



# 동적 데이터, 페이지 또는 구성요소 단위로 캐싱하라!



## 연재 순서

1. 동적 콘텐츠 캐싱 기술 (이번호)
2. 글로벌 로드밸런싱 기술
3. 콘텐츠 동기화 기술



이윤근 씨디네트웍스 기술연구소 개발팀장

최근에는 라스트 마일 구간인 가입자 회선 구간이 메트로 이터넷이나 VDSL 등의 보급으로 대역폭이 계속 늘어나는 것과 동시에 이러한 회선을 충분히 활용하는 고용량 콘텐츠도 보편화돼 가고 있다.

지난해 초부터 포털업체들은 고화질 VOD관을 개설해 1M급 스트리밍 서비스를 제공하고 있고, 가입자 회선 역시 날이 갈수록 발전하고 있다. 하지만 ISP들간의 연동 구간인 미들 마일 구간은 가입자 구간의 발전과 콘텐츠의 고용량화를 따라가고 있지 못하고 있는 상황이다. 이런 이유로 최근에는 미들 마일 구간의 병목구간 회피와 재난복구가 가능한 트래픽의 분산처리를 담당하는 CDN서비스에 대한 관심이 커지고 있다.

CDN에 대해서는 많은 지면에서 소개가 있었기에 여기서는 CDN의 요소 기술들에 대해 몇 가지 살펴보고자 한다. 먼저 1회에서는 여러 네트워크에 분산 배치되는 캐시 서버에 대해 알아보자. 오늘날 보편적으로 이용되고 있는 캐시서버는 정적인 데이터를 캐싱하는 서버들이지

만, 현재 많은 사이트가 정적인 데이터보다는 동적으로 화면을 구성하게 구축돼 있으므로 이러한 사이트에 적용될 수 있는 동적인 데이터의 캐싱에 대해 알아보겠다.

### 동적 데이터 캐싱의 필요성

캐시서버가 처리할 수 있는 데이터의 유형으로는 HTTP 정적 데이터, HTTP 동적 데이터, FTP 데이터, 스트리밍 데이터 등이 있는데, 일반적으로 많이 이용되고 있는 캐시서버는 HTTP 정적 데이터를 처리하는 것이다.

HTTP 정적 데이터나 FTP 데이터, 스트리밍 데이터 등은 모두 정적인 데이터들이므로 캐시서버의 효과를 많이 볼 수 있는 영역이 될 것이다. 캐시서버라는 것이 처음 사용자에게서 서비스 요청을 받으면 오리진 서버에서 실제 데이터를 가져다가 대신 서비스를 해주고, 그 데이터를 로컬 캐시 디스크에 저장해둔다. 이후 같은 데이터에 대해 다른 사용자에게서 요청을 받으면 오리진 서버에서 데이터를 가져오는 것이 아니라, 로컬에 저장해둔 데이터로 서비스를 해주게 된다.

캐시서버의 또 다른 장점으로는 데이터 관리정책에 따라 최신의 데이터를 가지고 있다는 것이다.

최근에는 e-비즈니스가 활성화됨에 따라 웹페이지의 구성이 단순한 정적인 페이지로 된 구성에서 ASP, JSP, PHP 등과 같은 동적 스크립트 기술을 이용해 다양한 정보들을 보여주고 있다. 일반적인 콘텐츠 제공업체를 비롯해 여러 포털업체들은 이 단계를 넘어 사용자들에게 개인화된 맞춤페이지를 제공하는 경우도 많이 있다.

