

(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
2017 年 9 月 28 日 (28.09.2017)



(10) 国际公布号
WO 2017/162034 A1

- (51) 国际专利分类号:
H04L 29/08 (2006.01)
- (21) 国际申请号: PCT/CN2017/076039
- (22) 国际申请日: 2017 年 3 月 9 日 (09.03.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610165378.6 2016 年 3 月 22 日 (22.03.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Cayman (KY)。
- (72) 发明人: 蒋海涛 (JIANG, Haitao); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 林洁 (LIN, Jie); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, 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, 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,

[见续页]

(54) Title: LOADING METHOD AND SYSTEM

(54) 发明名称: 一种负载方法和系统

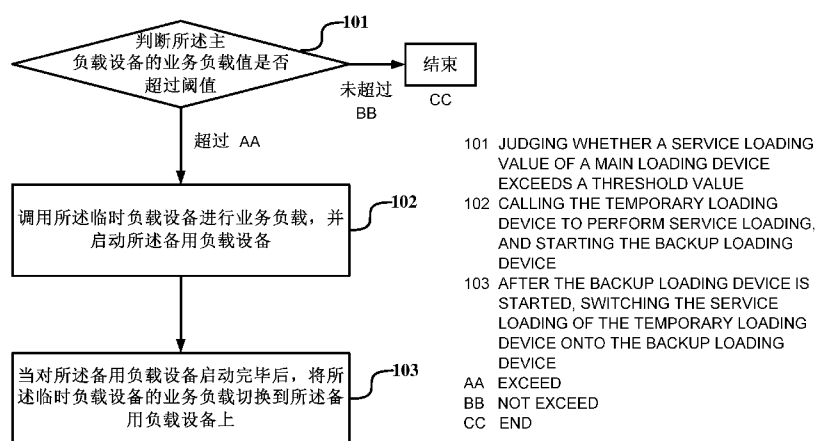


图 1

(57) Abstract: Provided are a loading method and system. The method comprises: judging whether a service loading value of a main loading device exceeds a threshold value; if so, calling a temporary loading device to perform service loading, and starting a backup loading device; and after the backup loading device is started, switching the service loading of the temporary loading device onto the backup loading device, so that the backup loading device performs service loading. The present invention avoids the abnormality and unavailability of a service, and even a suspending phenomenon caused by the fact that during a backup loading starting process, service loads continue to increase and exceed a service loading capability of a main loading device.

(57) 摘要: 本发明提供了一种负载方法和系统, 其中, 所述方法包括: 判断所述主负载设备的业务负载值是否超过阈值; 如果超过, 调用所述临时负载设备进行业务负载, 并启动所述备用负载设备; 当对所述备用负载设备启动完毕后, 将所述临时负载设备的业务负载切换到所述备用负载设备上, 以使所述备用负载设备进行业务负载。本发明避免了在备用负载启动的过程中由于业务负载持续上升, 超过主负载设备的业务负载能力时而引起的服务异常、不可用, 甚至挂起现象。



WO 2017/162034 A1



RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, **本国际公布:**

CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, — 包括国际检索报告(条约第 21 条(3))。
TG)。

一种负载方法和系统

本申请要求 2016 年 03 月 22 日递交的申请号为 201610165378.6、发明名称为“一种负载方法和系统”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。

5 技术领域

本发明涉及网络技术领域，尤其涉及一种负载方法和系统。

背景技术

云服务中线上服务器业务负载上升达到阈值的时候，就需要扩容出新的服务器来应
10 对新增负载。

现有技术中的扩容方案是横向扩容，也就是启动更多的扩容服务器来支撑当前业务负载。

在实现本发明的过程中，发明人发现现有技术至少存在如下问题：

扩容新服务器的等待时间通常是几分钟或者更长，如果在此期间业务负载持续上升，
15 一旦超出线上服务器的处理能力，就会造成服务异常、不可用，甚至是服务挂起的现象。

发明内容

本发明提供一种负载方法，当主负载设备的业务负载值超过阈值时，调用临时负载设备进行业务负载，并启动备用负载设备，当所述备用负载设备启动完毕后将所述临时
20 负载设备的业务负载切换到所述备用负载设备上。本发明避免了在备用负载启动的过程中由于业务负载持续上升，超过主负载设备的业务负载能力时而引起的服务异常、不可用，甚至挂起现象。

本发明提供一种负载方法，所述方法应用于包括主负载设备、临时负载设备和备用负载设备的系统中，其中，所述主负载设备为当前正在进行业务负载的设备，所述临时
25 负载设备为随时可以进行临时业务负载的设备，所述备用设备是进行启动后才能进行业务负载的设备，所述方法包括：

