

27 Aug, 2024

On Prem MDM to MDM SaaS Data Migration and Data Validation

- Abhisek Bhowmik, Principal Consultant, IPS
- Gaurav Misra, Principal Consultant, IPS

Where data & AI come to 

Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our [Success Portal](#) - where you can download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

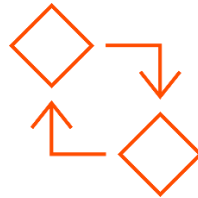
Feature Rich Success Portal



**Bootstrap trial and
POC Customers**



**Enriched Customer
Onboarding
experience**



**Product
Learning Paths
and Weekly
Expert Sessions**



**Informatica
Concierge**



**Tailored training
and content
recommendations**

More Information



Success Portal

<https://success.informatica.com>



Communities & Support

<https://network.informatica.com>



Documentation

<https://docs.informatica.com>



University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.



MDM Modernization

Data Migration & Data Validation (MDE C360 to SaaS C360)

Speaker: Gaurav Mishra & Abhisek Bhowmik , IPS

Where data & AI come to **LIFE**

Our Speaker



Abhisek Bhowmik

Principal Consultant, IPS



Gaurav Mishra

Principal Consultant, IPS

Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.

Agenda

1 Complexities

2 Data Migration Approach

3 XREF Only / BO & XREF / Rematching

4 Initial Data Load – Seeder / Non-Seeder

5 Cross Reference Record

6 Survivorship

7 Data Validation Approach

8 High Level Design

9 Dependencies

10 Benefits

Complexities



The complexity of MDM data migration & validation is largely because MDM systems typically contain large volumes of critical business data that must be accurately transferred and mapped to the new system



The conversion from the relational model to a non-relational model necessitates design changes in the implemented data model. Because in the SaaS world focusing on the business model to represent the business entity.



The process of data migration & validation can be time-consuming, demanding careful planning, execution, and validation to guarantee the accuracy, completeness, and security of the data

Data Migration

Data Migration Approach

1

XREF only

- Less complexity
- Cannot guarantee survivorship for any table/entity will mimic on-prem

- Make Row ID as Business ID
- Load By Row ID

2

BO & XREF

- Guaranteed Golden Record consistency
- Licensing concern, extra unconsolidated records

- Make Row ID as Business ID or Generate New Business ID for the record coming from BO
- Load BO record using highest trust system and decay period 2 days
- Load all XREF by Business ID
- One extra record will be created

3

Rematching

- No usage of Load by ID
- Cannot guarantee survivorship for any table/entity will mimic on-prem

- Load all XREF record
- Generate Business ID
- Execute Match & Merge

XREF Only

Option 1a

Identify Seeder Record
ROWID_OBJECT =
ORIG_ROWID_OBJECT

Migrate Seeder
Record from
XREF

Mapped ROWID_OBJECT
is Business ID
& SRC_LUD to Source
Last Update

Register the
Golden Record

Identify Non-Seeder Record
ROWID_OBJECT <>
ORIG_ROWID_OBJECT

Migrate Non-
Seeder Record
from XREF

Load By ROWID & Map
Source Last Update

Update the
Golden Record

Recalculating the Golden
Record

Option 1b

Identify Seeder Record
ROWID_OBJECT =
ORIG_ROWID_OBJECT

Migrate Seeder
Record from
XREF

SRC_LUD to Source Last
Update, Store
ROWID_OBJECT as an
Legacy MDM ID in SaaS

Register the
Golden Record

Generate New Business ID

Extract Business
ID & ROWID

Identify Non-Seeder Record
ROWID_OBJECT <>
ORIG_ROWID_OBJECT

Migrate Non-
Seeder Record
from XREF

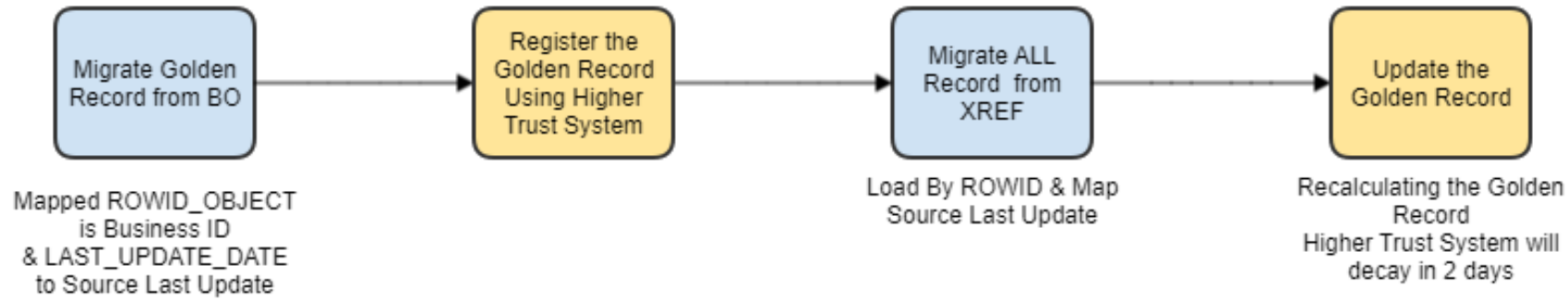
Identify the Business ID
against the
ROWID_OBJECT,
Load By Business ID &
Map Source Last Update

