



9 March 2022

Architect Workshop

*Master Data Management - Fueling
the Business or a Burden to IT?*

Housekeeping

- This session will be recorded
- Please feel free to ask any question via the Q&A option (not via chat) – they will be answered asap, some questions may be taken at the end of the session
- Please interact with us via the poll questions asked during the presentation
- End time is 16:30 CET

Agenda

1

Business drivers
for MDM

2

Architectural
principles and
design patterns

3

Customer case

4

Q&A

Business drivers for MDM

Marielle Verschoor

MDM Solution Architect – EMEA/LatAm



Informatica®

Business Drivers for Master Data Management



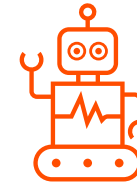
Customer Experience



Governance & Compliance



Operational Efficiency



Digital Transformation



Advanced Analytics



Mergers & Acquisitions

Many business initiatives include migrations



Customer Experience

Migration to Salesforce

Migration to Cloud

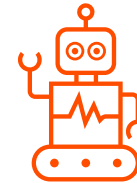
ERP consolidation

HR Consolidation to
Workday

Migration to more realtime
ecosystem

Migration to S4 Hana

...



Digital Transformation



Operational Efficiency

69% of organizations are
migrating business critical
applications to the cloud*

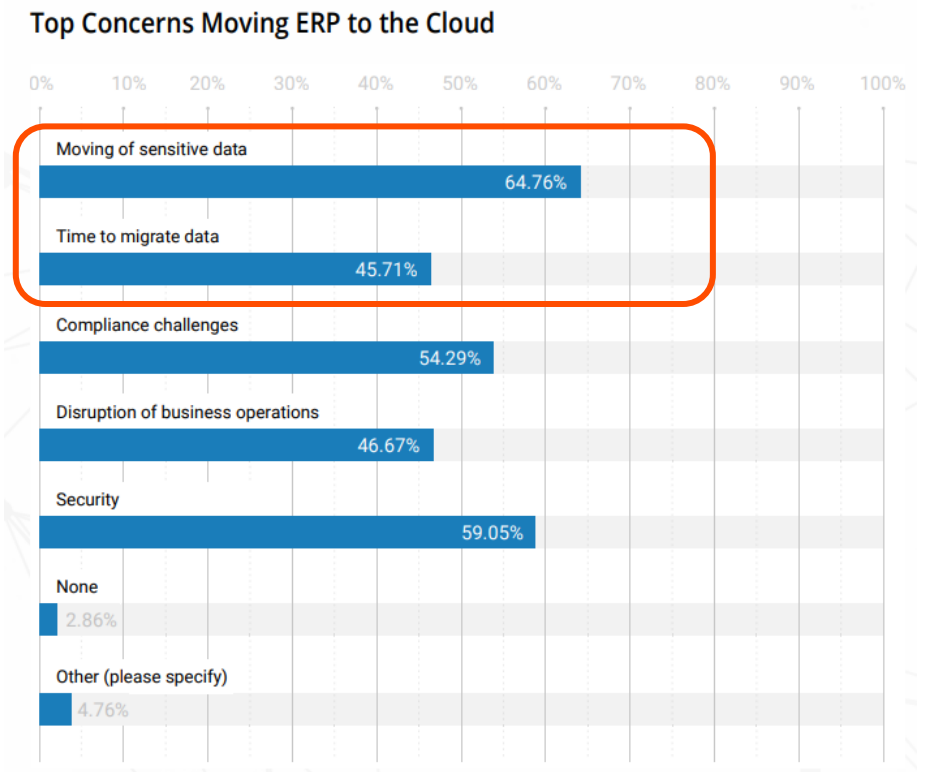
64% of organizations are
planning or in the middle of
an ERP cloud migration
project*