判断所述主负载设备的业务负载值是否超过阈值；

如果超过，调用所述临时负载设备进行业务负载，并启动所述备用负载设备；

当对所述备用负载设备启动完毕后，将所述临时负载设备的业务负载切换到所述备
30 用负载设备上，以使所述备用负载设备进行业务负载。

所述备用负载设备为多个，且不同的所述备用负载设备的负载能力不同，其中，业务负载能力越高的所述备用负载设备的启动时间越长；

启动所述备用负载设备，具体包括：

确定当前业务负载的增长速率；

- 5 根据所述增长速率和各个所述备用负载设备的启动时间启动对应负载能力的所述备用负载设备。

所述备用负载设备为多个，且各个所述备用负载设备的负载能力相同；

启动所述备用负载设备，具体包括：

确定当前业务负载的增长速率；

- 10 根据所述增长速率和所述备用负载设备的启动时间启动对应数量的所述备用负载设备。

将所述临时负载设备的业务负载切换到所述备用负载设备上，具体包括：

关闭所述临时负载设备进行业务负载的应用；

将所述临时负载设备的业务负载切换到所述备用负载设备上。

- 15 在将所述临时负载设备的业务负载切换到所述备用负载设备上后，所述方法还包括：
当所述备用负载设备的业务负载为零时，关闭所述备用负载设备。

一种系统，所述系统中包括主负载设备、临时负载设备和备用负载设备，其中，所述主负载设备为当前正在进行业务负载的设备，所述临时负载设备为随时可以进行临时业务负载的设备，所述备用设备是进行启动后才能进行业务负载的设备，所述系统包括：

- 20 判断模块，用于判断所述主负载设备的业务负载值是否超过阈值；

调用模块，如果超过阈值，用于调用所述临时负载设备进行业务负载；启动模块，用于启动所述备用负载设备；

切换模块，当对所述备用负载设备启动完毕后，用于将所述临时负载设备的业务负载切换到所述备用负载设备上，以使所述备用负载设备进行业务负载。

- 25 所述备用负载设备为多个，且不同的所述备用负载设备的负载能力不同，其中，业务负载能力越高的所述备用负载设备的启动时间越长；

所述启动模块，具体用于：

确定当前业务负载的增长速率；

根据所述增长速率和各个所述备用负载设备的启动时间启动对应负载能力的所述备

- 30 用负载设备。

所述备用负载设备为多个，且各个所述备用负载设备的负载能力相同；

所述启动模块，具体用于：

确定当前业务负载的增长速率；

根据所述增长速率和所述备用负载设备的启动时间启动对应数量的所述备用负载设

5 备。

所述切换模块，具体用于：

关闭所述临时负载设备进行业务负载的应用；

将所述临时负载设备的业务负载切换到所述备用负载设备上。

所述系统，还包括：

10 关闭模块，在将所述临时负载设备的业务负载切换到所述备用负载设备上后，当所述备用负载设备的业务负载为零时，用于关闭所述备用负载设备。

本发明中当主负载设备的业务负载值超过阈值时，调用临时负载设备进行业务负载，并启动备用负载设备，当所述备用负载设备启动完毕后将所述临时负载设备的业务负载切换到所述备用负载设备上。本发明避免了在备用负载启动的过程中由于业务负载持续

15 上升，超过主负载设备的业务负载能力时而引起的服务异常、不可用，甚至挂起现象。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本发明

20 明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图1为本发明实施例中的一种负载方法；

图2为本发明实施例中在线服务器、扩容服务器和缓冲服务器位置结构示意图；

图3为本发明实施例中的一种系统的结构示意图。

25

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技

30 术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范

围。

为了解决现有技术中提出的技术问题，本发明提出了一种负载方法，具体方法如图 1 所示，所述方法应用于包括主负载设备、临时负载设备和备用负载设备的系统中，其中，所述主负载设备为当前正在进行业务负载的设备，所述临时负载设备为随时可以进行临时业务负载的设备，所述备用设备是进行启动后才能进行业务负载的设备，所述方法包括以下步骤：

步骤 101，判断所述主负载设备的业务负载值是否超过阈值，如果超过，则执行步骤 102，如果没有超过，则结束流程。

具体的，所述主负载设备和所述备用负载设备能够提供稳定的负载业务，并为用户提供稳定的、高品质的服务，所述主负载设备和所述备用负载设备的稳定性、业务负载能力、提供业务的能力，以及硬件配置和硬件性能要高于所述临时负载设备，所述主负载设备和所述备用负载设备是横向启动，也就是说，在主负载设备的负载能力超过阈值时，才会启动所述备用负载设备，进一步的，所述方法具体可用于横向启动的场景里。

所述临时负载设备为时刻开启状态，即处于随时可以进行业务负载的状态，在所述主负载设备超过阈值时立即使用所述临时负载设备进行业务负载，以使所述临时负载设备为所述主负载设备分担负载，并且所述临时负载设备处于开启状态时其所消耗的成本相对于所述主负载设备和所述备用负载设备处于开启状态时所消耗的成本要低很多，所述成本包括如：维护成本、消耗成本，以及资源占用成本等。

当所述主负载设备的业务负载值接近阈值时，表示所述主负载设备能够提供的业务负载能力已经很少，如果所述业务负载在持续增长，那么会超过所述主负载设备的业务负载能力，需要横向启动所述备用负载设备。

