Tech Talk Series

PowerCenter Users: Architect Your Future to the Cloud

Session 1 of 4:

How to Overcome Cloud Challenges with a Cloud Native Solution

Waqar Ahmed
EMEA-LA Segment Leader, Data Warehouse/Lake & App Modernization, Informatica

Top of Mind

Modernize & Standardize Integration

On average an organization deploys 4-6 integration tools serving different integration patterns

Productivity for ALL Users

Citizen Integrators, Low Code/No-Code, AI-Powered Automation

Vendor Neutrality & Scale

Integration is sticky & leads to vendor lock in. Don't let your integration platform dictate your future architecture

Microservices-based

API driven

Multi-tenant

Multi-cloud

"Integration is the backbone of your architecture and should act as an enabler of digital transformation"

Cloud-Native Data Management Critical for your Warehouse, Lake -- Challenges Persist

Cost Overruns

of organizations using cloud data management will encounter budget overruns resulting in their questioning the value of using cloud services

Resource Constraints

year-over-year growth in the number of open data engineering positions

Complexity

of organizations are still struggling to operationalize AI within their enterprise

Source: 1– Aptum cloud impact study | 2– Advanced Global Research, May 28, 2020 | 3- Venturebbeat

Informatica

Problem #1: Cost Overruns

80%

of organizations **overshoot** their cloud budgets due to **ineffective cloud cost management**

Source: Cloud Cost Management, LLC

What we hear from customers:

- Difficulty managing compute costs
- Lack of control of users and usage
- Easy access to add more compute
- Increasing data transfer costs

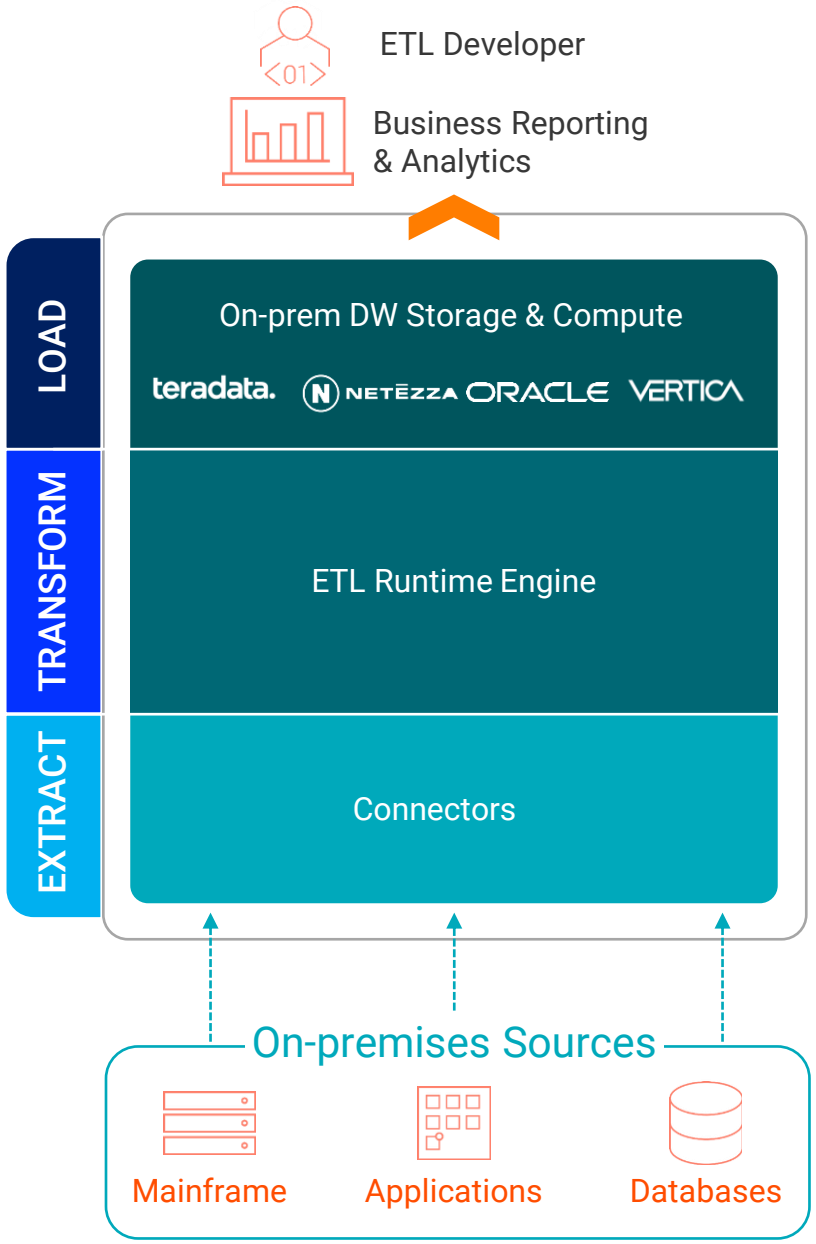
The solution:

