

## (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2014 年 1 月 30 日 (30.01.2014)



(10) 国际公布号  
WO 2014/015782 A1

- (51) 国际专利分类号:  
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2013/079855
- (22) 国际申请日: 2013 年 7 月 23 日 (23.07.2013)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201210261331.1 2012 年 7 月 26 日 (26.07.2012) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。
- (72) 发明人: 伍海君 (WU, Haijun); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。朱会灿 (ZHU, Kuaican); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。邓大付 (DENG, Dafu); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。李锐 (LI, Rui); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。邹永强 (ZOU, Yongqiang); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。董乘宇 (DONG, Shengyu); 中国广东省深圳市福田区

振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。阙太富 (QUE, Taifu); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。王磊 (WANG, Lei); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。杨绍鹏 (YANG, Shaopeng); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。张书鑫 (ZHANG, Shuxin); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。赵大勇 (ZHAO, Dayong); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。刘畅 (LIU, Chang); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。陈晓东 (CHEN, Xiaodong); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。张银锋 (ZHANG, Yin-feng); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。

- (74) 代理人: 北京德琦知识产权代理有限公司 (DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路 1 号学院国际大厦 7 层, Beijing 100083 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR,

[见续页]

(54) Title: DISTRIBUTED FILE SYSTEM, FILE ACCESSING METHOD, AND CLIENT

(54) 发明名称: 分布式文件系统、文件访问方法以及客户端

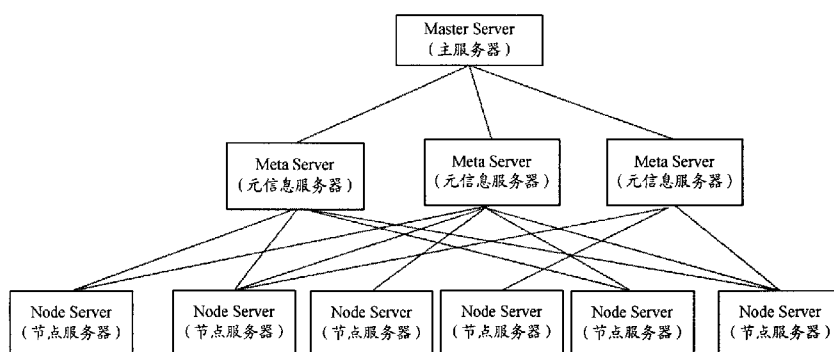


图 1 / FIG.1

(57) Abstract: Embodiments of this application provide a distributed file system, a file accessing method, and a client. The method comprises: accessing a file directory stored by a master server, and obtaining route information, of a meta server associated with a to-be-accessed file, from the master server; and accessing the meta server by using the obtained route information, obtaining meta information of the to-be-accessed file from the meta server, and accessing the to-be-accessed file on multiple node servers by using the obtained meta information.

(57) 摘要: 本申请实施例提供了分布式文件系统、文件访问方法以及客户端。其中, 该方法包括: 访问主服务器 Master Server 存储的文件目录, 并从 Master Server 获取与需要访问的文件相关联的元信息服务器 Meta Server 的路由信息; 利用获取的路由信息访问 Meta Server, 并从该 Meta Server 中获取需要访问的文件的元信息, 利用获取的元信息从多个节点服务器 Node Server 上访问需要访问的文件。

WO 2014/015782 A1



CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

- (84) **指定国** (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,

AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

**本国际公布:**

- 包括国际检索报告(条约第 21 条(3))。

## 分布式文件系统、文件访问方法以及客户端

### 技术领域

本申请涉及数据存储技术，特别涉及分布式文件系统、文件访问方法以及客户端。

### 5 发明背景

目前，在工业界具有典型代表性的分布式文件系统是 Google 公司的分布式文件系统，简称 GFS（Global File System）。该 GFS 主要由一个主服务器（Master Server）和众多块服务器（Chunk Server）组成。其中，Master Server 主要负责存储文件目录、以及文件目录中各文件的元（meta）信息。这里，meta 信息主要包括：文件的大小、文件被分成的数据块的数量、文件被分成的数据块所处的 Chunk Server 等。Chunk Server 主要负责存储文件被分成的数据块。通常，一个文件会按照一定大小划分为多个数据块，每一个数据块称为一个 Chunk，这些数据块分散存储在不同的 Chunk Server。