Update the
Golden Record

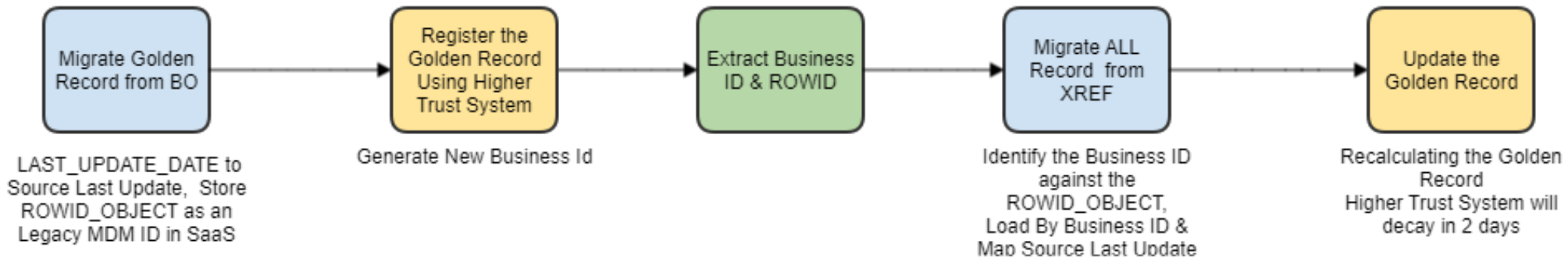
Recalculating the Golden
Record

BO & XREF

Option 2a

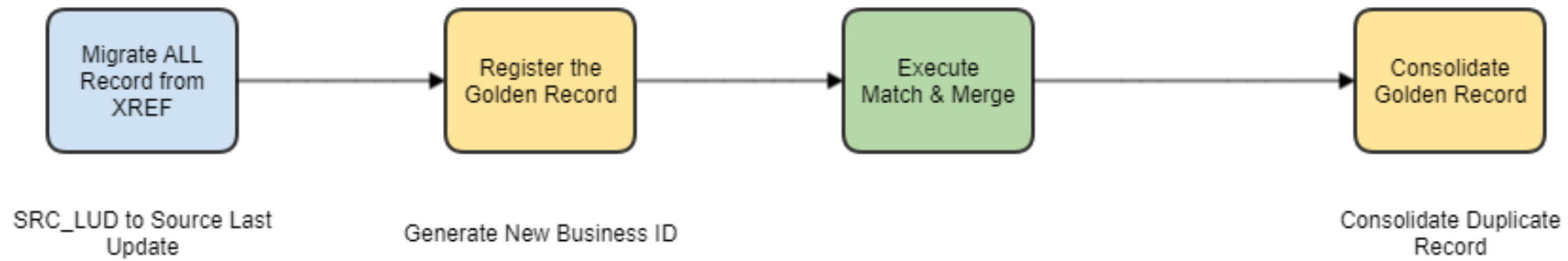


Option 2b



Rematching

Option 3



Initial Data Load

Seeder Mapping (ROWID_OBJECT = ORIG_ROWID_OBJECT)

```
SELECT * FROM c_bo_prt_y_xref
WHERE rowid_system='$$v_source_system'

and exists (select 1 from c_bo_prt_y prt_y where
prt_y.rowid_object = xref.rowid_object and
prt_y.prt_y_typ='$$v_prt_y_type')

and $$v_filter_condition

and hub_state_ind=1
```

```
SELECT * FROM
c_bo_prt_y_rle_phn_comm_xref
WHERE
rowid_system='$$v_source_system'

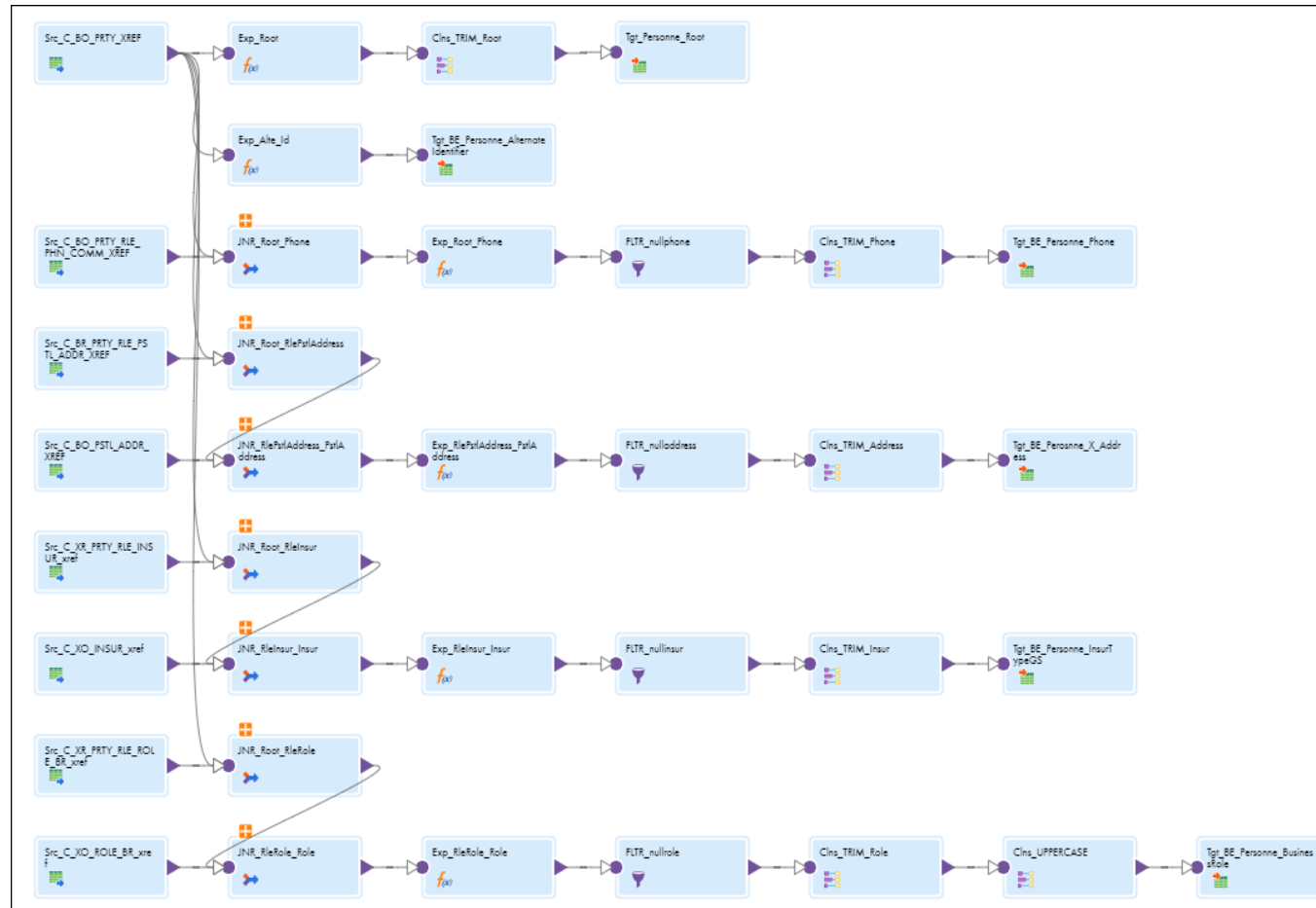
and hub_state_ind=1

ORDER BY ROWID_OBJECT
```

```
SELECT * FROM C_XO_ROLE_BR_xref
WHERE
rowid_system='$$v_source_system'

and hub_state_ind=1

ORDER BY ROWID_OBJECT
```



