

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
H04L 12/56 (2006.01)



## [12] 发明专利申请公布说明书

[21] 申请号 200710042895.5

[43] 公开日 2007 年 12 月 26 日

[11] 公开号 CN 101094173A

[22] 申请日 2007.6.28

[21] 申请号 200710042895.5

[71] 申请人 上海交通大学

地址 200240 上海市闵行区东川路 800 号

[72] 发明人 胡志萍 何援军 曹 健 陆朝俊  
徐志麟

[74] 专利代理机构 上海交达专利事务所

代理人 王锡麟 王桂忠

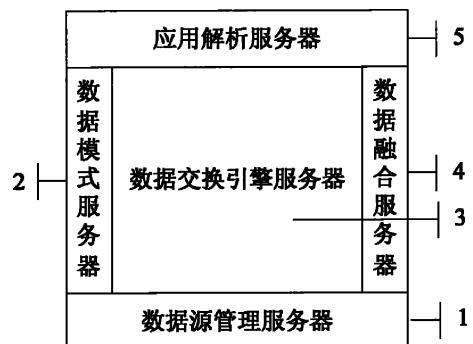
权利要求书 3 页 说明书 7 页 附图 3 页

[54] 发明名称

分布式异构环境下的数据交换集成系统

[57] 摘要

一种电信技术领域的分布式异构环境下的数据交换集成系统，包括：数据源管理服务器、数据模式服务器、数据交换引擎服务器、数据融合服务器和应用解析服务器。数据源管理服务器负责与外部异构数据源建立联系；数据模式服务器管理异构数据源数据的局部模式和系统的全局模式，以及之间的映射规则；数据交换引擎服务器对输入数据进行数据局部模式和全局模式之间的映射；数据融合服务器对存有数据冲突的数据进行融合；应用解析服务器将外部应用系统的任务需求进行解析，并对返回的数据进行集成拼接，同时接收外部应用系统对任务完成的评价，数据融合服务器依此修改融合规则和数据源的可信度。本发明模块清晰，具有动态性、通用性及实时性。



---

1、一种分布式异构环境下的数据交换集成系统，包括：数据源管理服务器、数据模式服务器、数据交换引擎服务器、数据融合服务器和应用解析服务器，其特征在于：

所述数据源管理服务器负责与有数据交换集成需求的外部异构数据源建立联系，接收外部异构数据源的注册信息，并将该数据源数据的局部模式传送到数据模式服务器，同时接收外部异构数据源输出的待集成数据，并将这些数据传送到数据交换引擎服务器；

所述数据模式服务器连接到数据源管理服务器上，该数据模式服务器管理由数据源管理服务器提供的各异构数据源的局部模式和系统的全局模式，以及局部模式和全局模式之间的映射规则；

所述数据交换引擎服务器连接到数据源管理服务器上，该数据交换引擎服务器通过数据模式服务器中的映射规则实现了数据的全局模式和局部模式之间的映射，并调用相应的数据源操作；

所述数据融合服务器连接到数据交换引擎服务器上，该数据融合服务器对数据交换引擎服务器传递的不同数据源提供的存有数据冲突的数据根据融合规则和数据源可信度进行融合，将融合后数据传递给应用解析服务器；

所述应用解析服务器连接到数据交换引擎服务器上，该应用解析服务器将外部应用系统对多个异构数据源任务需求解析为多个对单一数据源的任务需求，并对数据融合服务器返回的采集数据进行集成拼接，将结果返回给外部应用系统；同时，应用解析服务器接收外部应用系统对任务完成的评价，数据融合服务器依此修改融合规则和数据源的可信度。

2、根据权利要求1所述的分布式异构环境下的数据交换集成系统，其特征是，所述的数据源管理服务器包括：系统连接单元、数据管理单元和数据源存储单元，其中系统连接单元负责与外部异构数据源进行连接，外部异构数据源通过系统连接单元提供的接口与整个系统进行连接，从而实现整个系统在各种信息系统中的即插即用；数据管理单元负责外部异构数据源的注册，从系统连接单元已

---

连接的外部异构数据源获得数据源的地址、局部数据模式、数据源连接属性信息，以及向数据模式服务器提供描述本数据源数据的局部模式；数据源存储单元负责存储数据管理单元中外部异构数据源的注册信息。

