



Informatica® Multidomain MDM
10.5

Overview Guide

© Copyright Informatica LLC 2001, 2022

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, the Informatica logo, and ActiveVOS are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2022-09-29

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrices.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
Chapter 1: Introduction to Informatica MDM Hub	7
Master Data Management.	7
Master Data and Master Data Management.	7
Customer Case Studies.	8
Key Adoption Drivers for Master Data Management.	8
Informatica MDM Hub as the Enterprise MDM Platform.	8
About Informatica MDM Hub.	8
Core Capabilities.	9
Chapter 2: Informatica MDM Hub Architecture	10
Core Components.	10
Hub Store.	10
Hub Server.	11
Process Server.	11
Hub Console.	11
Hierarchy Manager.	11
Security Access Manager.	12
Repository Manager.	12
Services Integration Framework.	12
Informatica Data Director.	13
Workflow Manager.	13
Entity 360 Framework.	14
Informatica MDM Configuration Tools.	15
Hub Console.	15
IDD Configuration Manager.	15
Provisioning Tool.	16
When to Use the Configuration Tools.	17
Chapter 3: Key Concepts	18
Inbound and Outbound Data Flows.	18

Main Inbound Data Flow (Reconciliation)	19
Main Outbound Data Flow (Distribution)	20
Batch and Real-Time Processing	20
Batch Processing	20
Land Process	21
Stage Process	22
Load Process	22
Tokenize Process	22
Match Process	23
Consolidate Process	23
Publish Process	23
Real-Time Processing	24
Databases in the Hub Store	24
Content Metadata	24
Base Objects	24
Cross-Reference (XREF) Tables	25
History Tables	25
Workflow Integration and State Management	25
Hierarchy Management	25
Relationships	26
Hierarchies	26
Entities	26
Timeline	26
Index	27

Preface

Read the Informatica® *Multidomain MDM Overview Guide* to learn about the Multidomain MDM architecture and key concepts. The *Multidomain MDM Overview Guide* includes information about master data management, the MDM Hub, core components, and tools. Also, learn about the main inbound and outbound data flows for the MDM Hub, batch processing, and databases in the MDM Hub Store.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

CHAPTER 1

Introduction to Informatica MDM Hub

This chapter includes the following topics:

- [Master Data Management, 7](#)
- [Informatica MDM Hub as the Enterprise MDM Platform, 8](#)

Master Data Management

This section introduces master data management as a discipline for improving data reliability across the enterprise.

Master Data and Master Data Management

Master data is a collection of common, core entities along with their attributes and their values that are considered critical to a company's business, and that are required for use in two or more systems or business processes. Examples of master data include customer, product, employee, supplier, and location data. Complexity arises from the fact that master data is often strewn across many channels and applications within an organization, invariably containing duplicate and conflicting data.

Master Data Management (MDM) is the controlled process by which the master data is created and maintained as the *system of record* for the enterprise. MDM is implemented in order to ensure that the master data is validated as correct, consistent, and complete. Optionally, MDM can be implemented to ensure that Master Data is circulated in context for consumption by internal or external business processes, applications, or users.

Ultimately, MDM is deployed as part of the broader Data Governance program that involves a combination of technology, people, policy, and process. The following steps comprise the iterative process of implementing an MDM solution.

Step 1: Policy

Determine who the data domain and policy makers are. The data domain and policy makers then develop policy definitions, strategies, objectives, metrics, and a revision process.

Step 2: Process

Process executers define data usage, management processes, and protocols – for people, applications, and services – including how to store, archive, and protect data.

Step 3: Controls

Process managers create controls to enforce and monitor policy compliance and to identify policy exceptions.

Step 4: Audit

Auditors review, access, and report historical performance of the system. Auditor reports then feed into governance and policy revision (step 1).

Organizations are implementing master data management solutions to enhance data reliability and data maintenance procedures. Tight controls over data imply a clear understanding of the myriad data entities that exist across the organization, data maintenance processes and best practices, and secure access to the usage of data.

Customer Case Studies

The Informatica web site (<http://www.informatica.com>) provides case studies that describe how Informatica customers have benefited by deploying Informatica MDM Hub in their organizations.

Key Adoption Drivers for Master Data Management

Organizations are implementing master data management solutions to achieve the following goals:

- Regulatory compliance, such as financial reporting and data privacy requirements.
- Avoid corporate embarrassments. For example, you can improve recall effectiveness and avoid mailing to deceased individuals.
- Cost savings by streamlining business processes, consolidating software licenses, and reducing the costs associated with data administration, application development, data cleansing, third-party data providers, and capital costs.
- Productivity improvements across the organization by reducing duplicate, inaccurate, and poor-quality data, helping to refocus resources on more strategic or higher-value activities.
- Increased revenue by improving visibility and access to accurate customer data, resulting in increased yields for marketing campaigns and better opportunities for cross-selling and up-selling to customers and prospects.
- Strategic goals, such as customer loyalty and retention, supply chain excellence, strategic sourcing and contracting, geographic expansion, and marketing effectiveness.

Informatica MDM Hub as the Enterprise MDM Platform

This section describes Informatica MDM Hub (hereafter referred to as *Informatica MDM Hub*) as an MDM platform.

About Informatica MDM Hub

Informatica MDM Hub is the best platform available today for deploying MDM solutions across the enterprise. Informatica MDM Hub offers an integrated, model-driven, and flexible enterprise MDM platform that can be used to create and manage all kinds of master data.

