be applied at different points in the delivery system, in real time or offline, to add intelligence to the packaging and distribution of content.
- **Immersive media enhancement**—Applying analytics to immersive media applications opens opportunities in both 360-degree content display and AR applications. For example, sports presentations can benefit from tracking players or the ball, using AI to control the encoding for providing a selective view of the action. Interactive applications include being able to automatically overlay video content with relevant information (such as a player's batting average or basketball foul shot percentage), or an international football team's ball possession or passing accuracy statistics.

Immersive Media Use Case

Innovation in AR and VR solutions is changing the way that human beings interact with the world. These technologies are also inspiring innovators to introduce new products, services, and business models into the visual computing market, creating opportunities for CSPs, CoSPs, and broadcast companies.

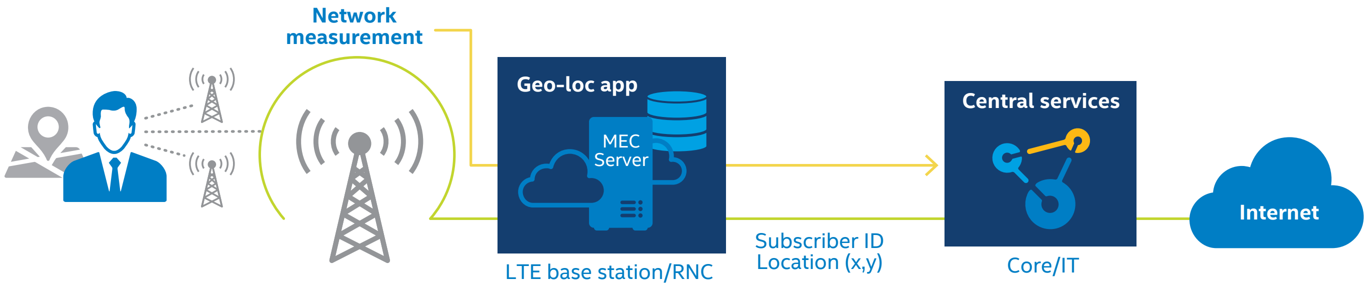
Potential revenues, as projected by International Data Corporation, show strong growth opportunities, with USD 45 billion in sales of headset hardware forecast by 2021. Software sales in this market also look to be substantial, expected to exceed USD 35 billion by 2025, according to Statistica.[6]

Advanced immersive media applications are just beginning to gain traction in the industry with the more practical uses leaning toward AR, such as guiding worker tasks through smart glasses. Consumer-driven interests and entertainment tend to be aligned toward VR. On the broadcast side, 3D renderings of stadiums and other types of facilities can be integrated into media presentations, whether live or on demand. Streaming of live concert experiences or sports using 360-degree video content is another application that is already being implemented by companies, including Rivet Media and Wowza Media Systems. Recent advances in moving media workloads to the network edge are providing for more quality experiences by addressing the bandwidth concerns and latency issues that have been a challenge for VR scenarios.

Enabling Hardware and Software

Intel Quick Sync Video enables many of the immersive scenarios, such as 360-degree video, by using Intel® Graphics Technology for rapid media processing, encoding, and decoding at accelerated rates without burdening the CPU. Another useful component is ProMedia* Xpress, a tool from Harmonic that delivers exceptional transcoding performance without the need for dedicated hardware accelerators. Employing Intel® multi-core processor architectures, it is a valuable resource for producing 360-degree VR video. Intel Media Server Studio optimizes enterprise-caliber video delivery and supports cloud graphics applications. Embree, a collection of high-performance ray tracing kernels originally developed by Intel and released to open source, will become increasingly valuable as immersive applications mature, increasing the demand for photorealistic rendering. As model sizes increase dramatically, immersive applications can benefit from CPU-based rendering that gives access to full system memory.

Location-Based Services



- Active device location is provided based on cell locations using standard multi-access edge computing APIs in real time
- Locates specific users and understands how the crowd is distributed
- Applicable to smart city, stadium, geo-fencing, retail, public safety, and advertising

Examples of Immersive Media Use Cases

The following immersive media use cases are among the more promising in this sector:

- **360-degree live streaming**: Companies such as Rivet and Wowza Media Systems are making it possible to live-stream events, such as concerts or sporting matches, as well as VR content using widely available off-the-shelf devices.
- **AR-guided service procedures**: AR offers a strong potential for guiding factory workers, warehouse employees, and technicians in daily procedures, equipment operation, and maintenance tasks, using instructions overlaid on AR head-mounted displays or smart glasses. Guidance can be delivered by video, audio, or as text lists of steps in a procedure. More sophisticated applications can provide immediate feedback by observing and commenting on worker actions.
- **VR enhanced location-based experiences**: On-premises immersive experiences are being used broadly in a number of consumer-focused ways. The number of VR arcades has been growing globally, as have the number of businesses augmenting tourist destination experiences.

Get Ready for Visual Cloud Opportunities

Visual cloud solutions—including workloads such media processing and delivery, media analytics, cloud gaming, cloud graphics, and immersive media—unlock opportunities for CSPs, CoSPs, and enterprises to attract new business customers and make maximum use of compute, storage, and network resources. By building a modern IT infrastructure—ideally incorporating existing general-purpose Intel architecture-based building blocks—media processing workloads can be handled cost effectively and efficiently, providing tremendous flexibility throughout the data center to adapt to changing requirements.

Data centers are seeing increasing volumes of media operations and what is typically referred to as *new media*—a rapidly evolving class of digital content—is closely associated with the visual cloud. To gain greater flexibility in deploying and reprovisioning servers to handle visual cloud workloads, service providers can obtain interoperable software from an expansive open source community that is developing innovative solutions to streamline deployment of visual computing use cases. Dynamic, high-volume workloads that include immersive media or media analytics are also well served by emerging technologies under development, as well as by the latest development toolkits and code bases.

Intel offers comprehensive, interoperable end-to-end media solutions powered by the Intel Xeon Scalable processor family. This platform is richly supported by a vibrant ecosystem and contributions from an active open source community. To keep pace with rapidly growing visual cloud opportunities, Intel is collaborating with industry leaders to enable innovative solutions through the latest cloud architectures, advanced networks, and agile, scalable compute platforms.

For more information visit **http://www.intel.com/visualcloud**.

[1] *Workload Optimization: A Guide for Cloud Service Providers.* Intel, 2018. https://plan.seek.intel.com/WorkloadOpteGuide_REG

[2] http://www.itprotoday.com/cloud-data-center/intel-owns-95-percent-cloud-and-you-didnt-know-it-until-now

[3] *Cloud Gaming Market Forecasts from 2018 to 2023*. Knowledge Sourcing Intelligence, 2016

[4] https://www.marketsandmarkets.com/PressReleases/iva.asp

[5] Tractica, 2Q, 2017

[6] https://www.statista.com/chart/4602/virtual-and-augmented-reality-software-revenue/