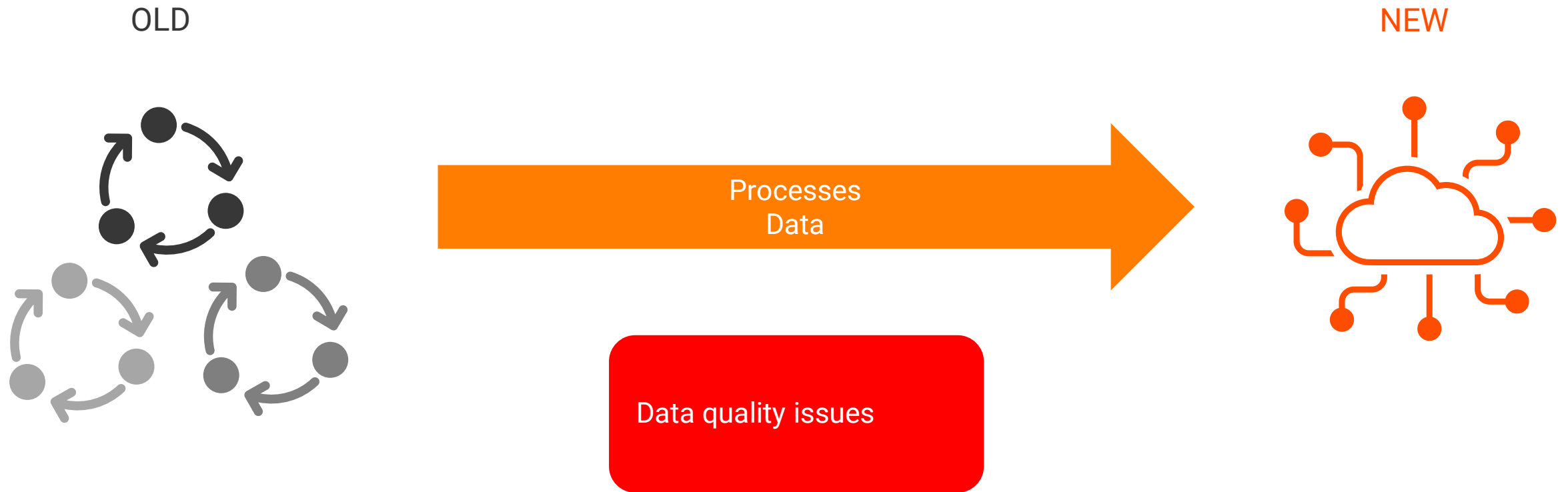
Mergers & Acquisitions

Data determines success of migrations

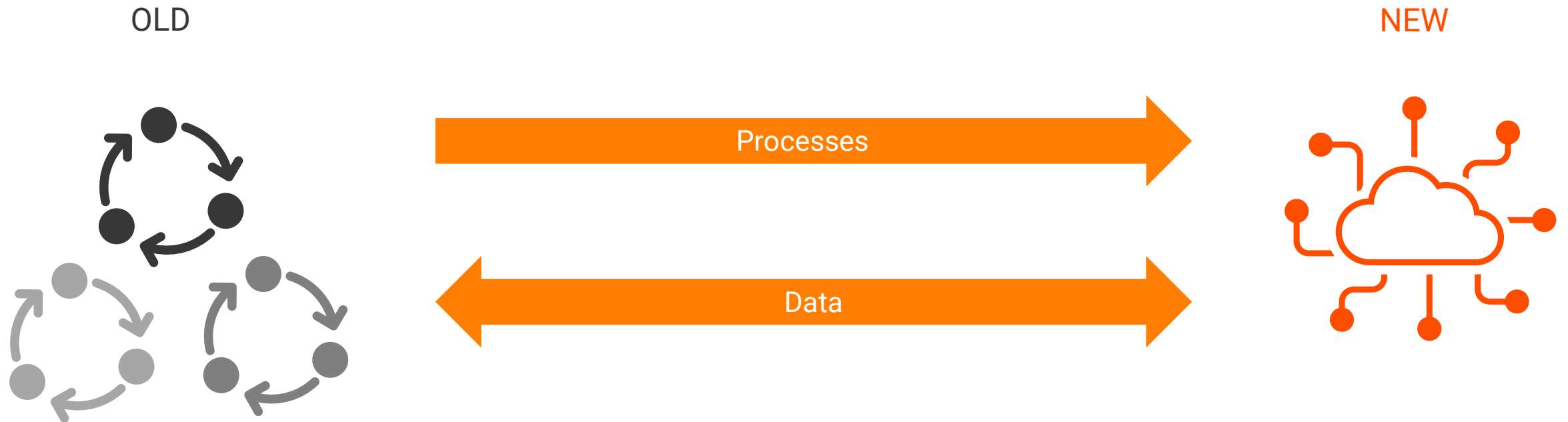
- Data accuracy defines success and acceptance of the new system
 - 90% of CIOs reported data migration projects falling short due to cloud on-prem complexity[#]
- Data issues cause budget & timeline overruns; 38% of projects are over budget and 47% over time*
 - 13% of budget overruns caused by data issues*
 - 19% of timeline overruns caused by data issues*
- Prevent duplicate & inaccurate data



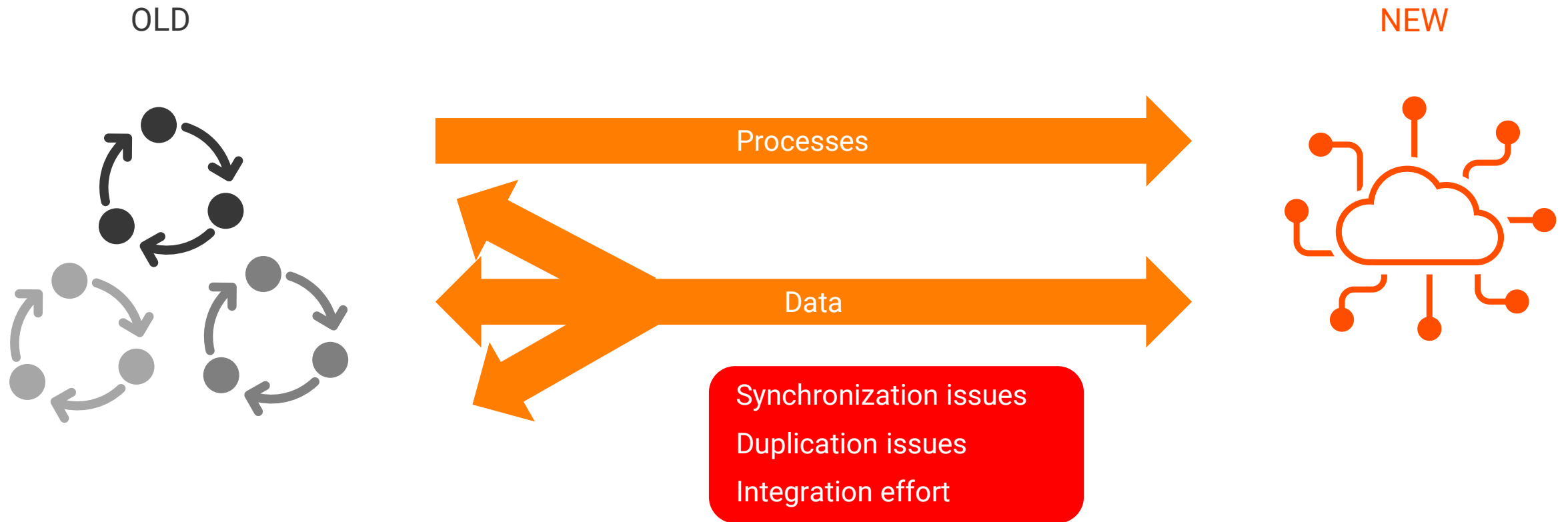
High level migration scenario



In fact - high level migration scenario



In fact - high level migration scenario



Migration considerations

Process migration considerations

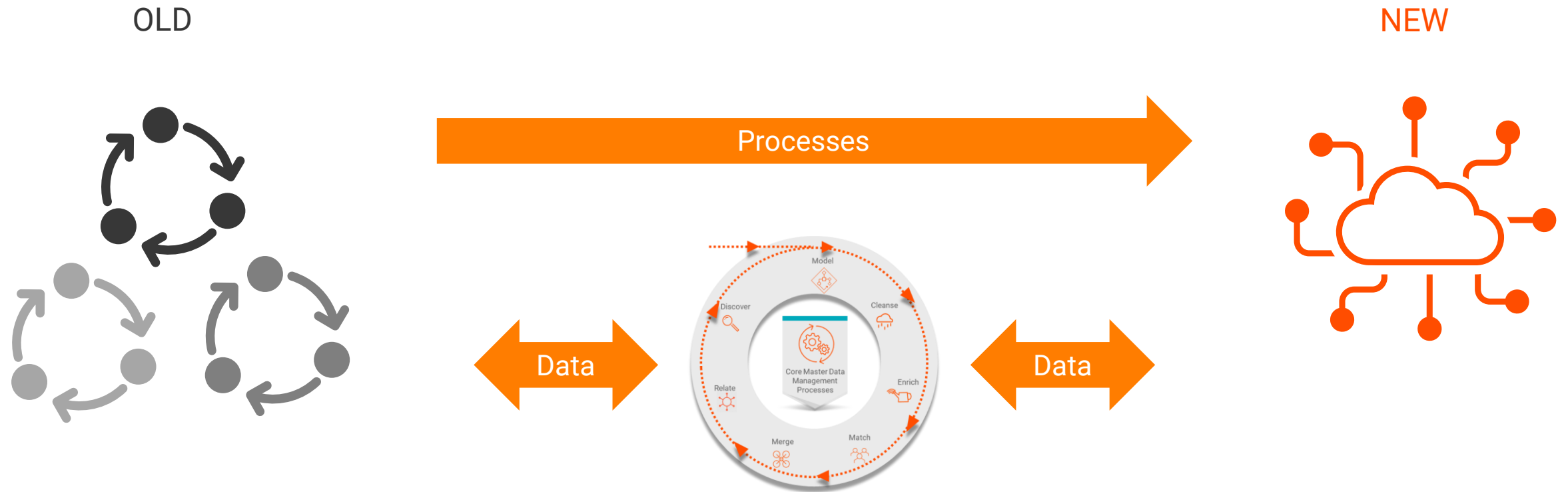
- More systems to be consolidated **increases risk & complexity** of data migration
- Migrations are not a big bang approach; old and new systems **need to run simultaneously**
- To be migrated processes are typically **diverse and spread** across different business entities