Informatica MDM Hub implements these characteristics in the following ways:

Integrated

Informatica MDM Hub provides a single code-base with all data management technologies, and handles all entity data types in all modes (for operational and analytical use).

Model-Driven

Informatica MDM Hub models an organization's business definitions according to its own requirements and style. All metadata and business services are generated on the organization's definitions. Informatica MDM Hub can be configured with history and lineage.

Flexible

Informatica MDM Hub implements all types of MDM styles registry. Reconciled trusted source of truth and styles can be combined within a single hub. Informatica MDM Hub also coexists with legacy hubs.

Core Capabilities

As data arrives at the hub, it is often not standardized. This standardization includes name corrections (for example, Mike to Michael), address standardizations (for example, 123 Elm St., NY NY to 123 Elm Street, New York, NY), as well as data transformations (one data model to another). The data can be enriched or augmented with data from third-party data providers such as D&B and Acxiom. Informatica MDM Hub provides out-of-the-box integration with major third-party data providers within its user interface.

After data standardization and enrichment, common records are identified by rapidly matching against each other. Once common records are identified, you can either link them as a registry style or merge the best attributes from the matched records to create the Best Version of the Truth. This reconciliation process, achieved within the Informatica Trust Framework and governed by configured business rules, provides the best attributes from contributing systems.

Relating people and organizations is a key requirement for many organizations. Informatica MDM Hub's Hierarchy Management capabilities let users group people into households and companies into corporate hierarchies.

Informatica MDM Hub also provides GUI-based functionality, enabling users to define and configure business rules that affect how data is cleansed, matched, and merged. This data management workflow presents the exceptions or non-automated matches to the data steward for resolution.

All data in the Informatica MDM Hub is available based on the entitlement rules that are put in place, ensuring that only authorized users can view or modify the data and, if necessary, mask important data (such as tax ID numbers).

One common goal of sharing the data in Informatica MDM Hub is to synchronize it with contributing source systems as well as downstream systems. Informatica MDM Hub can be configured to handle these synchronizations in real time, near-real time, or batch mode. If in real time or near-real time mode, Informatica MDM Hub is smart enough to avoid loop backs with the system that initiated the change in the first place.

With the complete view of the client and their transactions, users can configure notification events that are triggered when data changes and can kick off a workflow process, an email, or invoke a web service. This allows organizations to respond to changes as they happen.

Finally, Informatica MDM Hub can be configured to share data using pre-configured web services, or organizations can assemble higher-level functions by orchestrating multiple services.

CHAPTER 2

Informatica MDM Hub Architecture

This chapter includes the following topics:

- [Core Components, 10](#)
- [Hierarchy Manager, 11](#)
- [Security Access Manager, 12](#)
- [Repository Manager, 12](#)
- [Services Integration Framework, 12](#)
- [Informatica Data Director, 13](#)
- [Workflow Manager, 13](#)
- [Entity 360 Framework, 14](#)
- [Informatica MDM Configuration Tools, 15](#)

Core Components

The Informatica MDM Hub consists of the following core components:

- Hub Store
- Hub Server
- Process Server
- Hub Console

Hub Store

The Hub Store is where business data is stored and consolidated. The Hub Store contains common information about all of the databases that are part of a Informatica MDM Hub implementation. The Hub Store resides in a supported database server environment.

The Hub Store contains:

- all the master records for all entities across different source systems
- rich metadata and the associated rules needed to determine and continually maintain only the most reliable cell-level attributes in each master record

- logic for data consolidation functions, such as merging and unmerging data

Hub Server

The Hub Server is the run-time component that manages core and common services for the Informatica MDM Hub. The Hub Server is a J2EE application, deployed on the application server, that orchestrates the data processing within the Hub Store, as well as integration with external applications.

Process Server

The Process Server cleanses and matches data and performs batch jobs such as load, recalculate BVT, and revalidate. The Process Server is deployed in an application server environment.

The Process interfaces with cleanse engines to standardize the data and to optimize the data for match and consolidation.

Hub Console

The Hub Console is the Informatica MDM Hub user interface that comprises a set of tools for administrators and data stewards. Each tool allows users to perform a specific action, or a set of related actions, such as building the data model, running batch jobs, configuring the data flow, configuring external application access to Informatica MDM Hub resources, and other system configuration and operation tasks.

The Hub Console is packaged inside the Hub Server application. It can be launched on any client machine through a URL using a browser.

Note: The available tools in the Hub Console depend on your Informatica license agreement.

Hierarchy Manager

Use the Hierarchy Manager to manage relationship data across disparate source systems. For example, in the originating source systems, records often have existing hierarchies, such as customer-to-account, sales-to-account, or product-to-sales. You can use the Hierarchy Manager to view these relationships and to define new relationships. You can also search, navigate, and consolidate relationship data.

Administrators and data stewards access the Hierarchy Manager using different workbench tools.

The following table lists the roles and describes the workbench tool that is used by each role:

Role	Tool	Purpose
Administrator	Model workbench > Hierarchies	Configure the elements required to view and manipulate data relationships in the Hierarchy Manager, such as entity types, hierarchies, relationships types, packages, and profiles.
Data Steward	Data Steward workbench > Hierarchy Manager	Create, manage, search, navigate, and consolidate relationship data in the Hub Store.

Note: When you deploy the Hub Server, the deploy process also installs the run-time component of the Hierarchy Manager in the J2EE application server environment.