步骤 102，调用所述临时负载设备进行业务负载，并启动所述备用负载设备。

具体的，由于启动所述备用负载设备需要一定时间，为了避免在所述备用负载设备的过程中，当前的业务负载超过所述主负载设备的业务负载能力，因此调用所述临时负载设备，由于所述临时负载设备时刻处于开启状态，因此所述临时负载设备可以随时在所述主负载设备超过其负载能力时为所述主负载设备分担业务负载，在所述备用负载设备启动的过程中起到一个过渡作用，在此期间，所述临时负载设备能够为稳定的负载业务，并为用户提供稳定的、高品质的服务。

其中，所述备用负载设备为多个，并且多个所述备用负载设备的负载能力可以相同，也可以不同，并且负载能力越高的所述备用负载设备启动时间也越长。

当所述备用负载设备为多个，且不同的所述备用负载设备的负载能力不同时，启动所述备用负载设备，具体包括：

确定当前业务负载的增长速率；

根据所述增长速率和各个所述备用负载设备的启动时间启动对应负载能力的所述备用负载设备。

具体的，为了保证启动的所述备用负载设备能够承担超出所述主负载设备的业务负载，并且保证合理的启动所述备用负载设备，需要根据所述增长速率和各个所述备用负载设备的启动时间来启动所述备用负载设备，这样可以节省系统的资源和成本，例如：所述备用负载设备有 3 个，业务负载能力分别为：10、15、20，启动时间分别为 3 分钟、4 分钟、5 分钟，如果当前业务负载的增长速率为 1，那么需要启动业务负载能力为 10 的所述备用负载设备；如果当业务负载的增长速率为 3，并且如果超过业务负载能力为 10 的所述备用负载设备的阈值，但是没有超过业务负载能力为 15 的所述备用负载设备的阈值，那么需要启动业务负载能力为 15 的所述备用负载设备，如果未超过业务负载能力为 10 的所述备用负载设备的阈值，那么需要启动业务负载能力为 10 的所述备用负载设备。

其中，在确定系统中使用何种级别的业务负载量的所述备用负载设备可以根据实际情况确定。

当所述备用负载设备为多个，且各个所述备用负载设备的负载能力相同时，启动所述备用负载设备，具体包括：

确定当前业务负载的增长速率；

根据所述增长速率和所述备用负载设备的启动时间启动对应数量的所述备用负载设备。

具体的，由于多个所述备用负载设备的启动时间相同，在启动时启动任何一个都可以，但是为了保证启动的所述备用负载设备能够负载超过所述主负载设备的业务负载，并且保证启动合理个数的所述备用负载设备，需要根据所述增长速率和各个所述备用负载设备的启动时间来启动所述备用负载设备的个数，例如：所述备用负载设备有 3 个，业务负载能力都为 10，启动时间都为 3 分钟，如果当前业务负载的增长速率为 1，那么需要启动的个数为 1 个；如果当前业务负载的增长速率为 3，并且如果超过了一个所述备用负载设备的阈值，那么需要启动的个数为 2 个，如果未超过一个所述备用负载设备的阈值，那么需要启动的个数为 1 个；如果当前业务负载的增长速率为 4，那么需要启

动的个数为 2 个。

其中，在确定系统中使用何种级别的业务负载量的所述备用负载设备可以根据实际情况确定。

在所述备用负载设备为多个的情况下，当所述主负载设备的业务负载值超过阈值后，
5 启动所述备用负载设备时，也可以任意启动多个所述备用负载设备中的一个，如果启动的所述备用负载设备的负载能力不能满足超出所述主负载设备负载能力的业务负载，那么任意启动另一个所述备用负载设备，直到启动的所述备用负载设备的负载能力能够满足超出所述主负载设备负载能力的业务负载。当然所述备用负载设备也可以为一个，所述备用负载设备的负载能力是根据日常的业务负载量确定的，即所述备用负载设备的业务负载能力足够强大，在所述主负载设备的业务负载值超过阈值后，由所述临时负载设备暂时进行业务负载，当所述备用负载设备启动完成后，将所述临时负载设备上负载的业务切换到所述备用负载设备上业务负载。

步骤 103，当对所述备用负载设备启动完毕后，将所述临时负载设备的业务负载切换到所述备用负载设备上，以使所述备用负载设备进行业务负载。

15 具体的，为了保证为能够保证更好的进行业务负载，以及为用户提供更好的服务，当对所述备用负载设备启动完毕后，需要将所述临时负载设备的业务负载切换到所述备用负载设备上，进而让所述备用负载设备进行业务负载。

