



(12) 发明专利申请

(10) 申请公布号 CN 103916450 A

(43) 申请公布日 2014. 07. 09

(21) 申请号 201310670444. 1

(22) 申请日 2013. 12. 10

(71) 申请人 中国海洋石油总公司

地址 100010 北京市东城区朝阳门北大街
25 号

(72) 发明人 侯晓峰 王小玲 曾泉 应洁
刘海 彭晓枫 桂笛 李文湘
刘朝晖 傅光明

(74) 专利代理机构 北京双收知识产权代理有限
公司 11241

代理人 路远

(51) Int. Cl.

H04L 29/08 (2006. 01)

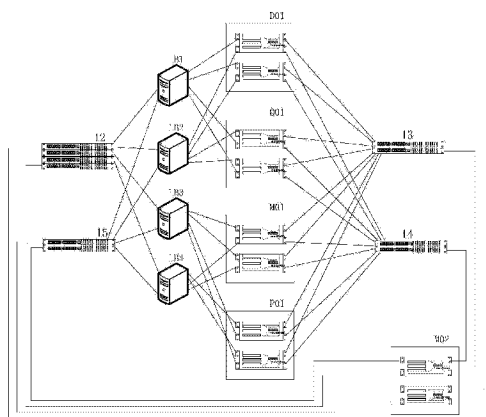
权利要求书2页 说明书5页 附图3页

(54) 发明名称

高性能数据仓库基础架构

(57) 摘要

一种高性能数据仓库基础架构,包括第二核心交换机、虚拟局域网交换机、第一光纤交换机和第三光纤交换机,还包括负载均衡器、开发服务器集群、测试服务器集群、采集与调测服务器集群和生产服务器集群,各集群分别与光纤交换机建立通信链路,分别通过负载均衡器与第二核心交换机和虚拟局域网交换机建立通信链路。形成完整的硬件体系,形成与各专业系统对应的数据仓库开发硬件集群、测试硬件集群、采集调测硬件集群和生产控制硬件集群,各硬件集群基于现有的可扩展技术架构,不仅能更好地支持和满足各二级单位对于自主开发、自主管理、自主运维的需求,规避各二级单位自主开发的数据安全风险,实现安全可靠的开发权限下放,而且具有较强的扩展性。



1. 一种高性能数据仓库基础架构,包括第二核心交换机(12)、虚拟局域网交换机(15)、第一光纤交换机(13)和第二光纤交换机(14),其特征在于:还包括负载均衡器、开发服务器集群(D01)、测试服务器集群(Q01)、采集与调测服务器集群(M01)和生产服务器集群(P01),各服务器集群分别与第一光纤交换机(13)和第二光纤交换机(14)建立通信链路,各服务器集群分别通过负载均衡器与第二核心交换机(12)和虚拟局域网交换机(15)建立通信链路。

2. 根据权利要求1所述的高性能数据仓库基础架构,其特征在于:所述各硬件服务器集群中的每个服务器都包括四个数据端口,每个服务器的第三数据端口连接第一光纤交换机(13),每个服务器的第四数据端口连接第二光纤交换机(14),各负载均衡器的请求侧数据端口分别与第二核心交换机(12)和虚拟局域网交换机(15)建立通信链路。

3. 根据权利要求2所述的高性能数据仓库基础架构,其特征在于:所述开发服务器集群(D01)中服务器的第一数据端口连接第一负载均衡器(LB1)的负载分配侧的相应数据端口,开发服务器集群(D01)中服务器的第二数据端口连接第二负载均衡器(LB2)的负载分配侧的相应数据端口。

4. 根据权利要求3所述的高性能数据仓库基础架构,其特征在于:所述测试服务器集群(Q01)中服务器的第一数据端口连接第二负载均衡器(LB2)的负载分配侧的相应数据端口,测试服务器集群(Q01)中服务器的第二数据端口连接第一负载均衡器(LB1)的负载分配侧的相应数据端口。

5. 根据权利要求4所述的高性能数据仓库基础架构,其特征在于:所述采集与调测服务器集群(M01)中服务器的第一数据端口连接第三负载均衡器(LB3)的负载分配侧的相应数据端口,采集与调测服务器集群(M01)中服务器的第二数据端口连接第四负载均衡器(LB4)的负载分配侧的相应数据端口。

