MDM Reference Architecture

Onepoint Data Values People

Contents

What is MDM?

Entities like Customer, Product, Accounts, Suppliers, and Locations etc. are key ingredients that fuel the enterprise engine. The processes used to operate the enterprise depend on these entities and their quality (to what extent can these entities be trusted). These key entities are referred to as 'Master Data'.

Most enterprises are slowly realising the importance of improving the quality of their master data and managing its lifecycle more efficiently to optimise business and operations. The discipline/ capability (people, process and technology) of managing the master data within an organisation is called Master Data Management

Investment in maturing the Master Data Management capability provides the following benefits to the enterprise:

- Provides a source of master data that can be trusted to be correct and authentic by systems and people within an organisation.
- Improved data governance by establishing data principles, policies and standards, thereby improving the quality of data.
- Reduces fragmentation of the data entities (copies of master data stored at multiple locations while keeping validity and consistency).

Use of MDM

IBM in their book 'Enterprise Master Data Management' describes three patterns how the managed master data entities can be used:

Collaborative Use:

In this pattern of use, different business functions or teams of people collaborate to manage the lifecycle of the master data. Defined workflows ensure adherence to processes and that the right approvals are sought. Role-based access control determines which roles can manage which aspect of the master data attributes. Typically, in this pattern of use, the MDM system broadcasts the master data to the transactional system(s). Managing catalogues of products is an example of this pattern of use.

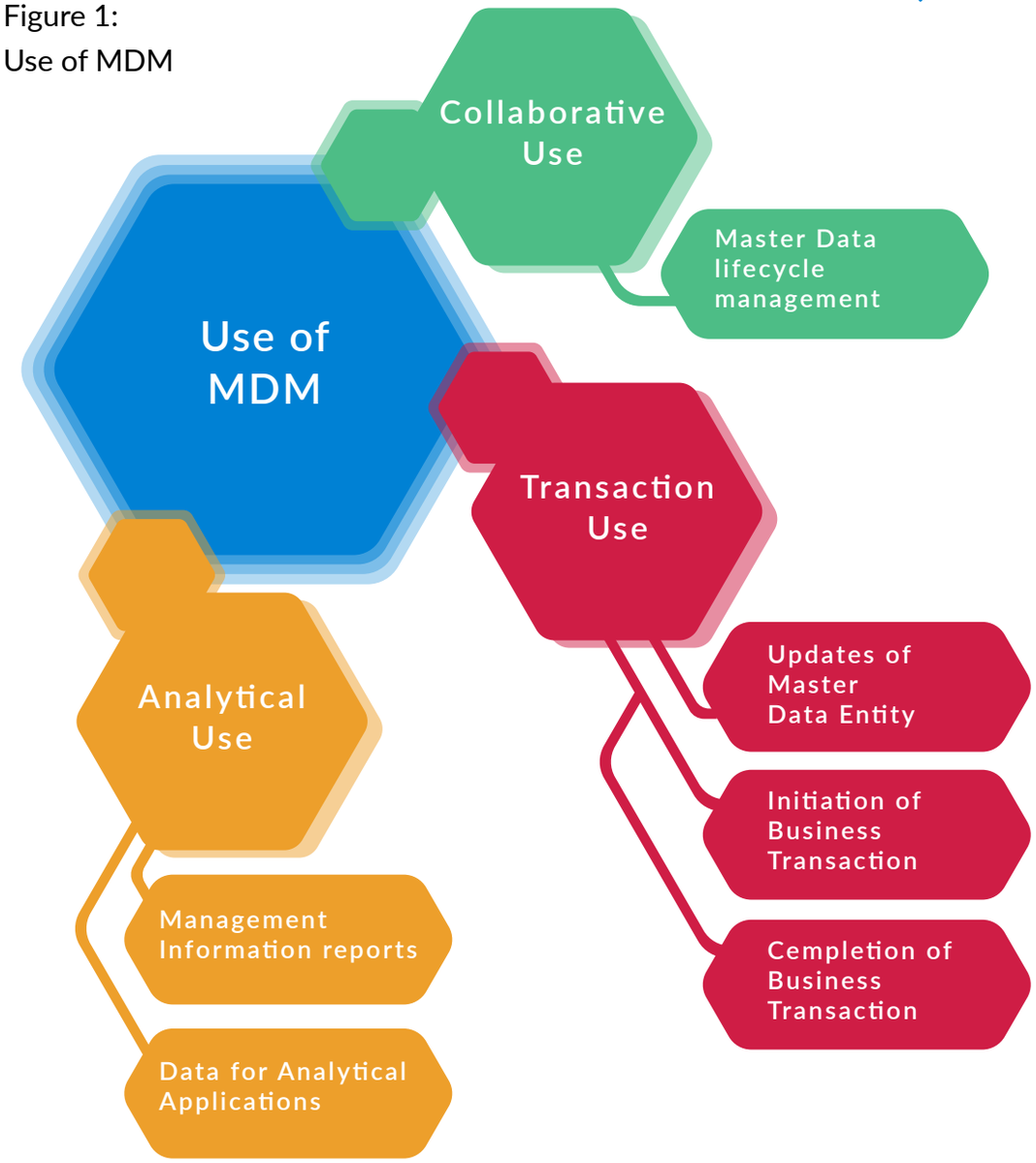
Transaction Use:

In this pattern of use the MDM system participates in end-to-end business transactions. Multiple applications and users need quick access in order to read and update the master data entity; to initiate and complete business transactions such as orders and payments etc. Here the MDM system exposes data services that can participate within an orchestrated business process. MDM systems supporting this method of use might have the need to support several hundred transactions per second on millions of master data records.

Analytical Use:

In this pattern of use, the role of MDM is to produce Management Information (MI) reports or to provide data to an analytical application such as a Data Warehouse (DWH).

Figure 1:
Use of MDM



MDM Hub Implementation Style

When implementing an MDM system one of the key decisions to be made is the choice of MDM Hub architecture. Gartner and others describe three styles of MDM Hub and choice depends on multiple considerations, such as:

- Time and budget available
- Read or Read/Write capability on Master Data
- Data consistency
- Data fragmentation

Registry Style

MDM systems implemented using this style provide a read-only view of the master data for downstream systems. Here the MDM Hub stores only a thin slice of the MDM Data (Source System Identifiers, Foreign Key Identifiers, key attributes needed for matching). Other data attributes that make up the master data entity are spread across multiple business systems. The Hub is responsible for cleansing and de-duplicating the master data entity and maintaining a reference to source master data attributes that it does not manage from the other business systems.

The main advantage of this architecture style is the low time and cost required to implement an MDM solution as well as the reduced impact on the source data systems.

The key disadvantage is performance: the time required for the registry Hub to aggregate data from multiple source systems and provide a complete master data record.

Transaction Style

MDM systems implemented using this style enable read/write operations on the master data. In this architecture style the Hub stores and manages all attributes that define the master data entity. The Hub is regarded as the single, authoritative source of the master data. Changes to the master data are enriched, cleansed, de-duplicated and pushed by the Hub to the various source systems.

This architecture style is more expensive and time consuming compared to the Registry style. In addition, it directly impacts source systems as these need to change to either implement a two-way synchronisation to maintain consistency of master data, or devolve responsibility of the master data entities by sourcing the data directly from the MDM Hub.

Transaction Style Implementation

Transaction Style can be implemented as:

Co-Existence

Master data entities co-exist in both the Hub and the source systems with a two way synchronisation between the two. Master data entities can be created/ updated either in the MDM Hub or the source system. The MDM Hub standardises, cleanses, enriches and de-duplicates the master data and pushes the data back to the source systems to rationalise the master data they hold.

