

(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
2019 年 10 月 31 日 (31.10.2019)



(10) 国际公布号
WO 2019/205406 A1

(51) 国际专利分类号:
H04W 28/02 (2009.01)

(21) 国际申请号: PCT/CN2018/104151

(22) 国际申请日: 2018 年 9 月 5 日 (05.09.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810363678.4 2018年4月22日 (22.04.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(72) 发明人: 刘丹(LIU, Dan); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(74) 代理人: 北京市京大律师事务所(BEIJING JINGDA LAW FIRM); 中国北京市西城区车公庄大街甲 4 号 A1706, Beijing 100080 (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, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, SA, SC, SD, SE, SG, SK, SL,

(54) Title: HIGHLY CONCURRENT SERVICE REQUEST PROCESSING METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 高并发业务请求处理方法、装置、计算机设备和存储介质

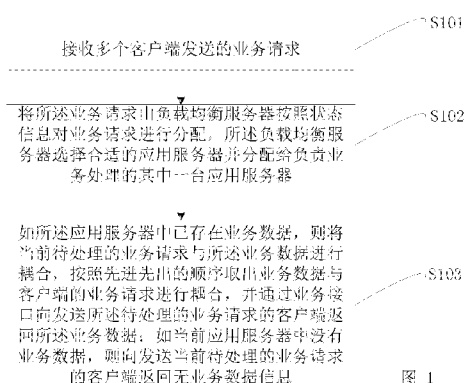


图 1

S101 RECEIVE SERVICE REQUESTS SENT BY MULTIPLE CLIENTS
S102 ALLOCATE THE SERVICE REQUESTS BY MEANS OF A LOAD-BALANCING SERVER ACCORDING TO STATE INFORMATION, WHEREIN THE LOAD-BALANCING SERVER SELECTS APPROPRIATE APPLICATION SERVERS AND ALLOCATES THE SERVICE REQUESTS TO ONE OF THE APPLICATION SERVERS RESPONSIBLE FOR SERVICE PROCESSING
S103 IF THERE IS ALREADY SERVICE DATA IN THE APPLICATION SERVER, COUPLE A CURRENT SERVICE REQUEST TO BE PROCESSED WITH THE SERVICE DATA, TAKE OUT THE SERVICE DATA IN FIRST-IN FIRST-OUT ORDER AND COUPLE SAME WITH THE SERVICE REQUEST OF A CLIENT, AND RETURN, THROUGH A SERVICE INTERFACE, THE SERVICE DATA TO THE CLIENT SENDING THE SERVICE REQUEST TO BE PROCESSED; AND IF THERE IS NO SERVICE DATA IN THE APPLICATION SERVER AT PRESENT, RETURN INFORMATION INDICATING THAT THERE IS NO SERVICE DATA TO THE CLIENT SENDING THE CURRENT SERVICE REQUEST TO BE PROCESSED

(57) Abstract: The present application relates to a highly concurrent service request processing method and apparatus, and a computer device and a storage medium. The method comprises: receiving service requests sent by multiple clients; allocating the service requests by means of a load-balancing server according to state information; if there is already service data in an application server, coupling a current service request to be processed with the service data, taking out the service data in first-in first-out order and coupling same with the service request of a client, and returning, through a service interface, the service data to the client sending the service request to be processed; and if there is no service data in the application server at present, returning, to the client sending the current service request to be processed, information indicating that there is no service data. The method improves the thread access efficiency of highly concurrent requests, reduces the amount of request data lost and ensures the integrity of a service data request process.

(57) 摘要: 本申请涉及一种高并发业务请求处理方法、装置、计算机设备和存储介质, 方法包括: 接收多个客户端发送的业务请求; 将业务请求由负载均衡服务器按照状态信息对业务请求进行分配; 如应用服务器中已存在业务数据, 则将当前待处理的业务请求与业务数据进行耦合, 按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合, 并通过业务接口向发送待处理的业务请求的客户端返回业务数据; 如当前应用服务器中没有业务数据, 则向发送当前待处理的业务请求的客户端返回无业务数据信息。上述方法提高了高并发请求线程访问效率, 减少请求数据的损失量, 保证了业务数据请求过程的完整性。

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, ST, 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))。

