



Informatica® Multidomain MDM 10.3

概览导向

Informatica Multidomain MDM 概览导向

10.3

2018 年 9 月

© 版权所有 Informatica LLC 2001, 2019

本软件和文档仅根据包含使用与披露限制的单独许可协议提供。未事先征得 Informatica LLC 同意，不得以任何形式、通过任何手段（电子、影印、录制或其他手段）复制或传播本文档的任何部分。

美国政府权利交付给美国政府客户的程序、软件、数据库及相关文档和技术数据是指适用的联邦采购条例和政府机构特定补充条例中定义的"商业计算机软件"或"商业技术数据"。因此，使用、复制、披露、修改和改编应遵循适用的政府合同中规定的限制和许可条款、政府合同条款的适用范围以及 FAR 52.227-19 商用计算机软件许可中规定的额外权利。

Informatica、Informatica 标志和 ActiveVOS 是 Informatica LLC 在美国和世界其他许多司法管辖区的商标或注册商标。欲获得 Informatica 商标的最新列表，请访问 <https://www.informatica.com/trademarks.html>。其他公司和产品名称可能是其各自所有者的商业名称或商标。

本软件和/或文档中的若干部分受第三方所拥有的版权约束。所需的第三方声明随产品一起提供。

本文档中的信息如有更改，恕不另行通知。如发现本文档中有什么问题，请通过以下电子邮件地址向我们报告：infa_documentation@informatica.com。

Informatica 产品根据对应协议的条款和条件进行担保。INFORMATICA 按"原样"提供本文档中的信息，无任何明示或暗示的担保，包括但不限于任何适销性和特定用途适用性担保，也没有任何非侵权担保或条件。

发布日期: 2019-05-28

目录

前言	5
Informatica 资源	5
Informatica Network	5
Informatica 知识库	5
Informatica 文档	5
Informatica 产品可用性矩阵	6
Informatica Velocity	6
Informatica Marketplace	6
Informatica 全球客户支持部门	6
第 1 章： Informatica MDM Hub 简介	7
主数据管理	7
主数据和主数据管理	7
客户案例研究	8
采用主数据管理的主要动因	8
作为企业 MDM 平台的 Informatica MDM Hub	8
关于 Informatica MDM Hub	8
核心功能	8
第 2 章： Informatica MDM Hub 体系结构	10
核心组件	10
Hub 存储	10
Hub 服务器	11
进程服务器	11
Hub 控制台	11
层次结构管理器	11
安全访问管理器	11
存储库管理器	12
服务集成框架	12
Informatica Data Director	12
工作流管理器	13
Entity 360 框架	14
Informatica MDM 配置工具	14
Hub 控制台	14
IDD 配置管理器	15
Provisioning Tool	15
何时使用配置工具	16
第 3 章： 重要概念	17
传入和传出数据流	17

主要传入数据流（调节）	18
主要传出数据流（分发）	19
批处理和实时处理	19
批处理	19
登录进程	20
暂存进程	20
加载进程	21
标志化进程	21
匹配进程	21
合并进程	22
发布进程	22
实时处理	22
Hub 存储中的数据库	22
内容元数据	23
基础对象	23
交叉引用 (XREF) 表	23
历史记录表	23
workflow集成和状态管理	23
层次结构管理	24
关系	24
层次结构	24
实体	24
时间轴	24
索引	25

前言

本文档对 Informatica^(R) Multidomain MDM 进行了概述，介绍了该产品的体系结构，并定义了在其组织中使用时需要了解的一些重要概念。

本文档旨在向参与 Informatica Multidomain MDM 实施的人员介绍一些重要的 Informatica Multidomain MDM 概念。本文的目标读者是负责在组织中管理、实施或使用 MDM Hub 的人员，包括（但不限于）项目经理、安装人员、开发人员、管理员、系统集成人员、数据库管理员、数据管理者以及 Informatica Multidomain MDM 实施过程中涉及的其他技术专家。本文档的目标是，使用户能够简要而全面地了解该产品套件的概况，并说明如何在产品文档集中查找有关特定主题的详细信息。

Informatica 资源

Informatica Network

Informatica Network 囊括了 Informatica 全球客户支持部门、Informatica 知识库和其他产品资源。要访问 Informatica Network，请访问 <https://network.informatica.com>。

成员可以执行以下操作：

- 在一个位置访问您的所有 Informatica 资源。
- 在知识库中搜索文档、常见问题和最佳实践等产品资源。
- 查看产品可用性信息。
- 查看支持案例。
- 查找当地的 Informatica 用户组网络并与您的伙伴进行协作。

Informatica 知识库

使用 Informatica 知识库可在 Informatica Network 中搜索文档、入门知识文章、最佳实践和 PAM 等产品资源。

要访问知识库，请访问 <https://kb.informatica.com>。如果您对知识库有任何疑问、意见或建议，请与 Informatica 知识库团队联系，电子邮件地址为 KB_Feedback@informatica.com。

Informatica 文档

要获取有关产品的最新文档，请浏览 Informatica 知识库，网址为 https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx。

如果您对此文档有任何疑问、意见或建议，请与 Informatica 文档团队联系，电子邮件地址为 infa_documentation@informatica.com。

Informatica 产品可用性矩阵

产品可用性矩阵 (PAM) 指明了产品版本支持的操作系统版本、数据库以及其他类型的数据源和目标。如果您是 Informatica Network 成员，您可以访问 PAM，网址为 <https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Informatica Velocity

Informatica Velocity 收集了 Informatica 专业服务开发的一系列提示和最佳实践。Informatica Velocity 基于数以百计的数据管理项目的实际经验而开发，汇集了我们曾在世界各地组织就职的顾问在成功规划、开发、部署和维护数据管理解决方案方面的知识。

如果您是 Informatica Network 成员，您可以访问 Informatica Velocity 资源，网址为 <http://velocity.informatica.com>。

如果您对 Informatica Velocity 有任何疑问、意见或建议，请通过 ips@informatica.com 与 Informatica 专业服务联系。

Informatica Marketplace

Informatica Marketplace 是一个论坛，该论坛中提供的解决方案可补充、扩展或增强您的 Informatica 实现。您可以利用 Informatica 开发人员和合作伙伴提供的数以百计解决方案中的任何方案，提高生产率，加快项目的实现时间。您可以访问 Informatica Marketplace，网址为 <https://marketplace.informatica.com>。

Informatica 全球客户支持部门

您可以通过电话或 Informatica Network 上的联机支持与全球支持中心联系。

要查找您当地的 Informatica 全球客户支持部门电话号码，请访问 Informatica 网站，链接为：
<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>。

