



(12) 发明专利

(10) 授权公告号 CN 106874334 B

(45) 授权公告日 2021.07.09

(21) 申请号 201610664982.3

审查员 鞠博

(22) 申请日 2016.08.12

(65) 同一申请的已公布的文献号

申请公布号 CN 106874334 A

(43) 申请公布日 2017.06.20

(73) 专利权人 创新先进技术有限公司

地址 开曼群岛大开曼岛乔治镇医院路27号

开曼企业中心

(72) 发明人 韩涛 袁自诚

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

代理人 解婷婷 龙洪

(51) Int. Cl.

G06F 16/00 (2019.01)

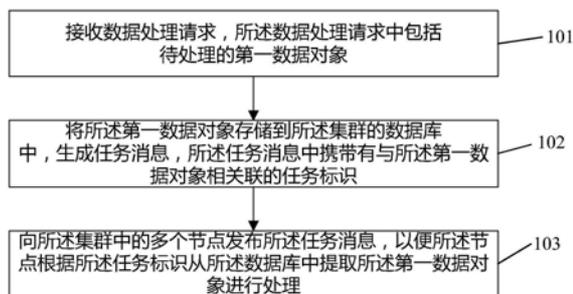
权利要求书4页 说明书18页 附图5页

(54) 发明名称

一种数据处理方法及装置、资讯处理系统

(57) 摘要

本申请公开了一种数据处理方法及装置、资讯处理系统,所述数据处理方法,应用于集群,所述集群包括多个节点,包括:接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。本文不仅实现了数据的异步处理,而且避免了新起线程池和线程队列的使用,降低了复杂度,提升了处理效率,也优化清晰了数据处理流程,且能够充分利用集群中的闲置节点资源,避免资源浪费。



1. 一种数据处理方法,其特征在于,应用于集群,所述集群包括多个节点,包括:
接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;
将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;
向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理;
其中,所述发布任务消息的节点与接收所述任务消息的节点,为同一节点或不同节点。
2. 根据权利要求1所述的数据处理方法,其特征在于,
在将所述第一数据对象存储到所述集群的数据库中之前,还包括:
为所述第一数据对象设置第一状态标识,所述第一状态标识用于标识所述第一数据对象为待处理的状态;
在所述节点根据所述任务标识从所述数据库中提取所述第一数据对象之后,还包括:
为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。
3. 根据权利要求1所述的数据处理方法,其特征在于,
在将所述第一数据对象存储到所述集群的数据库中,生成任务消息之后,还包括:
发送第一响应消息给所述第一数据对象的请求方,所述第一响应消息用于指示所述第一数据对象已收到。
4. 根据权利要求1所述的数据处理方法,其特征在于,在接收所述数据处理请求之后,生成任务消息之前,还包括:
验证所述数据库中是否已存储有所述数据处理请求中待处理的第一数据对象;
如果是,则从所述数据库中查询所述第一数据对象相关联的任务标识,根据所述任务标识提取所述第一数据对象,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态,并对所述第一数据对象进行处理;
如果不是,则创建任务流水,将所述任务流水的任务标识与所述第一数据对象相关联后存储到所述集群的数据库中。
5. 一种数据处理方法,其特征在于,应用于集群,所述集群包括多个节点;包括:
所述节点接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;所述任务消息,为接收到数据处理请求,并将数据处理请求中包括的待处理的第一数据对象存储到所述集群的数据库中之后生成的;
所述节点根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理;
其中,所述接收任务消息的节点与发布任务消息的节点,为同一节点或不同节点。
6. 根据权利要求5所述的数据处理方法,其特征在于,从数据库中提取所述第一数据对象之后,所述进行处理之前,还包括:
所述节点为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。
7. 根据权利要求5或6所述的数据处理方法,其特征在于,从所述集群的数据库中提取所述第一数据对象,包括:

在确定所述第一数据对象被设置为第二状态标识时,从所述数据库中提取所述第一数据对象。

8. 根据权利要求5所述的数据处理方法,其特征在于,从数据库中提取所述第一数据对象并进行处理之后,还包括:

所述节点为所述第一数据对象设置第四状态标识或第五状态标识,所述第四状态标识用于标识所述第一数据对象为处理失败的状态,所述第五状态标识用于标识所述第一数据对象为处理成功的状态。

9. 根据权利要求5所述的数据处理方法,其特征在于,从数据库中提取所述第一数据对象并进行处理之后,还包括:

所述节点将处理完成的所述第一数据对象保存到所述集群的数据库中。

10. 根据权利要求5所述的数据处理方法,其特征在于,从数据库中提取所述第一数据对象并进行处理之后,还包括:

所述节点将处理完成的所述第一数据对象保存到所述集群外的数据库中,以供展示时使用。

11. 根据权利要求5所述的数据处理方法,其特征在于,从数据库中提取所述第一数据对象并进行处理之后,还包括:

发送第二响应消息给所述第一数据对象的请求方,所述第二响应消息用于指示所述第一数据对象已处理完成。

12. 一种数据处理装置,其特征在于,应用于集群,所述集群包括多个节点;包括:

第一接收单元,用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

第一处理单元,用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

发布单元,用于向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理;

其中,所述发布任务消息的节点与接收所述任务消息的节点,为同一节点或不同节点。

13. 根据权利要求12所述的数据处理装置,其特征在于,

所述第一处理单元,还用于在将所述第一数据对象存储到所述集群的数据库中之前,为所述第一数据对象设置第一状态标识,所述第一状态标识用于标识所述第一数据对象为待处理的状态;以及,用于在根据所述任务标识从所述数据库中提取所述第一数据对象之后,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

14. 根据权利要求12所述的数据处理装置,其特征在于,还包括:

第一发送单元,用于发送第一响应消息给所述第一数据对象的请求方,所述第一响应消息用于指示所述第一数据对象已收到。

15. 一种数据处理装置,包括处理器和存储器;其特征在于,应用于集群,所述集群包括多个节点;所述存储器用于存储用于数据处理的程序;所述用于数据处理的程序在被所述处理器读取执行时,执行如下操作:

接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理;

其中,所述发布任务消息的节点与接收所述任务消息的节点,为同一节点或不同节点。

16. 一种数据处理装置,其特征在于,应用于集群,所述集群包括多个节点,所述数据处理装置应用于所述节点,包括:

第二接收单元,用于接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;所述任务消息,为接收到数据处理请求,并将数据处理请求中包括的待处理的第一数据对象存储到所述集群的数据库中之后生成的;

第二处理单元,用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理;

其中,所述接收任务消息的节点与发布任务消息的节点,为同一节点或不同节点。

17. 根据权利要求16所述的数据处理装置,其特征在于,

所述第二处理单元,还用于在从数据库中提取所述第一数据对象之后,所述进行处理之前,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

18. 根据权利要求16或17所述的数据处理装置,其特征在于,

所述第二处理单元,用于在确定所述第一数据对象被设置为第二状态标识时,从所述数据库中提取所述第一数据对象。

19. 根据权利要求16所述的数据处理装置,其特征在于,

所述第二处理单元,还用于从数据库中提取所述第一数据对象并进行处理之后,为所述第一数据对象设置第四状态标识或第五状态标识,所述第四状态标识用于标识所述第一数据对象为处理失败的状态,所述第五状态标识用于标识所述第一数据对象为处理成功的状态。

20. 根据权利要求16所述的数据处理装置,其特征在于,

所述第二处理单元,还用于从数据库中提取所述第一数据对象并进行处理之后,将处理完成的所述第一数据对象保存到所述集群的数据库中。

21. 根据权利要求16所述的数据处理装置,其特征在于,所述第二处理单元,还用于从数据库中提取所述第一数据对象并进行处理之后,将处理完成的所述第一数据对象保存到所述集群外的数据库中,以供展示时使用。

22. 根据权利要求16所述的数据处理装置,其特征在于,还包括:

第二发送单元,用于在所述第二处理单元对所述第一数据对象处理完成之后,发送第二响应消息给所述第一数据对象的请求方,所述第二响应消息用于指示所述第一数据对象已处理完成。

23. 一种数据处理装置,包括处理器和存储器;其特征在于,应用于集群,所述集群包括多个节点,所述数据处理装置应用于所述节点;所述存储器用于存储用于数据处理的程序;所述用于数据处理的程序在被所述处理器读取执行时,执行如下操作:

接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;所

述任务消息,为接收到数据处理请求,并将数据处理请求中包括的待处理的第一数据对象存储到所述集群的数据库中之后生成的;

根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理;

其中,所述接收任务消息的节点与发布任务消息的节点,为同一节点或不同节点。

24.一种资讯处理系统,其特征在于,应用于集群,所述集群包括多个节点,所述多个节点包括:第一节点和第二节点;其中,

所述第一节点,用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;并用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;以及,用于向所述集群中的多个所述第二节点发布所述任务消息;

所述第二节点,用于接收所述第一节点发布的所述任务消息;并用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理;

其中,所述第一节点和所述第二节点,为同一节点或不同节点。

一种数据处理方法及装置、资讯处理系统

技术领域

[0001] 本申请涉及网络技术领域,尤指一种数据处理方法及装置、资讯处理系统。

背景技术

[0002] 网络中,资讯处理系统之间业务数据的实时交互传递一般采用异步交互方式。

[0003] 相关技术中,资讯处理系统与外部系统异步交互的过程是,当前节点的应用系统层接收业务请求,将请求处理的业务数据投入到当前节点上操作系统层已有的线程池或操作系统层新起的线程上,由操作系统层的线程池或新起的线程对业务数据进行处理,虽然实现了异步处理,但需要管理线程池的资源和运行调度,或对主线程外的新线程进行管理,系统复杂度高,处理效率低,且不能有效利用资讯处理系统中的其他节点资源,发生问题时还需要人工排查,系统维护成本也比较高。

