Informatica

30-11-2021

Architect Workshop

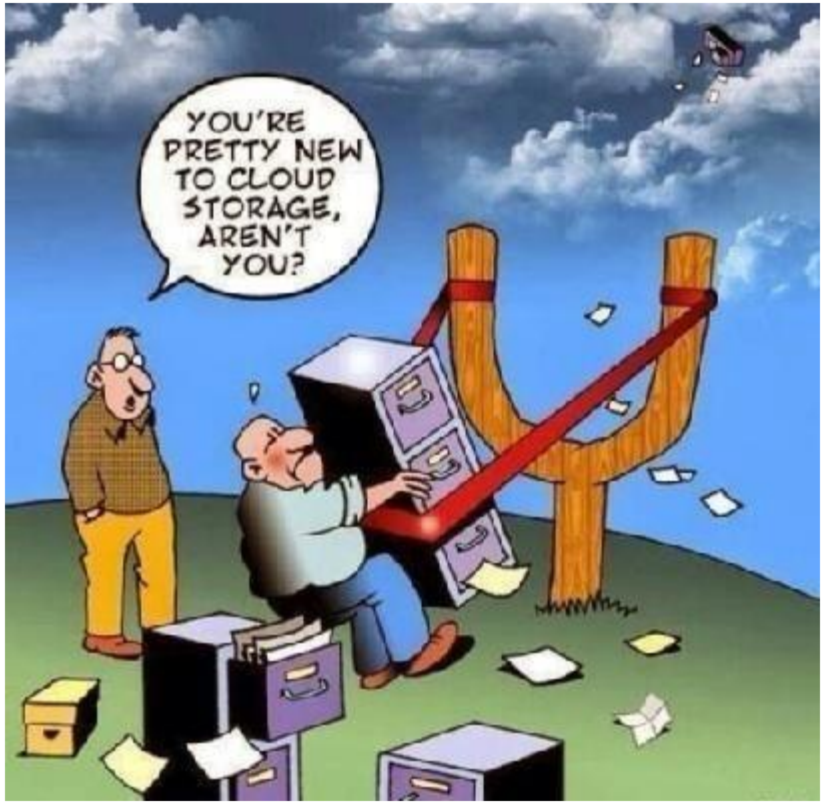
Architectural Considerations for modernizing Data and Applications to the Cloud