如果您是 Informatica Network 成员，您可以使用联机支持，网址为 <http://network.informatica.com>。

第 1 章

Informatica MDM Hub 简介

本章包括以下主题：

- [主数据管理, 7](#)
- [作为企业 MDM 平台的 Informatica MDM Hub, 8](#)

主数据管理

本节介绍了主数据管理，并将其作为一项准则，用于提高企业内部的数据可靠性。

主数据和主数据管理

主数据是指一组常见核心实体及其属性和值，这些信息对于公司业务至关重要，需要在两个或更多系统或业务流程中使用。主数据示例包括客户、产品、员工、供应商和位置数据。通常，主数据会分布在组织内的多个渠道和应用程序中，因此必然会包含重复的和冲突的数据，这使情况非常复杂。

主数据管理 (MDM) 是指用于创建和维护主数据的受控进程（与企业的记录系统一样）。实施 MDM 的目的是，确保主数据在经过验证后正确、一致且完整。此外，还可以通过 MDM 确保主数据可在相关环境中传递，以供内部或外部业务流程、应用程序或用户使用。

最后，MDM 会作为更广泛的数据管理程序的一部分进行部署，该程序会将技术、人员、策略和进程结合在一起。以下步骤构成了实施 MDM 解决方案的迭代过程。

步骤 1：策略

确定数据域和策略制定者。然后，数据域和策略制定者会制定策略定义、策略、目标、度量和修订过程。

步骤 2：过程

过程执行者可为人员、应用程序和服务定义数据用法、管理流程和协议，包括如何存储、存档和保护数据。

步骤 3：控制

过程管理者可创建控制机制以实施和监视策略合规性，并识别策略异常。

步骤 4：审计

审计人员可查看、访问并报告系统的历史性能。然后，审计报告会反馈给管理和策略修订（步骤 1）。

许多组织都在实施主数据管理解决方案，以提高数据可靠性并改进数据维护过程。对数据进行严格控制，反映了各个组织对其大量数据实体、数据维护流程、最佳实践以及对数据使用的安全访问具有了清晰的认识。

客户案例研究

Informatica 网站 (<http://www.informatica.com>) 提供了一些案例研究，介绍了 Informatica 客户如何通过在其组织中部署 Informatica MDM Hub 而受益。

采用主数据管理的主要动因

各个组织实施主数据管理解决方案旨在实现以下目标：

- 遵从法规，如遵从财务报表和数据隐私要求。
- 避免使企业陷入尴尬。例如，可以提高回访的有效性，避免向已故个人发送邮件。
- 简化业务流程，整合软件许可证，减少与数据管理、应用程序开发、数据清理、第三方数据提供商以及资本成本相关的费用，从而达到节省成本的目的。
- 减少重复的、不准确的和质量不高的数据，提高整个组织的工作效率，让资源集中精力执行战略性更强或价值更高的活动。
- 提高准确客户数据的透明度以及访问能力，进而提升市场营销活动的成效，并增加向客户和潜在客户进行交叉销售和追加销售的机会，从而增加企业收入。
- 战略目标，如客户忠诚度和保留能力、卓越的供应链运转能力、战略采购和承包、地域扩张和营销成效。

作为企业 MDM 平台的 Informatica MDM Hub

本节介绍了作为 MDM 平台的 Informatica MDM Hub（以下称为 *Informatica MDM Hub*）。

关于 Informatica MDM Hub

Informatica MDM Hub 是在企业内部部署 MDM 解决方案的最佳可用平台。Informatica MDM Hub 提供了一个灵活的集成企业 MDM 平台，该平台由模型驱动，可用于创建和管理各种主数据。

Informatica MDM Hub 通过以下方式实现这些特性：

集成

Informatica MDM Hub 提供了一个含有所有数据管理技术的代码库，可处理所有模式下的所有实体数据类型，以供操作和分析。

模型驱动

Informatica MDM Hub 可按照组织的要求和样式为其业务定义建立模型。所有元数据和业务服务都是根据组织的定义而生成的。可以为 Informatica MDM Hub 配置历史记录和沿袭。

灵活

Informatica MDM Hub 可实现所有类型的 MDM 样式注册表。经过调节的受信任数据源和样式可以合并在一个 Hub 中。Informatica MDM Hub 也可与旧版 Hub 共存。

核心功能

数据在到达 Hub 时通常未进行标准化。此标准化包括名称更正（例如，将 Mike 更正为 Michael）、地址标准化（例如，将 123 Elm St., NY NY 标准化为 123 Elm Street, New York, NY）以及数据转换（将一个数据模型转换为另一个数据模型）。可以使用第三方数据提供程序（如 D&B 和 Acxiom）提供的数据来补充或增加数据。Informatica MDM Hub 在其用户界面中与主要的第三方数据提供程序进行了 Out-of-the-box 集成。

在对数据进行标准化和补充之后，可通过在记录之间快速进行匹配来确定公用记录。确定公用记录后，可以将这些记录链接为注册表样式，也可以将匹配的记录中的最佳属性合并在一起，以创建最佳数据版本。此调节进程可提供相关系统中的最佳属性，该进程可在 Informatica 信任框架中实现，并受所配置的业务规则控制。

对于许多组织而言，将人员和组织相互关联起来是一项关键要求。通过 Informatica MDM Hub 的层次结构管理功能，可以使用户组人员进入住户层次结构，而公司则进入法人层次结构。

Informatica MDM Hub 还提供了基于 GUI 的功能，可以使用户定义和配置一些业务规则，用于控制清理、匹配和合并数据的方式。此数据管理工作流会向数据管理者提供有关异常或非自动匹配的信息，以便进行解决。

系统将根据已配置的权利规则来提供 Informatica MDM Hub 中的所有数据，以确保只有授权用户才能查看或修改数据，并在必要时屏蔽重要数据（如税号 ID）。

在 Informatica MDM Hub 中共享数据的一个共同目标是，将该数据与相关源系统以及下游系统进行同步。可以对 Informatica MDM Hub 进行配置，以便以实时模式、准实时模式或批处理模式来处理这些同步。如果采用实时或准实时模式，则 Informatica MDM Hub 可以灵活地避免首先启动更改的系统出现环回。

Informatica MDM Hub 还能动态地将事务和活动数据汇总为一个中心记录，从而利用内置在 Hub 中的联合查询技术。这样，组织就可以在 Hub 中仅存储引用数据，并可以实时访问所有事务数据。

通过完整的客户端及其事务视图，用户可以配置在数据发生更改时所触发的通知事件，并可启动工作流程或电子邮件，或者调用 Web 服务。这样，组织可以在发生更改时做出响应。