Data migration consequences

- Data migration needs to be efficient to **phase out old systems** as soon as possible
- Data of old and new systems needs to be **kept in synch continuously** and system independently
- Data migration needs to be **consistent** across different data domains
- Data from different systems needs to be **checked** for quality, duplicates and merged

Efficient & cost effective

Improved migration scenario



Architecture principles & Design patterns

Barry Wildhagen

MDM Solution Architect – EMEA/LatAm



Data for Specific Business Applications

(circa 1960-2000)

IBM
COGNOS

SAP
Business Objects

MicroStrategy



Analytics

Baan

SAP

ORACLE

Vantive

PeopleSoft

SIEBEL



Applications

IBM
Biginsights

TERADATA

Greenplum

NETEZZA

ORACLE
EXADATA

Pivotal

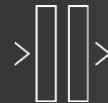


Databases

JMS

IBM MQ

TIBCO



Messaging

IBM

hp

DELL



Compute

IBM

hp

EMC²



Storage

Data for Enterprise-Wide Business Processes

(Last 15-20 Years)

IBM COGNOS
SAP Business Objects
MicroStrategy



Analytics



Power BI



tableau



Qlik



DOMO



Baan
SAP
ORACLE
Vantive
PeopleSoft
SIEBEL



Applications



salesforce



NETSUITE



CONCUR



workday



veeva



eloqua



Microsoft Dynamics



Marketo

IBM BigInsights
Greenplum
ORACLE EXADATA
TeraData
Netezza
Pivotal



Databases



cloudera



HORTONWORKS



Google BigQuery



MAPR



MEMSQL



snowflake



KUDU



SQL

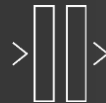


amazon REDSHIFT



amazon REDSHIFT

JMS
IBM MQ
TIBCO



Messaging



kafka



confluent



Azure Event Hub



Amazon Kinesis

IBM
hp
DELL



Compute



hadoop



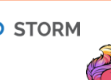
hive



cassandra



TEZ



STORM



Flink



SPARK



kafka



Google Compute Engine



Azure HDInsight



Amazon Web Services EC2 & EMP

IBM
hp
EMC²



Storage



vmware



hadoop



HBASE



MAPR



Azure Data Lake Store



Azure Blob Storage

Data Powers Digital Modernization & Transformation



AI/ML



Analytics



Applications



Databases



Messaging



Compute



Storage



What role does MDM play in migration architecture?

Lower cost, minimize risk, increase speed

1. **Pre-migration; Clean up** dirty data before migration, consolidation or upgrade
2. **During migration; Simplify** migration architecture
3. **Post-migration; Maintain data consistency** across new and old systems
4. **Repeat; Reuse data, mappings and rules** for the next migration project

What are typical steps for these data migrations?

1. Identify and access
2. Analysis of the source and target system(s)
3. Model the data, definitions and requirements
4. Define cleansing, enrichment, transformation rules
5. Map source to target attributes and crosswalks
6. Test load data into target system
7. Monitor quality and exceptions
8. Migrate data to new system and audit
9. Keep data in sync
10. Decommission source system

1. Identify and access; connectivity is key



<https://www.informatica.com/products/cloud-integration/connectivity/connectors.html#>

2. Analysis of the source and target system(s); Profile



Profile to examine structure and context



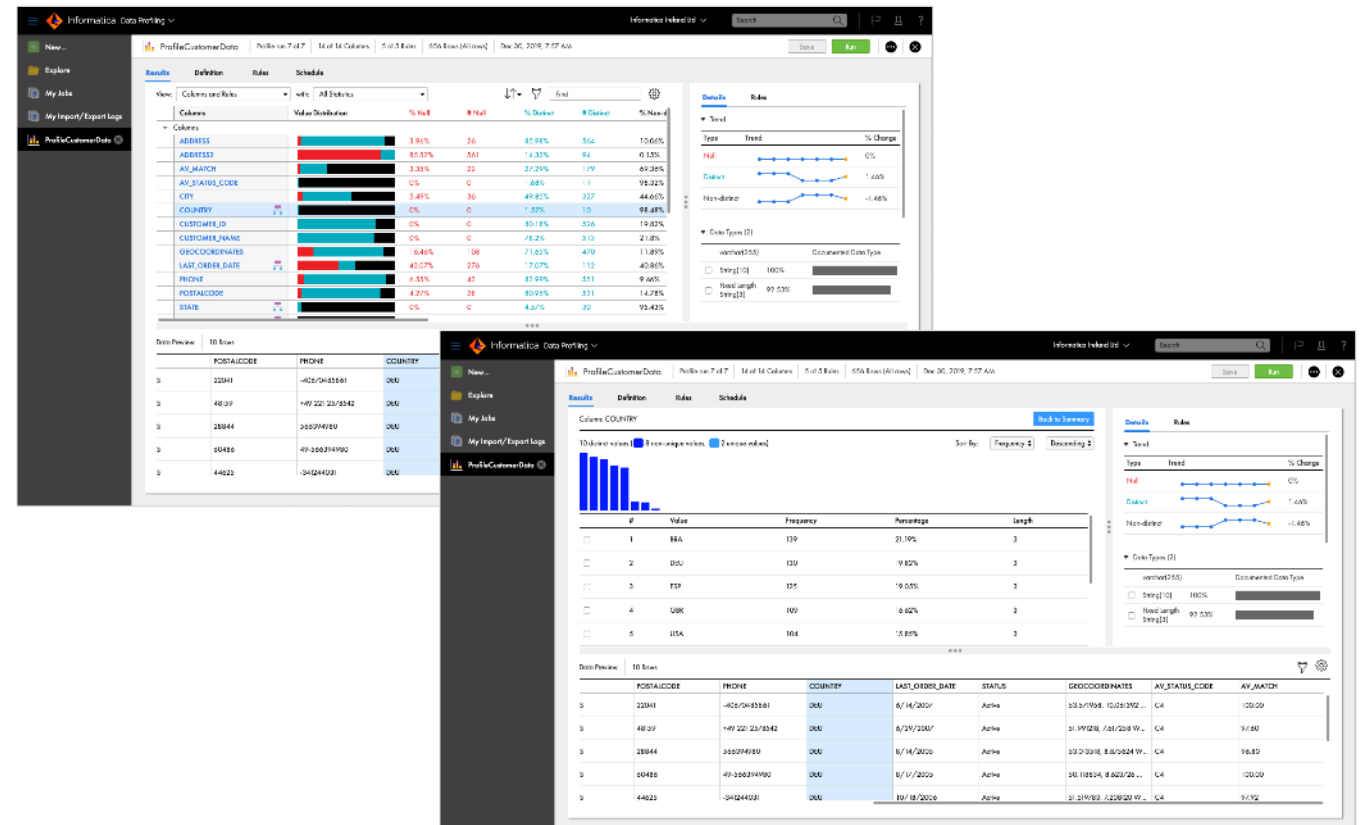
Catalog your data assets for easy discovery