高并发业务请求处理方法、装置、计算机设备和存储介质

本申请要求于 2018 年 05 月 22 日提交中国专利局、申请号为 201810363678.4、发明名称为“高并发业务请求处理方法、装置、计算机设备
5 和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及互联网技术领域，尤其涉及高并发业务请求处理方法、装置、计算机设备和存储介质。

10

背景技术

高并发请求线程包括客户向商业机构发出购买商品请求、联系客服等请求和客服联系客户进行售后处理或者回访等请求，当出现多个请求同时发送时，往往导致连接线路拥挤，访问速度大大降低。比如在电商平台进行网络购物，
15 特别在在“双十一”、“双十二”等大型促销活动或是畅销商品的抢购业务中，出现了许多秒杀商品、秒杀红包、秒杀抽奖等商业活动。这些活动通常是在短时间内进行，并产生大量的访问量，因此具有高并发业务的特点，会对服务商的网络服务器，诸如应用服务器、数据库等产生极大的负载压力。

由于客户请求数量的暴增，请求线路高并发，系统反应缓慢，成为商业领域
20 内的一大难题，众多公司逐渐对自己领域内的客服联系客户或者客户发出的请求，采取各种提高交互效率的手段，一般系统采取高并发异步线程同时请求，通过大量的校验判断和缓存技术提高数据高并发请求效率。

现有技术在处理客户信息高并发请求时的缺陷主要有，校验和缓存程序繁琐，减慢高并发请求处理效率；请求访问线路较差，容易导致丢包。

25

发明内容

有鉴于此，有必要针对现有技术的不足，提供一种高并发业务请求处理方法、装置、计算机设备和存储介质。

一种高并发业务请求处理方法，包括：

接收多个客户端发送的业务请求；

将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

一种高并发业务请求处理装置，包括：

接收单元，设置为接收多个客户端发送的业务请求；

分配单元，设置为将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

返回单元，设置为如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

接收多个客户端发送的业务请求；

将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述

业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

- 5 一种存储有计算机可读指令的存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

接收多个客户端发送的业务请求；

- 10 将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

- 15 如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

- 20 上述高并发业务请求处理方法、装置、计算机设备和存储介质，通过接收多个客户端发送的业务请求，将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器，当所述应用服务器为多台时，所述业务接口根据负载均衡，通过负载均衡服务器将接收到的所述业务请求分配给各台应用服务器，每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载均衡服务器发送自己的状态信息，每台应用服务器定时统计自己的状态信息，并定时向所有负载均衡服务器发送自己的状态信息，每台应用服务器定时统计自己的状态信息，并定时向负载均衡服务器组的至少一台负载均衡服务器发送自己的状态信息；所述负载均衡服务器再定时向其它负载均衡服务器发送所有应用服务器的状态信息。如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客

户端返回所述业务数据，如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息，提高了高并发请求线程访问效率，减少请求数据的损失量，保证了业务数据请求过程的完整性。

5 附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本申请的限制。

图 1 为本申请一个实施例中提供的高并发业务请求处理方法的流程图；

10 图 2 为本申请一个实施例中负载均衡器处理方法示意图；

图 3 为本申请一个实施例中负载均衡服务器处理方法流程图；

图 4 为本申请一个实施例中高并发业务请求处理装置的结构框图。

具体实施方式

15 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本技术领域技术人员可以理解，除非特意声明，这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是，本申请的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。

作为一个较好的实施例，如图 1 所示，一种高并发业务请求处理方法，该处理方法包括以下步骤：

25 步骤 S101，接收多个客户端发送的业务请求；

这里不同客户端发送的业务请求，可以是用于请求处理同一业务的，也可以是用于请求不同业务的。网关可以并发地接收多个客户端同时发来的业务请求，可以理解的是，这里所谓的并发地接收多个客户端的业务请求，可以是在

同一时间区间内接收到多个客户端的业务请求，比如在 1 秒内接收到 100 万业务请求，可以视为 1 秒内并发地接收 100 万业务请求。

步骤 S102，将业务请求分配给负责业务处理的其中一台应用服务器；

当仅包括一台应用服务器时，直接把接收到的高并发业务请求分配到该应用服务器中。当所述应用服务器为多台时，所述业务接口根据负载均衡，通过负载均衡服务器将接收到的所述业务请求分配给各台应用服务器，当所述负载均衡服务器与所述地址解析协议请求相匹配时，负载均衡服务器将物理地址反馈至所述客户端，客户端根据所述负载均衡服务器的物理地址，将业务请求发送至所述负载均衡服务器，负载均衡服务器将所述业务请求调度给所述负载均衡服务器所连接的负载值最小的节点服务器进行处理。

