Stack-C

Master Data Management Strategy

Author: Yogesh Joshi

Designation: MDM Architect

Date: 19 April 2019

**Project Details**

|  |  |
| --- | --- |
| **Project Name:** | **Stack-C Master Data Management Implementation** |
| **Account:** | **Micro Focus International** |
| **Vertical:** |  |
| **Location/s:** | **Newbury, UK** |
| **Customer Name:** | **Micro Focus International** |
| **Delivery Manager / Email ID:** | [Mayuresh.tendulkar@microfocus.com](mailto:Mayuresh.tendulkar@microfocus.com) |
| **Project Manager / Email ID:** |  |
| **Quality coordinator / Email ID:** |  |
| **Customer Contact Information:** | **Joe Alfini /** [joe.alfini@microfocus.com](mailto:joe.alfini@microfocus.com) |
| **Gareth Davies /** [Gareth.Davies@microfocus.com](mailto:Gareth.Davies@microfocus.com) |
| **Stephen Warner /** [Stephen.Warner@microfocus.com](mailto:Stephen.Warner@microfocus.com) |
|  |

**Revision History**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version (x.yy)** | **Date of Revision** | **Description of Change** | **Reason for Change** | **Affected Sections** | **Approved By** |
| **1.0** | **4-Mar-2019** | **NA** | **Initial** | **1,6,7,8** |  |
| **1.1** | **11-Mar-2019** | **First Draft** | **First Draft submitted for internal review** | **2,3,4,5** |  |
| **1.2** | **14-Mar-2019** | **Section updates, formatting** | **Second draft with all sections completed** | **All** |  |
| **1.3** | **22-Mar-2019** | **Section updates, formatting** | **Third draft incorporating peer review comments** | **All** |  |
| **1.4** | **26-Mar-2019** | **Section updates, formatting** | **Fourth draft incorporating EA review comments** | **All** |  |
| **1.5** | **1-Apr-2019** | **Section updates, formatting** | **Fifth draft incorporating Murteza & Anup P’s comments** | **8,11,13** |  |
| **1.6** | **2-Apr-2019** | **Section updates, formatting, New subsection** | **Sixth draft incorporating Frederic’s comments** | **1,2,8** |  |
| **1.7** | **4-Apr-2019** | **Section updates, rephrasing** | **Seventh draft incorporating Rob’s comments** | **1,2,3** |  |
| **1.8** | **9-Apr-2019** | **Section updates, formatting** | **Eighth draft incorporating EA, Steve comments** | **All** |  |
| **1.9** | **12-Apr-2019** | **Section updates** | **Ninth draft incorporating Rob P’s comments** | **All** |  |
| **2.0** | **17-Apr-2019** | **Section 8.3.2 updated** | **John Hakim’s feedback incorporated** | **8.3.2** |  |
| **3.0** | **19-Apr-2019** | **Formatting, Spellcheck, Grammar** | **Feedback from Stephan, Issued for final review** | **All** |  |

**References**

|  |  |
| --- | --- |
| **Document Name** | **Description** |
| [Stack-C Data Migration Strategy](https://microfocusinternational.sharepoint.com/:w:/r/teams/StackC-Programme/Data/001%20Data%20Migration/Data%20Migration%20Strategy/DM02%20Data%20Migration%20Strategy%20-%20Stack%20C%20001%20(15-MAR-2019).docx?d=wbc0e701f9f5e4bd9b962b90a6ede85b3&csf=1&e=bNrwDB) | The aim of this document is to provide the overall approach to be followed for the Data Migration activity (including data cleansing).  This document outlines the Data Profiling, Data Cleansing, Data Conversion and Data Validation approach to be taken in the Stack-C Program. |
| [Stack-C Architecture](https://microfocusinternational.sharepoint.com/teams/StackC-Programme/architecture/Published?web=1) | These documents provide Applications to Process mapping, Application Tracker, Stack-C Architecture Landscape, Business Models, Enterprise Architecture, Interface Catalogue, Stack-C TO BE State Architecture, Process Library (latest version), Technical / Business SME |
| Stack-C Program Charter |  |

**Reviewers**

|  |  |
| --- | --- |
| **Name** | **Role** |
| Joe Alfini | Stack-C Data Architecture Work Stream Lead |
| Robert Pidgeon | Stack-C Data Migration Work Stream Lead |
| Danial Clarke | Stack-C Enterprise Architect |
| Frederic Glenat | QA Lead |
| John Hakim | Stack-C Data Architecture Solution Architect |
| Mayuresh Tendulkar | Stack-C Data Delivery Manager (Wipro) |
| Wipro Mission Quality Team | WMQ Team |

**Approvers**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Role | Approver Level | Approved On |
| Joe Alfini | Stack-C Data Architecture Work Stream Lead | L1 |  |
| Robert Pidgeon | Stack-C Data Migration Work Stream Lead | L1 |  |
| Stephen Warner | Stack-C Data Work Stream Leader | L2 |  |

**Abbreviations**

|  |  |
| --- | --- |
| **Abbreviation** | **Description** |
| 2NF | 2nd Normal Form |
| 3NF | 3rd Normal Form |
| API | Application programming interface |
| BAU | Business As Usual |
| BI | Business Intelligence |
| CDM | Canonical Data Model |
| CPQ | Configure Price Quote |
| CRM | Customer Relationship Management |
| CRUD | Create, Read, Update, Delete |
| D&B | Dun & Bradstreet |
| DE | Data Entity |
| DG | Data Governance |
| DME | Data Migration Engine |
| DQ | Data Quality |
| EDC | Enterprise Data Catalogue |
| EDW | Enterprise Data Warehouse |
| ERP | Enterprise Resource Planning |
| ESB | Enterprise Service Bus |
| ETL | Extract Transform Load |
| GDPR | General Data Protection Regulation |
| GRID | MDM Key Mapping of unique record ID in Informatica MDM to source / legacy record IDs |
| GUID | Informatica MDM BaseObject rowed\_object column – Global Unique Identifier for a given MDM domain |
| IDQ | Informatica Data Quality |
| IFRS | International Financial Reporting Standards |
| IPC | Informatica Power Center |
| ISO | International Organisation for Standardisation |
| KYC | Know Your Customer |
| LASA | Legal Account & Sales Account |
| M&A | Mergers & Acquisitions |
| MDM | Master Data Management |
| OOB | Out-Of-Box |
| POS | Point of Sale |
| PPA | Product Pricing Analytics – This is Micro Focus internal legacy system which provides limited MDM capabilities for product data in Stack-A |
| RPL | Restricted Party List |
| SE | Software Entitlement |
| SFDC | SalesForce |
| SIF | Services Integration Framework |
| SKU | Stock Keeping Unit |
| SME | Subject Matter Expert |
| SOP | Standard Operating Procedure |
| UI | User Interface |
| UOM | Unit of Measurement |
| VPN | Virtual Private Network |

Table of Contents

[1. Overview 9](#_Toc6779815)

[2. Executive Summary 10](#_Toc6779816)

[3. Introduction 11](#_Toc6779817)

[3.1 Summary Recommendation 13](#_Toc6779818)

[4. MDM Scope Boundary 14](#_Toc6779819)

[4.1 In Scope 14](#_Toc6779820)

[4.2 Out of Scope 14](#_Toc6779821)

[5. Strategic Principles 15](#_Toc6779822)

[6. Master and Reference Data 16](#_Toc6779823)

[6.1 Reference Data 16](#_Toc6779824)

[6.2 Master Data 16](#_Toc6779825)

[6.3 External Data 17](#_Toc6779826)

[7. Master Data Quality 18](#_Toc6779827)

[7.1 Data Profiling & Current Data Quality Status 18](#_Toc6779828)

[7.1.1 Product Data 18](#_Toc6779829)

[7.1.2 Customer / Account Data 18](#_Toc6779830)

[7.1.3 Finance Data 18](#_Toc6779831)

[7.1.4 Employee Data 18](#_Toc6779832)

[7.2 Data Quality Dimensions 19](#_Toc6779833)

[7.3 Preventive Data Quality Procedures 19](#_Toc6779834)

[7.4 Reactive Data Quality Methodology 19](#_Toc6779835)

[8. Master Data Management 20](#_Toc6779836)

[8.1 Current State of Master Data at Micro Focus 20](#_Toc6779837)

[8.2 Proposed MDM Implementation Style for Micro Focus 21](#_Toc6779838)

[8.2.1 MDM Domain Outline Data Architecture 21](#_Toc6779839)

[8.3 Stack-C MDM Domain Management 22](#_Toc6779840)

[8.3.1 Product MDM Domain Management 22](#_Toc6779841)

[8.3.2 Customer / Account MDM Domain Management 23](#_Toc6779842)

[8.3.3 Manage duplicate Customer / Account & Merge 23](#_Toc6779843)

[8.3.4 Data Governance Reviews 24](#_Toc6779844)

[8.3.5 Finance MDM Domain Management 24](#_Toc6779845)

[8.3.6 Employee MDM Domain Management - 25](#_Toc6779846)

[8.3.7 Reference Data Management 25](#_Toc6779847)

[8.3.8 Other Master Data Domain Management - not covered above 25](#_Toc6779848)

[8.4 Data Stewardship Agency 26](#_Toc6779849)

[9. Master Data Governance 27](#_Toc6779850)

[10. Required Solution Capabilities 29](#_Toc6779851)

[10.1 Data Validation upon Creation 29](#_Toc6779852)

[10.2 De-duplication 29](#_Toc6779853)

[10.3 Approval 29](#_Toc6779854)

[10.4 Data Enrichment 29](#_Toc6779855)

[10.5 Audit Trail 29](#_Toc6779856)

[10.6 Post Creation Quality Monitoring 30](#_Toc6779857)

[10.7 Master Data Distribution 30](#_Toc6779858)

[10.8 Semantic Catalogue 30](#_Toc6779859)

[11. High-Level Business Process Flow – Mastered & Non-Mastered Entities 31](#_Toc6779860)

[11.1 High-Level Process Flow – Initial Data Load (Day 1) – Mastered Data Entities 31](#_Toc6779861)

[11.1.1 Data Migration Approach 31](#_Toc6779862)

[11.1.2 Product Data Migration Approach - Example 32](#_Toc6779863)

[11.1.3 Use of Informatica MDM Capabilities for De-Duplication 33](#_Toc6779864)

[11.1.4 The first pass MDM data migration pattern is 34](#_Toc6779865)

[11.1.5 Customer Life History of Data 35](#_Toc6779866)

[11.2 High-Level Process Flow – Master Data Process (Day 2) – Mastered Data Entities 37](#_Toc6779867)

[11.2.1 Overview 37](#_Toc6779868)

[11.2.2 Example Product Master Data – Day 2 38](#_Toc6779869)

[11.3 High-Level Process Flow – Initial Data Load (Day 1) – Non-Mastered Data Entities 39](#_Toc6779870)

[11.3.1 Data Migration & Cleansing Approach 40](#_Toc6779871)

[11.4 High-Level Process Flow – Master Data Process (Day 2) – Non-Mastered Data Entities 40](#_Toc6779872)

[11.5 Summary - High-Level Process Flow – Master & Non-Master Entities 41](#_Toc6779873)

[12. Tool capabilities 42](#_Toc6779874)

[13. Assumptions 45](#_Toc6779875)

[14. Best Practices 46](#_Toc6779876)

[15. Appendix 47](#_Toc6779877)

[15.1 Data Domains Summary 47](#_Toc6779878)

# Overview

Master Data is a foundational component of the Stack-C environment. A coherent Master Data Management (MDM) strategy to manage that data is crucial to the success of this transformation journey. This document outlines an approach towards MDM and the available enabling tools within the Stack-C program for maintaining the Master Data. The MDM solution design phase will help to identify the necessary steps to implement a successful MDM process for critical Master Data domains at Micro Focus.