将所述临时负载设备的业务负载切换到所述备用负载设备上，具体包括：

关闭所述临时负载设备进行业务负载的应用；

20 将所述临时负载设备的业务负载切换到所述备用负载设备上。

具体的，所述临时负载设备是通过其运行的应用来进行业务负载的，当对所述备用负载设备启动完毕后只需要将所述应用关闭，就会释放业务负载，系统会将释放的业务负载切换到启动了的所述备用负载设备上，这样不用关闭所述临时负载设备就能达到释放业务负载的效果，并且相对于关闭所述临时负载设备的方法更加合理和节省时间，同时
25 在调用所述临时负载设备时，通过启动所述临时负载设备中的所述应用就可以调用所述临时负载设备了，或者在所述临时负载设备关闭所述应用释放业务负载后，所述临时负载设备可以主动加载所述应用，以在调用所述临时负载设备时可以立即使用。当超出所述备用负载设备的负载能力时，可以按照上述方法再调用所述临时负载设备进行业务负载，并启动其他所述备用负载设备。

30 在将所述临时负载设备的业务负载切换到所述备用负载设备上后，所述方法还包括：

当所述备用负载设备的业务负载为零时，关闭所述备用负载设备。

具体的，为了节约成本，以及避免不必要的资源浪费，当所述备用负载设备的业务负载为零时，即所述备用负载设备不在进行业务负载和为用户提供服务时，需要关闭所述备用负载设备，当再次超出所述主负载设备时，可以按照上述方法再调用所述临时负载设备进行业务负载，并启动所述备用负载设备。

所述系统可以为云服务系统，所述云服务系统中包括在线服务器、扩容服务器和缓冲服务器，其中，在线服务器对应本实施例中的主负载设备、扩容服务器对应本实施例中的备用负载设备、缓冲服务器对应本实施例中的临时负载设备，当然也可以是其他系统，凡是在主负载设备超过阈值需要横向启动备用服务器，且启动备用服务器的过程中会造成服务异常、不可用，甚至是服务挂起的现象的系统均属于本发明的保护范围。

本发明中当主负载设备的业务负载值超过阈值时，调用临时负载设备进行业务负载，并启动备用负载设备，当所述备用负载设备启动完毕后将所述临时负载设备的业务负载切换到所述备用负载设备上。本发明避免了在备用负载启动的过程中由于业务负载持续上升，超过主负载设备的业务负载能力时而引起的服务异常、不可用，甚至挂起现象，实现了零等待、零风险的业务负载。

为了进一步阐述本发明的技术思想，现结合具体的应用场景，对本申请的技术方案进行说明，具体的，云服务器包括在线服务器、扩容服务器和缓冲服务器，具体如图 2 所示，其中，所述在线服务器为当前正在进行业务负载的服务器，所述缓冲服务器为随时可以进行临时业务负载的服务器，所述扩容服务器是进行扩容后才能进行业务负载的服务器，其中，所述缓冲服务器放于缓存池中，如果当前业务负载超过云服务中在线服务器的负载阈值，需要横向对扩容服务器进行扩容时，所述方法包括：

1、从缓存池中调用所述缓冲服务器，以快速对超出所述在线服务器负载能力的业务负载进行负载，这样可以在超出所述在线服务器的负载能力时所述云服务仍可以为用户提供服务，实现零等待、零风险的目的。

2、对扩容服务器进行扩容，如果云服务器中有多个所述扩容服务器，且多个所述扩容服务器的业务负载能力不同，根据当前的业务负载增长速率和各个所述扩容服务器的扩容时间扩容对应负载能力的所述扩容服务器；如果云服务器中有多个所述扩容服务器，且多个所述扩容服务器的业务负载能力相同，根据当前的业务负载增长速率和所述扩容服务器的扩容时间扩容对应个数的所述扩容服务器，具体方法已在上述进行了详细说明，在此不再一一赘述。

3、当对所述扩容服务器扩容完毕后，立即让所述扩容服务器上线。

4、关闭所述缓冲服务器进行业务负载的应用，并将所述缓冲服务器放回所述缓存池中，放入所述缓存池中的所述缓冲服务器可以自动启动进行业务负载的应用，但是不进行业务负载，也可以在调用所述缓冲服务器后，所述缓冲服务器在加载所述应用，以对业务负载进行负载。

5、将所述缓冲服务器负载的业务负载切换到所述扩容服务器中进行业务负载。当超出所述扩容服务器的业务负载能力时，可以按照上述方法再次调用所述缓冲服务器，并启动其他所述扩容服务器。

6、当所述扩容服务器的业务负载为零时，关闭所述扩容服务器。当超出所述在线服务器的业务负载能力时，可以按照上述方法再次调用所述缓冲服务器进行业务负载，并启动其他所述扩容服务器。

基于与上述方法同样的申请构思，本发明提出了一种系统，如图3所示，所述系统中包括主负载设备、临时负载设备和备用负载设备，其中，所述主负载设备为当前正在进行业务负载的设备，所述临时负载设备为随时可以进行临时业务负载的设备，所述备用设备是进行启动后才能进行业务负载的设备，所述系统包括：

