



(12) 发明专利申请

(10) 申请公布号 CN 102387220 A

(43) 申请公布日 2012. 03. 21

(21) 申请号 201110432930. 0

(22) 申请日 2011. 12. 22

(71) 申请人 乐视网信息技术(北京)股份有限公司

地址 100026 北京市朝阳区光华路4号东方梅地亚中心C座8层

(72) 发明人 李茗 邵长松

(74) 专利代理机构 北京驰纳智财知识产权代理事务所(普通合伙) 11367

代理人 谢亮 唐与芬

(51) Int. Cl.

H04L 29/08(2006. 01)

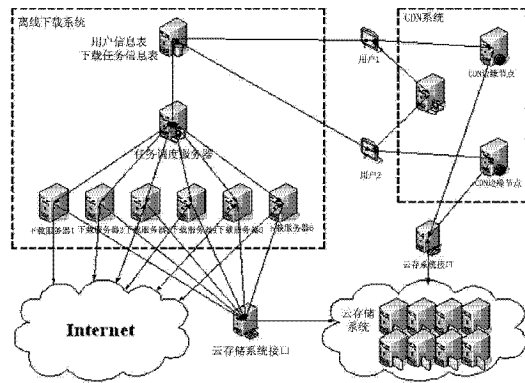
权利要求书 1 页 说明书 5 页 附图 1 页

(54) 发明名称

一种基于云存储的离线下载的方法及其系统

(57) 摘要

本发明提供的一种基于云存储的离线下载的方法及其相应的系统替代用户完成实际的下载工作,下载的文件可以永久存储在离线下载系统本地,即便将来该文件的源URL失效,用户仍然可以通过将下载请求提交到离线下载系统来获取到该文件。这就保证了下载的永久有效性,并且系统侧的工作流程对于用户是完全透明的。此外,本发明提供的基于云存储的一种离线下载的方法及其相应的系统完成实际下载工作后,通知用户,用户使用与离线下载系统之间的高速通道将文件下载到自己的本地计算机中,这就保证了数据传输的高速性。



1. 一种基于云存储的离线下载系统,其特征在于,所述系统包括:
用户系统,分布式下载系统,云存储系统和 CDN 内容分发系统,其中,
所述用户系统包括用户信息列表和下载任务信息列表,所述用户信息列表记录每个注册用户的信息,所述下载任务信息列表记录每个下载任务的信息;
所述分布式下载系统包括任务调度服务器和下载服务器集群,所述任务调度服务器实时记录所述下载服务器集群的负载状态,所述下载服务器集群中的每台下载服务器执行实际的下载任务;
所述云存储系统通过云存储系统接口与其它系统进行数据通讯,并且所述每个注册用户在所述云存储系统中都有其相应的私有存储空间;
所述 CDN 内容分发系统包括 CDN 调度服务器和 CDN 边缘节点,所述 CDN 边缘节点通过所述云存储系统接口与所述云存储系统进行数据通讯,所述 CDN 调度服务器与每个用户进行数据通讯。
2. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的注册用户的信息包括用户唯一的标识符。
3. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的注册用户的信息还包括所述云存储系统中的用户私有空间存储路径。
4. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的注册用户的信息还包括所述云存储系统中的用户私有空间总计容量的大小。
5. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的注册用户的信息还包括所述云存储系统中的用户私有空间当前剩余容量大小。
6. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的下载任务的信息包括提交所述下载任务的用户的唯一的标识符。
7. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的下载任务的信息还包括所述下载任务的下载地址的链接。
8. 一种如权利要求 1 所述的基于云存储的离线下载系统,其特征在于,所述记录的下载任务的信息还包括所述下载任务的下载状态。
9. 一种如权利要求 8 所述的基于云存储的离线下载系统,其特征在于,所述下载任务的下载状态包括未下载的状态。
10. 一种如权利要求 8 所述的基于云存储的离线下载系统,其特征在于,所述下载任务的下载状态还包括正在下载中的状态。