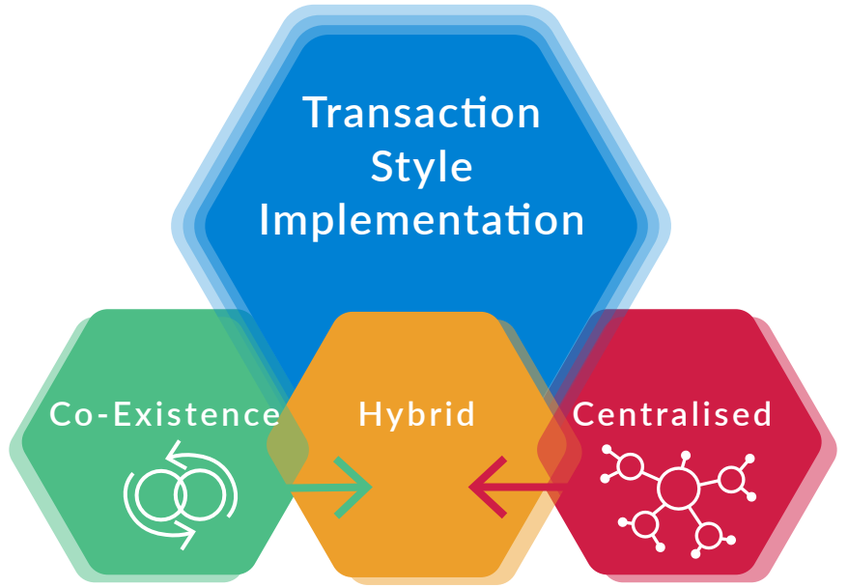
Centralised

Here the Master data entity is stored and managed solely within the MDM Hub. Creation and update of master data entity is only done via the MDM system. Source systems interact with the MDM system when they require the master data record.

Hybrid

This is the 'Goldilocks' or hybrid architecture style. Here the Hub does not store all attributes that define the master data as in Transaction Style but stores more than the Registry Style. The main aim of the Hybrid is to resolve performance issues faced by the Registry Style.

Figure 2:
Transaction Style Implementation



MDM Reference Architecture

The MDM Reference Architecture is a vendor-neutral, product-agnostic Reference Architecture that supports multiple methods of MDM use, and multiple architecture styles. It aims to define the key architecture building blocks (ABB) that will enable an organisation to govern and maintain accurate master data for all users and applications.

Architecture Principles

The proposed Reference Architecture is guided and influenced by the following principles:

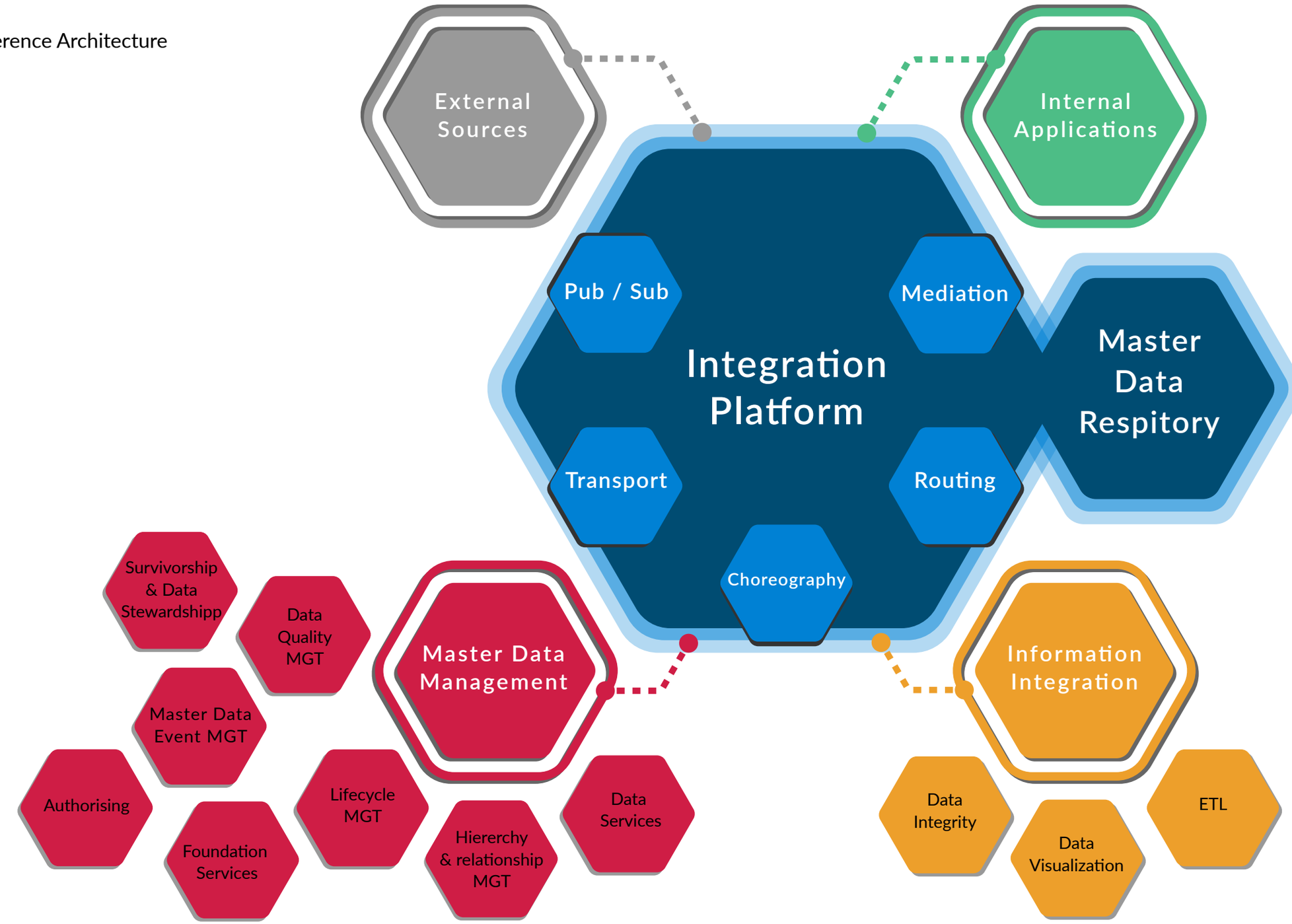
- Securely managing complete lifecycle of the managed master data entity. The data must be managed in line with prevailing privacy and data protection legislation.
- Being the single authoritative source for the managed master data, assuring the integrity, quality and uniqueness of the master data record.
- Decoupling master data from enterprise applications ensuring the master data record can be accessed by the business applications in a performant manner.
  Being a flexible data hub that can be extended to support changing business requirements and/ or regulations.