最后，可以对 Informatica MDM Hub 进行配置，以便使用预先配置的 Web 服务来共享数据，组织也可以通过协调多个服务来整合更高层次的功能。

第 2 章

Informatica MDM Hub 体系结构

本章包括以下主题：

- [核心组件, 10](#)
- [层次结构管理器, 11](#)
- [安全访问管理器, 11](#)
- [存储库管理器, 12](#)
- [服务集成框架, 12](#)
- [Informatica Data Director, 12](#)
- [工作流管理器, 13](#)
- [Entity 360 框架, 14](#)
- [Informatica MDM 配置工具, 14](#)

核心组件

Informatica MDM Hub 由以下核心组件组成：

- Hub 存储
- Hub 服务器
- 进程服务器
- Hub 控制台

Hub 存储

Hub 存储用于存储和合并业务数据。Hub 存储包含有关 Informatica MDM Hub 实施中所有数据库的通用信息。Hub 存储驻留在受支持的数据库服务器环境中。

Hub 存储包含以下内容：

- 不同源系统中所有实体的所有主记录
- 大量元数据以及相关规则，用于确定并持续维护每个主记录中最可靠的单元格级别属性
- 数据合并功能逻辑，如合并和取消合并数据

Hub 服务器

Hub 服务器是管理 Informatica MDM Hub 的核心和通用服务的运行时组件。Hub 服务器是在应用程序服务器上部署的一个 J2EE 应用程序，用于协调 Hub 存储中的数据处理，以及与外部应用程序的集成。

进程服务器

进程服务器清理和匹配数据，并执行批处理作业，例如加载，重新计算 BVT 和重新验证。进程服务器在应用程序服务器环境中进行部署。

进程与清理引擎相连接，从而将数据标准化，并对数据进行优化，以便于匹配和合并。

Hub 控制台

Hub 控制台是指 Informatica MDM Hub 的用户界面，其中包含一组供管理员和数据管理者使用的工具。每个工具都可以使用户执行一个特定操作或一组相关操作，例如，构建数据模型、运行批处理作业、配置数据流、配置外部应用程序对 Informatica MDM Hub 资源的访问权限，以及其他系统配置和操作任务。

Hub 控制台打包在 Hub 服务器应用程序中。可在任何客户端计算机上通过浏览器和 Sun 的 Java Web Start 使用 URL 启动该控制台。

注意: Hub 控制台提供哪些工具取决于您的 Informatica 许可协议。

层次结构管理器

使用 层次结构管理器 管理各种源系统中的关系数据。例如，在始发源系统中，记录往往已经存在层次结构，如客户至帐户、销售至帐户或产品至销售。您可以使用 层次结构管理器 查看这些关系并定义新的关系。您也可以搜索、导航以及整合关系数据。

管理员和数据管理者可使用各种工作台工具访问 层次结构管理器。

下表列出了各种角色以及每种角色使用的工作台工具。

角色	工具	目的
管理员	模型工作台 > 层次结构	配置在 层次结构管理器 中查看和操作数据关系所需的元素，例如，实体类型、层次结构、关系类型、包以及配置文件。
数据管理者	数据管理者工作台 > 层次结构管理器	在 Hub 存储 中创建、管理、搜索、导航以及整合关系数据。

注意: 部署 Hub 服务器 时，部署进程同时会在 J2EE 应用程序服务器环境中安装 层次结构管理器 的运行时组件。

安全访问管理器

Informatica 安全访问管理器 (SAM) 是 Informatica MDM Hub 的一部分，它提供了全面而精细的安全机制，确保只有经过身份验证和授权的用户才可访问 Informatica MDM Hub 数据、资源和功能。安全访问管理器提供了

一种安全决策机制，并可与第三方安全提供程序产品集成，以便为访问 Informatica MDM Hub 的用户提供安全服务（身份验证、授权和用户配置文件服务）。

注意：配置和实施 Informatica MDM Hub 安全性的方式取决于您组织的特定安全要求、IT 部署环境以及您组织的安全策略、过程和最佳实践。

存储库管理器

存储库管理器是 Hub 控制台中的一个工具，它可以使管理员管理其 Informatica MDM Hub 实施中的元数据。元数据用于描述各种架构设计和配置组件，例如，基础对象和关联列、清理函数、匹配规则和 Hub 存储中的映射等。

管理员可以使用存储库管理器执行以下任务：

- 验证 Informatica MDM Hub 存储库中的元数据，并生成有关需要引起注意的 *问题*（即，物理架构和逻辑架构之间的差异或问题）的报告。
- 比较存储库并生成描述存储库间差异的更改列表。
- 将设计对象从一个存储库复制到另一个存储库，例如，将设计对象从开发升级到生产，或者在各个 Informatica MDM Hub 实施之间导出/导入设计对象。在分布式开发环境中，开发人员可以使用存储库管理器工具共享和重新使用设计对象。
- 将存储库的元数据导出到 XML 文件，以供后续导入或存档。
- 使用存储库的图形模型视图来查看架构。

有关存储库管理器的详细信息，请参阅《*Multidomain MDM 存储库管理器指南*》。

服务集成框架

服务集成框架 (SIF) 是 Informatica MDM Hub 的一部分，用于与外部程序和应用程序相连接。通过 SIF，外部应用程序可以使用以下任一体系结构变体实现请求/响应交互：

- 使用 SOAP 协议的松散耦合 Web 服务。
- 基于 Enterprise JavaBeans (EJB) 或 XML 的紧密耦合 Java 远程过程调用。
- 基于 Java Message Service (JMS) 的异步消息。

利用这些功能，Informatica MDM Hub 可以支持多种数据访问模式，并通过 SIF SDK 公开众多 Informatica MDM Hub 数据服务，还能根据 Informatica Hub 中的数据更改生成事件。这样就有利于与外部应用程序和数据源进行传入和传出集成，并可在同步和异步模式下使用。

Informatica Data Director

Informatica Data Director (IDD) 是 Informatica MDM Hub 的数据管理应用程序，业务用户可以通过它有效地创建、管理、使用和监视主数据。Informatica Data Director 是一个基于 Web 的任务型应用程序，它通过工作流来工作，并可以进行自定义和配置，该应用程序具有一个基于 Web 的配置向导，可根据您组织的数据模型创建易于使用的界面。

集成任务管理可确保所有数据更改在影响“最佳数据版本”之前都会自动路由到相应的人员以获得批准。通过对任务进行路由，Informatica Data Director 仪表板可向业务用户提供已分配任务的视图，同时，以图形方式显示关键度量，例如，工作效率和数据质量趋势等。