Remy van der Kleij

Joris Zwart

Rick Mutsaers

Topic – Architectural Considerations for modernizing to the Cloud



Agenda

1 What to consider

2 Experiences

3 Modernizing your analytics

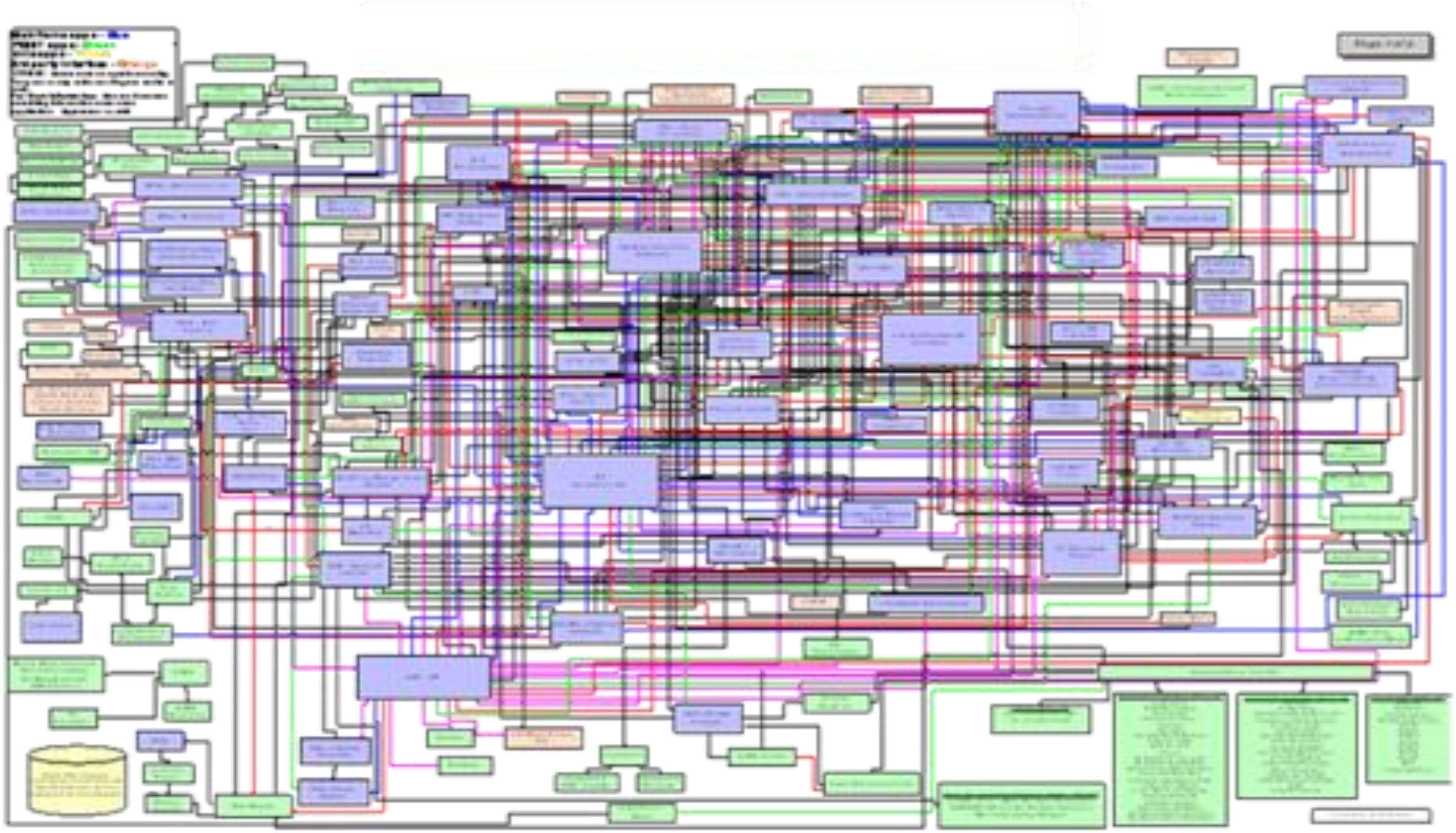
4 Closing

What to consider

Modernizing towards the Cloud

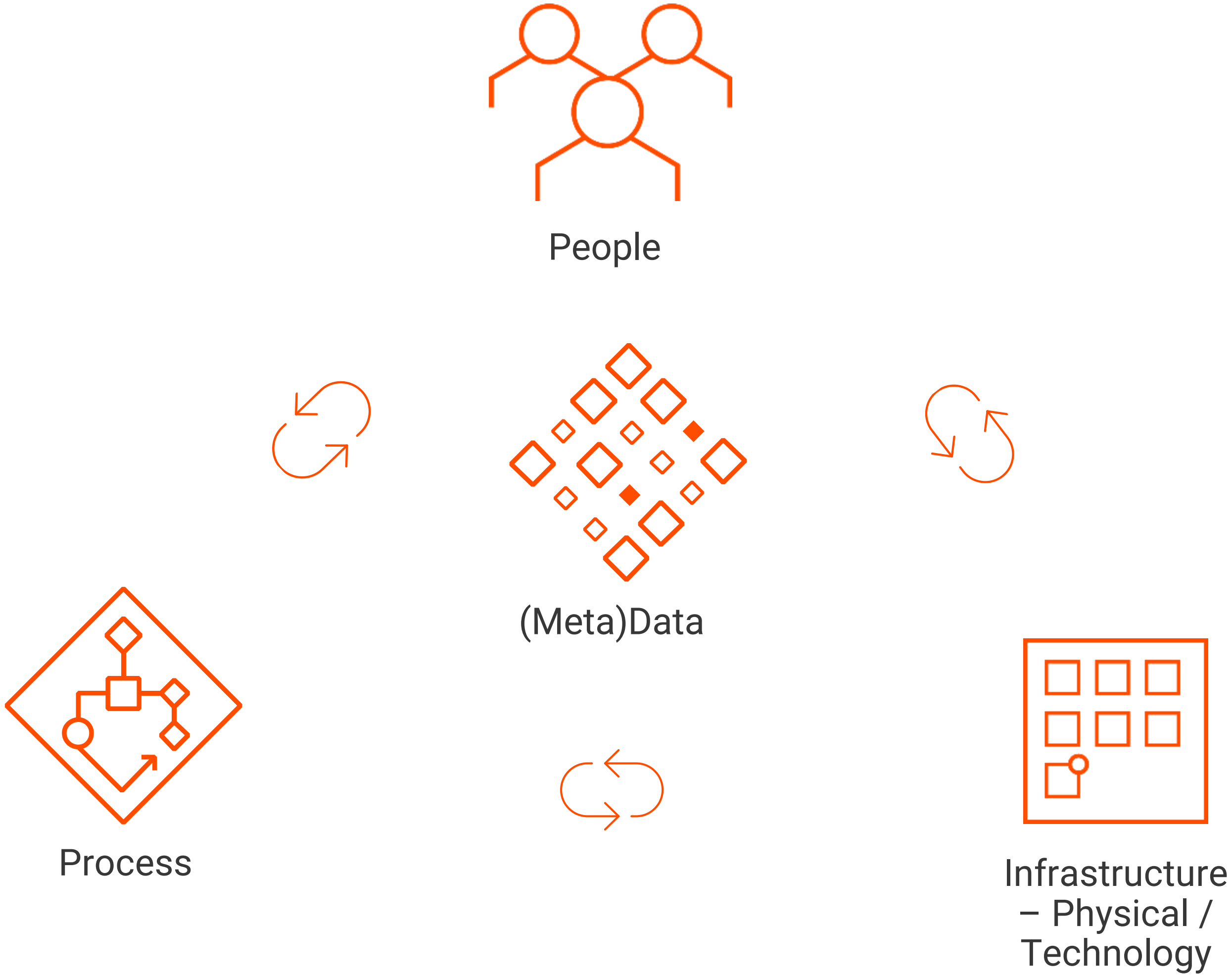
Remy van der Kleij
Solution Architect, Informatica

Looks familiar as your starting point?



Data integration architecture from an actual customer

How to make sure this doesn’t happen again



People

(Meta)Data

Process

Infrastructure – Physical / Technology

Key information elements to capture for modernization

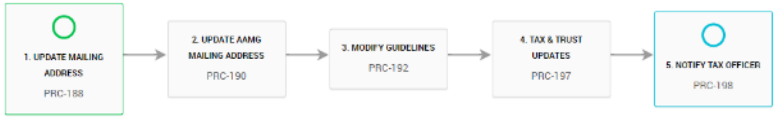
- Involved Systems
- Critical data elements & definitions
- Data lineage & traceability
- Business process details
- Business rules & quality validation

+ Stakeholders for these assets
+ Collaboration between stakeholders

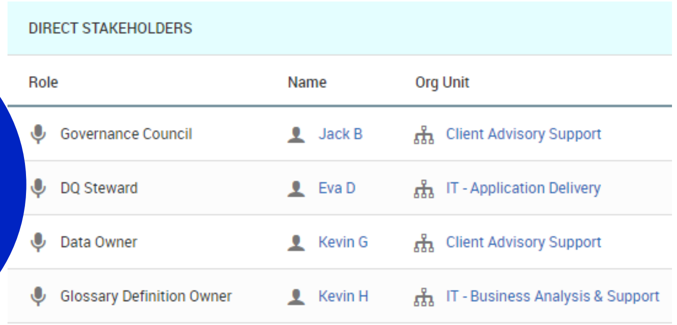
Need to eliminate gaps between People, Processes and Data

Security Policy

Process



Stakeholders



DIRECT STAKEHOLDERS

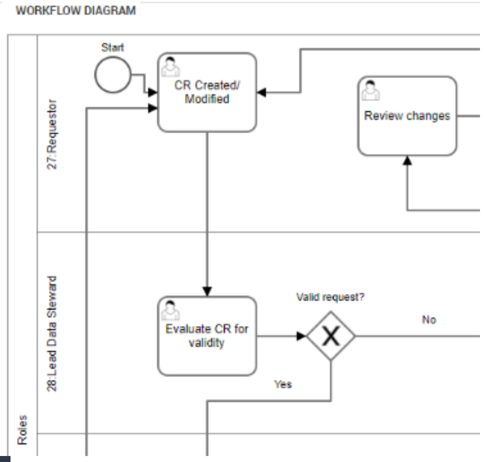
Role	Name	Org Unit
Governance Council	Jack B	Client Advisory Support
DQ Steward	Eva D	IT - Application Delivery
Data Owner	Kevin G	Client Advisory Support
Glossary Definition Owner	Kevin H	IT - Business Analysis & Support

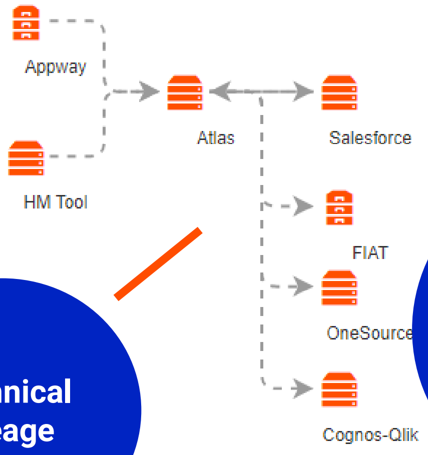
Quality Scores

Data Quality

How do I Understand the true impact of a Data Asset?

Change Request

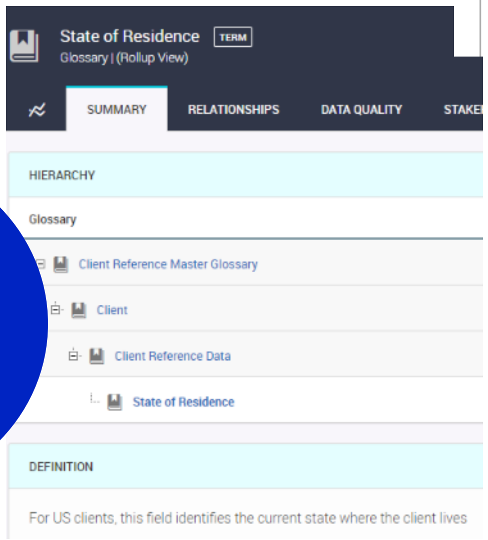




Technical lineage

Business Lineage

Glossary



State of Residence TERM

Glossary | (Rollup View)

SUMMARY RELATIONSHIPS DATA QUALITY STAKE

HIERARCHY

Glossary

- Client Reference Master Glossary
  - Client
    - Client Reference Data
      - State of Residence

DEFINITION

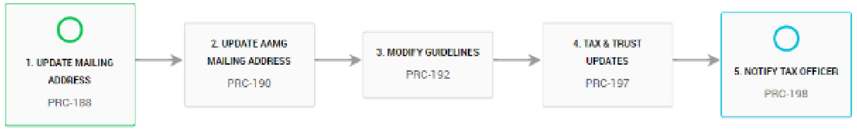
For US clients, this field identifies the current state where the client lives

Informatica

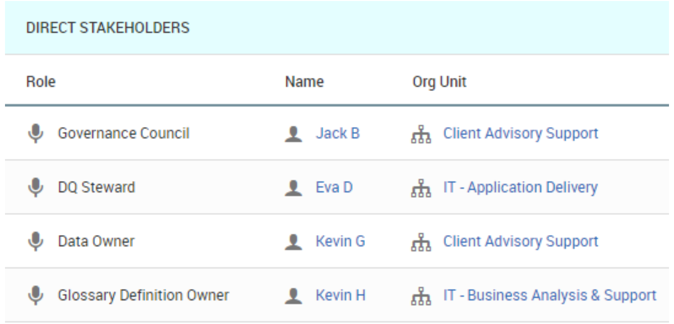
How do I understand the true context of a Process?