3、根据权利要求 1 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的数据模式服务器包括：数据模式引擎单元、模式存储单元和映射规则存储单元，其中数据模式服务器负责收集各外部异构数据源提交的局部模式，并按照提供者和类型进行存储，通过映射工具将各局部数据模式合成为全局模式，数据模式引擎单元负责将输入的局部模式合成到全局模式中，以及局部模式和全局模式之间的映射，模式和映射规则的增加、删除、修改、查找操作，模式存储单元存储系统中的所有局部模式和全局模式，映射规则存储单元存储系统中的所有局部模式和全局模式之间的映射。

4、根据权利要求 1 或 3 所述的分布式异构环境下的数据交换集成系统，其特征是，所述局部模式由每个外部异构数据源各自提供，描述本数据源所提供的数据内容；所述全局模式为系统对所有局部模式进行合成所生成的，面向外部应用系统描述当前整个系统所提供的数据内容，使整个数据交换集成系统对外部应用系统提供数据的统一表示格式，屏蔽了外部异构数据源的异构性。

5、根据权利要求 1 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的数据融合服务器包括：融合引擎单元、规则管理单元和可信度管理单元，融合引擎单元根据输入待融合数据以及规则库的融合规则和相应数据源的可信度信息对数据进行融合操作，并将结果传送到应用解析服务器；规则管理单元完成规则存储和规则管理，其中规则存储是指存储待融合数据的融合规则，规则管理是指响应系统对规则库进行创建、维护操作，并按照外部应用系统返回的要求对规则库中的规则进行增加、删除、修改、查找操作；可信度管理单元完成可信度存储和可信度管理，其中可信度存储是指存储各异构数据源的可信度数据，可信度管理是指响应系统对数据源可信度数据进行创建、维护操作，并按照外部应用系统返回的要求对数据源可信度数据进行增加、删除、修改、查找操作。

6、根据权利要求 5 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的数据融合服务器接收到数据交换引擎服务器传递过来的有冲突的集成

---

数据，判断是否存在数据冲突，依据是规则管理单元提供的融合规则以及可信度管理单元提供的对应数据源可信度数据进行融合。

7、根据权利要求 5 或 6 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的可信度管理单元还根据本次融合结果和本次融合过程中各数据源的服务质量对数据源可信度进行修改，实现可信度的动态化。

8、根据权利要求 1 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的数据交换引擎服务器，是一个交换引擎单元，该交换引擎单元根据数据模式服务器中映射规则对输入数据进行数据局部模式和全局模式之间的映射，对数据源管理服务器输入的采集数据实现由局部模式到全局模式的映射，对应用解析服务器输入的任务参数实现由全局模式到局部模式的映射；

9、根据权利要求 1 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的应用解析服务器，实现了对多源异构数据源的统一访问，应用解析服务器将外部应用系统对多源异构数据源任务需求解析为多个对单一数据源的任务需求，并实现最终数据集成任务结果的拼接，将结果返回给外部应用系统，同时应用解析服务器接收外部应用系统对任务完成的评价，数据融合服务器依此修改融合规则和数据源的可信度。

10、根据权利要求 1 所述的分布式异构环境下的数据交换集成系统，其特征是，所述的数据源管理服务器、数据模式服务器、数据交换引擎服务器、数据融合服务器和应用解析服务器位于同一台计算机中或分布在同一个网络环境中的多台计算机中。

## 分布式异构环境下的数据交换集成系统

### 技术领域

本发明涉及的是一种电信技术领域的系统，具体地说是一种分布式异构环境下的数据交换集成系统。

### 背景技术

信息化时代和全球经济竞争的环境下，随着信息化的不断发展，各行业、单位或者机构、部门内部逐渐实现了业务信息化。但由于业务和功能归属不同，所采用的计算机环境由不同平台组成，形成了大量分布式异构数据源，这些孤立的异构数据源也被称为“信息孤岛”。对多个异构数据源提供统一的表示、存储、查询和管理的工作称为信息集成。当前，不同人不同组织开发的针对不同领域的数据交换集成系统重复开发，系统灵活性差，增加了系统开发和维护的开销。现在缺乏各信息领域都能通用的灵活的，可扩展的数据交换集成系统。因此，为了从根本上解决数据交换集成系统的灵活性、可扩展性和重用性，研发一个开放的，可以适应信息集成技术和需求不断发展的数据交换集成系统，是数据交换集成技术在信息领域的一个重要应用趋势。