步骤 S103，如应用服务器中已存在业务数据，则将当前待处理的业务请求与业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送待处理的业务请求的客户端返回业务数据；如当前应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

接收业务请求后，选择应用服务器对业务请求进行处理，待处理的业务请求可以是兑奖、红包、秒杀抽奖等。如果应用服务器存在业务数据，则可以将该业务数据与该客户端发送的该业务请求进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，如客户端发送的是秒杀抽奖业务请求，如果应用服务器中当前有兑奖数据，则将兑奖数据与该客户端的该秒杀抽奖请求耦合，并向该客户端返回中奖信息，如果应用服务器中当前无兑奖数据，则向该客户端返回未中奖信息。同样地，应用服务器需要将客户端的中奖信息存储到数据库或缓存中，以便后续操作。如果应用服务器中有多个兑奖数据时，例如可以采用队列缓存多个兑奖数据，在与客户端的业务请求耦合时，例如可以按照先进先出的顺序取出兑奖数据与客户端的业务请求进行耦合。并通过业务接口向该客户端返回该业务数据，而如果该应用服务器中当前无业务数据，则通过业务接口向该客户端返回无业务数据信息，即业务请求未通过。此外，为了方便后续作业，应用服务器还需要将得到业务数据的客户端及业务数据信

息存储到数据库或缓存中。业务数据例如可以为兑奖数据、红包数据、秒杀抽奖数据等。如客户端发送的是秒杀抽奖业务请求，如果应用服务器当前有兑奖数据，则将兑奖数据与该客户端的该秒杀抽奖请求耦合，并向该客户端返回中奖信息；如果应用服务器中当前无兑奖数据，则向该客户端返回未中奖信息。

- 5 同样地，应用服务器需要将客户端的中奖信息存储到数据库或缓存中，以便后续操作。另外，如果应用服务器中有多个兑奖数据时，例如可以采用队列缓存多个兑奖数据，在与客户端的业务请求耦合时，例如可以按照先进先出的顺序取出兑奖数据与客户端的业务请求进行耦合。

如果该应用服务器中当前无业务数据，则通过业务接口向该客户端返回无
10 业务数据信息，即业务请求未通过。另外，为了方便后续应用，应用服务器还需要将得到业务数据的客户端信息及业务数据信息存储到数据库或缓存中。在一些实施例中，可以是传统的网络存储系统采用集中的存储服务器存放所有数据。在另一些实施例中，可以采用分布式内存缓存，就是将数据分散存储在多台独立的设备上，分布式内存缓存采用可扩展的系统结构，利用多台存储服务
15 器分担存储负荷，利用位置服务器定位存储信息，常见的包括 REDIS 缓存、MHMCACHED 缓存等。

在一个实施例中，该方法进一步包括：对所述业务请求进行限流，获取预设存储区域中预先存储的全局状态数据，所述全局状态数据用于表征业务的全局状态，全局状态数据是服务器对不被限流的业务请求进行处理得到的，将所
20 述业务请求中未被限流的业务请求发送给应用服务器。

通过限流措施对应用服务器的业务请求进行限流时，进一步分散高并发业务时的负载压力，在另一些实施例，还可以通过业务请求承载能力比应用服务器强的网关来接收客户端发送的请求，比如，应用服务器的业务请求承载能力为每秒承载 100 万次业务请求，那么可以选择每秒承载 1000 万次业务请求的网
25 关。在另一些实施例，还可以设置缓存数据，用于更进一步分散高并发业务时的负载压力，可以是传统的网络存储系统采用集中的存储服务器存放所有数据。也可以采用分布式内存缓存，就是将数据分散存储在多台独立的设备上，分布式内存缓存采用可扩展的系统结构，利用多台存储服务器分担存储负荷，利用