Drill down to see details and filter results



Compare profile runs and identify trends over time



3. Model the data, definitions and requirements



Out of the box
Accelerators



Easy to extend in
new modeling
canvas



Ref360 code lists
for lookups from
MDM model



Smart Fields:
context aware
data types

All Attributes

Basic Fields

- Text
- Integer
- Double
- Decimal
- Boolean

First Name *	Middle Name	Last Name *	Full Name	Title	Birth Place
Birth Date	Gender	Image	Prefix Name	Suffix Name	Designation
Marital Status	Address	Status	Phone	Tax Details	Documents
Alternate Names	Email	Education	Identifier	Social Media	Employment
Citizenship	Financial	Insurance	Specialization	Lifestyle	Loyalty

Properties: Address

General

Survivorship

Display Name: * Address

Internal ID: * PostalAddress

Description: Defines all the fields in a standardized, global address. When you want to support more than one address in a record, select

Allow multiple values: ☒ Allow multiple values

4. Define cleansing, enrichment, transformation rules



Cloud DQ – DQ rules and reference data



DQ Rules associated directly with attributes



Trust downgrades tied to DQ rule results



Contact data verification

	Input	Operator	Condition	Action
if	firstName	is not	NULL	then Merge(firstName)
if	Result of # 1	is not	NULL	then Merge(PREVIOUS)
if	Result of # 2	is not	NULL	then REMOVE_MULTIPLE
or	if	No rule statement is valid	then	Do Nothing

5. Map source to target attributes and crosswalks



Orchestrated data flows



Extract data in batch with Data Integration

The screenshot displays the Informatica Data Integration (DI) console. The top navigation bar includes the Informatica logo, 'Data Integration', and user information 'Prathiba'. The left sidebar contains navigation links: 'New...', 'Home', 'Explore', 'Bundles', 'My Jobs', 'My Import/Export Logs', and 'Organization-Read'. The main workspace is divided into three panes. The top pane, titled 'Design', shows a visual mapping flow from a 'Source' to a 'Target' (MDM) through several intermediate components: 'ElectronicAddress', 'Indicator', 'Insurance', 'AlternateName', 'Document', and 'KeyExecutive'. The bottom pane, titled 'Properties', is currently showing the 'Field Mapping' tab for the 'MDM' target. It contains two tables: 'Response Structure' and 'Output Fields'. The 'Response Structure' table lists elements from the source response, and the 'Output Fields' table lists the corresponding target fields. The 'Field Mapping' tab is selected, and the 'Organization' element is expanded, showing its fields and their cardinalities. The 'Output Fields' table shows the mapping of these fields to the target structure.

Element Name	Cardinality
Organization	0-1
name	0-1
numberOfEmployees	0-1
companyType	0-1
dateOfIncorporation	0-1
stateOfIncorporation	0-1
countryOfIncorporation	0-1
imageUrl	0-1

Field Name	Actions	Mapped Field
root		/root
PK_root		system generated
sourceSystem		/htype/sourceSystem
sourcePKey		/htype/sourcePKey
name		/htype/Organization/name
numberOfEmployees		/htype/Organization/numberOfEmployees
companyType		/htype/Organization/companyType
dateOfIncorporation		/htype/Organization/dateOfIncorporation

6. Test load data into target system



Convert early



Use real
production data



Run multiple
mock
migrations



Establish
benchmarks

The screenshot shows the Informatica Monitor interface with a list of jobs. The table below represents the data visible in the screenshot.

Instance Name	Location	Subtasks	Start Time	End Time	Rows Processed	Status
Mapping_External_Match_Test-10	Configuration		Mar 8, 2022, 3:34 PM	Mar 8, 2022, 3:34 PM	0	Failed
Mapping_External_Match_Test-9	Configuration		Mar 8, 2022, 1:21 PM	Mar 8, 2022, 1:21 PM	0	Failed
Mapping_External_Match_Test-8	Configuration		Mar 8, 2022, 1:16 PM	Mar 8, 2022, 1:16 PM	0	Failed
Mapping_External_Match_Test-7	Configuration		Mar 8, 2022, 1:10 PM	Mar 8, 2022, 1:11 PM	0	Failed
Mapping_External_Match_Test-6	Configuration		Mar 8, 2022, 12:54 PM	Mar 8, 2022, 12:55 PM	0	Failed
Mapping_External_Match_Test-5	Configuration		Mar 8, 2022, 12:53 PM	Mar 8, 2022, 12:54 PM	0	Failed
Mapping_External_Match_Test-4	Configuration		Mar 8, 2022, 12:51 PM	Mar 8, 2022, 12:52 PM	0	Failed
Mapping_External_Match_Test-3	Configuration		Mar 8, 2022, 12:50 PM	Mar 8, 2022, 12:50 PM	0	Failed
Mapping_External_Match_Test-2	Configuration		Mar 8, 2022, 12:38 PM	Mar 8, 2022, 12:38 PM	0	Failed
Mapping_External_Match_Test-1	Configuration		Mar 8, 2022, 12:37 PM	Mar 8, 2022, 12:37 PM	200	Success
Taskflow_Match_External-685810969493...	Configuration	1 task	Mar 8, 2022, 12:27 PM	Mar 8, 2022, 12:30 PM	View Subtasks	Success
Mapping_Match_External-1	Configuration		Mar 8, 2022, 12:22 PM	Mar 8, 2022, 12:23 PM	0	Failed
Persons ECommerce Profile - run - 1-12	Data Profiling	1 task	Mar 2, 2022, 2:59 PM	Mar 2, 2022, 2:59 PM		Failed
Profile_CRM_Cust - run - 1-3	Data Profiling	1 task	Feb 25, 2022, 12:48 PM	Feb 25, 2022, 12:48 PM		Failed
PreviewMapping_RULE_SPECIFICATION-4	SystemTEMP		Feb 8, 2022, 1:23 PM	Feb 8, 2022, 1:23 PM	3	Success
Profile_CRM_Cust - run - 1-2	Data Profiling	1 task	Feb 4, 2022, 6:23 PM	Feb 4, 2022, 6:23 PM		Failed
Profile_CRM_Cust - run - 1-1	Data Profiling	1 task	Feb 4, 2022, 6:21 PM	Feb 4, 2022, 6:21 PM		Failed
Profile_Orgs - run - 1-2	Default	1 task	Jan 25, 2022, 10:47 PM	Jan 25, 2022, 10:48 PM		Failed
Profile_Orgs - run - 1-1	Default	1 task	Jan 25, 2022, 10:46 PM	Jan 25, 2022, 10:46 PM		Failed
Outlier - Profile_Persons - run - 1-2	Data Profiling	1 task	Jan 25, 2022, 10:39 PM	Jan 25, 2022, 10:39 PM	0	Success
Profile_Persons - run - 1-1	Data Profiling	3 tasks	Jan 25, 2022, 10:38 PM	Jan 25, 2022, 10:39 PM	17300	Success
Outlier - Profile_Test - run - 1-2	Data Profiling	1 task	Jan 25, 2022, 10:34 PM	Jan 25, 2022, 10:34 PM	0	Success

7. Monitor quality and exceptions



Track data quality improvements over time



Provide value to business users from onset



Align data quality/governance and data migration efforts



Enable IT to focus on exceptions

The screenshot displays the Informatica Data Profiling tool interface. The main window shows a comparison of two profile runs (Run 3 and Run 7) for the 'ProfileCustomerData' table. The table lists various columns and their associated quality metrics, including % Null, # Null, % Distinct, # Distinct, % Non-distinct, # Non-distinct, and # Patterns. The columns are: ADDRESS, ADDRESS2, AV_MATCH (Added), AV_STATUS_CODE (Added), CITY, COUNTRY, CUSTOMER_ID, CUSTOMER_NAME, GEOCOORDINATES (Added), LAST_ORDER_DATE, PHONE, POSTALCODE, STATE, and STATUS. The table also includes a section for 'CheckCompleteness' with input columns: STATUS, COUNTRY, and LAST_ORDER_DATE. The right-hand pane shows details for the 'ADDRESS' column, including values in Run 7, distinct values, non-distinct values, data types in Run 7, and patterns in Run 7.