Quality Scores

Business/ Quality Rules



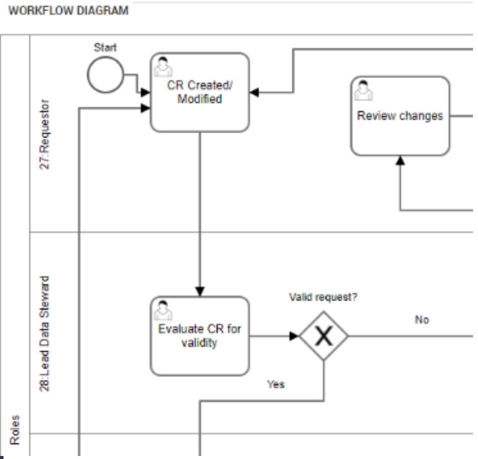
Stake holders



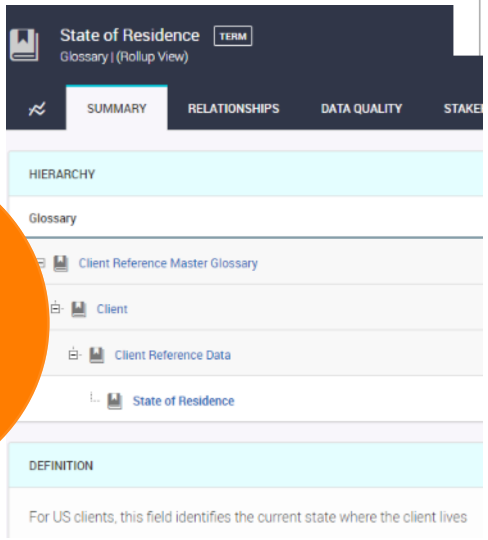
DIRECT STAKEHOLDERS

Role	Name	Org Unit
Governance Council	Jack B	Client Advisory Support
DQ Steward	Eva D	IT - Application Delivery
Data Owner	Kevin G	Client Advisory Support
Glossary Definition Owner	Kevin H	IT - Business Analysis & Support

Change Request


WORKFLOW DIAGRAM

Data elements



State of Residence TERM

Glossary | (Rollup View)

SUMMARY RELATIONSHIPS DATA QUALITY STAKE

HIERARCHY

Glossary

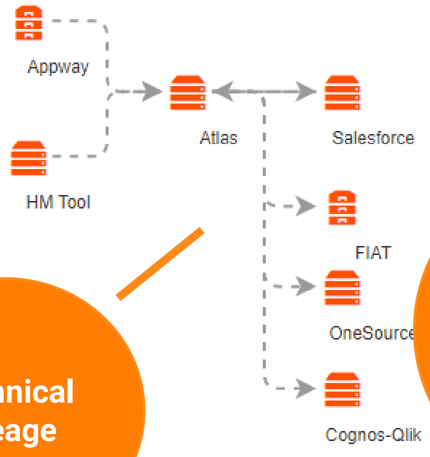
- Client Reference Master Glossary
  - Client
    - Client Reference Data
      - State of Residence

DEFINITION

For US clients, this field identifies the current state where the client lives

Systems & Business Lineage

Technical lineage



Informatica

Business value opportunities

- Capture knowledge in central repository instead of project artifacts
  - Reuse, collaborate, improve data literacy
  - Build and maintain data catalog
- Eliminate waste
  - Remove legacy systems and “shadow-IT”
  - Standardize technology choices
  - Replace Point2Point integrations by Hub&Spoke architecture
- Address data quality issues
- Optimize to benefit from cloud-native technologies and cost models

Experiences

Modernizing towards the Cloud

Joris Zwart
Senior Manager – Architecture
Practice, Informatica

Topics

- How to approach a Cloud modernization project
- What not to do
- The Cloud native mindset
- Further scope for modernizations
- Practical considerations

How to approach a Cloud modernization project

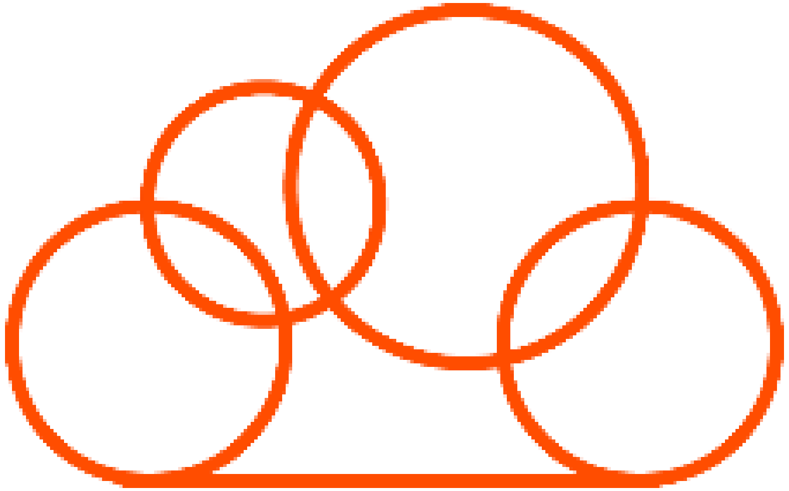
- The most important: Have a clear goal. What do you want to get out of a move to the Cloud?
  - Think both of what benefits you intend to get from the Cloud and what improvements you could make generally.
- Try to find a relatively isolated part of your data landscape that is representative and that can serve as a pilot
- As much as possible, plan to migrate coherent pieces to avoid too much data movement back and forth between Cloud and on-premise
- Try to move the metadata together with the data itself

What not to do

- Avoid lift & shift projects
- Don’t try to do a big bang; work in (fairly short) sprints
  - Think of an MVP that demonstrates the benefits of moving to the Cloud in the short term and whets people’s appetite for further migration/modernization
  - Don’t try to design everything all at once, but verify that your design can be expanded into something bigger.
  - Don’t assume that everything needs to be converted – you will probably have some developments that are obsolete
- Try not to make a maze between Cloud and on-premise systems

The Cloud native mindset

- Don’t worry too much where applications or data are physically located
  - Of course, you still need to mind the GDPR!
  - With bigger data streams you would need to consider performance implications
- Resources are dynamic
  - Don’t spend too much effort sizing systems – adding resources is easy
  - Don’t think of systems as static
  - Big data clusters (CDI-E, DataBricks, etc.) are easy and (relatively) cheap in the Cloud – consider them even where you would not have done so on-premise
- Think serverless – abstract the service away from the machine it runs on



Further scope for modernizations

Moving to the Cloud is often a major reengineering effort; think of what else you might want to improve at the same time

Migrating to different (native Cloud) applications and databases
ETL to ELT – see the next part of this presentation
Checking the quality of data to be moved and correcting it
Governance/cataloging
Moving to less or no coding
Real-time processing
API-based interfacing From flat files to XML/JSON

Practical considerations

- Performance between Cloud and on-premise (latency)
- Security is often a major topic (User-level, system-level)
- Network/internet access should not become a single point of failure
- Operating Costs – inefficiency can be hidden for longer on-premise but will incur immediate costs with SaaS
- Be pragmatic. It’s all about your requirements in the end, but make sure to distinguish what you need from what you think you need

Modernizing your analytics

And the impact on your DI landscape

Rick Mutsaers
Principal Architect

Informatica

Topics

- What to consider when moving DI to the cloud?
  - Deployment options: from ETL to ELT to distributed processing…
    Why is a different approach needed and which one to choose?
  - If data model changes, how useful are your data pipelines?
    (re-architect versus re-platform)
  - Optimizations due to newer DI tool capabilities. Maybe replace some simple data pipelines by pattern driven pipelines
- Approach to moving DI to cloud. Aka What does it take?
- Informatica’s Migration Factory approach, how does that work and how does it benefit.