位置服务器定位存储信息，常见的包括 REDIS 缓存、MHMCACHED 缓存等。

在一个实施例中，当应用服务器为多台时，业务接口根据负载均衡，通过负载均衡服务器将接收到的业务请求分配给各台应用服务器。

如图 2 所示，负载均衡服务器(Load-Balancing Server, LBS)采用负载均衡技术，将高并发的业务请求分配到各应用服务器中，例如负载均衡服务器将用户 PC 发来的业务请求分配到应用服务器 1~3 中。业务接口根据各应用服务器的业务处理状态，确定将高并发业务请求具体分配到哪台应用服务器中。需要说明的是，具体使用的负载均衡技术可以根据实际需求而进行选择，本申请不以此为限。通过将高并发的业务请求均衡分配到各应用服务器中，可以保证整个系统的高并发响应速度。

在一个实施例中，每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载均衡服务器发送自己的状态信息。

预先为解析服务器有一个设置好的负载情况的阈值：若其负载情况超过该阈值则当接收到客户端的请求后，选择同组其他应用服务器来处理；若其负载情况小于该阈值则自行直接处理客户端的请求，开始解析客户端发送来的请求，并重新统计自身的状态信息更新列表，发送给其他解析服务器，其他解析服务器收到状态信息后更新列表。如果当前负载情况大于阈值，则该应用服务器查看其维护的解析服务器负载情况列表，先选出满足条件的应用服务器，如负载情况小于设定阈值的解析服务器，再在所有满足条件的解析服务器中通过赌轮盘的方式选择一个解析服务器。所谓赌轮盘的方式就是指负载情况越小越容易被选择到，但负载情况最低的应用服务器并非绝对会被选择到。赌轮盘的决策算法为现有技术，在此不再赘述。当前的解析服务器选择得到解析服务器 S_t 后，将其路径返回给客户端，客户端重新向解析服务器 S_t 发送请求消息，并附带标识表示其为重定向消息，如此解析服务器 S_t 将直接执行该请求而不再进行负载重定向。

在一个实施例中，每台应用服务器定时统计自己的状态信息，并定时向所有负载均衡服务器发送自己的状态信息。

分配给客户端的解析服务器解析客户端的请求，并为客户端选择一台负载

情况较低的应用服务器。

如图 3 所示，在一个实施例中，步骤 S201，每台应用服务器定时统计自己的状态信息，并定时向负载均衡服务器组的至少一台负载均衡服务器发送自己的状态信息；步骤 S202，负载均衡服务器再定时向其它负载均衡服务器发送所有应用服务器的状态信息。

解析服务器查看其维护的应用服务器状态信息列表，选出满足条件的应用服务器，如负载情况小于设定阈值的应用服务器，在所有满足条件的应用服务器中使用一定的方法进行决策，选出最终的应用服务器。如赌轮盘决策法，应用服务器负载情况越小者越容易被选择到。

如图 4 所示，在一个实施例中，提出了一种高并发业务请求处理装置，该处理装置包括：

接收单元，设置为接收多个客户端发送的业务请求；

分配单元，设置为将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

返回单元，设置为如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

在一个实施例中，还包括：限流单元，设置为对所述业务请求进行限流，获取预设存储区域中预先存储的全局状态数据，所述全局状态数据用于表征业务的全局状态，全局状态数据是服务器对不被限流的业务请求进行处理得到的，将所述业务请求中未被限流的业务请求发送给应用服务器。

在一个实施例中，所述分配单元还设置为当所述应用服务器为多台时，所述业务接口根据负载均衡，当所述负载均衡服务器与所述地址解析协议请求相匹配时，负载均衡服务器将物理地址反馈至所述客户端，客户端根据所述负载均衡服务器的物理地址，将业务请求发送至所述负载均衡服务器，负载均衡服

务器将所述业务请求调度给所述负载均衡服务器所连接的负载值最小的节点服务器进行处理。

在一个实施例中，所述分配单元还设置为每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载均衡服务器发送自己的状态信息。

5 在一个实施例中，所述分配单元还设置为每台应用服务器定时统计自己的状态信息，并定时向所有负载均衡服务器发送自己的状态信息。

10 在一个实施例中，所述分配单元还设置为每台应用服务器定时统计自己的状态信息，并定时向负载均衡服务器组的至少一台负载均衡服务器发送自己的状态信息；所述负载均衡服务器再定时向其它负载均衡服务器发送所有应用服务器的状态信息。