判断模块31，用于判断所述主负载设备的业务负载值是否超过阈值；

调用模块32，如果超过阈值，用于调用所述临时负载设备进行业务负载；启动模块33，用于启动所述备用负载设备；

切换模块34，当对所述备用负载设备启动完毕后，用于将所述临时负载设备的业务负载切换到所述备用负载设备上，以使所述备用负载设备进行业务负载。

所述备用负载设备为多个，且不同的所述备用负载设备的负载能力不同，其中，业务负载能力越高的所述备用负载设备的启动时间越长；

所述启动模块，具体用于：

确定当前业务负载的增长速率；

根据所述增长速率和各个所述备用负载设备的启动时间启动对应负载能力的所述备用负载设备。

所述备用负载设备为多个，且各个所述备用负载设备的负载能力相同；

所述启动模块，具体用于：

确定当前业务负载的增长速率；

根据所述增长速率和所述备用负载设备的启动时间启动对应数量的所述备用负载设备。

备。

所述切换模块，具体用于：

关闭所述临时负载设备进行业务负载的应用；

将所述临时负载设备的业务负载切换到所述备用负载设备上。

5 所述系统，还包括：

关闭模块，在将所述临时负载设备的业务负载切换到所述备用负载设备上后，当所述备用负载设备的业务负载为零时，用于关闭所述备用负载设备。

本发明中当主负载设备的业务负载值超过阈值时，调用临时负载设备进行业务负载，并启动备用负载设备，当所述备用负载设备启动完毕后将所述临时负载设备的业务负载
10 切换到所述备用负载设备上。本发明避免了在备用负载启动的过程中由于业务负载持续上升，超过主负载设备的业务负载能力时而引起的服务异常、不可用，甚至挂起现象，实现了零等待、零风险的业务负载。

本领域普通技术人员可以理解：实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时，执行包括上述各方法实施例的步骤；而前述的存储介质包括：**ROM、RAM、**
15 **磁碟或者光盘**等各种可以存储程序代码的介质。

以上所描述的装置实施例仅仅是示意性的，其中作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到至少两个网络单元上。可以根据实际的需要选择
20 其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下，即可以理解并实施。

最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施
25 例技术方案的范围。

权利要求书

1、一种负载方法，其特征在于，所述方法应用于包括主负载设备、临时负载设备和备用负载设备的系统中，其中，所述主负载设备为当前正在进行业务负载的设备，所述临时负载设备为随时可以进行临时业务负载的设备，所述备用设备是进行启动后才能进行业务负载的设备，所述方法包括：

判断所述主负载设备的业务负载值是否超过阈值；

如果超过，调用所述临时负载设备进行业务负载，并启动所述备用负载设备；

当对所述备用负载设备启动完毕后，将所述临时负载设备的业务负载切换到所述备用负载设备上，以使所述备用负载设备进行业务负载。

2、如权利要求 1 所述方法，其特征在于，所述备用负载设备为多个，且不同的所述备用负载设备的负载能力不同，其中，业务负载能力越高的所述备用负载设备的启动时间越长；

启动所述备用负载设备，具体包括：

确定当前业务负载的增长速率；

根据所述增长速率和各个所述备用负载设备的启动时间启动对应负载能力的所述备用负载设备。

3、如权利要求 1 所述方法，其特征在于，所述备用负载设备为多个，且各个所述备用负载设备的负载能力相同；

启动所述备用负载设备，具体包括：

确定当前业务负载的增长速率；

根据所述增长速率和所述备用负载设备的启动时间启动对应数量的所述备用负载设备。

4、如权利要求 1 所述方法，其特征在于，将所述临时负载设备的业务负载切换到所述备用负载设备上，具体包括：

关闭所述临时负载设备进行业务负载的应用；

将所述临时负载设备的业务负载切换到所述备用负载设备上。

5、如权利要求 1 所述方法，其特征在于，在将所述临时负载设备的业务负载切换到所述备用负载设备上后，所述方法还包括：

当所述备用负载设备的业务负载为零时，关闭所述备用负载设备。

6、一种系统，其特征在于，所述系统中包括主负载设备、临时负载设备和备用负

载设备，其中，所述主负载设备为当前正在进行业务负载的设备，所述临时负载设备为随时可以进行临时业务负载的设备，所述备用设备是进行启动后才能进行业务负载的设备，所述系统包括：

判断模块，用于判断所述主负载设备的业务负载值是否超过阈值；

5 调用模块，如果超过阈值，用于调用所述临时负载设备进行业务负载；启动模块，用于启动所述备用负载设备；

切换模块，当对所述备用负载设备启动完毕后，用于将所述临时负载设备的业务负载切换到所述备用负载设备上，以使所述备用负载设备进行业务负载。

7、如权利要求 6 所述系统，其特征在于，所述备用负载设备为多个，且不同的所述备用负载设备的负载能力不同，其中，业务负载能力越高的所述备用负载设备的启动时间越长；