이런 HTTP 동적 데이터의 경우에 캐시를 적용하는 것은 위에서 설명한 정적인 데이터를 캐싱하는 것과는 사정이 다르다. 캐시서버가 정적인 데이터에 대해서는 많은 효과가 있겠지만, 동적인 데이터에 있어서는 인터넷 여러 곳에 분포돼 있는 캐시서버로 사전 적재를 하는데 많은 어려움이 있는 것이다. 또한 일반 사용자의 요청에 의해 저장돼 있는 동적인 데이터가 다른 사용자에게 제공할 수 있는지에 대한 여부도 확실하지 않다. 예를 들어 많은 웹사이트에서 제공하는 게시판, 뉴스, 주식정보, 날씨 등의 정보는 정적인 자료가 아니라 매 시각마다 변할 수 있는 동적인 자료이고, 이런 자료들이 시시각각으로 변하고, 사용자에게 따라 맞춤화 된 다른 정보를 제공해야 하기 때문이다.

일반사용자들에게 효과적인 개인화 된 맞춤식 웹페이지를 제공하기 위해서는 정확성, 안정성, 관련성이 보장돼야 한다. 정확성이란 시시각각으로 변하는 소스 데이터가 일반 사용자들이 원하는 시점에 항상 최신 데이터를 보여주어야 한다는 것을 말하고, 안정성이란 지정된 사용자들에게만 정보가 보여주어야만 한다는 것이다. 또 관련성이란 일반사용자의 성향이나 호감을 가질 만한 것들을 보여주어야 한다는 말이다. 이런 개인화 된 맞춤식 웹페이지를 운영하기 위한 요구조건은 데이터를 생성하거나 전송하는데 드는 비용을 최소화한다는 정책과는 어느 정도 상반되는 점이 있다.

### 페이지 단위 캐싱, 서버 부하 감소 효과

동적인 페이지로 서비스하는데 있어서, 데이터를 생성하거나 전송하는데 드는 비용을 최소화하기 위해서는 캐시서버 운영이 대안이 될 수 있으며 위에서 이야기한 정확성, 안정성, 관련성이 담보가 돼야 한다. 이런 것을 해결하면서 캐시를 도입할 수 있는 방법으로는 다음과 같은 2가지 방법이 있다.

첫 번째는 페이지 단위로 캐싱을 하는 것이다. 페이지 단위의 캐시는 특정한 사용자 그룹에 적합한 웹 콘텐츠를 페이지 단위로 캐시에 저장함으로써 구현할 수 있다. 이는 일반적인 캐시보다는 좀더 많은 처리능력을 요구하지만, 특정 사용자마다의 콘텐츠를 캐시할 수 있게 된다. 페이지 단위로 캐시할 수 있는 기본적인 원리는 콘텐츠를 요청하는 HTTP 헤더정보(브라우저의 언어나, 지역적인 정보, 혹은 개인화 된 정보 등)에 기초를 하게 된다.

예를 들면 브라우저의 언어가 한글이라면 한글로 된 콘텐츠의 내용을 보여주고, 영어로 돼 있다면 영문 콘텐츠의 내용을 보여줄 수 있다. 또는 개인화 된 정보에 개인의 이름이 들어가 있다면 일반적인 초기 화면에 “안녕

하세요? 홍길동님"이라는 메시지를 담은 페이지를 보여줄 수도 있을 것이다. 물론 개인마다 이런 페이지를 개별적으로 캐싱하고 있어야 한다는 부담은 있지만, 사용자들이 들어올 때마다 이런 일반적인 페이지를 동적으로 생성하는데 드는 부하를 줄일 수 있을 것이다.

페이지 단위 캐시의 이점을 살펴보면 다음과 같다.