此外，Informatica Data Director 可利用 Informatica 的安全访问管理器 (SAM) 模块，并提供一个全面而灵活的安全框架，以实现属性级和数据级安全性。通过该功能，客户可以通过加强策略合规性并确保对关键信息的访问来在开放和安全之间达到难得的平衡。

利用 Informatica Data Director，数据管理者和其他业务用户可以执行以下操作：

- **创建主数据。**通过独立工作或在业务线之间协同工作，用户可以将新的实体和记录添加到 Hub 存储中。Informatica Data Director 提供了诸如在数据输入过程中进行内联数据清理以及重复记录识别和解决等功能，可以使用户主动验证、增加和补充其主数据。
- **管理主数据。**用户可以批准和管理对主数据的更新，使用拖放来管理层次结构，解决可能的匹配以及合并重复项，并为其他用户创建和分配任务。
- **使用主数据。**用户可以从一个中央位置搜索所有主数据，然后查看主数据的详细信息和层次结构。用户还可以将 UI 组件嵌入到业务应用程序中。
- **监视主数据。**用户可以跟踪主数据的沿袭和历史记录，审计主数据的合规性，并使用可自定义的仪表板显示最相关的信息。

使用 Informatica Data Director，公司可以主动管理数据以降低质量成本，更快地查找准确信息以提高生产效率，提供完整而统一的数据和沿袭视图以保证合规性，根据对主数据关系的洞察来采取行动以增加收入。

workflows 管理器

使用 workflows 管理器 可注册业务流程管理 (BPM) 工具作为 workflows 引擎，并将 workflows 引擎映射到 操作引用存储。

默认的预定义 workflows 引擎是 Multidomain MDM 随附的许可版本的 ActiveVOS^(R)。安装过程会将此版本的 ActiveVOS Server 与 MDM Hub 和 Data Director 集成起来，并部署预定义的 MDM workflows、任务类型和角色。

Informatica ActiveVOS workflows 引擎支持以下适配器：

- 适用于通过业务服务对业务实体运行的任务的适配器。此适配器的名称是 **BE ActiveVOS**。
- 适用于通过 SIF API 对主题区域运行的任务的适配器。此适配器的名称是 **Informatica ActiveVOS**。

也可以选择集成 BPM 工具的独立实例：

Informatica ActiveVOS

如果您在自己的环境下运行的是 Informatica ActiveVOS 的独立实例，可手动将您的实例与 MDM Hub 和 Data Director 集成。可以部署预定义的 MDM workflows，也可以创建自定义 workflows。有关详细信息，请参阅《*Multidomain MDM Data Director - ActiveVOS 集成指南*》。

第三方 BPM 工具

如果您在自己的环境下运行的是第三方实例，可手动将您的实例与 MDM Hub 和 Data Director 集成。可以部署预定义的 MDM workflows，也可以创建自定义 workflows。有关详细信息，请参阅《*Multidomain MDM 业务流程管理器适配器 SDK 实施指南*》。

重要说明: Informatica 建议迁移到基于业务实体的 ActiveVOS workflows 适配器。Siperian workflows 适配器已弃用。Informatica 目前仍然支持该已弃用的适配器，但它将过时，Informatica 将在未来版本中停止提供支持。MDM Hub 支持主 workflows 引擎和辅助 workflows 引擎。您可以从 Siperian workflows 适配器迁移到基于业务实体的 ActiveVOS workflows 适配器。

Entity 360 框架

Entity 360 框架使用业务实体模型来支持可自定义的记录视图布局和基于 Web 的服务。

业务实体模型

Entity 360 框架依赖业务实体模型。业务实体代表对某个组织有意义的实体，例如客户、产品、帐户或位置。您可以基于在操作引用存储中定义的架构信息创建业务实体模型。有关创建业务实体模型的详细信息，请参阅《*Multidomain MDM 置备工具指南*》。

注意: 业务实体模型类似于 Informatica Data Director 应用程序中的主题区域。

记录视图布局

定义了业务实体模型后，您可以为每个业务实体模型创建自定义记录视图布局。记录视图布局可以同时显示主数据和外部数据源（例如，Twitter 订阅源或从 CRM 系统提取的数据）。您可以在置备工具中创建布局，而在 Informatica Data Director 中显示该布局。有关创建布局的详细信息，请参阅《*Multidomain MDM 置备工具指南*》。

业务实体服务

您可以使用业务实体服务直接对主数据执行操作。业务实体服务支持 Enterprise Java Bean、REST 和 SOAP。例如，您可以使用业务实体服务直接读取、转换和写入主数据。有关业务实体服务的详细信息，请参阅《*Multidomain MDM 业务实体服务指南*》。

Informatica MDM 配置工具

配置 Informatica MDM 时，可以使用以下工具：

1. **Hub 控制台。**可定义 Informatica MDM 导入、清理、管理和发布数据所需的一切内容。您必须先定义架构和基础对象，然后才能使用其他工具。
2. **IDC 配置管理器。**可以通过配置 Informatica Data Director 应用程序来为业务用户创建用户界面。
3. **置备工具。**创建业务实体模型。定义了业务实体模型后，您可以为业务用户创建自定义记录视图，以显示主数据的子集和来自外部数据源的信息。可以使用业务实体服务与主数据交互。

Hub 控制台

可以使用 Hub 控制台定义 Informatica MDM 导入、清理、管理和发布数据所需的一切内容。Hub 控制台包含一组工作台，每个工作台都包含相应工具。其中部分工具用于配置目的，其他工具则用于管理目的和数据管理。

以下工作台用于配置目的：

- “配置”工作台。可在 Hub 控制台中配置操作引用存储的数据库以及用户、安全提供程序、消息队列和工具访问。
- “模型”工作台。可配置数据模型，包括操作引用存储的架构以及源系统、信任、查询、清理函数、映射和层次结构。
- “安全访问管理器”工作台。可配置对资源的安全访问，还可以配置用户角色和用户组。
- “实用程序”工作台。可配置批处理组，还可以配置审计和调试行为。

有关详细信息，请参阅《*Multidomain MDM 配置指南*》。

IDD 配置管理器

可以使用 IDD 配置管理器创建、更新和管理 Informatica Data Director 应用程序。

在应用程序中，您可以基于在操作引用存储中定义的架构信息来定义主题区域。主题区域代表对某个组织有意义的实体，例如客户。一个主题区域包含一个根记录以及一定数量的子记录和孙记录（通过一对一或一对多关系相关联）。

有关详细信息，请参阅《*Multidomain MDM Data Director 实施指南*》。

Provisioning Tool

Use the Informatica MDM Provisioning tool to create business entity models based on the schema information that you defined in an Operational Reference Store. The business entity model is a foundational component of the Entity 360 framework.