Security Access Manager

Informatica Security Access Manager (SAM) is the part of Informatica MDM Hub that provides comprehensive and highly-granular security mechanisms to ensure that only authenticated and authorized users have access to Informatica MDM Hub data, resources, and functionality. Security Access Manager provide a mechanism for security decisions, and can integrate with security providers third-party products that provide security services (authentication, authorization, and user profile services) for users accessing Informatica MDM Hub.

Note: The way in which you configure and implement Informatica MDM Hub security is governed by your organization's particular security requirements, by the IT environment in which it is deployed, and by your organization's security policies, procedures, and best practices.

Repository Manager

The Repository Manager is a tool in the Hub Console that allows administrators to manage metadata in their Informatica MDM Hub implementation. Metadata describes the various schema design and configuration components such as base objects and associated columns, cleanse functions, match rules, and mappings in the Hub Store.

Using the Repository Manager, administrators can perform the following tasks:

- Validate the metadata in a Informatica MDM Hub repository and generate a report of *issues* (discrepancies or problems between the physical and logical schemas) that warrant attention.
- Compare repositories and generate change lists that describe the differences between them.
- Copy design objects from one repository to another such as promoting a design object from development to production, or exporting/importing design objects between Informatica MDM Hub implementations. In a distributed development environment, developers can use the Repository Manager tool to share and re-use design objects.
- Export the repository's metadata to an XML file for subsequent import or archival purposes.
- Visualize the schema using a graphical model view of the repository.

For more information about the Repository Manager, see the *Multidomain MDM Repository Manager Guide*.

Services Integration Framework

The Services Integration Framework (SIF) is the part of Informatica MDM Hub that interfaces with external programs and applications. SIF enables external applications to implement the request/response interactions using any of the following architectural variations:

- Loosely coupled web services using the SOAP protocol.
- Tightly coupled Java remote procedure calls based on Enterprise JavaBeans (EJBs) or XML.
- Asynchronous Java Message Service (JMS)-based messages.

These capabilities enable Informatica MDM Hub to support multiple modes of data access, expose numerous Informatica MDM Hub data services through the SIF SDK, and produce events based on data changes in the Informatica Hub. This facilitates inbound and outbound integration with external applications and data sources, which can be used in both synchronous and asynchronous modes.

Informatica Data Director

The Informatica Data Director (IDD) is a data governance application for Informatica MDM Hub that enables business users to effectively create, manage, consume, and monitor master data. Informatica Data Director is web-based, task-oriented, workflow-driven, highly customizable, and highly configurable, providing a web-based configuration wizard that creates an easy-to-use interface based on your organization's data model.

Integrated task management ensures that all data changes are automatically routed to the appropriate personnel for approval prior to impacting to the 'best version of the truth.' As tasks are routed, the Informatica Data Director Dashboard provides business users with a view of assigned tasks, while also providing a graphical view into key metrics such as productivity and data quality trending.

In addition, Informatica Data Director leverages Informatica's Security Access Manager (SAM) module, providing a comprehensive and flexible security framework - enabling both attribute and data level security. With this, customers can strike that elusive balance between open and secure by strengthening policy compliance and ensuring access to critical information.

Informatica Data Director enables data stewards and other business users to:

- **Create Master Data.** Working individually or collaboratively across lines of business, users can add new entities and records to the Hub Store. Offering capabilities such as inline data cleansing and duplicate record identification and resolution during data entry, Informatica Data Director enables users to proactively validate, augment, and enrich their master data.
- **Manage Master Data.** Users can approve and manage updates to master data, manage hierarchies using drag and drop, resolve potential matches and merge duplicates, and create and assign tasks to other users.
- **Consume Master Data.** Users can search for all master data from a central location, and then view master data details and hierarchies. Users can also embed UI components into business applications.
- **Monitor Master Data.** Users can track the lineage and history of master data, audit their master data for compliance, and use a customizable dashboard that shows them the most relevant information.

With the Informatica Data Director, companies can reduce cost of quality by proactively managing data, improve productivity by finding accurate information faster, enable compliance by providing a complete, consistent view of data and lineage, and increase revenue by acting on master data relationship insights.

Workflow Manager

Use the Workflow Manager to register a business process management (BPM) tool as a workflow engine and to map the workflow engine to Operational Reference Stores.

The default, predefined workflow engine is the licensed version of the ActiveVOS[®] Server that is included with Multidomain MDM. The installation process integrates this version of the ActiveVOS Server with the MDM Hub and Data Director, and deploys predefined MDM workflows, task types, and roles.

The Informatica ActiveVOS workflow engine supports the following adapters:

- An adapter for tasks that operate on business entities through business services. The adapter name is **BE ActiveVOS**.
- An adapter for tasks that operate on subject areas through SIF APIs. The adapter name is **Informatica ActiveVOS**.

You can also choose to integrate standalone instances of BPM tools:

Informatica ActiveVOS

If you run a standalone instance of Informatica ActiveVOS in your environment, you can manually integrate your instance with the MDM Hub and Data Director. You can deploy the predefined MDM workflows or create custom workflows. For more information, see the *Multidomain MDM Data Director - ActiveVOS Integration Guide*.

Third-party BPM tool

If you run a third-party instance in your environment, you can manually integrate your instance with the MDM Hub and Data Director. You can deploy the predefined MDM workflows or create custom workflows. For more information, see the *Multidomain MDM Business Process Manager Adapter SDK Implementation Guide*.