Optimization Engine

Intelligent Cost Controls

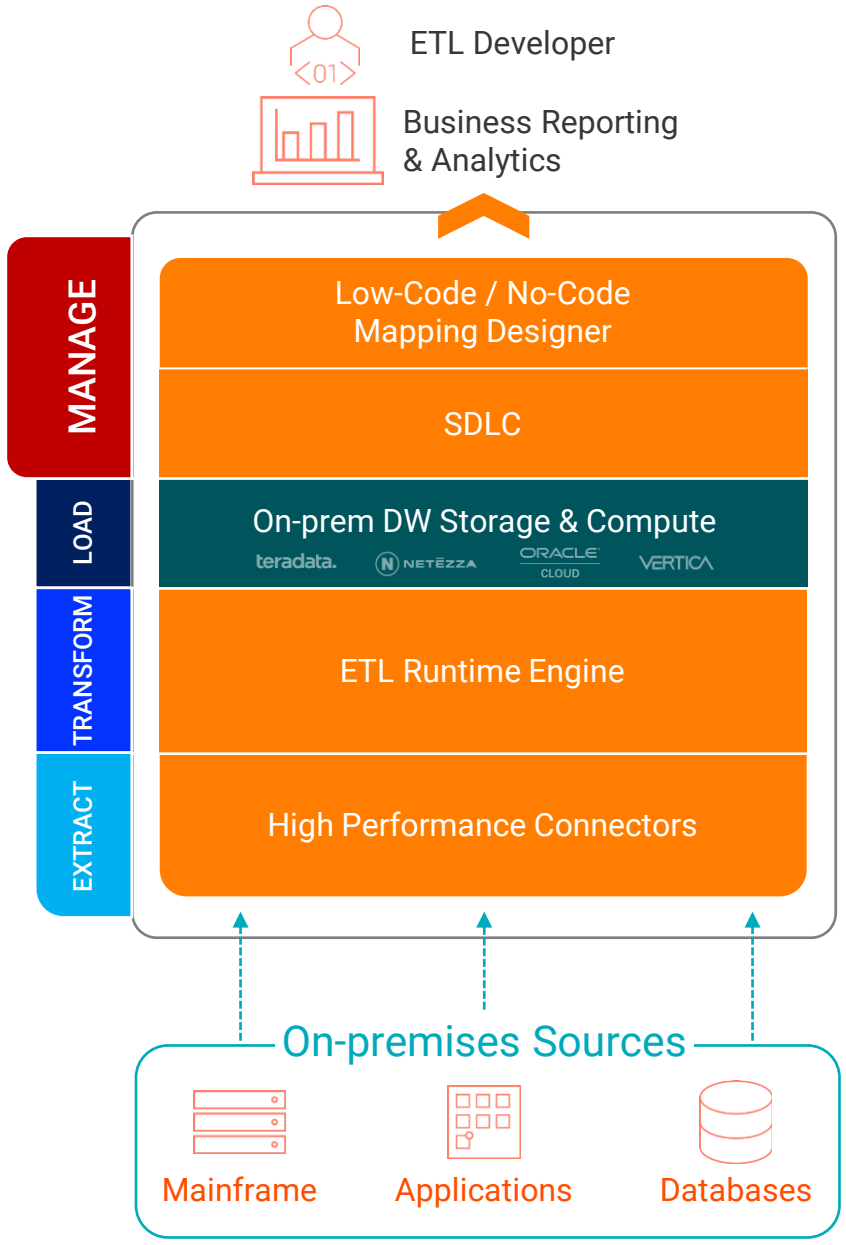
Consumption-based Pricing

Old World Traditional ETL for Star Schema



Informatica

Informatica for Traditional ETL for Star Schema



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

TRANSFORM

LOAD

EXTRACT

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

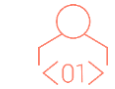
SaaS Apps sources + On-premises Sources + Real-time / Streaming Sources

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

MANAGE

TRANSFORM

LOAD

EXTRACT

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

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

workday. servicenow. salesforce

ORACLE CLOUD SAP

Mainframe Applications Databases

IoT Machine Data Logs

Use Our Optimization Engine to Control Cost Overruns

Data Transfer Costs

Costs associated with getting data in and out of CDW's or public Clouds

Compute Hours

Costs associated with the number of servers and amount of time the servers are used

Most cloud data integration uses Spark

Informatica offers FASTER Spark and More Cost-Efficient Processing

	Informatica sends the work to a large number of Spark servers	Informatica sends the work to a Cloud Ecosystem	Informatica sends the work to a Cloud Data Warehouse	The work is done on an Informatica server
Execution	Serverless Processing Auto-Scaling Elastic Spark Cluster	Ecosystem Pushdown Optimization ELT	CDW Pushdown Optimization ELT	Traditional ETL

aws · Google Cloud · Azure · Azure Synapse Analytics · Google Big Query · amazon REDSHIFT · snowflake · Informatica