Main trends in analytics drive cloud modernization

Scale

- More analytics projects are needed to supply business with the right info for decision making
- More data needs to be collected and processed, existing on-prem solutions are not scalable and cost effective enough

Agility & elasticity

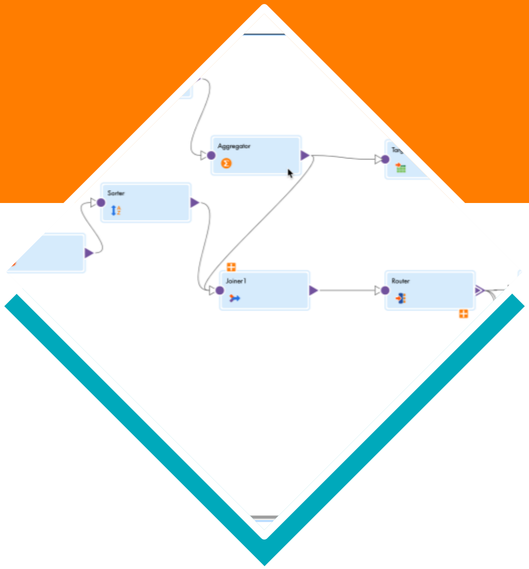
- Business needs answers fast and can't wait months for a new datamart to be created
- On-premise architectures simply not suited for agile development
- Processing demands vary across time so elastic solution that can scale up and down is needed

Cost control

- In on-premise world, infrastructure is scaled based on largest workloads so most of time capacity is unused, can't benefit from economies of scale.
- On-premise storage is expensive, cloud storage much cheaper at scale

This drives modernization to the cloud

Analytics modernization is more than migrate to CDW



DB migration

Modernize your on-premise Data warehouse to modern cloud datawarehouse

ETL migration

Modernize your ETL data pipelines into modern ELT/ETL/Spark data pipelines

Analytics migration

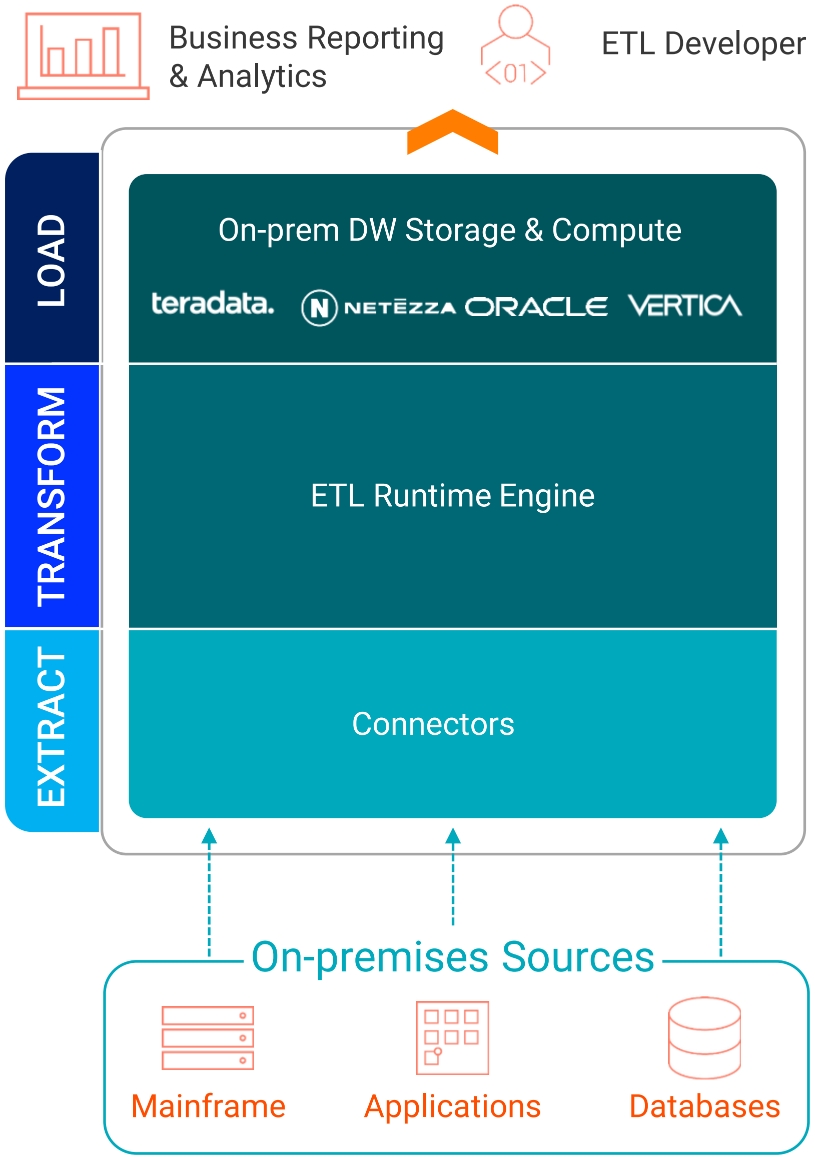
Modernize your reports and dashboards to cloud native analytics solution

Hardware migration

Modernize your hardware and storage to cloud managed virtual machines and storage

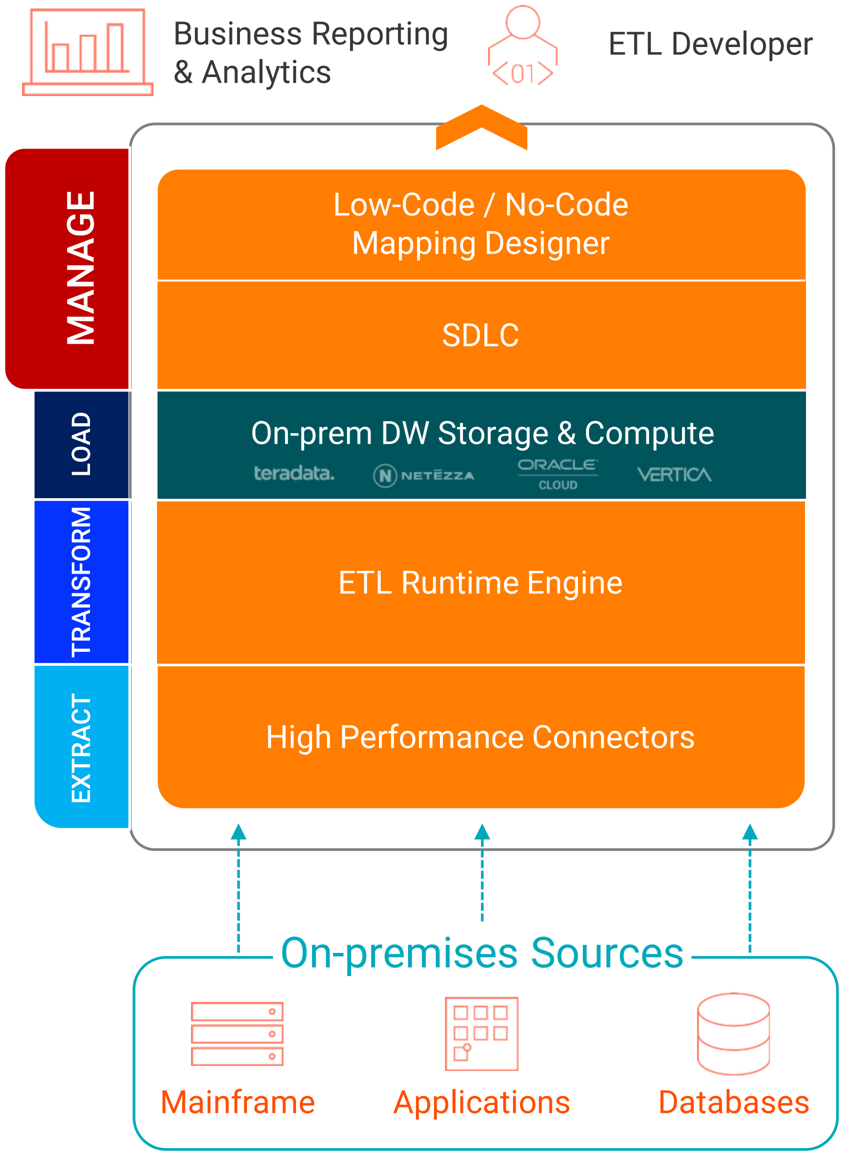
Why do we need a different way of integrating data?