[0004] 综上,在实现本申请的过程中,发明人发现,对于业务处理的异步交互中系统复杂度高、不能有效利用其他节点资源以及系统维护成本高的问题,目前还没有好的解决方案。

发明内容

[0005] 为了解决上述技术问题,本申请提供了一种数据处理方法及装置、资讯处理系统。

[0006] 一方面,本申请提供了一种数据处理方法,应用于集群,所述集群包括多个节点,包括:

[0007] 接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

[0008] 将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

[0009] 向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。

[0010] 其中,在将所述第一数据对象存储到所述集群的数据库中之前,还包括:

[0011] 为所述第一数据对象设置第一状态标识,所述第一状态标识用于标识所述第一数据对象为待处理的状态;

[0012] 在所述节点根据所述任务标识从所述数据库中提取所述第一数据对象之后,还包括:

[0013] 为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

[0014] 其中,在将所述第一数据对象存储到所述集群的数据库中,生成任务消息之后,还包括:

[0015] 发送第一响应消息给所述第一数据对象的请求方,所述第一响应消息用于指示所述第一数据对象已收到。

[0016] 其中,在接收所述数据处理请求之后,生成任务消息之前,还包括:

[0017] 验证所述数据库中是否已存储有所述数据处理请求中待处理的第一数据对象;

[0018] 如果是,则从所述数据库中查询所述第一数据对象相关联的任务标识,根据所述任务标识提取所述第一数据对象,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态,并对所述第一数据对象进行处理;

[0019] 如果不是,则创建任务流水,将所述任务流水的任务标识与所述第一数据对象相关联后存储到所述集群的数据库中。

[0020] 另一方面,本申请还提供了一种数据处理方法,应用于集群,所述集群包括多个节点;包括:

[0021] 所述节点接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;

[0022] 所述节点根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。

[0023] 其中,从数据库中提取所述第一数据对象之后,所述进行处理之前,还包括:

[0024] 所述节点为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

[0025] 其中,从所述集群的数据库中提取所述第一数据对象,包括:

[0026] 在确定所述第一数据对象被设置为第二状态标识时,从所述数据库中提取所述第一数据对象。

[0027] 其中,从数据库中提取所述第一数据对象并进行处理之后,还包括:

[0028] 所述节点为所述第一数据对象设置第四状态标识或第五状态标识,所述第四状态标识用于标识所述第一数据对象为处理失败的状态,所述第五状态标识用于标识所述第一数据对象为处理成功的状态。

[0029] 其中,从数据库中提取所述第一数据对象并进行处理之后,还包括:

[0030] 所述节点将处理完成的所述第一数据对象保存到所述集群的数据库中。

[0031] 其中,从数据库中提取所述第一数据对象并进行处理之后,还包括:

[0032] 所述节点将处理完成的所述第一数据对象保存到所述集群外的数据库中,以供展示时使用。

[0033] 其中,从数据库中提取所述第一数据对象并进行处理之后,还包括:

[0034] 发送第二响应消息给所述第一数据对象的请求方,所述第二响应消息用于指示所述第一数据对象已处理完成。

[0035] 一方面,本申请提供了一种数据处理装置,应用于集群,所述集群包括多个节点;包括:

[0036] 第一接收单元,用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

[0037] 第一处理单元,用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

[0038] 发布单元,用于向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。

[0039] 其中,所述第一处理单元,还用于在将所述第一数据对象存储到所述集群的数据库中之前,为所述第一数据对象设置第一状态标识,所述第一状态标识用于标识所述第一

数据对象为待处理的状态;以及,用于在根据所述任务标识从所述数据库中提取所述第一数据对象之后,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

[0040] 其中,还包括:第一发送单元,用于发送第一响应消息给所述第一数据对象的请求方,所述第一响应消息用于指示所述第一数据对象已收到。

[0041] 另一方面,本申请还提供了一种数据处理装置,包括处理器和存储器;其特征在于,应用于集群,所述集群包括多个节点;所述存储器用于存储用于数据处理的程序;所述用于数据处理的程序在被所述处理器读取执行时,执行如下操作:

[0042] 接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

[0043] 将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

[0044] 向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。

[0045] 一方面,本申请还提供了一种数据处理装置,应用于集群,所述集群包括多个节点,所述数据处理装置应用于所述节点,包括:

[0046] 第二接收单元,用于接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;

[0047] 第二处理单元,用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。

[0048] 其中,所述第二处理单元,还用于在从数据库中提取所述第一数据对象之后,所述进行处理之前,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

[0049] 其中,所述第二处理单元,用于在确定所述第一数据对象被设置为第二状态标识时,从所述数据库中提取所述第一数据对象。

[0050] 其中,所述第二处理单元,还用于从数据库中提取所述第一数据对象并进行处理之后,为所述第一数据对象设置第四状态标识或第五状态标识,所述第四状态标识用于标识所述第一数据对象为处理失败的状态,所述第五状态标识用于标识所述第一数据对象为处理成功的状态。

[0051] 其中,所述第二处理单元,还用于从数据库中提取所述第一数据对象并进行处理之后,将处理完成的所述第一数据对象保存到所述集群的数据库中。

[0052] 其中,所述第二处理单元,还用于从数据库中提取所述第一数据对象并进行处理之后,将处理完成的所述第一数据对象保存到所述集群外的数据库中,以供展示时使用。

[0053] 其中,还包括:第二发送单元,用于在所述第二处理单元对所述第一数据对象处理完成之后,发送第二响应消息给所述第一数据对象的请求方,所述第二响应消息用于指示所述第一数据对象已处理完成。

[0054] 另一方面,本申请还提供了一种数据处理装置,包括处理器和存储器;其特征在于,应用于集群,所述集群包括多个节点,所述数据处理装置应用于所述节点;所述存储器用于存储用于数据处理的程序;所述用于数据处理的程序在被所述处理器读取执行时,执行如下操作:

- [0055] 接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;
- [0056] 根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。
- [0057] 另一方面,本申请还提供了一种资讯处理系统,应用于集群,所述集群包括多个节点,所述多个节点包括:第一节点和第二节点;其中,
- [0058] 所述第一节点,用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;并用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;以及,用于向所述集群中的多个所述第二节点发布所述任务消息;
- [0059] 所述第二节点,用于接收所述第一节点发布的所述任务消息;并用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。
- [0060] 本申请包括以下优点:
- [0061] 本申请通过集群中的不同节点执行或同一节点在不同时刻执行数据的处理,不仅实现了数据的异步处理,同时避免了新起线程池和线程队列的使用,降低了复杂度,提升了处理效率,也优化清晰了数据处理流程,且能够充分利用集群中的闲置节点资源,避免资源浪费。
- [0062] 当然,实施本申请的任一产品必不一定需要同时达到以上所述的所有优点。

附图说明

- [0063] 附图用来提供对本申请技术方案的进一步理解,并且构成说明书的一部分,与本申请的实施例一起用于解释本申请的技术方案,并不构成对本申请技术方案的限制。
- [0064] 图1为本申请一种数据处理方法的流程图;
- [0065] 图2为本申请一种数据处理装置的组成结构示意图;
- [0066] 图3为本申请另一种数据处理方法的流程图;
- [0067] 图4为本申请另一种数据处理装置的组成结构示意图;
- [0068] 图5为本申请资讯处理系统的组成结构示意图;
- [0069] 图6为本申请实施例接收数据处理请求到的发布任务消息的数据处理流程示意图;
- [0070] 图7为本申请实施例发布任务消息之后数据处理的流程示意图;
- [0071] 图8为本申请实施例新闻相关的数据处理流程示意图。

具体实施方式

- [0072] 为使本申请的目的、技术方案和优点更加清楚明白,下文中将结合附图对本申请的实施例进行详细说明。需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互任意组合。
- [0073] 在本申请一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。
- [0074] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的

示例。

[0075] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括非暂存电脑可读媒体 (transitory media),如调制的数据信号和载波。

[0076] 在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行。并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0077] 实施例一

[0078] 如图1所示,本申请提供了一种数据处理方法,应用于集群,所述集群包括多个节点,可包括:

[0079] 步骤101,接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

[0080] 步骤102,将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

[0081] 步骤103,向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。

[0082] 在一种实现方式中,可在将所述第一数据对象存储到所述集群的数据库之前,为所述第一数据对象设置第一状态标识,所述第一状态标识用于标识所述第一数据对象为待处理的状态;在所述节点根据所述任务标识从所述数据库中提取所述第一数据对象之后,还可为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。如此,可通过状态标识来标识第一数据对象的处理状态,便于在处理时可通过识别第一数据对象的状态决定是否继续处理。

[0083] 在一种实现方式中,所述任务标识可以是任务流水号。

[0084] 在一种实现方式中,所述第一数据对象可以是资讯。例如,所述第一数据对象可以是新闻、供求、动态、技术、政策、评论、观点和学术等类型的资讯。

[0085] 在一种实现方式中,可以在将所述第一数据对象存储到所述集群的数据库中,生成任务消息之后,发送第一响应消息给所述第一数据对象的请求方,所述第一响应消息用于指示所述第一数据对象已收到。如此,可以及时将第一数据对象的处理情况反馈给请求方。