6. 根据权利要求5所述的高性能数据仓库基础架构,其特征在于:所述生产服务器集群(P01)中服务器的第一数据端口连接第四负载均衡器(LB4)的负载分配侧的相应数据端口,生产服务器集群(P01)中服务器的第二数据端口连接第三负载均衡器(LB3)的负载分配侧的相应数据端口。

7. 根据权利要求1至6任一所述的高性能数据仓库基础架构,其特征在于:还包括第二采集与调测服务器集群(M02),其中的服务器都包括四个数据端口,服务器的第一数据端口连接第二核心交换机(12),每个服务器的第二数据端口连接虚拟局域网交换机(15),每个服务器的第三数据端口连接第一光纤交换机(13),每个服务器的第四数据端口连接第二光纤交换机(14)。

8. 根据权利要求1至7任一所述的高性能数据仓库基础架构,其特征在于:所述负载均衡器为服务器的运算存储器空间形成的负载均衡装置(AP),开发服务器集群(D01)中包括一个开发服务器,测试服务器集群(Q01)中包括一个测试服务器,采集与调测服务器集群(M01)中包括一个采集与调测服务器,生产服务器集群(P01)中包括两个生产服务器,负载均衡装置(AP)中设置4个负载请求分配模块,第二核心交换机(12)和虚拟局域网交换机(15)分别与各负载请求分配模块建立数据访问路由,第一负载请求分配模块(MS01)分别与开发服务器和测试服务器的第一数据端口建立数据访问路由,第二负载请求分配模块(MS02)分别与开发服务器和测试服务器的第二数据端口建立数据访问路由,第三负载请求

分配模块(MS03)分别与采集与调测服务器和两个生产服务器的第一数据端口建立数据访问路由,第四负载请求分配模块(MS04)分别与采集与调测服务器和两个生产服务器的第二数据端口建立数据访问路由。

9. 根据权利要求1至7任一所述的高性能数据仓库基础架构,其特征在于:在开发服务器集群(D01)上安装开发环境系统,在测试服务器集群(Q01)上安装测试环境系统,在采集与调测服务器集群(M01)上安装真实环境调试系统,在生产服务器集群(P01)上安装正式生产系统;在生产服务器集群(P01)中还包括请求交互装置(100)、身份认证装置(200)、性能保障装置(300)和数据表现装置(400),

请求交互装置(100),用于向用户提供标准的互联网数据接口,为PRD系统与用户间建立通信链路;

身份认证装置(200),用于判断用户登录信息真伪,对真实用户给与PRD系统中相应权限的数据访问凭证;

性能保障装置(300),用于根据用户的访问凭证将PRD系统对用户请求的响应过程调度至高速内存环境处理;

数据表现装置(400),用于将响应过程处理的数据转换为相关的图形数据,反馈至请求交互装置(100)。

高性能数据仓库基础架构

技术领域

[0001] 本发明涉及一种数据存储系统,特别是涉及一种大数据的存储系统。

背景技术

[0002] 对于一个集团公司来说,都会存在多个不同地域的分公司,分公司的级别可以从二级、三级直到五级,每一家分公司作为独立的机构更愿意将自己的数据放在单独的服务器上,形成一个单独的数据仓库,这种按公司来划分的数据仓库结构,称之为分布式数据仓库系统。由于每一个分布式数据仓库系统都会由几个相关专业系统形成,其数据作用域受软件系统和硬件系统的结构限制,使得数据的融合和利用受到地域限制,主要表现在:

[0003] 各分布式数据仓库难以管控。集团公司的生产数据合并报表难以在这种分布式环境下得到很好的出具。合并报表主要的的数据之一就是组织机构,每台服务器上都存在一份单独的组织机构图。某一时点,我们假设这个组织机构数据在所有分布式数据仓库中都是一样的。过了1个月,美国分公司并购了ABC公司,那么该项修改在美国分公司的数据仓库中得到了更新,但是,集团的组织机构数据中并不存在ABC公司。非洲分公司和远东分公司的数据仓库中也不存在ABC公司。无法保证各分布式数据仓库的主数据一致性。