Old World traditional ETL for star schema



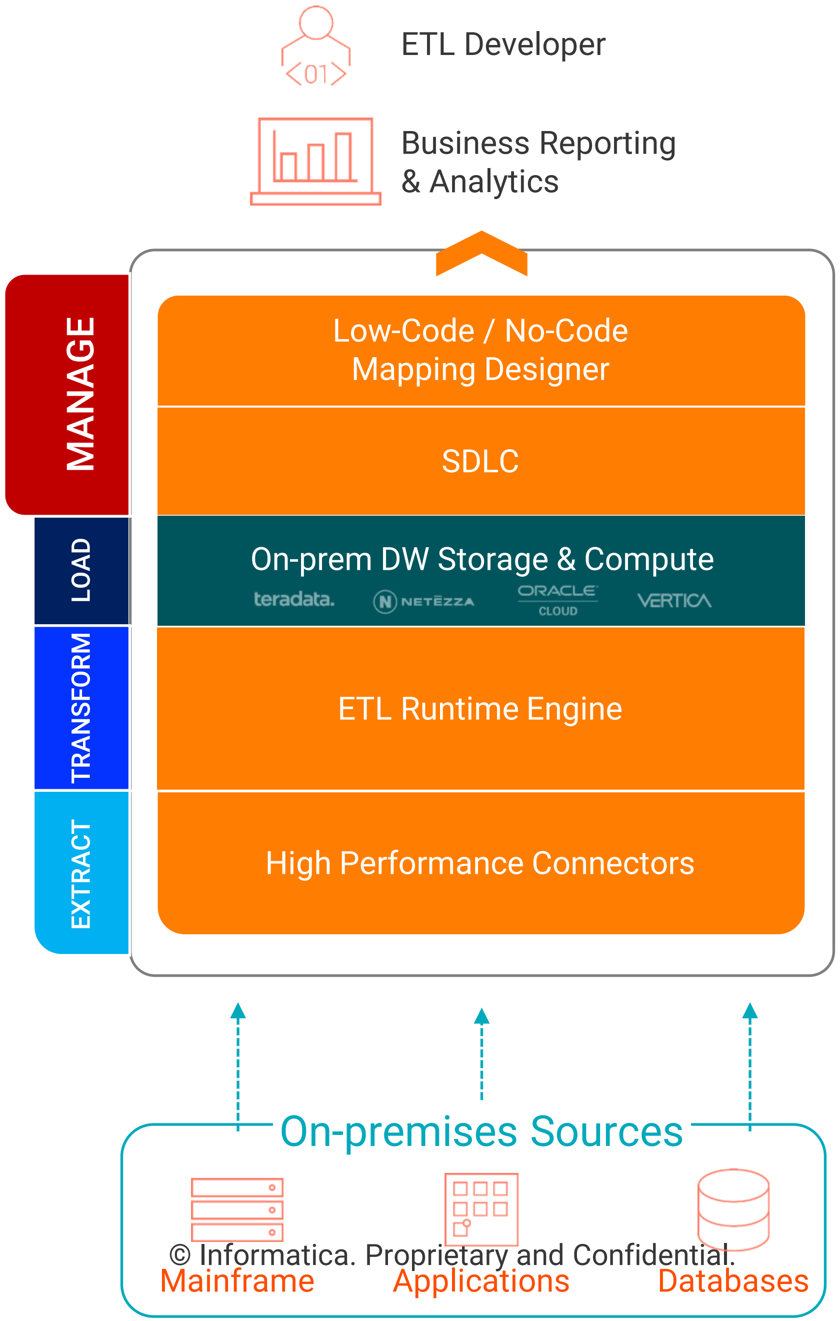
Why do we need a different way of integrating data?