在一个实施例中，提出了一种计算机设备，所述计算机设备包括存储器和处理器，存储器中存储有计算机可读指令，计算机可读指令被处理器执行时，使得处理器执行所述计算机程序时实现上述各实施例里高并发业务请求处理方法中的步骤。

15 在一个实施例中，提出了一种存储有计算机可读指令的存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行上述各实施例里高并发业务请求处理方法中的步骤。其中，存储介质可以为非易失性存储介质。

20 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：只读存储器（ROM，Read Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁盘或光盘等。

25 以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请一些示例性实施例，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变

形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

1、一种高并发业务请求处理方法，包括：

接收多个客户端发送的业务请求；

将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

2、根据权利要求 1 所述的高并发业务请求处理方法，其中，还包括：对所述业务请求进行限流，获取预设存储区域中预先存储的全局状态数据，所述全局状态数据用于表征业务的全局状态，全局状态数据是服务器对不被限流的业务请求进行处理得到的，将所述业务请求中未被限流的业务请求发送给应用服务器。

3、根据权利要求 1 所述的高并发业务请求处理方法，其中，由负载均衡服务器按照状态信息对业务请求进行分配包括：

当所述应用服务器为多台时，所述业务接口根据负载均衡，通过负载均衡服务器将接收到的所述业务请求分配给各台应用服务器，当所述负载均衡服务器与地址解析协议请求相匹配时，负载均衡服务器将物理地址反馈至所述客户端，客户端根据所述负载均衡服务器的物理地址，将业务请求发送至所述负载均衡服务器，负载均衡服务器将所述业务请求调度给所述负载均衡服务器所连接的负载值最小的节点服务器进行处理。

4、根据权利要求 3 所述的高并发业务请求处理方法，其中，每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载均衡服务器发送自己的状态信息。

5、根据权利要求 4 所述的高并发业务请求处理方法，其中，每台应用服务器定时统计自己的状态信息，并定时向所有负载均衡服务器发送自己的状态信

息。

6、根据权利要求 4 所述的高并发业务请求处理方法，其中，每台应用服务器定时统计自己的状态信息，并定时向负载均衡服务器组的至少一台负载均衡服务器发送自己的状态信息；所述负载均衡服务器再定时向其它负载均衡服务器发送所有应用服务器的状态信息。

7、一种高并发业务请求处理装置，包括：

接收单元，设置为接收多个客户端发送的业务请求；

分配单元，设置为将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

返回单元，设置为如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

8、根据权利要求 7 所述的高并发业务请求处理装置，其中，还包括：

限流单元，设置为对所述业务请求进行限流，获取预设存储区域中预先存储的全局状态数据，所述全局状态数据用于表征业务的全局状态，全局状态数据是服务器对不被限流的业务请求进行处理得到的，将所述业务请求中未被限流的业务请求发送给应用服务器。

9、根据权利要求 7 所述的高并发业务请求处理装置，其中，所述分配单元还设置为当所述应用服务器为多台时，所述业务接口根据负载均衡，当所述负载均衡服务器与所述地址解析协议请求相匹配时，负载均衡服务器将物理地址反馈至所述客户端，客户端根据所述负载均衡服务器的物理地址，将业务请求发送至所述负载均衡服务器，负载均衡服务器将所述业务请求调度给所述负载均衡服务器所连接的负载值最小的节点服务器进行处理。

10、根据权利要求 9 所述的高并发业务请求处理装置，其中，所述分配单元还设置为每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载

均衡服务器发送自己的状态信息。

11、根据权利要求 10 所述的高并发业务请求处理装置，其中，所述分配单元还设置为每台应用服务器定时统计自己的状态信息，并定时向所有负载均衡服务器发送自己的状态信息。

5 12、根据权利要求 10 所述的高并发业务请求处理装置，其中，所述分配单元还设置为每台应用服务器定时统计自己的状态信息，并定时向负载均衡服务器组的至少一台负载均衡服务器发送自己的状态信息；所述负载均衡服务器再定时向其它负载均衡服务器发送所有应用服务器的状态信息。

