



(12)发明专利

(10)授权公告号 CN 103605810 B

(45)授权公告日 2017.02.01

(21)申请号 201310666885.4

(51)Int.Cl.

(22)申请日 2013.12.10

G06F 17/30(2006.01)

(65)同一申请的已公布的文献号

(56)对比文件

申请公布号 CN 103605810 A

CN 202886956 U, 2013.04.17,
CN 203733118 U, 2014.07.23,
US 6408292 B1, 2002.06.18,

(43)申请公布日 2014.02.26

审查员 高沛沛

(73)专利权人 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 王小玲 应洁 李文湘 陈旸华
杨凯 谢正华 赵晖 陈凯 任芳
李源

(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

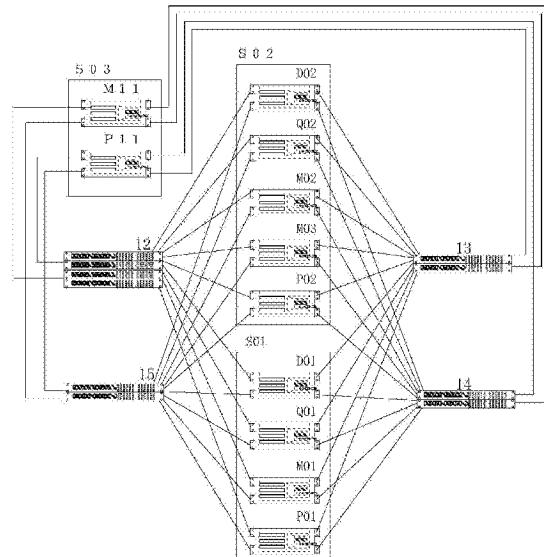
权利要求书2页 说明书5页 附图4页

(54)发明名称

数据仓库的数据处理基础架构

(57)摘要

一种数据仓库的数据处理基础架构，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二光纤交换机，还包括第一并行专业系统硬件服务器群、数据仓库硬件服务器群和第二并行专业系统硬件服务器群，各服务器群分别与第二核心交换机、虚拟局域网交换机、第一光纤交换机和第二光纤交换机建立通信链路。这样软件系统在业务数据建模的时候考虑维度设计时就可以区分并行系统的相同模型下的数据。在进行主数据加载考虑数据流设计时，从数据处理基础架构获得可区分的并行两系统的主数据。



1. 一种数据仓库的数据处理基础架构,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:还包括第一并行专业系统硬件服务器群(S01)、数据仓库硬件服务器群(S02)和第二并行专业系统硬件服务器群(S03),各服务器群分别与第二核心交换机(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)建立通信链路;

所述第一并行专业系统硬件服务器群(S01)包括第一开发服务器(D01)、第一测试服务器(Q01)、第一采集与调试服务器(M01)和第一生产服务器(P01),各服务器分别包括四个数据端口,每个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);

所述数据仓库硬件服务器群(S02)包括第二开发服务器(D02)、第二测试服务器(Q02)、第二采集与调试服务器(M02)、第三采集与调试服务器(M03)和第二生产服务器(P02),各服务器分别包括四个数据端口,每个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);

所述第二并行专业系统硬件服务器群(S03)包括第四采集与调试服务器(M11)和第三生产服务器(P11),各服务器分别包括四个数据端口,每个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);

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;

在第一开发服务器(D01)和第二开发服务器(D02)上分别安装开发环境系统,在第一测试服务器(Q01)和第二测试服务器(Q02)上分别安装测试环境系统,在第一采集与调试服务器(M01)、第二采集与调试服务器(M02)和第四采集与调试服务器(M11)上分别安装真实环境调试系统,在第一生产服务器(P01)、第二生产服务器(P02)和第三生产服务器(P11)上分别安装正式生产系统;还包括配置装置(100)、信号装置(200)和传输装置(300);

配置装置(100),用于接收并行系统中,相同软件专业系统的数据和信号,完成对数据和信号来源的定义,并输出;

信号装置(200),用于接收并行系统的数据和信号,完成相同软件专业系统中相同类型数据的标志和区分,并输出;