Old World traditional ETL for star schema – Informatica's approach



Informatica

Informatica for Traditional ETL for Star Schema



ETL Developer

Business Reporting & Analytics

MANAGE

LOAD

TRANSFORM

EXTRACT

Low-Code / No-Code Mapping Designer

SDLC

On-prem DW Storage & Compute

teradata. NETEZZA ORACLE CLOUD VERTICA

ETL Runtime Engine

High Performance Connectors

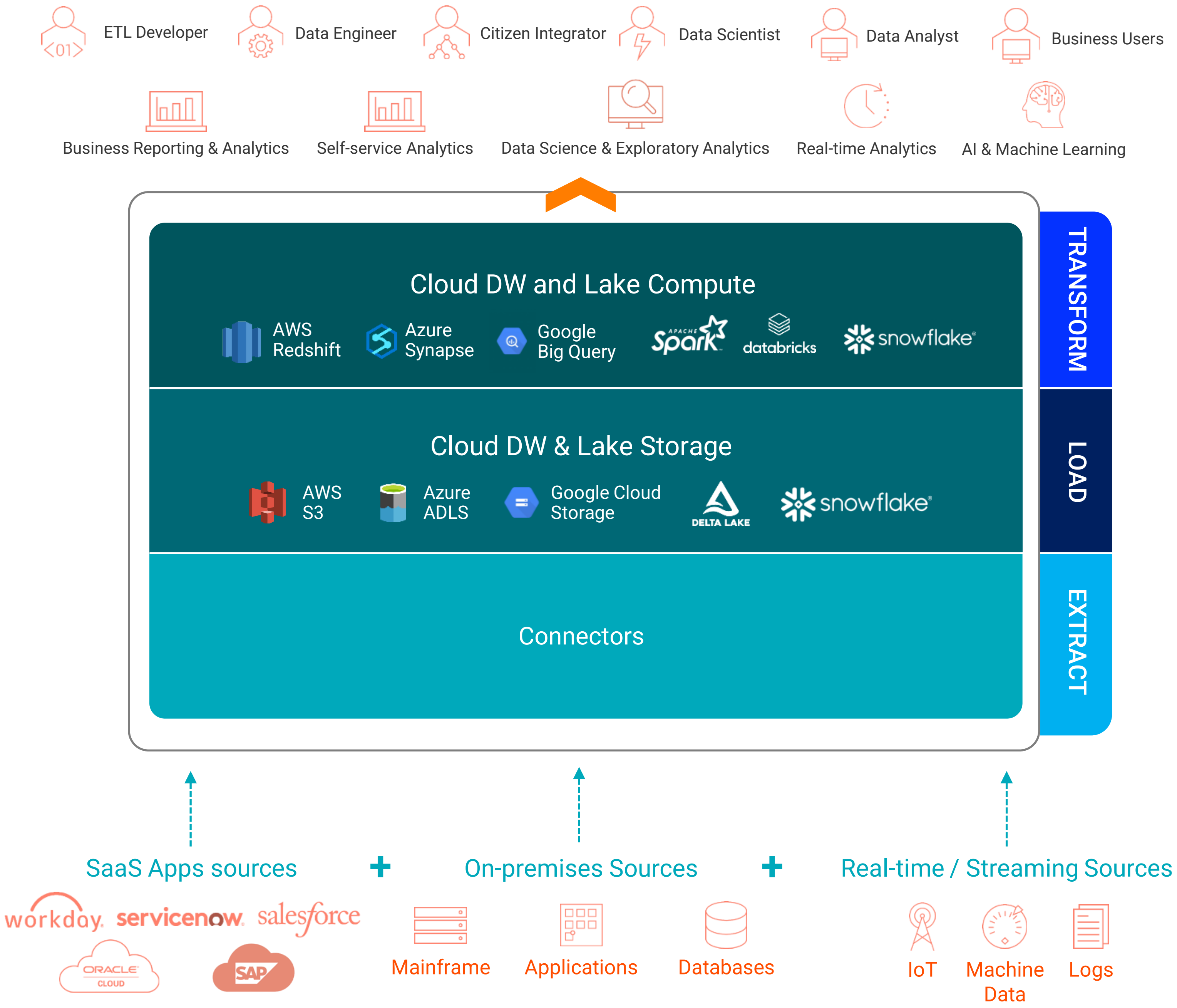
On-premises Sources

Mainframe

Applications

Databases

New World Modern Cloud ELT for Cloud DW/L



ETL Developer

Data Engineer

Citizen Integrator

Data Scientist

Data Analyst

Business Users

Business Reporting & Analytics

Self-service Analytics

Data Science & Exploratory Analytics

Real-time Analytics

AI & Machine Learning

Cloud DW and Lake Compute

AWS Redshift

Azure Synapse

Google Big Query

Apache Spark

databricks

snowflake

Cloud DW & Lake Storage

AWS S3

Azure ADLS

Google Cloud Storage

DELTA LAKE

snowflake

Connectors

TRANSFORM

LOAD

EXTRACT

SaaS Apps sources + On-premises Sources + Real-time / Streaming Sources

workday. servicenow. salesforce ORACLE CLOUD SAP

Mainframe

Applications

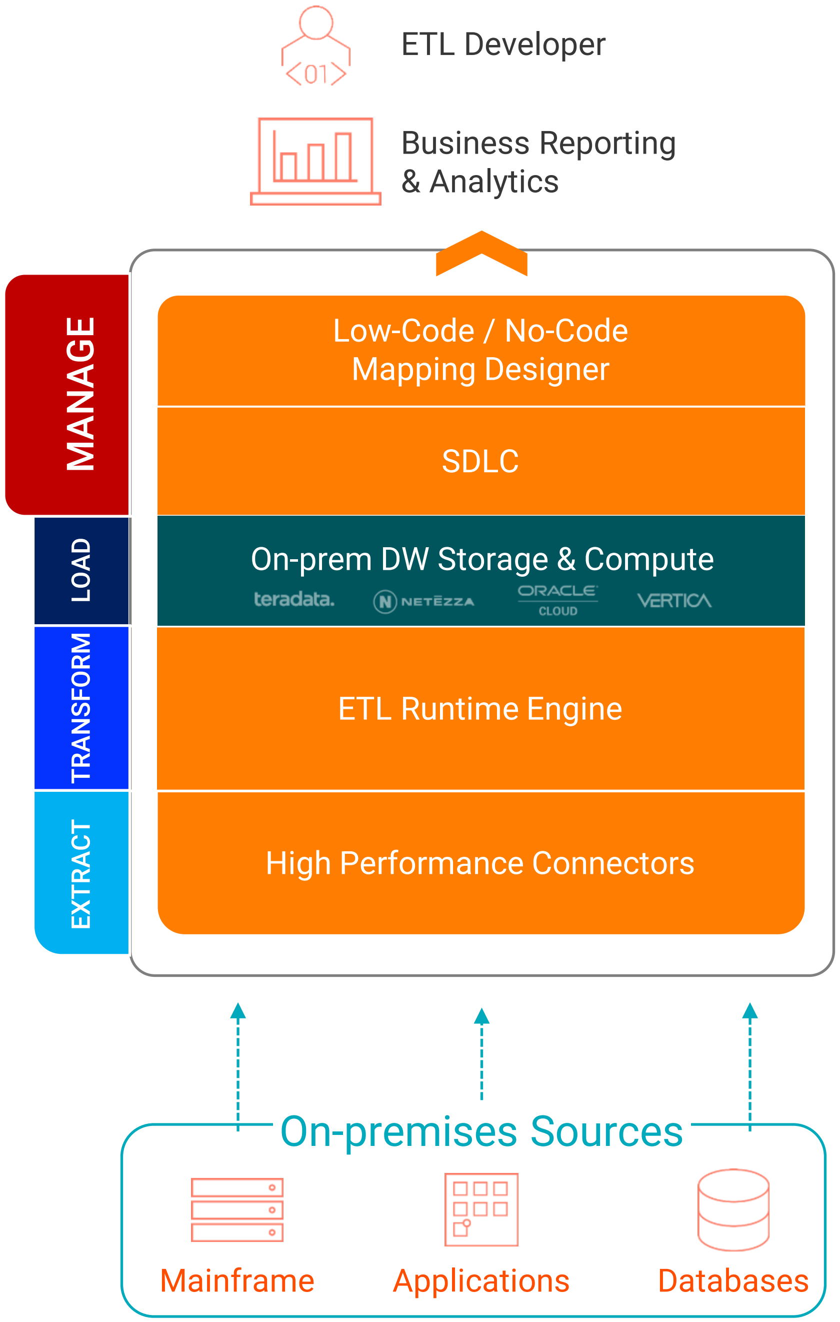
Databases

IoT

Machine Data

Logs

Informatica for Traditional ETL for Star Schema



ETL Developer

Business Reporting & Analytics

MANAGE

LOAD

TRANSFORM

EXTRACT

Low-Code / No-Code Mapping Designer

SDLC

On-prem DW Storage & Compute

teradata. NETEZZA ORACLE CLOUD VERTICA

ETL Runtime Engine

High Performance Connectors

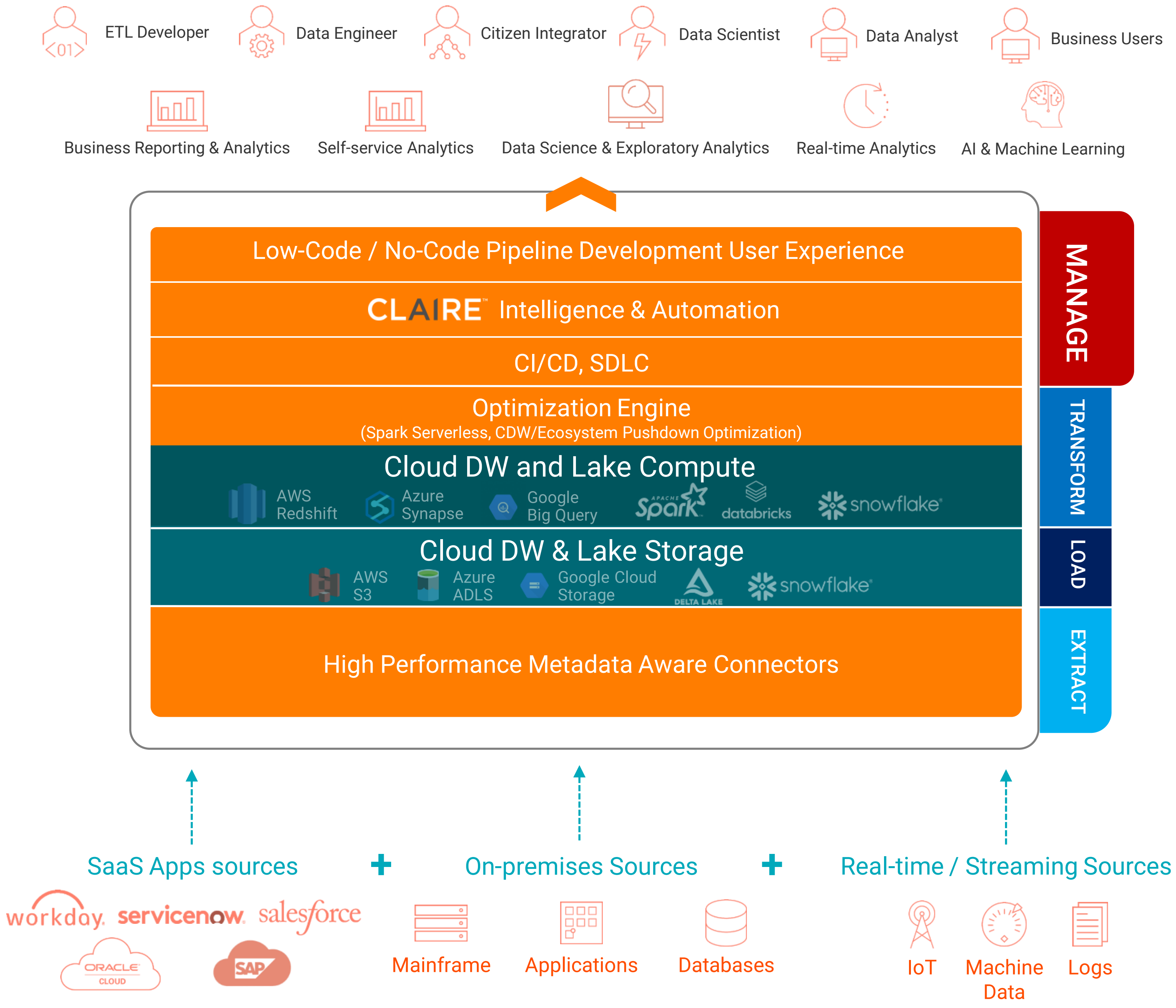
On-premises Sources

Mainframe

Applications

Databases

Informatica for Modern Cloud ELT for Cloud DW/L



ETL Developer

Data Engineer

Citizen Integrator

Data Scientist

Data Analyst

Business Users

Business Reporting & Analytics

Self-service Analytics

Data Science & Exploratory Analytics

Real-time Analytics

AI & Machine Learning

Low-Code / No-Code Pipeline Development User Experience

CLAIRE Intelligence & Automation

CI/CD, SDLC

Optimization Engine

(Spark Serverless, CDW/Ecosystem Pushdown Optimization)