Columns	% Null	# Null	% Distinct	# Distinct	% Non-distinct	# Non-distinct	# Patterns
ADDRESS	▲ 3.18%	▲ 21	▼ 4.44%	▼ 12	▲ 1.26%	▲ 10	-
ADDRESS2	▲ 29.95%	▲ 207	▼ 9.22%	▼ 56	▼ 20.73%	▼ 132	▼ 1
AV_MATCH (Added)	-	-	-	-	-	-	-
AV_STATUS_CODE (Added)	-	-	-	-	-	-	-
CITY	▲ 2.98%	▲ 20	▼ 3.68%	▼ 14	▲ 0.7%	▲ 13	▲ 1
COUNTRY	-	-	▼ 1.15%	▼ 7	▲ 1.15%	▲ 26	▼ 4
CUSTOMER_ID	-	-	▼ 2.39%	-	▲ 2.39%	▲ 19	-
CUSTOMER_NAME	-	-	▼ 2.18%	▲ 1	▲ 2.18%	▲ 18	-
GEOCOORDINATES (Added)	-	-	-	-	-	-	-
LAST_ORDER_DATE	▼ 1.25%	-	▲ 0.27%	▲ 5	▲ 0.99%	▲ 14	-
PHONE	▲ 0.43%	▲ 4	▼ 0.15%	▲ 15	▼ 0.28%	-	-
POSTALCODE	▲ 0.97%	▲ 7	▼ 0.53%	▲ 12	▼ 0.44%	-	▲ 1 ▼ 1
STATE	▼ 41.29%	▼ 263	▼ 10.5%	▼ 66	▲ 51.79%	▲ 348	▲ 1 ▼ 5
STATUS	▼ 1.88%	▼ 12	▼ 0.48%	▼ 3	▲ 2.36%	▲ 34	▲ 2 ▼ 5
CheckCompleteness Input Column: STATUS	-	-	▼ 0.16%	▼ 1	▲ 0.16%	▲ 20	▼ 1
CheckCompleteness Input Column: COUNTRY	-	-	▼ 0.01%	-	▲ 0.01%	▲ 19	-
CheckCompleteness Input Column: LAST_ORDER_DATE	-	-	▼ 0.01%	-	▲ 0.01%	▲ 19	-