# Executive Summary

A key objective of this MDM strategy is to build a reliable data foundation with clean & trusted Master Data for successful Stack-C implementation (*Refer to* [***Appendix 15.1***](#_Data_Domains_Summary) *to refer specific Reference & Master Data entities*)

Informatica MDM is the tool of choice to support Master Data Management in Micro Focus. Data Governance team from Chief Data Officers’ office will be leading the initiative to setup stewardship & accompanying processes, policies and rules.

Hybrid MDM Model is the optimum pattern for Micro Focus. New records are created in their originating application (“system of record”) and then passed to the MDM solution as a secondary host for onward propagation to downstream consumer applications

This hybrid MDM model will differ for each data domain as follows

|  |  |  |
| --- | --- | --- |
| Data Domain | System of Record | Secondary Host |
| Product | SalesForce CPQ | MDM Solution |
| Customer / Account | SalesForce | MDM Solution |
| Finance | NetSuite | MDM Solution (no data stewardship in MDM) |
| Employee | Workday | MDM Solution (no data stewardship in MDM) |

As a principle, Master & Reference Data entities will not exist in MDM solution if

* These entities are used only by “system of record” application and
* These entities does not need any advanced data management capabilities like Matching, Enrichment, Governance, and Distribution etc.

# Introduction

MDM programs have four moving parts, which need careful coordination and design to ensure the success of the program

|  |  |
| --- | --- |
| Data Stewardship | Use of MDM tool provides the opportunity for automated matching of inbound Master Data changes, which results in either the creation of unique new Master Data or alternatively search & provide potential matching records with less than 100% match.  In the latter case, human intervention by nominated persons (known as Data Stewards) is required to adjudicate these possible low confidence matches.   * For Customer data, Micro Focus SalesOps team require data stewards who know the local country / sub region customer data well and are located around the world. These Data Stewards are therefore, best placed to adjudicate customer matches. * Micro Focus FinOps and HR teams may require a different data stewardship based on geography to master FINANCE data. * Micro Focus data operations team is already responsible for PRODUCT data. |
| Data Models | Each MDM domain needs its data attributes defined e.g. Customer Name, Address etc., and will be described using a data model. |
| Workflow | In Micro Focus operating model, there are certain end-to-end value streams / business processes, which will have touch points with MDM domains in Stack-C applications. These business processes will be stewarded by creating workflows across MDM tool, Enterprise Service Bus (ESB) and Stack-C applications.  These processes may include orchestration of MDM steps for matching data validation & subsequent decisions for Data Stewards i.e. to create a new record or to use existing matching record available in MDM.  Review and approving of match results will be done by Workflow processes |
| Data Life History | Where and how Master Data is Created, Read, Updated, and Deleted (CRUD) is required to sync-up the Stack-C intra-system event communications (Enterprise Service Bus (ESB) Event Messages) for a given MDM domain. By modelling a CRUD matrix of which systems are involved the required orchestration will be build. |

Reference Data is

* A subset of Master Data E.g. Currency, Country etc.
* Consumed across business applications
* Managed and governed centrally

An enterprise MDM tool is the right place to manage such Reference Data. (*Refer to* [***Appendix 15.1***](#_Data_Domains_Summary)*to refer specific Reference & Master Data entities*)

The MDM solution should facilitate:

* Creation & maintenance of Master Data plus facilitate data stewardship in the situation where a given MDM domain is mastered in a system of engagement.
* Numbering services for the data objects that need unique identifiers
* Advanced data management capabilities (Data Cleansing, De-Duplication, Data Enrichment etc.)

At Micro Focus, the proposed MDM solution will have two components:

1. **Master systems of record** i.e. business applications such as SalesForce (SFDC), NetSuite, Workday, and SalesForce CPQ for
   * Management of Master and Reference Data, which are consumed by only these businesses applications & not shared across other Stack-C applications in Micro Focus.
   * Maintenance & validation of Master and Reference Data using out-of-the-box functionalities
2. An **enterprise MDM Application** for
   * Management of Master and Reference Data consumed by multiple target applications & shared across other Stack-C applications in Micro Focus.
   * Maintenance of Master Data, where advanced data management functionalities like, De-Duplication, Data Enrichment, and Data Governance.

The Master Data strategy proposed in this document will provide Micro Focus with the tools and maintenance functions to Master Data maintenance and governance for its most critical Master Data objects. This will drive standardisation of the information in the Stack-C applications.

Micro Focus will use Informatica MDM software tool in the migration of applications from Stack-B to Stack-C thus mastering data for the first time, using MDM tool. The integration of MDM into the overall Data Migration architecture pattern will be reusable and hence repeatable to migrate Stack-A to Stack-C as well as any future Mergers and Acquisitions (M&A) applications to Stack-C.

Post Stack-B to Stack-C migration, MDM tool will assist in Business as Usual (BAU) Master Data processes using Enterprise Service Bus (ESB) to synchronise Master Data to “systems of record” (Stack-C applications) and consumption of Master Data in Stack-C applications.

Although the aim is to use common MDM architecture patterns wherever possible, not all MDM domains may have the same pattern largely due to the differing Data Stewardship requirements (SalesOps, FinOps, Data Management teams etc.) and hybrid MDM architecture pattern.

## Summary Recommendation

In summary, this document proposes the following recommendations

* Use a fit-for-purpose Master Data tool for managing Master Data, rather than a traditional approach for managing all Master Data centrally and use Informatica MDM tool to master data domains with advanced data management requirements like Cleansing, Data Quality, Governance, Validation & Distribution.
* Implement all relevant Master Data Management processes prior to Go-Live including the data stewardship of migrated data via Informatica MDM wherever applicable.
* Identify and execute Data Quality rules during data migration to ensure an appropriate level of Data Quality in the Stack-C system of record.
* Stack-C program should not attempt to address all Data Management & Data Quality issues but rather follow a pragmatic approach based on business impact and resource utilisation.
* Setup a Data Stewardship Agency as an international virtual business, facing Micro Focus organisation (to support SalesOps, FinOps etc.) during both Data Migration and BAU operations to coordinate MDM data stewardship interventions and subsequent approvals of Master Data merges.
* Chief Data Officer within Micro Focus to lead Data Governance.
* Identify Micro Focus data owners and assigned roles to route all the errors (in the creation of master data from other applications) to appropriate business data stewards / business operations users.

# MDM Scope Boundary

## 

## In Scope

The strategy will cover both Reference and Master Data and how each will be managed, as well as considering the implications for other applications in use within the wider IT estate, both in terms of constraints and opportunities.

The Master Data domains and Stack-C systems of record in scope for MDM tool are

|  |  |
| --- | --- |
| Domain | System of record |
| Product | SalesForce Configure Price Quote (CPQ) |
| Customer / Account | SalesForce (SFDC) |
| Finance (Class, Department, Legal Entity, Location, Critical Mass Territory, Subsidiary & Chart of Accounts) | NetSuite |
| Employee | Workday |

As part of data migration, each MDM domain data model will be created to map from migrated Stack-B & Stack-A to Canonical Data Model (CDM) which in turn mapped to Stack-C applications.

## Out of Scope

This document will not describe the following

* Rationalisation of Business Process Capabilities
* Hardware selection and sizing
* Enterprise Architecture Strategy
* MDM Tool selection & any point of view
* Operating systems
* Database system selection
* Security and network components for MDM environments without functional impact (Firewalls / VPN, CITRIX etc.)
* Enterprise Data Warehouse (EDW) / Business Intelligence (BI) Strategy

# Strategic Principles

The proposed Master Data process / solution / platform for Stack-C will need to comply with strategic guiding principles defined for the Stack-C program. These principles are as below

|  |  |
| --- | --- |
| Principle | Description |
| Fit for Purpose | Robust and compliant, but not necessarily best in class / best of breed  MDM solution will faithfully Master the Data without interrupting BAU operations when taking on-board future Master Data.  The solution will be easily and quickly scalable to support the Feb-2020 and Nov-2020 Go Live dates. |
| Fit for Standard | Out of the box, not customised  Change the business as required to align to standard industry practices (unless key market differentiator) |
| Two-level architecture | Functional domains with clear accountability, incorporating applications as building blocks to support end-to-end processes |
| Single master product  per functional domain | Develop a strategic partnership with a few (4) ‘master product’ vendors  Master product is the reference for processes and data within the domain  All other applications within the domain, are - ‘secondary host’ to the ‘master’ product |
| Strategic vendor  eco-system | ‘Master’ and ’ ‘secondary host’ applications or add-ons are selected within the strategic vendor’s ecosystem  The strategic vendor is responsible for technology aggregation, and ensuring end-to-end value |
| Prioritise ecosystem  integration | Preference for the integration of - ‘secondary host’ applications with the master product is through OOB connectors  Integration effectiveness is prioritised over functional capabilities |
| Optimise integration  patterns | Where there is no ecosystem integration, an ESB or ETL alternative can be used |

# Master and Reference Data

In the following sections, the definitions of master & Reference Data for Stack-C program will be set out which will be used in later sections of this document as well as in subsequent design & delivery details.

## Reference Data

Reference Data defines the set of permissible values used by other data fields. Reference data gain in value when they are widely re-used and widely referenced. Typically, they do not change overly much in terms of definition, apart from occasional revisions.

Examples of reference data include (but not limited to)

* Units of measurement (UOM)
* Country codes
* Corporate codes
* Fixed conversion rates e.g., weight, temperature, and length
* Calendar structure and constraints

It is therefore critical to control the Reference Data to ensure quality and consistency through the environments. Reference Data is fundamental to the operation of the system, Master Data set up and Data Quality.

Micro Focus MDM platform will be feeding Master and Reference Data to multiple target systems, which will consume that in different formats as constrained by that application.