Important: Informatica recommends that you migrate to the business entity-based ActiveVOS workflow adapter. The Siperian workflow adapter is deprecated. Informatica will continue to support the deprecated adapter, but it will become obsolete and Informatica will drop support in a future release. The MDM Hub supports a primary workflow engine and a secondary workflow engine. You can migrate from the Siperian workflow adapter to the business entity-based ActiveVOS workflow adapter.

Entity 360 Framework

The Entity 360 framework uses business entity models to support customizable record view layouts and web-based services.

Business Entity Models

The Entity 360 framework depends on business entity models. A business entity represents an entity with significance to an organization, such as customers, products, accounts, or locations. You create business entity models based on the schema information that you defined in an Operational Reference Store. For more information about creating business entity models, see the *Multidomain MDM Provisioning Tool Guide*.

Note: A business entity model is similar to a subject area in an Informatica Data Director application.

Record View Layouts

With business entity models defined, you can create customized record view layouts for each business entity model. The record view layout can display both master data and external data sources, such as a Twitter feed or an extract from your CRM system. You create the layout in the Provisioning tool, and you display the layout in Informatica Data Director. For more information about creating layouts, see the *Multidomain MDM Provisioning Tool Guide*.

Business Entity Services

You can use business entity services to act on the master data directly. Business entity services support Enterprise Java Beans, REST, and SOAP. For example, you can use business entity services to read, transform, and write master data directly. For more information about business entity services, see the *Multidomain MDM Business Entity Services Guide*.

Informatica MDM Configuration Tools

When you configure Informatica MDM, you can use the following tools:

1. **Hub Console.** Define everything that Informatica MDM requires to import, cleanse, manage, and publish data. You must define the schema and base objects before you use the other tools.
2. **IDD Configuration Manager.** Create a user interface for business users by configuring an Informatica Data Director application.
3. **Provisioning Tool.** Create business entity models. With business entity models defined, you can create customized record views for business users that display a subset of master data plus information from external data sources. You can use business entity services to interact with the master data.

Hub Console

Use the Hub Console to define everything that Informatica MDM requires to import, cleanse, manage, and publish data. The Hub Console contains a set of workbenches, each of which contains tools. Some of the tools are for configuration purposes, while others are for administration and for managing data.

Use the following workbenches for configuration purposes:

- Configuration workbench. Configure databases for the Operational Reference Stores, users, security providers, message queues, and access to tools in the Hub Console.
- Model workbench. Configure the data model, including the schema for an Operational Reference Store, source systems, trust, queries, cleanse functions, mappings, and hierarchies.
- Security Access Manager workbench. Configure secure access to resources, and configure user roles and user groups.
- Utilities workbench. Configure batch groups, and configure audit and debugging behavior.

For more information, see the *Multidomain MDM Configuration Guide*.

IDD Configuration Manager

Use the IDD Configuration Manager to create, update, and manage Informatica Data Director applications.

In an application, you define subject areas based on the schema information that you defined in an Operational Reference Store. A subject area represents an entity with significance to an organization, such as customers. A subject area has a root record and some number of child records and grandchild records that are related through one-to-one or one-to-many relationships.

For more information, see the *Multidomain MDM Data Director Implementation Guide*.

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.

Define

You can create and edit the following definitions:

Definition	Description
Business entities	Define business entity models. Business entities represent entities with significance to an organization. Organizations commonly define business entity types to represent customers, suppliers, employees, products, and accounts.
Business entity views	Define business entity view models. Business entity views represent a condensed version of a business entity.
Reference entities	Define reference entities. Reference entities are business entities that are associated with lookup base objects.
Relationships	Define the relationships between the base objects in the business entities.
Transformations	Define structural transformations and data transformations.
Task configurations	Define who can accept tasks, define who can launch business processes, and define the default task properties.
Extensions	Define SOAP services and WSDL URLs.

Design

You can create and design the following user interface items:

User Interface Item	Description
Applications	Create an application for use in Data Director.
Components	Create user interface components to manage business entities.
Layouts	Design user interface layouts.

When to Use the Configuration Tools

Based on your environment, you use a different set of configuration tools.

The following table describes the types of environments and identifies which tools you use:

Environment	Description	Tools
Informatica MDM	You use MDM components. You do not use Informatica Data Director or Business Entity Services.	Hub Console
Informatica MDM with Informatica Data Director	You use MDM components. You also use Informatica Data Director to create a standard user interface for business users. Note: This option is supported for upgrade customers who want to maintain the behavior of existing IDD applications, including custom tabs and user exits.	1. Hub Console 2. IDD Configuration Manager
Informatica MDM with Informatica Data Director and the Entity 360 Framework	You use MDM components. You also use Informatica Data Director with the Entity 360 framework enabled.	1. Hub Console 2. IDD Configuration Manager 3. Provisioning tool
Informatica MDM with Business Entity Services	You use MDM components. You also use business entity services to make calls to the MDM Hub from a custom application.	1. Hub Console 2. Provisioning tool

CHAPTER 3

Key Concepts

This chapter includes the following topics:

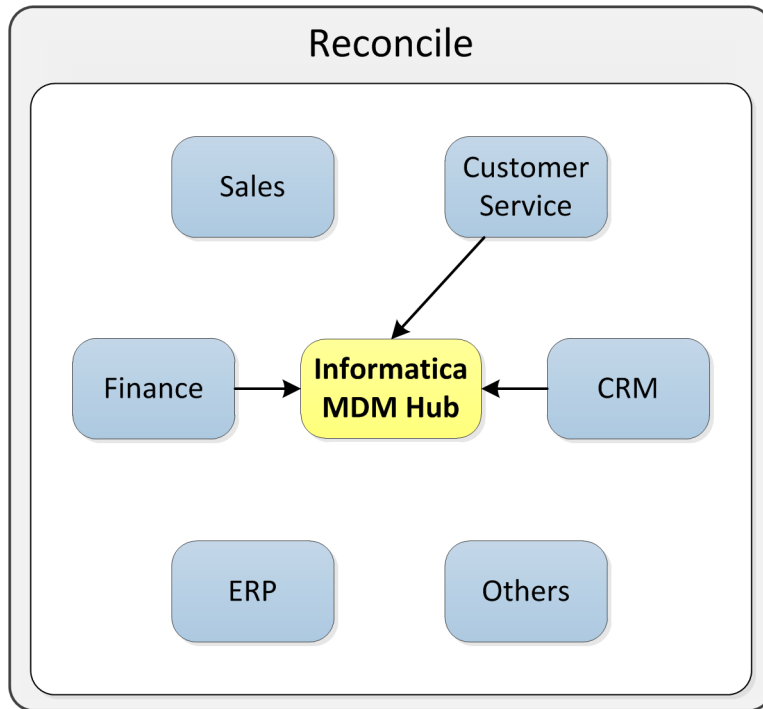
- [Inbound and Outbound Data Flows, 18](#)
- [Batch and Real-Time Processing, 20](#)
- [Batch Processing, 20](#)
- [Real-Time Processing, 24](#)
- [Databases in the Hub Store, 24](#)
- [Content Metadata, 24](#)
- [Workflow Integration and State Management, 25](#)
- [Hierarchy Management, 25](#)
- [Timeline, 26](#)

Inbound and Outbound Data Flows

This section describes the main inbound and outbound data flows for Informatica MDM Hub.

Main Inbound Data Flow (Reconciliation)

The main inbound flow into Informatica MDM Hub is called *reconciliation*.



In Informatica MDM Hub, business entities such as customers, accounts, products, or employees are represented in tables called *base objects*. For a given base object:

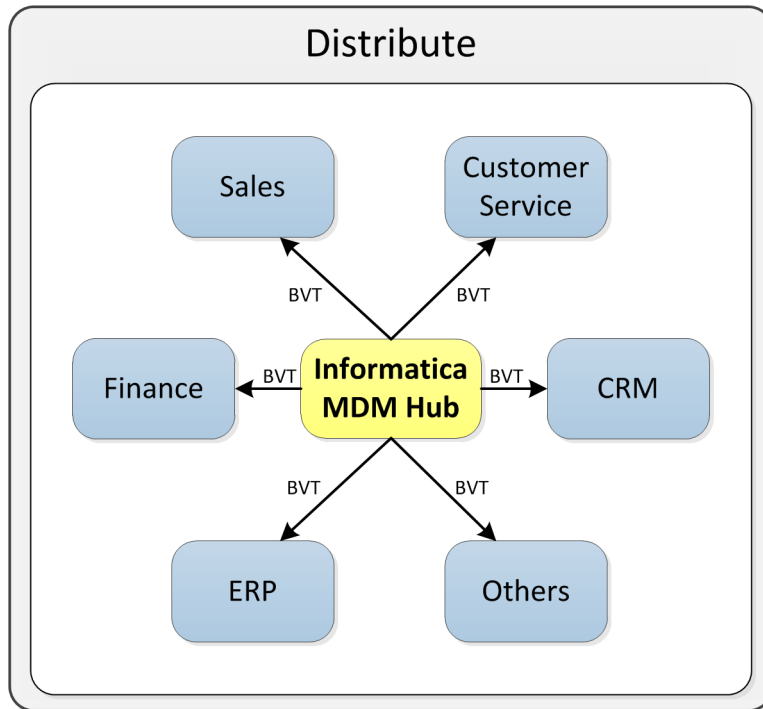
- Informatica MDM Hub obtains data from one or more *source systems*, an operational system or third-party application that provides data to Informatica MDM Hub for cleansing, matching, consolidating, and maintenance. Reconciliation can involve cleansing the data beforehand to optimize the process of matching and consolidating records. Cleansing is the process by which data is standardized by validating, correcting, completing, or enriching it.
- An individual entity (such as a specific customer or account) can be represented by multiple records (multiple versions of the truth) in the base object
- Informatica MDM Hub then reconciles multiple versions of the truth to arrive at the master record, the best version of the truth, for each individual entity. Consolidation is the process of merging duplicate records to create a *consolidated record* that contains the most reliable cell values from the source records.

For example, suppose the billing, finance, and customer relationship management applications all have different billing addresses for a given customer. Informatica MDM Hub can be configured to determine which data represents the best version of the truth based on the relative reliability of column data from different source systems based on such factors as the age of the data (the customer's most recent purchase).

The Hub reconciles and consolidates source records from different systems into a master record. Data in the master record might derive from a single record (such as the most recent billing address from the billing system), or it might represent a composite of data from different records.

Main Outbound Data Flow (Distribution)

The main outbound flow out of Informatica MDM Hub is called *distribution*. Once the master record is established for a given entity, Informatica MDM Hub can then (optionally) distribute the master record data to other applications or databases.



For example, if an organization's billing address has changed in Informatica MDM Hub, then Informatica MDM Hub can notify other systems in the organization (through JMS messaging) about the updated information so that master data is synchronized across the enterprise.