- 특별한 사용자 그룹을 위해 각 페이지를 개별적으로 생성할 필요가 없어 캐시를 함으로써 페이지를 생성하기 위한 서버의 로드를 줄일 수 있다.
- 콘텐츠 그 자체가 아무리 일반적인 내용이라 하더라도 유일한 데이터를 가진 HTTP 헤더정보를 변경하는 것만으로도 일반 사용자들에게 개인화 된 정보를 서비스 받는다는 인상을 줄 수 있다. 하지만 페이지 단위의 캐시는 이런 이점에도 불구하고 약간의 문제점을 가지고 있다.
- 사용자들의 선호도를 상세하게 반영하기 힘들다 : 캐시되는 정보들은 브라우저의 언어나 지역적인 정보, 혹은 개인화 된 정보 등과 같은 HTTP 헤더정보에 의존하기 때문에 동적인 페이지를 구성하는데 한계가 존재한다. 애플리케이션 서버와 대화 없이 HTTP 헤더정보만으로 동적으로 구성되는 페이지는 사용자들에게 개인화 된 맞춤형 페이지를 제공하기 어렵다.
- 낮은 캐시 히트율 : 다양한 개인화 된 콘텐츠를 캐시에 저장하므로 캐시된 데이터는 그만큼 사용률이 떨어지게 된다. 하나의 콘텐츠에 대해서 캐시 데이터들은 사용자들의 개인화 된 정보를 부가적으로 가지게 되며, 각각이 서로 다른 콘텐츠로 인식이 되므로 개인화 된 정보만큼이나 캐시 데이터를 많이 보유해야 한다. 캐시 데이터를 저장하는 디스크의 용량의 한계와 개인화 된 데이터의 노출이 특정 사용자 그룹에게만 이루어지므로 캐시서버의 히트율이 떨어지게 된다. 캐시에 대한 히트율이 떨어지게 되면 자연스럽게 실제 애플리케이션 서버로 요청이 들어가게 되는 확률이 다소 높아지게 된다.

### 페이지 구성요소 단위 캐싱, 비용 절감 가능

동적인 페이지를 캐싱하는 또 다른 방법으로는 하나의 동적인 페이지를 각 구성요소 단위로 쪼개어 캐싱하는 방법이다. <그림1>에서와 같이 동적으로 생성되는 페이지 내에는 다른 부분보다는 특히 동적으로 생성해야 하는 구성 요소들이 존재한다. 예를 들면 로고나 헤더, 그리고 광고와 같은 것들은 정적인 데이터들이고, 이들은 정적인 캐시서버를 통해 충분히 서비스를 받을 수가 있다. 이렇게 동적인 페이지라 하더라도 그 페이지 내에는 항상 정적인 데이터들이 존재하기 마련이다. 이런 구성 요소들은 맞춤형페이지나 개인화 된 페이지에 모두 나타나게 된다.

최신뉴스, 날씨, 주식시세, 스포츠 점수 등과 같은 동적인 구성 요소들은 소스 데이터들이 시시각각으로 변화하기 때문에 그에 따라 매우 다양하게 변화하게 된다. 최신 뉴스는 기자들이 매시간 기사를 올리고, 주식 시세, 스포츠 점수 등도 계속해서 바뀌는 내용들이다. 이런 내용들은 분이나 초단위로 변경될 수도 있고, 시간단위로 변경이 될 수도 있다. 사용자의 개인화 된 정보를 반영하면 특정개인이 관심 있는 주식종목이 있을 수 있고, 사는 지역에 따라 관심이 있는 지역 날씨가 다를 수 있다. 이런 정보들을 이용해 사용자들의 관련성과 연계해 동적이 페이지를 구성할 수 있다.

정적인 데이터만을 캐시하거나, 동적인 데이터를 페

<그림1> 동적 페이지의 구성 요소



이지 단위로 캐시를 하는 대신에 동적인 페이지를 여러 구성 요소로 쪼개어 캐시를 함으로써, 웹서버나 애플리케이션 서버에서 데이터를 가져올 때 구성 요소를 하나의 페이지로 가정하고 요청함으로써 구성요소 캐시도 가능하다.

이렇게 동적인 페이지들을 구성요소 단위로 캐싱을 하게 되면 다음과 같은 이점이 있다.