传输装置(300),用于将并行系统中各软件专业系统的数据和信号进行匹配,完成数据校验后传输至数据仓库的对应软件专业系统。

2.利用权利要求1所述的数据仓库的数据处理基础架构,进行数据仓库软件专业系统部署时,包括以下步骤:

在数据处理基础架构上的相应服务器上部署开发、测试、采集调试和生产专业软件系统;

在配置装置中设置区分并行专业系统相同对象模型的参考数据;

在信号装置中设置区分并行专业系统相同类型数据的标志信号;

在进行数据仓库各系统的传输流沿开发-测试-采集调试-生产顺序传输前,将采集调试系统复制到另一个采集与调试服务器,并激活;

通过传输装置完成各系统间数据与信号传输,并行系统的两组差异化的采集调试系统和生产系统的数据在数据仓库的相应系统中融合。

数据仓库的数据处理基础架构

技术领域

[0001] 本发明涉及一种数据处理的硬件结构,特别是涉及一种海量数据处理的硬件结构。

背景技术

[0002] 在集团企业中实施ERP中,集团中的有限公司会采用独立的ERP专业系统,例如在集团公司采用完整的SAP的开发、测试、采集调试和生产等专业系统,而在不同的二级企业也会采用完整的开发、测试、采集调试和生产系统,但是集团层面建立数据仓库时对来自不同专业系统的数据处理会出现以下问题:

[0003] 由于数据来源于两套业务系统,集团和有限公司的业务数据,如采购订单、销售订单等等,可能会出现编码范围重复。

[0004] 由于数据来源于两套业务系统,集团和有限公司的主数据,包括人力数据、物料编码等等,可能会出现编码重复。

发明内容

[0005] 本发明的目的是提供一种数据仓库的数据处理基础架构,利用改变数据仓库与并行系统数据处理的基础硬件架构,解决现有并行的业务系统数据无法在数据仓库进行有效数据处理的技术问题。

[0006] 本发明的数据仓库的数据处理基础架构,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二光纤交换机,还包括第一并行专业系统硬件服务器群、数据仓库硬件服务器群和第二并行专业系统硬件服务器群,各服务器群分别与第二核心交换机、虚拟局域网交换机、第一光纤交换机和第二光纤交换机建立通信链路。

[0007] 所述第一并行专业系统硬件服务器群包括第一开发服务器、第一测试服务器、第一采集与调试服务器和第一生产服务器,各服务器分别包括四个数据端口,每个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。

[0008] 所述数据仓库硬件服务器群包括第二开发服务器、第二测试服务器、第二采集与调试服务器、第三采集与调试服务器和第二生产服务器,各服务器分别包括四个数据端口,每个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。

[0009] 所述第二并行专业系统硬件服务器群包括第四采集与调试服务器和第三生产服务器,各服务器分别包括四个数据端口,每个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。

[0010] 还包括日志服务器,第一存储服务器、第二存储服务器和磁带机,实例服务器,核心数据库服务器;核心数据库服务器的第一组电数据端口与第一核心交换机建立冗余数据链路,核心数据库服务器的第二组电数据端口与第二核心交换机建立冗余数据链路,核心数据库服务器的第三组电数据端口与虚拟局域网交换机的不同网段建立数据链路,实例服务器的第一组电数据端口与第一核心交换机建立冗余数据链路,实例服务器的第二组电数据端口与虚拟局域网交换机的不同网段建立数据链路。核心数据库服务器的第一组光数据端口与第一光纤交换机建立数据链路,核心数据库服务器的第二组光数据端口与第二光纤交换机建立数据链路。第一光纤交换机和第二光纤交换机通过光数据端口分别与日志服务器,第一存储服务器、第二存储服务器和磁带机建立数据链路。

[0011] 在第一开发服务器和第二开发服务器上分别安装开发环境系统,在第一测试服务器和第二测试服务器上分别安装测试环境系统,在第一采集与调试服务器、第二采集与调试服务器和第四采集与调试服务器上分别安装真实环境调试系统,在第一生产服务器、第二生产服务器和第三生产服务器上分别安装正式生产系统;还包括配置装置、信号装置和传输装置;

[0012] 配置装置,用于接收并行系统中,相同软件专业系统的数据和信号,完成对数据和信号来源的定义,并输出;