Batch and Real-Time Processing

Informatica MDM Hub has a well-defined data management flow that proceeds through distinct processes in order for the data to get reconciled and distributed. Data can be processed by Informatica MDM Hub into two different ways: batch processing and real-time processing. Many Informatica MDM Hub implementations use a combination of both batch and real-time processing as applicable to the organization's requirements.

Batch Processing

In the MDM Hub, a batch job is a program that completes a discrete unit of work when it runs. This discrete unit of work is called a process. Processes are multithreaded. Batch jobs can run in parallel on all the child base objects that are in the match path of the parent base object.

For example, you can use batch processing the first time you load business data into the Hub Store. Batch processing is the most efficient way to load a large number of records from source systems.

The data that you load from source systems goes through the following series of processes:

Step 1: Land

Transfers data from a source system that is external to the MDM Hub to landing tables in the Hub Store. Part of the reconciliation process described in [“Main Inbound Data Flow \(Reconciliation\)” on page 19](#).

Step 2: Stage

Retrieves data from the landing table, cleanses it, and copies it into a staging table in the Hub Store. Part of the reconciliation process.

Step 3: Load

Loads data from the staging table into the corresponding Hub Store table called the base object. Part of the reconciliation process.

Step 4: Tokenize

Generates match tokens in a match key table that the match process uses to identify candidate base object records for matching.

Step 5: Match

Compares records for points of similarity (based on match rules), determines whether records are duplicates, and flags duplicate records for consolidation. Part of the reconciliation process.

Step 6: Consolidate

Merges data in duplicate records to create a consolidated record that contains the most reliable cell values from the source records. Part of the reconciliation process.

Step 7: Publish

Publishes the best version of the truth to other systems or processes that use outbound JMS message queues. Part of the distribution process described in [“Main Outbound Data Flow \(Distribution\)” on page 20](#).

For more information about batch processes, see the *Multidomain MDM Configuration Guide*, *Multidomain MDM Services Integration Framework Guide*, *Multidomain MDM Data Steward Guide*, and the *Multidomain MDM Javadoc*.

Land Process

The land process transfers data from a source system to landing tables in the Hub Store. A *landing table* provides intermediate storage in the flow of data from source systems into Informatica MDM Hub. In effect, landing tables are “where data lands” from contributing source systems.

The land process populates landing tables using one of the following methods:

Batch processing

A third-party ETL (Extract-Transform-Load) tool or other external process writes the data into one or more landing tables. Such tools or processes are not part of the Informatica MDM Hub suite of products.

On-line, real-time processing

An external application populates landing tables in the Hub Store. This application is not part of the Informatica MDM Hub suite of products.

The land process is *external* to Informatica MDM Hub and is executed using an external batch process such as a third-party ETL (Extract-Transform-Load) tool, or in on-line, real-time mode in which an external application directly populates landing tables in the Hub Store. Subsequent processes for managing data are internal to Informatica MDM Hub.

Stage Process

The stage process reads data from a landing table, cleanses the data, and moves the cleansed data into a staging table in the Hub Store. The MDM Hub uses the staging table as a temporary, intermediate storage in the flow of data from landing tables into base objects.

Mappings help the transfer and cleansing of data between landing and staging tables during the stage process. A mapping defines which landing table column the MDM Hub must use to populate a column in the staging table. A mapping defines the standardization and verification that the MDM Hub must perform before it populates a staging table.

The MDM Hub standardizes and verifies data by using the cleanse functions that you configure. Use cleanse functions for specialized cleansing functionality, such as address verification, address decomposition, gender determination, text casing, and white space compression. The output of the cleanse function becomes the input to the target column in the staging table.

Load Process

The load process loads data from the staging table into the corresponding Hub Store table, called a *base object*.

If a column in a base object derives its data from multiple source systems, Informatica MDM Hub uses *trust* to help with comparing the relative reliability of column data from different source systems. For example, the Orders system might be a more reliable source of billing addresses than the Sales system.

Trust provides a mechanism for measuring the confidence factor associated with each cell based on its source system, change history, and other business rules. Trust takes into account the age of data, how much its reliability has decayed over time, and the validity of the data. Trust is used to determine survivorship (when two records are consolidated) and whether updates from a source system are sufficiently reliable to update the master record.

Trust is often used in conjunction with validation rules, which tell Informatica MDM Hub the condition under which a data value is not valid. When data meets the criterion specified by the validation rule, then the trust value for that data is downgraded by the percentage specified in the validation rule. For example:

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

Tokenize Process

The *tokenize process* generates match tokens that are used subsequently by the match process to identify candidate base object records for matching. *Match tokens* are strings that represent both encoded (match key) and unencoded (raw) values in the match columns of the base object. *Match keys* are fixed-length, compressed, and encoded values, built from a combination of the words and numbers in a name or address, such that relevant variations have the same match key value.

The generated match tokens are stored in a *match key table* associated with the base object. For each record in the base object, the tokenize process stores one or more records containing generated match tokens in the match key table. The match process depends on current data in the match key table, and will run the tokenize process automatically if match tokens have not been generated for any of the records in the base object. The tokenize process can be run before the match process, automatically at the end of the load process, or manually, as a batch job or stored procedure.

The Hub Console allows users to investigate the distribution of match keys in the match key table. Users can identify potential *hot spots* in their data (high concentrations of match keys that could result in *overmatching*) where the match process generates too many matches, including matches that are not relevant.