10 13、一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

接收多个客户端发送的业务请求；

15 将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

14、根据权利要求 13 所述的计算机设备，其中，所述计算机可读指令被所述处理器执行时，还使得所述处理器执行以下步骤：

25 对所述业务请求进行限流，获取预设存储区域中预先存储的全局状态数据，所述全局状态数据用于表征业务的全局状态，全局状态数据是服务器对不被限流的业务请求进行处理得到的，将所述业务请求中未被限流的业务请求发送给应用服务器。

15、根据权利要求 13 所述的计算机设备，其中，由负载均衡服务器按照状态信息对业务请求进行分配时，使得所述处理器执行以下步骤：

当所述应用服务器为多台时，所述业务接口根据负载均衡，通过负载均衡服务器将接收到的所述业务请求分配给各台应用服务器，当所述负载均衡服务器与地址解析协议请求相匹配时，负载均衡服务器将物理地址反馈至所述客户端，客户端根据所述负载均衡服务器的物理地址，将业务请求发送至所述负载均衡服务器，负载均衡服务器将所述业务请求调度给所述负载均衡服务器所连接的负载值最小的节点服务器进行处理。

16、根据权利要求 15 所述的计算机设备，其中，所述计算机可读指令被所述处理器执行时，还使得所述处理器执行以下步骤：

每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载均衡服务器发送自己的状态信息。

17、一种存储有计算机可读指令的存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

接收多个客户端发送的业务请求；

将所述业务请求由负载均衡服务器按照状态信息对业务请求进行分配，所述负载均衡服务器选择合适的应用服务器并分配给负责业务处理的其中一台应用服务器；

如所述应用服务器中已存在业务数据，则将当前待处理的业务请求与所述业务数据进行耦合，按照先进先出的顺序取出业务数据与客户端的业务请求进行耦合，并通过业务接口向发送所述待处理的业务请求的客户端返回所述业务数据；如当前所述应用服务器中没有业务数据，则向发送当前待处理的业务请求的客户端返回无业务数据信息。

18、根据权利要求 17 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时，还使得一个或多个处理器执行以下步骤：

对所述业务请求进行限流，获取预设存储区域中预先存储的全局状态数据，所述全局状态数据用于表征业务的全局状态，全局状态数据是服务器对不被限流的业务请求进行处理得到的，将所述业务请求中未被限流的业务请求发送给应用服务器。

19、根据权利要求 17 所述的存储介质，其中，由负载均衡服务器按照状态

信息对业务请求进行分配时，使得一个或多个处理器执行以下步骤：

当所述应用服务器为多台时，所述业务接口根据负载均衡，通过负载均衡服务器将接收到的所述业务请求分配给各台应用服务器，当所述负载均衡服务器与地址解析协议请求相匹配时，负载均衡服务器将物理地址反馈至所述客户端，客户端根据所述负载均衡服务器的物理地址，将业务请求发送至所述负载均衡服务器，负载均衡服务器将所述业务请求调度给所述负载均衡服务器所连接的负载值最小的节点服务器进行处理。

20、根据权利要求 19 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时，还使得一个或多个处理器执行以下步骤：