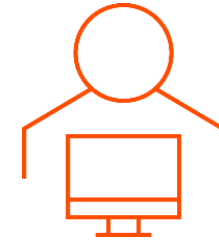
8. Migrate data to new system and audit



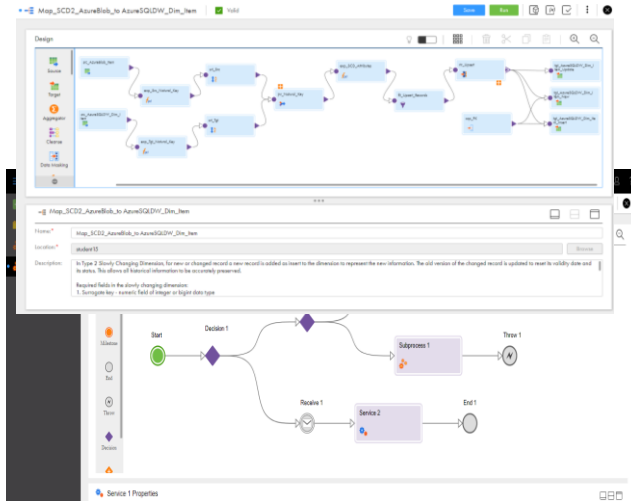
Developers



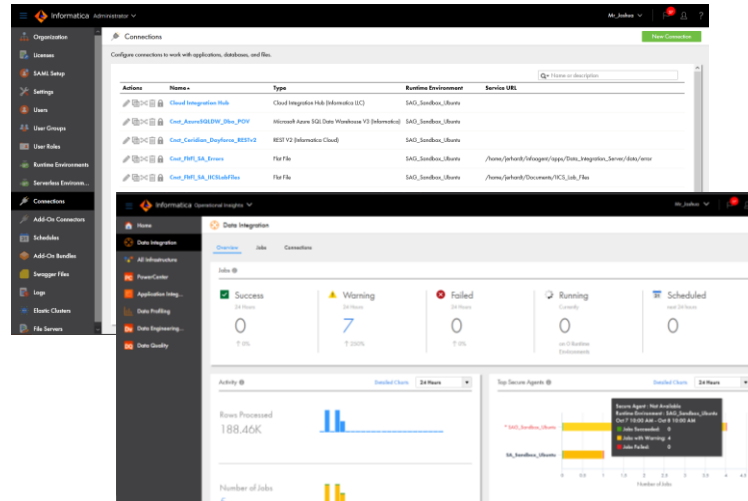
Admins



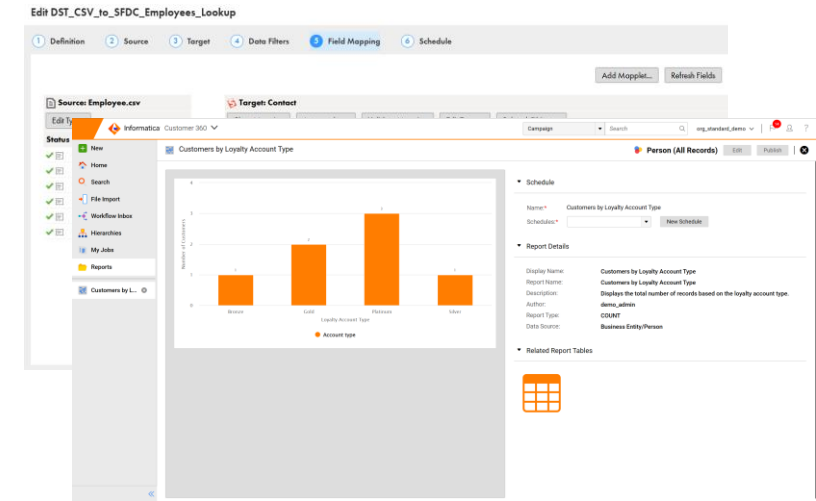
Business Users



On standby

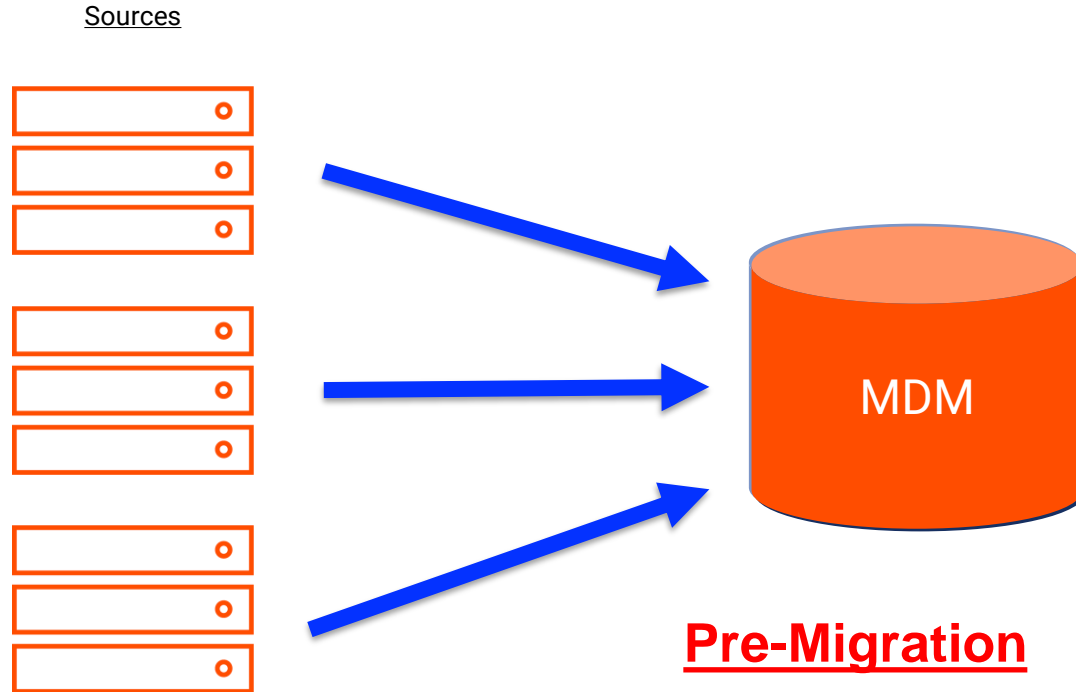


Monitoring migration jobs



Audit the new system

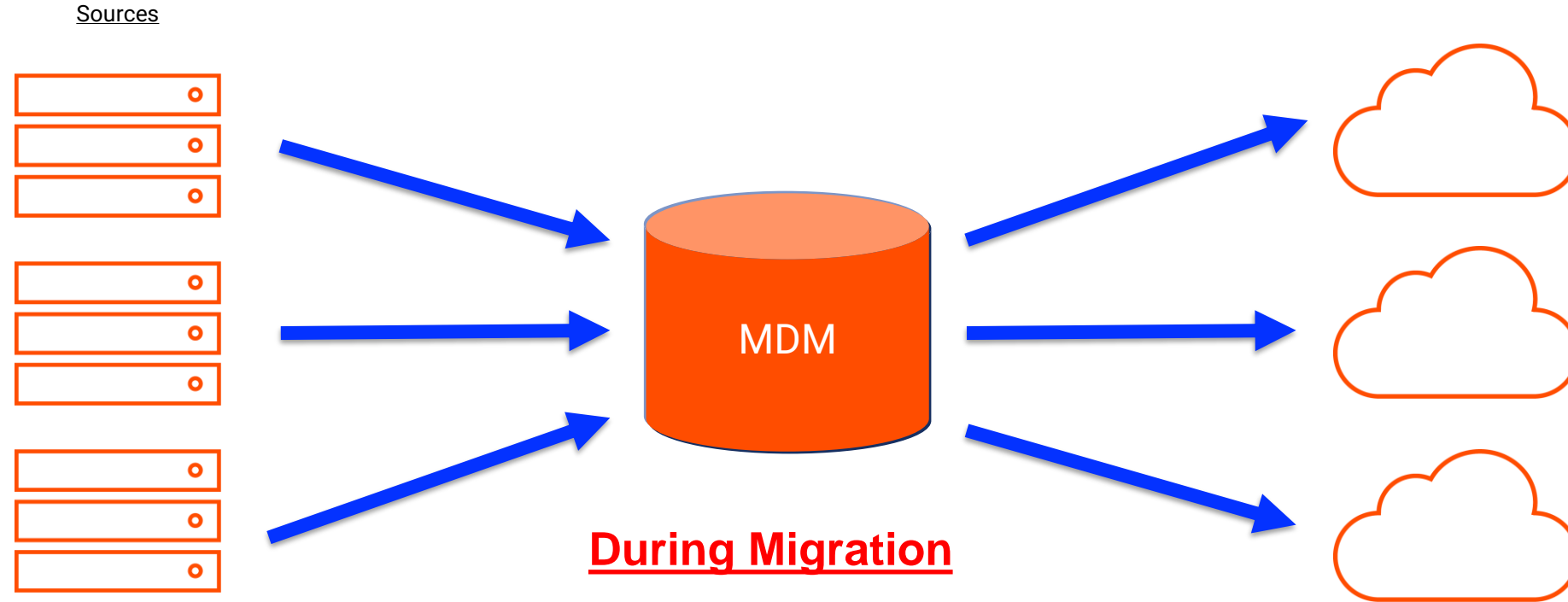
9. Keep data in sync (pre-migration)



Pre-Migration

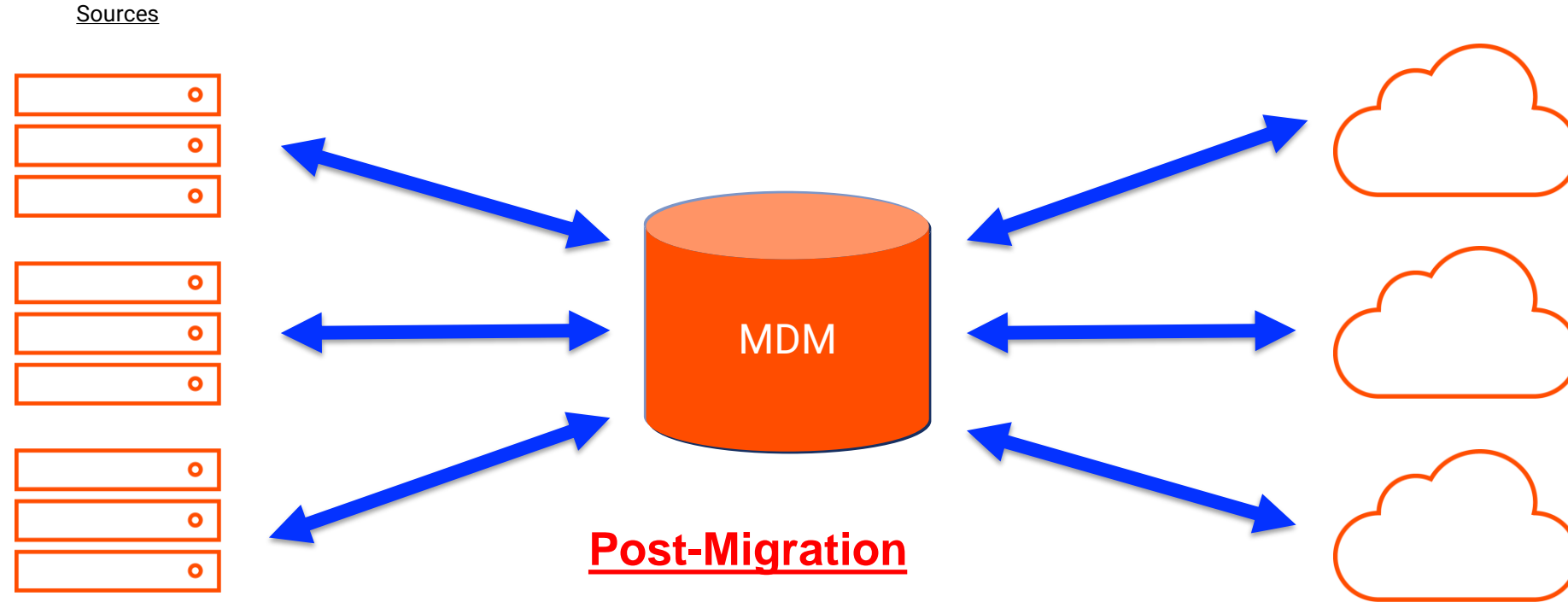
1. **Cleanse, standardize, and enrich** dirty data
2. **Remove duplicates** and create a best version of the truth
3. **Centrally manage** data cleansing/mastering rules

