lt momento

Building the Ultimate Viewer Experience

The Business Impact of a Better Viewer Experience





Live streaming video is key to your company's success with stream quality impacting advertising revenue, broadcasting rights, business deals, and brand reputation. While content can attract viewers and sponsorships, poor quality can undermine key metrics like audience retention and brand image. Despite significant advancements in video streaming technology over the past two decades, there's still room to enhance the viewer experience and move closer to the Ultimate Viewer Experience.

The Quest for the Ultimate Viewer Experience

Building the ultimate viewer experience requires four key components:

No Buffering:

Buffering typically occurs due to insufficient bandwidth or network congestion and is a visible clue to the user that something has gone wrong with their stream, negatively impacting their viewing experience. Tracking and measuring these events through a metric like "Zero Buffer Rate" (the percentage of viewers who don't experience buffering over the course of an event) is critical to creating the Ultimate Viewer Experience.

High-Quality Video:

Adaptive bitrate streaming allows the playback device to select the highest bitrate variant (highest quality) stream that fits within the customer's internet connection. Receiving the highest bitrate streams and having an experience that meets or ideally exceeds the customer's expectations for a broadcast stream leads to increased customer satisfaction.

Low Latency:

Latency is important in the time trailing broadcast, Time to First Frame, and overall latency between the playback device and content source. The time trailing broadcast should be minimized as much as possible to protect against other sources, such as social media, spoiling a match's outcome. Still, it is also becoming more important than ever due to the rising importance of sports betting. Reducing the communications latency between the playback device and content source can improve the chances for the playback device to receive the best quality video available by allowing the largest amount of time possible to download high bitrate variants.

Consistency:

Creating a workflow that performs well across the metrics won't matter unless the workflow is reliable when there are a few viewers, as well as when you are dealing with a thundering herd of viewers at key moments in an event and breaking viewership records.



Even though many consumers choose their platforms based on content offerings and ease of use,

47%

have reported unsubscribing due to buffering issues. Just in terms of subscriptions, this could cost companies \$390 million per month.¹

Live Events Are Vital in the Lives of Viewers

Large-scale live event streaming is a core facet of media; from sporting events to live musicals, political rallies, and even board meetings, viewers like watching the world around them in real time. Even with a world of video on demand, people continue to tune into these live events, and using metrics such as Zero Buffer Rate (the number of customers that watch a stream and do not have a buffering event) can help you make sure you are delivering the Ultimate Viewer Experience.

In the summer of 2024, over **23 billion minutes** of the Paris Olympics were watched, and **16.8 billion** of those minutes were live-streamed over digital platforms.² These iconic cultural moments are unmissable with an increasingly large number of over-the-top viewers.

51% of consumers rate live sports as a top-watched genre, and

of sports fans prefer to watch live events at home rather than in person.³

li momento

The Importance of Live Streaming

While stream quality can outsize the bottom line, it is also more challenging for live events. Not only is the audience more plugged into the moment, but the high view count and traffic spikes can also put more pressure on the distribution infrastructure. A solution must account for a broad set of unique obstacles so concerts, announcements, and games can provide the best viewing experience possible.

Live events are cultural moments that are more time-sensitive and attract a bigger audience than other forms of media. The high number of viewers signing in simultaneously creates demand spikes across all areas of the infrastructure, from sign-in and user concurrency checks to the actual content being delivered.

Streaming technology, especially adaptive bitrate streaming (ABR), places significant demands on broadcasters. As viewers with different devices and internet speeds require various video qualities, more requests are generated for the same content. This increased load on the infrastructure can strain the system, impacting overall performance.

The content creation process begins at the source and moves towards the origin server. Conversely, the viewer's experience starts at the player and moves towards the origin. The origin server acts as a crucial bridge between these two processes. It continuously updates playlists for live streams to inform players about available video variants. The CDN then requests the appropriate video files from the origin server based on the player's needs. Due to its central role, the origin server can become a bottleneck, potentially impacting the overall viewing experience.



Even beyond the surges, the traffic often remains high as viewers watch the whole live event;

69% of sports fans watch until the end of the game, even if their team is losing.³

Similar to the cascading impact of tail latencies, read and write requests have compounding impacts on stream quality. As a manifest is made of all the segments, a single issue delivering the segment or reading the segment at the origin impacts the quality. With these obstacles that separate live streams from other forms of content, it's clear that they need a purpose-built infrastructure to deliver the Ultimate Viewer Experience.



Delivering the Ultimate Viewer Experience

To deliver the Ultimate Viewer Experience, an origin must exhibit several key characteristics:

Consistent Performance:

Origin server read and write latency should be consistently shorter than segment lengths to avoid buffering. It should maintain stable performance, avoiding sudden spikes or drops in response times. Understanding tail latencies, or the worst-case performance, is crucial for ensuring a reliable streaming experience.

Hot Item Mitigation:

The origin server must effectively manage hot items or frequently accessed content to handle sudden traffic surges. This involves replication and load balancing techniques to distribute the load across multiple servers and prevent performance degradation without operator intervention.

Scalability and Flexibility:

The origin server should be designed to handle the dynamic nature of live streaming, scaling seamlessly to accommodate sudden traffic spikes. It should also support various content tiers, allowing for efficient storage and delivery of live and on-demand content.