[0004] 各分布式数据仓库的数据同步存在时间差。分布式数据仓库系统必须选择一个同步时间点,将所有网络连接的多个数据仓库进行同步更新,才能保证所有数据仓库中数据的准确性。这种分布式数据仓库局部的人员原则上只能查看局部数据,为了其他目的而查看全局数据将面临2个问题,一是数据一致性问题,同步的时间与ABC公司更新的时间必定有一个时间差,在这个时间差之间数据很难做到一致。

[0005] 各分布式数据仓库的数据同步过程中的数据安全及完整性传输存在潜在隐患。分布式数据仓库数据传输要素包括传输频率、传输的合法性(跨越作用域),无法确保信息来源于正确的地方,无法设定权限保证只有需要了解ABC公司更新的数据仓库才能接收到。

[0006] 目前形成分布式数据仓库系统的专业系统通常利用SAP系统软件,必然会出现开发、测试、采集和生产等复杂的系统间连接和与具体生产过程的软硬件衔接,这会造成集团层面数据仓库系统无法进行软硬件两方面的有效优化。

发明内容

[0007] 本发明的目的是提供一种高性能数据仓库基础架构,解决分布式数据仓库由于硬件架构分散,造成软件系统受作用域和物理地域限制,数据的一致性和同步性,以及数据安全性无法保证的技术问题。

[0008] 本发明的高性能数据仓库基础架构,第二核心交换机、虚拟局域网交换机、第一光纤交换机和第二光纤交换机,还包括负载均衡器、开发服务器集群、测试服务器集群、采集与调测服务器集群和生产服务器集群,各服务器集群分别与第一光纤交换机和第二光纤交换机建立通信链路,各服务器集群分别通过负载均衡器与第二核心交换机和虚拟局域网交换机建立通信链路。

[0009] 所述各硬件服务器集群中的每个服务器都包括四个数据端口,每个服务器的第三数据端口连接第一光纤交换机,每个服务器的第四数据端口连接第二光纤交换机,各负载均衡器的请求侧数据端口分别与第二核心交换机和虚拟局域网交换机建立通信链路。

[0010] 所述开发服务器集群中服务器的第一数据端口连接第一负载均衡器的负载分配侧的相应数据端口,开发服务器集群中服务器的第二数据端口连接第二负载均衡器的负载分配侧的相应数据端口。

[0011] 所述测试服务器集群中服务器的第一数据端口连接第二负载均衡器的负载分配侧的相应数据端口,测试服务器集群中服务器的第二数据端口连接第一负载均衡器的负载分配侧的相应数据端口。

[0012] 所述采集与调测服务器集群中服务器的第一数据端口连接第三负载均衡器的负载分配侧的相应数据端口,采集与调测服务器集群中服务器的第二数据端口连接第四负载均衡器的负载分配侧的相应数据端口。

[0013] 所述生产服务器集群中服务器的第一数据端口连接第四负载均衡器的负载分配侧的相应数据端口,生产服务器集群中服务器的第二数据端口连接第三负载均衡器的负载分配侧的相应数据端口。

[0014] 还包括第二采集与调测服务器集群,其中的服务器都包括四个数据端口,服务器的第一数据端口连接第二核心交换机,每个服务器的第二数据端口连接虚拟局域网交换机,每个服务器的第三数据端口连接第一光纤交换机,每个服务器的第四数据端口连接第二光纤交换机。

[0015] 所述负载均衡器为服务器的运算存储器空间形成的负载均衡装置,开发服务器集群中包括一个开发服务器,测试服务器集群中包括一个测试服务器,采集与调测服务器集群中包括一个采集与调测服务器,生产服务器集群中包括两个生产服务器,负载均衡装置中设置 4 个负载请求分配模块,第二核心交换机和虚拟局域网交换机分别与各负载请求分配模块建立数据访问路由,第一负载请求分配模块分别与开发服务器和测试服务器的第一数据端口建立数据访问路由,第二负载请求分配模块分别与开发服务器和测试服务器的第二数据端口建立数据访问路由,第三负载请求分配模块分别与采集与调测服务器和两个生产服务器的第一数据端口建立数据访问路由,第四负载请求分配模块分别与采集与调测服务器和两个生产服务器的第二数据端口建立数据访问路由。