15 但是，由于 GFS 只有一个 Master Server 提供文件目录和文件 meta 信息的查询，这会限制文件的并发访问量，并且，因单一 Master Server 的内存限制，也会限制 GFS 存储的文件数目。

### 发明内容

本申请实施例提供了分布式文件系统、文件访问方法以及客户端，  
20 以提高单一集群（cluster）的文件数目以及文件的并发访问量。

本申请实施例提供的技术方案包括：

一种分布式文件系统，包括：

主服务器 Master Server，用于存储文件目录、以及与文件目录中各文件相关联的 Meta Server 的路由信息；在存储的文件目录中存在客户端需要访问的文件时，从存储的路由信息中找到与该文件相关联的 Meta Server 的路由信息并提供给客户端，以使客户端利用所述 Master Server 提供的路由信息访问 Meta Server；

Meta Server，用于存储相关联的文件的元信息，Meta Server 的数量大于等于 1；在接收到客户端的访问时，提供该客户端需要访问的文件的元信息给客户端，以使客户端利用所述 Meta Server 提供的元信息从 Node Server 访问需要访问的文件。

节点服务器 Node Server，用于存储文件被划分成的数据块和/或数据块的备份，Node Server 的数量大于等于 1。

一种文件访问方法，包括：

访问主服务器 Master Server 存储的文件目录，并从 Master Server 获取与需要访问的文件相关联的元信息服务器 Meta Server 的路由信息；

利用获取的路由信息访问 Meta Server，并从该 Meta Server 中获取需要访问的文件的元信息；

利用获取的元信息从多个节点服务器 Node Server 上访问需要访问的文件。

一种应用于文件访问的客户端，包括：

第一访问单元，用于访问主服务器 Master Server 存储的文件目录，并从 Master Server 获取与本客户端需要访问的文件相关联的元信息服务器 Meta Server 的路由信息；

第二访问单元，用于利用所述第一访问单元获取的路由信息访问 Meta Server，并从该 Meta Server 中获取本客户端需要访问的文件的元信息；

第三访问单元，用于利用获取的元信息从多个节点服务器 Node Server 上访问需要访问的文件。

由以上技术方案可以看出，在本发明实施例中，文件目录和文件元信息分开存储，即，客户端只在 Master Server 访问文件目录和与文件目录中各文件相关联的 Meta Server 的路由信息，而去 Meta Server 访问具体文件的元信息，这相比单个 Master Server 既要提供文件目录访问，又要提供文件元信息的访问而言，能提供更高的每秒访问次数（QPS），也能提供更高的并发访问量，同时由于本发明实施例中 Master Server 只存储文件目录，这能保证本发明实施例中的分布式文件系统存储更多的文件。

### 附图简要说明

图 1 为本发明实施例提供的分布式文件系统结构示意图；

图 2 为本发明实施例提供的文件访问方法流程示意图；

图 3 为本发明实施例提供的客户端结构示意图；

图 4 为本发明另一实施例提供的客户端结构示意图。

### 实施本发明的方式

为了使本发明的目的、技术方案和优点更加清楚，下面结合附图和具体实施例对本发明进行详细描述。

本发明实施例提供的分布式文件系统具有图 1 所示的结构，其包括：一个主服务器（Master Server）、至少一个元信息服务器（Meta Server）、以及至少一个节点服务器（Node Server）。其中，Meta Server 以及 Node Server 的数量可根据集群规模设置，本发明实施例并不具体限定。

从图 1 所示的分布式系统可以看出，本发明实施例提供的分布式系

统采用了三层架构，其中，最上层为 Master Server，中间层为至少一个 Meta Server，最下层为至少一个 Node Server，基于此，本发明实施例提供的分布式系统可称为三层架构的分布式文件系统。

另外，由于本发明实施例提供的分布式文件系统中，Meta Server 以及 Node Server 的数量可根据集群规模设置，当根据需求需要扩展集群规模时，该 Meta Server 以及 Node Server 的数量也需要扩展，基于此，还可将本发明实施例提供的分布式系统称为扩展的分布式文件系统，简称 XFS（eXtensible File System）。