Robust Security:

A robust security infrastructure is essential to protect content and prevent unauthorized access. The origin server should implement strong security measures, such as access controls, without compromising performance or ease of management.

By selecting an origin server that meets these criteria, you can significantly improve the quality and reliability of your live-streaming service, ultimately enhancing the viewer experience.

The Challenges of Object-oriented Storage as a Live Streaming Origin

Origins are essential for managing the files that make up the live video stream, constantly storing the updated video files and the variant manifests, and delivering them to the CDN. Because CDN points of perspective (PoPs) collapse requests, they can put a lot of read pressure on the origin. A failed cache hit, read request error, and latency in the origin can have an exaggerated effect on the delivery of content to viewers. Consider that a single request for a CDN's origin shield could impact tens of thousands of viewers. Milliseconds and retry requests add up, so origins need to render requests quickly, regardless of spikes in the volume of requests.

Many organizations use object-based storage for their origin servers, but the base infrastructure wasn't built for live streaming. Object-based storage is ultra-reliable, durable, and cost-effective, but there is a mismatch when the goal is lower time-to-first-byte (TTFB), consistent tail latencies, and near-zero error rates. Typical object-based storage solutions have SLAs of around 99.9%, meaning that 0.1% or one request out of a thousand will result in an error for both reads and writes. If you consider that to stream a typical American Football game with seven variants and 2s segments, there will be approximately 100,000 writes to the origin; this means that roughly a little more than 100 will result in errors, and there are many more reads from the CDNs which will result in many hundreds of read errors in even the simplest of scenarios. These errors will likely be retried successfully, but precious time will be lost.



Tail latencies and response times, significantly longer than the average response time, are also a concern for cloud object-based storage systems. It is common to find p99 times that are 5x and p999 times that are 7x longer than p50 times. This means that randomly, the CDN will experience a response from an object-based storage origin that takes 7 times longer than the median response time. These delays can impact the streaming content moving through the system and, with origin shields, can impact a significantly large number of viewers.

Finally, though HTTP/2 has been used for over 10 years, W3Tech has found that only around 35% of web servers use the technology.⁴ HTTP/2 has many great technological advances, including header packing, bundling requests, and faster response times. However, upon implementing HTTP/2, many companies were forced to roll back to standard HTTP. They have found that their underlying database and caching infrastructure cannot efficiently process bundled requests fast enough, resulting in overall system slowdowns.



To overcome these challenges, a purpose-built serverless solution is ideal for live streaming origins. Serverless architectures offer the flexibility to scale seamlessly to meet the demands of live events, ensuring low latency and high performance. By eliminating the need for traditional infrastructure management, serverless solutions can significantly reduce operational overhead and improve overall reliability.

By understanding and addressing these challenges, you can select the optimal origin solution to deliver a seamless and high-quality live streaming experience.

li momento

Momento's Focus on the Relentless Pursuit of Quality Enables the Ultimate Viewer Experience

Momento Media Storage is a purpose-built live origin service designed to overcome the limitations of traditional object-based storage origins and deliver the Ultimate Viewer Experience. With a relentless focus on speed, scalability, and reliability, Momento empowers content streamers to deliver high-quality, low-latency live streams.

Momento's innovative architecture, which includes a high-performance RAM cache and an intelligent datastore, enables it to handle the demanding requirements of live streaming. The RAM cache ensures rapid response times for frequently accessed content, while the intelligent datastore automatically tiers content to the appropriate storage tier based on its access patterns. This approach minimizes latency and maximizes cost efficiency.

A key way that Momento Media Storage addresses the technology challenges of live origins is with consistently fast response times. By significantly outperforming object-based storage systems, Momento Media Storage provides significantly faster response times, especially for GETs/READs, the most common requests

Solution	p50	p90	p99	p999
Object Store PUT	146ms	219ms	328ms	530ms
Momento PUT	23.7ms	27.7ms	36.4ms	52ms
Object Store GET	119ms	302ms	502ms	775ms
Momento GET	5.6ms	6.1ms	6.9ms	7.6ms

* 4MiB items, 100 requests per second

Built by a team of engineers from DynamoDB with streaming media backgrounds, Momento Media Storage is designed to improve upon the key metrics and address the evolving needs of live streaming. By leveraging cutting-edge technology and a deep understanding of streaming workflows, Momento empowers content streamers to deliver the Ultimate Viewing Experience.



Your audience expects a seamless, high-quality live-streaming experience. Buffering, glitches, and crashes can quickly erode viewer engagement and damage your brand reputation.

Momento Media Storage is designed to address these challenges. By leveraging advanced technology and a deep understanding of live streaming, Momento provides a robust and reliable solution for delivering exceptional live streams.

Build the Ultimate Viewer Experience with Momento Media Storage.

Learn More About Momento



2 "NBCUniversal's Presentation of Spectacular Paris Olympics Dominates Media Landscape Across All Platforms," NBCUniversal, August 13, 2024. https://www.nbcuniversal.com/article/nbcuniversals-presentation-spectacular-paris-olympics-dominates-media-landscape-across-all-platforms

3 "Behind the Remote: A Deep Dive into Sports Viewership & Fandom," Effectv, October 2, 2024. https://www.effectv.com/insights/research-reports/behind-the-remote-sports-viewership-fandom/

4 "Usage statistics of HTTP/2 for websites," W3 Techs. https://w3techs.com/technologies/details/ce-http2