经对现有技术文献的检索发现，中国专利申请号 0410057020.9，发明名称：数据交换方法，公开号：CN1738309，该专利的数据交换系统中，系统通过定义可扩展标记语言的文本格式，对交换数据进行封装，实现数据交换。但不能解决数据交换中出现的语义异构和数据冲突。该数据交换系统：(1) 添加新的异构数据源时，需要重新定义可扩展标记语言，可扩展性差，难于实现该方法的即插即用。(2) 需根据不同的应用领域定义可扩展标记语言文本格式，限制了该方法的应用领域，影响了数据交换集成系统的通用性。(3) 由于不同数据源对同一数据项提供的数据可能会产生冲突，信息集成需要消除数据冲突，而该系统没有解决数据交换集成中出现的数据冲突问题。

### 发明内容

本发明针对现有技术的不足与缺陷，提供一种分布式异构环境下的数据交换集成系统，使其能够在分布式异构环境下为信息集成领域提供一种灵活、通用的、即插即用的数据交换集成系统，解决当前信息集成领域需要对数据交换集成功能进行重复开发，以及现有数据交换集成系统不能满足信息领域需求，以及不灵活、可扩展性差的问题，推动数据交换集成的应用和发展。

本发明是通过以下技术方案实现的，本发明包括：数据源管理服务器、数据模式服务器、数据交换引擎服务器、数据融合服务器和应用解析服务器。

所述数据源管理服务器负责与有数据交换集成需求的外部异构数据源建立联系，接收外部异构数据源的注册信息，并将该数据源数据的局部模式传送到数据模式服务器，同时接收外部异构数据源输出的待集成数据，并将这些数据传送到数据交换引擎服务器；

所述数据模式服务器连接到数据源管理服务器上，该数据模式服务器管理由数据源管理服务器提供的各异构数据源的局部模式和系统的全局模式，以及局部模式和全局模式之间的映射规则；

所述数据交换引擎服务器连接到数据源管理服务器上，该数据交换引擎服务器通过数据模式服务器中的映射规则实现了数据的全局模式和局部模式之间的映射，并调用相应的数据源操作；

所述数据融合服务器连接到数据交换引擎服务器上，该数据融合服务器对数据交换引擎服务器传递的不同数据源提供的存有数据冲突的数据根据融合规则和数据源可信度进行融合，将融合后数据传递给应用解析服务器；

所述应用解析服务器连接到数据交换引擎服务器上，该应用解析服务器将外部应用系统对多个异构数据源任务需求解析为多个对单一数据源的任务需求，并对数据融合服务器返回的采集数据进行集成拼接，将结果返回给外部应用系统。同时，应用解析服务器接收外部应用系统对任务完成的评价，数据融合服务器依此修改融合规则和数据源的可信度。

外部异构数据源和外部应用系统是与本发明系统交互的两个外部系统，本发明的任务需求来自外部应用系统，集成的数据来自外部异构数据源，集成结果返回给外部应用系统。

所述的数据源管理服务器包括：系统连接单元、数据管理单元和数据源存储单元，该数据源管理服务器主要负责与外部异构数据源进行连接和数据交互，并负责外部异构数据源的注册，提供对外部异构数据源的相关描述信息。其中系统连接单元根据数据源存储单元的注册信息与外部异构数据源进行连接，外部异构数据源可以通过系统连接单元提供的接口与本发明系统进行连接，从而实现本发明系统在各种信息系统中的即插即用；数据管理单元主要负责外部异构数据源的注册，从系统连接单元已连接的外部异构数据源获得数据源的地址、局部数据模式、数据源连接属性等相关信息，以及向数据模式服务器提供描述本数据源数据的局部模式；数据管理单元将数据源注册信息传递给数据源存储单元，数据源存储单元负责存储数据管理单元中外部异构数据源的注册信息。