E.g., Country code in NetSuite is 50 Char while SalesForce may configure & maintain ISO country list (2-Char). MDM typically masters the standardised list of Reference Data & maintain key references / mappings against target systems. Depending upon target systems, MDM can send relevant information to the target systems. Reference Data in MDM will be as below.

|  |  |  |
| --- | --- | --- |
| MDM Value – (ISO – 3 CHAR Country Code) | System Name | System Value |
| USA | NetSuite | United States |
| USA | SalesForce | US |

## Master Data

## 

**Master data** represents the business objects that contain the most valuable, agreed upon information shared across an organization. It can cover relatively static Reference Data, Transactional, Unstructured, Analytical, Hierarchical, and Metadata. Master Data is usually non-transactional in nature (but not in all the cases)

|  |  |
| --- | --- |
| Consideration | Description |
| Behaviour | Master data can be described by the way that it interacts with other data. In an operational system, master data is usually involved with transactional data. This relationship between master data and transactional data is viewed as a noun/verb relationship.  Transactional data capture the verbs, such as sale, delivery, purchase, email, and revocation; master data are the nouns. |
| Life Cycle | Master data is described by the way that it is created, read, updated, deleted, and searched. This life cycle is called the CRUD cycle and is different for different master-data element types and companies |
| Cardinality | As cardinality (the number of elements in a set) decreases, the likelihood of an element being treated as a master-data element |
| Lifetime | Master Data tends to be less volatile than transactional data. As it becomes more volatile, it typically is considered more transactional. |
| Value | Each attribute added to MDM carries a significant IT and business cost. This includes the initial costs of integrating, cleansing, validating, matching and synchronising the attribute and the on-going maintenance, profiling and stewardship thereafter. Therefore, all attributes within MDM should bring high business value. |
| Volatility | Master Data & related attributes are less volatile than transactional data and as a result, it changes less frequently. |
| Reuse | One of the primary drivers of MDM is reuse. If a Master Data (entity or attribute) reused in multiple systems, MDM system needs to manage it. More the systems uses it, the higher the business case to master it. |

## External Data

Certain data will be within the control of third parties (E.g., global lookup data from Dun & Bradstreet (D&B) for legal entity identification & customer corporate hierarchy information). Failure to ensure that there is an accurate and timely update to the representation of that data in Micro Focus systems will be crucial to the effective execution of business processes. Thorough analysis is required when devising business processes & managing underlying data

Micro Focus business & data operations team uses D&B Reference Data to identify legal entity & create subsequent sales hierarchy. Informatica MDM tool provides standard adapters to connect & refer D&B data using real-time & batch integration process. However, this integration has significant cost implications in terms of license fees, integration cost, per record reference cost etc.

# Master Data Quality

Data quality management is one of the key components and enablers of Micro Focus MDM strategy.

## Data Profiling & Current Data Quality Status

Perform Data Profiling to analyse the quality of data in legacy sources as a pre-requisite for Data Cleansing. Perform Data Profiling and Data Cleansing iteratively; prior to data being extracted, transformed and loaded into Target Systems.

Data Profiling will enable:

* The assessment of data quality, as a basis for the execution of Data Cleansing activities
* The monitoring of data quality improvement due to ongoing Data Cleansing activities for the duration of the project
* Validate data quality readiness at progress gates.

### Product Data

Based on initial conversations with Micro Focus business SMEs & data operations teams,

* Product data is most accurate & complete.
* For Stack-A PPA, the current master source for product data, holds the majority of Product Master Data.
* For Stack-B, the current master source for product data is JDE, Oracle & Pivotal.

### Customer / Account Data

Customer / Account Master Data has issues regarding duplication. This is primarily because it is managed in multiple systems, lack of common definitions across Micro Focus, absence of rules & validation during data management process etc.

### Finance Data

No major data quality issues reported for financial entity data. However, current Data Quality maturity levels can be identified once data profiling results are available.

### Employee Data

Employee data has recently been migrated to Workday. Hence, it is assumed that all data quality issues have been identified & addressed during this migration.

## Data Quality Dimensions

The below data quality dimensions will be used for assessing the quality of data at Micro Focus

* **Accuracy**: Determines the extent to which data objects correctly reflect Micro Focus business for which they were designed such as the address of a customer, Bill to, Ship to etc.
* **Completeness**: Determines the extent to which data is not missing. E.g., an order is not completed without a price and quantity.
* **Conformity**: Determines the extent to which data conforms to a specified format. E.g., the order date must be in the format YYYY/MM/DD.
* **Consistency**: Determines uniformity amongst underlying data object.
* **Integrity**: Determines the extent to which data is not missing important relationship linkages.
* **Timeliness**: Determines the extent to which data is sufficiently up-to-date for the task at hand.
* **Uniqueness**: Determines the extent to which the data for a set of columns is not repeated.

## Preventive Data Quality Procedures

This is also termed as Data Quality at source & applied for the data in motion. MDM process will be centralised in the business applications like SalesForce, NetSuite, and Workday etc. or Informatica MDM after Stack-C is operational. Having a single system as “system of record” per domain will address the majority of Data Quality issues like consistency, integrity, uniqueness etc.

To ensure data quality at these “system of record” applications, it is recommended

* To design data management process & configure user interfaces aligned with the process
* To include data providers (Business Operations Team) & data processors (Data Operations Team) in a decision to establish what information is feasible to collect, what information can be derived / defaulted based on configurable rules.
* To develop & document instructions (SOPs) for data processors (Data Operations Team) to execute the Master Data process.

## Reactive Data Quality Methodology

This methodology is applied for the data at rest. It involves continuously monitoring the quality of critical data elements using Informatica Data Quality (IDQ) for the Micro Focus Stack-C program. It may not be always possible to apply Data Quality of source data because of multiple reasons like increased complexity in a business application, performance issues causing dissatisfaction for Data Operations & Business Operations teams, availability of resources and identification of Data Quality requirements etc. Reactive Data Quality includes the development of Data Quality reports based on dimensions defined above & define the actions / procedures to address issues identified. This methodology is regarded as a continuous improvement process.

# Master Data Management

## Current State of Master Data at Micro Focus

Currently, Micro Focus manages Master Data with varying degrees of rigour and success, both in terms of applications and data entities. There is no central hub for Master Data, but there is a de-facto master application for most entities – E.g. – PPA for Stack-A Product Master Data, LASA for Customer Account, Workday for Employee Data, OneStream / NetSuite for Financial Entity Data, with changes sent via interfaces to other applications that need the data.

This has generally worked well enough in terms of Master Data consistency between applications, but there have been specific instances where inconsistency between the business rules implemented have led to issues in the receiving system and subsequent knock-on effects to process. It is generally agreed that data ownership (for all types of data) is not centralised within the organisation, with very few data entities having a named person or people or data organisation who take responsibility for ensuring quality or for agreeing to the changes.

Micro Focus links data to the processes and / or acquired applications that created it. In most areas, the data exists to serve a single process and / or application and is ONLY valued by those concerned with that process and / or application. This has led to a significant amount of duplication, variation in data standards, data definitions and ambiguity in ownership.

There is an increasing need to make holistic decisions across Micro Focus. Decisions can no longer be taken in isolation to understand their systemic impact across the business needs. The fragmentation and silos of Master Data prevent it from being valued in the decision-making process.

Micro Focus treats Master Data as a valued asset and maturity levels concerning data ownership and governance varies per process and / or application.

A separate Data Architecture organisation is established to

* Define data standards which includes data definitions, how data is created / used / shared in business operational process, logical data model across Micro Focus
* Define which & where attributes are mastered in Stack-C with tools & process deployed in a complementary fashion in order to achieve effective Master Data Management throughout Micro Focus

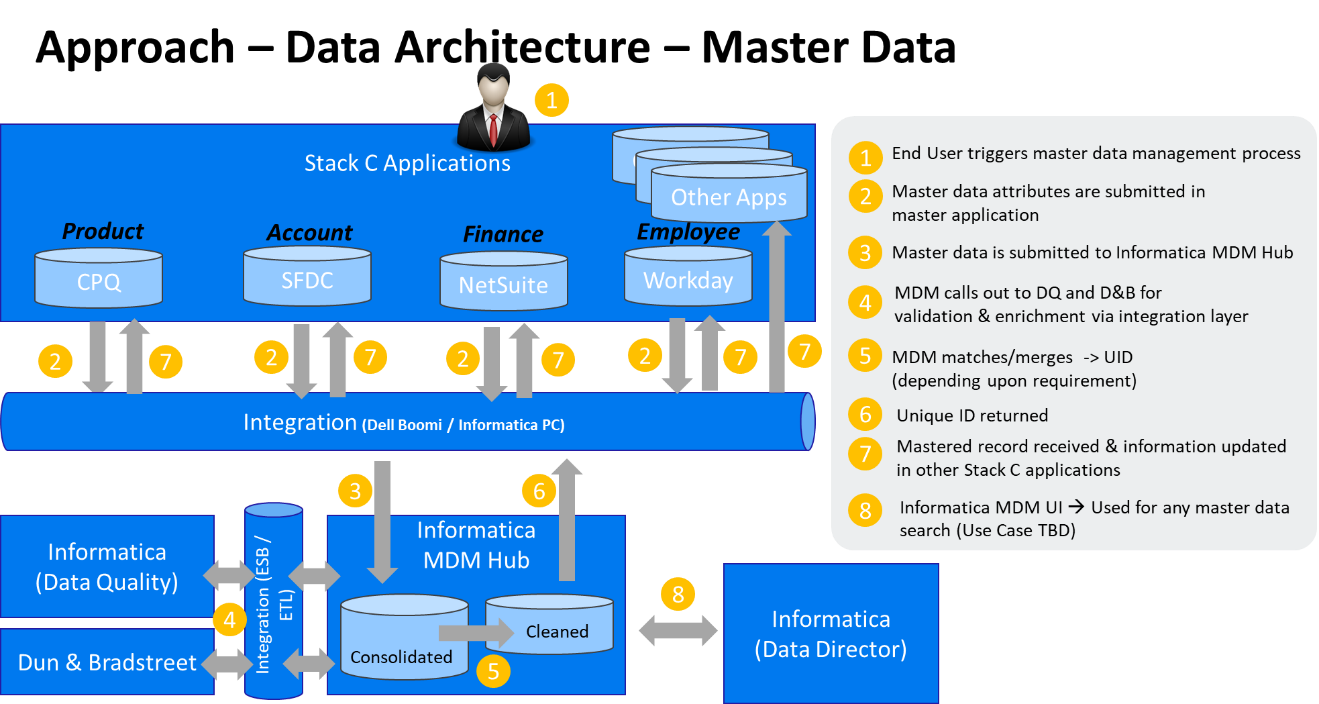
## Proposed MDM Implementation Style for Micro Focus

Key principle for Stack-C program is ‘Fit for Standard’ i.e. to use out of box contents with configurations. Our recommendation is to use similar business applications in Stack-C as the system of record for Master Data domains.

End users will be using these applications for Master Data interactions, bundle Master Data maintenance, Governance and Data Quality as a service (built on fit-for-purpose tools).

### MDM Domain Outline Data Architecture

We advocate the following hybrid approach for Master Data architecture at Micro Focus in a Stack-C world.