定义

可以创建和编辑以下定义：

业务实体

定义业务实体模型。业务实体代表对某个组织很重要的实体。组织通常会定义业务实体类型来表示客户、供应商、员工、产品和帐户。

业务实体视图

定义业务实体视图模型。业务实体视图代表精简版的业务实体。

引用实体

定义引用实体。引用实体是与查找基础对象关联的业务实体。

关系

定义业务实体中的基础对象之间的关系。

转换

定义结构转换和数据转换。

任务配置

定义谁可以接受任务、谁可以启动业务流程以及默认任务属性。

扩展

定义 SOAP 服务和 WSDL URL。

设计

可以创建和设计以下用户界面项：

应用程序

创建应用程序供在 Data Director 中使用。

组件

创建用户界面组件以管理业务实体。

布局

设计用户界面布局。

何时使用配置工具

请根据您的环境使用一组不同的配置工具。

下表介绍了环境类型并指定了应使用的工具：

环境	说明	工具
Informatica MDM	您使用 MDM 组件，而不使用 Informatica Data Director 或业务实体服务。	Hub 控制台
Informatica MDM 与 Informatica Data Director	您使用 MDM 组件，还使用 Informatica Data Director 为业务用户创建标准用户界面。 注意: 此选项适用于希望保持现有 IDD 应用程序的行为（包括自定义选项卡和用户退出）的升级客户。	1. Hub 控制台 2. IDD 配置管理器
Informatica MDM 与 Informatica Data Director 和 Entity 360 框架	您使用 MDM 组件，还使用 Informatica Data Director 并启用了 Entity 360 框架。	1. Hub 控制台 2. IDD 配置管理器 3. 置备工具
Informatica MDM 与业务实体服务	您使用 MDM 组件，还使用业务实体服务从自定义应用程序对 MDM Hub 执行调用。	1. Hub 控制台 2. 置备工具

第 3 章

重要概念

本章包括以下主题：

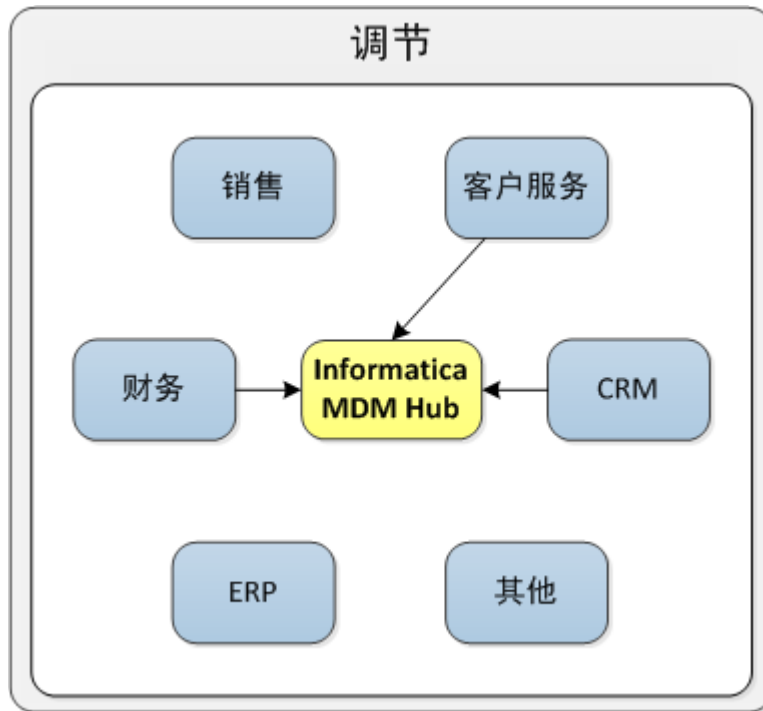
- [传入和传出数据流, 17](#)
- [批处理和实时处理, 19](#)
- [批处理, 19](#)
- [实时处理, 22](#)
- [Hub 存储中的数据库, 22](#)
- [内容元数据, 23](#)
- [工作流集成和状态管理, 23](#)
- [层次结构管理, 24](#)
- [时间轴, 24](#)

传入和传出数据流

本节介绍了 Informatica MDM Hub 的主要传入和传出数据流。

主要传入数据流（调节）

向 Informatica MDM Hub 中传入主要传入数据流称为 *调节*。



在 Informatica MDM Hub 中，客户、帐户、产品或员工等业务实体将呈现在称为 *基础对象* 的表中。对于给定的基础对象：

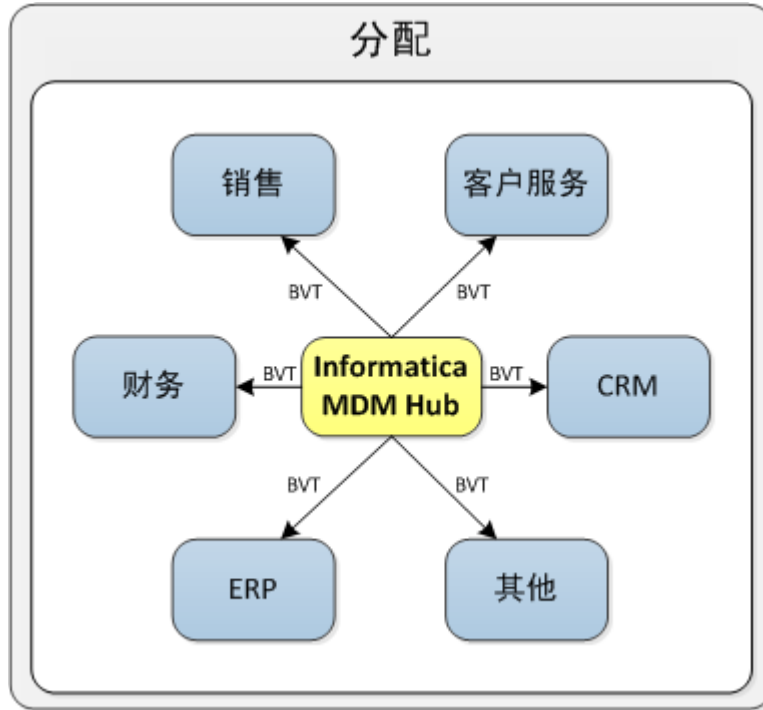
- Informatica MDM Hub 可从一个或多个 *源系统* 获取数据，该系统可以是一个运营系统或第三方应用程序，用于向 Informatica MDM Hub 提供数据，以进行清理、匹配、合并和维护。调节可能会涉及预先清理数据，以优化记录匹配和合并过程。清理是指通过验证、更正、完成或补充数据以使数据标准化的过程。
- 一个实体（例如，特定客户或帐户）可通过基础对象中的多条记录（多个数据版本）来表示。
- 然后，Informatica MDM Hub 将调节多个数据版本，以便为每个单独实体构建主记录（最佳数据版本）。合并是指合并重复记录以创建 *合并的记录* 的过程，该合并的记录包含源记录中最可靠的单元格值。