[0086] 相关技术中,应用B与应用A建立链接,获取待处理的数据对象,在数据处理完成之后才会返回响应消息给应用A,这样的话,在数据处理过程中应用B与应用A之间的链接将会一直处于被占用的状态。对于资讯的处理来讲,处理时间较长,在这较长的时间内一直占用链接,必将会造成通信资源的浪费。

[0087] 本申请中,获取待处理的第一数据对象之后即刻回复响应消息给请求方,请求方收到响应消息之后,可将相关链接断开。如此,在数据处理的长时间内不会一直占用链接,

节省了通信资源。

[0088] 在一种实现方式中,在接收所述数据处理请求之后,生成任务消息之前,还可以验证所述数据处理请求中的第一数据对象是否为重复接收,也就是说,验证所述数据库中是否已存储有所述第一数据对象;如果是(即之前收到过所述第一数据对象),从所述数据库中查询所述第一数据对象相关联的任务标识,根据所述任务标识提取所述第一数据对象,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态,并对所述第一数据对象进行处理;如果不是(即之前没有收到过所述第一数据对象),创建任务流水,将所述任务流水的任务标识与所述第一数据对象相关联后存储到所述集群的数据库中。

[0089] 如图2所示,本申请还提供了一种数据处理装置,应用于集群,所述集群包括多个节点;包括:

[0090] 第一接收单元21,用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

[0091] 第一处理单元22,用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

[0092] 发布单元23,用于向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。

[0093] 在一种实现方式中,所述第一处理单元还可用于在将所述第一数据对象存储到所述集群的数据库中之前,为所述第一数据对象设置第一状态标识,所述第一状态标识用于标识所述第一数据对象为待处理的状态;以及,还可用于在根据所述任务标识从所述数据库中提取所述第一数据对象之后,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

[0094] 在一种实现方式中,所述数据处理装置还可包括:第一发送单元24,用于发送第一响应消息给所述第一数据对象的请求方,所述第一响应消息用于指示所述第一数据对象已收到。

[0095] 实际应用中,第一接收单元21、第一处理单元22、发布单元23以及第一发送单元24分别可以是软件、硬件或两者的结合。所述第一接收单元21可以通过信号接收器或类似设备实现;所述第一处理单元22可以通过数据处理器、CPU或类似设备实现;所述发布单元23和/或第一发送单元24可以通过信号发射器或类似设备实现。上述数据处理装置可以是信号接收器、数据处理器以及信号发射器组成的设备,也可以是具有上述第一接收单元21、第一处理单元22、发布单元23以及第一发送单元24相应功能的服务器或计算设备。

[0096] 在一种实现方式中,上述数据处理装置应用于所述集群中的节点上,所述节点可以是服务器或计算设备。

[0097] 在一种实现方式中,本申请还提供了一种数据处理装置,包括处理器和存储器;应用于集群,所述集群包括多个节点;所述存储器用于存储用于数据处理的程序;所述用于数据处理的程序在被所述处理器读取执行时,执行如下操作:

[0098] 接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;

[0099] 将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;

[0100] 向所述集群中的多个节点发布所述任务消息,以便所述节点根据所述任务标识从所述数据库中提取所述第一数据对象进行处理。

[0101] 需要说明的是,所述用于数据处理的程序在被所述处理器读取执行时还可执行图1所示方法的所有操作,具体可参考上文相关说明。

[0102] 此外,本申请实施例还提供一种计算机可读存储介质,存储有计算机可执行指令,所述计算机可执行指令被执行时实现所述数据处理方法。

[0103] 实施例二

[0104] 如图3所示,本申请还提供了另一种数据处理方法,应用于集群,所述集群包括多个节点;包括:

[0105] 所述节点接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;

[0106] 所述节点根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。

[0107] 在一种实现方式中,从数据库中提取所述第一数据对象之后,所述进行处理之前,所述节点还可为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。如此,可根据第一数据对象的状态进行后续处理。

[0108] 在一种实现方式中,可在确定所述第一数据对象被设置为第二状态标识时,从所述数据库中提取所述第一数据对象。如此,便于识别需要处理的第一数据对象,也便于在处理过程中基于处理进度更新第一数据对象的处理状态。

[0109] 在一种实现方式中,从数据库中提取所述第一数据对象并进行处理之后,所述节点可以为所述第一数据对象设置第四状态标识或第五状态标识,所述第四状态标识用于标识所述第一数据对象为处理失败的状态,所述第五状态标识用于标识所述第一数据对象为处理成功的状态。如此,可以通过标识第一数据对象的处理状态,便于后续通知请求方相应第一数据对象的处理状态。

[0110] 在一种实现方式中,从数据库中提取所述第一数据对象并进行处理之后,所述节点可将处理完成的所述第一数据对象保存到所述集群的数据库中。以便在需要时从集群的数据库中查询第一数据对象是否被处理过以及处理的状态等。

[0111] 在一种实现方式中,从数据库中提取所述第一数据对象并进行处理之后,所述节点可将处理完成的所述第一数据对象保存到所述集群外的数据库中,以供展示时使用。如此,以便在应用中展示第一数据对象时可直接从相应的数据库中读取。

[0112] 在一种实现方式中,从数据库中提取所述第一数据对象并进行处理之后,还可发送第二响应消息给所述第一数据对象的请求方,所述第二响应消息用于指示所述第一数据对象已处理完成。如此,可将第一数据对象的处理结果同时给请求方,以便请求方将该处理结果反馈给所述第一数据对象的发布应用,该发布应用可基于第一数据对象的处理结果确定是否需要向当前应用重新发布。例如,所述第一数据对象可以是资讯,所述发布应用可以是发布该资讯的应用,所述当前应用可以是所述发布应用获取资讯加工之后再展示的应用。

[0113] 如图4所示,本申请还提供了一种数据处理装置,应用于集群,所述集群包括多个节点,所述数据处理装置应用于所述节点,包括:

[0114] 第二接收单元41,用于接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;

[0115] 第二处理单元42,用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。

[0116] 在一种实现方式中,所述第二处理单元42,还可用于在从数据库中提取所述第一数据对象之后,所述进行处理之前,为所述第一数据对象设置第二状态标识,所述第二状态标识用于标识所述第一数据对象为处理中的状态。

[0117] 在一种实现方式中,所述第二处理单元42,可用于在确定所述第一数据对象被设置为第二状态标识时,从所述数据库中提取所述第一数据对象。

[0118] 在一种实现方式中,所述第二处理单元42,还可用于从数据库中提取所述第一数据对象并进行处理之后,为所述第一数据对象设置第四状态标识或第五状态标识,所述第四状态标识用于标识所述第一数据对象为处理失败的状态,所述第五状态标识用于标识所述第一数据对象为处理成功的状态。

[0119] 在一种实现方式中,所述第二处理单元,还可用于从数据库中提取所述第一数据对象并进行处理之后,将处理完成的所述第一数据对象保存到所述集群的数据库中。

[0120] 在一种实现方式中,所述第二处理单元42,还可用于从数据库中提取所述第一数据对象并进行处理之后,将处理完成的所述第一数据对象保存到所述集群外的数据库中,以供展示时使用。

[0121] 在一种实现方式中,还包括:第二发送单元43,用于在所述第二处理单元42对所述第一数据对象处理完成之后,发送第二响应消息给所述第一数据对象的请求方,所述第二响应消息用于指示所述第一数据对象已处理完成。

[0122] 实际应用中,第二接收单元41、第二处理单元42以及第二发送单元43分别可以是软件、硬件或两者的结合。所述第二接收单元41可以通过信号接收器或类似设备实现;所述第二处理单元42可以通过数据处理器、CPU或类似设备实现;所述第二发送单元43可以通过信号发射器或类似设备实现。上述数据处理装置可以是信号接收器、数据处理器以及信号发射器组成的设备,也可以是具有上述二接收单元41、第二处理单元42以及第二发送单元43相应功能的服务器或计算设备。

[0123] 在一种实现方式中,上述数据处理装置应用于所述集群中的节点上,所述节点可以是服务器或计算设备。

[0124] 在一种实现方式中,本申请还提供了一种数据处理装置,包括处理器和存储器;应用于集群,所述集群包括多个节点,所述数据处理装置应用于所述节点;所述存储器用于存储用于数据处理的程序;所述用于数据处理的程序在被所述处理器读取执行时,执行如下操作:

[0125] 接收任务消息,所述任务消息携带有与待处理的第一数据对象相关联的任务标识;

[0126] 根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。

[0127] 需要说明的是,所述用于数据处理的程序在被所述处理器读取执行时还可执行图3所示方法的所有操作,具体可参考上文相关说明。

[0128] 此外,本申请实施例还提供一种计算机可读存储介质,存储有计算机可执行指令,

所述计算机可执行指令被执行时实现所述数据处理方法。

[0129] 实施例三

[0130] 如图5所示,本申请还提供了一种资讯处理系统,应用于集群,所述集群包括多个节点,所述多个节点至少包括:第一节点51和第二节点52;其中,

[0131] 所述第一节点51,用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;并用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;以及,用于向所述集群中的多个所述第二节点发布所述任务消息;