正常情况下，文件元信息的存储量会比文件目录的存储量大很多，为了扩展分布式文件系统，本发明实施例将文件目录和文件元信息分开存储，具体为：将文件目录单独用 Master Server 存储，而将文件元信息用 Meta Server 存储，其中，为了将文件目录中的文件与 Meta Server 存储的该文件的元信息关联起来，还需要在 Master Server 中存储与文件目录中各文件相关联的 Meta Server 的路由信息。

下面对图 1 所示分布式文件系统中各个功能单元进行描述：

Master Server，用于存储文件目录、以及与文件目录中各文件相关联的 Meta Server 的路由信息；

而每一 Meta Server，用于存储与其相关联的文件的元信息，所述文件的元信息至少包括：文件的长度、文件被划分成的数据块的数量、文件被划分成的各数据块以及各数据块的备份所处的 Node Server；优选地，本发明实施例中，所述文件的元信息还可进一步包括：文件创建时间、文件创建者、文件被划分成的各个数据块的摘要信息等，本发明实施例并不具体限定。

而每一 Node Server，用于存储数据块和/或其它数据块的备份。

其中，每一 Node Server 可以仅存储同一文件被划分成的一个或多个

数据块，但禁止同时存储同一文件被划分成的某一数据块和该数据块的备份。也就是说，一个数据块不能和其备份同时存储在同一 Node Server 上。

以图 1 所示分布式文件系统为例，则，针对 Master Server 存储的文件目录中的文件（记为文件 1），如果该文件 1 被划分成 5 个数据块，则为了提高系统的容错能力，还需要针对这 5 个数据块进行备份，基于上面描述的 Node Server 存储数据块及数据块的备份的特点，在本发明实施例中，可将该 5 个数据块和其备份分散放在不同的 Node Server 上即可。其中，文件 1 如何被划分成数据块可依据现有划分方式，本发明不再赘述。

在本发明实施例中，一个数据块可存在多个备份，针对一个数据块的多个备份，为了提高系统容错能力，优选地，本发明实施例并不将该多个备份同时存放在同一个 Node Server，而是分散存放在不同的 Node Server 上，即同一数据块的所有备份不存放在同一 Node Server。并且，针对同一文件被划分成的不同数据块的备份，基于容错能力的考虑，也不存放在同一个 Node Server。

基于上述 Master Server、Meta Server 以及 Node Server 存储的信息，则，当客户端需要访问一文件时，Master Server 在存储的文件目录中存在客户端需要访问的文件时，从存储的路由信息中找到与该文件相关联的 Meta Server 的路由信息并提供给客户端，以使客户端利用所述 Master Server 提供的路由信息访问 Meta Server；而当 Meta Server 接收到客户端的访问时，提供该客户端需要访问的文件的元信息给客户端，以使客户端利用所述 Meta Server 提供的元信息访问需要访问的文件。

如此，即可完成客户端的文件访问。可以看出，客户端只在 Master Server 访问文件目录和与文件目录中各文件相关联的 Meta Server 的路由

信息,而去 Meta Server 访问具体文件的元信息,这相比单个 Master Server 既要提供文件目录访问,又要提供文件元信息的访问而言,能提供更高的每秒访问次数(QPS),也能提供更高的并发访问量,同时由于 Master Server 仅存储文件目录,这能够扩展该 Master Server 存储的文件目录,

5 保证本发明实施例提供的分布式文件系统存储更多的文件。

需要说明的是,在本发明实施例中,由于 Master Server 仅存储文件目录和与文件目录中各文件相关联的 Meta Server 的路由信息,并不存储文件的元信息,这相比单个 Master Server 既要提供文件目录,又要提供文件元信息而言,不再因单一 Master Server 的内存有限而限制集群的文件数目,而是可以灵活扩展文件数目,也可相应灵活扩展 Meta Server