- **향상된 히트율** : 캐시된 콘텐츠는 사용자의 요청을 많은 부분 수용할 수 있다. 로고나 헤더, 광고와 같은 데이터는 모든 사용자의 지역 날씨가 최신뉴스와 같은 요청에 포함돼 있으므로 좀더 많은 히트율이 발생할 것이다.
- **정확성에 대한 개선** : 각 구성 요소들의 유효성에 대한 활용도가 높아진다. 일반적인 캐시와는 다르게 일부 구성 요소들이 변경됨에 따라 전체적인 내용이 버려지는 것이 아니라 각 구성 요소 각각을 캐시하기 때문에 활용도가 높아진다. 또 그만큼 콘텐츠를 새로 생성하는 것이 아니기 때문에 콘텐츠의 생성과 전송시간 단축에 따라 사용자에 대한 만족도가 높아지게 된다.
- **비용 절감** : 각 구성 요소들은 애플리케이션 서버를 통해 한번만 생성되더라도 많은 사용자의 요구사항을 맞춰줄 수 있기 때문에 개입된 페이지를 매번 생성하는 것보다는 하드웨어, 소프트웨어의 비용을 절감할 수 있다.

현재 널리 쓰이고 있는 동적인 데이터를 캐싱하는 방법은 페이지 내에 동적 데이터 부분을 특정 태그로 감싸는 것이다. 이러한 태그는 HTML 태그는 아니고, 보통 동적 캐싱을 지원하는 솔루션 업체에서 만들어서 제공하고 있다.

이런 동적 캐싱은 웹서버 앞단에서만 수행되는 것이 아니라 인터넷의 옛지인 라스트 마일 구간에도 동적 캐싱 서버를 구축해 운영이 가능하다. 이런 모델로 미국 CDN 사업자인 아카마이, 미러 이미지, 디지털 아일랜드 등과 오라클이 공동 제안해 표준화가 진행되고 있다. 관련된 태그는 ESI 사이트(Edge Side Include : [www.esi.org](http://www.esi.org))에서 자세하게 살펴볼 수 있다.

이런 페이지 구성요소 단위별 동적 캐싱을 지원하기 위해서는 웹마스터 혹은 웹 기획자가 동적인 구성요소의 특성을 잘 파악해서 기획을 해야 한다. 예를 들어 최신 뉴스의 경우 필드에서 기사를 작성해서 온라인 상으로 매 시간 기사를 입력하고 있다. 이렇게 매시각 변화하는 기사내용을 캐시서버가 어느 정도 캐싱을 하고 있을 것인가에 대한 판단이 필요하게 되는 것이다. 이를 위해서는 일정기간 기사가 입력되는 주기를 살펴보고 패턴을 찾아야 할 것이다. 특히 스포츠 경기의 점수의 경우에는 언제 누가 점수를 낼지 모르는 상황이다. 이런 경우에는 캐싱의 주기는 좀더 짧아져야 할 것이다.

날씨의 경우는 갑작스럽게 기온이나 날씨가 변화하는 경우도 있겠으나, 일반적으로는 일정시간을 유지하고 변화가 서서히 일어나므로 캐싱 주기를 다른 것에 비해서 약간 길게 설정해도 무방할 것이다. 이렇듯 콘텐츠의 내용과 특성에 따라 웹마스터나 웹 기획자는 콘텐츠가 캐싱 서버에 저장되는 주기를 결정해야 하며, 사이트 개발자 역시 사이트 구축 시 웹마스터나 웹 기획자가 결정한 주기와 관련된 태그를 설정해야 한다.

이런 페이지 구성요소 단위별 동적 캐싱을 지원하기 위해서는 기존 사이트 구축시보다 구축단계에서 좀 더 많은 손길이 필요하겠지만, 캐싱을 효과적으로 이용함으로 인해 하드웨어나 소프트웨어의 자원을 절감할 수 있다. 🌐

#### 참 고 사 이 트

· Persistence Danami (<http://www.persistence.com/>) : 동적 캐싱 솔루션 업체  
· Chulney Technology (<http://www.chulneytech.com/>) : 동적 캐시 솔루션 업체  
· NetFusion 기사 (<http://www.nwfusion.com/news/2002/0930specialfocus.html>)

· Spider Cache (<http://www.spidercache.com/>) : 동적 캐시 솔루션 업체  
· ESI - Edge Side Include (<http://www.esi.org>) : 동적캐시 표준화 모임