Match Process

The match process identifies data that conforms to the match rules that you have defined. These rules define duplicate data for Informatica MDM Hub to consolidate. *Matching* is the process of comparing two records for points of similarity. If sufficient points of similarity are found to indicate that the two records are probably duplicates of each other, then Informatica MDM Hub flags those records for consolidation.

In a base object, the columns to be used for comparison purposes are called *match columns*. Each match column is based on one or more columns from the base object. Match columns are combined into *match rules* to determine the conditions under which two records are considered to be similar enough to consolidate. Each match rule tells Informatica MDM Hub the combination of match columns it needs to examine for points of similarity. When Informatica MDM Hub finds two records that satisfy a match rule, it records the primary keys of the records, as well as the match rule identifier. The records are flagged for either automatic or manual consolidation according to the category of the match rule.

External match is used to match new data with existing data in a base object, test for matches, and inspect the results without actually loading the data into the base object. External matching is used to pretest data, test match rules, and inspect the results before running the actual match process on the data.

Consolidate Process

After duplicate records have been identified in the match process, the consolidate process merges duplicate records into a single record.

Informatica MDM Hub	Master ID	First Name	MN	Last Name	Address	City	State	Zip
	M-0001	Abel	Noel	Willan	161 Washington Ave.	Buffalo	NY	14263
Sales	SFA_ID	First Name	MN	Last Name	Address	City	State	Zip
	12345	Abel		Willan	161 Washington Ave.	Buffalo	NY	14263
Accounts	Cust_ID	First Name	MN	Last Name	Address	City	State	Zip
	502068	Abel	Noel	Willan	161 Washington Ave.	Buffalo	NY	14263
Marketing	Target_ID	First Name	MN	Last Name	Address	City	State	Zip
	willan05	Abel	N	Willan	Elm & Carlston Streets	Buffalo	NY	14263

The goal in Informatica MDM Hub is to identify and eliminate all duplicate data and to merge them together into a single, consolidated master record containing the most reliable cell values from the source records. For more information about the consolidate process, see the *Multidomain MDM Configuration Guide*.

Publish Process

The publish process can be configured to publish the BVT to an outbound JMS message queue. Other external systems, processes, or applications that listen on the message queue can retrieve the message and process it accordingly. For more information about the publish process, see "Configuring the Publish Process" in the *Multidomain MDM Configuration Guide*.

Real-Time Processing

For real-time processing, applications that are external to the MDM Hub invoke the MDM Hub operations through the Services Integration Framework (SIF) interface. SIF provides APIs for various MDM Hub services, such as reading, cleansing, matching, inserting, and updating records.

In MDM Hub implementations, real-time processing is used as appropriate. For example, real-time processing can be used to update data in the Hub Store whenever a record is added, updated, or deleted in a source system. Real-time processing can also be used to handle incremental data loads. Incremental data loads are loads that occur after the initial data is loaded into the Hub Store.

For more information about SIF, see the *Multidomain MDM Services Integration Framework Guide* and the *Multidomain MDM Javadoc*. The MDM Hub can generate events to notify external applications when specific data changes occur in the Hub Store.

Databases in the Hub Store

The Hub Store is a collection of databases that contain configuration settings and data processing rules. The Hub Store contains the following databases:

MDM Hub Master Database

Contains the MDM Hub environment configuration settings, such as user accounts, security configuration, Operational Reference Store registry, and message queue settings. The Hub Store consists of one or more MDM Hub Master Database.

Operational Reference Store

Contains the master data, content metadata, and the rules to process and manage the master data. You can configure separate Operational Reference Stores for different geographies, different organizational departments, and for the development and production environments. The Hub Store consists of one or more Operational Reference Store.

Note: The term, database that is used in the context of the MDM Hub Master Database and the Operational Reference Store refers to user schemas and must not be confused with database systems.

Content Metadata

For each base object in the schema, Informatica MDM Hub automatically maintains support tables containing *content metadata* about data that has been loaded into the Hub Store. For more information about content metadata and support tables, see “Building the Schema” in the *Multidomain MDM Configuration Guide*.

Base Objects

A *base object* (sometimes abbreviated as *BO*) is a table in the Hub Store that is used to describe central business entities, such as customers, accounts, products, employees, and so on. The base object is the end-point for consolidating data from multiple source systems. In a Informatica MDM Hub implementation, the *schema* (or data model) for an organization typically includes a collection of base objects.

The goal in Informatica MDM Hub is to create the *master record* for each instance of each unique entity within a base object. The master record contains the *best version of the truth* (abbreviated as *BVT*), which is a record that has been consolidated with the best, most-trustworthy cell values from the source records. For example, for a Customer base object, you want to end up with a master record for each individual customer. The master record in the base object contains the best version of the truth for that customer.

Cross-Reference (XREF) Tables

Cross-reference tables, sometimes referred to as XREF tables, are used for tracking the lineage of data, which systems and which records from those systems contributed to consolidated records, and also for tracking versions of data.

For each source system record, Informatica MDM Hub maintains a cross-reference record that contains an identifier for the system that provided the record, the primary key value of that record in the source system, and the most recent cell values provided by that system. In case of timeline-enabled base objects, the associated XREF tables include the period start and end date values for the records. If the same column (for example, phone number) is provided by multiple source systems, the XREF table contains the value from every source system.

Each base object record has one or more cross-reference records. XREF tables are used for merge and unmerge operations, delete management (removing records that were contributed by a particular source system), and to manage versions of business entities and relationships.