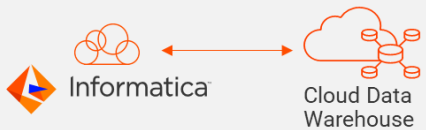
Optimization Engine

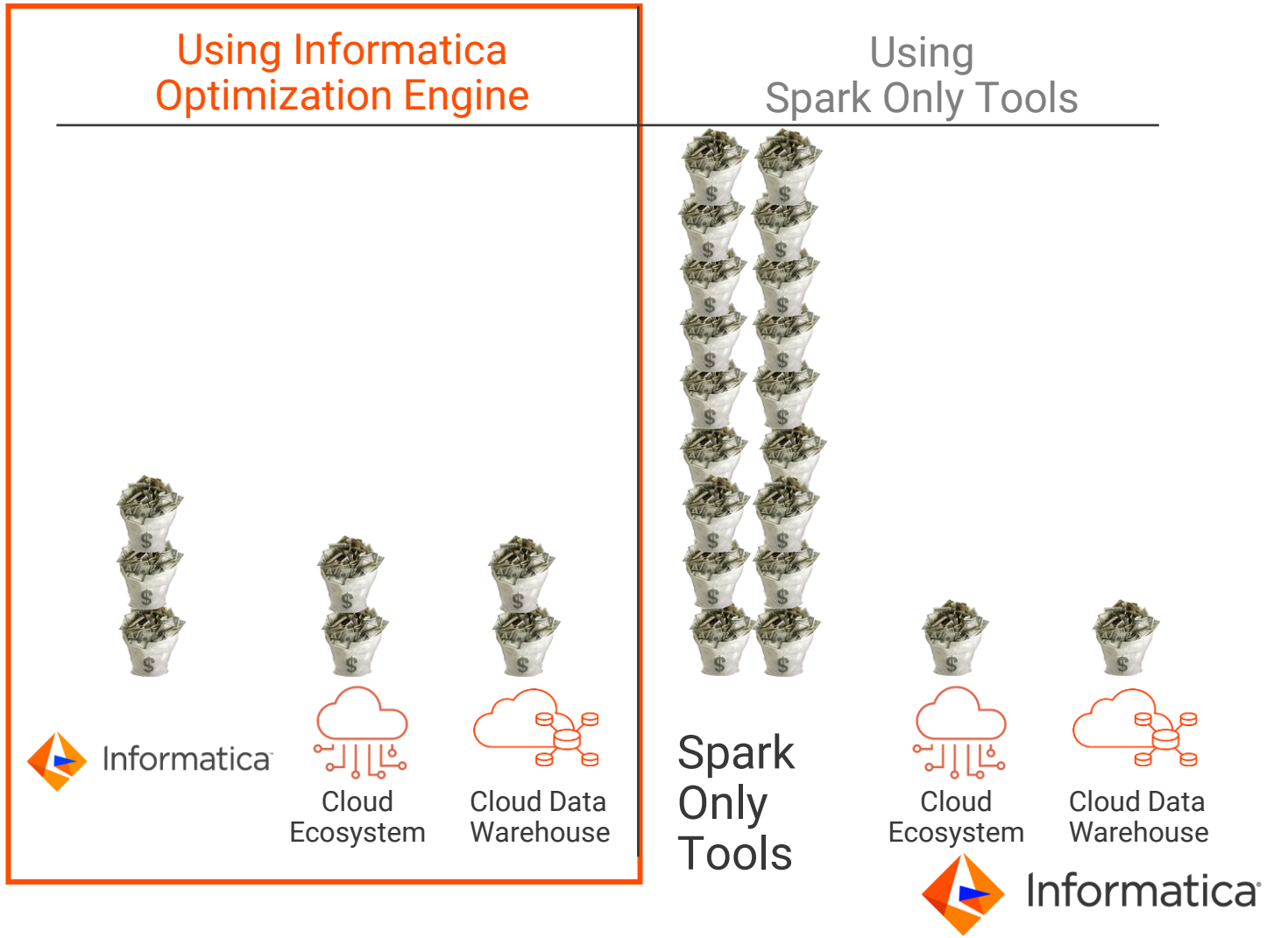
Informatica Sends the Work to the Most Cost Effective Option