The rationale for parsing Master Data in Informatica MDM repository as a system of records, while not making it as a system of engagement is

* Type of applications like ERP, CRM, and CPQ etc. in Stack-C; are transaction oriented applications and have a complex database structure to hold Master Data. Accessing such Master Data from these applications in a large-scale enterprise like Micro Focus is not a good practice
* Canonical, simple models and integration architectures are recommended for Master Data maintenance & distribution will be available via Informatica MDM layer
* Single & central view of Master Data is available to support & accelerate data acquisition during M&A

Key highlights of proposed hybrid architecture for Master Data architecture at Micro Focus are as below

* Standards and Policies for Data Governance can be enforced when new Master Data records are requested and maintained in the business landscape
* The MDM layer provides a single view but does not own the centralization of the Master Data
* Every new Master Data request is checked at the point of entry for duplicates (applicable for customer account); Informatica MDM solution then identifies matching records. User will be presented with two options

a) To use existing record found by MDM using matching rules and option

b) To create a new record in case of low confidence during matching by MDM

In either case, we do not recommend using hard error & discontinue the Master Data process after record matching process.

* Key-mapping and linkage information is available at the Informatica MDM layer.
* Deploy data standards during data entry and no customisation or bespoke development is required at master application, which aligns with strategic / guiding principle.

In summary, we recommend the usage of a fit-for-purpose Master Data tool for managing Master Data, rather than a traditional approach for managing Master Data centrally using the Informatica MDM solution. Use Informatica MDM layer for De-Duplication, Data Validation, Enrichment, and Distribution purpose.

## Stack-C MDM Domain Management

### Product MDM Domain Management

For the initial load from source (Stack-A / Stack-B) to Stack-C

* Product data will be loaded & mastered into SalesForce CPQ and then pushed to MDM

For BAU

* SalesForce CPQ will be used for Product Data.
* All product data changes will be pushed to MDM and published to make them available to downstream systems.
* SalesForce CPQ will be the primary system for product data however Software Entitlement (SE) & other downstream systems may own & master a few of the attributes locally which will flow into MDM and be available to downstream systems as needed.

### Customer / Account MDM Domain Management

For the initial load from source (Stack-A / Stack-B) to Stack-C

* The Account data will be loaded from the Data Migration Engine (DME) into MDM
* Then a bulk match and merge process will be executed in MDM
* Post this, the GUID & GRID will be pushed back to the DME
* Account data will be pushed from DME to SalesForce & NetSuite
* The SalesForce & NetSuite IDs are synced to MDM

For BAU

* SalesForce will be used for Customer / Account Data
* On entry of Customer data in SalesForce, a SalesForce search will try and match to existing Accounts or add a new Account in SalesForce
* The latter will be routed for the attention of the appropriate data steward for review and possible merging of unverified accounts
* MDM will publish Account changes to make it available for downstream systems.
* Any updates to account attributes in the scope of MDM will be synchronised to MDM from SalesForce via Dell Boomi integration and vice versa

### Manage duplicate Customer / Account & Merge

The approach to the Customer Account merge process is as follows

* The Sales Rep. creates an unverified account in SalesForce
* The duplicate check takes place in SalesForce using SalesForce matching algorithms
* If there is a duplicate, the Sales Rep. must use the account already in "verified" state to create an opportunity
* If there is no duplicate, then the Sales Rep. will submit the account to Account business ops.
* The account is also sent in MDM in "unverified" state
* MDM will create a task for every unverified account submitted by SalesForce and send back a tracking number (task ID) to SalesForce
* Account Business Ops can verify the account in MDM. If they approve in MDM, at the same time, Account Business Ops enrich the account with D&B data in MDM, including full lineage in MDM up to global ultimate
* Account Business Ops submit only three levels of that hierarchy (Global Ultimate, Domestic Ultimate, and the legal entity we are transacting with) to SalesForce as Accounts for planning / compensation purposes. Only the legal entity we are transacting with will be in a verified state
* Account Business Ops do any additional Restricted Party List (RPL) and credit checks in MDM
* Once the account is fully approved, SalesForce will be notified by MDM for all 3 Accounts, and SalesForce will mark the requested account (the legal entity we are transacting with) as verified
* MDM will also send the approved account information to downstream applications with the exception of NetSuite where the account is required only once the order is generated in SalesForce
* If the requested account is rejected by Business Ops, the Account will be marked as inactive, and a "rejection" flag sent to SalesForce
* If an account is merged in SalesForce, it is always the new account to be merged with the old account (Master Account in SalesForce). If a new opportunity is created against the new account, it will also be mapped to the old account. But since an opportunity is managed within SalesForce, no downstream impact is expected

### Data Governance Reviews

In addition to the BAU checks, Data Governance processes will require periodic reviews of potential duplicates of VERIFIED SalesForce accounts that exist in MDM. These will be validated by running match rules against the records in MDM and generating a list of potential duplicates that will be checked by the MDM Data Stewards. Any actual duplicates will be resolved by means of a (to be defined) cross-functional and cross-application Standard Operating Procedure (SOP) that initially will be an end-to-end manual process.

### Finance MDM Domain Management

For the initial load from source (Stack-A / Stack-B) to Stack-C

* The Finance data will be loaded into Informatica MDM and NetSuite

For BAU

* NetSuite will be the master for Finance MDM domain data entities plus some of the financial Reference Data.
* NetSuite will push all changes for Reference Data to MDM. MDM will publish finance / Reference Data changes to make it available to downstream systems.
* For clarity, the product hierarchy, which is of interest to FinOps, will be loaded and maintained in Informatica MDM.

### Employee MDM Domain Management -

Whilst EMPLOYEE was in scope at the start of the design phase, further discussions with the Workday team and concerns around data privacy (GDPR) have resulted in the decision to keep Employee data within the Workday application. Stack-C applications, which require Employee data, will be fed the appropriate changes via the integration layer (Dell Boomi).

### Reference Data Management

The two patterns are:

* MDM is a secondary host i.e. Reference Data created in a Stack-C application and pushed to MDM for mastering and consumption by downstream applications
* MDM is a primary i.e. Reference Data required by more than one Stack-C application, created in MDM for onward distribution to downstream applications

### Other Master Data Domain Management - not covered above

There is no one-size-fit-all solution and the decision on where and how to manage a Master Data object will need to be determined only after a careful analysis of the object itself, its source and intended destination(s), desired maintenance procedures, potential enrichment requirements, current data quality, data sensitivity etc.

Key characteristics of business applications are:

* Out-of-the-box, full support for standard data structures necessitates little if any additional data modelling effort for standard master objects
* Data maintenance and validation is provided through built-in maintenance transactions
* Weak out-of-the-box data quality/de-duplication/enrichment functionality; expensive custom development required.

Given characteristics of business applications (SalesForce, NetSuite, Workday, and SalesForce CPQ etc.), these applications are best suited for

* Management of Master & Reference Data which are consumed by only these businesses applications & not shared across Micro Focus
* Maintenance & validation of Master & Reference Data using out-of-the-box functionalities

## Data Stewardship Agency

As part of the Micro Focus Data Governance, initiative led by Chief Data Officer and Data Architecture Group a BAU global data stewardship agency will be setup.

It will be consulted on Data Governance and will have a different role in Master Data Management solution. It will operate autonomously on maintaining Master Data at a global level, and staffed from employees and partners from within the business in order to locally co-ordinate stewardship of regional Master Data.

Concisely, the Micro Focus employee(s) / partner(s) who are nearest the location of a Customer / Account business relationship are best placed to resolve “Know Your Customer” (KYC) efforts and importantly adjudicate any low confidence Master Data matches that have been routed by the Informatica MDM solution to data stewards.

One of the main advantages of working autonomously at a local / regional level for stewarding Master Data is the less time to market for fixing Master Data issues, and therefore maximising the speed in which Micro Focus / partners can for instance quote a product price to prospects.

In addition to Customer & Product, proposed agency will also assist in the stewardship of the other in-scope Micro Focus Master Data domains (Finance & Employee).

Data Stewardship Agency will also be responsible for Reference Data definitions & Reference Data values along with Reference Data Management process.

Product Master Data will be the focus of the existing Global Data Operations Team.

Attached link to the slides are the first draft of those general principals, roles and responsibilities.

These slides are still work-in-progress so in case updated this is the [link to the SharePoint folder](https://microfocusinternational.sharepoint.com/:f:/t/StackC-Programme/Eju4WkVXJEZLrE86IjE3JDABD40fQXatHz5aATUP6NcD-A?e=Lnq968).

# Master Data Governance

In practical terms, the Data Governance means putting personnel, policies, procedures, and organisational structures in place to make data accurate, consistent, secure, and available to accomplish Micro Focus’ mission

As discussed above, additionally a BAU Data Stewardship Agency will be set up with a global organisation structure, and the named data stewards will be recorded in the Informatica MDM solution

Sound data governance program will allow Micro Focus to achieve the following objectives

* Enable better decision-making
* Reduce cost and increase overall operational efficiency
* Ensure transparency of data and processes
* Increase the potential revenue
* Establish standard and repeatable processes
* Educate staff to adopt a common approach to a data issue
* Create a more data-aware culture
* Consistency of data quality
* Helps enforce compliance with regulatory requirements

Data Governance steered by Data Architecture stream is integral to Micro Focus’ ability to improve their overall Data Quality, comply with an ever-increasing regulatory / legislation requirement, and ultimately identify business opportunities to exploit the data.

In addition, a robust Data Governance program with well-defined business process and an established formal training program can remediate some of the following challenges that have been identified by Micro Focus

* Lack of adherence to data standards
* Lack of understanding for critical standardised attributes.
* Identification of common attributes and their utilisation across Micro Focus systems.

To have a sound Data Governance program it will be critical to establish a formal Data Governance training program. Following is an example of requirements of a sound Data Governance training program

* Executive support to drive a data-aware culture.
* Well documented governance policies and procedures.
* Clear definitions of roles and responsibilities.

Based on initial conversations with multiple business & data operations teams at Micro Focus, following are key areas of concerns which needs to be addressed as part of Stack-C program

* No consistent & standardised data management process across the organisation due to multiple systems for Master Data
* No standard & accepted Master Data and attribute definition for Master Data domains
* Initial stages in formalising the Data Governance group by institutionalising Data Architecture Group. However, roles and responsibilities for data management (e.g. ownership, stewardship, etc.) should be fully identified during the functional design phase
* MDM is one of the critical success factors for Stack-C program but there is a lack of clarity on Data Quality monitoring in BAU phase. Reactive Data Quality monitoring using Data Quality reports is highly recommended

Micro Focus has taken a few steps towards achieving good governance of its data assets. However, it is in a nascent state. A robust Data Governance foundation needs to be built through a structured approach to bring it to a level that can sustain continuous growth and address future challenges. Micro Focus must demonstrate a firm commitment towards its Data Governance program with a dedicated org structure via Data Architecture group and well-defined metrics. This is critical to the success of Stack-C program.