所述的数据模式服务器包括：数据模式引擎单元、模式存储单元和映射规则存储单元。整个系统中数据模式包括全局模式和局部模式：局部模式由每个外部异构数据源各自提供，描述本数据源所提供的数据内容；全局模式为系统对所有局部模式进行合成所生成的，面向外部应用系统描述当前整个系统所提供的数据内容，使整个数据交换集成系统对外部应用系统提供数据的统一表示格式，屏蔽了外部异构数据源的异构性。数据模式服务器负责收集各外部异构数据源提交的局部模式，并按照提供者和类型进行存储，通过映射工具将各局部数据模式合成为全局模式。数据模式引擎单元负责将输入的局部模式合成到全局模式中，以及局部模式和全局模式之间的映射，模式和映射规则的增加、删除、修改、查找操作；模式存储单元存储系统中的所有局部模式和全局模式；映射规则存储单元存储系统中的所有局部模式和全局模式之间的映射。

所述的数据融合服务器包括：融合引擎单元、规则管理单元和可信度管理单元，融合引擎单元根据输入待融合数据以及规则库的融合规则和相应数据源的可信度信息对数据进行融合操作，并将结果传送到应用解析服务器。规则管理单元完成规则存储和规则管理，其中规则存储是指存储待融合数据的融合规则，规则管理是指响应系统对规则库进行创建、维护等操作，并可以按照外部应用系统返回的要求对规则库中的规则进行增加、删除、修改、查找操作。可信度管理单元完成可信度存储和可信度管理，其中可信度存储是指负责存储各异构数据源的

---

可信度数据，可信度管理是指负责响应系统对数据源可信度数据进行创建、维护等操作，并可以按照外部应用系统返回的要求对数据源可信度数据进行增加、删除、修改、查找操作。可信度管理单元还可以根据本次融合结果和本次融合过程中各数据源的服务质量对数据源可信度进行修改，实现可信度的动态化。

所述的数据交换引擎服务器，是一个交换引擎单元，该交换引擎单元根据数据模式服务器中映射规则对输入数据进行数据局部模式和全局模式之间的映射，对数据源管理服务器输入的采集数据实现由局部模式到全局模式的映射，对应用解析服务器输入的任务参数实现由全局模式到局部模式的映射。

所述的应用解析服务器包括：解析单元、集成单元和反馈单元，实现了对多源异构数据源的统一访问。解析单元将外部应用系统对多源异构数据源任务需求解析为多个对单一数据源的任务需求，集成单元实现最终数据集成任务结果的集成拼接，将结果返回给外部应用系统。同时反馈单元接收外部应用系统对任务完成的评价，数据融合服务器依此修改融合规则和数据源的可信度。

本发明能够准确有效的对各类异构数据源进行动态交换集成。本发明是一种在互联网、广域网和局域网环境中，具有开放性和低成本的数据交换集成应用解析、数据模式管理、数据采集、数据源管理、数据融合、数据交换集成能力的跨网络、跨平台、跨数据库的分布式异构环境下的数据交换集成系统。本发明具有根据他、融合规则和数据源的可信度解决数据源冲突的功能，提高了数据集成的准确性。另外，本发明可以针对不同的数据融合领域进行融合规则定制，从而达到可以对不同应用领域的数据进行融合。同时，本发明可以以即插即用的方式对数据交换集成系统中新添加的异构数据源进行集成，适应信息领域发展对数据交换集成技术的需要。系统采用开放、模块化、可扩展的架构，数据源管理服务器、数据模式服务器、数据交换引擎服务器、数据融合服务器和应用解析服务器可以位于一台计算机中，也可以是分布在网络环境中的多台设备中，易于以即插即用的方式集成至信息系统中，减少了构建信息集成系统的开销，并适应未来增长需要。

本发明的特点在于：1) 准确性：由于不同数据源对于同一数据项提供的数据可能会产生冲突，本发明通过数据融合消除数据冲突，提高数据集成的准确性；

2) 实时性：数据是存储在分布式异构数据源中的，数据交换集成系统按应用需求动态采集分布式异构数据源上存储的数据，不存在数据延时的问题；3) 通用性：本发明具有很强的通用性，不同领域的异构数据源以即插即用的方式通过接口与数据交换集成系统进行交互，适应信息集成发展的需求。