一种基于云存储的离线下载的方法及其系统

技术领域

[0001] 本发明涉及通信技术领域,更具体地,涉及一种基于云存储的离线下载的方法及其系统。

背景技术

[0002] 下载是将网络上计算机中的数据复制到本地计算机的行为。文件下载在个人电脑上是很成熟的技术,用户从网络上下载自己感兴趣的文件并保存到个人电脑。互联网中的下载技术主要有三种:服务端下载 P2S,点对点下载 P2P 以及智能网格 P2SP。

[0003] P2S 是指用户计算机向服务器请求下载,并通过协议将服务器上的数据复制到用户计算机的下载技术。P2S 包括 HTTP 和 FTP 两种协议。

[0004] P2P 下载方式中,所有参与下载的用户计算机同时也充当提供下载的角色,即同时也上传。对于一个完整的资源,可包含很多数据块。当很多用户下载同一个资源时,获得的数据块都不尽相同,因此每一个用户计算机都可以将自己已有的数据提供给其他没有该数据的计算机下载。在网络带宽足够的情况下,当参与下载的计算机越多时,每一个计算机下载该资源的速度就越快。

[0005] P2SP 融合了 P2S 和 P2P 两种下载方式的优点,通过多媒体检索数据库这个桥梁把原本孤立的服务器和其镜像资源和 P2P 资源整合到了一起。这样下载速度更快,同时下载资源更丰富,下载稳定性更强。

[0006] 然而,不管是哪一种下载方式,传统的下载方式由于网络环境的错综复杂,在下载过程中往往难以保持持续的高速下载,耗费时间较长。另外,如果一个下载 URL 链接失效,则很可能再也无法下载到想要的文件,不能保证下载链接的永久有效。

[0007] 而且个人电脑一般是固定的,不可以随时随地上网,随着技术的进步,移动通讯设备终端已经具有浏览网页的功能,更多的用户通过移动通讯设备终端随时随地访问网站。但是,通过移动通讯设备终端下载文件,仍然存在很大的问题。其中最大的问题在于移动通讯设备终端是通过窄带与网络连接的,因此下载文件时速度较慢。同时,在现今的通讯环境下,通过移动通讯网络访问网站,用户下载所产生的流量是直接和话费挂钩的,如果下载的文件较大,则所需要的费用和金钱就比较多。

发明内容

[0008] 本发明要解决的技术问题下载的高速性和待下载文件的永久有效性,以提高下载的用户体验度。为了解决上述问题,本发明提供了一种基于云存储的离线下载的方法及其相应的系统,其中服务器端建立专用的离线下载系统,用户想要下载一个文件时,只要将下载请求提交到离线下载系统即可,由离线下载系统完成时间的下载工作。

[0009] 本发明提供一种基于云存储的离线下载系统,包括用户系统,分布式下载系统,云存储系统和 CDN 系统,其中,所述用户系统包括用户信息列表和下载任务信息列表。

[0010] 所述用户信息列表记录每个注册用户的信息,所述记录的注册用户的信息包括

用户唯一的标识符,所述云存储系统中的用户私有空间存储路径,所述云存储系统中的用户私有空间总计容量的大小,所述云存储系统中的用户私有空间当前剩余容量大小。

[0011] 所述下载任务信息列表记录每个下载任务的信息,所述记录的下载任务的信息包括提交所述下载任务的用户的唯一的标识符,所述下载任务的下载地址的链接,所述下载任务的下载状态。

[0012] 其中所述下载状态包括未下载,正在下载中,下载完成以及下载失败。

[0013] 所述分布式下载系统包括任务调度服务器和下载服务器集群。其中,

所述任务调度服务器实时记录所述下载服务器集群的负载状态,所述下载服务器集群中的每台下载服务器执行实际的下载任务。

[0014] 所述云存储系统通过云存储系统接口与其它系统进行数据通讯。其中,所述云存储系统采用 GFS 架构,所述每个注册用户在所述云存储系统中都有其相应的私有存储空间。