10 和 Node Server 的数量。

以根据需求需要扩展 Meta Server 的数量为例,该扩展的每一 Meta Server 与分布式文件系统原有的 Meta Server 功能类似,比如,假如当前扩展的 Meta Server 记为 Server1 和 Server2,以 Server1 为例,Server2

15 原理类似,则,

该 Server1 可以存储与其相关联的文件的元信息,其中,与 Server1 相关联的文件可为 Master Server 的文件目录中已存在的文件,以与 Server1 相关联的文件为 Master Server 的文件目录中已存在的一文件(记为文件 1)为例,则 Server1 存储文件 1 的元信息,其中,Server1 存储文件 1 的元信息,能够为已存储了文件 1 的元信息的 Meta Server 提供备

20 份,以提高系统容错能力。

当然,作为本发明实施例的扩展,与该 Server1 相关联的文件也可以为 Master Server 的文件目录中不存在的一文件,而是根据需求需要扩展的文件。基于此,Server1 就存储该需要扩展的文件的元信息。相应地,

25 Master Server 还用于在文件目录中添加与扩展的 Meta Server 比如

Server1 相关联的文件，以及接收并存储该扩展的 Meta Server 比如 Server1 之间的路由信息。

至于本发明实施例中根据需求扩展的每一 Node Server，其与分布式文件系统已有的 Node Server 功能类似，用于存储文件被划分的数据块和/或数据块的备份。其中，扩展的每一 Node Server 存储的数据块可为 Master Server 的文件目录中已存在的一文件被划分的数据块或数据块备份，也可为新扩展的一文件被划分的数据块或数据块备份，具体如何存放数据块，可依据当前实际情况设定，这里不再赘述。

还有，在本发明实施例中，由于 Master Server 只存储文件目录和与文件目录中各个文件相关联的 Meta Server 的路由信息，而该文件目录和与文件目录中各个文件相关联的 Meta Server 的路由信息占用的存储空间不大，尤其是当文件目录中文件的名称简短比如为简短的数字或字符编码时，这能够保证 Master Server 存储更多的文件目录和与文件目录中各个文件相关联的 Meta Server 的路由信息，扩展集群规模。另外，作为本发明实施例的一个扩展，在本发明实施例中，还可将 Master Server 上的文件目录、以及与文件目录中各文件相关联的 Meta Server 的路由信息存储至其他可快速访问的分布系统上，该分布系统的存储空间远大于 Master Server，这能保证存储更多的文件目录、以及更多与文件目录中各个文件相关联的 Meta Server 的路由信息，整个集群的并发访问能力将进一步提升。

另外，在本发明实施例中，因 Meta Server 的数量并非唯一，这样，若其中一个或多个 Meta Server 出现故障，不影响其他正常 Meta Server，部分文件仍能读写，容错能力更强。

至此，完成图 1 所示分布式文件系统的描述。

下面对本发明实施例提供的方法进行描述：

基于图 1 所示的分布式文件系统，本发明实施例提供了一种文件访问方法。参见图 2，图 2 为本发明实施例提供的文件访问方法流程图。该图 2 所示的流程可由客户端执行，具体如图 2 所示，包括以下步骤：

步骤 201，访问 Master Server 存储的文件目录，并从 Master Server  
5 获取与需要访问的文件相关联的 Meta Server 的路由信息；

步骤 202，利用获取的路由信息访问 Meta Server，并从该 Meta Server 中获取需要访问的文件的元信息。

优选地，本发明实施例中，所述文件的元信息至少包括：文件的长度、文件被划分成的数据块的数量、文件被划分成的各数据块以及各数  
10 据块的备份所处的 Node Server。

步骤 203，利用获取的元信息从多个 Node Server 上访问需要访问的文件。

至此，完成图 2 所示方法的描述。从图 2 所示流程可以看出，客户端只在 Master Server 访问文件目录和与文件目录中各文件相关联的  
15 Meta Server 的路由信息，而去 Meta Server 访问具体文件的元信息，这相比单个 Master Server 既要提供文件目录访问，又要提供文件元信息的访问而言，能提供更高的每秒访问次数（QPS），也能提供更高的并发访问量。