History Tables

History tables are used for tracking the history of changes to a base object and its lineage back to the source system. Informatica manages several different history tables, including base object and cross-reference history tables, to provide detailed change-tracking options, including merge and unmerge history, history of the pre-cleansed data, history of the base object, and history of the cross-reference.

Workflow Integration and State Management

You can ensure that updated entity data goes through a change-approval workflow before the updated records contribute the Best Version of the Truth (BVT) records.

The MDM Hub supports BPM workflow tools by storing predefined system states, ACTIVE, PENDING, and DELETED, for base object records and cross-reference records. By enabling state management on your data, the MDM Hub integrates with workflow integration processes and tools. The MDM Hub ensures that only approved, active records contribute to the best version of the truth. The MDM Hub tracks intermediate stages of the process as pending records. For more information, see "State Management" in the *Multidomain MDM Configuration Guide*.

Hierarchy Management

The Hierarchy Manager allows users to manage hierarchy data that is associated with the records managed in the MDM Hub. For more information, see the *Multidomain MDM Configuration Guide*.

Relationships

In Hierarchy Manager, a *relationship* describes the affiliation between two specific entities. Hierarchy Manager relationships are defined by specifying the relationship type, hierarchy type, attributes of the relationship, and dates for when the relationship is active. Information about a Hierarchy Manager entity is stored in a *relationship base object*. A *relationship type* describes classes of relationships. A relationship type defines the types of entities that a relationship of this type can include, the direction of the relationship (if any), and how the relationship is displayed in the Hub Console.

Hierarchies

A *hierarchy* is a set of relationship types. These relationship types are not ranked, nor are they necessarily related to each other. They are merely relationship types that are grouped together for ease of classification and identification. The same relationship type can be associated with multiple hierarchies. A *hierarchy type* is a logical classification of hierarchies.

Entities

In Hierarchy Manager, an *entity* is any object, person, place, organization, or other thing that has meaning and can be acted upon in your database. Examples include a specific person's name, a specific checking account number, a specific company, a specific address, and so on. Information about a Hierarchy Manager entity is stored in an *entity base object*, which you create and configure in the Hub Console. An *entity type* is a logical classification of one or more entities. Examples include doctors, checking accounts, banks, and so on. All entities with the same entity type are stored in the same entity object.

Timeline

Timeline lets you manage versions of business entities and their relationships.

The versions of business entities and their relationships are defined in terms of their effective periods. Timeline provides two dimensional visibility into data based on effective periods and history, and equips you with the ability to track past, present, and future changes to data.

You can enable timeline for base objects through the MDM Hub console. When you enable timeline for base objects, state management and history are also enabled by default.

Versions are maintained in the cross-reference tables associated with the timeline-enabled business entities and their relationships. For more information, see the *Multidomain MDM Configuration Guide* .

INDEX

A

about batch processing [20](#)
ActiveVOS Process Server
 default workflow engine [13](#)

B

base objects [19, 24](#)
batch processing
 consolidate process [23](#)
 land process [21](#)
 load process [22](#)
 match process [23](#)
 publish process [23](#)
 stage process [22](#)
 tokenize process [22](#)
best version of the truth (BVT) [19](#)
BPM [13](#)
BPM tool and state management [25](#)
business process management [13](#)

C

cleanse functions [22](#)
configuration
 tools [15](#)
consolidate process [23](#)
consolidated record [19](#)
content metadata [24](#)
cross-reference tables [25](#)

D

data model [24](#)
distribution [20](#)

E

entities [26](#)
Entity 360 framework [14](#)
ETL tools [21](#)
external match [23](#)
extraction-transformation-load tools [21](#)

F

framework, Entity 360 [14](#)

H

hierarchies [26](#)
Hierarchy Manager (HM) [11](#)
history tables [25](#)
hotspots [22](#)
Hub Console [11](#)
Hub Server [11](#)
Hub Store [10](#)

I

incremental data loads [24](#)
Informatica Data Director [13](#)
Informatica MDM Hub
 about Informatica MDM Hub [8](#)
 core capabilities [9](#)
initial data loads [20](#)
introduction [7](#)

J

JMS message queues [23](#)

L

land process [21](#)
landing tables [21](#)
load process [22](#)

M

mappings [22](#)
master data [7](#)
Master Data Management (MDM) [7](#)
master records [19](#)
match columns [23](#)
match key tables [22](#)
match keys [22](#)
match process [23](#)
match rules [23](#)
match tokens [22](#)
MDM Hub Master Database [24](#)
merging duplicate records [23](#)
message queues [23](#)

O

Operational Reference Store [24](#)
overmatching [22](#)

P

preface [5](#)
Process Server [11](#)
Process Server, ActiveVOS [13](#)
publish process [23](#)

R

real-time processing
 about real-time processing [24](#)
reconciliation [19](#)
relationships [26](#)
Repository Manager [12](#)

S

schema [24](#)
Security Access Manager (SAM) [12](#)
Services Integration Framework (SIF) [12](#), [24](#)
source systems [19](#)
stage process [22](#)
staging tables [22](#)
state management [25](#)

T

tasks
 state management [25](#)
timeline [26](#)
tokenize process [22](#)
trust [22](#)

V

validation rules [22](#)

W

Workflow Manager [13](#)
workflows
 tasks and state management [25](#)
 Workflow Manager [13](#)

X

XREF tables [25](#)