[0015] 所述 CDN 内容分发系统包括 CDN 调度服务器和 CDN 边缘节点,其中,所述 CDN 边缘节点通过所述云存储系统接口与所述云存储系统进行数据通讯,所述 CDN 调度服务器与每个用户进行数据通讯。

[0016] 此外,本发明还提供一种基于云存储的离线下载方法,其特征在于,包括:

用户系统接收用户提交到所述用户系统中的下载请求;

用户系统将所述接收到的下载请求转换为下载任务,并将所述下载任务存储到下载任务信息列表中;

任务调度服务器从所述用户系统中取得新的下载任务,并将取得的新的下载任务分配给下载服务器;

任务调度服务器将所述用户系统中的所述下载任务信息列表中将已经分配给所述下载服务器的下载任务的下载任务的任务状态更新为“正在下载中”;

所述下载服务器执行实际的下载工作;

所述下载服务器完成指定的下载工作后将下载的文件存储到云存储系统中;

所述下载服务器完成指定的下载工作后向所述任务调度服务器发送下载完成的信息;

所述任务调度服务器在接收下载完成的信息后,将所述用户系统中的所述下载任务信息列表中将已经完成的下载任务的任务状态更新为“下载完成”;

用户通过用户系统获得请求下载的文件。

[0017] 优选的是,所述任务调度服务器根据取得的所述下载任务,获得待下载文件的大小,所述任务调度服务器将待下载文件从逻辑上分为 N 个分片,从所述下载服务器集群中选择 N 台当前负载任务相对最少的下载服务器;

优选的是,所述 N 台下载服务器分别接收所述任务调度服务器分配的对应的 N 个下载任务;

优选的是,所述任务调度服务器在接收所述 N 片下载任务全部完成后,将所述用户系统中的所述下载任务信息列表中将已经完成的下载任务的任务状态更新为“下载完成”;

优选的是,所述用户系统将已经下载完成的文件的 ID 和 CDN 调度地址发送给用户,所述用户根据所述 CDN 调度地址访问所述 CDN 调度服务器,所述 CDN 调度服务器将符合条件

的 CDN 边缘节点地址发送给所述用户；

优选的是，所述用户将所述已经下载完成的文件的 ID 发送给接收到的所述 CDN 边缘节点获得请求下载的文件。

[0018] 优选的是，所述 CDN 边缘节点通过所述云存储系统的接口从云存储系统中下载所述请求下载的文件，并在下载任务完成后将所述下载文件缓存中所述 CDN 边缘节点内。

[0019] 与现有技术相比，本发明的优点在于：本发明提供的一种离线下载的方法及其相应的系统替代用户完成实际的下载工作，下载的文件可以永久存储在离线下载系统本地，即便将来该文件的源 URL 失效，用户仍然可以通过将下载请求提交到离线下载系统来获取到该文件。这就保证了下载的永久有效性，并且系统侧的工作流程对于用户是完全透明的。此外，本发明提供的一种离线下载的方法及其相应的系统完成实际下载工作后，通知用户，用户使用与离线下载系统之间的高速通道将文件下载到自己的 PC，这就保证了数据传输的高速性。

附图说明

[0020] 为了使本发明便于理解，现在结合附图描述本发明的具体实施例。

[0021] 图 1 为本发明中一种基于云存储的离线下载系统整体结构图。

具体实施方式

[0022] 下面结合附图和优选的实施方式对本发明作进一步详细描述。

[0023] 本发明在服务器一侧建立专用的离线下载系统，用户想要下载一个文件时，只要将下载请求提交到离线下载系统即可，由离线下载系统完成实际的下载工作。离线下载系统完成下载工作后，将下载下来的文件存储在离线下载系统本地，并通知用户下载已完成，可以到离线下载系统获取。用户使用与离线下载系统之间的高速传输通道将文件下载到自己的本地计算机中。