Cloud DW and Lake Compute

AWS Redshift

Azure Synapse

Google Big Query

Spark

databricks

snowflake

Cloud DW & Lake Storage

AWS S3

Azure ADLS

Google Cloud Storage

DELTA LAKE

snowflake

High Performance Metadata Aware Connectors

MANAGE

TRANSFORM

LOAD

EXTRACT

SaaS Apps sources

workday. servicenow. salesforce

ORACLE CLOUD

SAP

+

On-premises Sources

Mainframe

Applications

Databases

+

Real-time / Streaming Sources

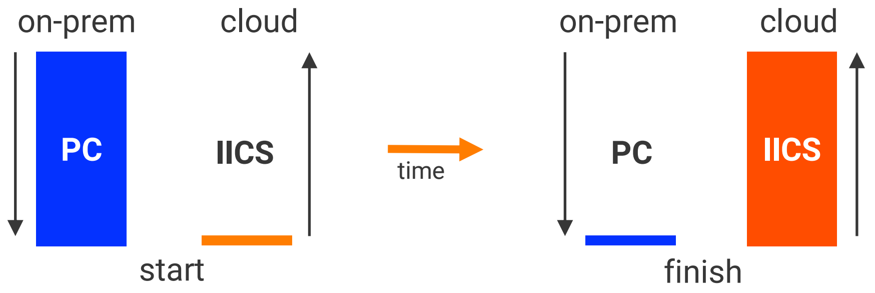
IoT

Machine Data

Logs

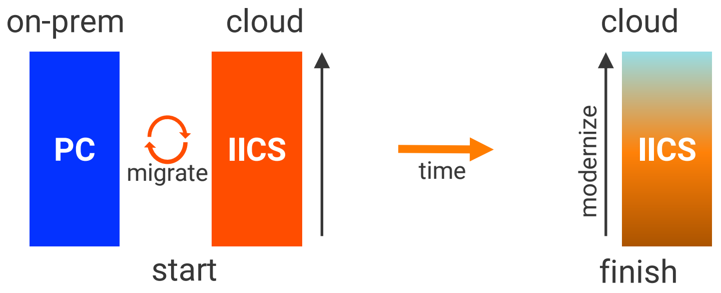
Modernization options

Start from scratch or re-use what was already built?



Re-architect

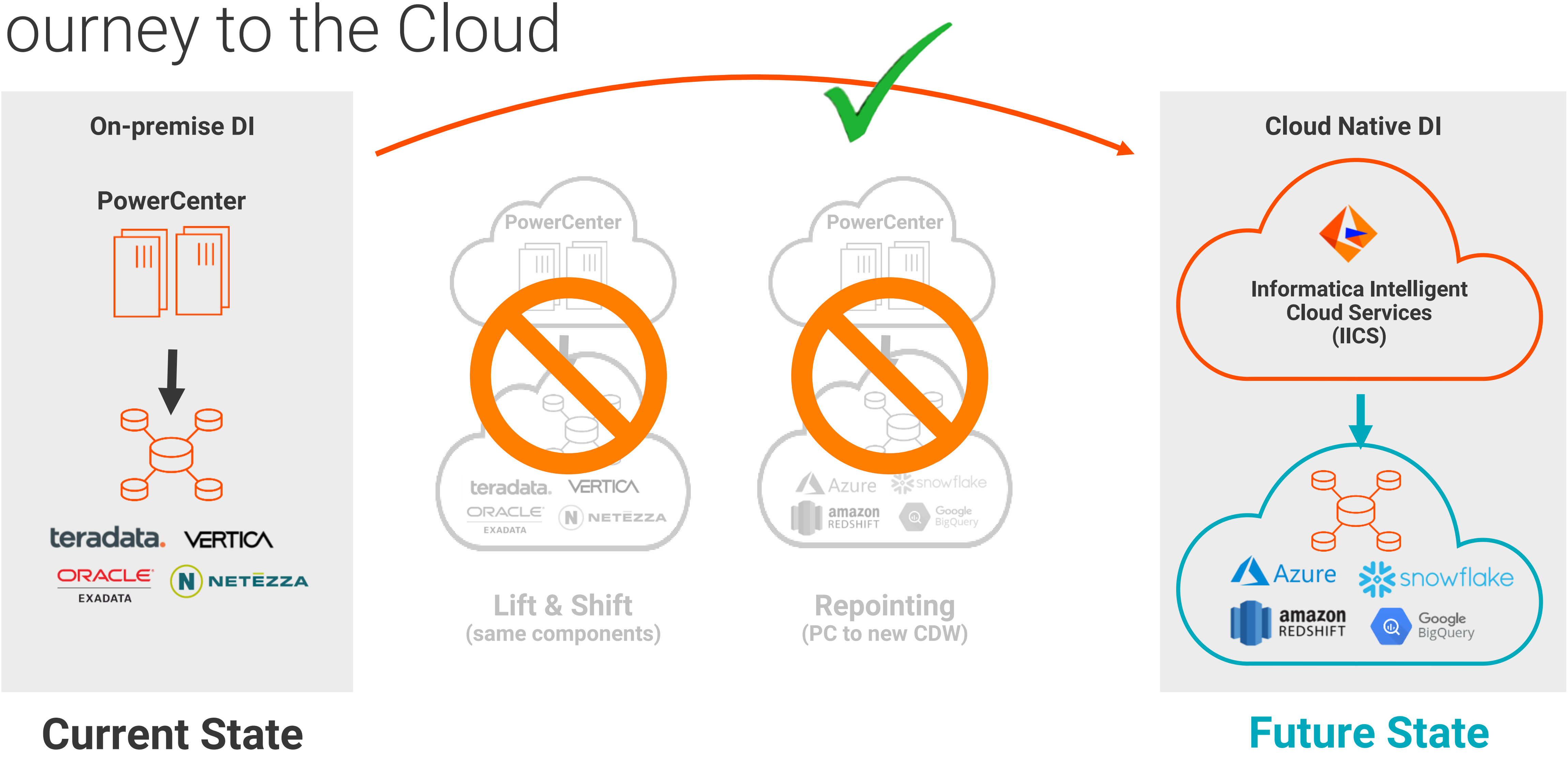
- Pro
  - Start fresh, fix modeling issues
- Con
  - Very time consuming (consider how long it took to build current architecture)
  - High cost
  - New tools = new challenges
  - Double cost during development of new solution



Re-platform

- Pro
  - Get rid of current infrastructure **fast**, avoid double cost
  - **Impact** to business **low** since reports/dashboards don’t change, only underlying technology changes
  - Build experience with new tech before modernize
  - Prioritize modernization without double cost
- Con
  - Migration cost might be unexpected
  - Takes a bit longer to modernize (because you migrate first)

Journey to the Cloud



On-premise DI

PowerCenter

Lift & Shift
(same components)

Repointing
(PC to new CDW)

Cloud Native DI

Informatica Intelligent Cloud Services (IICS)