#### 附图说明

图 1 为本发明实施例的系统框架图；

图 2 为本发明实施例的工作总体流程图；

图 3 为本发明实施例数据融合服务器的框架图；

图 4 为本发明实施例的数据流程图；

图 5 为本发明实施例的系统功能示意图。

#### 具体实施方式

下面结合附图对本发明的实施例作详细说明：本实施例在以本发明技术方案为前提下进行实施，给出了详细的实施方式和过程，但本发明的保护范围不限于下述的实施例。

本实施例系统以微机为设备平台，对不同信息领域中的异构数据源进行数据交换集成，且支持安全可靠的数据传输。

如图 1 所示，本实施例包括数据源管理服务器 1、数据模式服务器 2、数据交换引擎服务器 3、数据融合服务器 4 和应用解析服务器 5。其中：

所述数据源管理服务器 1 负责与有数据交换集成需求的外部异构数据源建立联系，接收外部异构数据源的注册信息，并将该数据源数据的局部模式传送到数据模式服务器 2，同时接收外部异构数据源输出的待集成数据，并将这些数据传送到数据交换引擎服务器 3；

所述数据模式服务器 2 连接到数据源管理服务器 1 上，该数据模式服务器 2 管理由数据源管理服务器 1 提供的各异构数据源的局部模式和系统的全局模式，以及局部模式和全局模式之间的映射规则；

所述数据交换引擎服务器 3 连接到数据源管理服务器 1 上，该数据交换引擎服务器 3 通过数据模式服务器 2 中的映射规则实现了数据的全局模式和局部模式之间的映射，并调用相应的数据源操作；

所述数据融合服务器 4 连接到数据交换引擎服务器 3 上，该数据融合服务器 4 对数据交换引擎服务器 3 传递的不同数据源提供的存有数据冲突的数据根据融合规则和数据源可信度进行融合，将融合后数据传递给应用解析服务器 5；

所述应用解析服务器 5 连接到数据交换引擎服务器 3 上，该应用解析服务器 5 将外部应用系统对多个异构数据源任务需求解析为多个对单一数据源的任务需求，并对数据融合服务器 4 返回的采集数据进行集成拼接，将结果返回给外部应用系统。同时，应用解析服务器 5 接收外部应用系统对任务完成的评价，数据融合服务器 4 依此修改融合规则和数据源的可信度。

如图 2 所示，是本实施例的工作总体流程图，具体描述如下：

在步骤 20 中，由应用解析服务器 5 接收到外部应用系统的服务请求。解析外部应用系统的任务需求，将对多源数据源服务需求解析为多个对单一数据源的任务需求，配置单一数据源任务的方法和属性。

在步骤 21 中，数据交换引擎服务器 3 接收到应用解析服务器 5 传递的任务，将针对单一数据源的任务参数，由全局模式映射为局部模式；根据数据源管理服务器 1 的数据源注册信息调用相应数据源。

在步骤 22 中，数据源管理服务器 1 接收到数据交换引擎服务器 3 传递来的数据集成任务，按异构数据源目标地址分解任务，建立通信对话通道，向异构数据源分发任务子包。创建采集任务、管理数据采集进程，存储数据采集结果，上传数据采集结果。

在步骤 23 中，数据交换引擎服务器 3 将数据源管理服务器 1 传来的采集数据进行由局部模式到全局模式的映射。

在步骤 24 中，数据融合服务器 4 接收到数据交换引擎服务器 3 传递过来的有冲突的集成数据，判断是否存在数据冲突，依据是规则管理单元提供的融合规则以及可信度管理单元提供的对应数据源可信度数据进行融合。

在步骤 25 中，由应用解析服务器 5 对采集到的数据进行数据集成，并将集成的数据传递给外部数据集成应用。

在步骤 26 中，返回外部应用系统对本次任务的评价，依此修改数据源的可信度和融合规则。

如图 3 所示,是本实施例的数据融合框架图,系统通过融合引擎与应用解析服务器 55 和数据交换引擎服务器 3 连接,从数据交换引擎服务器 3 中获得待融合数据及相关信息,并返回最终融合结果到应用解析服务器 5。融合引擎根据规则管理单元提供的融合规则和可信度管理单元提供的数据源可信度进行融合。

如图 4 所示,是本实施例的数据流程图。根据系统的全局模式,应用解析服务器 5 将针对多个数据源的单一任务分解为多个针对单一数据源的任务;由数据交换引擎服务器 3 将全局模式的任务参数转换为相应数据源的局部模式任务参数;数据源管理服务器 1 根据数据源的操作命令和任务参数完成数据采集;数据交换引擎服务器 3 将采集的数据由局部模式转换为全局模式;经数据融合服务器 4 解决数据冲突问题;再将数个全局模式的数据集成统一的全局模式数据,传递给外部应用系统。

