

How Intel Sports built a cloud system for the future

Authors

Phani Vadali

Director of Technology

Waldo Bastian

Sr. Cloud Solutions Architect

Abstract

Sports and the media industry are in the process of a huge technology revolution to create content and experiences to connect with fans. However, most of the production of games and events still happen on location and at the venues. When the coronavirus pandemic struck, the industry was greatly impacted because of the limits to the physical presence of production teams onsite and in venues. Even before the pandemic shook the world, Intel Sports had taken innovative steps to transfer our volumetric video processing to the Cloud, thereby eliminating onsite personnel to produce volumetric content in Intel® True View at the venues. In this paper, we will dig deeper to show how to successfully run Cloud operations for the sports and media industry.

Introduction

Intel Sports is pioneering a next-level viewing experience that allows fans to feel more, know more, and share more of the games they love. Intel True View is Intel Sports' volumetric video platform for data capture, processing, and production. Fans can immerse into their favorite sport by watching the most exciting plays of a game in unbounded perspectives. Intel Sports and our 40+ league and team partners empower fans who are unable to attend live and in-person at the stadium, get down on the field right next to their favorite players and be part of the action. Intel Sports' True View platform creates volumetric video that enables infinite storytelling by capturing every point of view of a play from dozens of camera sensors located in the venue.

Each volumetric frame takes ~30 secs to be produced with the on-premise servers. While a single frame volumetric clip is exciting, Intel Sports is transitioning True View to a near real-time (30 frames per second) volumetric video. To process the 30 fps from cameras, the overall system needs to be able to handle thousands of frames/sec to create the volumetric video.