# Required Solution Capabilities

Within the Stack-C Program, there are some key capabilities (I.e. Governance, Stewardship, and Data Quality etc.), that will need to be met in order to ensure good quality Master and Reference Data. Different types of data objects may utilise one or more of these capabilities – this will need to be defined for each data object during High-Level Design.

## Data Validation upon Creation

The application of validation rules upon creation will ensure that mandatory fields are populated and the invalid combinations of data are prevented. It is envisaged that this capability will be required in the system of record for all master objects (current scope – Product, Customer, Finance Entities & Employee)

## De-duplication

Checks that duplicate records are either prevented or flagged may be applied either prior to creation or prior to use. Communal area for this type of functionality at Micro Focus will be for Customer Account records.

## Approval

Some Master Data is so important to Micro Focus, that approval prior to use will be required. A single person may approve some objects while multiple users may require others. An example of this may be for a new financial entity (Profit Center, Subsidiary). The process will allow for both data quality checks and segregation of duties.

Additionally, where a low confidence Master Data match is deduced by Informatica MDM solution as it is entered via the user interface (UI) of a Stack-C system, then these are automatically routed in real-time to the assigned data steward for resolution.

## Data Enrichment

Some Master Data objects may require enrichment after creation. This may be appropriate where a skeleton record is acceptable upon creation with additional information added after the event, either manually by specialist users or automatically by the system (e.g. address completion or geotagging, D&B Lookup data). The GUID (Global Unique Identifier) and the GRID (Pointer to the current master in Informatica MDM for a given MDM domain) are always returned from the Informatica MDM solution for a unique record and can be used to enrich & link new Master Data created in a Stack-C system.

## Audit Trail

For many Master Data objects, it will be necessary to keep an audit trail of changes made.

## Post Creation Quality Monitoring

For some Master Data objects, pro-active validation may not be essential, but post creation validation and reporting will be necessary. Where this is the case, processes will be required to ensure that checks are complete and quality issues actioned in a timely manner. Capabilities in this area may include the provision of workflows and / or data quality scorecards.

## Master Data Distribution

Where multiple systems can consume Master Data for transactions, the ability to receive updates to that data and to distribute to consumers will be required. Wipro recommends using Informatica MDM solution as the central hub for data distribution in a canonical format & target system are to consume the required information via the integration layer (Dell Boomi).

## Semantic Catalogue

A common language, with misunderstandings and the resultant inefficiencies as to what certain terms mean can often divide an organisation. E.g., Item Pool in JDE, Product Group in SAP are available for the same functional usage but semantics used is different. Data Catalogue will help to group such terms together & make it available in a single repository. This catalogue will also help data mapping during data migration.

With third party software used with different terms used for the same thing, or the same term used for different things worsens this issue. Having a central catalogue that defines what terms mean helps to reduce these misunderstandings. Informatica Enterprise Data Catalogue (EDC) can be further explored for detail requirements at Micro Focus.

# High-Level Business Process Flow – Mastered & Non-Mastered Entities

## High-Level Process Flow – Initial Data Load (Day 1) – Mastered Data Entities

### 

### Data Migration Approach

Initial Master Data load to MDM hub is termed as Day 1 for Stack-C program, which will comprise of following key activities

* **Data Profiling** – The process of analysing the legacy data to align to data standards and relevancy rules that require to be fixed via Data Cleansing
* **Data Cleansing** – The process of cleansing data in Legacy Systems that will be migrated to target systems to ensure the data is valid, accurate, up-to-date and fit for business purposes
* **Data Mapping** – The process of mapping source to target attributes and defining code translation rules etc.
* **Data Extract** – The process of extracting data from the Legacy Systems
* **Data Transform** – The process of converting the extracted data from the Legacy Systems to be able to successfully load the data into the Target Systems. Master Data is created via the Informatica MDM solution, and the GUID and GRID returned are used to migrate the Master Data in the Data Management Engine (DME) as well as the transactions. The Informatica MDM solution will route any low confidence duplicates to the correct data steward responsible for that Master Data domain and where applicable geographic region.
* **Data Load** – The process of loading the extracted (and transformed) data into the Target systems.
* **Data Reconciliation and Validation** – The process of validating and reconciling the uploaded data in the Target Systems against data extracted from Legacy Systems.This includes pre-load and post-load validations steps.

Above steps may have variations depending upon approach chosen for each of the objects. In principle, the data migration team will get the data from all identified sources into a single canonical format before transformation & cleansing rules are executed. During MDM functional design discussions, different migration approaches will be discussed & optimal approach is selected based on pros & cons like the risk of data loss if any, timelines, resource availability, reuse of existing work assets, effort investment needed for fixing the source etc.

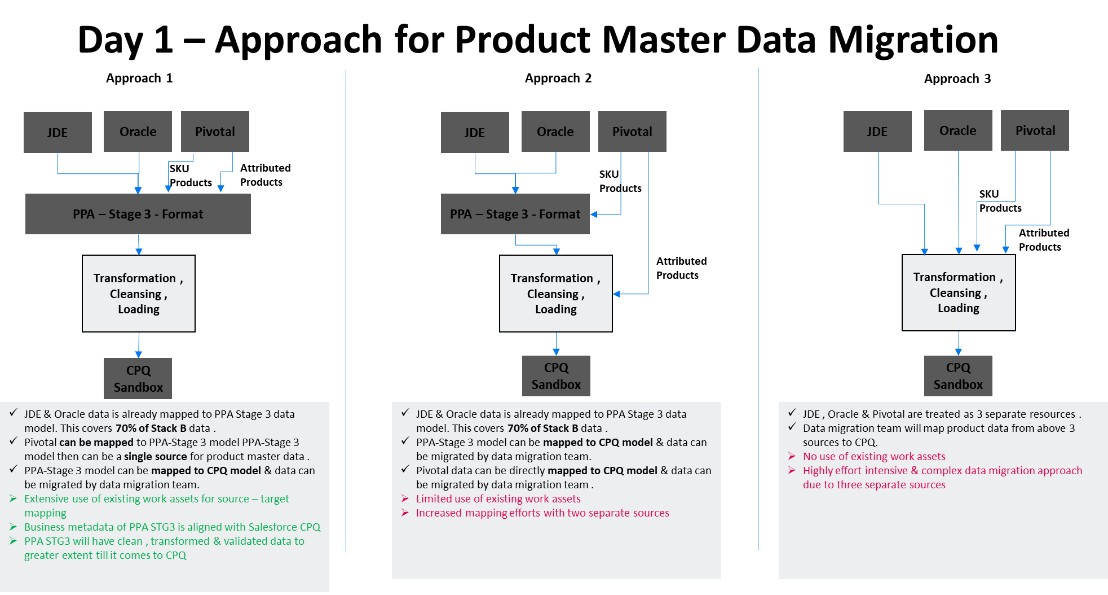
Following will be the Master Data sources identified so far that will provide the data for MDM domains. Please note that this list is based on the initial conversations done with the business & data operations teams, data migration team & is subject to change

|  |  |
| --- | --- |
| Master Data Domain | Data Sources |
| Product | Stack-B – JDE, Oracle, Pivotal  Stack-A – PPA |
| Customer Account | Stack-B – Pivotal  Stack-A – SalesForce  LASA – Customer Hierarchy Data |
| Finance Entities (Entity, Account , Department , Class , Location) | SUN / Manufactured Data by Finance |
| Employee | Workday – Both Stack-A & Stack-B |

### 

### Product Data Migration Approach - Example

This approach was used for the previous Product data migration & was discussed in detail with the business & data operations teams.



Each MDM domain will be considered on its own merits and a possible similar Day 1 data load process will be developed for each MDM domain in scope & will be included in respective functional solution design documents

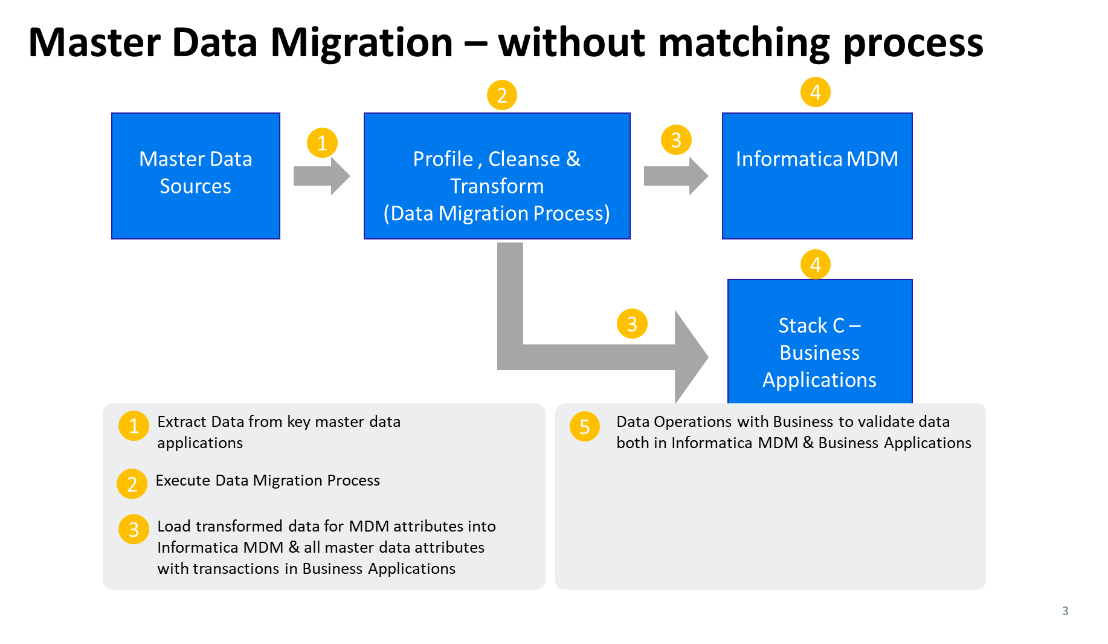
### Use of Informatica MDM Capabilities for De-Duplication

As mentioned in above business applications in Stack-C will continue to be systems of record & Master Data will be parsed into Informatica MDM repository. This involves two separate target systems for the data migration team.