Child Record Join Condition

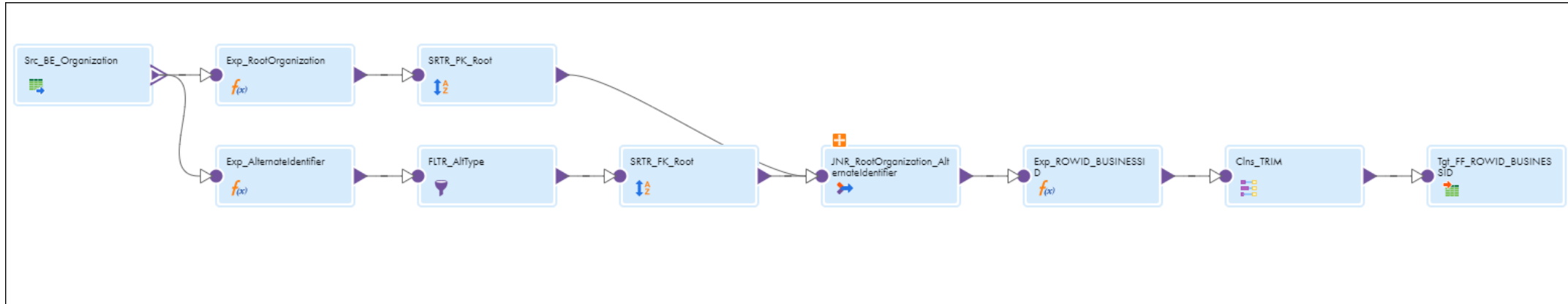
Example:

C_BO_PARTY_XREF.ROWID_XREF =
C_BO_PARTY_ALT_ID_XREF.PRTY_FK

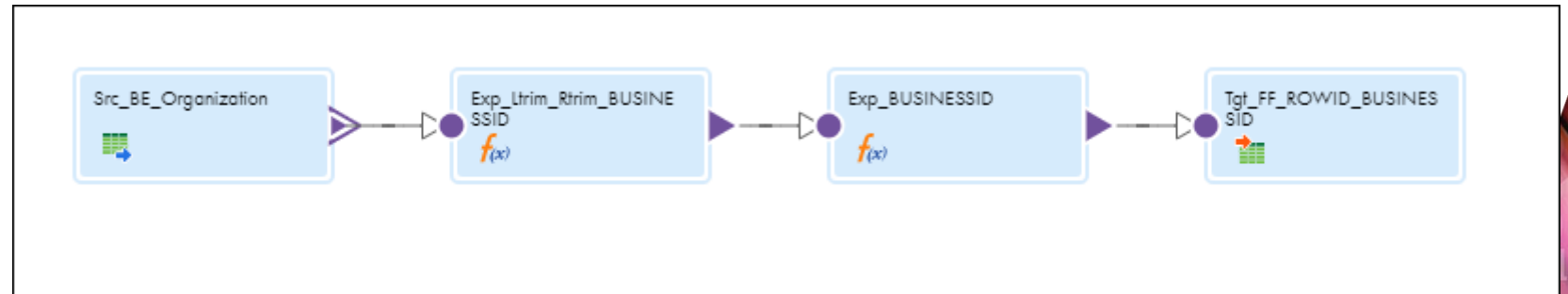
C_BO_PARTY_XREF.ROWID_XREF =
C_BO_PARTY_ALT_NM_XREF.PRTY_FK

Extract Business Id/Row ID Object

Seeder mapping with Business-id as Blank/Empty



Seeder mapping with Business-id as Row-id object



Initial Data Load

Non-Seeder Mapping (ROWID_OBJECT <=> ORIG_ROWID_OBJECT)

```
SELECT * FROM c_bo_prty_xref
xref
```

```
WHERE
rowid_system='$$v_source_system'
```

```
and exists (select 1 from c_bo_prty
prty where prty.rowid_object =
xref.rowid_object and
prty.prty_tpy='$$v_prty_tpy')
```

```
and $$v_filter_condition
```

```
and hub_state_ind=1
```

```
SELECT * FROM
c_bo_PRTY_RLE_PHN_COMM
_XREF
```

```
WHERE
rowid_system='$$v_source_syst
em'
```

```
and hub_state_ind=1
```

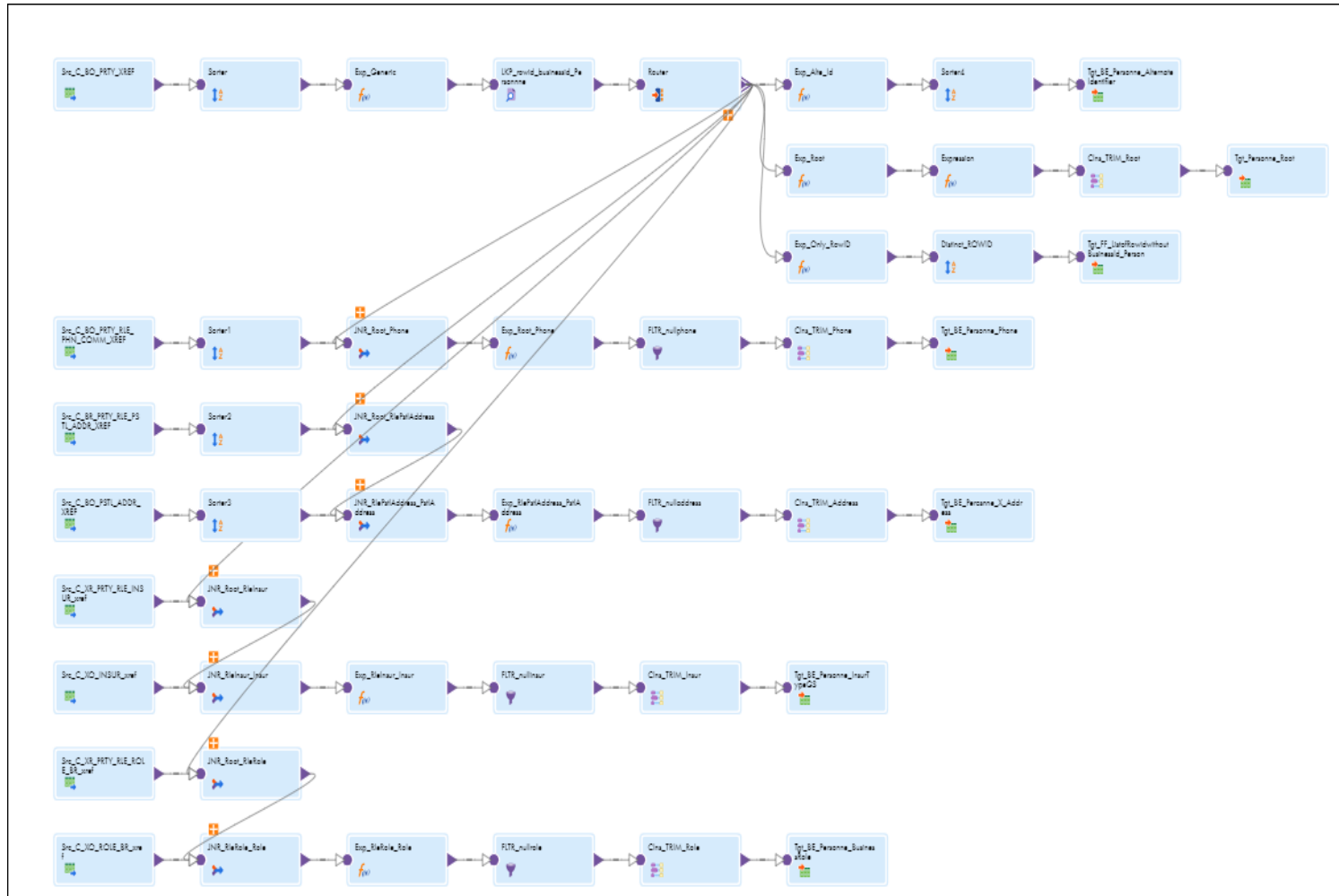
```
ORDER BY ROWID_OBJECT
```

```
SELECT * FROM
C_XO_ROLE_BR_xref
```

```
WHERE
rowid_system='$$v_source_syst
em'
```

```
and hub_state_ind=1
```

```
ORDER BY ROWID_OBJECT
```