[0132] 所述第二节点52,用于接收所述第一节点发布的所述任务消息;并用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。

[0133] 本实施例中,所述第一节点51还可以用于接收所述第一节点发布的所述任务消息;并用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理。所述第二节点52也可用于接收数据处理请求,所述数据处理请求中包括待处理的第一数据对象;并用于将所述第一数据对象存储到所述集群的数据库中,生成任务消息,所述任务消息中携带有与所述第一数据对象相关联的任务标识;以及,用于向所述集群中的多个所述第二节点发布所述任务消息。也就是说,实际应用中,集群中可以是同一节点发布并接收任务消息,也可以是不同节点执行任务消息的发布和接收。

[0134] 所述第二节点,用于接收所述第一节点发布的所述任务消息;并用于根据所述任务标识,从所述集群的数据库中提取所述第一数据对象并进行处理

[0135] 相关技术中的异步处理是:集群中的一个节点接收到数据处理请求之后,将其中待处理的数据对象保存到数据库,然后在该节点上新起线程或使用已有的线程池中线程队列,从数据库中捞取所述数据对象进行处理。在数据对象处理完成之后,再反馈响应消息给发出上述数据处理请求的请求方。

[0136] 本申请中,集群中的节点接收数据处理请求,将其中待处理的数据对象保存到数据库;然后,生成任务消息,在该任务消息中携带所述数据对象相关联的任务标识,将所述任务消息发布出去以便集群中的其他节点可接收所述任务消息。在将任务消息发布之后,所述节点反馈第一响应消息给发出所述数据处理请求的请求方。集群中的任意节点接收所述任务消息,根据其中的任务标识从第一数据库中提取所述数据对象进行处理,在处理完成后,反馈第二相应消息给所述请求方,该第二响应消息中包含所述数据对象的处理结果。如此,本申请利用消息机制,将数据处理解耦为两步,通过集群中的不同节点执行或同一节点在不同时刻执行,不仅实现了数据的异步处理,而且避免了新起线程池和线程队列的使用,降低了集群系统的复杂度,提升了处理效率,也优化清晰了数据处理流程,且能够充分利用集群中的闲置节点资源,避免资源浪费。

[0137] 在一些实现方式中,集群可通过调用消息中间件来实现所述任务消息在发布和接收之间的传递。具体的,可在消息中间件中预配置对应所述任务消息的主题,并配置集群中全部或部分节点为该主题的订阅者和发布者。集群中的节点发布任务消息之后,消息中间件接收所述任务消息并将该任务消息投递出去,作为该任务消息订阅者的节点接收所述任务消息,根据其中的任务标识从集群的数据库中提取数据对象进行处理。例如,可以在消息中间件的服务器上可以定义对应所述任务消息的第一主题,并配置所述第一主题的发布者

和订阅者为集群中的全部或部分节点(两个或两个以上的节点)。集群中的节点根据所述第一主题生成所述任务消息并发布,消息中间件接收所述任务消息,并向所述第一主题的订阅者投递所述任务消息,集群中的多个节点作为所述订阅者将会接收到所述任务消息,该任务消息中包含所述任务流水号。如此,通过消息中间件实现任务消息在发布和接收之间的传递,可实现所述任务消息的可靠传递,可以降低集群中节点之间的耦合度,提高所述集群的吞吐量、可扩展性和高可用性。

[0138] 在一些实现方式中,在集群中所述任务消息的发布与接收之间的传递过程,可以通过如下方式实现:

[0139] 在第一服务器和集群中所有节点上预配置对应所述任务消息的第一消息类型,并在所述第一服务器配置所述第一消息类型的发送端地址列表和接收端地址列表,将集群所有节点的地址同时写入所述第一消息类型的所述发送端地址列表和接收端地址列表;

[0140] 集群中的第一节点接收数据处理请求之后,将所述数据处理请求中待处理的第一数据对象存储到集群的数据库中,基于预配置的所述第一消息类型生成所述任务消息,该任务消息的内容包含与所述第一数据对象相关联的任务标识(如任务流水号等),并将该任务消息发布;在发布所述任务消息时,还携带集群中节点的地址;

[0141] 所述第一服务器接收所述任务消息,根据所述任务消息所对应第一消息类型的接收端地址列表将所述任务消息投递出去;

[0142] 集群中的第二节点接收所述第一服务器投递的任务消息,根据所述任务消息中的任务标识从集群的数据库提取所述第一数据对象并进行处理。

[0143] 其中,所述第一服务器可以是具有消息传递功能的服务器,例如:可以是消息中间件的服务器等。所述第一消息类型可由主题(TOPIC)/事件编码(eventcode)唯一标识。

[0144] 示例一

[0145] 如图6所示,从接收数据处理请求到发布任务消息的数据处理过程可以包括:

[0146] 步骤601,应用A发布新闻,将新发布的新闻发送给应用B;

[0147] 步骤602,应用B中的网关服务器接收所述新闻,对所述新闻验签;

[0148] 具体的,验签可以包括:验证所述新闻是否合法,如果合法,则对所述新闻进行报文解析,如果不合法,则验签失败,可以返回验签失败的响应消息给所述应用A,本流程结束。

[0149] 步骤603,验签成功,网关服务器向资讯处理集群发送数据处理请求,所述数据处理请求包含所述新闻;

[0150] 步骤604,资讯处理集群中一个节点接收所述数据处理请求,验证所述新闻是否接收过,如果是则继续步骤612,如果没有接收过,则继续步骤605;

[0151] 实际应用中,应用A会将同一新闻重复发送,相应的,网关服务器也会以不同的数据处理请求发送同一新闻,因此,在本示例中先验证是否之前收到过所述新闻,如果收到过(即重收)则可以就进入步骤612,如果之前没有收到过(也就是说非重复接收)再继续步骤605,为该新闻创建任务流水。如此,可避免为同一新闻重复创建任务流水,一方面减少冗余,防止对同一新闻进行多次重复处理;另一方面也可避免同一新闻在数据库重复存储而占用存储空间。

[0152] 一种实现方式中,验证是否为重发可以通过如下方式实现:验证数据库是否已存

在与其中待处理的新闻相关联的任务流水,如果已存在与其中待处理的新闻相关联的任务流水,则所述新闻为重发,如果不存在与其中待处理的新闻相关联的任务流水,则所述新闻为非重发。

[0153] 一种实现方式中,验证是否为重发还可以通过如下方式实现:验证数据库是否已存在与所述数据处理请求中的新闻相匹配的新闻,如果已存在,则所述新闻为重发,如果不存在,则所述新闻为非重发。

[0154] 步骤605,所述节点为所述数据处理请求创建异步处理任务的任务流水,该任务流水与所述数据处理请求中携带的新闻相关联,并将其中待处理的新闻存储到资讯处理集群的数据库中;

[0155] 步骤606,验证是否创建任务流水成功;

[0156] 步骤607,如果创建任务流水成功,则并为所述新闻设置第一状态标识,所述第一状态标识用于标识所述新闻为待处理的状态;

[0157] 步骤608,如果创建任务流水失败,则为所述新闻设置第三状态标识,所述第三状态标识用于标识任务执行失败的状态;在一些实现方式中,创建任务流水失败可以是所述新闻的任务流水在执行过程中的某个环节执行失败,例如,可以是如下的一项或几项:任务流水号生成失败;将任务流水写入数据库时失败(例如,写入的格式不符合预定要求);在将任务流水与所述新闻的图片和/或文本进行关联时执行失败;在为所述新闻设置状态标识时执行失败。

[0158] 也就是说,在任务流水的执行过程中失败,则将所述新闻设置为第一状态标识。

[0159] 在一些实现方式中,步骤608之后可直接结束当前流程,也可以返回步骤605重新建立任务流水,还可以直接进入图7所示的数据处理流程。

[0160] 步骤609,所述节点生成任务消息并向资讯处理集群中的多个节点发布所述任务消息,所述任务消息中携带有与所述任务流水的任务流水号,通知资讯处理集群中其他节点处理任务,以便其他节点根据所述任务流水号从所述数据库中提取所述新闻进行处理;

[0161] 步骤610,所述节点发出第一响应消息给网关服务器,所述第一响应消息用于指示所述数据处理请求受理成功;

[0162] 步骤611,网关服务器将所述第一响应消息进行报文组装、签名等处理后转发给所述应用A,通知应用A所述新闻已受理成功;

[0163] 在一些实现方式中,应用A在接收所述第一响应消息之后,还可以将自身与应用B中网关服务器之间的链接断开或者将该链接设置为休眠状态,如此,在处理新闻的过程中应用之间的链接不会一直处于被占用的状态,能够节省网络资源,避免网络资源的浪费。

[0164] 此后,应用A与应用B之间链接可暂时断开。

[0165] 步骤612,所述节点在第一数据库中查询所述新闻相关联的任务流水,以获取所述新闻的流水信息;

[0166] 步骤613,所述节点为所述新闻设置第二状态标识,所述第二状态标识用于标识所述新闻为处理中的状态;

[0167] 步骤614,对所述任务流水进行处理,进入图7所示的数据处理流程。

[0168] 示例二

[0169] 如图7所示,在发布任务消息之后,数据处理的流程可以包括如下步骤:

[0170] 步骤701,资讯处理集群中节点接收任务消息,根据所述任务消息中的任务流水号加载相应的异步处理任务,从第一数据库中查询任务流水信息;