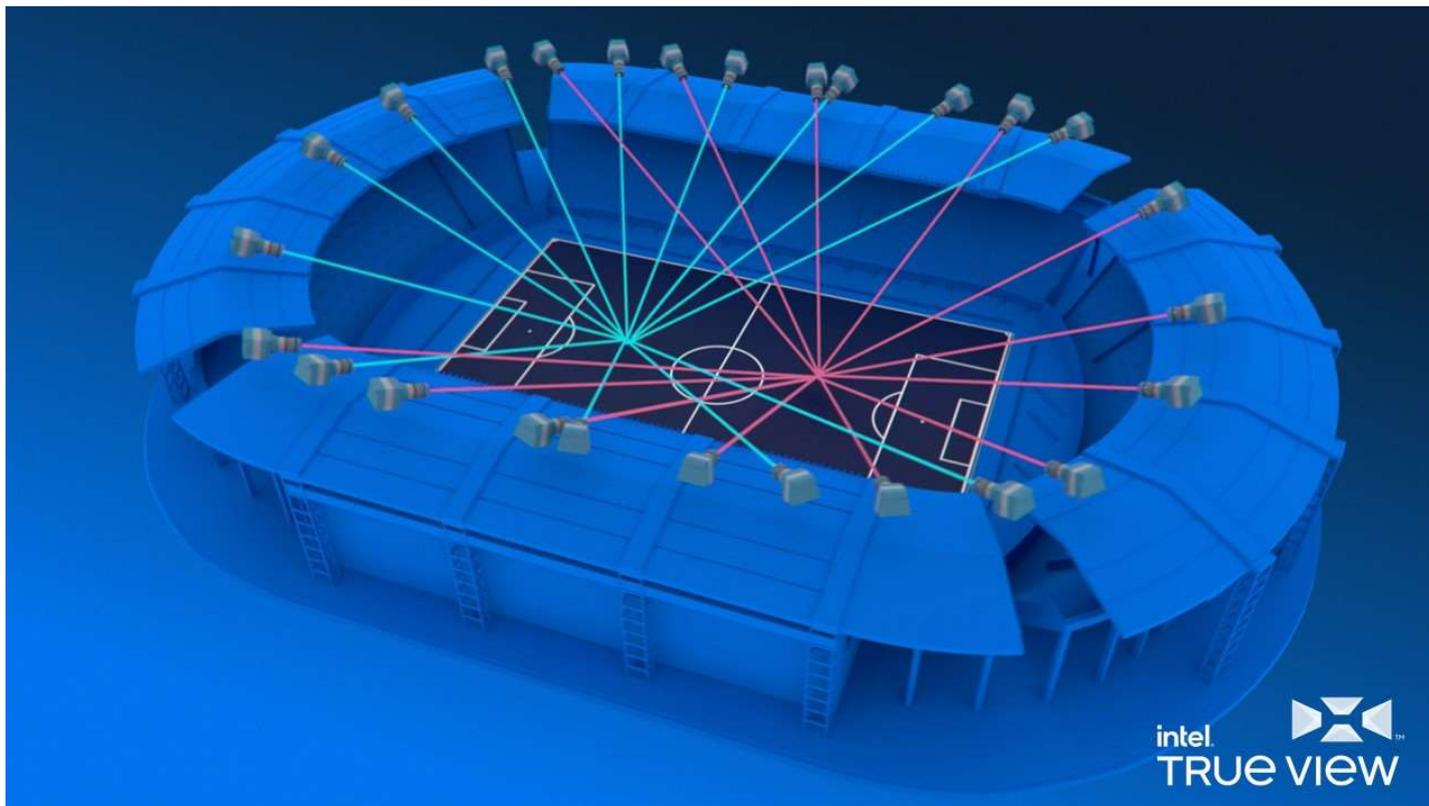


Figure 1. Intel True View in-venue camera system illustration

The massive amount of volumetric data (up to 200 Terabytes of raw data per game) is in the form of voxels which capture height, width, depth, and relative attributes that are needed for point cloud formation. Volumetric video created from the point cloud generates volumetric data that is rendered into high-fidelity volumetric content in the form of virtual camera videos. Introducing a large data center in each venue to process this massive amount of data would be cost prohibitive and impossible due to the physical space, logistics and resources required.

Instead, to support such an advanced system, Intel Sports engineered a cloud compute solution to create volumetric clips to support real time production and scalability across multiple games. It is a distributed system with microservices dedicated to unique algorithmic steps required for creating the volumetric video as described in the diagram below. It is paramount to have a reliable cloud solution that can process vast amounts of volumetric data into content and allow fans to experience their own personal perspective of the game.

In the next section, we will discuss how to enable Cloud Operations for the media production teams and how to operate the system while a game is happening.

Cloud Operations for the Media Industry

Cloud Operations for live sports and the media industry has its own unique challenges and requirements. To operate effectively,

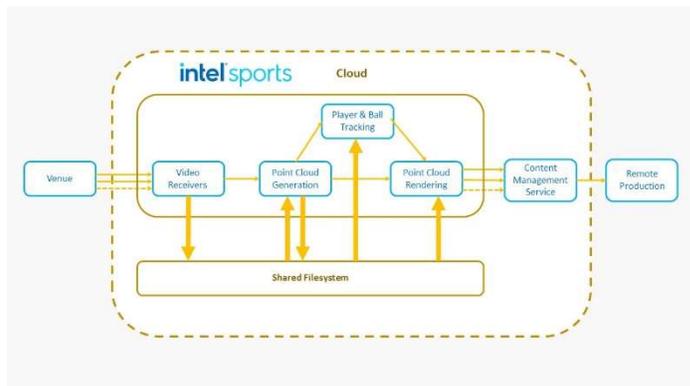


Figure 2. True View Pipeline

Cloud infrastructure for these use-cases needs to be fault tolerant, reliable, and resilient. Automation and building appropriate self-serviceable applications become critical to achieve these goals.

Intel Sports has built applications with these requirements in mind to support content production teams to become self-sufficient in producing games with zero to minimal support from their engineering teams. These applications help the content production teams manage all the different aspects of the game operations. Additionally, these applications help manage the cloud by bringing up the servers needed to produce the games, provisioning them and ensuring they are reliable and available during the game.

On-demand Cloud Infrastructure

While most of the cloud workloads run 24 hours per day, 7 days per week and see peaks due to heavy traffic, cloud workloads in the media industry are typically on-demand and event based. Intel Sports' content production team produces content across the globe at various hours, day and night. Our cloud infrastructure is tightly coupled with this schedule and is automated to support multiple games in different geographic locations.

We considered all these unique requirements while designing our Cloud Operations tool. The production team uses the tool to scale up the "deployment," which is an encapsulation of all the servers and other resources needed to successfully create the content. The scale up process includes acquiring all the cloud instances, provisioning them and deploying our services on to the servers to get them ready for the game. Once the content production is complete, similar processes to shut down the infrastructure are taken. Supporting a cloud infrastructure of this size is expensive and managing usage allows us to keep costs manageable and within budget.

During the content production itself, we have to consider how to auto-scale, auto-recover from failures, and monitor all the cloud resources that are being used for the production. It is crucial to build the cloud infrastructure to ensure that the system can process content to meet the latency service level agreements (SLAs) and ensure fault tolerance.

