

www.coremedia.com

The Content Distribution Network (CDN) Challenge

A Hybrid Approach to Dynamic, High Performance Web Caching

Content Distribution Networks (CDNs) are a popular and effective means of increasing the performance and reliability of high traffic websites while reducing total cost of ownership. However, as the Web evolves from a primarily static publishing environment to a dynamic, personalized, multi-touchpoint experience, many companies are being forced to confront the limitations of this approach.

This brief document will examine both the strengths and weaknesses of CDN Web caching and show how companies can combine CDN technology with CoreMedia's own high performance caching infrastructure to:

- Deliver more page impressions
- Reduce time-to-Web for critical content
- Support the need for increasingly dynamic and "contextual" content
- Reduce total cost of ownership

The Challenge

The online experience has never been more complex. Online today means more than just a static website. Visitors expect to do more online and they want to do it faster and more efficiently. Web sites must be accessible from a wide range of mobile devices and must compete against a growing array of new competitors including blogs, social networks, and content aggregation portals.

Web publishers are looking for ways to enhance site performance, and expand brand reach, and retain visitors. Yet, it's becoming increasingly difficult to attract and retain audiences online. The only way to remain competitive is to meet (or even exceed) your visitors' expectations.

In order to deliver successful online experiences to your visitors, your Web delivery infrastructure must be:

- **Fast:** Low latency and the ability to deliver content quickly to anyone, anywhere in the world
- **Timely:** Support for frequent updates and reduced time-to-Web for breaking news and other time sensitive content
- **Robust:** High throughput to serve a large number of unique users in parallel
- **Scalable:** Able to accommodate many page impressions per user per day as well as unexpected surges in traffic
- **Engaging:** Able to deliver pages with many rich media objects — including high resolution formats
- **Reliable:** High availability and protected against distributed denial-of-service (DDOS) attacks and other threats

Companies that can meet these requirements are rewarded with shorter load times and greater page impressions which lead to happier visitors, higher rates of conversion, and improved SEO. Those who can't meet them lose visitors, sales, and revenues to those who can.

Yet, at the same time that websites are being pushed to innovate and increase performance, economic pressures are driving them to cut costs, consolidate operations, and increase productivity. Companies need a solution that increases site performance while simultaneously lowering their total cost of ownership.

The History: How Companies Traditionally Approached This Problem

Until recently, the only effective approach to this problem has been to work with a Content Distribution Network. CDNs were first introduced in the late 90s as a way to increase the reliability and scalability of a site without requiring additional investments in expensive hardware, software, and maintenance.

A CDN is a geographically distributed service that transparently shifts Web content from the "origin" servers in company's data center to an optimized network of "edge" servers that are located closer to the user who is requesting the content. The CDN administrators use a variety of techniques to cache content, balance server loads, and automatically route requests to the fastest (or cheapest) server location.

Content on a CDN is typically cached for a fixed time period and is only updated when a pre-determined time-to-live (TTL) setting has been exceeded. For example, long tail or archival content may only need to be refreshed once a day, whereas breaking news content may need to be updated every couple of minutes.

The benefits of this approach are significant. The CDN cache shifts the performance burden from a company's Web servers to the CDN's servers. Companies can also take advantage of the CDN's vast shared network to reduce costs and scale virtually to meet unexpected demand. Companies that deploy a CDN to publish their Web content can experience:

- Shorter latency and higher overall throughput since the content is located physically closer to the website visitor
- Improved multimedia delivery performance
- Increased scalability and reliability
- Decreased load on internal systems leading to significantly lower total cost of ownership

Despite their power and convenience, CDNs can't solve all the challenges faced by high traffic dynamic websites.

The limitations of relying solely on a CDN to distribute your Web content include:

- Most CDNs cannot handle highly personalized or context-aware content
- CDNs typically rely on simple time-based caching, which leads to increased time-to-Web
- Content that is extremely dynamic and must be constantly updated (e.g. a stock ticker, breaking news, social content, a personalized home page, or a dynamically generated micro-site) cannot be effectively cached on a CDN

At the end of the day, any CDN deployment is going to involve a tradeoff between increased throughput and scalability on the one hand versus faster time-to-Web and more dynamic content on the other.

The following table summarizes the different benefits of each approach:

Without a CDN	With a CDN
<ul style="list-style-type: none"> → Every request goes directly to company's data center 	<ul style="list-style-type: none"> → The CDN's caching servers take care of most of the requests
<ul style="list-style-type: none"> → Direct correlation between the number of requests and the overall server load 	<ul style="list-style-type: none"> → Origin servers only need to take care of requests coming from the CDN
<ul style="list-style-type: none"> → Full network bandwidth is required 	<ul style="list-style-type: none"> → Reduced traffic into the company's data center

CDNs provide users with a range of strategies for balancing the need for great throughput and reduced cost against the need for reduced time-to-Web. For example, the edge server can make frequent conditional requests to the origin to see if the status of a Web resource has changed before its explicit TTL setting has been exceeded. This helps address some of the time-to-Web limitations but does little to address the need for personalization or other forms of dynamic content. Another approach is to perform a partial flush of the CDN cache when urgent updates are required. However, this is really only suitable for exceptional situation — such as updates required by legal or regulatory mandates. The unpredictable latency involved in flushing and replacing the edge cache makes it an unsuitable method for regular updates.