10 每台负载均衡服务器定时统计自己的状态信息，并定时向其他负载均衡服务器发送自己的状态信息。

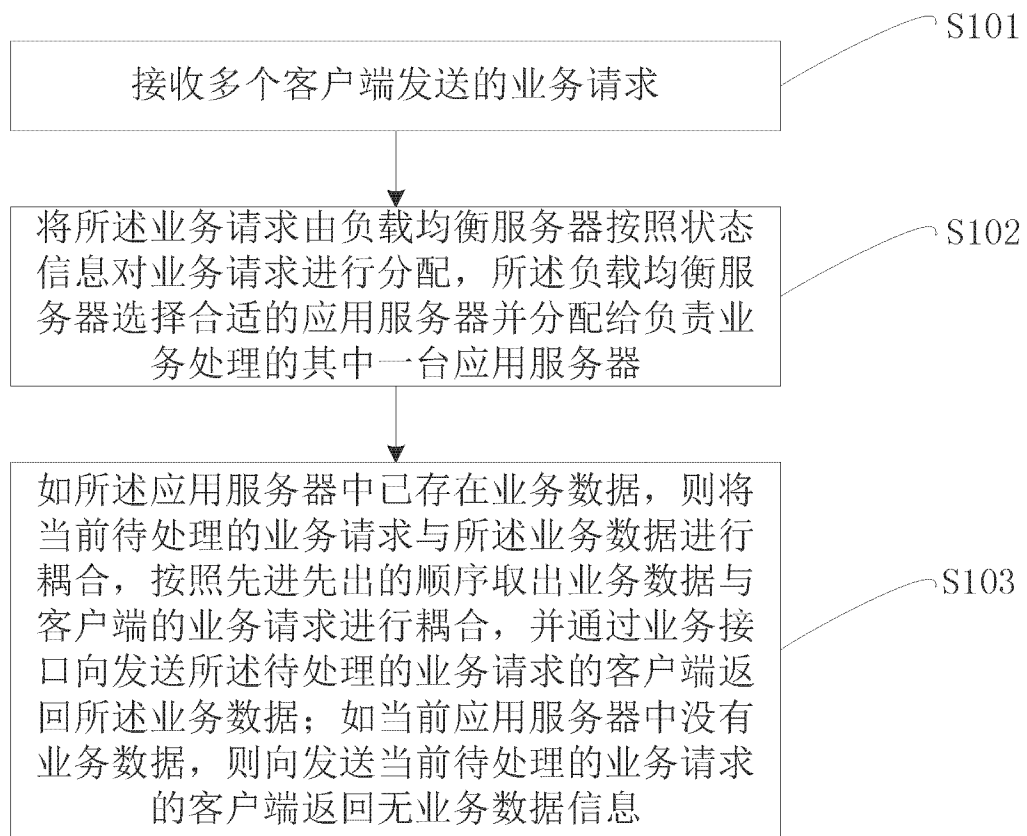


图 1

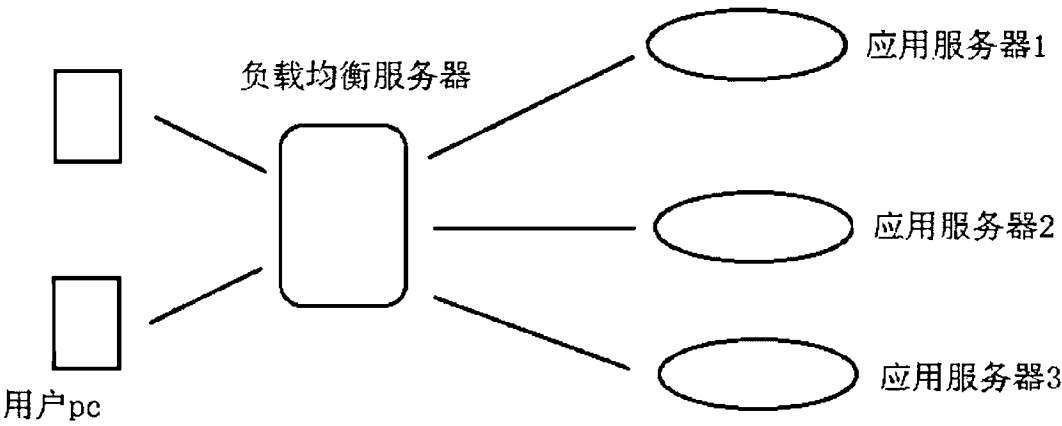


图 2

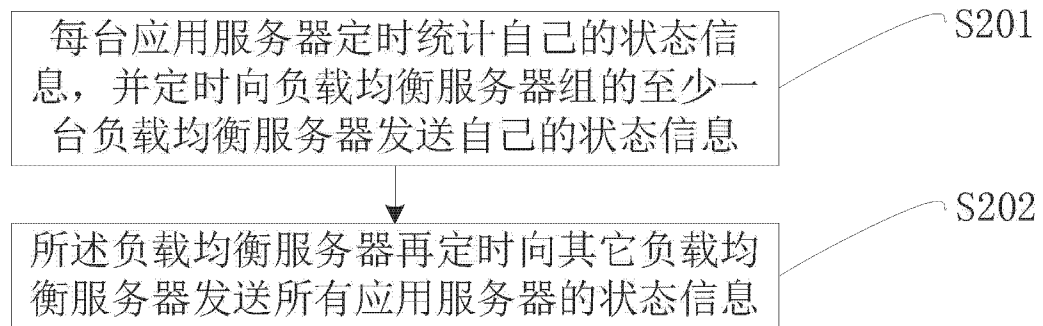


图 3

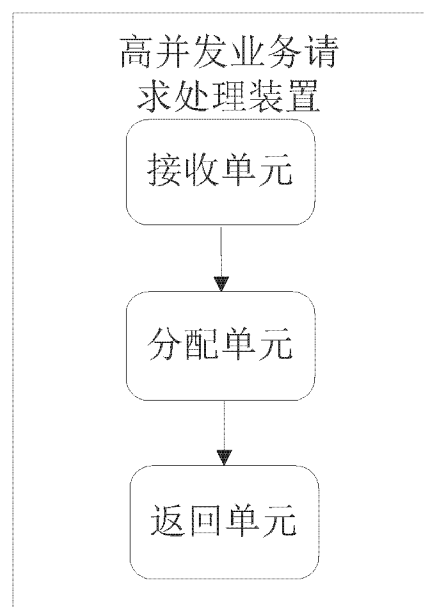


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/104151

A. CLASSIFICATION OF SUBJECT MATTER

H04W 28/02(2009.01)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)

H04W,H04Q,H04L

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)

CNPAT, WPI, EPODOC, GOOGLE: 并发, 并行, 请求, 服务器, 分配, 选择, 负载均衡, intercurrent, parallel, request, server, allocat+, select+, load balance+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107277088 A (TAIKANG INSURANCE GROUP CO., LTD. ET AL.) 20 October 2017 (2017-10-20) claims 1-12, and description, paragraphs 25-28	1-20
A	CN 103220354 A (GUANGDONG EASTONE CENTURY TECHNOLOGY CO., LTD.) 24 July 2013 (2013-07-24) entire document	1-20
A	CN 105072182 A (BEIJING JIAXUN FEIHONG ELECTRICAL CO., LTD.) 18 November 2015 (2015-11-18) entire document	1-20
A	CN 105791370 A (HUAWEI TECHNOLOGIES CO., LTD.) 20 July 2016 (2016-07-20) entire document	1-20
A	WO 2008009185 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 24 January 2008 (2008-01-24) entire document	1-20

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

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“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)

“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

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search

29 November 2018

Date of mailing of the international search report

03 January 2019

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)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/104151