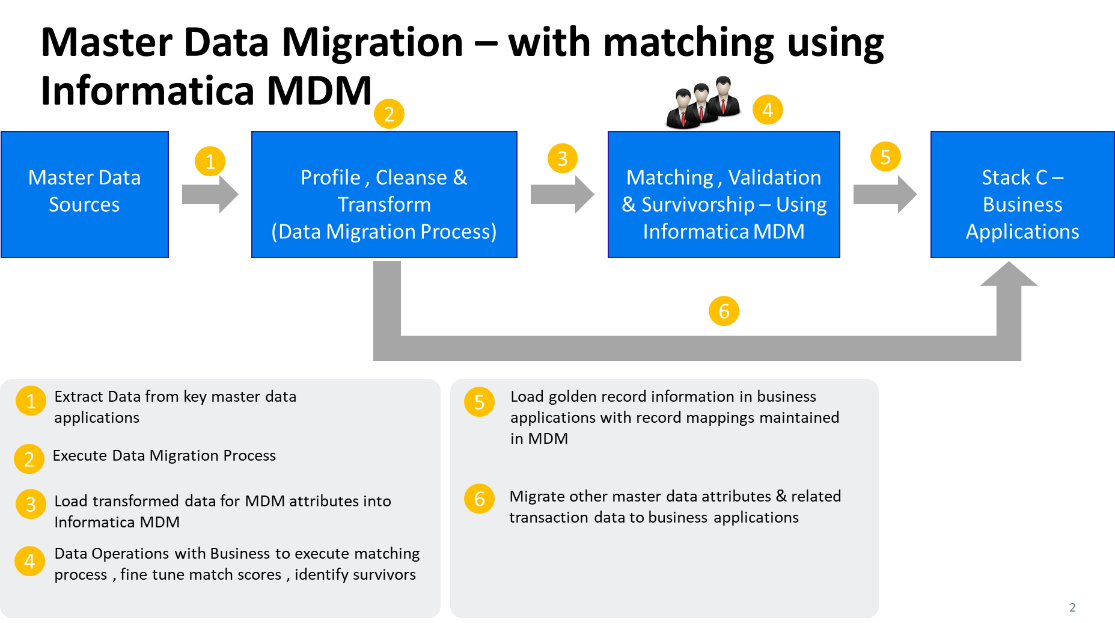
For the data migration, the mastering of Master Data domains (Products, Customer, Finance, and Employee) will be the responsibility of coordinated efforts of the Data Stewardship Agency plus the Data Architecture Team. They will both work with the Micro Focus business users (in SalesOps, FinOps etc.) to ensure

* The correct data attributes are captured and mastered and
* That the correct look-up / mapping tables are provided to the Data Migration Team

For BAU Day 2 operations the responsibility will fall to the Data Stewardship Agency



However, for Customer domain, it is also required to have matching process executed, duplicates identified & survivorship rules executed. This will ensure that clean, unique golden customer data is available for Stack-C Go-Live.



### 

### The first pass MDM data migration pattern is

* Bulk load into MDM from the data migration engine (DME) canonical model.
* Run bulk auto dedupe and match in MDM.
* Data stewards then can handle any low confidence MDM matches / potential duplicates, which have been routed to different MDM user groups for adjudication.
* Enrich the next downstream staged database of the DME by returning from MDM the Master Data GRID + GUID for respective parent and transactional data.
* Load unique data including the MDM GRID + GUID into target system Stack-C.

To be defined in the detailed design:

* If we run out of time for the data stewards to complete their manual matching / merging activity during the data migration, then the effect of flushing migrated data through into Stack-C would result in unique, matched, plus also any remaining potential MDM duplicates being loaded into Stack-C, the latter thus creating duplicates in Stack-C. Data stewards would then finish their manual matching + merge activity as a BAU activity, and any MDM merges would also require a merge in the Stack-C system e.g. merge of 2 customers in SalesForce.
* WThe second pass data migration e.g. Stack-A migration or IT system migration pattern from any interim M&A activity is as per the first pass MDM data migration pattern described above.

### Customer Life History of Data

The design pattern is **Source 🡪 MDM 🡪** **Stack-C System of Engagement application**

Data migration of Master Data to MDM & Business Application where matching capabilities of MDM are used, is further illustrated as below. Based on the current understanding, this process can be followed for Customer data migration.

This is an indicative flow, which will get refined further during MDM Functional & Data Migration design

**Step 1** – Data View in the source system



**Step 2** – Source data loaded in MDM & matching rules executed. Match score & match groups are created by MDM. Please note Match Scores are not available easily match Groups are created internally during Merge and are included here for information on what is going on behind the scenes in Informatica MDM.



**Step 3** – MDM matching results reviewed by data stewards & GUID (MDM ID) created. In case of match score lower than the threshold, separate GUID (MDM ID) is created. This is based on data steward’s decision.



**Step 4 –** MDM maintains key references based on De-Duplication exercise.



**Step 5** – MDM pushes unique records and GUIDs (MDM IDs) to SalesForce & SalesForce generates customer accounts (with SalesForce internal ID)



**Step 6** – MDM updates key mapping with SalesForce IDs



**Step 7** – Open transaction data for customer accounts is extracted from the source system. We have considered only open order value in this example for simplification.



**Step 8** – Data migration team migrates transaction data referring to key mapping in MDM for record ID references.



## High-Level Process Flow – Master Data Process (Day 2) – Mastered Data Entities

### 

### Overview

Business as usual (BAU) process is termed Day 2 for the Stack-C data migration program. Master Data Management (MDM) has five key business processes – Create, Read, Update, Delete / Archive & Mass Updates. There can be different variants for each of these processes. E.g., Create Bill-to customer is a variant of customer create process. Create SaaS Product is a variant of the product create process.

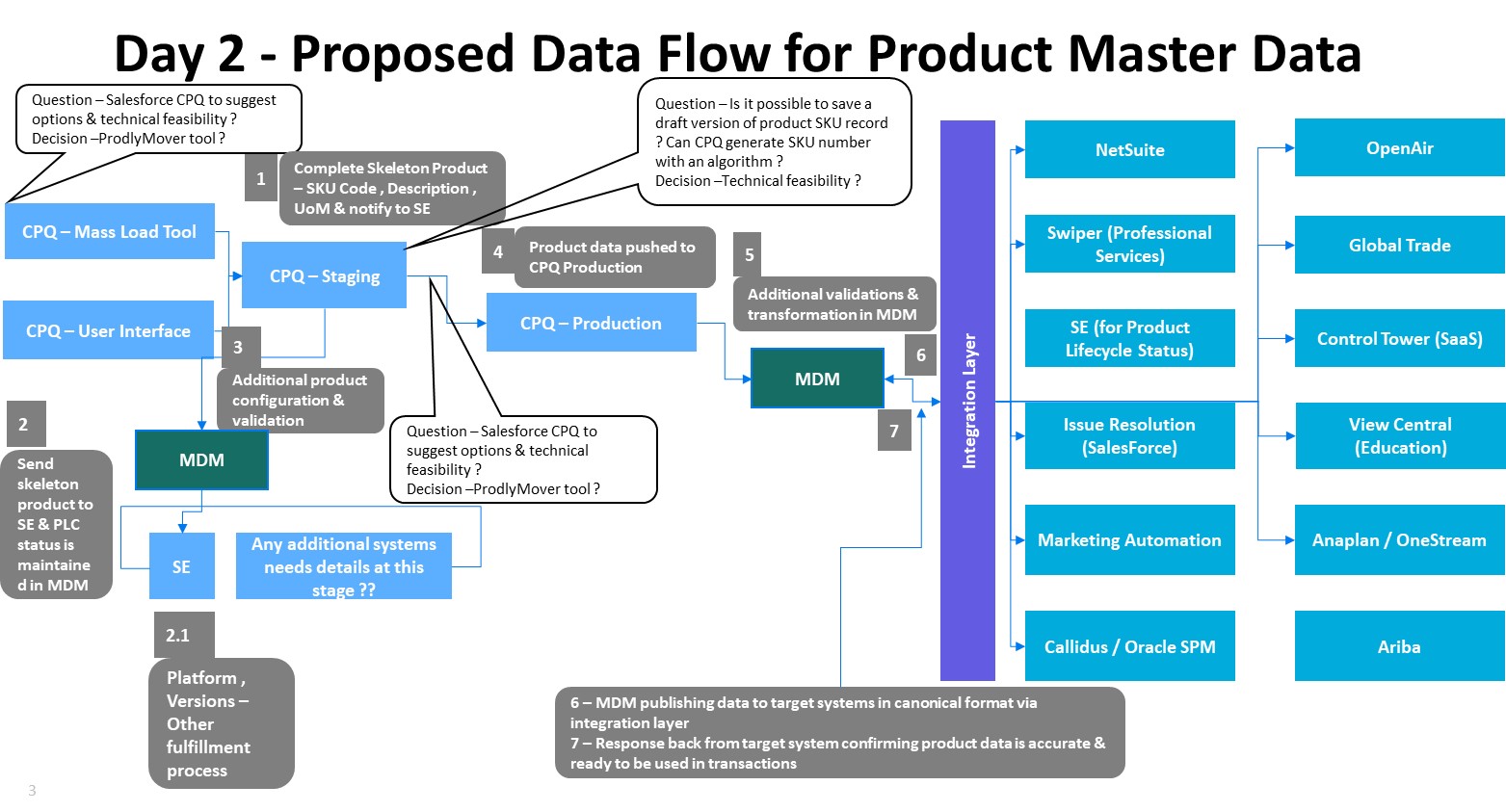
Business applications in Stack-C will continue to be the system of engagement & Master Data will be parsed into Informatica MDM repository. We recommend creating data management process flows as below during the functional design phase. This is to ensure that

* All Master Data variants for each domain are covered for Stack-C
* Variations (if any) in the data governance / workflow process can be identified
* Data quality rules & validations can be identified per variants
* The new record is compared against existing MDM records for duplicates. MDM will route new records to appropriate data stewards for any incidence of low confidence match scores. The assertion is that we will pursue duplicate prevention in all cases.

### Example Product Master Data – Day 2

Product Master Data is the most critical data element for the success of Stack-C program. Following diagram depicts the process flow. This flow will be used as a template to define Day 2 process for other data domains in scope. This process does not cover the activities done by the R&D Manager & his / her interactions with the business & data operations teams.

This is just a couple of steps that were created to know the interaction between CPQ and SE. However, this is not the full data flow/process flow for new products. Process team will be working with the product operations team to define the process flows for new products, new versions etc.



* The business team has a requirement to create product data using web-based interface (single record creation) and mass data load creation using SalesForce CPQ.
* Initial data gets created in a staging environment with a minimal set of attributes (termed as skeleton product) such as SKU number, description, unit of measurement, product type etc.
* Using a staging environment makes sure that product configurations are validated & product is ready for a quote when in production.
* Skeleton product gets pushed to Informatica MDM at this stage with specific product life cycle status (e.g. ‘Unreleased’).
* Informatica MDM will then publish via the integration layer the Product related information to the Software Entitlement (SE) system in order to update attributes such as platform, version & complete fulfilment process.
* This parallel processing will accelerate the SE process & be ready for fulfilment once product configuration, pricing set up is available in the CPQ – Production system.
* Data Operations will then complete any remaining product configurations & pricing set up, data validations if any. Once complete, the product is pushed to CPQ Production.
* CPQ Production sends the required information (a subset of product attributes) to Informatica MDM.
* Informatica MDM is a central hub, which then publishes the data on the integration layer to all target systems identified using the integration layer. It may also be required to configure response back from consuming systems to Informatica MDM in order to make sure that it has received the data and is ready for further processing.

This is high-level Day 2 Master Data process for the Product MDM domain. Details of all MDM domains process will be available in the functional solution design document.