[0171] 步骤702,确定所述任务流水信息相关联的新闻被设置为第二状态标识,(即确定所述新闻为处理中的状态),从所述第一数据库中提取所述新闻;

[0172] 在一种实现方式中,在提取任务流水信息和所述新闻之前,还根据所述任务流水号查询所述第一数据库中所述新闻是否被设置为第二状态标识或第一状态标识,如果所述新闻被设置为第二状态标识,则直接从所述第一数据库中提取所述新闻,并开始执行业务处理。如果所述新闻被设置为第一状态标识,则可以先将所述新闻设置为第二状态标识,在确定所述新闻被设置为第二状态标识时从所述第一数据库中提取所述新闻。如果所述新闻不是被设置第二状态标识和第一状态标识,则可以直接跳转到步骤712,并在所述第二响应消息中所述新闻的处理结果为处理失败;或者,可以直接结束当前流程。

[0173] 步骤703,转换新闻数据,如果转换成功,则继续步骤704,如果异常,则继续步骤709;

[0174] 步骤704,进行安全扫描,如果安全扫描成功,则继续步骤705,如果异常,则继续步骤709;

[0175] 步骤705,保存新闻图片,如果保存新闻成功,则继续步骤706,如果异常,则继续步骤709;

[0176] 步骤706,替换所述新闻中正文图片中的imgkey为CDN路径,如果成功,则继续步骤707,如果异常,则继续步骤709;

[0177] 步骤707,将处理后的新闻写入第一数据库,如果成功,则继续步骤708,如果异常,则继续步骤709;

[0178] 步骤708,将处理后的新闻写入第二数据库,如果成功,则继续步骤710,如果异常,则继续步骤709;

[0179] 步骤709,为所述新闻设置第四状态标识,所述第四状态标识用于标识所述新闻为处理错误的状态,继续步骤712;

[0180] 步骤710,所述节点对所述新闻进行文本去重;

[0181] 步骤711,所述节点为所述新闻设置第五状态标识,所述第五状态标识用于标识所述新闻为处理成功的状态;

[0182] 步骤712,所述节点发送第二响应消息给网关服务器,所述第二响应消息用于指示所述新闻的处理结果;

[0183] 其中,如果所述新闻最终被设置为第五状态标识,则所述第二响应消息中指示所述新闻的处理结果为处理成功;如果所述新闻最终被设置为第四状态标识或者在步骤702中确定所述新闻不是被设置第二状态标识和第一状态标识(例如,被设置为第三状态标识),则所述第二响应消息中指示所述新闻的处理结果为处理异常(即处理失败)。

[0184] 步骤713,网关服务器将所述第二响应消息进行报文组装、签名等处理后转发给所述应用A,通知应用A所述新闻已处理完成。

[0185] 在一些实现方式中,网关服务器接收所述第二响应消息之后,如果与应用A之间的链接已断开,可先与应用A建立链接,如果与应用A之间的链接已休眠,可先将该链接激活,然后通过新建立的链接或激活的链接将所述第二响应消息转发给应用A。如此,只在传送消

息时占用应用之间的链接,使得网络资源得到了充分利用,避免了网络资源的浪费,节省了网络资源。

[0186] 步骤714,应用A接收所述第二响应消息,并记录所述新闻的处理结果。

[0187] 在一些实现方式,应用A在接收到所述第二响应消息之后,解析获得新闻的处理结果并记录。对于处理错误的新闻,应用A还可以将所述新闻重发给应用B的网关服务器,由网关服务器再重发数据处理请求给资讯处理集群,以便资讯处理集群对所述新闻再次进行处理,如此,可避免某些新闻被漏处理,提升新闻处理的准确率,以便应用B将更全面的资讯展示给用户,提升用户体验。

[0188] 需要说明的是,图7所示流程中步骤703至步骤708的执行顺序可根据实际需要调整。步骤710也可以在步骤703之前,也可以设置在步骤703与步骤708之间,对此也可根据实际需要调整,本文不做限制。

[0189] 在一些实现方式中,图7所示流程中步骤703至步骤708的新闻数据处理过程中,如其中部分步骤在执行时出现异常,其他步骤仍可执行,也可停止执行,可直接进入步骤709,将所述新闻的状态标识更新为状态标识U,即处理错误,并在步骤712中的第二响应消息中包含处理失败的结果。

[0190] 示例三

[0191] 如图8所示,数据处理的过程可以包括:

[0192] 步骤801,应用A发布新闻,将新发布的新闻发送给应用B;

[0193] 其中,所述新闻可以包括标题、正文和图片。

[0194] 步骤802-804,应用B的网关服务器接收所述新闻,对所述新闻进行验签和报文解析后,向资讯处理集群发送数据处理请求,所述数据处理请求包含所述新闻;

[0195] 这里,所述数据处理请求可以是HTTP请求,可以包含所述新闻的数据和事先约定的加密封文(以防止网络劫持篡改)。该数据处理请求可以通过HTTP方式传递。

[0196] 在一些实现方式中,应用A和应用B之间交互也可以为通过HTTP方式。

[0197] 步骤805,资讯处理集群的第一节点接收所述数据处理请求,创建与其中的新闻相关联的任务流水,生成任务消息,并将所述新闻与所述任务流水的信息对应存储到第一数据库中,所述任务消息携带有所述任务流水的任务流水号;

[0198] 在一些实现方式中,接收所述数据处理请求之后,还可以验证是否为重发,具体实现过程可参照上文中步骤604的相关记载,不再赘述。

[0199] 在一些实现方式中,还可以验证所述任务流水是否创建成功,具体实现过程可参照上文步骤606至步骤608的相关记载,不再赘述。

[0200] 在一些实现方式中,所述任务流水表示新闻的处理进程,任务流水通过任务流水号标识,该任务流水号与所述新闻相关联。

[0201] 在一些实现方式中,将所述新闻及其相关联的任务流水的信息对应存储到第一数据库可以通过如下方式实现:保存时,对接收到的“新闻数据”进行分类,将所述新闻的正文和图片分开存放到第一数据库中;同时,在第一数据库中写入一条任务流水信息,并将该任务流水信息与所述新闻的正文和图片关联起来,所述任务流水信息中包含所述新闻相关联的任务流水号和为所述新闻设置的状态标识。

[0202] 在一些实现方式中,所述任务流水信息可以包括如下字段:任务流水号、状态标识

和新闻ID;其中,新闻ID字段用于记录相关联新闻的ID;状态标识字段用于记录为新闻设置的状态标识,所述状态标识用于标识所述新闻的处理状态(即所述新闻处理到了哪个环节);任务流水号字段用于记录任务流水号。

[0203] 在一些实现方式中,将任务流水与新闻进行关联的方式可以是:将新闻存储在新闻表中,该新闻表中设置新闻ID字段,该新闻ID字段记录有新闻的ID;将任务流水的任务流水号存储在任务流水信息中,并在任务流水信息中设置新闻ID字段,该新闻ID字段也记录新闻的ID,如此,通过新闻的ID可将新闻表与任务流水信息相关联,也就是将新闻与任务流水关联。

[0204] 在一些实现方式中,所述任务流水号可以是时间戳,该时间戳为所述任务流水的建立时间,该时间戳的精确度可以到毫秒级,因此,通过任务流水号即可区分不同的任务流水。

[0205] 本实施例中,所述新闻相关联的状态标识可以包括如下类型:

[0206] I-初始化:标识所述新闻为待处理的状态,也就是说“刚收到新闻,任务刚刚创建,还没执行”;

[0207] F-任务执行失败:标识所述新闻为处理失败的状态,也就是说“新闻在某环节处理失败,不符合业务要求”;例如,任务流水号生成失败;未能成功将任务流水写入数据库或写入的格式不符合预定要求;在将任务流水与所述新闻的图片和/或文本进行关联时执行失败;在为所述新闻设置状态标识时执行失败;诸如此类。在任务流水的执行过程中失败时将所述新闻设置为状态标识F。

[0208] P-处理中:标识所述新闻为处理中的状态,也就是说“收到新闻了,任务开始执行,但尚未执行完成”;

[0209] U-处理发生了错误:标识所述新闻为处理失败的状态,也就是说“新闻在某环节处理时发生了错误”;例如,在对所述新闻进行数据转换、安全扫描、图片的保存和路径替换、处理后的新闻数据写入数据库等操作中发生错误。此时,为所述新闻设置状态标识U。

[0210] S-成功:标识所述新闻为处理成功的状态,也就是说“任务处理成功,新闻可成功放开,外部用户可见”。

[0211] 在本步骤中,还为所述新闻设置状态标识I,该状态标识I用于标识所述新闻为待处理的状态。具体的,设置状态标识I的方式可以是:在资讯处理集群数据中将所述新闻相关联的任务流水信息中状态标识字段设置为状态标识“I”。

[0212] 在一些实现方式中,如果本步骤中出现了任务执行失败的情况,则可重新执行本步骤,直到任务流水信息成功写入数据库并与所述新闻成功建立关联;或者,可以将所述新闻设置为状态标识F,并直接跳转到步骤818;或者,在有任务流水号的情况下,仍生成携带该任务流水号的任务消息,并将所述新闻设置为状态标识F;或者,可以直接结束本流程。

[0213] 步骤806,资讯处理集群的第一节点发布所述任务消息,通知资讯处理集群中其他节点处理任务,使得资讯处理集群中的所有节点都可以从所述第一数据库中提取所述新闻进行处理;