Pushdown

Spark

Traditional ETL

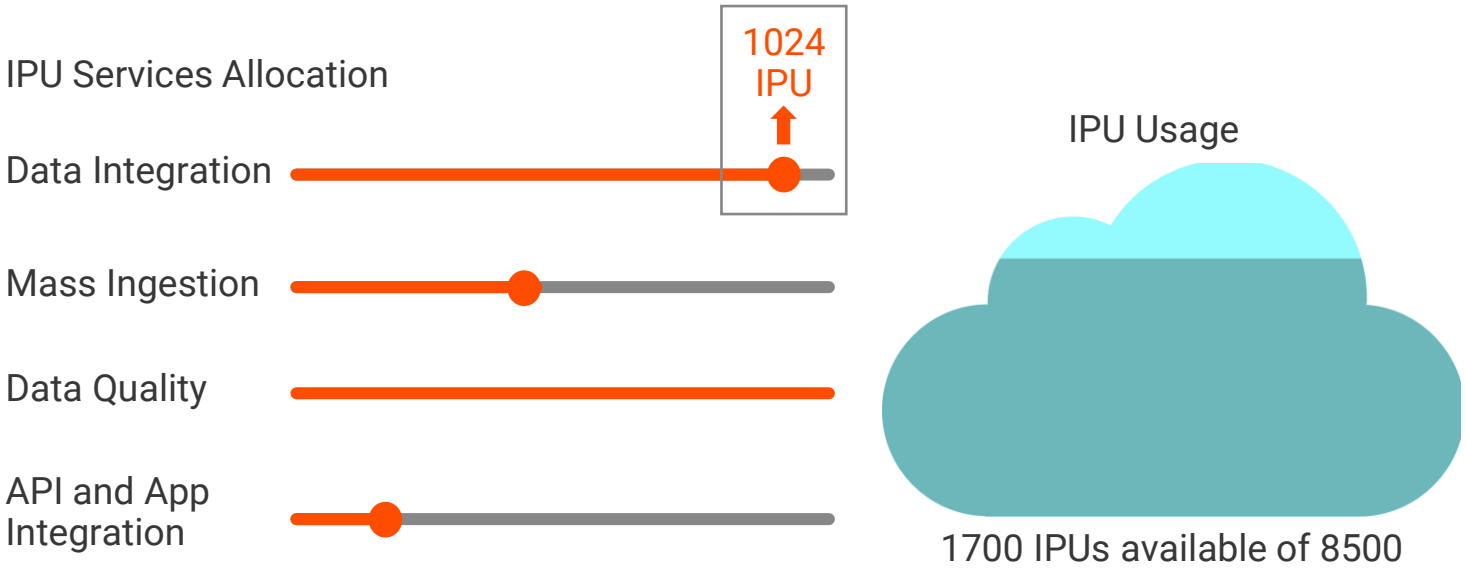




Get access to all our cloud services

Dynamically scale usage up/down to fit your needs

Use AWS/Azure/Google Cloud Pre-Commit Dollars



1. Our pricing is Usage-based
   The cost of a service is based on how much the service is used
2. You have Full Access
   All cloud native services
3. Use Any Service at Any Time
   Every month, change services and usage as needed

Informatica

Problem #2: Resource Constraints

87%

of organizations experiencing or expecting to face talent shortage in software development, including data & analytics talent

Source: Daxx, A Software Development Outsourcing Company based on Europe

What we hear from customers:

- Not enough resources & 2 or more resources for every job
- Difficulty finding specialized skills
- Lack of automation
- Lack of tools for non-tech users

The solution:

Self-Service

Bundles, Templates, Wizards

Intelligent Automation

Self Service: GUI Development vs. Hand-Coding

Spark Code Example

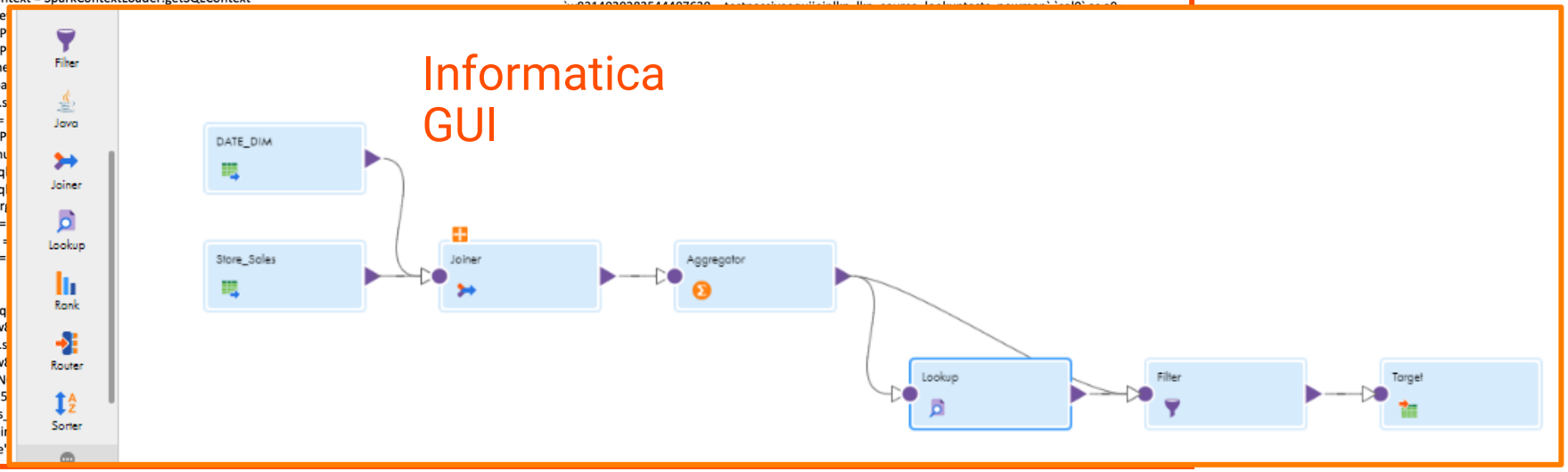
```
import org.apache.spark._
import org.apache.spark.rdd._
import org.apache.spark.storage.StorageLevel._
import org.apache.spark.sql._
import org.apache.spark.sql.types._
import org.apache.spark.sql.functions._
import org.apache.spark.sql.functions.*
import java.io._
import java.sql.Timestamp
import scala.reflect.ClassTag
import org.apache.spark.sql.catalyst.expressions.Caster
import org.apache.spark.sql.catalyst.expressions.JavaCaster
import org.apache.spark.util.LongAccumulator
import org.apache.spark.scheduler.SparkListener
import org.apache.spark.SparkEnv
import org.apache.spark.sql.Row

object Spark0 {
 def main(s:Array[String]) {
   val sc = SparkContextLoader.getSparkContext
   val sqlContext = SparkContextLoader.getSQLContext
   val ls = ne
   ls.relay(JP
   ls.relay(JP
   val lis = ne
   sc.addSpa
sqlContext.s
   val accs =
   ls.relay(JP
   lis.accumu
   import sq
   import sq
   import or
   val icast =
   val acast =
   val jcast =

   try {
   Tuple2(sq
`default`.`w
sqlContext.s
`default`.`w
`col2` STRIN
'hdfs://cdh5
ookupTests_
'pwx.mappi
'auto.purge'
```