如图 5 所示,是本实施例的系统功能示意图。管理数据源包括数据源的添加、查询、删除等操作。通过数据模式服务器 2 可以对数据模式和映射规则进行管理,数据模式管理包括局部模式的添加、删除、查询等操作和全局模式的生成、浏览操作;映射规则管理包括映射规则的添加、删除、查询操作。数据融合管理通过数据融合管理器设置/查询规则库信息和数据源可信度信息;设置规则库信息和数据源可信度信息包括创建、修改、增加、删除等操作;查询规则库信息和数据源可信度信息即可以根据提供的查询条件进行任意组合对所需信息进行查询。系统以全局模式视图对外部应用系统提供统一的数据视图,可以屏蔽外部数据源的异构性,使外部应用系统通过对统一的全局模式数据的操作来处理外部数据源的数据。

本实施例中,数据源管理服务器、数据模式服务器、数据交换引擎服务器、数据融合服务器和应用解析服务器位于同一台计算机中或分布在同一个网络环境中的多台计算机中。

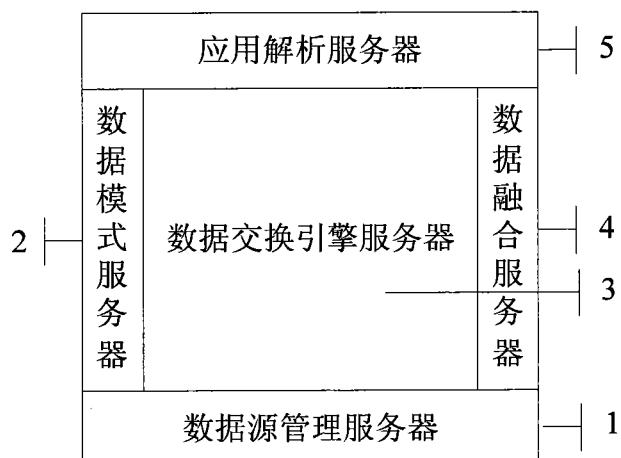


图 1

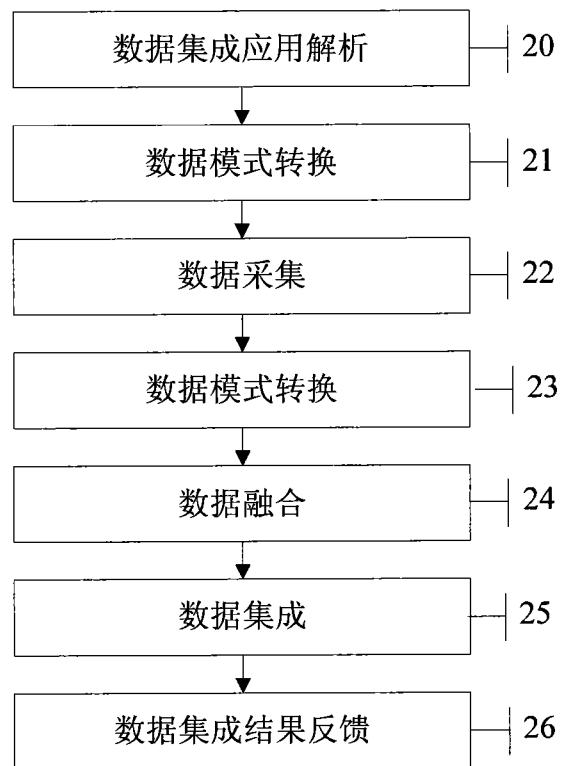


图 2

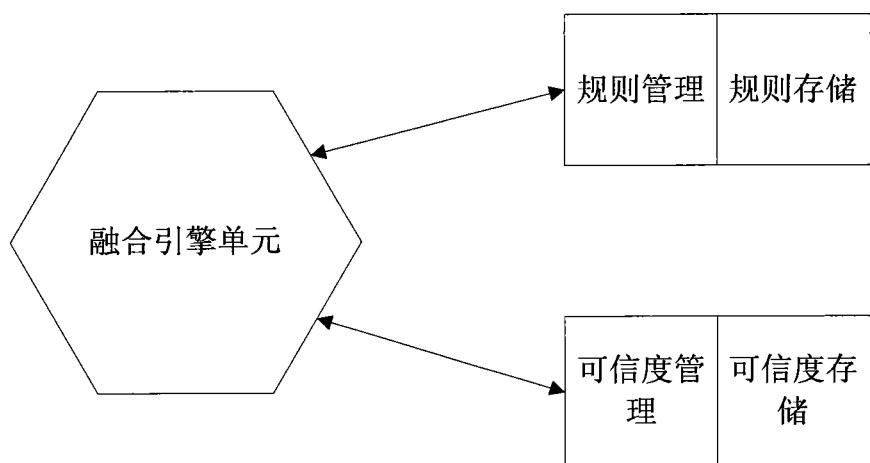


图 3

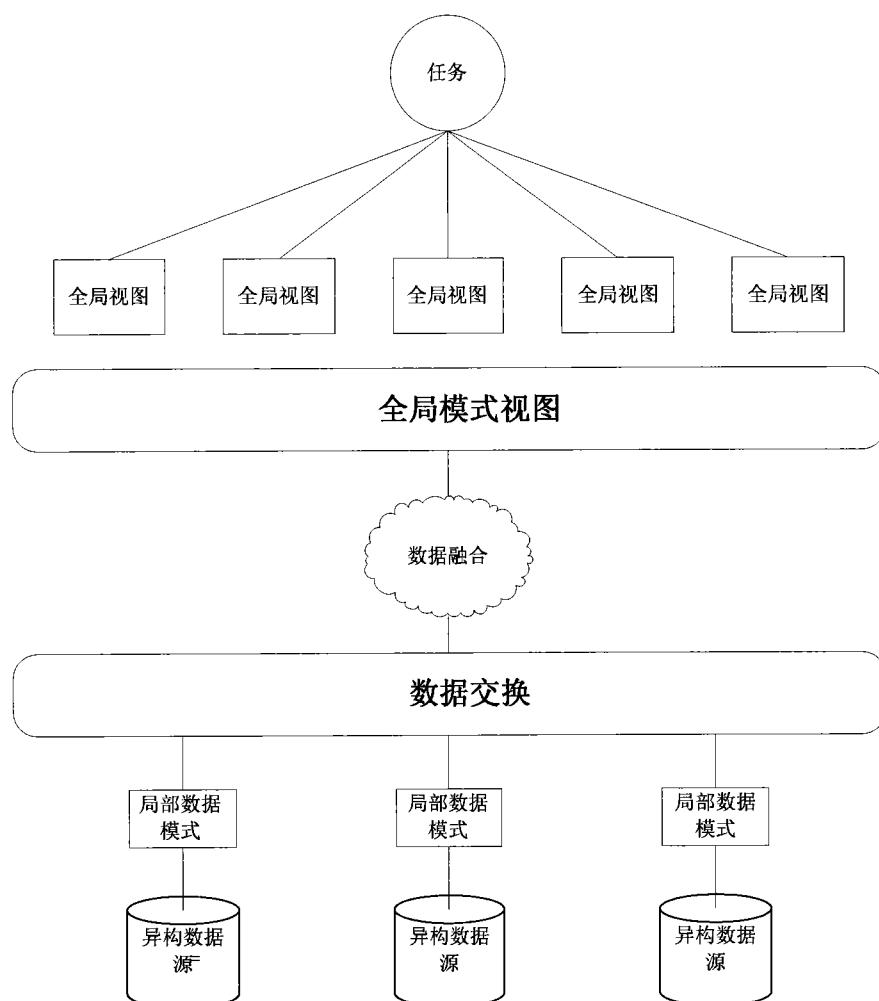


图 4

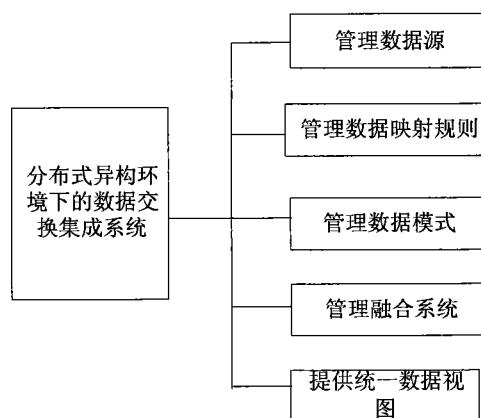


图 5