例如，假设帐单、财务和客户关系管理应用程序对于某个给定客户具有不同的帐单地址。可以对 Informatica MDM Hub 进行配置，使其根据不同源系统中列数据的相对可靠性以及诸如客户最近购买情况等数据的存在期限等因素来确定哪条数据代表最佳数据版本。

Hub 将对源记录进行调节，并将其从不同系统合并到主记录。主记录中的数据可能源于单一记录（例如，帐单系统中的最近帐单地址），也可能表示来自不同记录的复合数据。

主要传出数据流（分发）

Informatica MDM Hub 的主要传出流称为分发。为给定实体构建了主记录之后，Informatica MDM Hub 可以选择将该主记录数据分发到其他应用程序或数据库。



例如，如果一个组织的账单地址在 Informatica MDM Hub 中发生了更改，则 Informatica MDM Hub 可以通过 JMS 消息传递通知该组织中的其他系统有关更新的信息，从而使主数据在整个企业内保持同步。

批处理和实时处理

Informatica MDM Hub 具有一个定义明确的数据管理流，可通过不同的进程来工作，以便于调节和分发数据。Informatica MDM Hub 可通过两种不同的方式来处理数据：批处理和实时处理。某些 Informatica MDM Hub 实施可根据组织需要将批处理和实时处理结合起来使用。

批处理

在 MDM Hub 中，批处理作业是一个在运行期间完成离散工作单元的程序。此离散工作单元称作进程。进程会由多线程处理。批处理作业可在父基础对象匹配路径中的所有子基础对象上并行运行。

例如，首次将业务数据加载到 Hub 存储时就可以使用批处理。批处理是从源系统加载大量记录的最有效方式。

从源系统加载的数据将经过以下一系列过程：

步骤 1：登录

将来自 MDM Hub 外部源系统的数据传输到 Hub 存储的登录表中。这是调节进程的一个步骤，在 [“主要传入数据流（调节）”](#) 页面上 18 中进行了介绍。

步骤 2：暂存

从登录表中检索数据，对该数据进行清理，然后将其复制到 Hub 存储的暂存表中。这是调节进程的一个步骤。

步骤 3：加载

将数据从暂存表加载到相应的 Hub 存储表（即基础对象）中。这是调节进程的一个步骤。

步骤 4：标志化

在匹配进程使用的匹配键表中生成匹配标志，以识别用于匹配的候选基础对象记录。

步骤 5：匹配

根据匹配规则比较记录的相似性，确定记录是否重复，并对要合并的重复记录进行标记。这是调节进程的一个步骤。

步骤 6：合并

合并重复记录中的数据以创建合并的记录，该记录包含源记录中最可靠的单元格值。这是调节进程的一个步骤。

步骤 7：发布

将最佳数据版本发布到其他使用传出 JMS 消息队列的系统或进程。这是分发进程的一个步骤，在 [“主要传出数据流（分发）”](#) 页面上 19 中进行了介绍。

有关批处理的详细信息，请参阅《*Multidomain MDM 配置指南*》、《*Multidomain MDM 服务集成框架指南*》、《*Multidomain MDM 数据管理者指南*》以及 *Multidomain MDM Javadoc*。

登录进程

登录进程用于将源系统中的数据传输到 Hub 存储中的登录表。登录表为将数据从源系统传输到 Informatica MDM Hub 提供了一个中间存储。实际上，登录表是相关源系统中的数据所登录的位置。

登录进程将使用以下方法填充登录表：

批处理

第三方 ETL（提取-转换-加载）工具或其他外部进程将数据写入一个或多个登录表。此类工具或进程不属于 Informatica MDM Hub 产品套件。

联机实时处理

外部应用程序在 Hub 存储 中填充登录表。该应用程序不属于 Informatica MDM Hub 产品套件。

登录进程是 Informatica MDM Hub 的 *外部* 进程，它是使用外部批处理进程（如第三方 ETL（提取-转换-加载）工具）或以联机实时模式执行的，在该模式中，外部应用程序会直接填充 Hub 存储中的登录表。用于管理数据的后续进程是 Informatica MDM Hub 的 *内部* 进程。

暂存进程

暂存进程用于读取登录表中的数据、清理数据并将所清理的数据移入 Hub 存储中的暂存表。MDM Hub 使用暂存表作为数据从登录表传输到基础对象过程中的临时中间存储。

映射有助于在暂存进程期间在登录表和暂存表之间传输和清理数据。映射定义了 MDM Hub 填充暂存表中的列时必须使用的登录表列。映射定义了 MDM Hub 在填充暂存表之前必须执行的标准化和验证。

MDM Hub 使用您配置的清理函数对数据进行标准化和验证。使用具有专业清理功能的清理函数，例如，地址验证、地址分解、性别鉴定、文本大小写以及空格压缩。清理函数的输出将成为暂存表中目标列的输入。

注意：您可以在数据直接从源移入 MDM Hub 暂存表的 Informatica 平台上执行暂存进程。Informatica 平台暂存进程不是批处理进程。

加载进程

加载进程用于将暂存表中的数据加载到相应的 Hub 存储表（称为*基础对象*）中。

如果基础对象中的某个列从多个源系统派生其数据，则 Informatica MDM Hub 将使用*信任*来帮助比较来自不同源系统的列数据的相对可靠性。例如，对于帐单地址，订单系统可能是比销售系统更可靠的来源。

信任机制将根据每个单元格的源系统、更改历史记录和其他业务规则来度量与该单元格相关联的置信度因子。信任会考虑数据的存在期限、数据可靠性随时间的衰减程度以及数据的有效性。可以使用信任来确定幸存状况（在合并两个记录时），并确定来自源系统的更新是否足够可靠，以便于更新主记录。

信任通常与验证规则结合使用，用于告知 Informatica MDM Hub 有关数据值无效的条件。当数据满足验证规则所指定的条件时，该数据的信任值就会按验证规则中指定的百分比相应地降级。例如：

```
Downgrade trust on First_Name by 50% if Length < 3
```