[0016] 在开发服务器集群上安装开发环境系统,在测试服务器集群上安装测试环境系统,在采集与调测服务器集群上安装真实环境调试系统,在生产服务器集群上安装正式生产系统;在生产服务器集群中还包括请求交互装置、身份认证装置、性能保障装置和数据表现装置,

[0017] 请求交互装置,用于向用户提供标准的互联网数据接口,为 PRD 系统与用户间建立通信链路;

[0018] 身份认证装置,用于判断用户登录信息真伪,对真实用户给与 PRD 系统中相应权限的数据访问凭证;

[0019] 性能保障装置,用于根据用户的访问凭证将 PRD 系统对用户请求的响应过程调度至高速内存环境处理;

[0020] 数据表现装置,用于将响应过程处理的数据转换为相关的图形数据,反馈至请求

交互装置。

[0021] 本发明的高性能数据仓库基础架构中,将各分布式数据仓库的硬件结构一致化,并形成完整的硬件体系,形成与各专业系统对应的数据仓库开发硬件集群、测试硬件集群、采集调试硬件集群和生产控制硬件集群,各硬件集群基于现有的可扩展技术架构,不仅能更好地支持和满足各二级单位对于自主开发、自主管理、自主运维的需求,规避各二级单位自主开发的数据安全风险,实现安全可靠的开发权限下放,而且具有较强的扩展性。

[0022] 利用数据仓库基础架构,各专业软件系统可以建立并改进企业级数据仓库软件架构,建立清晰的针对不同应用的数据仓库层级,包括数据仓库层,信息交换层,数据集市层,报表分析层,决策支持层等,提供灵活的数据接口,适应二级单位的业务变化。

[0023] 在数据仓库基础架构的基础上,可以利用集中式数据仓库的硬件体系和各集群的高度可扩展性,灵活实现硬件资源的分配和优化,满足数据仓库容量的与日俱增,同时保证复杂专业软件系统的高效运行。

[0024] 下面结合附图对本发明的实施例作进一步说明。

附图说明

[0025] 图 1 为本发明高性能数据仓库基础架构实施例的硬件体系结构示意图;

[0026] 图 2 为本发明高性能数据仓库基础架构实施例的硬件连接示意图;

[0027] 图 3 为利用本发明高性能数据仓库基础架构的数据仓库软件系统的结构示意图;

[0028] 图 4 为本发明高性能数据仓库基础架构的硬件连接另一实施例的结构示意图。

具体实施方式

[0029] 如图 1 所示,本实施例中包括第一核心交换机 11、第二核心交换机 12、虚拟局域网交换机 15、第一光纤交换机 13 和第二光纤交换机 14,还包括日志服务器 19,第一存储服务器 16、第二存储服务器 17 和磁带机 18,实例服务器 51,核心数据库服务器 52,核心数据库服务器 52 的第一组电数据端口与第一核心交换机 11 建立冗余数据链路,核心数据库服务器 52 的第二组电数据端口与第二核心交换机 12 建立冗余数据链路,核心数据库服务器 52 的第三组电数据端口与虚拟局域网交换机 15 的不同网段建立数据链路,实例服务器 51 的第一组电数据端口与第一核心交换机 11 建立冗余数据链路,实例服务器 51 的第二组电数据端口与虚拟局域网交换机 15 的不同网段建立数据链路。核心数据库服务器 52 的第一组光数据端口与第一光纤交换机 13 建立数据链路,核心数据库服务器 52 的第二组光数据端口与第二光纤交换机 14 建立数据链路。第一光纤交换机 13 和第二光纤交换机 14 通过光数据端口分别与日志服务器 19,第一存储服务器 16、第二存储服务器 17 和磁带机 18 建立数据链路。日志服务器 19、磁带机 18、第一存储服务器 16 与光纤交换机在本地部署,第二存储服务器 17 在异地部署。

[0030] 以上硬件架构,通过交换机数据端口在核心数据库服务器 52、实例服务器 51 和存储服务器间建立链路网络,使得核心生产系统获得最可靠的运维环境,使得各专业系统在物理架构上有效分离,任何一个项目系统的改变都不会对核心系统直接产生干扰。