Reference Architecture

The Reference Architecture illustrated in Figure 3: MDM Reference Architecture provides architecture building blocks that can manage the lifecycle of master data as well as the quality and integrity of the data, making master data actionable and providing data services to enable the consumption and distribution of data. The design is guided by the principles outlined above.

MDM Reference Architecture

Figure 3:
MDM Reference Architecture



Key components of the Reference Architecture:

Transaction Systems and Data Sources

These are internal and external transactional systems that interact with the master data management system to request the master data entities or enrich the master data entities. These systems may, in addition, store the master data information in a registry or Co-existence Style implementation.

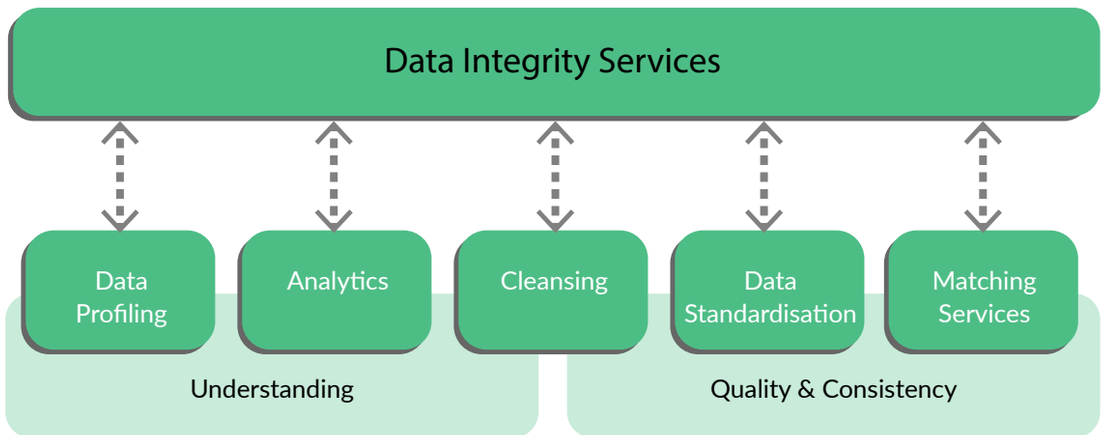
Integration Platform

The integration platform enables the Transaction Systems and Data Sources to interact with the Master Data Management system.

Information Integration

The Information Integration component provides Data Integrity Services, Extract, Transform, Load (ETL) services and Data Virtualisation services. Data Integrity Services include data profiling, analysis, cleansing, data standardisation, and matching services. Data profiling and analysis services are critical for understanding the quality of master data across enterprise systems, and for defining data validation, data cleansing, matching, and standardisation logic required to improve master data quality and consistency. ETL services support the initial and incremental extract, transform, and load of data from one or more source systems to populate the MDM system. It also helps to load data from the MDM system to analytical platforms such as Data Warehouse in batch mode. The Data Virtualisation services enables access to data distributed across multiple data sources abstracting details from the data requestor.

Figure 4:
Information Integration - Data Integrity Services



Key components of the Reference Architecture:

Master Data Management

The Master Data Management component comprises of the following subcomponents

Data Services:

This sub component provides a service interface to access the master data entities managed by the MDM system. The service interface support techniques such as messaging, method calls, web services, and batch processing.