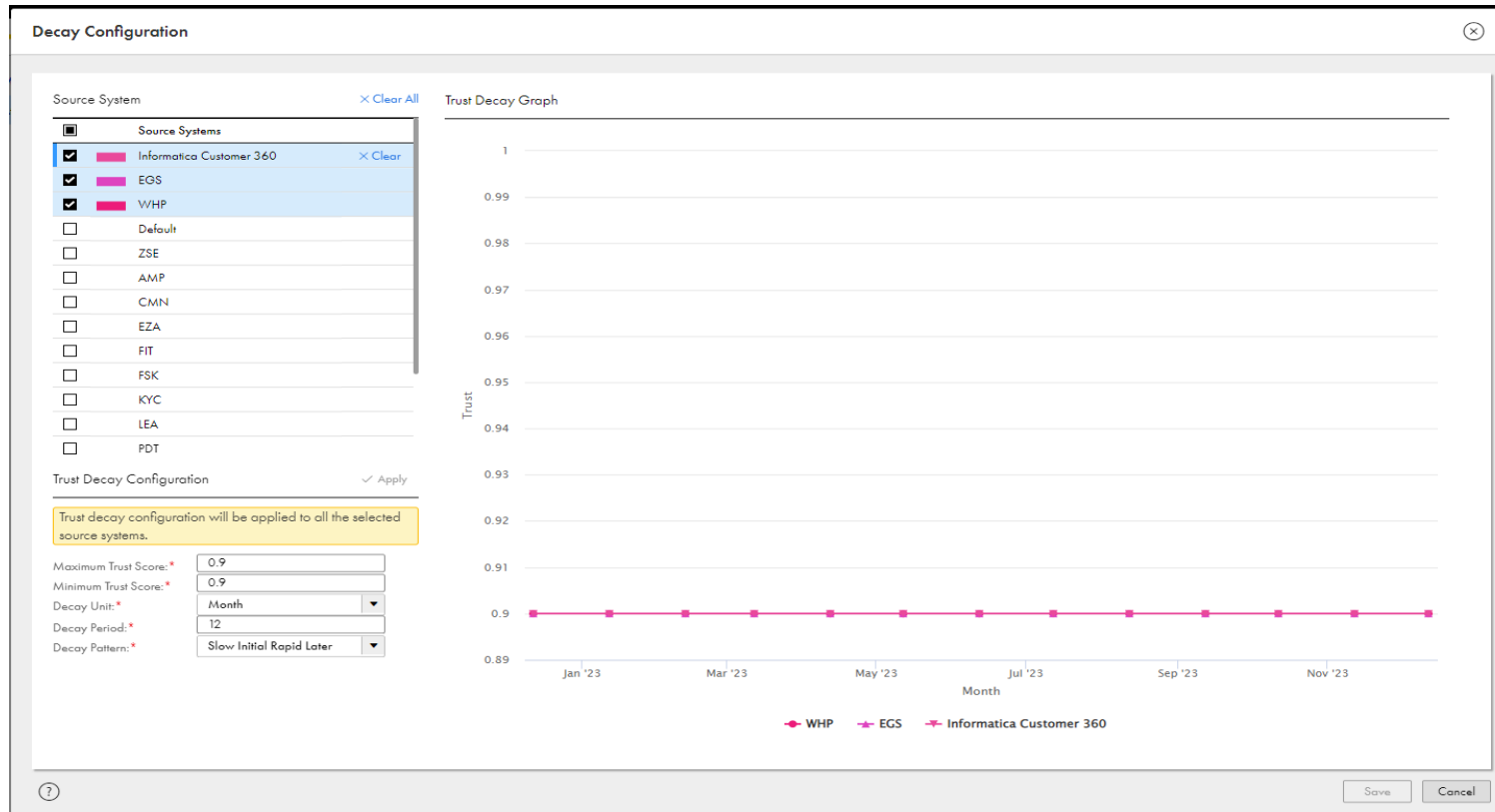
Child Record Join Condition

Example:

```
C_BO_PRTY_XREF.ROWID_XREF =
C_BO_PRTY_ALT_ID_XREF.PRTY_FK
```

```
C_BO_PRTY_XREF.ROWID_XREF =
C_BO_PRTY_ALT_NM_XREF.PRTY_FK
```

Survivorship



Applicable for Option 2

Source System

1. IDL Source

Decay

Minimum Trust Score: 0.9

Maximum Trust Score: 0.9

Decay Unit: Days

Decay Period: 2

Decay Pattern: Rapid Initial Slow Later

- This Will add one extra record for IDL source to mimic the golden record as on-prem (Optional)
- Requires license for extra unconsolidated record