[0013] 信号装置,用于接收并行系统的数据和信号,完成相同软件专业系统中相同类型数据的标志和区分,并输出;

[0014] 传输装置,用于将并行系统中各软件专业系统的数据和信号进行匹配,完成数据校验后传输至数据仓库的对应软件专业系统。

[0015] 利用所述的数据仓库的数据处理基础架构,进行数据仓库软件专业系统部署时,包括以下步骤:

[0016] 在数据处理基础架构上的相应服务器上部署开发、测试、采集调试和生产专业软件系统;

[0017] 在配置装置中设置区分并行专业系统相同对象模型的参考数据;

[0018] 在信号装置中设置区分并行专业系统相同类型数据的标志信号;

[0019] 在进行数据仓库各系统的传输流沿开发-测试-采集调试-生产顺序传输前,将采集调试系统复制到另一个采集与调试服务器,并激活;

[0020] 通过传输装置完成各系统间数据与信号传输,并行系统的两组差异化的采集调试系统和生产系统的数据在数据仓库的相应系统中融合。

[0021] 本发明的数据仓库的数据处理基础架构中,将并行系统中的源数据生成的专业系统的硬件架构归一化,使得在源数据生成系统统一的基础上,区别专业系统不同分支形成的专业系统,并将树形结构的专业系统与数据仓库的硬件架构对接,从硬件架构上保证了软件专业系统中数据可溯源,可以根据硬件架构确定数据的差异性和相关性。这样软件系统在业务数据建模的时候考虑维度设计时就可以区分并行系统的相同模型下的数据。在进行主数据加载考虑数据流设计时,从数据处理基础架构获得可区分的并行两系统的主数据。

[0022] 下面结合附图对本发明的实施例作进一步说明。

附图说明

- [0023] 图1为本发明数据仓库的数据处理基础架构的硬件体系结构示意图；
- [0024] 图2为本发明数据仓库的数据处理基础架构的并行系统连接结构示意图；
- [0025] 图3为利用本发明数据仓库的数据处理基础架构的软件系统的结构示意图；
- [0026] 图4为利用本发明数据仓库的数据处理基础架构进行软件系统优化时的方法流程图。

具体实施方式

[0027] 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。

[0028] 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。

[0029] 如图2所示，本实施例中包括第一并行专业系统硬件服务器群S01、数据仓库硬件服务器群S02和第二并行专业系统硬件服务器群S03，第一并行专业系统硬件服务器群S01包括第一开发服务器D01、第一测试服务器Q01、第一采集与调试服务器M01和第一生产服务器P01，各服务器分别包括四个数据端口，每个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。

[0030] 数据仓库硬件服务器群S02包括第二开发服务器D02、第二测试服务器Q02、第二采集与调试服务器M02、第三采集与调试服务器M03和第二生产服务器P02，各服务器分别包括四个数据端口，每个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。

[0031] 第二并行专业系统硬件服务器群S03包括第四采集与调试服务器M11和第三生产服务器P11，各服务器分别包括四个数据端口，每个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。

[0032] 通过第二核心交换机12和虚拟局域网交换机15将数据仓库的硬件架构与配置完整的第一并行专业系统硬件架构完善连接,使得数据和传输可以根据需要在相应的服务器间传输,并且将第一并行专业系统硬件架构与作为源数据系统子集的第二并行专业系统硬件架构完善连接,为源数据的复制和统一提供了完整的硬件结构。同时利用网络网段分割,保证了各专业系统地独立性,避免了具体生产过程或采集调试过程的系统间干扰。

[0033] 第一光纤交换机13第二光纤交换机14保证了各系统中服务器与存储器间传输链路的可靠性和独立性,最大限度保障了各服务器的动态数据吞吐量。

[0034] 如图3所示,具体的ERP专业软件系统以SAP为例,在第一开发服务器D01和第二开发服务器D02上分别安装开发环境(DEV)系统,在第一测试服务器Q01和第二测试服务器Q02上分别安装测试环境(QAS)系统,在第一采集与调试服务器M01、第二采集与调试服务器M02和第四采集与调试服务器M11上分别安装真实环境调试(MOCK)系统,在第一生产服务器P01、第二生产服务器P02和第三生产服务器P11上分别安装正式生产(PRD)系统。在软件系统中还包括配置装置100、信号装置200和传输装置300。