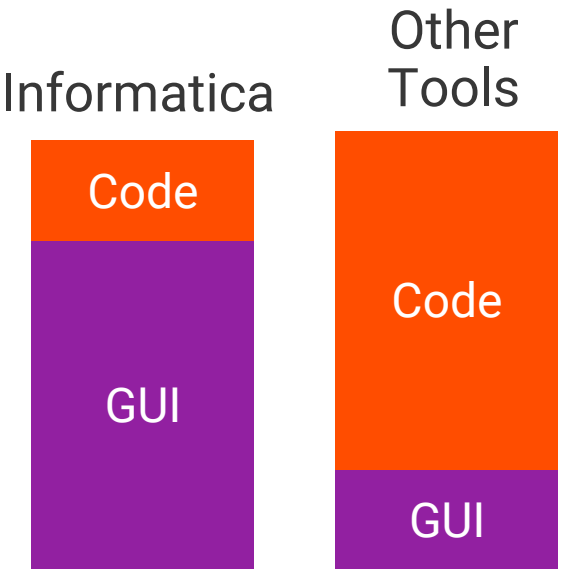
```
   Tuple2(sqlContext.sql(Params.resolve("DROP TABLE IF EXISTS
`default`.`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap`",
"MM/DD/YYYY HH24:MI:SS")), sqlContext.sql(Params.resolve("CREATE TABLE
`default`.`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap` (`col0`
DECIMAL(18, 2), `col1` STRING, `col2` STRING) ROW FORMAT
SERDE  hdfs://cdh52.vm.com:8020//tmp/sess8314020283544407639//W8314020283544407639_COPY
_testPassiveEquiJoinLkp_lkp_source_LookupTests_newMap' TBLPROPERTIES
('.columns.types'='decimal(18,2),string,string',
'pwx.mapping.file.path'='./testPassiveEquiJoinLkp_lkp_source_MAPPING_1204153312640770_1204153
976631823.bin', 'auto.purge'='true', '.columns'='col0,col1,col2')", "MM/DD/YYYY HH24:MI:SS")));
   Tuple2(sqlContext.sql(Params.resolve("DROP TABLE IF EXISTS
`default`.`w8314020283544407639__testpassiveequijoinlkptgt_lookuptests_newmap`", "MM/DD/YYYY
HH24:MI:SS")), sqlContext.sql(Params.resolve("CREATE TABLE
`default`.`w8314020283544407639__testpassiveequijoinlkptgt_lookuptests_newmap` (`col0` STRING,
`col1` STRING) ROW FORMAT LOCATION
'hdfs://cdh52.vm.com:8020//tmp/sess8314020283544407639//W8314020283544407639__testPassive
EquiJoinLkpTgt_LookupTests_newMap' TBLPROPERTIES ('.columns.types'='string,string',
'pwx.mapping.file.path'='./testPassiveEquiJoinLkpTgt_MAPPING_1204153304711020_12041540159301
76.bin', 'pwx.skip.serialization'='true', 'auto.purge'='true', '.columns'='col0,col1')", "MM/DD/YYYY
HH24:MI:SS")));
   val v0 = updatePartitions(asBlock(sqlContext.sql(Params.resolve("SELECT
```