Monitoring

To ensure a correctly working system, it's important to have insight into what is happening at every level of the system. This begins at the infrastructure level to ensure the servers are healthy and operate within the constraints of the operating environment. CPU & GPU loads are monitored to ensure the servers are not overloaded, network usage is monitored to ensure the data can efficiently flow from one server to the next without incurring unneeded delays, and the storage throughput is closely watched. For each of the observed metrics, we establish a healthy operating range where we want our system to operate. If the metrics stray out of this range for a prolonged time, alerts are triggered. The alerts serve two purposes: first, to make sure that the operator is aware of any issues as soon as they arise, and second, and more importantly, to use automated responses and build automated recovery procedures to correct issues that have known solutions.

Apart from the infrastructure, we also closely monitor the service availability. The Kubernetes container orchestration framework will monitor and correct service issues by itself, but we still want to provide the operator insight into issues as they occur. Some problems might arise out of the configuration of the application and cannot be solved by Kubernetes on its own.

The last layer of monitoring happens at the application level. All the micro-services involved in the processing of the volumetric

content generate application-level monitoring events that are processed in real-time by a distributed analytics application. The analytics application generates specific metrics used to monitor the quality and performance of the volumetric processing. Just like the infrastructure and service level monitoring, it too can generate alerts and trigger automatic recovery processes for known application-level error conditions.

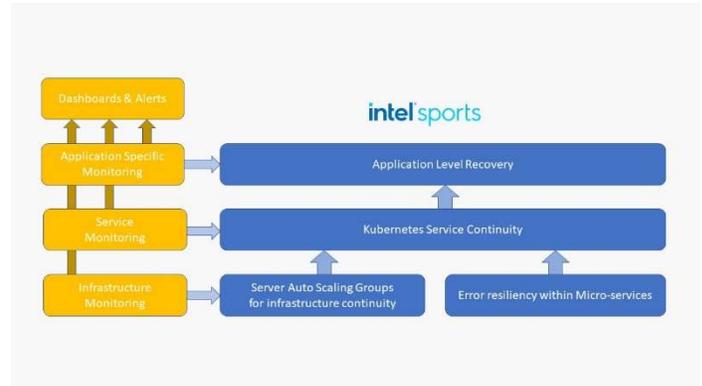


Figure 3. Different levels of Monitoring

Auto Recovery & Resiliency

In addition to monitoring, resiliency is provided at multiple layers in the stack. It starts at the micro-service level where each service will try to recover from errors optimally. The next level of resiliency is provided by the Kubernetes container orchestration framework which ensures that enough micro-service instances will always run. If a single micro-service instance fails a periodic health check it will immediately be replaced with a new healthy instance.

A similar approach is taken for the compute infrastructure where health checks are used to ensure that each compute instance is healthy. Server scaling groups provided by the cloud service provider are used to maintain the desired number of healthy compute instances. If an issue is detected with any one of them, the scaling group will take it out of service and replace it with a new healthy instance. On top of these service and infrastructure level safeguards, Intel Sports has developed its own real-time monitoring framework that continuously monitors application-level events generated by the volumetric processing pipeline. If it detects an anomaly, it can use the services provided by the service and infrastructure levels to either restart micro-services or replace servers depending on the type of error detected. By adding another layer of redundancy to the infrastructure, we aim to ensure there is minimal service disruption.

Summary

The sports and media industry will continue to go through a much-needed technical transformation and to meet the increasing demands for content from fans, it will require the move of various workloads to the Cloud for remote processing. To design the cloud infrastructure for these workloads, we must

consider all the unique requirements for a fast and uninterrupted production.

These unique requirements include designing a system that can transfer content from the venue to the cloud with very small latency and ensure the infrastructure is fast, reliable, scalable, and efficient. Intel Sports has successfully broken this technical barrier enabling leaps and bounds of innovation in this space. It is critical to consider all of the Cloud features such as auto-scaling, auto-recovery, and monitoring to provide a robust system to content production teams to support their operations.

In 2020, the volumetric video content for the NFL, including Super Bowl LV, was created by Intel Sports using these tools and technology. While the production team was not allowed into venues due to tight pandemic restrictions, being Cloud ready enabled our team to continue creating amazing storytelling content for fans. In 2021, Intel Sports is scaling our Cloud operations to continue to provide immersive media experiences in True View for our partners' fans to re-live memorable experiences seconds, days and years after they happen.



© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries

Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.



Figure 4. Intel True View showcasing the NFL Super Bowl LV in True View on Twitter