[0214] 步骤807,资讯处理集群的第一节点返回受理成功消息给网关服务器;

[0215] 步骤808,网关服务器将所述受理成功消息转发给应用A;

[0216] 其中,所述受理成功消息为上文中第一响应消息的一个示例,表示所述新闻已被

受理成功。

[0217] 在一些实现方式中,应用A在接收所述受理成功消息之后,还可以将自身与应用B中网关服务器之间的链接断开或者将该链接设置为休眠状态,如此,在处理新闻的过程中应用之间的链接不会一直处于被占用的状态,能够节省网络资源,避免网络资源的浪费。

[0218] 步骤809,资讯处理集群中的节点(可以是上文中的第一节点,也可以是资讯处理集群中除所述第一节点之外的其他节点)接收所述任务消息;

[0219] 步骤810,资讯处理集群中的节点根据所述任务消息中的任务流水号,从第一数据库中提取任务流水信息和所述新闻,开始执行业务处理;

[0220] 此时,还为所述新闻设置状态标识P,该状态标识P表示所述新闻处于处理中的状态。具体的,设置状态标识P的方式可以是:将所述新闻相关联的任务流水信息中“状态标识”的字段设置为状态标识“P”。

[0221] 在一些实现方式中,提取任务流水信息和所述新闻时,还可根据所述任务流水号查询所述第一数据库中所述新闻是否被设置为状态标识P或状态标识I,如果所述新闻被设置为状态标识P,则直接从所述第一数据库中提取任务流水信息和所述新闻,并开始执行业务处理。如果所述新闻被设置为状态标识I,则可以先将所述新闻设置为状态标识P,再从所述第一数据库中提取任务流水信息和所述新闻,并开始执行业务处理。如果所述新闻不是被设置状态标识P和状态标识I(例如,被设置为状态标识F),则不执行业务处理,可以直接进入步骤818或步骤819,或者直接结束当前流程。所述新闻不是被设置状态标识P和状态标识I之后的执行过程为可选,因此,未在图8中示出。

[0222] 步骤811,资讯处理集群中的节点通过安全扫描系统对所述新闻进行安全扫描;

[0223] 在一些实现方式中,需要通过与文本安全扫描系统交互来实现所述安全扫描。例如,可以将新闻发送给文本安全扫描系统,由文本安全扫描系统对所述新闻进行安全扫描之后再所述新闻返回给资讯处理集群中的所述节点。实际应用中,安全扫描的功能也可内置在资讯处理集群的节点中,由节点直接进行安全扫描。

[0224] 在一些实现方式中,安全扫描具体可以是扫描有无违禁词。

[0225] 步骤812,资讯处理集群中的节点将所述新闻中的图片保存到CDN;

[0226] 在一些实现方式中,保存图片到CDN是将所述新闻的图片放到专门保存图片的服务器(如CDN)上;

[0227] 具体的,资讯处理集群中的节点在将图片保存到CDN之后,还记录所述图片的CDN路径。

[0228] 步骤813,资讯处理集群中的节点替换所述新闻中正文部分的图片标签;

[0229] 在一些实现方式中,替换所述新闻中正文部分的图片标签具体可以是:把所述新闻中正文部分的图片标签替换为资讯处理集群识别的标签,例如,替换所述所述新闻中正文图片的imgkey为CND路径。

[0230] 步骤814,资讯处理集群中的节点对所述新闻的正文数据进行文本去重处理;

[0231] 在一些实现方式中,文本去重处理可以是验证当前新闻与其他新闻是否重复。

[0232] 步骤815,资讯处理集群中的节点对所述新闻进行新闻模型转换;

[0233] 在一些实现方式中,转换新闻模型可以是对所述新闻重新排版;也就是说,将所述新闻按照应用B预定的模式重新排版。

[0234] 步骤816,资讯处理集群中的节点将处理后的新闻插入第一数据库的新闻表中;

[0235] 在一些实现方式中,第一数据库中的新闻表用于搜索使用。可通过搜索第一数据库中的新闻表提取新闻进行处理,或者在进行上述新闻处理的过程中需要查询新闻数据时可通过搜索第一数据库中的新闻表实现。

[0236] 在一些实现方式中,第一数据库通过新闻表保存新闻,并将新闻表与所述新闻相关联的任务流水信息相关联。实际应用中,可以使用一张“新闻表”记录一条“新闻”的数据,也可以使用一张“新闻表”记录多个不同“新闻”的数据。在上述新闻表的不同行或不同列中记录所述新闻的正文和图片,并以“新闻ID”标识新闻与具体数据之间的关系,如此,可将所述新闻的正文和图片分开存储,有利于提升数据的处理效率。

[0237] 步骤817,资讯处理集群中的节点将处理后的新闻插入第二数据库的新闻表中,便于应用B的展示系统从所述第二数据库读取所述新闻并展示;

[0238] 其中,第二数据库中也可以通过新闻表的形式记录所述新闻第二数据库中的新闻表为用户可看的数据表,也就是说,第二数据库中存储的新闻是可以直接展示的,第二数据库中存储的新闻是已经处理完成并且处理成功的新闻。在mobilem中新闻表存储所述新闻的方式与第一数据库相同,不再赘述。

[0239] 步骤818,资讯处理集群中的节点更新第一数据库中所述新闻相关联的状态标识,也就是说,更新与所述新闻相关联的任务流水信息;

[0240] 本步骤中,更新第一数据库中所述新闻相关联的状态标识为U/F/S,如果在步骤817之后,所述新闻处理成功,则更新状态标识为S;如果在步骤817之后,所述新闻在处理过程中发生错误,则更新状态标识为U;如果在上述步骤805至步骤817之中,所述新闻相关联的任务流水执行失败,则更新状态标识为F。

[0241] 步骤819,资讯处理集群中的节点发送结果回执消息给网关服务器,所述结果回执消息用于指示所述新闻的处理结果;

[0242] 如果所述新闻最终被设置为状态标识S,则在所述结果回执消息中指示所述新闻的处理结果为处理成功;如果所述新闻最终被设置为状态标识U,则在所述结果回执消息中指示所述新闻的处理结果为处理失败。如果所述新闻最终被设置为状态标识F,则在所述结果回执消息中指示所述新闻的处理结果为处理失败。

[0243] 步骤820,网关服务器将所述结果回执消息转发给所述应用A,通知应用A所述新闻已处理完成。

[0244] 其中,所述结果回执消息为上文中第二响应消息的一个示例,所述结果回执消息用于指示所述新闻已被处理完成,该结果回执消息中可包含所述新闻的处理结果。

[0245] 在一些实现方式中,网关服务器接收所述结果回执消息之后,如果与应用A之间的链接已断开,可先与应用A建立链接,如果与应用A之间的链接已休眠,可先将该链接激活,然后通过新建立的链接或激活的链接将所述结果回执消息转发给应用A。如此,在需要传送消息时才启用应用之间的链接资源,使得网络资源得到了充分利用,避免了网络资源的浪费,节省了网络资源。

[0246] 在一些实现方式,应用A可以接收所述结果回执消息,并记录所述新闻的处理结果。对于处理错误的新闻,应用A还可以将所述新闻重发给应用B的网关服务器,由网关服务器再重发数据处理请求给资讯处理集群,以便资讯处理集群对所述新闻再次进行处理,如

此,可避免某些新闻被漏处理,提升新闻处理的准确率,以便应用B将更全面的资讯展示给用户,提升用户体验。

[0247] 需要说明的是,图8所示流程中步骤811至步骤815、以及的执行顺序可根据实际需要调整,本文不做限制。步骤818也可以在步骤815之后,也可以在步骤817中执行,还可在步骤816之后执行,其执行顺序可根据实现需要调整,本文不做限制。

[0248] 需要说明的是,图8所示步骤811至步骤816中的新闻数据处理过程中,如其中部分步骤在执行时出现异常,其他步骤仍可执行,也可停止执行,直接可跳过步骤817直接进入步骤818,将所述新闻的状态标识更新为状态标识U,即处理错误,并在步骤819中的结果回执消息中包含处理失败的结果。

[0249] 本实施例中,应用A实际可以是发布新闻等资讯的服务器或者集群;网关服务器实际可以是应用B的网关系统,该网关系统实际可以是服务器或服务器集群,第一数据库是用于保存需要处理的数据对象(如待处理或正在处理的或已经处理完成的新闻等资讯),第二数据库是用于保存待展示的数据对象(如已处理成功可展示给用户的新闻等资讯)。在一些实现方式中,第一数据库和第二数据库均可以为常规关系型数据库(mysql)。

[0250] 本文提到的资讯处理集群可以是包含多个机器的集群,每个机器可作为一个节点,每个节点都可独立处理数据,也可以多个节点共同协作来处理数据业务。

[0251] 在一些实现方式中,资讯处理集群可通过调用消息中间件来实现所述任务消息在发布和接收之间的传递。通过消息中间件,可实现所述任务消息的可靠传递,可以降低集群中节点之间的耦合度,提高集群的吞吐量、可扩展性和高可用性。具体的,可在消息中间件中预配置对应所述任务消息的主题,并配置资讯处理集群中全部或部分节点为该主题的订阅者和发布者。资讯处理集群中的节点发布任务消息之后,消息中间件接收所述任务消息并将该任务消息投递出去,作为该任务消息订阅者的节点接收所述任务消息,根据其中的任务标识从第一数据库中提取数据对象进行处理。例如,可以在消息中间件的服务器上可以定义对应所述任务消息的第一主题,并配置所述第一主题的发布者和订阅者为资讯处理集群中的全部或部分节点(两个或两个以上的节点)。资讯处理集群中的节点根据所述第一主题生成所述任务消息并发布,消息中间件接收所述任务消息,并向所述第一主题的订阅者投递所述任务消息,资讯处理集群中的多个节点作为所述订阅者将会接收到所述任务消息,该任务消息中包含所述任务流水号。