## High-Level Process Flow – Initial Data Load (Day 1) – Non-Mastered Data Entities

Key requirements for on-boarding or mastering a data entity into Informatica MDM solution are Data Matching / De-Duplication, Data Distribution, Data Enrichment using third party Reference Data, governance using workflow & data distribution to multiple systems along with the considerations, like value, volatility, reuse, consensus & behaviour. While Informatica MDM solution has, technical capability of master all data entities identified my Micro Focus, it is critical to qualify them based on the criteria described.

This section describes an approach for non-mastered entities. E.g., Project - is a list of codes that relate to activities with a specific purpose, defined scope, resource profile, budget and duration. NetSuite primarily uses Project data entity. Following sections will describe the approach for data migration & BAU data management process for such non-mastered entities.

### Data Migration & Cleansing Approach

Initial Master Data load to target is termed as Day 1 for Stack-C program. Data migration team has defined the data migration process as a part of the data migration strategy. This process will comprise of following key activities

* **Data Profiling** – The process of analysing the legacy data to align to data standards and relevancy rules that require to be fixed via Data Cleansing
* **Data Cleansing** – The process of cleansing data in Legacy Systems that will be migrated to target systems to ensure the data is valid, accurate, up-to-date and fit for business purposes
* **Data Mapping** – The process of mapping source to target attributes and defining code translation rules etc.
* **Data Extract** – The process of extracting data from the Legacy Systems
* **Data Transform** – The process of converting the extracted data from the Legacy Systems to be able to successfully load the data into the Target Systems.
* **Data Load** – The process of loading the extracted (and transformed) data into the Target systems.
* **Data Reconciliation and Validation** – The process of validating and reconciling the uploaded data in the Target Systems against data extracted from Legacy Systems. This includes pre-load and post-load validations steps

## High-Level Process Flow – Master Data Process (Day 2) – Non-Mastered Data Entities

The data management process for Non-Mastered data entities will be executed in respective business applications. Business applications in Stack-C like SalesForce and SalesForce CPQ are the primary applications for these entities & will use standard out-of-the-box features & functionalities to manage (CRUD Process) Master Data. These applications offer

* Basic data quality validations like mandatory field checks
* Intuitive user interface
* Lookup configurations to ensure data quality
* Role-based access

It is important for the business applications to use the features available for non-mastered entities to make sure that Master Data is managed to maintain expected data quality during their individual application design process.

Master Data for Non-Mastered entities will flow through the integration layer along with transaction data (in batch / real-time mode).

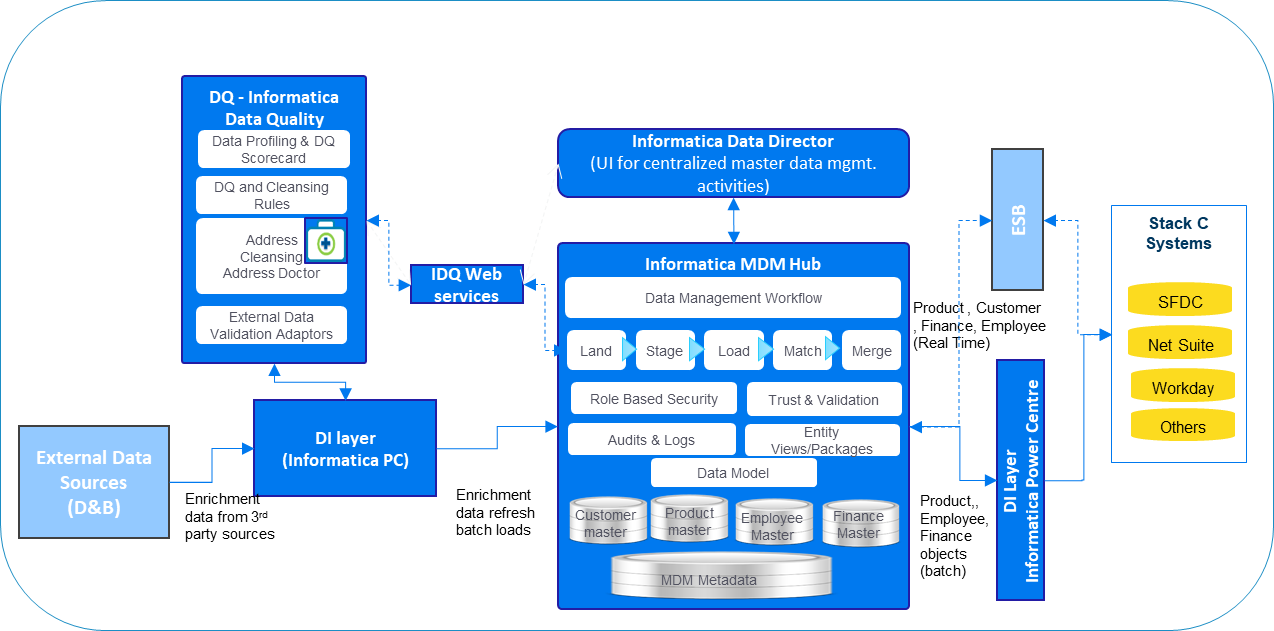
Data Architecture Group along with Data Stewardship Agency will monitor data quality in BAU mode.

## Summary - High-Level Process Flow – Master & Non-Master Entities

|  |  |  |
| --- | --- | --- |
|  | Mastered Entity (E.g. Customer Account) | Non-Mastered Entity(E.g. Project) |
| Initial Data Migration & Data Load (Day 1) | * Use the data migration engine to extract, transform & load Master Data into Informatica MDM * Matching & De-Duplication using Informatica MDM * Create key mapping in Informatica MDM * Data validation & reconciliation * Load Master Data into target business applications * Load transaction data referring key mapping into target business applications | * Use the data migration engine to extract, transform & load Master Data into Informatica MDM * Use basic data cleansing features from data migration engine * Data validation & reconciliation * Load Master Data & transaction data into target business applications |
| Business as Usual (Day 2) | * Initiate Master Data Management process in master business application (E.g. SalesForce) * Basic data quality checks like mandatory fields, lookup etc. done in the user * Validate if a record already exists in Informatica MDM * Data Stewards to take appropriate decisions for duplicate records & other exceptions * Master Data gets committed to the business application & Informatica MDM * Informatica MDM is used to further enrich & validate data * Master Data is shared across target systems from Informatica MDM via the integration layer & is consumed during transactions | * Initiate Master Data Management process in master business application (E.g. SalesForce) * Basic data quality checks like mandatory fields, lookup etc. done in the user * Data Stewards to take appropriate decisions for data quality exceptions * Master Data gets committed to the business application & is consumed during transactions * Master Data is shared across target systems via the integration layer in case of such requirements |

# Tool capabilities

The reference solution architecture for Informatica MDM solution is as below. This is high-level architecture and usage of various components will vary based on domain-specific functional solution design. The contextualised MDM reference architecture will be available in the individual functional design document for Customer, Product, Finance, and Employee data



Required solution capabilities are mapped to solution architecture components as below –

|  |  |
| --- | --- |
| Required Solution Capabilities | Solution Architecture Component Details |
| Data Validation upon Creation | The proposed solution on Informatica suite of products has an integrated framework that allows for data to be matched & validated as it enters the hub in all modalities (batch, near real time, or real-time). The assertion is that we will pursue duplicate prevention.  Batch processing is executed in the background as part of a data processing script.  Real-time matching & validation, on the other hand is done by putting a record on a message queue or by calling the Informatica’ s APIs through Web services or EJB interfaces. |
| De-duplication | The MDM Hub in the proposed solution has fully integrated the SSAName3 matching engine. This industry-leading match engine has been combined with MDM Hub’s powerful rules-based matching interface to perform both deterministic ('exact') matching, as well as probabilistic ('fuzzy') matching based on sensitivity levels and includes set-based matching for both.  Fuzzy matching utilises various matching algorithms based on the type of data being matched, with built-in algorithms for phonetic spellings or partial fields. The statistical distribution of probabilistic matches is handled through pre-built "populations" for individual countries. E.g., the USA population is aware of common nicknames such as "Bob" for "Robert". Furthermore, match columns of either or both types can be combined based on business requirements. Multiple match rules can then be created with sophisticated matching including null-Post Creation DQ Monitoring, matching, non-equal matching, segment matching and more. Informatica also allows fine-grained tuning of the ranges to search, tightness of match and other parameters for the balancing degree of matches vs. amount of processing (performance). |
| Approval | Proposed solution architecture offers Informatica Active VOS, which can be used to configure data review & approval process. |
| Data Enrichment | The proposed solution on the Informatica platform provides integration with multiple data services in multiple ways and modes. Including prebuilt adapters to major popular data cleansing and matching products.  For example - Integrating D&B data can be accomplished both through batch and through real-time interfaces.  In batch mode, D&B is considered a source and it is processed the same way as any other data. It is match and merged using the rules configured in the Hub. A customary practice is to create merge rules in the Trust Framework that score D&B higher than internal sources for attributes that D&B is likely to have better insight into.  For real-time integration, Informatica has a partner, who in conjunction with D&B has created a real-time adapter to dynamically enrich customer data with attributes from D&B. |
| Audit Trail | The proposed MDM Hub automatically creates and maintains various metadata tables to ensure the highest quality, reliability and audit capability for all data stored in the hub. This includes a complete history of raw, cleansed /merged data and any changes, manual or automatic, made to the data and its relationships. |
| Post-Creation DQ Monitoring | Informatica Data Director dashboard can be used for monitoring the data life cycle. Data stewards can use it to monitor data quality, and the default DQ chart displays the level of data quality for a particular individual, product, or organisation. |
| Master Data Distribution | ETL jobs configured in Informatica PC will extract relevant Master Data records exposed by MDM hub through packages and will push this data to target systems periodically.  For real-time data distribution, the consuming systems will be integrated with MDM hub for Master Data through SIF web services exposed by the MDM hub. These systems will use the integration layer to integrate with MDM solution. Informatica MDM hub exposes all base objects configured in its repository through web services to other systems. These fine-grained web services will have to be orchestrated to create composite web services for other systems usage |
| Semantic Catalogue | Informatica Enterprise Data Catalogue allows for easy import of business glossary assets such as terms, policies, and classifications from Informatica Axon™. Add rich business context to the data by associating business terms with the right technical metadata. Informatica Enterprise Data Catalogue will even recommend term associations. This allows business and IT stewards to manage business metadata that includes efficient human workflow automation.  Informatica Enterprise Data Catalogue also supports import of business glossary assets from Informatica Business Glossary and third-party tools |

# Assumptions

Following assumptions are working hypotheses for this document

* Throughout the document, a basic set of business requirements are assumed that are based on our current knowledge of Micro Focus and the knowledge of best practice in the software industry. As such, the requirements and suggestions made are based on best practice and the information available to us.
* New business processes driven through the detailed design of business applications like SalesForce CPQ, SalesForce, NetSuite, Workday could lead to enhancements and revisions of this document.
* It has been assumed that in the end all business applications (SalesForce CPQ, SalesForce, NetSuite, and Workday) will be implemented & will be part of Stack-C program.
* In the end, it is expected that Informatica suite of products (MDM, IDQ, EDC, and Power Center) will be used as a platform of choice for Data Management & Governance.