Cross Reference Record

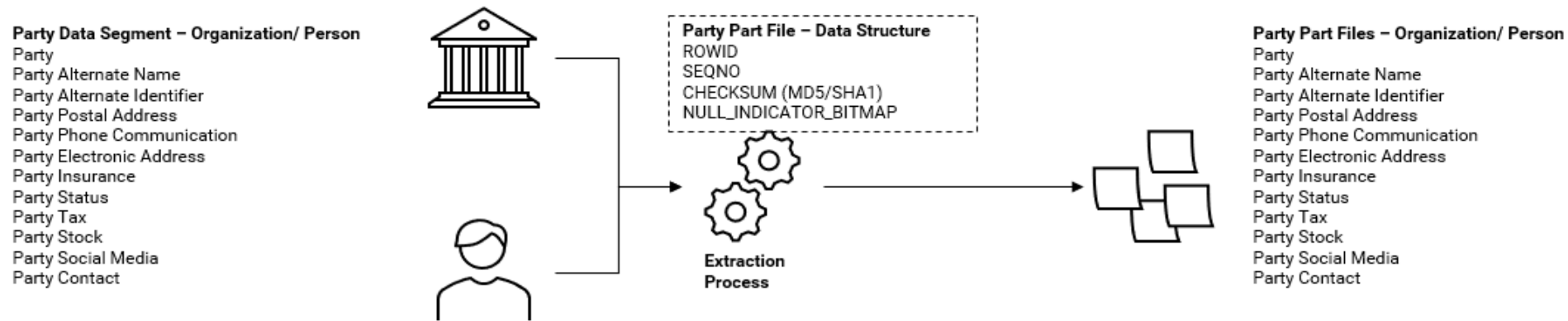
BERNARD MAURIN					
Person Références croisées Historiques Relations Graphe					
Source Records 4 of 4 source records					Acti
BERNARD MAURIN	BERNARD MAURIN Source System: WHP	BERNARD MAURIN Source System: WHP	BERNARD MAURIN Source System: WHP	BERNARD MAURIN Source System: WHP	
▼ System Fields					
Business ID	11143718				
ID	: 63932080ba1d...	WBH-KYC-563085	WBH-KYC-563113	WBH-KYC-563070	WBH-KYC-563108
Updated On	: 12/09/2022	12/09/2022	12/09/2022	12/09/2022	12/09/2022
Updated By	: atanu_gs_dev	atanu_gs_dev	atanu_gs_dev	atanu_gs_dev	atanu_gs_dev
Record State	: Active	Active	Active	Active	Active
Population	:				
▼ General Information					
Prenom	: BERNARD	BERNARD	BERNARD	BERNARD	BERNARD
Deuxieme prenom	:				
Nom	: MAURIN	MAURIN	MAURIN	MAURIN	MAURIN
Nom Complet	: BERNARD MAU...	BERNARD MAURIN	BERNARD MAURIN	BERNARD MAURIN	BERNARD MAURIN
Prefixe	:				
Suffixe	:				
Civilite	:				
Designation	:				

Data Validation

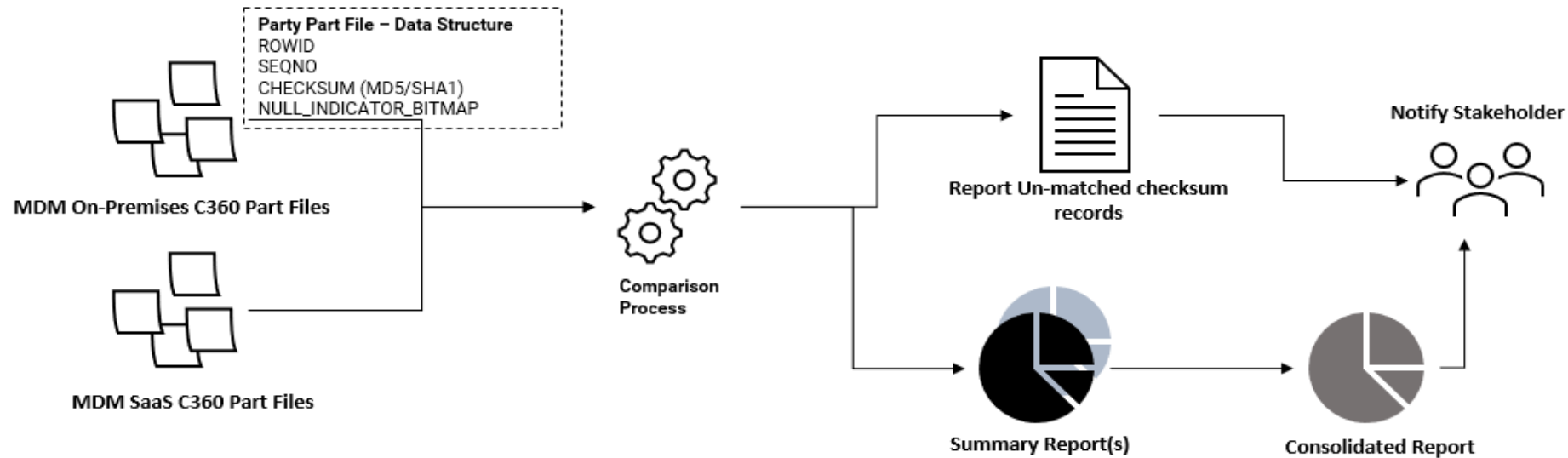
Data Validation Approach

- Establishing a standard framework for data reconciliation between on-premises vs SaaS
- Detail Reporting on data migration
- Certifying the completion of data migration to MDM SaaS platform
- Ensuring the seamless movement to MDM SaaS
- Validation framework offers utility to compare the Golden records from both systems
- Comparison process generates & verifies the CHECKSUM & NULL_INDICATOR_BITMAP
- Process reports the incident if there are any mismatch/error found.
- No business data writes over disk, ensuring higher data security
- At the end of the validation process, provides the un-matched report & final consolidated report

High Level Design



Step 1 & 2 : Extraction of part file from MDM C360 On Premises / MDM SaaS C360



Step 3,4,5,6 : Comparison analysis of part files of MDM on-premises vs MD SaaS, followed by summary reporting