[0252] 在一些实现方式中,在资讯处理集群中的节点之间传递所述任务消息的过程,可以通过如下方式实现:

[0253] 在第一服务器和资讯处理集群中所有节点上预配置对应所述任务消息的第一消息类型,并在所述第一服务器配置所述第一消息类型的发送端地址列表和接收端地址列表,将资讯处理集群所有节点的地址同时写入所述第一消息类型的所述发送端地址列表和接收端地址列表;

[0254] 资讯处理集群中的第一节点接收数据处理请求之后,将所述数据处理请求中待处理的第一数据对象存储到资讯处理集群的数据库中,基于预配置的所述第一消息类型生成所述任务消息,该任务消息的内容包含与所述第一数据对象相关联的任务标识(如任务流水号等),并将该任务消息发布;在发布所述任务消息时,还携带资讯处理集群中节点的地址;

[0255] 所述第一服务器接收所述任务消息,根据所述任务消息所对应第一消息类型的接收端地址列表将所述任务消息投递出去;

[0256] 资讯处理集群中的第二节点接收所述第一服务器投递的任务消息,根据所述任务消息中的任务标识从资讯处理集群的数据库提取所述第一数据对象并进行处理。

[0257] 其中,所述第一服务器可以是具有消息传递功能的服务器,例如:可以是消息中间件的服务器等。所述第一消息类型由主题(TOPIC)/事件编码(eventcode)唯一标识。

[0258] 本领域的技术人员应该明白,上述的本申请实施例所提供的装置的各组成部分,以及方法中的各步骤,它们可以集中在单个的计算装置上,或者分布在多个计算装置所组成的网络上。可选地,它们可以用计算装置可执行的程序代码来实现。从而,可以将它们存储在存储装置中由计算装置来执行,或者将它们分别制作成各个集成电路模块,或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样,本申请不限制于任何特定的硬件和软件结合。

[0259] 虽然本申请所揭露的实施方式如上,但所述的内容仅为便于理解本申请而采用的实施方式,并非用以限定本申请。任何本申请所属领域内的技术人员,在不脱离本申请所揭露的精神和范围的前提下,可以在实施的形式及细节上进行任何的修改与变化,但本申请的专利保护范围,仍须以所附的权利要求书所界定的范围为准。