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107277088	A	20 October 2017	None			
CN	103220354	A	24 July 2013	None			
CN	105072182	A	18 November 2015	None			
CN	105791370	A	20 July 2016	WO	2016101662	A1	30 June 2016
WO	2008009185	A1	24 January 2008	CN	101106483	A	16 January 2008

国际检索报告

国际申请号

PCT/CN2018/104151

A. 主题的分类

H04W 28/02 (2009.01) 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)

H04W, H04Q, H04L

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))

CNPAT, WPI, EPDOC, GOOGLE: 并发, 并行, 请求, 服务器, 分配, 选择, 负载均衡, intercurrent, parallel, request, server, allocat+, select+, load balance+

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	CN 107277088 A (泰康保险集团股份有限公司等) 2017年 10月 20日 (2017 - 10 - 20) 权利要求1-12, 说明书第25-58段	1-20
A	CN 103220354 A (广东宜通世纪科技股份有限公司) 2013年 7月 24日 (2013 - 07 - 24) 全文	1-20
A	CN 105072182 A (北京佳讯飞鸿电气股份有限公司) 2015年 11月 18日 (2015 - 11 - 18) 全文	1-20
A	CN 105791370 A (华为技术有限公司) 2016年 7月 20日 (2016 - 07 - 20) 全文	1-20
A	WO 2008009185 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 2008年 1月 24日 (2008 - 01 - 24) 全文	1-20

☐ 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

“&” 同族专利的文件

国际检索实际完成的日期

2018年 11月 29日

国际检索报告邮寄日期

2019年 1月 3日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

赵新蕾

传真号 (86-10) 62019451

电话号码 86-(10)-53961623

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/104151

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	107277088	A	2017年 10月 20日	无			
CN	103220354	A	2013年 7月 24日	无			
CN	105072182	A	2015年 11月 18日	无			
CN	105791370	A	2016年 7月 20日	WO	2016101662	A1	2016年 6月 30日
WO	2008009185	A1	2008年 1月 24日	CN	101106483	A	2008年 1月 16日

表 PCT/ISA/210 (同族专利附件) (2015年1月)