[0024] 特别地，当离线下载系统完成一次下载过程后，将已存储的下载文件与下载 URL 建立关联，并将该关联关系缓存。这样，当其他用户再次提交针对该 URL 的下载请求时，离线下载系统可以发现需要下载的文件已经存储在系统本地，则直接告知用户下载已完成，用户使用高速传输通道将文件下载到自己的本地计算机中。

[0025] 本发明提供的基于云存储的离线下载系统的用户系统中，维护着两张表：用户信息列表和下载任务信息列表。其中，

用户信息列表是对于每一个注册用户，在云存储系统中均为其分配一定大小的私有存储空间。用户信息列表中记录着所有注册用户的信息，每一个注册用户的信息为以下四元组：

<UserID、StorePath、TotalCapacity、RemainCapacity>

UserID：用户唯一标识符；

StorePath：云存储系统中的用户私有空间存储路径；

TotalCapacity：云存储系统中的用户私有空间总计容量大小；

RemainCapacity：云存储系统中的用户私有空间当前剩余容量大小。

[0026] 下载任务信息列表是用户系统接受用户提交的下载请求后，将请求转换成为一个

下载任务存储在任务信息列表里。每一个下载任务的信息为如下三元组：

<UserID、URL、State>

UserID :提交该下载任务的用户唯一标识；

URL :下载地址链接；

State :下载任务的状态。下载任务在整个下载过程中分为四个状态：“未下载”、“下载中”、“下载完成”、“下载失败”。

[0027] 本发明提供的基于云存储的离线下载系统的分布式下载系统由单台任务调度服务器和下载服务器集群组成。下载服务器集群中的每台下载服务器执行实际的下载工作，且每台下载服务器均有承担的下载任务的上限。任务调度服务器实时记录着每台下载服务器的负载状况。

[0028] 任务调度服务器定时从用户系统下载任务信息表中取得一个状态为“未下载”的任务，根据下载地址链接 URL，获取到待下载文件的大小，从逻辑上将该文件划分为 N 个文件块，根据记录的当前各台下载服务器的负载状况，选择负载相对最轻(当前承担的下载任务相对最少)的 N 台下载服务器，将取得的任务指派给此 N 台下载服务器，每台下载服务器负责一个文件块(分片)的实际下载。任务指派完毕后，调度服务器在用户系统任务信息表里将该下载任务的状态更新为“下载中”。下载服务器完成分片的实际下载后，将结果反馈给调度服务器。只有所有分片的下载都成功时，调度服务器才将任务信息表里的下载任务状态更新为“下载完成”，否则将下载任务标记为“下载失败”。

[0029] 下载服务器集群中的每台下载服务器负责实际的下载工作。对于一个下载任务，一台下载服务器负责待下载文件一个逻辑分片的下载；根据任务调度服务器的实时调度，单台下载服务器可以承担多个下载任务，但任务总数不可超过预先设定的上限。下载服务器完成每一个任务的逻辑分片的下载后，通知任务调度服务器该分片的下载已完成，并使用云存储系统提供的接口，将下载完成的数据写入云存储系统中相应的用户私有存储空间中。

[0030] 本发明提供的基于云存储的离线下载系统的云存储系统是文件存储的仓库，离线下载系统下载下来的文件，都存储在云存储系统中。云存储系统使用 GFS 架构，内部有着优良的备份机制、容灾机制，这些均对外界透明，本发明不涉及云存储系统的内部机制。外界系统使用云存储系统提供的 I/O 接口与其进行数据读、写交互。

[0031] 本发明提供的基于云存储的离线下载系统的 CDN 内容分发系统将下载下来的文件，分发至距离用户最近的 CDN 边缘节点，使用户就近访问，因为所述云存储系统从内部看是分布式的存储集群，但其分布式的特点，是为了满足备份与容灾，并不具备“分发”的性质，因此从外部看，云存储系统仍然是集中式存储，需要 CDN 系统的支持。