The CoreMedia Approach – a Hybrid Architecture

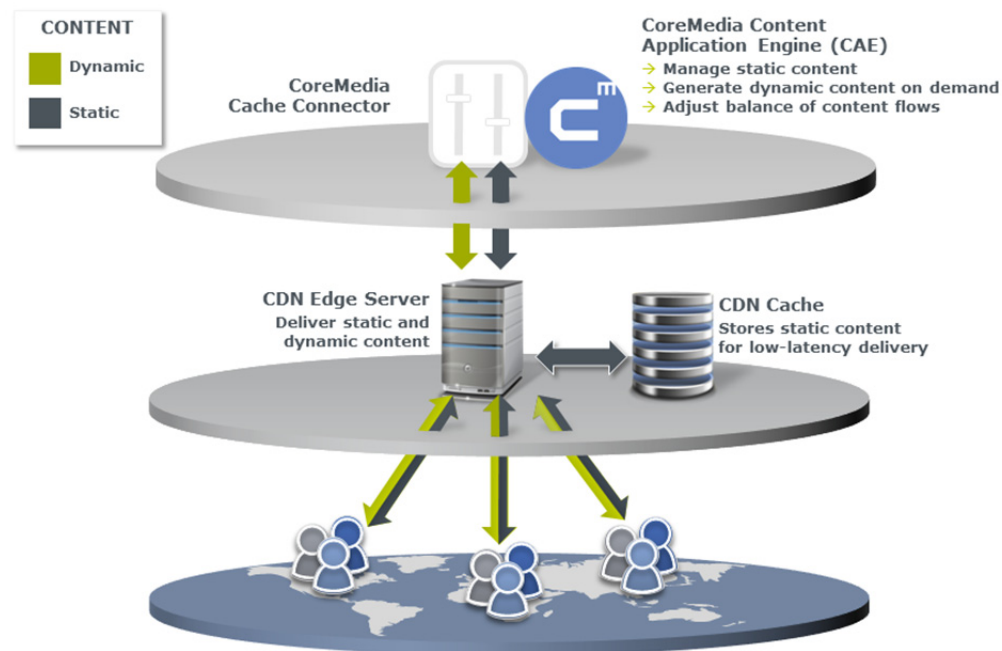
At CoreMedia, however, we believe that there is a better way.

CoreMedia’s approach to Web caching is designed to balance the speed and flexibility of our own sophisticated multi-tier caching infrastructure with the reduced costs and improved throughput of a CDN.

In a dynamic, contextually-aware Web environment, optimal performance depends on the ability to select and assemble, and publish relevant content on-demand. This can require immense processing power if memory is not used efficiently and cannot be handled effectively by most CDN caching configurations. Fortunately, there’s a better way to provide up-to-date content dynamically without incurring exorbitant hardware costs: CoreMedia’s Content Application Engine manages this by deploying a unique, multi-level caching system with an event-driven invalidation mechanism.

Multi-level means that you don’t just cache the final Web page. Page-level caching on its own provides limited performance benefits in a dynamic Web environment where every visitor is potentially seeing an entirely different assemblage of page elements.

In addition to page caching, you also need to cache content at the business object and content level as well. Not only does this ensure that all contextually-relevant content is immediately ready to be assembled into a new page for every visitor, it also guarantees that all relevant external data is cached along with the native Web content.



CoreMedia hybrid approach to caching allows for optimal CDN configuration for your implementation.

Event-driven invalidation means that the cache is only refreshed when something actually alters the status of a stored content object. When an event triggers a change in the content at one level, the system automatically communicates this change to all the other caching levels as well — including pages, business objects, and data. Smart caching thus keeps track of all dependencies between content items for instant updates and ultimate performance.

CoreMedia helps companies get the best of both worlds by supporting highly dynamic or frequently updated content on our own caching servers, while seamlessly integrating a CDN to take care of less time-sensitive content.

The CoreMedia approach allows any site administrator to fine-tune the mix of origin and CDN traffic by adjusting a simple set of parameters within the WCM system to achieve the perfect balance of time-to-Web, throughput, and dynamic delivery.

Our consultants will work with your team to determine the optimal CDN configuration for your company and will deploy a custom implementation that combines excellent support for static, cacheable requests on the edge servers for high volume pages with multi-level, event-driven caching for more dynamic content.

This approach enables extremely low time-to-Web while reducing resource consumption (origin CPU and bandwidth).

Our CDN integration expertise has already been used to optimize the performance of a range of high performance, global websites, including the official site of the 2010 Winter Olympics in Vancouver, Canada, where CoreMedia implemented a real time cache connector between the CDN and own Content Application Engine (CAE) server technology that improved performance by offloading a large percentage of requests to the edge servers while maintaining the timeliness of content stored in the CDN.

Other examples of CDN integration expertise include the news website of the Australian Broadcasting Corporation (ABC) and the Web properties of leading European publisher, M. DuMont Schauberg.

By enabling a sophisticated integration between the CDN and the full content delivery infrastructure in your data center, CoreMedia helps you to achieve the greatest cost savings and performance benefits from your CDN without sacrificing content freshness or personalization.

CoreMedia Corporation
118 Second Street
Fifth Floor
San Francisco, CA 94105
tel: +1.415.371.0400

Email: info@coremedia.com
Web: www.coremedia.com