[0035] 配置装置100,用于接收并行系统中,相同软件专业系统的数据和信号,完成对数据和信号来源的定义,并输出;

[0036] 信号装置200,用于接收并行系统的数据和信号,完成相同软件专业系统中相同类型数据的标志和区分,并输出;

[0037] 传输装置300,用于将并行系统中各软件专业系统的数据和信号进行匹配,完成数据校验后传输至数据仓库的对应软件专业系统。

[0038] 通过以上装置,配合安装在相应服务器上的SAP系统,就可以将并行系统中各软件专业系统中的数据进行向数据仓库中倒换和融合,并且在倒换和融合过程中,完成对并行系统中相同数据源系统数据和数据类型的辨识和标记,并对应传输到数据仓库的相应专业系统中,完成根据数据仓库业务逻辑进行的数据配置和标志,使得真实完整的并行业务数据可以在数据仓库的正式生产系统充分融合,实现数据仓库各特定业务。

[0039] 如图4所示,在本实施例数据仓库的数据处理基础架构上,进行数据仓库软件专业系统部署时的主要步骤包括:

[0040] 在数据处理基础架构上的相应服务器上部署开发、测试、采集调试和生产专业软件系统;

[0041] 在配置装置中设置区分并行专业系统相同对象模型的参考数据;

[0042] 在信号装置中设置区分并行专业系统相同类型数据的标志信号;

[0043] 通过传输装置完成各系统间数据和信号传输,并行系统的两组差异化的采集调试系统和生产系统的数据在数据仓库的相应系统中融合。

[0044] 完成各系统间传输流,两个差异化的采集调试系统对生产系统的真实运行环境进行数据模型、类型等关键数据进行对比测试。

[0045] 在进行数据仓库各系统的传输流沿开发-测试-采集调试-生产顺序传输前,将采集调试系统复制到另一个采集与调试服务器,并激活;

[0046] 通过以上方法步骤,当数据来源于两套并行业务系统时可以避免出现编码范围重复,主数据无法区分,对原系统地映射关系无法区分等问题,人力数据、物料编码、采购订单、销售订单等相同类型对象的数据都可以全部区分。同时满足

[0047] 进行模型维度调整,或者主数据流调整时,调整前后的对比测试可行,保证调整的模型及业务分析报表不受调整误差影响。

[0048] 具体地,再专业软件系统中的模拟实现包括通过模型维度设计,建立单独维度作为区分集团和有限公司的特定维度。在模型中,单独设计源系统标识的维度,维度中包含相关信息对象,可以对集团及有限公司的数据进行区分。在数据流中,通过编写代码的方式,实现集团和有限公司的数据划分。将来源于并行系统的数据,通过编码的方式,加以标识“B”,此种方法可以实现数据加载过程中的两部分数据划分。结合数据处理基础架构的硬件架构设计,利用备份系统,进行调整前后的对比测试。

[0049] 在调整后的数据模型、数据流等内容传输至MOCK系统前,将MOCK系统备份至新的服务器上,完成备份后,再进行内容传输。之后,采取两系统比对测试的方法,将MOCK备份系统的报表、数据模型与MOCK系统的报表、数据模型进行比对,从而保证调整前后的各并行系统报表,均不受调整影响。