优选地，本发明实施例还提供了一种应用于文件访问的客户端。

20 参见图 3，图 3 为本发明实施例提供的客户端结构图。如图 3 所示，所述客户端包括：

第一访问单元，用于访问 Master Server 存储的文件目录，并从 Master Server 获取与本客户端需要访问的文件相关联的 Meta Server 的路由信息；

25 第二访问单元，用于利用所述第一访问单元获取的路由信息访问

Meta Server, 并从该 Meta Server 中获取本客户端需要访问的文件的元信息。其中, 所述文件的元信息至少包括: 文件的长度、文件被划分成的数据块的数量、文件被划分成的各数据块以及各数据块的备份所处的 Node Server;

- 5        第三访问单元, 用于利用第二访问单元获取的元信息从多个 Node Server 上访问需要访问的文件。

至此, 完成图 3 所示的客户端描述。

- 图 4 为本发明另一实施例提供的客户端结构示意图。如图 4 所示, 该客户端至少包括: 存储器以及与存储器通信的处理器, 其中该存储器  
10        器中包括可由处理器执行的第一访问指令、第二访问指令和第三访问指令。

该第一访问指令, 用于访问主服务器 Master Server 存储的文件目录, 并从 Master Server 获取与本客户端需要访问的文件相关联的元信息服务器 Meta Server 的路由信息;

- 15        第二访问指令, 用于利用所述第一访问指令获取的路由信息访问 Meta Server, 并从该 Meta Server 中获取本客户端需要访问的文件的元信息;

第三访问指令, 用于利用所述第二访问指令获取的元信息从多个节点服务器 Node Server 上访问需要访问的文件。

- 20        优选地, 所述文件的元信息至少包括: 文件的长度、文件被划分成的数据块的数量、文件被划分成的各数据块以及各数据块的备份所处的 Node Server。

- 由以上技术方案可以看出, 在本发明实施例中, 文件目录和文件元信息分开存储, 即, 客户端只在 Master Server 访问文件目录和与文件目  
25        录中各文件相关联的 Meta Server 的路由信息, 而去 Meta Server 访问具

体文件的元信息，这相比单个 Master Server 既要提供文件目录访问，又要提供文件元信息的访问而言，能提供更高的 QPS，也能提供更高的并发访问量。

- 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡
- 5 在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明保护的范围之内。

## 权利要求书

1、一种分布式文件系统，其特征在于，所述分布式文件系统包括：

主服务器 Master Server，用于存储文件目录、以及与文件目录中各文件相关联的元信息服务器 Meta Server 的路由信息；在存储的文件目录  
5 中存在客户端需要访问的文件时，从存储的路由信息中找到与该文件相关联的 Meta Server 的路由信息并提供给客户端，以使客户端利用所述 Master Server 提供的路由信息访问 Meta Server；

Meta Server，用于存储与其相关联的文件的元信息，Meta Server 的数量大于等于 1；在接收到客户端的访问时，提供该客户端需要访问的  
10 文件的元信息给客户端，以使客户端利用所述 Meta Server 提供的元信息从 Node Server 访问需要访问的文件；

节点服务器 Node Server，用于存储文件被划分成的数据块和/或该文件其它数据块的备份，Node Server 的数量大于等于 1。

2、根据权利要求 1 所述的系统，其特征在于，所述文件的元信息至少包括：文件的长度、文件被划分成的数据块的数量、文件被划分成的  
15 各数据块以及各数据块的备份所处的 Node Server。

3、根据权利要求 1 所述的系统，其特征在于，每一 Node Server 禁止同时存储数据块和该数据块的备份；和/或，

每一 Node Server 禁止同时存储同一数据块的所有备份。

20 4、根据权利要求 1 至 3 任一所述的系统，其特征在于，所述分布式文件系统进一步包括扩展的 Meta Server 和/或 Node Server；

Master Server 还用于在文件目录中添加与扩展的 Meta Server 相关联的文件，以及接收并存储该扩展的 Meta Server 的路由信息；

扩展的每一 Meta Server 用于存储与其相关联的文件的元信息；

扩展的每一 Node Server 用于存放数据块和/或其它数据块的备份。

5、一种文件访问方法，其特征在于，该方法包括：

访问主服务器 Master Server 存储的文件目录，并从 Master Server 获取与需要访问的文件相关联的元信息服务器 Meta Server 的路由信息；