[0031] 如图 2 所示,本实施例中包括开发服务器集群 D01、测试服务器集群 Q01、采集与调测服务器集群 M01 和生产服务器集群 P01,还包括四个负载均衡器,硬件服务器集群中的每

个服务器都包括四个数据端口,每个服务器的第三数据端口连接第一光纤交换机 13,每个服务器的第四数据端口连接第二光纤交换机 14;

[0032] 开发服务器集群 D01 中服务器的第一数据端口连接第一负载均衡器 LB1 的负载分配侧的相应数据端口,开发服务器集群 D01 中服务器的第二数据端口连接第二负载均衡器 LB2 的负载分配侧的相应数据端口;测试服务器集群 Q01 中服务器的第一数据端口连接第二负载均衡器 LB2 的负载分配侧的相应数据端口,测试服务器集群 Q01 中服务器的第二数据端口连接第一负载均衡器 LB1 的负载分配侧的相应数据端口;采集与调测服务器集群 M01 中服务器的第一数据端口连接第三负载均衡器 LB3 的负载分配侧的相应数据端口,采集与调测服务器集群 M01 中服务器的第二数据端口连接第四负载均衡器 LB4 的负载分配侧的相应数据端口;生产服务器集群 P01 中服务器的第一数据端口连接第四负载均衡器 LB4 的负载分配侧的相应数据端口,生产服务器集群 P01 中服务器的第二数据端口连接第三负载均衡器 LB3 的负载分配侧的相应数据端口;

[0033] 各负载均衡器的请求侧数据端口分别与第二核心交换机 12 和虚拟局域网交换机 15 建立通信链路。

[0034] 以上安装在第二核心交换机 12、虚拟局域网交换机 15、第一光纤交换机 13 和第二光纤交换机 14 之间的硬件架构充分保证了数据仓库性能的稳定性和可伸缩性,可以保证集中式数据仓库的可靠性,通过集中管理可以对本实施例中的任一部分进行调整,在动态适应用户请求增长量同时,使得核心软件系统可以与网络门户等逻辑层应用分离,软硬件系统的优化得以深度实现。

[0035] 本实施例还包括第二采集与调测服务器集群 M02,其中的服务器都包括四个数据端口,服务器的第一数据端口连接第二核心交换机 12,每个服务器的第二数据端口连接虚拟局域网交换机 15,每个服务器的第三数据端口连接第一光纤交换机 13,每个服务器的第四数据端口连接第二光纤交换机 14。本集群可以通过第二核心交换机 12 和虚拟局域网交换机 15 仿真数据仓库系统外部环境,直接采集生产服务器集群 P01 的真实运行数据形成调测数据进行系统调试,提高系统数据一致性和完整性的检测。

[0036] 如图 3 所示,具体的 ERP 专业软件系统以 SAP 为例,在开发服务器集群 D01 上安装开发环境(DEV)系统,在测试服务器集群 Q01 上安装测试环境(QAS)系统,在采集与调测服务器集群 M01 上安装真实环境调试(MOCK)系统,在生产服务器集群 P01 上安装正式生产(PRD)系统。还包括请求交互装置 100、身份认证装置 200、性能保障装置 300 和数据表现装置 400 四个附加装置用于完成相应的系统功能。

[0037] 请求交互装置 100,用于向用户提供标准的互联网数据接口,为 PRD 系统与用户间建立通信链路;

[0038] 身份认证装置 200,用于判断用户登录信息真伪,对真实用户给与 PRD 系统中相应权限的数据访问凭证;

[0039] 性能保障装置 300,用于根据用户的访问凭证将 PRD 系统对用户请求的响应过程调度至高速内存环境处理;

[0040] 数据表现装置 400,用于将响应过程处理的数据转换为相关的图形数据,反馈至请求交互装置 100。

[0041] 具体的,请求交互装置 100 可以是富媒体网页服务器,或是综合 SSL 协议的门户网

站。身份认证装置 200 可以是 RADIUS 服务器,或是 LDAP 服务器。性能保障装置 300 可以是刀片服务器集群,或是 DDR5 内存阵列。数据表现装置 400 可以是配置自动化脚本的图形工作站,或是报表服务器。