[0050] 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。

[0051] 以上所述的实施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。

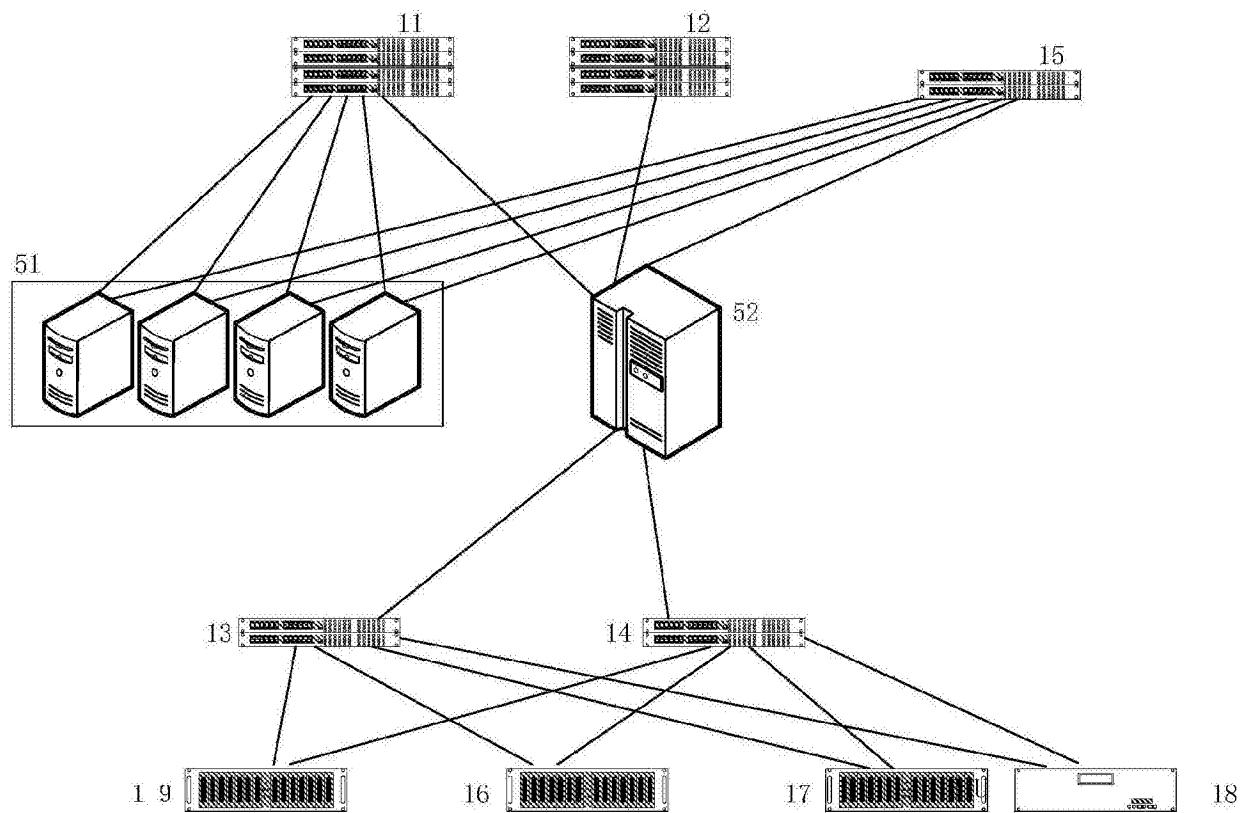


图1

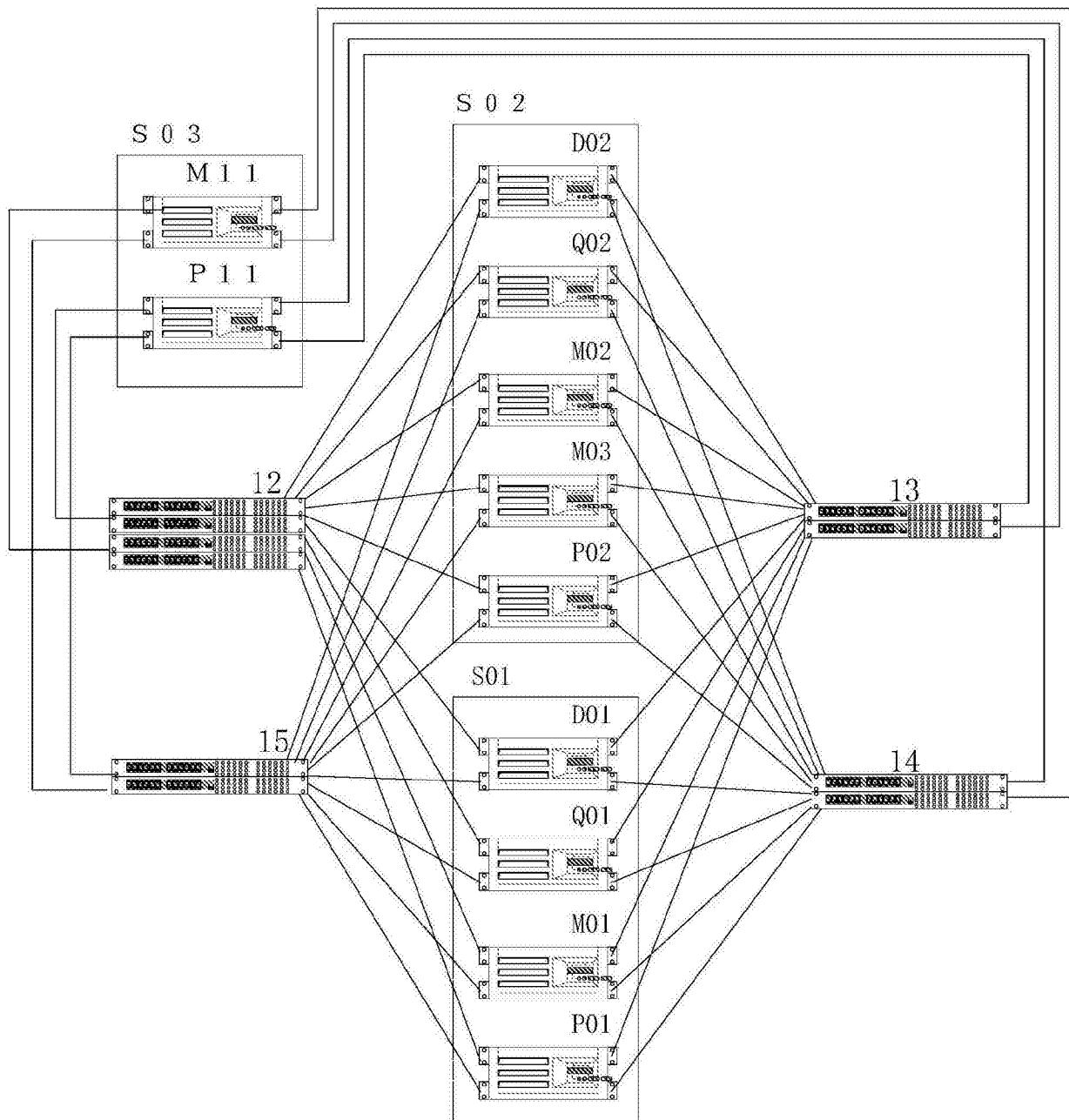