9. Keep data in sync (during migration)



1. **Avoid** spaghetti point-to-point integrations
2. **Simplify integration** with hub-and-spoke architecture
3. **Automate** data management of very large volumes of data

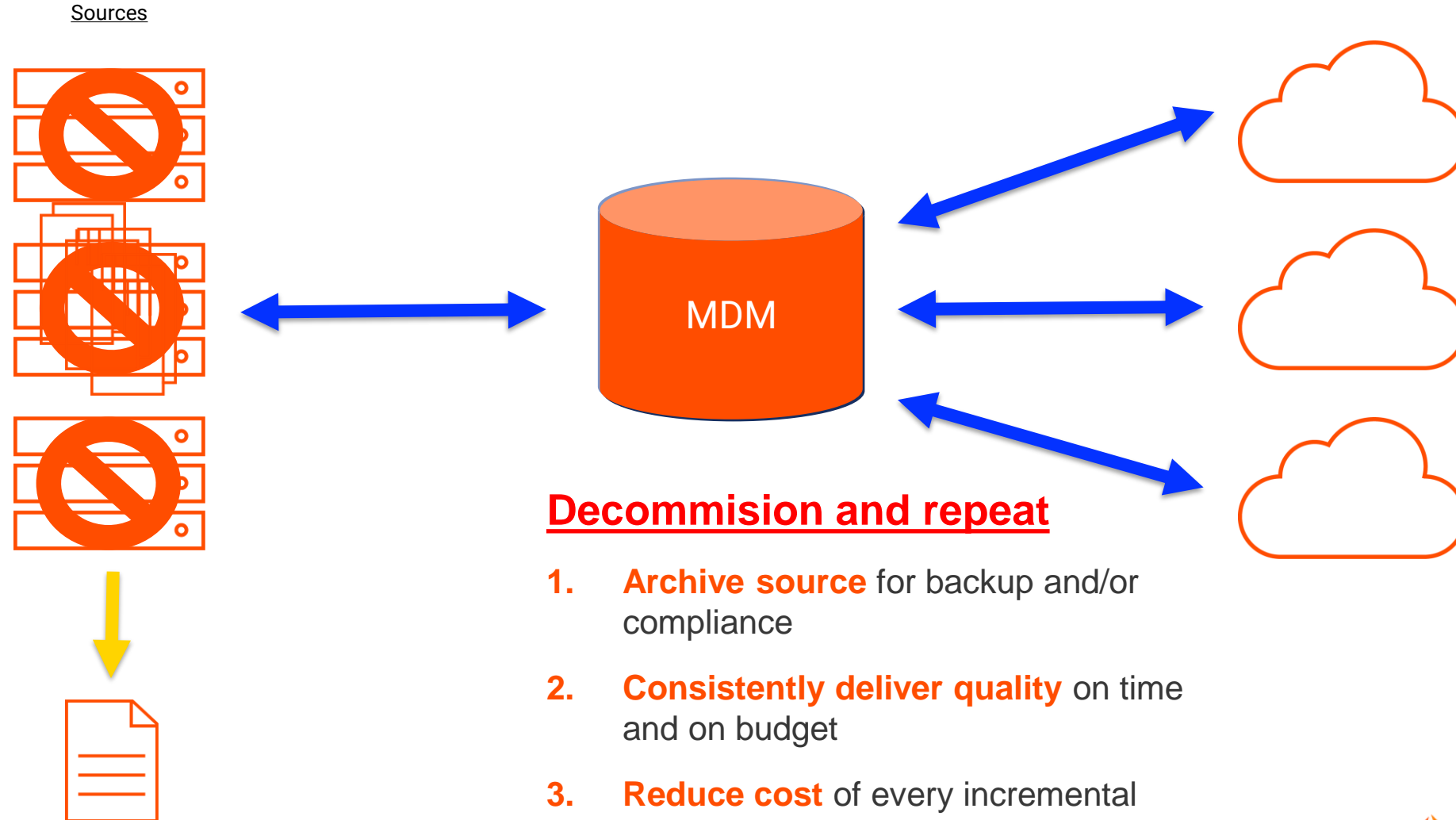
9. Keep data in sync (Post-migration)