[0042] 通过在数据仓库形成的数据层之上,利用以上装置为用户建立了接入层并在数据层和接入层之间形成了数据图形化,通过接入层的标准化和认证机制,使得图形化数据可以标准格式为用户在不同终端上引用,消除了对用户请求终端的限制,为用户理解数据提供了多视角的参考,保证了决策的实时性和准确性。

[0043] 如图 4 所示,在以上实施例基础上,在第二核心交换机 12、虚拟局域网交换机 15 与各服务器集群间利用服务器的运算存储器空间形成负载均衡装置 AP,用于对指定的不同服务器的运算存储器进行统一寻址。在负载均衡装置 AP 中加载若干个负载请求分配模块,形成由内存中的负载请求分配模块完成服务器中应用与交换机间的数据连接,实现通过交换机的用户请求可以高速的在内存中由各负载请求分配模块进行分配,充分发挥服务器的效能。

[0044] 在本实施例中,开发服务器集群 D01 中包括一个开发服务器,测试服务器集群 Q01 中包括一个测试服务器,采集与调测服务器集群 M01 中包括一个采集与调测服务器,生产服务器集群 P01 中包括两个生产服务器,负载均衡装置 AP 中设置 4 个负载请求分配模块,第二核心交换机 12 和虚拟局域网交换机 15 分别与各负载请求分配模块建立数据访问路由,第一负载请求分配模块 MS01 分别与开发服务器和测试服务器的第一数据端口建立数据访问路由,第二负载请求分配模块 MS02 分别与开发服务器和测试服务器的第二数据端口建立数据访问路由,第三负载请求分配模块 MS03 分别与采集与调测服务器和两个生产服务器的第一数据端口建立数据访问路由,第四负载请求分配模块 MS04 分别与采集与调测服务器和两个生产服务器的第二数据端口建立数据访问路由。

[0045] 考虑到通信链路的可靠性,每一个硬件服务器的任一数据端口都可以配置成一组,实现同一通信链路的冗余连接。

[0046] 以上所述的实施案例仅仅是对本发明的优选实施方式进行描述,并非对本发明的范围进行限定,在不脱离本发明设计精神的前提下,本领域普通技术人员对本发明的技术方案作出的各种变形和改进,均应落入本发明权利要求书确定的保护范围内。