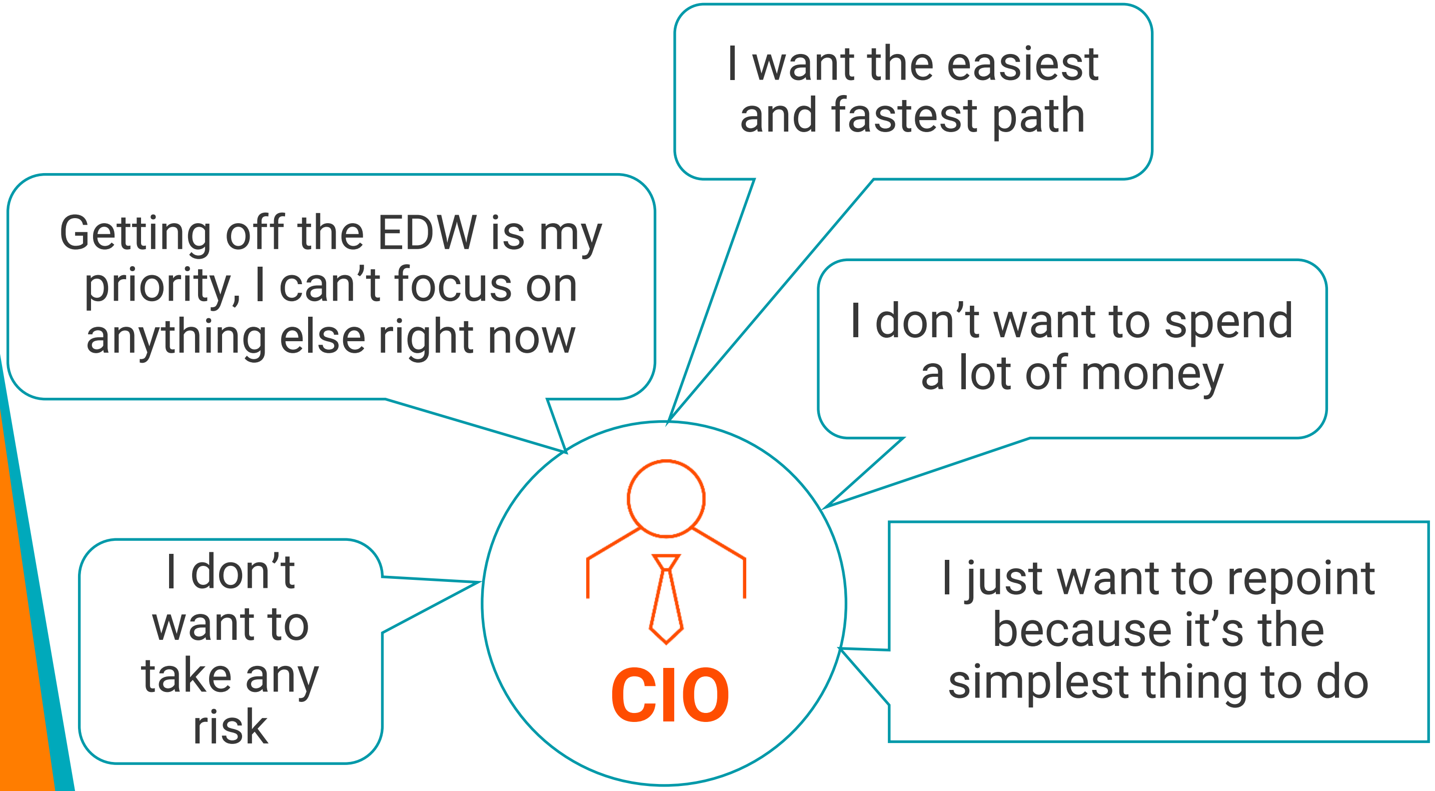
Current State

Future State

Informatica

The myth of repointing PowerCenter to the cloud for Cloud Data Warehouse Modernization

What customers say/expect



I want the easiest and fastest path

Getting off the EDW is my priority, I can't focus on anything else right now

I don't want to spend a lot of money

I don't want to take any risk

I just want to repoint because it's the simplest thing to do

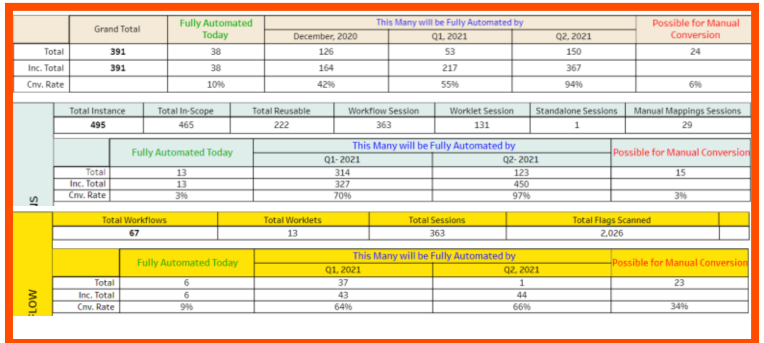
CIO

The reality

- It's much more **costly** to run than Cloud Native
- Lots of **manual work** to change mappings
- It's an **interim step** - you are effectively migrating twice
- No data lake or advanced modernization support
- The time to value is **slower** when repointing

Informatica’s approach to PowerCenter migration

Assess



	Grand Total	Fully Automated Today	This Many will be Fully Automated by			Possible for Manual Conversion
			December, 2020	Q1, 2021	Q2, 2021	
Total	**391**	38	126	53	150	24
Inc. Total	**391**	38	164	217	367	
Cnv. Rate		10%	42%	55%	94%	6%

Total Instance	Total In-Scope	Total Reusable	Workflow Session	Worklet Session	Standalone Sessions	Manual Mappings Sessions
495	465	222	363	131	1	29

	Fully Automated Today	This Many will be Fully Automated by		Possible for Manual Conversion
		Q1- 2021	Q2- 2021	
Total	13	314	123	15
Inc. Total	13	327	450	
Cnv. Rate	3%	70%	97%	3%

Total Workflows	Total Worklets	Total Sessions	Total Flags Scanned
67	13	363	2,026

	Fully Automated Today	This Many will be Fully Automated by		Possible for Manual Conversion
		Q1, 2021	Q2, 2021	
Total	6	37	1	23
Inc. Total	6	43	44	
Cnv. Rate	9%	64%	66%	34%

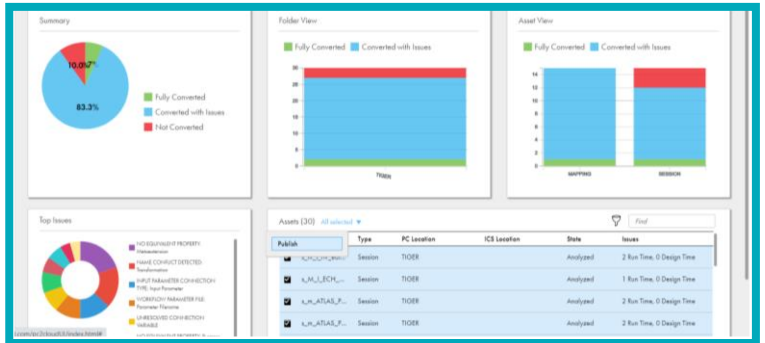
Analyzes all PowerCenter mappings, sessions and workflows

Identifies which assets can be fully or partially migrated to Informatica Cloud

Provides basis for level of effort required for migration

Intelligence

Migrate



	Type	PC Location	ICS Location	State	Issues
☑ [illegible]	Session	TIOER		Analyzed	2 Run Time, 0 Design Time
☑ s_M_I_ECH_...	Session	TIOER		Analyzed	1 Run Time, 0 Design Time
☑ s_m_ATLAS_F...	Session	TIOER		Analyzed	2 Run Time, 0 Design Time
☑ s_m_ATLAS_F...	Session	TIOER		Analyzed	2 Run Time, 0 Design Time

Performs initial migration data load (optional)

Migrates PowerCenter assets to IICS and repoints mappings from EDW to CDW

Unit testing of migrated assets

Automation

Benefits of our Informatica Migration Factory approach

Reduced Migration Time

- Automated Migration utilities
- Migration Factory approach
- Reduced Time-To-Market

Reduced Migration Cost

- Reduced time = less cost
- Higher quality code = less rework
- Less development resources

Reduced Migration Risk

- Code generation = less errors
- Predefined scope = less scope creep
- Shorter duration = less scheduling challenges

Summary

Metadata Driven
Identify dependencies to plan your Cloud Modernization Journey
1

Performance
Design to truly enable the scalability & elasticity in the cloud
2

Design before execution
Data Quality and Security as Key Design Principles
3

Smart Automation
Automated Conversion of Pipelines
4

Thank You!

Informatica®