5 利用获取的路由信息访问 Meta Server，并从该 Meta Server 中获取需要访问的文件的元信息；

利用获取的元信息从多个节点服务器 Node Server 上访问需要访问的文件。

6、根据权利要求 5 所述的方法，其特征在于，所述文件的元信息至少包括：文件的长度、文件被划分成的数据块的数量、文件被划分成的各数据块以及各数据块的备份所处的 Node Server。

7、根据权利要求 5 所述的方法，其特征在于，每一 Node Server 禁止同时存储数据块和该数据块的备份；和/或，

每一 Node Server 禁止同时存储同一数据块的所有备份。

15 8、一种应用于文件访问的客户端，其特征在于，所述客户端包括：

第一访问单元，用于访问主服务器 Master Server 存储的文件目录，并从 Master Server 获取与本客户端需要访问的文件相关联的元信息服务器 Meta Server 的路由信息；

第二访问单元，用于利用所述第一访问单元获取的路由信息访问 Meta Server，并从该 Meta Server 中获取本客户端需要访问的文件的元信息；

第三访问单元，用于利用所述第二访问单元获取的元信息从多个节点服务器 Node Server 上访问需要访问的文件。

9、根据权利要求 8 所述的客户端，其特征在于，所述文件的元信息至少包括：文件的长度、文件被划分成的数据块的数量、文件被划分成

的各数据块以及各数据块的备份所处的 Node Server。

1/2

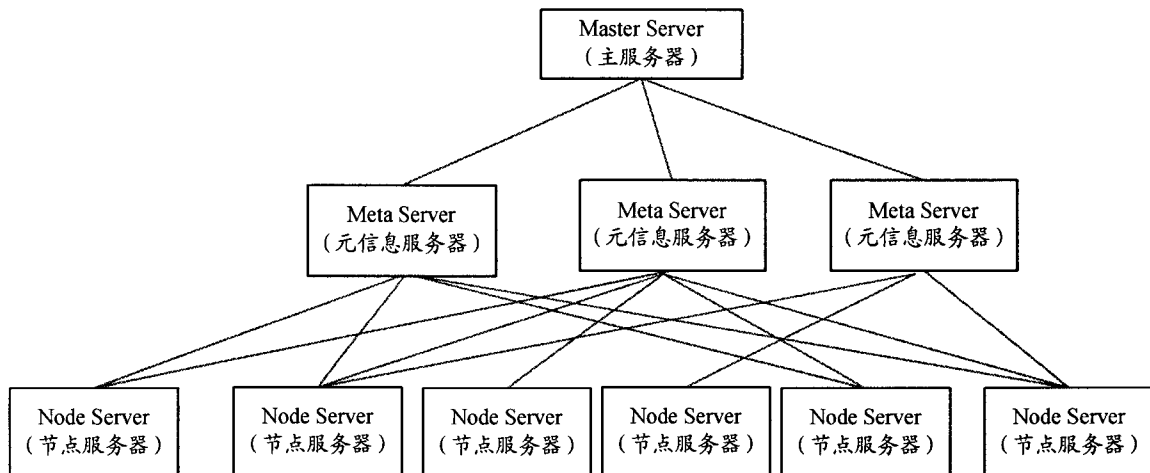


图 1

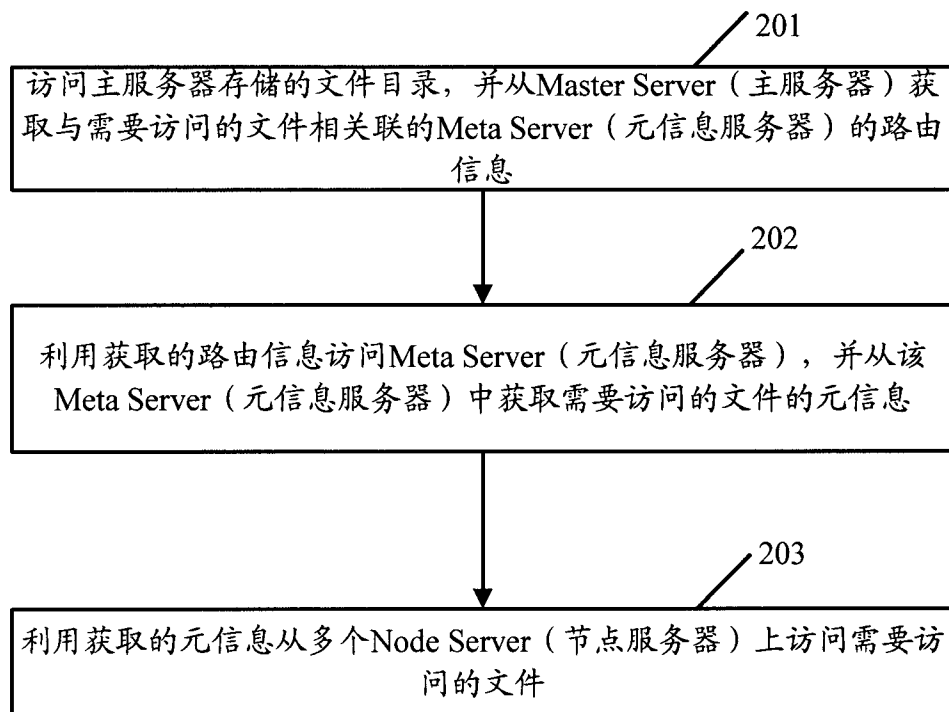


图 2

2/2

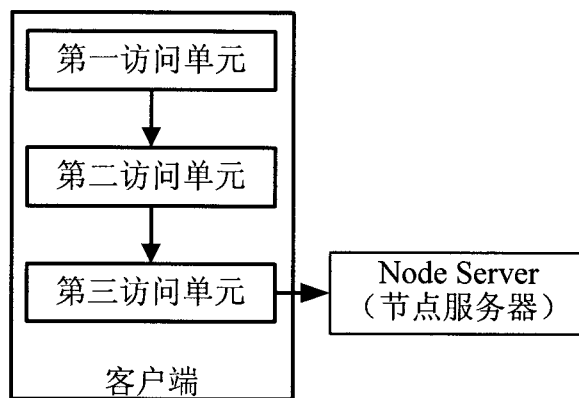


图 3

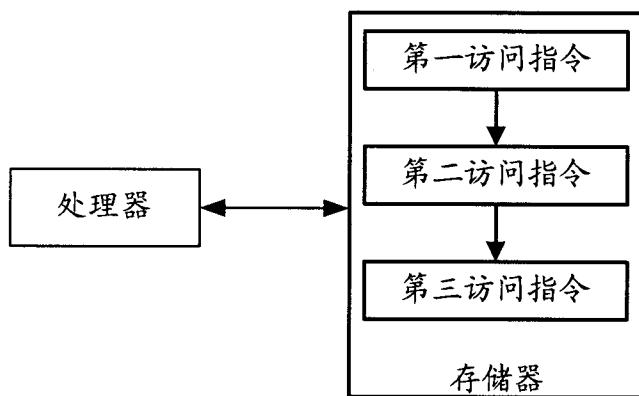


图 4

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2013/079855****A. CLASSIFICATION OF SUBJECT MATTER**

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F 17

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Data bases: CNPAT, CNKI, WPI, EPODOC, GOOGLE

Search terms: distributed file system, GFS, metainformation, global file system, server, master, meta, route, chunk, slave, secondary

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                              | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | CN 101997823 A (LENOVO (BEIJING) CO., LTD.), 30 March 2011 (30.03.2011), description, paragraphs [0002]-[0096], and figures 1-7 | 1-9                   |
| A         | CN 101576915 B (PEKING UNIVERSITY), 08 June 2011 (08.06.2011), the whole document                                               | 1-9                   |
| A         | CN 102420854 A (XIDIAN UNIVERSITY), 18 April 2012 (18.04.2012), the whole document                                              | 1-9                   |
| A         | WO 2010005460 A1 (AVERE SYSTEMS INC.), 14 January 2010 (14.01.2010), the whole document                                         | 1-9                   |
| A         | US 7487228 B1 (RED HAT INC.), 03 February 2009 (03.02.2009), the whole document                                                 | 1-9                   |