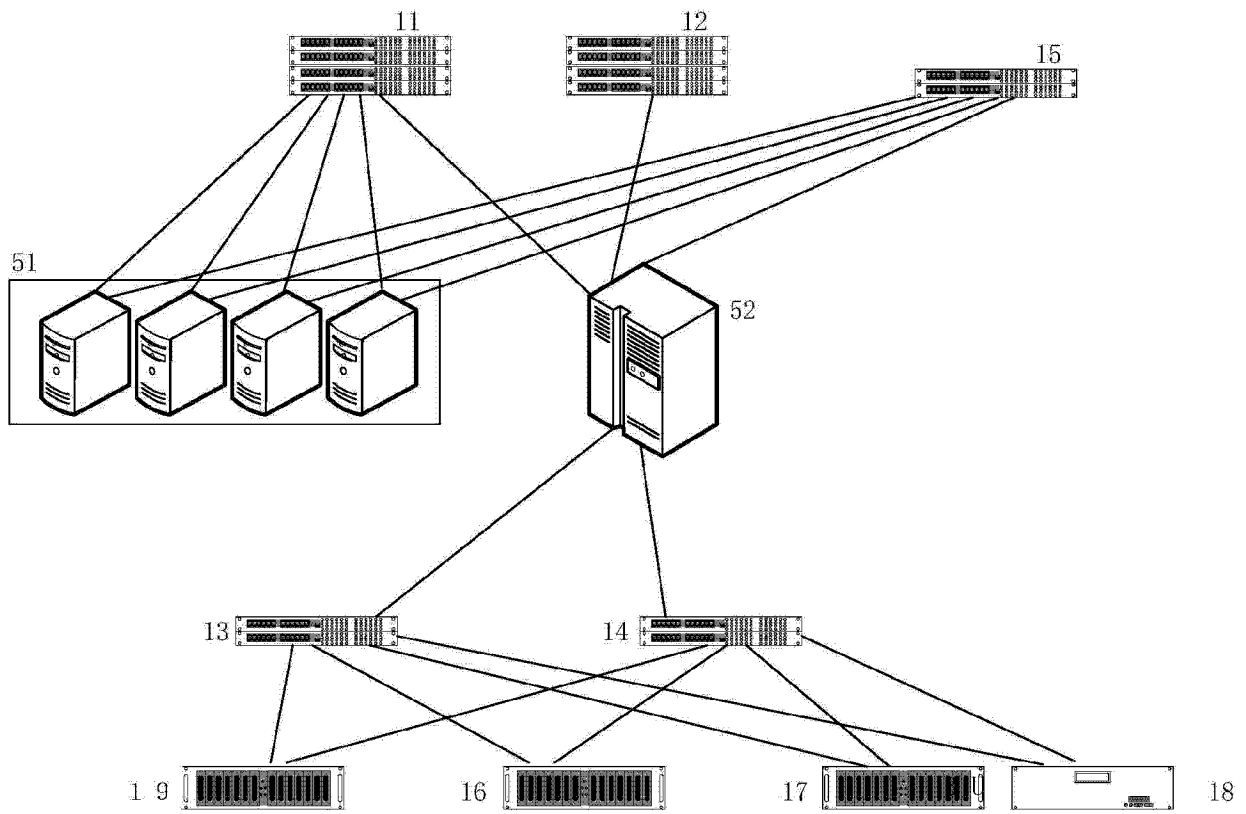


图 1

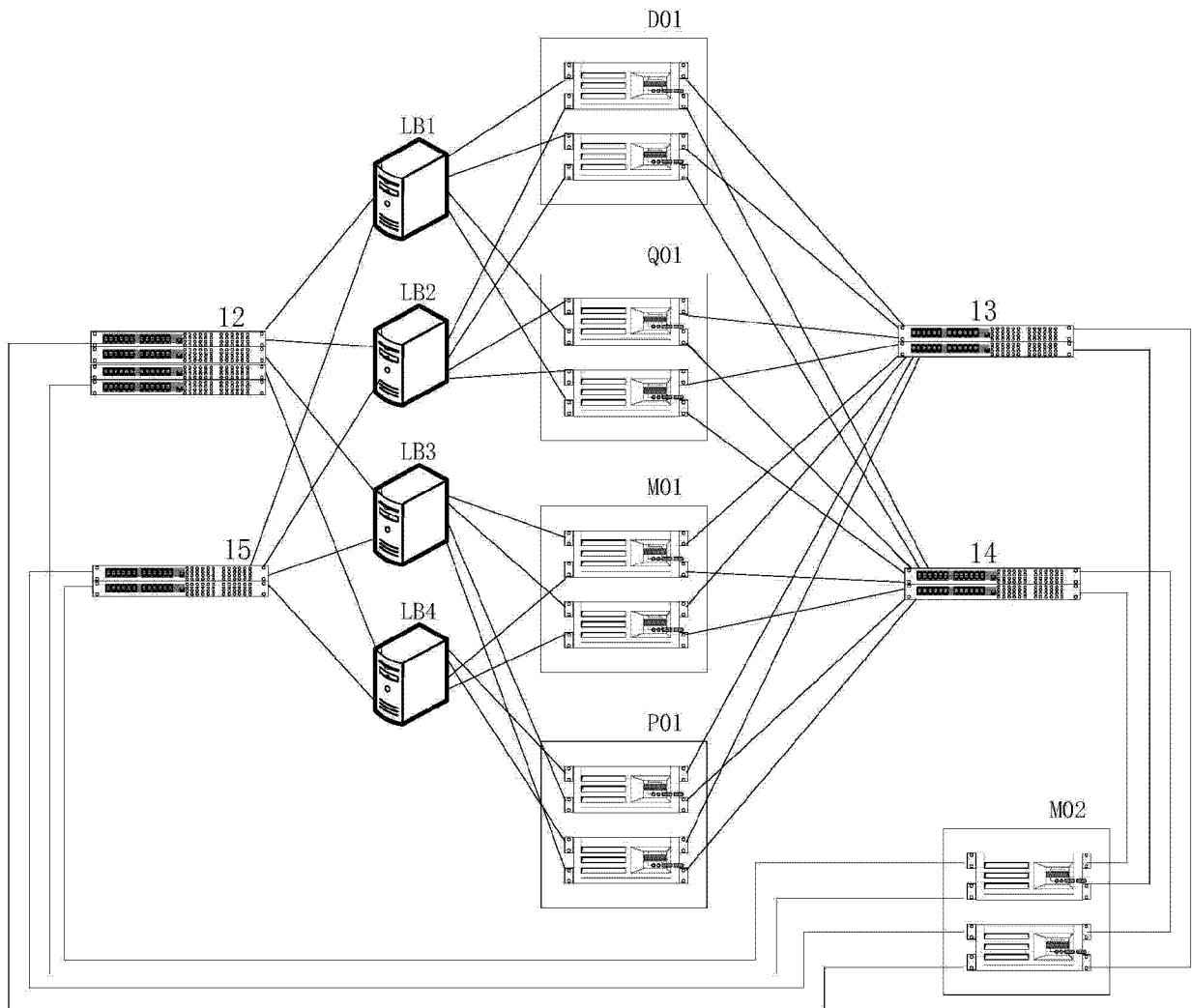


图 2