Lifecycle Management:

This subcomponent manages the lifecycle of the managed master data. Data Quality Management Services are called on by Lifecycle Management to enforce data quality rules and perform data cleansing, standardisation, and reconciliation.

MDM Event Management:

MDM Event Management Services are called on to detect any actions that should be triggered based upon business rules or data governance policies. Master data entities will be archived or deleted depending on privacy and data protection regulations

This subcomponent triggers actions based upon changes to attributes of the master data. Business systems can subscribe to these events and take appropriate action. The MDM Event Management enables business systems to listen to data, change events and synchronise data to ensure consistency. MDM Event management can also trigger notifications to data stewards advising them to approve master data changes as part of data governance.

Authoring:

Subcomponent to author, approve, manage, customise, and extend the definition of master data. This subcomponent supports the collaborative style of MDM use and can be invoked as part of a workflow process that creates, updates and approves master data.

Data Quality Management:

This subcomponent works with the data integrity service to validate, standardise and enforce data quality rules.

Foundation:

This subcomponent provides foundation services of security, search, privacy, audit logging, versioning and workflow.

Hierarchy & Relationship Management:

This subcomponent manages master data hierarchies, groupings and relationships.

Survivorship & Data Stewardship:

This subcomponent resolves conflicts when duplicate master data entities are identified by executing survivorship rules or pushing to a data steward for manual resolution.

Master Data Repository:

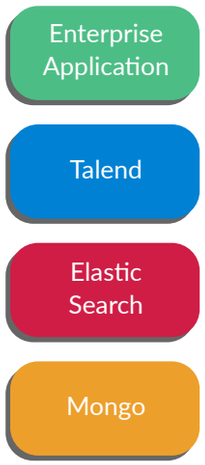
Repository that stores master data, both instance and master data definition, metadata for the MDM System, and history data that records changes to master data.

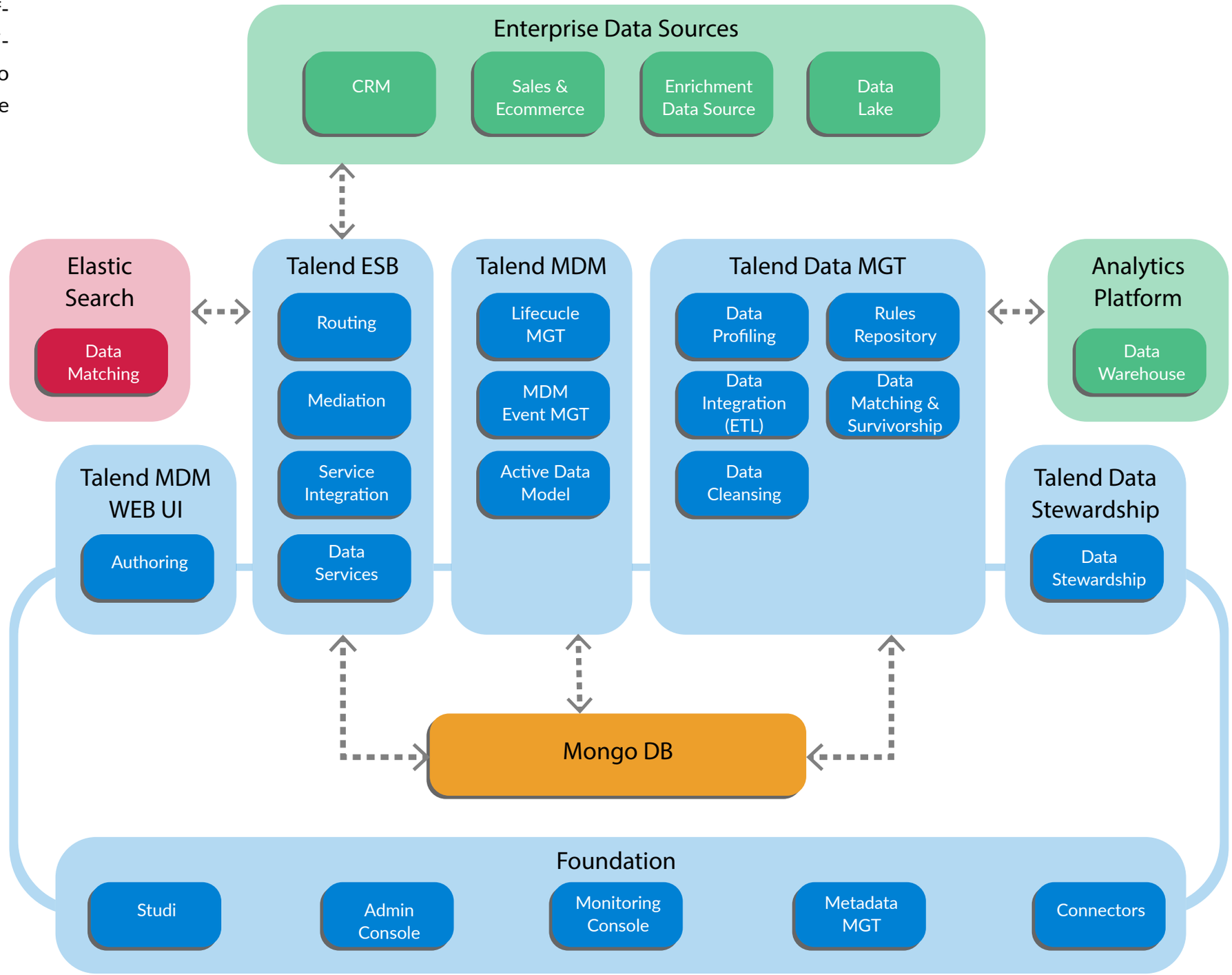
Large Scale and High Performance MDM based on Talend