[0032] 本发明的一个优选的离线下载的流程如下：

用户将下载请求提交到离线下载系统中的用户系统，假设下载地址链接为 URL。用户系统将下载请求转换为一个下载任务，存储在下载任务信息表中。

[0033] 离线下载系统中的任务调度服务器从用户系统中取得新的下载任务，通过下载链接 URL 获得待下载文件的大小，并从逻辑上划分为 N 个分片，根据自身的调度机制，从下载服务器集群中选择 N 台当前负载相对最少的下载服务器，分别将 N 个分片的下载任务指派给这 N 台下载服务器，并将用户系统下载任务信息表中的任务状态更新为“下载中”。

[0034] N 台下载服务器执行实际的下载工作,每台下载服务器完成指定分片的下载后,通过云存储系统接口,将分片数据写入云存储系统中,并向任务调度服务器反馈下载结果。

[0035] 全部分片下载成功后,用户需要下载的文件也即存储在了云存储系统中。任务调度服务器更新下载任务信息表中的任务状态为“下载完成”。

[0036] 用户系统将已下载文件的 ID 和 CDN 调度地址提供给用户 1,用户 1 根据调度地址访问 CDN 调度服务器,请求距离自己最近、负载状况最理想的一台 CDN 边缘节点,CDN 调度服务器将符合条件的一台 CDN 边缘节点的地址告知用户。

[0037] 用户将已下载文件的 ID 发送至指定的 CDN 边缘节点,索要该文件。如果 CDN 边缘节点尚无该文件的缓存,则通过高速通道,利用云存储系统的接口从云存储系统中下载该文件,下载完后先缓存在自己本地,再推送至用户。此时整个离线下载过程完成,用户获得了想要的下载的文件。

[0038] 以上结合本发明的具体实施例做了详细描述,但并非是对本发明的限制。凡是依据本发明的技术实质对以上实施例所做的任何简单修改,均仍属于本发明技术方案的范围。

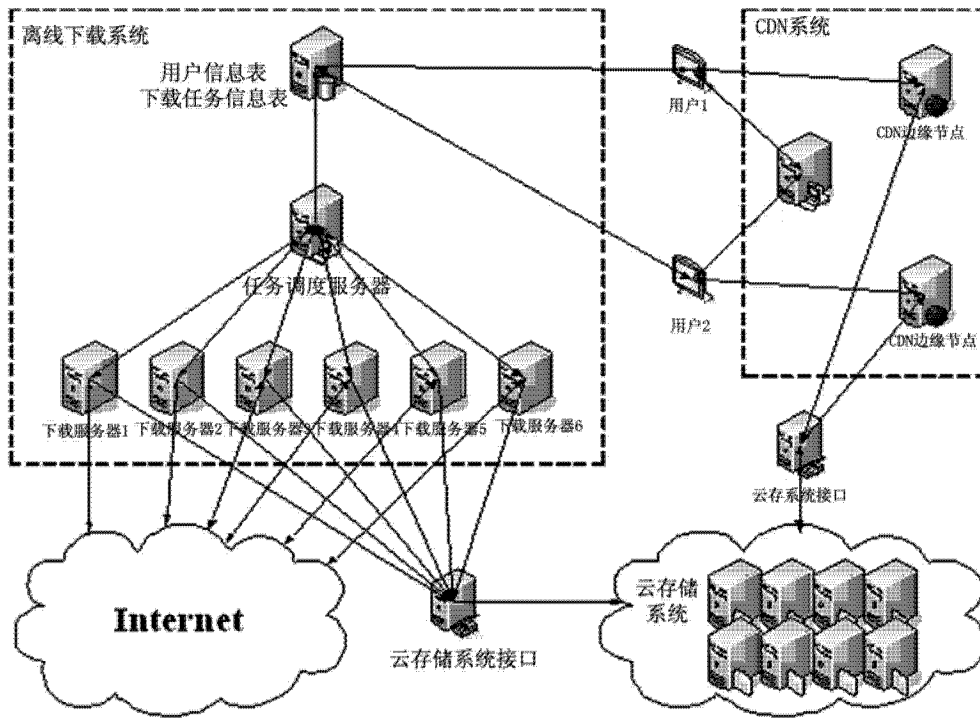


图 1