图2

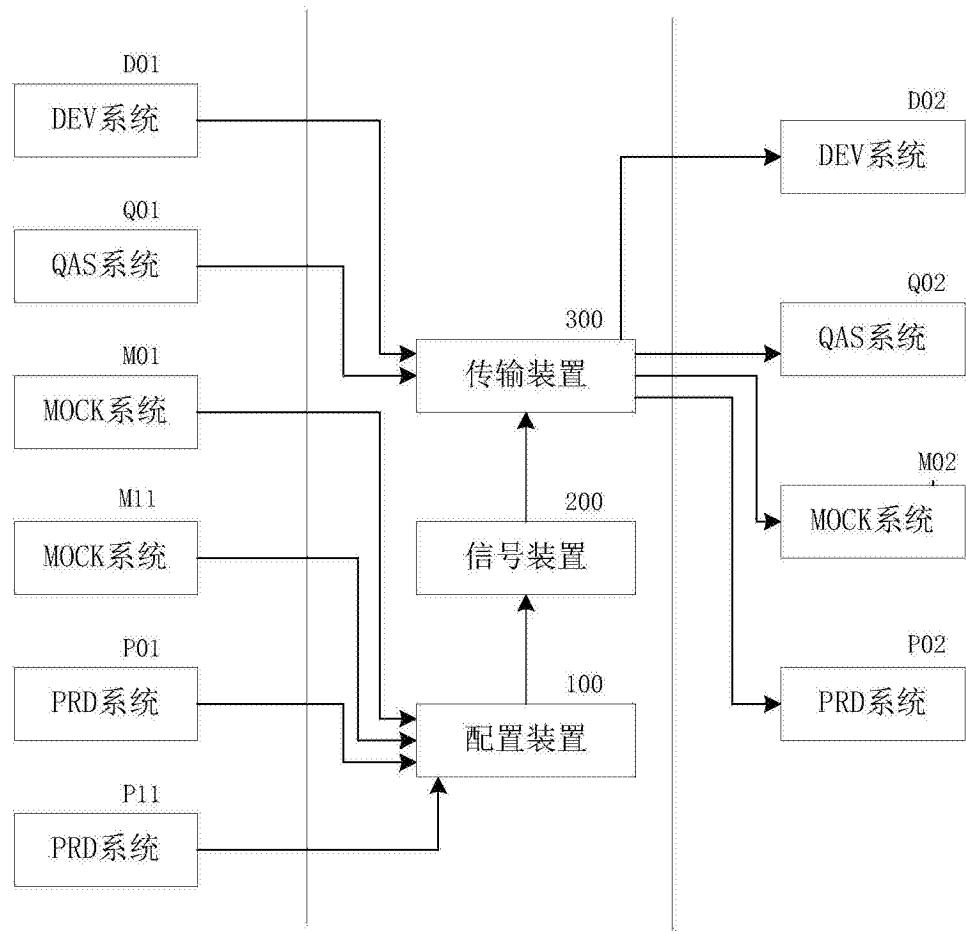


图3

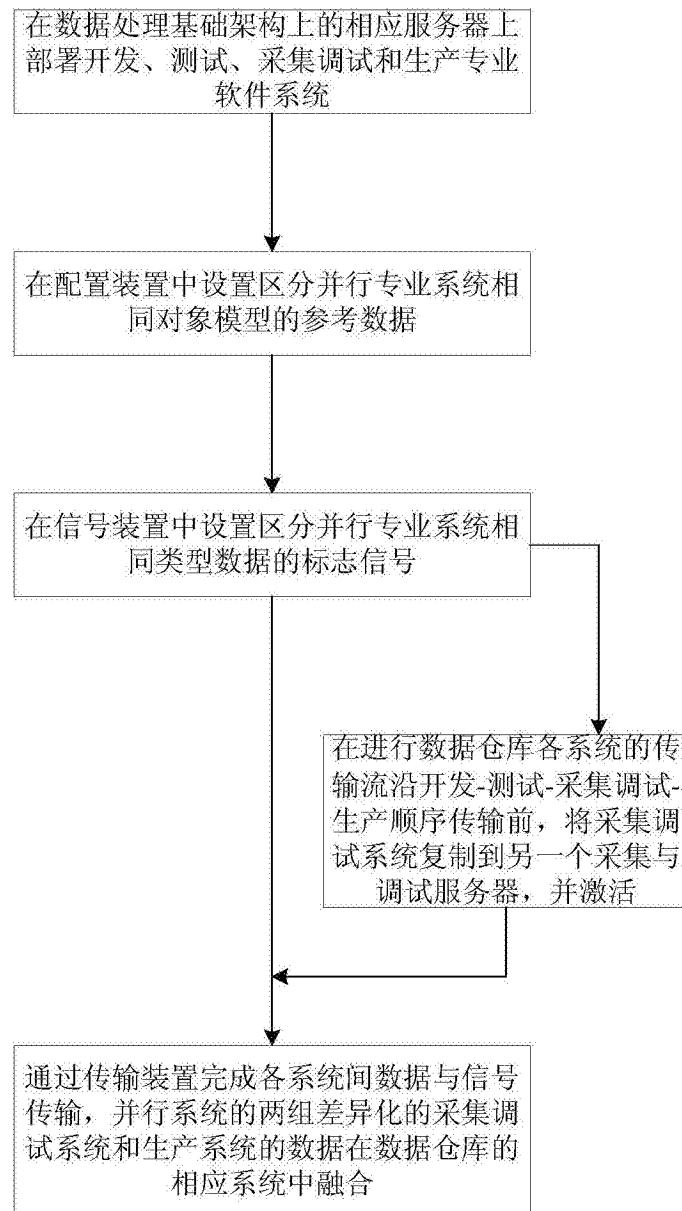


图4