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention                                              |
| "A" document defining the general state of the art which is not considered to be of particular relevance                                                                | "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| "E" earlier application or patent but published on or after the international filing date                                                                               | "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" document member of the same patent family                                                                                                                                                                                                    |
| "O" document referring to an oral disclosure, use, exhibition or other means                                                                                            |                                                                                                                                                                                                                                                  |
| "P" document published prior to the international filing date but later than the priority date claimed                                                                  |                                                                                                                                                                                                                                                  |

Date of the actual completion of the international search  
30 September 2013 (30.09.2013)Date of mailing of the international search report  
**31 October 2013 (31.10.2013)**Name and mailing address of the ISA/CN:  
State Intellectual Property Office of the P. R. China  
No. 6, Xitucheng Road, Jimenqiao  
Haidian District, Beijing 100088, China  
Facsimile No.: (86-10) 62019451Authorized officer  
**ZHANG, Xia**  
Telephone No.: (86-10) **61648105**

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

**PCT/CN2013/079855**

| Patent Documents referred<br>in the Report | Publication Date | Patent Family    | Publication Date |
|--------------------------------------------|------------------|------------------|------------------|
| CN 101997823 A                             | 30.03.2011       | None             |                  |
| CN 101576915 B                             | 08.06.2011       | None             |                  |
| CN 102420854 A                             | 18.04.2012       | None             |                  |
| WO 2010005460 A1                           | 14.01.2010       | US 2010011037 A1 | 14.01.2010       |
|                                            |                  | US 2011282922 A1 | 17.11.2011       |
|                                            |                  | US 2012271858 A1 | 25.10.2012       |
| US 7487228 B1                              | 03.02.2009       | None             |                  |

**A. 主题的分类**

G06F 17/30 (2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类

**B. 检索领域**

检索的最低限度文献(标明分类系统和分类号)

IPC: G06F 17

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

数据库: CNPAT,CNKI,WPI,EPODOC,GOOGLE

检索词: 分布式文件系统, GFS, 服务器, 主, 从属, 元数据, 元信息, 路由, global file system, server, master, meta, route, chunk, slave, secondary

**C. 相关文件**

| 类 型* | 引用文件, 必要时, 指明相关段落                                                         | 相关的权利要求 |
|------|---------------------------------------------------------------------------|---------|
| X    | CN101997823A(联想(北京)有限公司)30.3 月 2011(30.03.2011) 说明书第[0002]-[0096]段, 图 1-7 | 1-9     |
| A    | CN101576915B(北京大学)08.6 月 2011(08.06.2011)全文                               | 1-9     |
| A    | CN102420854A(西安电子科技大学)18.4 月 2012(18.04.2012)全文                           | 1-9     |
| A    | WO2010005460A1(AVERE SYSTEMS INC.)14.1 月 2010(14.01.2010)全文               | 1-9     |
| A    | US7487228B1(RED HAT INC.)03.2 月 2009(03.02.2009)全文                        | 1-9     |



其余文件在 C 栏的续页中列出。



见同族专利附件。

\* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&amp;” 同族专利的文件

国际检索实际完成的日期

30.9 月 2013(30.09.2013)

国际检索报告邮寄日期

31.10 月 2013 (31.10.2013)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

授权官员

张霞

电话号码: (86-10) 61648105

**国际检索报告**  
关于同族专利的信息

国际申请号  
**PCT/CN2013/079855**

| 检索报告中引用的<br>专利文件 | 公布日期       | 同族专利           | 公布日期       |
|------------------|------------|----------------|------------|
| CN101997823A     | 30.03.2011 | 无              |            |
| CN101576915B     | 08.06.2011 | 无              |            |
| CN102420854A     | 18.04.2012 | 无              |            |
| WO2010005460A1   | 14.01.2010 | US2010011037A1 | 14.01.2010 |
|                  |            | US2011282922A1 | 17.11.2011 |
|                  |            | US2012271858A1 | 25.10.2012 |
| US7487228B1      | 03.02.2009 | 无              |            |