所述启动模块，具体用于：

确定当前业务负载的增长速率；

15 根据所述增长速率和各个所述备用负载设备的启动时间启动对应负载能力的所述备用负载设备。

8、如权利要求 6 所述系统，其特征在于，所述备用负载设备为多个，且各个所述备用负载设备的负载能力相同；

所述启动模块，具体用于：

确定当前业务负载的增长速率；

20 根据所述增长速率和所述备用负载设备的启动时间启动对应数量的所述备用负载设备。

9、如权利要求 6 所述系统，其特征在于，所述切换模块，具体用于：

关闭所述临时负载设备进行业务负载的应用；

将所述临时负载设备的业务负载切换到所述备用负载设备上。

25 10、如权利要求 6 所述系统，其特征在于，所述系统，还包括：

关闭模块，在将所述临时负载设备的业务负载切换到所述备用负载设备上后，当所述备用负载设备的业务负载为零时，用于关闭所述备用负载设备。

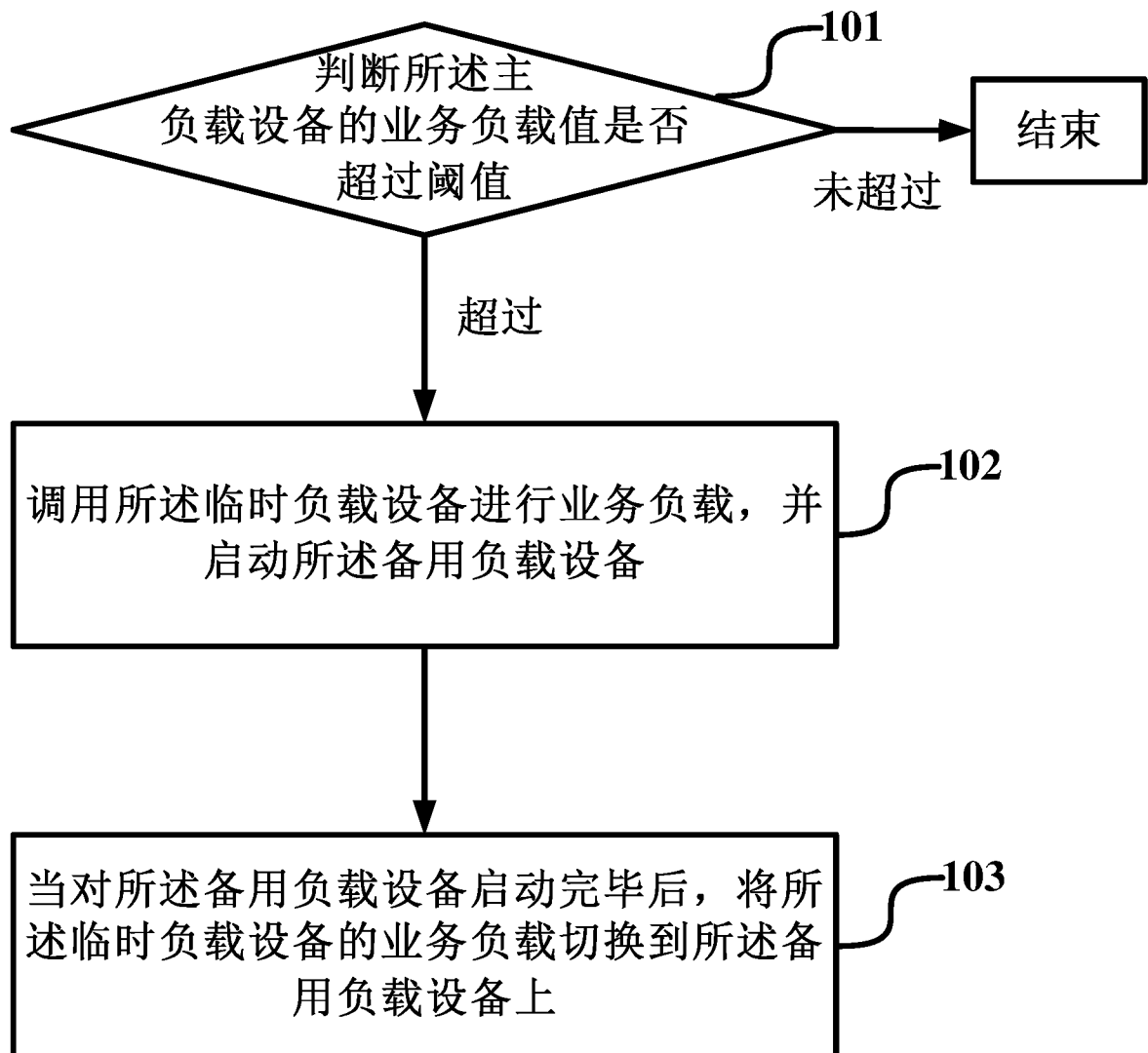


图 1

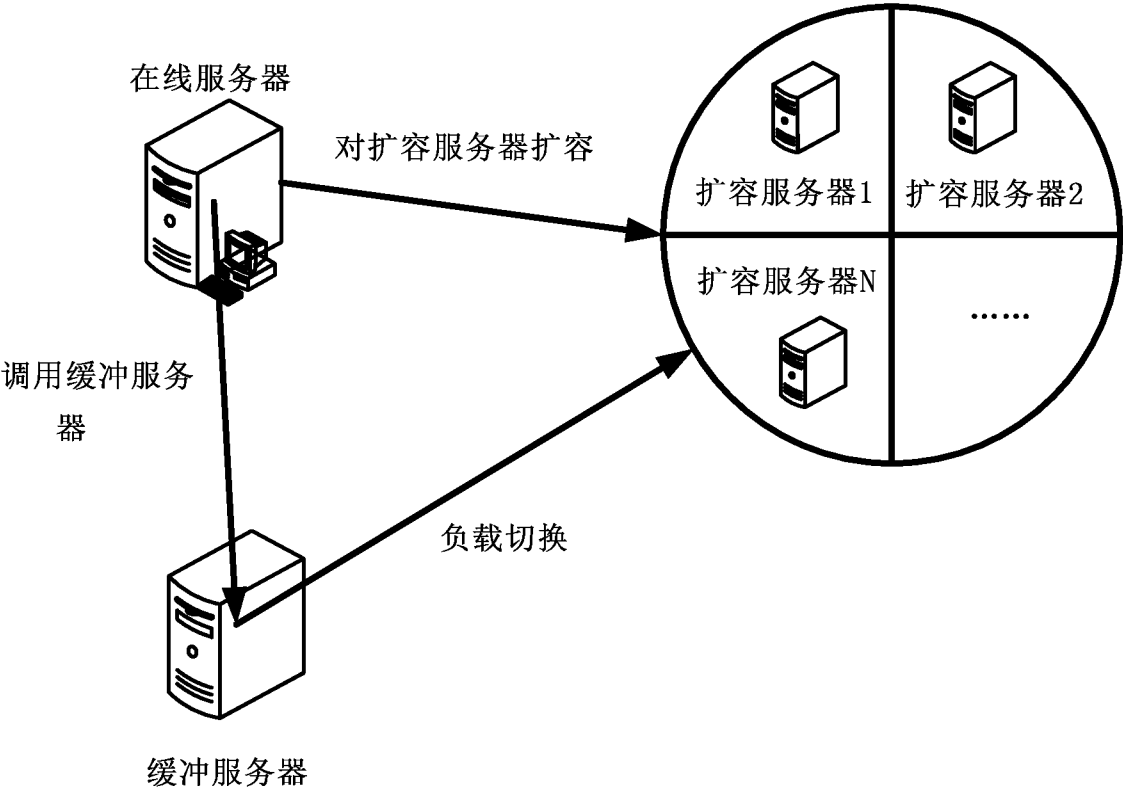


图 2

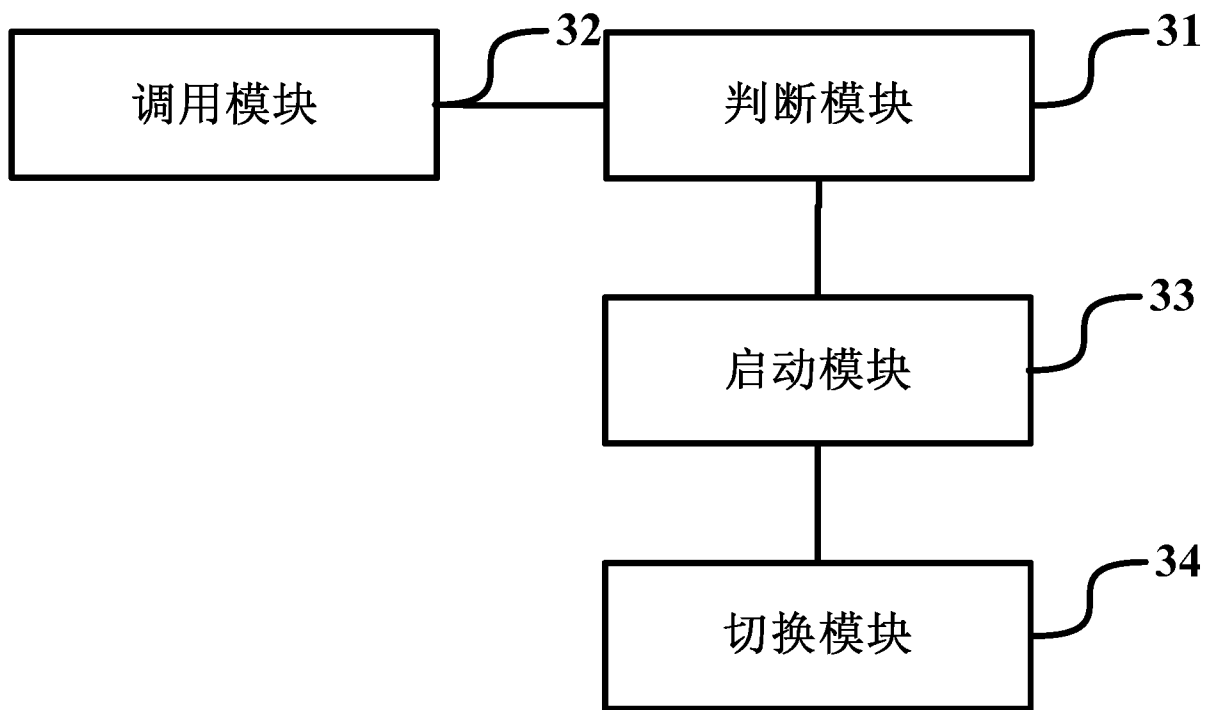


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/076039

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/08 (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)