Data Validation Report

Consolidated Report of Data Migration Activity

countMatched	countUnmatched	UnMatchedDueToMD5	UnMatchedDueToNull	UnMatchedDueToBoth	UnmatchedDueToMissingRow	ObjectName
6999	1	1	0	0	6	Party_ORG
15931	4053	4053	0	0	17	Party_ORG_Alternate_Identifier
3827	112	112	0	0	0	Party_ORG_Alternate_Name
11483	984	984	0	0	5	Party_ORG_Electronic_Address
15910	1423	1423	0	0	17	Party_ORG_Phone_Communication

Matched Report of a Part File

saas_ROWID	saas_CHECKSUM	saas_NULL_INDICATOR_BITMAP	saas_SEQNO	onprem_ROWID	onprem_CHECKSUM	onprem_NULL_INDICATOR_BITMAP	onprem_SEQNO	onprem_RowID_cls
7695	50BC07E9AA20D265A8F92C4F16CC46E5	1110011	T1	7695	50BC07E9AA20D265A8F92C4F16CC46E5	1110011	T1	7695
7515	DF2DE932ABEDD29B30E1DE9F6284ACF4	1110011	T1	7515	DF2DE932ABEDD29B30E1DE9F6284ACF4	1110011	T1	7515
9088	E62056DD61A79726A9CEBF99401147B6	1110011	T1	9088	E62056DD61A79726A9CEBF99401147B6	1110011	T1	9088
12738	4C51A96499256AA2C3501002EA9A7AE8	1110011	T1	12738	4C51A96499256AA2C3501002EA9A7AE8	1110011	T1	12738
12902	EF1C082740805B8D8155671240B19A0B	1110011	T1	12902	EF1C082740805B8D8155671240B19A0B	1110011	T1	12902
8250	AB322409C0F2E269769EDB9EA8282C55	1110011	T1	8250	AB322409C0F2E269769EDB9EA8282C55	1110011	T1	8250
9130	FA50A378B07606D634769775E8A28209	1110011	T1	9130	FA50A378B07606D634769775E8A28209	1110011	T1	9130
9330	ACB4CCAD0AF3D482DF76A2625531FCC2	1110011	T1	9330	ACB4CCAD0AF3D482DF76A2625531FCC2	1110011	T1	9330
9907	EF1AC705FF3D62F2D06BC530A209D328	1110011	T1	9907	EF1AC705FF3D62F2D06BC530A209D328	1110011	T1	9907
7982	873A4EA5C74328683D8E64983F3E7DE2	1110011	T1	7982	873A4EA5C74328683D8E64983F3E7DE2	1110011	T1	7982
8083	8631320C97B61A85E0F88A8CB30B86A2	1110011	T1	8083	8631320C97B61A85E0F88A8CB30B86A2	1110011	T1	8083

Data Validation Report

Unmatched Report of a Part File

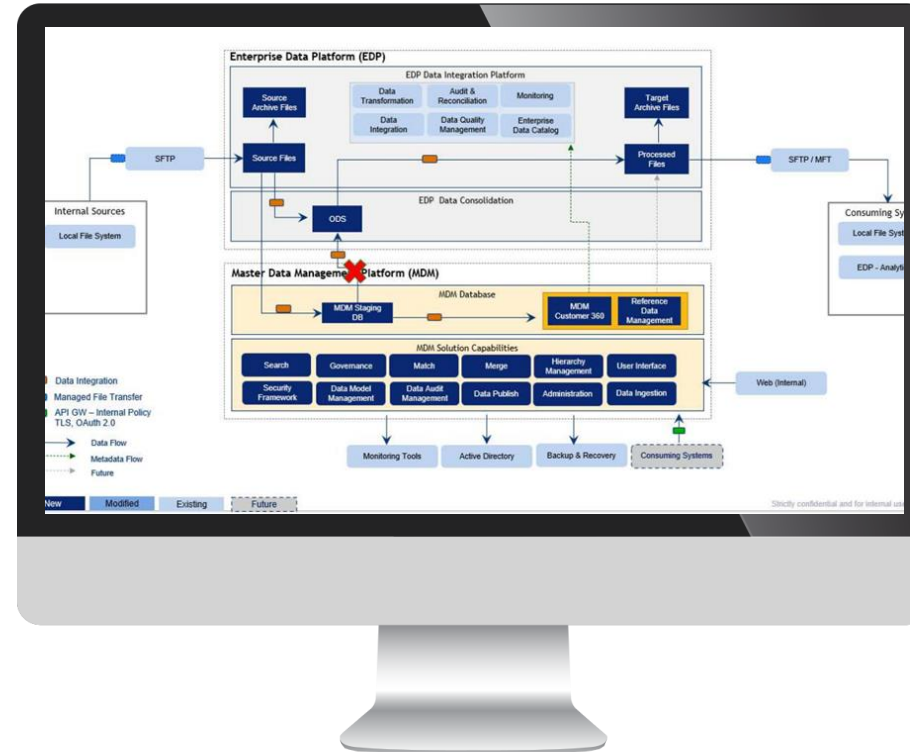
saas_ROWID	saas_CHECKSUM	saas_NULL_INDICATOR_BITMAP	saas_SEQNO	onprem_ROWID	onprem_CHECKSUM	onprem_NULL_INDICATOR_BITMAP	onprem_SEQNO	UnMatchedRemark
8212.0	F1B77021CDE4ECE7C843773E62D73592	1110011.0	T1	8212	42E9515B575B443135FED40FEE0BBEB9	1110011	T1	MD5 checksum does not match
nan	nan	nan	nan	488	D9BA36174D553B43EE2D5C9DCCA87508	1111011	T1	Row not available in SaaS
nan	nan	nan	nan	634	6DD8DE155BB4A7BC4B43958F6B301320	1111011	T1	Row not available in SaaS
nan	nan	nan	nan	4151	417091280BA69A22A05CDE5C859E0547	1110011	T1	Row not available in SaaS
nan	nan	nan	nan	6567	1DECF08EF94B59933E774E21EF3F9869	1110011	T1	Row not available in SaaS
nan	nan	nan	nan	7410	E1F19F27264BE6D2D71A7DAF07488183	1110011	T1	Row not available in SaaS
nan	nan	nan	nan	7412	A2A287F1C864A4E7007A2844F623439F	1110011	T1	Row not available in SaaS