# Best Practices

|  |  |
| --- | --- |
| Solution Component | Best Practices |
| Conceptual Data Model | Make sure the match requirements are well thrashed out as they will drive the model and the major effort required towards a successful implementation  Keep the database at a 3NF form and don’t normalise it completely as it limits flexibility and functionality  During design, it is useful to create a mix of a logical & physical model that contains key attributes & related notes  Articulate hierarchy and relationship definition. They have a direct impact on Informatica MDM data model and configuration |
| Data Enrichment, Standardisation & Validation | Use 3rd party tool for data enrichment like Address Doctor for address enrichment, and standardisation  Standardise all Reference Data elements like country code, state code, phone numbers, salutation  Use Appropriate IDQ cleansing capabilities & reference tables in MDM to store “Standard” reference values  Implement robust validation rules in the application where data is generated like POS, Portal, etc.  Load only valid data in MDM. Data should be corrected in source & re-loaded in MDM |
| Data Matching | Plan adequate iterations for doing match rule tuning  Include exact match column in fuzzy match rule for higher performance. Limit trust configuration for important columns only  “BOT” can learn and take actions on such records by learning from actions taken by human data steward in past  Consider data model design between 2NF & 3NF; database hardware capacity, data volume must be considered during the design stage |
| Data Integration | Consider real-time data ingestion if a source provides data in real-time  Use entity framework for coarse grain MDM APIs and SIF for fine-grained APIs  Profile source data and define appropriate data quality rules  Evolution of MDM architecture from one style to another will require changes to the data integration patterns  Define an integration layer to handle unplanned scenarios / Data rejection |

# Appendix

## Data Domains Summary

* Data Entity ID: Cross-reference to the Data Migration Canonical Data Dictionary.
* Data Entity: The data entity being mastered.
* Definition: An explanation of the data entity.
* Master / Reference: Whether the data entity is deemed to be Master Data or Reference Data.
* System of Record: Where the data entity is created.
* Mastered In: Where the data entity is mastered.
* Held in MDM: Whether the data entity is held in Informatica MDM or not.
* Application for Data Management Process:
* Data Owner: The business team(s) responsible for managing and mastering the data entity.
* Initial Data Cleansing Approach: How we first ingest the data into Stack-C

| Data Entity ID | Data Entity | Definition | Master / Reference | System of Record | Mastered In | Held in MDM | Application for Data Management Process | Data Owner | Initial Data Cleansing Approach |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DE-05 | Address | The address associated with a customer, contact, employee, partners or vendor record. This would also include an address associated with Micro Focus’ own physical locations. | Master | SalesForce (Customer, Contact, Partner, Vendor), NetSuite (Contact, Vendor), Workday (Employee) | Informatica MDM | Yes | SalesForce (Customer, Contact, Partner, Vendor), NetSuite (Contact, Vendor), Workday (Employee) | SalesOps (Customer, Partner), HR Ops(Employee), FinOps / DataOperations (Vendor) | Data Migration Engine, Informatica MDM |
| DE-09 | Bank Account | Bank account information, be this for employee, vendor, customer, partner or for Micro Focus itself | Master | NetSuite | NetSuite | No | NetSuite | SalesOps (Customer, Partner), HR Ops(Employee), FinOps / DataOperations (Vendor) | Data Migration Engine |
| DE-10 | Billing Schedule | The periodic billing of customers based on pre-defined intervals for recurring invoices | Master | NetSuite | NetSuite | No | NetSuite | SalesOps | Data Migration Engine |
| DE-11 | Configuration Rules (CPQ) | Configure, price quote software is a term used to describe software systems that help Micro Focus quote complex and configurable products. | Master | SalesForce CPQ | SalesForce CPQ | No | SalesForce CPQ | Product Data Operations | Data Migration Engine / Manual using a spreadsheet |
| DE-12 | Campaign | A specific record that defines a series of activities used in marketing a product or service | Master | SalesForce | SalesForce | No | SalesForce | SalesOps | Data Migration Engine |
| DE-13 | Chart of Accounts | A complete listing of every account in Micro Focus’ accounting system(s) | Master | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-15 | Community Data | Community forums, blogs and questions that promote the sharing of information between customers, partners and Micro Focus. | 0 | SalesForce | SalesForce | No | SalesForce | SalesOps | Data Migration Engine |
| DE-16 | Contact | Any individual that Micro Focus has a business relationship with, including customer, employee, vendor or partner contacts. This could also include marketing contacts. | Master | SalesForce, NetSuite | System of record | No | SalesForce, NetSuite | SalesOps (Customer, Partner) , HR Ops(Employee) , FinOps / DataOperations (Vendor) | Data Migration Engine, Informatica MDM |
| DE-17 | Contract | A legal agreement between Micro Focus and a third party, be this a customer, vendor, employee or partner | Master | SalesForce, SalesForce CPQ | SalesForce, SalesForce CPQ | No | SalesForce, SalesForce CPQ | SalesOps | Data Migration Engine |
| DE-18 | Cost Centre | A code that denotes the area of the business to which a cost or revenue is to be apportioned/allocated | Master | NetSuite (may be held as a different concept) | NetSuite | No | NetSuite | FinOps | Data Migration Engine |
| DE-20 | Currency | The multiple valid legal currencies against which transactions in the various sales regions may be captured, captured in relation to a base currency | Reference | Informatica MDM | Informatica MDM (Account (Customer / Partner) , Finance) | Yes | Informatica MDM | FinOps | Data Migration Engine |
| DE-22 | Customer | An existing or prospective buyer of Micro Focus’ products and services, captured along with its hierarchy in relation to other customer records. | Master | SalesForce | Informatica MDM | Yes | SalesForce | SalesOps | Data Migration Engine, Informatica MDM |
| DE-24 | Customer Hierarchy | A D&B based hierarchy is a structure consisting of customers. Each customer - with the exception of the uppermost customer refers to another customer at a higher level in the hierarchy. Customers that are assigned to higher-level customers are known as dependent customers. | Master | SalesForce | Informatica MDM | Yes | SalesForce | SalesOps | Data Migration Engine, Informatica MDM |
| DE-27 | Department | A functional area within Micro Focus, intrinsically linked to a cost centre | Master | NetSuite | Informatica MDM | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-29 | Division / Sales team | A sub-area within the sales regions | Reference | SalesForce | SalesForce | Yes | SalesForce | SalesOps | Data Migration Engine without data quality checks. |
| DE-30 | Employee | An individual who works part-time, full-time or on a contingent basis under a contract of employment and has recognised rights and duties. | Master | Workday | Workday | No (Out of scope – 1st April 19) | Workday | HR Ops | Not Applicable since data is already migrated |
| DE-40 | Legal Entity | A sub-corporate division of Micro Focus that can legally enter into, and be held in compliance with, a contract | Master | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-42 | Location | A office or warehouse location for Micro Focus | Master | NetSuite | Informatica MDM | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-46 | Partner | A company that collaborates with Micro Focus to market sell or implement our products or services. This record is captured in a hierarchy in relation to other existing partner records. | Master | SalesForce | Informatica MDM | Yes | SalesForce | SalesOps | Data Migration Engine |
| DE-47 | Payment Terms | Payment terms as the terms required for payment on a product are a function of the service offering of a vendor. | Reference | NetSuite | NetSuite | No | NetSuite | FinOps | Data Migration Engine |
| DE-49 | Product | A catalogue of the offerings that Micro Focus take to market, including services. | Master | SalesForce CPQ | SalesForce CPQ | Yes | SalesForce CPQ | Product Data Operations | Data Migration Engine, SalesForce CPQ |
| DE-51 | Product Hierarchy | The linkages between the top down a grouping of products that drive portfolio profit reporting | Reference | Informatica MDM | Informatica MDM | Yes | Informatica MDM | Product Management | Manual using a spreadsheet (only Reference Data) |
| DE-52 | Project | A list of project codes that relate to activities with a specific purpose, defined scope, resource profile, budget and duration | Master | NetSuite | NetSuite | No | NetSuite | FinOps | Data Migration Engine |
| DE-63 | Sales Region-CMT | A set of codes that describe the geographic region and route to a market/channel which Micro Focus operate within, in regards to sales.  Critical Mass Territory – Sales GEO structure. | Reference | SalesForce | Informatica MDM (Customer, Product) | Yes | SalesForce | SalesOps | Data Migration Engine / Manual using a spreadsheet |
| DE-66 | Tax Assignment | "Tax Assignment designations understood to be used to determine indirect tax (e.g. VAT, or US Sales & Use tax) status / classification. However, this should be confirmed on an application-by-application basis. Similar functionality could be used in principle to aid tax classification for direct tax (e.g. corporate income tax) purposes in a future-state group. However, Tax is not aware of such tax sensitization being used at present." | Reference | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-68 | Vendor | An external company, individual or an employee (if paid through Accounts Payable) that provides goods or services to Micro Focus | Master | NetSuite | Informatica MDM | No | NetSuite | FinOps | Data Migration Engine |
| DE-70 | Vendor Item | A catalogue of items (physical item or service item) that may be purchased by Micro Focus from a vendor. | Master | NetSuite | NetSuite | No | NetSuite | FinOps | Data Migration Engine |
| DE-71 | Class | This represents summary product structure and the revenue type. | Reference | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-72 | Plan Rates | Rates used for plan rate calculation | Reference | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-74 | Region / States | Region / States | Reference | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-xx | Price Rate | Is conversion rate used to calculate local pricing? | Reference | Informatica MDM | Informatica MDM | Yes | Informatica MDM | Pricing Center of Excellence | Data Migration Engine / Manual using a spreadsheet |
| DE-xx | BMU | Business Mark-up used to calculate local price. | Reference | Informatica MDM | Informatica MDM | Yes | Informatica MDM | Pricing Center of Excellence | Data Migration Engine / Manual using a spreadsheet |
| DE-xx | Countries | ISO countries | Reference | Informatica MDM | Informatica MDM | Yes | SalesForce, NetSuite | SalesOps | Data Migration Engine |
| DE-xx | Fiscal Period | Corporate Reporting Financial period. | Reference | NetSuite | NetSuite | Yes | NetSuite | FinOps | Data Migration Engine / Manual using a spreadsheet |
| DE-xx | Partner Exhibit | Defines the product and discount structure of a partner within each country. | Master | SalesForce CPQ | SalesForce CPQ | Yes | SalesForce CPQ | PartnerOps | Data Migration Engine / Manual using a spreadsheet |
| DE-xx | Worldwide Reference Price | Worldwide price used as a basis to calculate local pricing in a multiplier-pricing model. | Reference | SalesForce CPQ | SalesForce CPQ | Yes | SalesForce CPQ | Product Management | Data Migration Engine / Manual using a spreadsheet |