H04L; H04W; H04Q

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, CNKI, EPODOC, WPI: momentarily, current, cache, capacity expansion, extension, preset value, over, greater than, close, call, temporar+, primary, master, online, work, standby, backup, buffer?, scal+, stretch+, expand???, expansion, idle, spare, load?, server?, virtual machine, host?, VM?, threshold, preset, schedul+, handover+, start?up

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103856512 A (HUAWEI TECHNOLOGIES CO., LTD.), 11 June 2014 (11.06.2014), description, paragraphs [0064]-[0082] and [0098]-[0118], and figures 1, 2, 4 and 5	1-10
A	CN 105024842 A (TENCENT INC.), 04 November 2015 (04.11.2015), the whole document	1-10
A	CN 102868744 A (BEIJING UFIDA GOVERNMENT AFFAIRS SOFTWARE CO., LTD.), 09 January 2013 (09.01.2013), the whole document	1-10
A	US 2003079154 A1 (PARK, K.J. et al.), 24 April 2003 (24.04.2003), the whole document	1-10

☐ 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 07 April 2017 (07.04.2017)	Date of mailing of the international search report 03 May 2017 (03.05.2017)
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 LIU, Hanyan Telephone No.: (86-10) 010-62413201

International application No.
PCT/CN2017/076039

Form PCT/ISA/210 (patent family annex) (July 2009)

<p>A. 主题的分类</p> <p>H04L 29/08 (2006.01) i</p> <p>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</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献 (标明分类系统和分类号)</p> <p>H04L; H04W; H04Q</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNPAT, CNKI, EPODOC, WPI: 临时, 随时, 主, 工作, 在线, 当前, 备用, 备份, 缓冲, 缓存, 扩容, 扩展, 扩充, 空闲, 闲置, 负载, 服务器, 虚拟机, 主机, 阈值, 门限, 预设值, 超过, 大于, 接近, 调用, 切换, 启动, temporar+, primary, master, online, work, standby, backup, buffer?, scal+, stretch+, expand???, expansion, idle, spare, load?, server?, virtual machine, host?, VM?, threshold, preset, schedul+, handover+, start?up</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 103856512 A (华为技术有限公司) 2014年 6月 11日 (2014 - 06 - 11) 说明书第0064-0082、0098-0118段, 图1-2、4-5</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 105024842 A (深圳市腾讯计算机系统有限公司) 2015年 11月 4日 (2015 - 11 - 04) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 102868744 A (北京用友政务软件有限公司) 2013年 1月 9日 (2013 - 01 - 09) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 2003079154 A1 (PARK, KIE JIN等) 2003年 4月 24日 (2003 - 04 - 24) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 103856512 A (华为技术有限公司) 2014年 6月 11日 (2014 - 06 - 11) 说明书第0064-0082、0098-0118段, 图1-2、4-5	1-10	A	CN 105024842 A (深圳市腾讯计算机系统有限公司) 2015年 11月 4日 (2015 - 11 - 04) 全文	1-10	A	CN 102868744 A (北京用友政务软件有限公司) 2013年 1月 9日 (2013 - 01 - 09) 全文	1-10	A	US 2003079154 A1 (PARK, KIE JIN等) 2003年 4月 24日 (2003 - 04 - 24) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
A	CN 103856512 A (华为技术有限公司) 2014年 6月 11日 (2014 - 06 - 11) 说明书第0064-0082、0098-0118段, 图1-2、4-5	1-10															
A	CN 105024842 A (深圳市腾讯计算机系统有限公司) 2015年 11月 4日 (2015 - 11 - 04) 全文	1-10															
A	CN 102868744 A (北京用友政务软件有限公司) 2013年 1月 9日 (2013 - 01 - 09) 全文	1-10															
A	US 2003079154 A1 (PARK, KIE JIN等) 2003年 4月 24日 (2003 - 04 - 24) 全文	1-10															
<p>☐ 其余文件在C栏的续页中列出。</p> <p>☒ 见同族专利附件。</p>																	
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2017年 4月 7日</p>		<p>国际检索报告邮寄日期</p> <p>2017年 5月 3日</p>															
<p>ISA/CN的名称和邮寄地址</p> <p>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</p> <p>传真号 (86-10) 62019451</p>		<p>受权官员</p> <p>刘寒艳</p> <p>电话号码 (86-10) 010-62413201</p>															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/076039

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	103856512	A	2014年 6月 11日	无	
CN	105024842	A	2015年 11月 4日	无	
CN	102868744	A	2013年 1月 9日	无	
US	2003079154	A1	2003年 4月 24日	KR 20030034411	A 2003年 5月 9日