标志化进程

*标志化进程*用于生成匹配标志，匹配进程随后可使用这些标志来识别要匹配的候选基础对象记录。*匹配标志*是代表基础对象的匹配列中已编码（匹配键）和未编码（原始）值的字符串。*匹配键*是固定长度的压缩编码值，它由名称或地址中的字词和数字组合而成，可以使相关的变体具有相同的匹配键值。

生成的匹配标志存储在与基础对象相关联的*匹配键表*中。对于基础对象中的每个记录，标志化进程都会在匹配键表中存储一个或多个包含生成的匹配标志的记录。匹配进程会根据匹配键表中的当前数据来运行，如果没有为基础对象中的任何记录生成匹配标志，则匹配进程将自动运行标志化进程。标志化进程可以在匹配进程之前运行，它可以在加载进程之后自动运行，也可以作为批处理作业或存储过程手动运行。

通过 Hub 控制台，用户可以调查匹配键在匹配键表中的分布。用户可以识别其数据中可能存在的*作用点*（即，可能导致*过度匹配*的大量匹配键），在这些作用点中，匹配进程会生成过多匹配，包括不相关的匹配。

匹配进程

匹配进程用于识别符合已定义的匹配规则的数据。这些规则用于定义重复数据，以供 Informatica MDM Hub 进行合并。*匹配*是指比较两个记录的相似性的过程。如果发现足够的相似性，则表示这两个记录可能彼此重复，Informatica MDM Hub 会对这些记录进行标记，以便进行合并。

基础对象中用于比较的列称为*匹配列*。每个匹配列都以基础对象中的一个或多个列为基础。各个匹配列会组合在一起而形成*匹配规则*，用于确定将两个记录视为足够相似而可以合并的条件。每个匹配规则都会为 Informatica MDM Hub 提供有关需要检查相似性的匹配列组合的信息。如果 Informatica MDM Hub 发现两个记录符合某个匹配规则，它就会记录这两个记录的主键，以及该匹配规则的标识符。这些记录会根据匹配规则的类别标记为自动或手动合并。

外部匹配用于将新数据与基础对象中的现有数据进行匹配、对匹配进行测试并检查结果，而不会实际向基础对象中加载数据。可以使用外部匹配在对数据运行实际匹配进程之前预先对数据进行测试，并测试匹配规则和检查结果。

合并进程

在匹配进程中确定了重复的记录后，合并进程会将这些重复记录合并为一个记录。

Informatica MDM Hub	主 ID	名字	中间名	姓氏	地址	城市	州/省	邮政编码
	M-0001	Abel	Noel	Willan	161 Washington Ave.	Buffalo	NY	14263
销售	SFA_ID	名字	中间名	姓氏	地址	城市	州/省	邮政编码
	12345	Abel		Willan	161 Washington Ave.	Buffalo	NY	14263
客户	Cust_ID	名字	中间名	姓氏	地址	城市	州/省	邮政编码
	502068	Abel	Noel	Willan	161 Washington Ave.	Buffalo	NY	14263
市场营销	Target_ID	名字	中间名	姓氏	地址	城市	州/省	邮政编码
	willan05	Abel	N	Willan	Elm & Carlston Streets	Buffalo	NY	14263

Informatica MDM Hub 的目标是，确定并清除所有重复数据，并将这些数据合并为一个主记录，该主记录包含源记录中最可靠的单元格值。有关合并进程的详细信息，请参阅《*Multidomain MDM 配置指南*》。

发布进程

可以配置发布进程，以便向传出 JMS 消息队列发布 BVT。对消息队列进行侦听的其他外部系统、进程或应用程序可以检索消息，并进行相应处理。有关发布进程的详细信息，请参阅《*Multidomain MDM 配置指南*》中的“配置发布进程”。

实时处理

对于实时处理，Informatica MDM Hub 的外部应用程序会通过服务集成框架 (SIF) 接口来调用 Informatica MDM Hub 操作。SIF 为各种 Informatica MDM Hub 服务提供了 API，例如，读取、清理、匹配、插入和更新记录等。

在 Informatica MDM Hub 实施中，可根据需要使用实时处理。例如，在源系统中添加、更新或删除记录后，可以使用实时处理来更新 Hub 存储中的数据。还可以使用实时处理将增量数据加载（即，首次数据加载之后发生的数据加载）加载到 Hub 存储中。

有关 SIF 的详细信息，请参阅《*Informatica MDM Hub Services Integration Framework Guide*》和 Informatica MDM Hub Javadoc。Informatica MDM Hub 可以生成事件，以便当 Hub 存储中发生特定数据更改时通知外部应用程序。

Hub 存储中的数据库

Hub 存储 是指一组包含配置设置和数据处理规则的数据库。Hub 存储 包含以下数据库：

MDM Hub 主数据库

包含 MDM Hub 环境配置设置，如用户帐户、安全配置、操作引用存储 注册表以及消息队列设置。Hub 存储 包含一个或多个 MDM Hub 主数据库。

操作引用存储

包含主数据、内容元数据以及处理和管理主数据所采用的规则。您可以分别为不同地理位置、不同组织部门以及为开发环境和生产环境配置操作引用存储。Hub 存储 包含一个或多个 操作引用存储。

注意: MDM Hub 主数据库 和 操作引用存储 的上下文中使用的术语“数据库”指的是用户架构，不能与数据库系统混淆。

内容元数据

对于架构中的每个基础对象，Informatica MDM Hub 都会自动维护相应的支持表，这些表包含与加载到 Hub 存储中的数据相关的 *内容元数据*。有关内容元数据和支持表的详细信息，请参阅《*Multidomain MDM 配置指南*》中的“构建架构”。

基础对象

基础对象（有时缩写为 *BO*）是指 Hub 存储中用于描述中央业务实体（例如客户、帐户、产品、员工等）的表。基础对象是用于合并多个源系统中的数据的端点。在 Informatica MDM Hub 实施中，某一组织的 **架构**（或数据模型）通常包含一组基础对象。

Informatica MDM Hub 的目标是，为基础对象中的每个唯一实体的每个实例创建 **主记录**。主记录包含 **最佳数据版本**（缩写为 *BVT*），它是与源记录中的最佳且最可信的单元格值合并后的记录。例如，对于“客户”基础对象，需要为每个客户创建一个主记录。该基础对象中的主记录包含该客户的最佳数据版本。