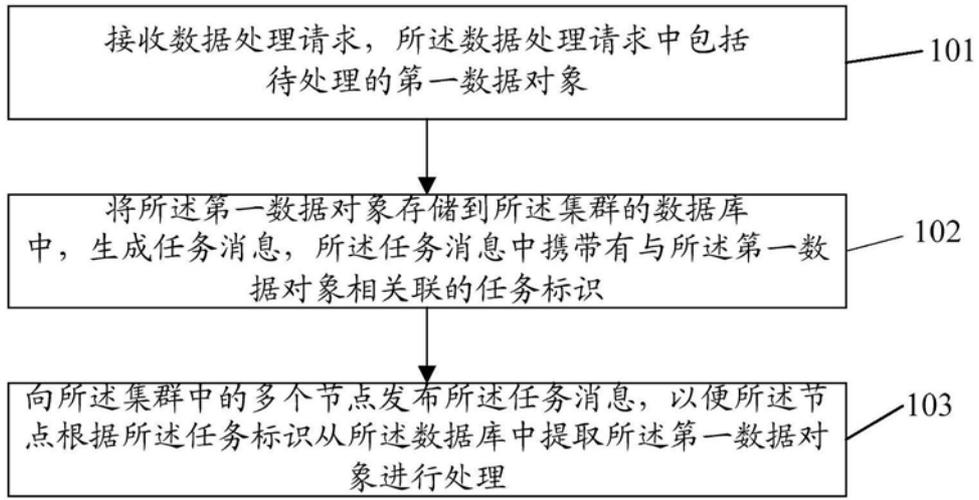


图1

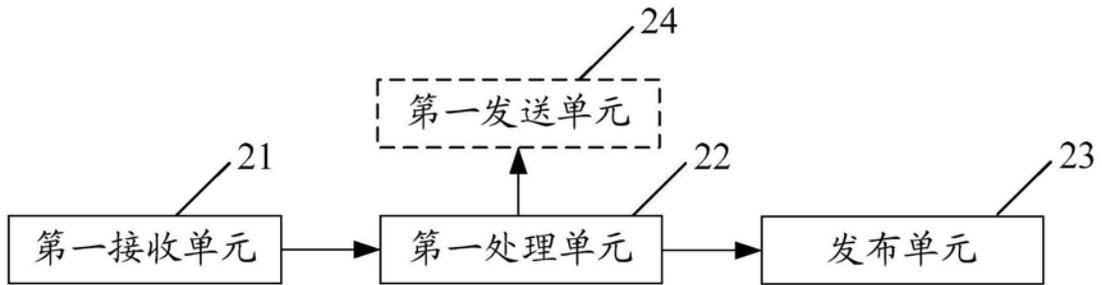


图2

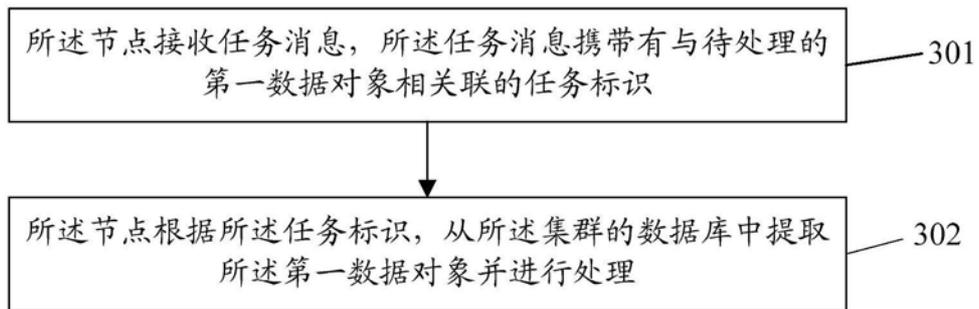


图3

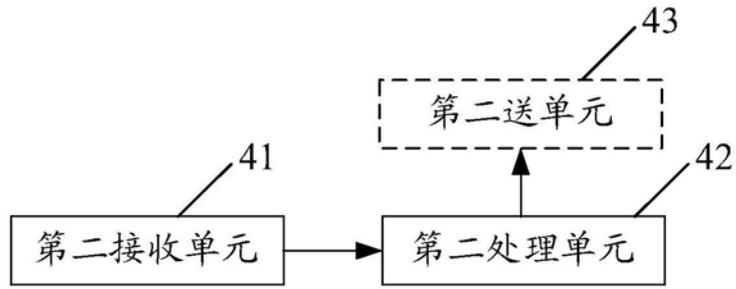


图4

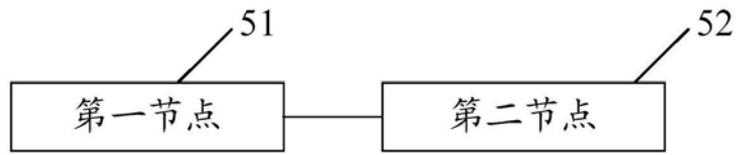


图5

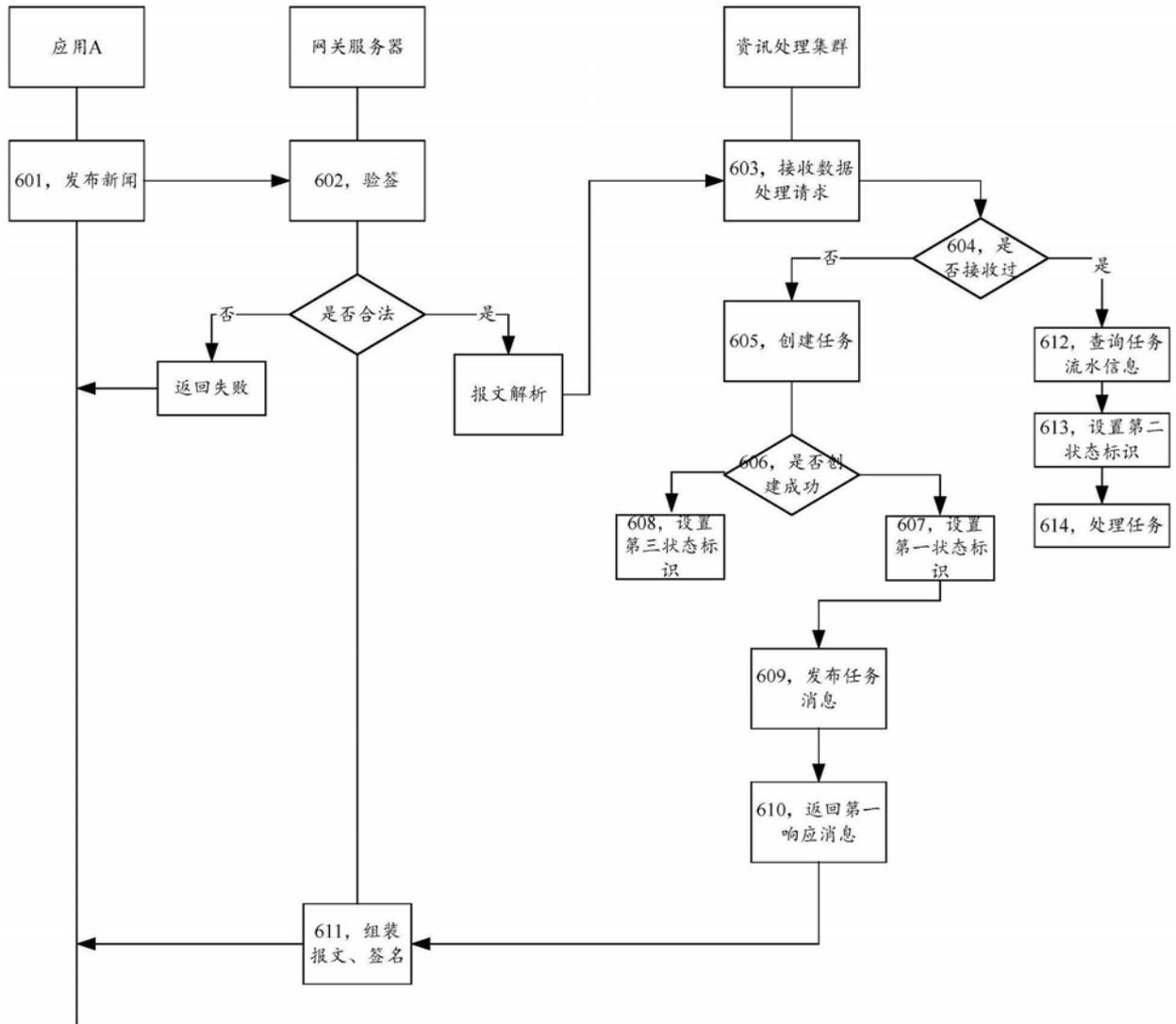


图6

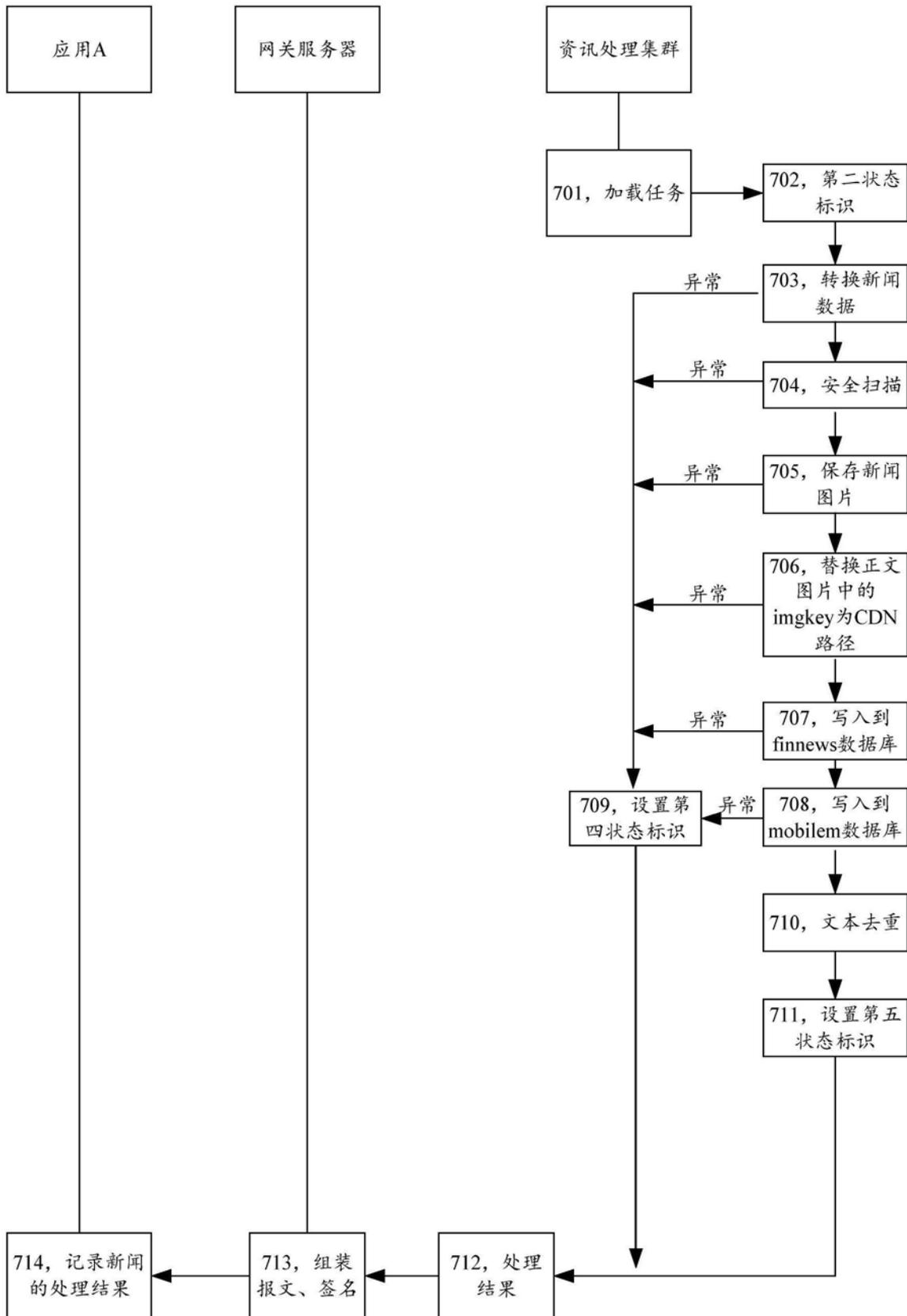


图7

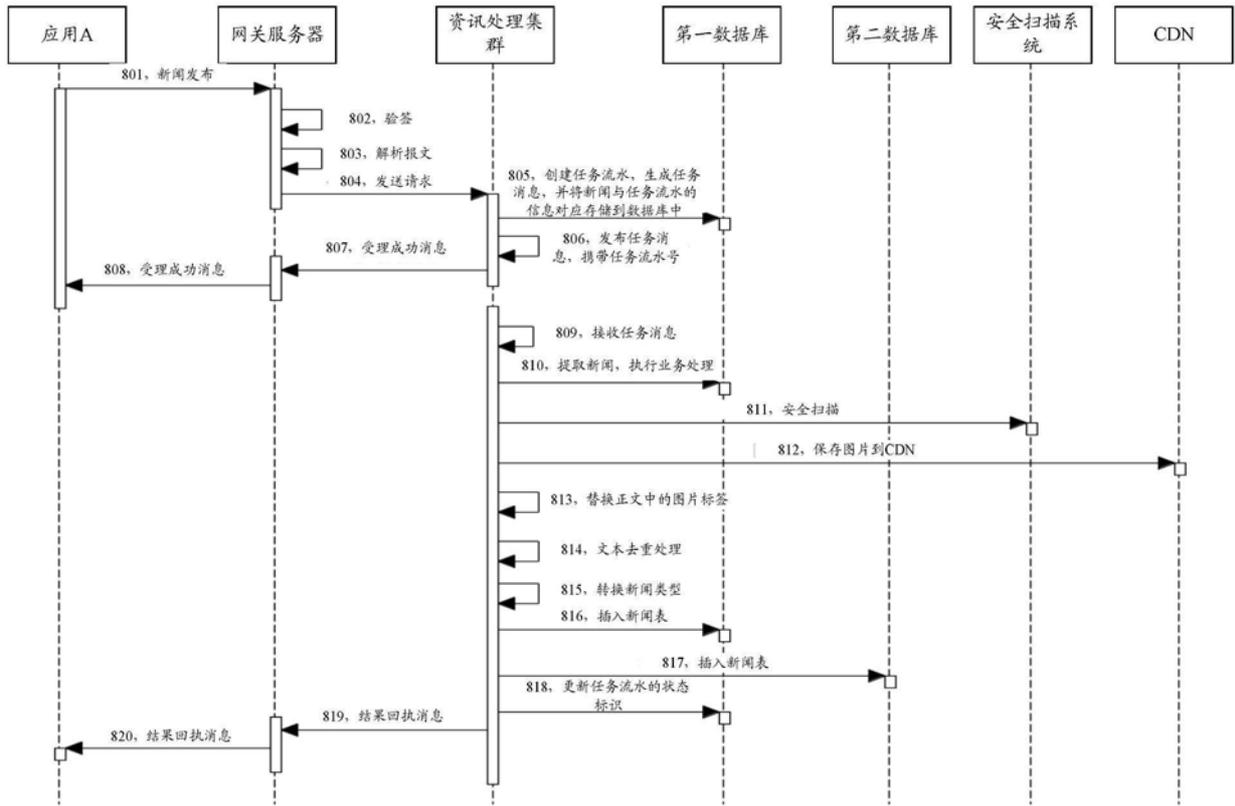


图8