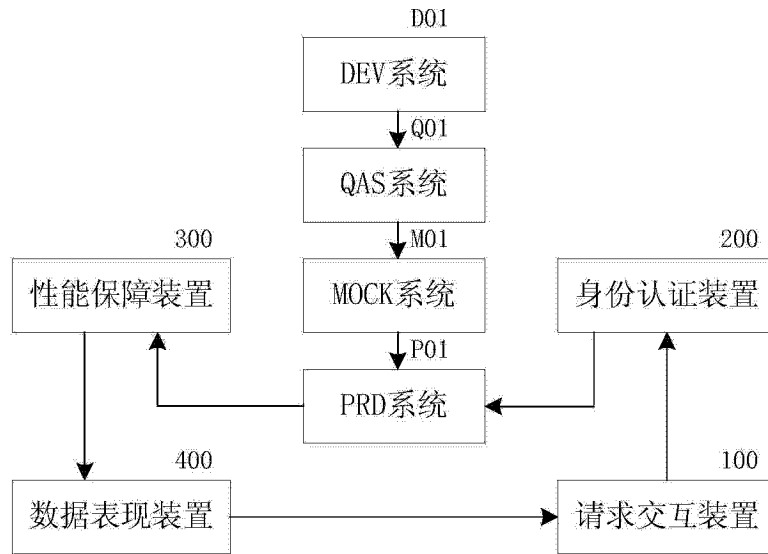


图 3

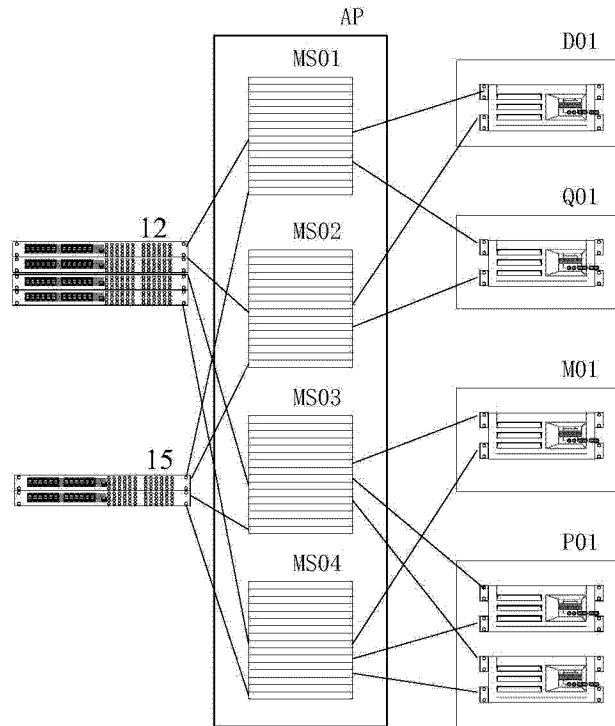


图 4