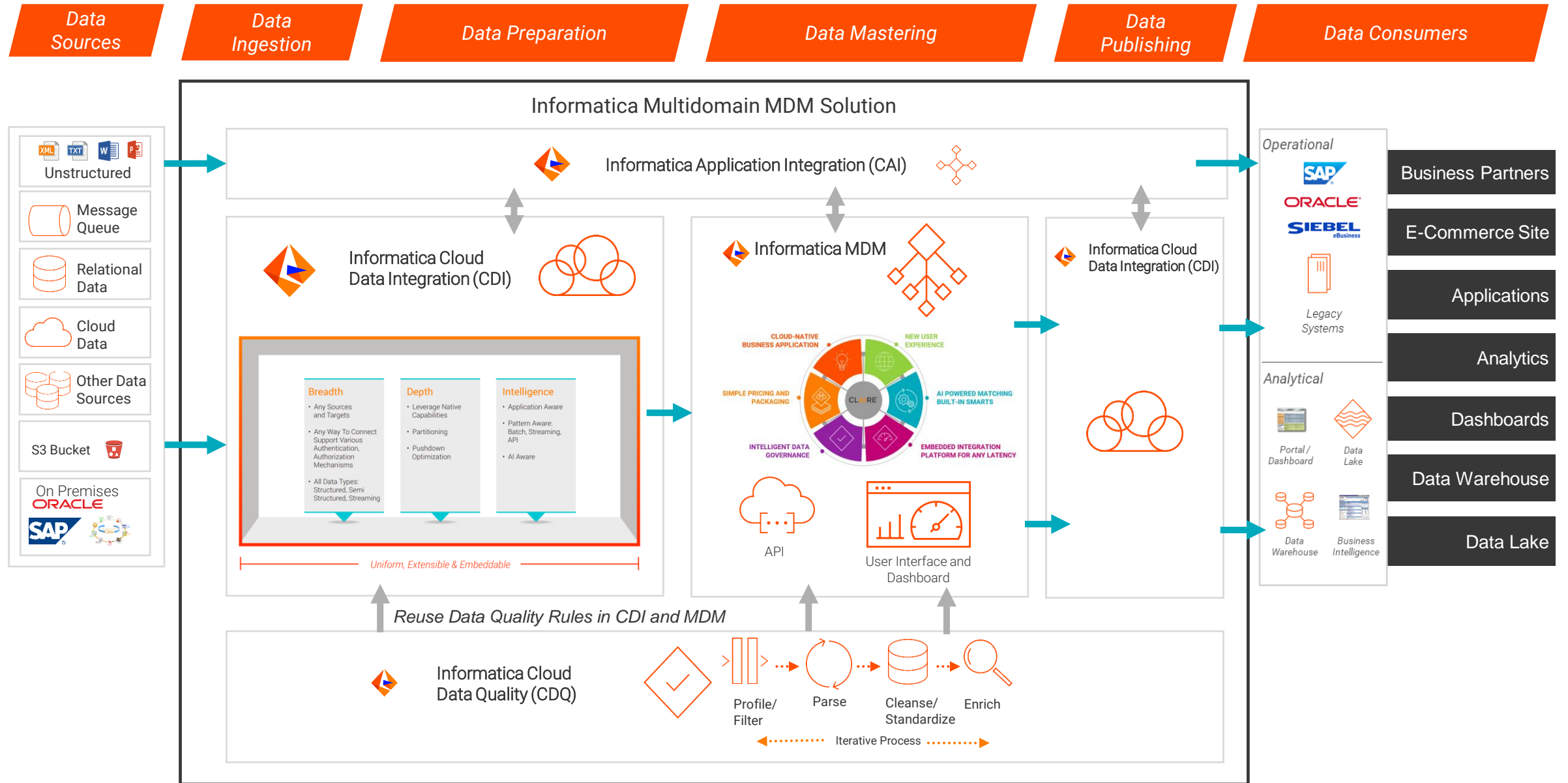
Post-Migration

1. **Synchronize clean data** from new systems with old systems
2. **Minimize disruption** to business processes still using old systems (until they are retired)
3. **Implement information governance** across the enterprise

10. Decommission source system



How does it all come together?





My Services



Application Integration



Application Integration
Console



Business 360 Console



Customer 360



Data Integration



Data Profiling



Data Quality



Reference 360



Administrator



Monitor



Operational Insights

[Show all services](#)

Customer Case

Accelerating Modernization

- One of the world's largest agricultural processors and food ingredient providers
- Operating in 160+ countries
- Rapid ongoing growth through many acquisitions
- Simplification and Reduction of costs by standardizing on ONE global ERP system



Guiding principles

Establish MDM principles

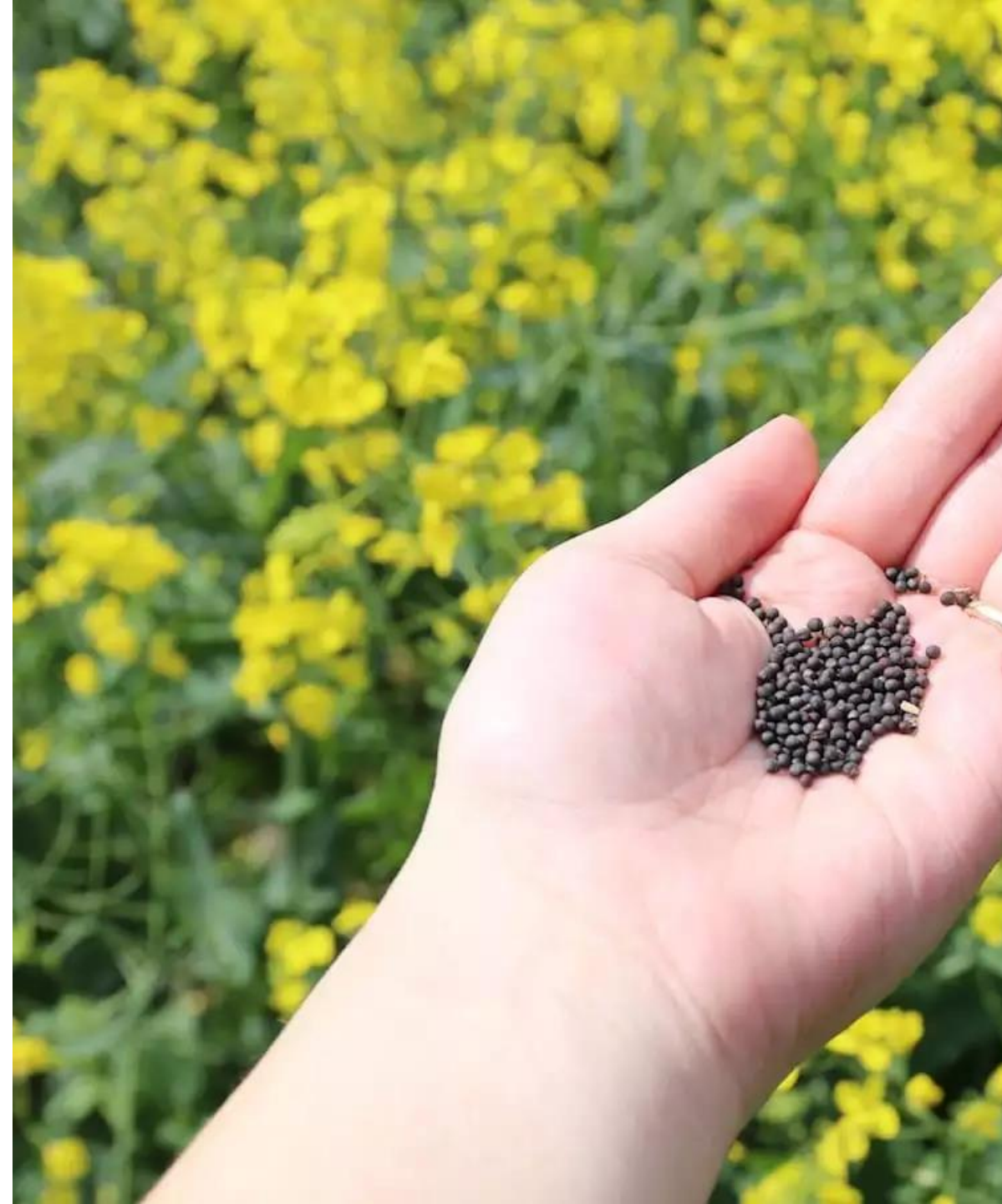
- Data is system agnostic
- Master data describes unambiguously the 'thing' it is describing
- Differentiate between global, local (regulatory), local (processing) requirements
- Govern master data not the MDM solution

Existing architectural principals

- Design for reusability
- Design for agility
- Design for adaptability
- Design for efficiency and effectiveness and a great user experience

The Benefits

- Drove digital transformation starting with visibility into end-to-end relationships and lineage between suppliers, products, customers, locations and employees.
- Ability to turn off legacy systems and reduce costs
- Streamlined processes; gained a single trusted view of suppliers across regions and systems; and enhanced accuracy of data.



Questions?

While we answer some of your questions
please feel free to also share your thoughts
about the session today

Thank You!