How does one instantiate the logical Reference Architecture, composed of architecture building blocks listed above, to provide a high performing and scaleable Master Data Management platform?

Figure 5:
Highly scaleable, Performant MDM

Key:





Large Scale and High Performance Customer MDM

Onepoint Data Values People

Talend

Talend MDM Platform:

talend

The Talend MDM platform is comprised of three pre-integrated components:

- **Talend ESB:**
  Providing Integration and Data Services

- **Talend Data Management:**
  Providing Data Integration and Integrity services

- **Talend MDM:**
  Managing the master data entities, their hierarchies & relationships and publishing events when they change. It also provides authoring and data stewardship services.

The three pre-integrated components are underpinned by a common foundation providing metadata management, am Administration and Monitoring console, a studio for design and development, and connectors to integrate with a diverse suite of enterprise data sources.

Elastic Search

Elastic Search:

elasticsearch

Elasticsearch enables high performance, distributed search supporting a two-step matching process. The first step of the matching process performs a coarse-grained match followed by a fine-grained match using Talend Data Management's matching component.

Mongo

Mongo DB:

mongoDB

A NoSQL database enabling high performance create, read, update and delete (CRUD) operations for hierarchal master data entities.

Contact Details

Onepoint Consulting Limited

Alpha House, Unit 14, 100 Villiers Road, London NW2 5PJ

+44 (0) 203 198 6699

contact@onepointltd.com

http://www.linkedin.com/in/lancym

www.Onepointltd.com

Thank you

Lancy Mendonca – Enterprise Architect

A Computer Science grad with a BSc in Physics, Lancy Mendonca has well over two decades of experience in creating and contributing to IT architectures for the telecom and healthcare industries.

Lancy is a TOGAF-certified Enterprise/ Solution Architect and a certified Java developer with an acute understanding of Enterprise Java and associated frameworks. He specialises in enterprise architecture, enterprise integration, solution architecture, technology governance, data strategy and architecture and TM Forum Frameworx.

After leading roles at Tech Mahindra, British Telecom and Carphone Warehouse, Lancy was appointed to head up the Solution Architecture and Design team at TalkTalk. There, he was on point in the delivery of a multi-million pound business transformation project for TalkTalk's fixed line business. After five years at TalkTalk, Lancy took on the Business Support Systems brief at Tesco Telecoms. He later helped to shape Truphone's Enterprise Application strategy.

Lancy joined Onepoint in 2013, bringing his unrivalled knowledge of Enterprise Architecture and Telecoms to multiple clients including UK Broadband, SITA, Travis Perkins and Network Rail Telecom. But with his keen mind, and an instinct for identifying and solving business problems, Lancy's skills and experience benefit Onepoint collaborators across the industry piste.

An avid reader, a dedicated supporter of church charities, and someone who donates his time to supporting young learners, Lancy is passionate about many things, but particularly the future of data and the possibilities it offers to modern businesses.

Onepoint Data Values People

Onepoint Data Values People