Sample Use Case

Data Migration & Data Validation

- Our client has been operating an MDE C360 10.5 HF2 instance in production since 2018.
- The On-Prem instance contains data from three primary source systems: Legacy, MIG, and SFDC.
- Initial analysis indicates no data has been onboarded via the UI.
- A single match ruleset is configured within the On-Prem system with default survivorship settings.
- The customer data managed is primarily B2B client information.

DEMO



Dependencies

Data mapping requirements are depending on the data migration rules, such as business logic, data quality rules etc.

Data extraction mappings are partially parameterized

Reports are publishing in CSV format file, contains the count and other details

Limited amount of development effort require to consider for Addition and Extension

Data conversion related transformation need to be identified and implemented prior to validation

Limited to MDM C360 data model and additional effort requires for custom solution

Benefits

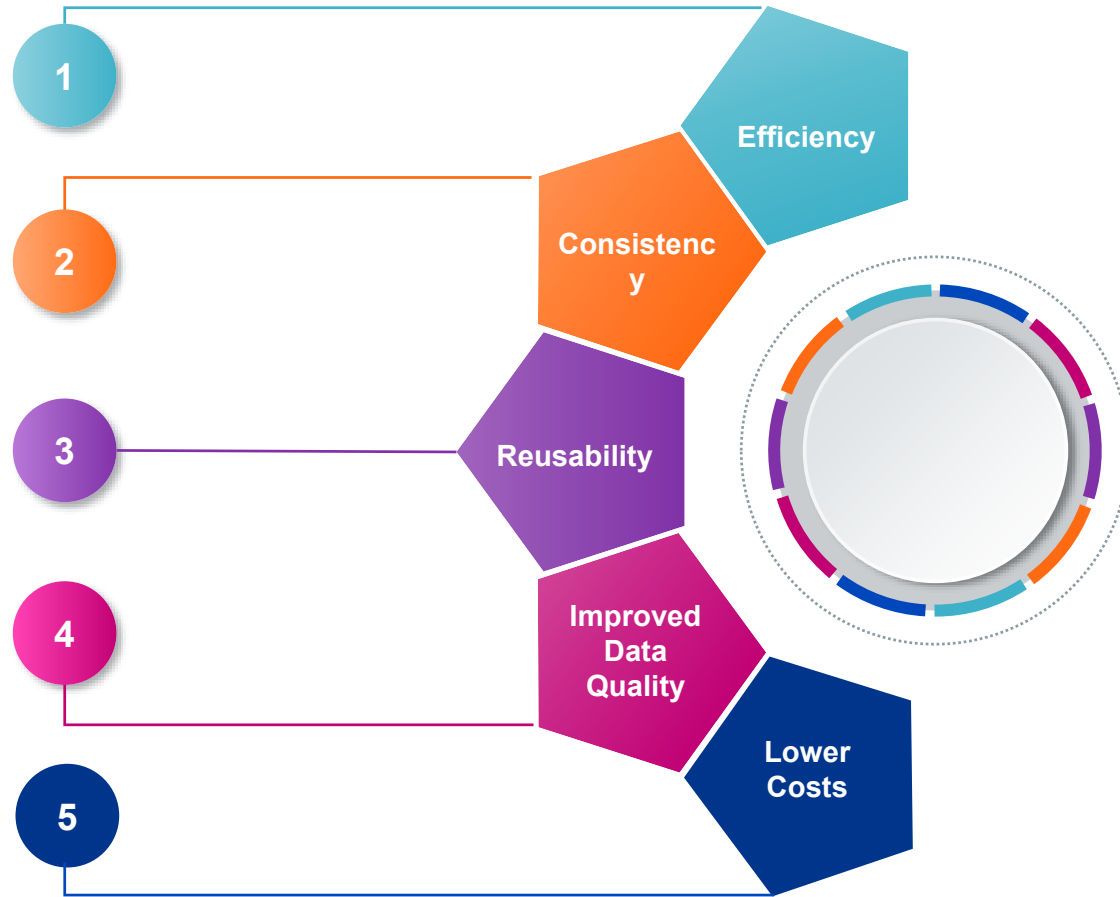
This reusable framework can streamline the migration process, reducing the time and effort required by up to **70%**.

By using a standardized framework, you can ensure that data is migrated consistently reducing the risk of errors and inconsistencies

Can be reused for future migrations, reducing the need to create new migration processes from scratch

Includes pre-built data quality checks, which can help to improve the accuracy and consistency of the migrated data

The framework can reduce the need for specialized expertise and reduce the risk of errors that can result in costly delays and rework



Q & A

Where data & AI come to