交叉引用 (XREF) 表

交叉引用表（有时称为 XREF 表）用于跟踪数据沿袭、合并的记录是由哪些系统以及这些系统中的哪些记录构成的，同时也用于跟踪数据版本。

对于每个源系统记录，Informatica MDM Hub 都会维护一个交叉引用记录，其中包含提供该记录的系统的标识符、该记录在源系统中的主键值以及该系统所提供的最新单元格值。如果基础对象启用了时间轴，则相关 XREF 表将包含期间开始日期和结束日期值。如果多个源系统提供了同一个列（如电话号码），则 XREF 表将包含每个源系统中的值。

每个基础对象记录都有一个或多个交叉引用记录。可以使用 XREF 表执行合并和取消合并操作和删除管理（删除特定源系统提供的记录），还可用于管理业务实体和关系版本。

历史记录表

历史记录表用于跟踪基础对象更改历史记录，以及基础对象与源系统之间的沿袭。Informatica 管理着多个不同的历史记录表，其中包括基础对象历史记录表和交叉引用历史记录表，用于提供详细的更改跟踪选项（包括合并和取消合并历史记录、预清理数据历史记录、基础对象历史记录以及交叉引用历史记录）。

workflow 集成和状态管理

您可以确保更新的实体数据先经过变更-批准工作流，然后再将更新的记录用于最佳数据版本 (BVT) 记录。

MDM Hub 可存储基础对象记录和交叉引用记录的预定义系统状态（如“活动”、“挂起”和“删除”），从而为 BPM 工作流工具提供支持。通过在数据上启用状态管理，MDM Hub 可与 workflow 集成进程和工具集成。MDM

Hub 可确保仅批准的活动记录可用于最佳数据版本。MDM Hub 以挂起记录的形式跟踪进程的中间阶段。有关详细信息，请参阅《*Multidomain MDM 配置指南*》中的“状态管理”。

层次结构管理

用户可以通过 层次结构管理器 管理与在 MDM Hub 中管理的记录相关联的层次结构数据。有关详细信息，请参阅《*Multidomain MDM 配置指南*》和《*Multidomain MDM 数据管理者指南*》。

关系

在层次结构管理器中，关系说明了两个特定实体之间的从属关系。可以通过指定关系类型、层次结构类型、关系属性以及关系处于活动状态的日期来定义层次结构管理器关系。有关层次结构管理器实体的信息存储在 *关系基础对象* 中。关系类型说明了关系的类。关系类型用于定义此类型的关系可包含的实体的类型、关系的方向（如果有）以及关系在 Hub 控制台中显示的方式。

层次结构

*层次结构*是指一组关系类型。这些关系类型不会划分等级，也不一定彼此相互关联。它们仅仅是分为一组的关系类型，以便于分类和识别。同一关系类型可与多个层次结构相关联。*层次结构类型*是指层次结构的逻辑分类。

实体

在层次结构管理器中，实体是指数据库中有意义且并可操作的任何对象、人员、地点、组织或其他事物。例如，特定人员的姓名、特定支票账号、特定公司、特定地址等。有关层次结构管理器实体的信息存储在 *实体基础对象* 中，您可以在 Hub 控制台中创建和配置该对象。*实体类型*是指一个或多个实体的逻辑分类。例如，医生、支票账户、银行等。同一实体类型的所有实体存储在同一实体对象中。

时间轴

通过时间轴可以管理业务实体及其关系的版本。

业务实体及其关系的版本按有效期来定义。时间轴会根据有效期和历史记录呈现二维格式的数据，使您可以跟踪过去、现在和将来对数据所做的更改。

可以通过 MDM Hub 控制台为基础对象启用时间轴。默认情况下，在为基础对象启用时间轴后，也会启用状态管理和历史记录。

版本将在与启用了时间轴的业务实体及其关系相关联的交叉引用表中进行维护。有关详细信息，请参阅《*Multidomain MDM 配置指南*》。

索引

A

ActiveVOS 进程服务器
默认工作流引擎 [13](#)
安全访问管理器 (SAM) [11](#)

B

标志化进程 [21](#)
BPM [13](#)
BPM 工具和状态管理 [23](#)

C

操作引用存储 [22](#)
层次结构 [24](#)
层次结构管理器 (HM) [11](#)
初始数据加载 [19](#)
存储库管理器 [12](#)

D

登录表 [20](#)
登录进程 [20](#)

E

Entity 360 框架 [14](#)
ETL 工具 [20](#)

F

发布进程 [22](#)
分发 [19](#)
服务集成框架 (SIF) [12](#), [22](#)

G

工作流
工作流管理器 [13](#)
任务和状态管理 [23](#)
工作流管理器 [13](#)
关系 [24](#)
过度匹配 [21](#)
关于批处理 [19](#)

H

合并重复记录 [22](#)

合并的记录 [18](#)
合并进程 [22](#)
Hub 存储 [10](#)
Hub 服务器 [11](#)
Hub 控制台 [11](#)

I

Informatica Data Director [12](#)
Informatica MDM Hub
关于 Informatica MDM Hub [8](#)
核心功能 [8](#)

J

架构 [23](#)
简介 [7](#)
交叉引用表 [23](#)
加载进程 [21](#)
基础对象 [18](#), [23](#)
进程服务器, ActiveVOS [13](#)
JMS 消息队列 [22](#)
进程服务器 [11](#)

K

框架, Entity 360 [14](#)

L

历史记录表 [23](#)

M

MDM Hub 主数据库 [22](#)

N

内容元数据 [23](#)

P

匹配标志 [21](#)
匹配规则 [21](#)
匹配键 [21](#)
匹配键表 [21](#)
匹配进程 [21](#)
匹配列 [21](#)

配置

工具 [14](#)

批处理

标志化进程 [21](#)

登录进程 [20](#)

发布进程 [22](#)

合并进程 [22](#)

加载进程 [21](#)

匹配进程 [21](#)

暂存进程 [20](#)

Q

清理函数 [20](#)

前言 [5](#)

R

任务

状态管理 [23](#)

S

时间轴 [24](#)

实时处理

关于实时处理 [22](#)

实体 [24](#)

数据模型 [23](#)

T

调节 [18](#)

提取-转换-加载工具 [20](#)

W

外部匹配 [21](#)

X

消息队列 [22](#)

信任 [21](#)

XREF 表 [23](#)

Y

验证规则 [21](#)

业务流程管理 [13](#)

映射 [20](#)

源系统 [18](#)

Z

暂存表 [20](#)

暂存进程 [20](#)

增量数据加载 [22](#)

状态管理 [23](#)

主记录 [18](#)

主数据 [7](#)

主数据管理 (MDM) [7](#)

最佳数据版本 (BVT) [18](#)

作用点 [21](#)