Informatica GUI



Hand coding vs Informatica

Sales by geography for every month in the financial year



Informatica

Self-Service with Simple, Role-based Tools

No Code/ No Build

Developer Canvas

Multi-Step Wizards

Reusable Templates

Operations Tools

Architect

Data Engineer

ETL Developer

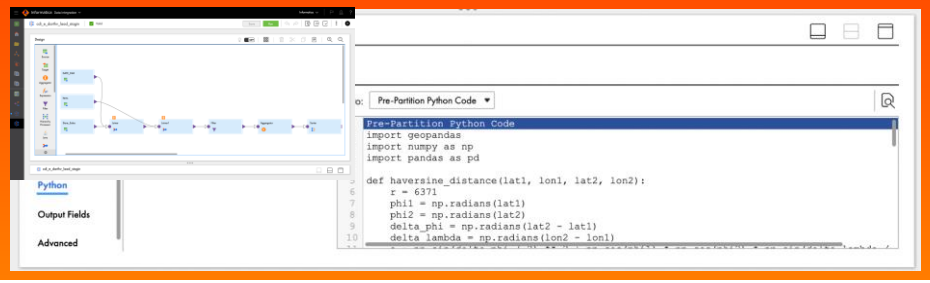
Data Scientist

IT Specialist

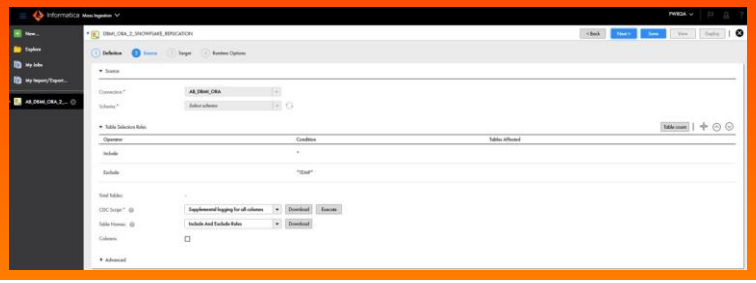
Business Analyst

IT Operations

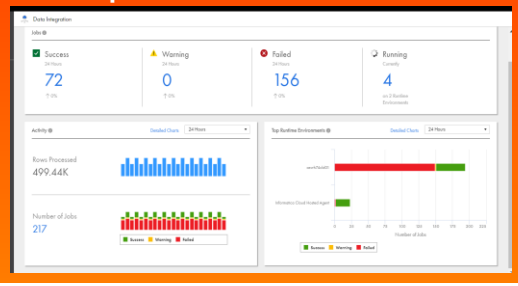
Developer GUI Canvas & Provision to Add Code



Analyst Wizard



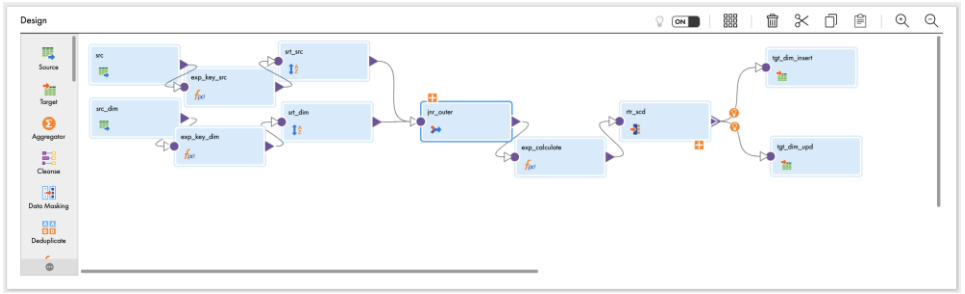
IT Operations Dashboard



Single pane of glass across all services

Self Service with Serverless

Start developing your data pipelines on day 1



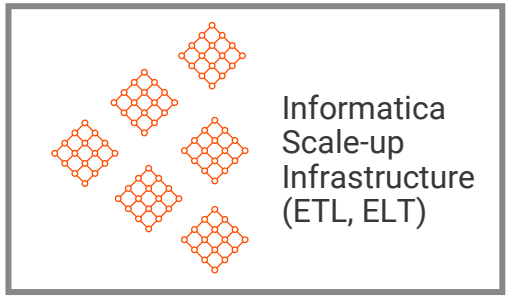
Developer

Analyst

Data Engineer

Open Informatica Design Canvas. Start Developing

- ✗ Talk to administrator
- ✗ Pick OS version
- ✗ Select hardware
- ✗ Define configuration options
- ✗ Spin up cluster
- ✗ Ramp down cluster



Informatica takes care of your Secure Compute Infrastructure

- ✓ Provision infrastructure
- ✓ Patch
- ✓ Upgrade
- ✓ Built-in unlimited scale
- ✓ High Availability
- ✓ Enterprise-grade resiliency

Informatica

Accelerators can reduce development by up to 80%

Multi-Step Wizards for Data Ingestion & Integration

Quick Start Bundles for Common Use Cases

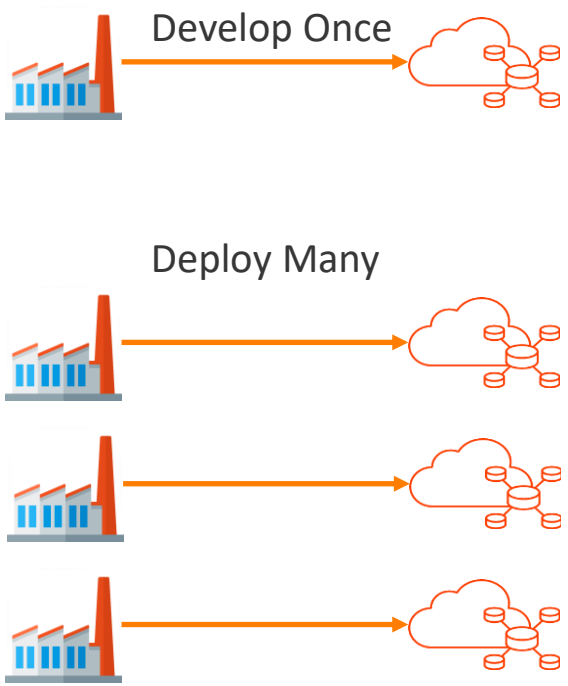
Create Custom Templates with Dynamic Mappings

4 STEP WIZARD

- Ingestion
- Replication
- Streaming
- Synchronization
- Masking

35+ PRE-BUILT

- Data Warehousing
- Data Quality
- Data Migration
- Snowflake
- Application Integration



Informatica

Data Integration Development

Intelligent Automation vs. Manual Design

Informatica

Hand-Coding & Other Tools

Use AI & ML to Automate Development:

- Claire recommends how to build a mapping
- Claire automatically recognizes different file formats
- Build once, deploy many reusable mappings

No AI & ML, throw bodies at the problem:

- Start every job from scratch
- Build each job manually
- No reusability

Problem #3: Complexity with technology & operationalization

80%

of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization

Source: DZone, October 2020

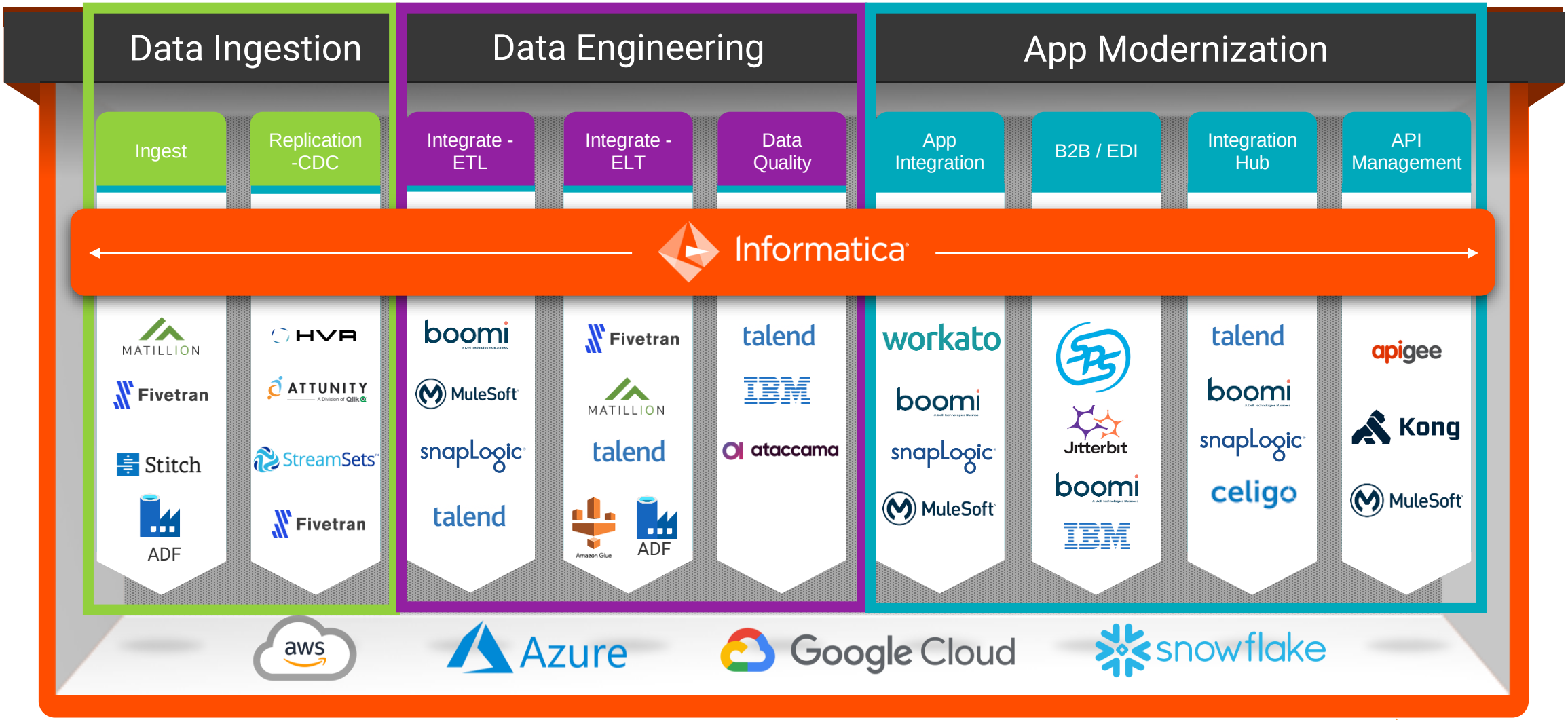
What we hear from customers:

- Disjointed tools that don't work together
- Development to production is arduous/costly
- Cloud and on-prem deployments don't interoperate
- Constantly evaluating new technologies

The solution:

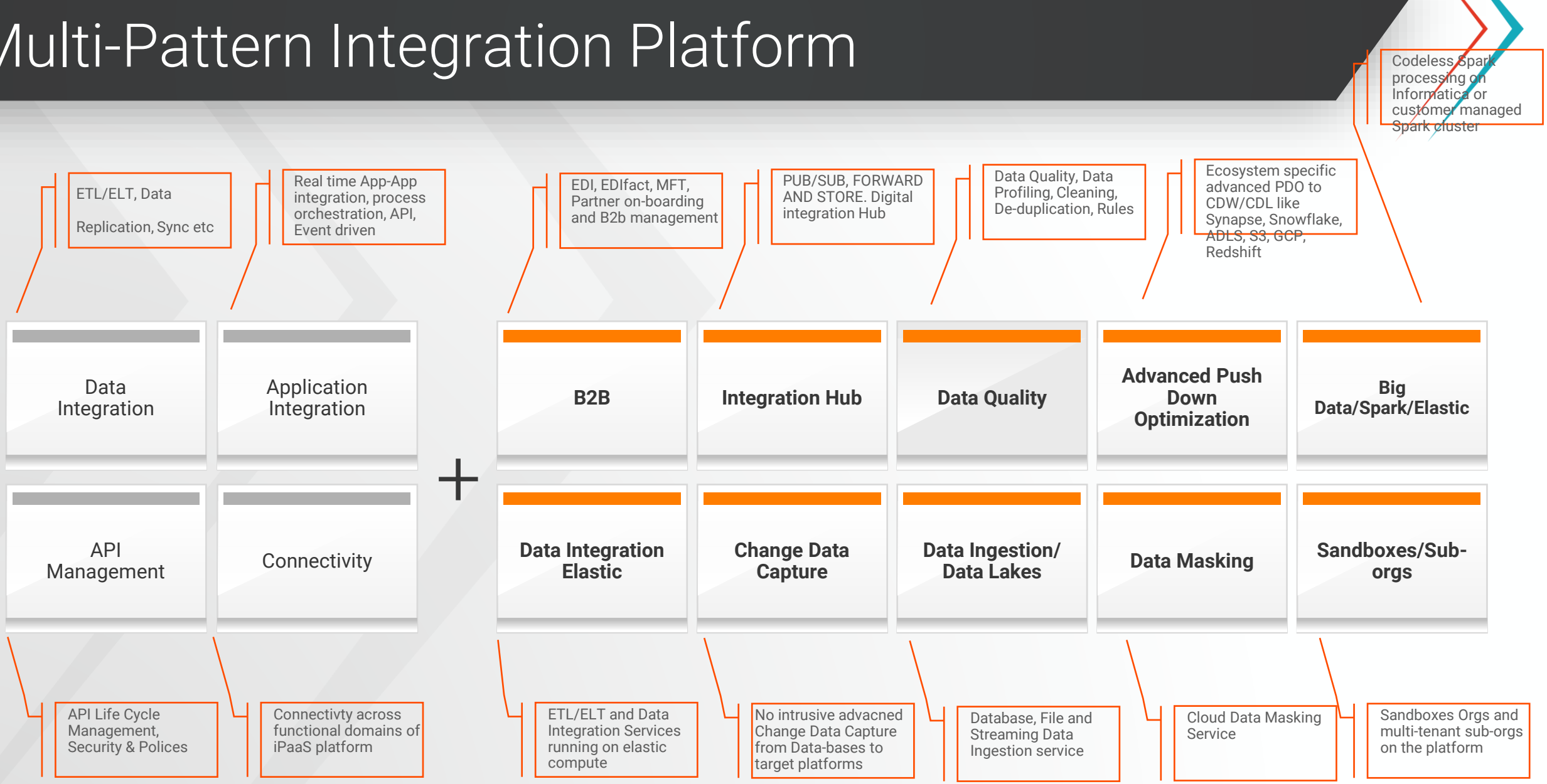
Single, Comprehensive Platform
Support for Multi-Cloud, On-Prem
Integrated Data Quality

End-to-End Cloud-Native Data Management



Informatica

Multi-Pattern Integration Platform



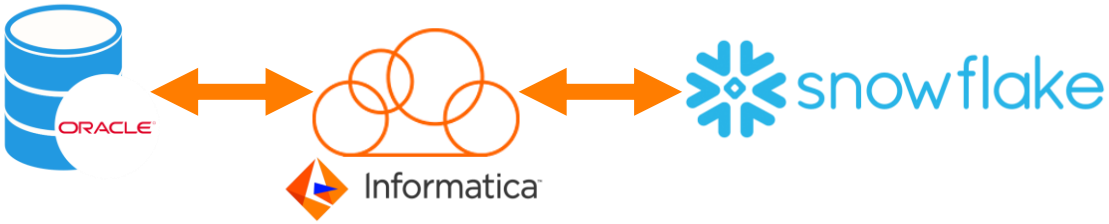
Informatica™

One set of tools for all your cloud and on-premises data and applications

Single Cloud

Multi-Cloud

Hybrid Cloud

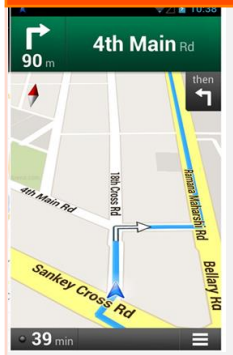


On-Premises

Informatica

Why Data Quality for Data Integration?

Profile



Turn by turn directions for data pipeline development instead of "heading west"

Standardize

Example 1:
KitKat
Kit-Cat
Kit Kat
→ **Kit-Kat**

Example 2:
North Central
Midwest
Great Lakes
→ **Central**

Parse

Text Data: Call Center Comments, Social Media, Product Descriptions

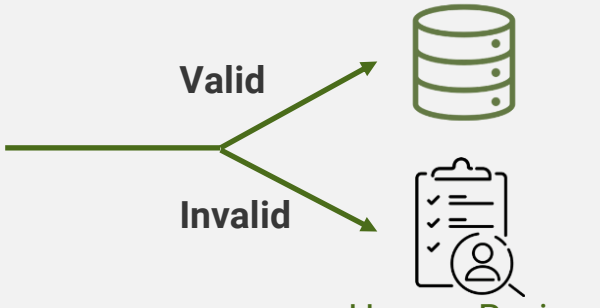
I love my Pink 64GB iPad Pro!!!

Color	Size	Product	Model
Pink	64 GB	iPad	Pro

Fuzzy Lookup/Match

Land O Lakes	Land O'Lakes
Biluxi	Biloxi
KitKat	Kit-Kat

Validate



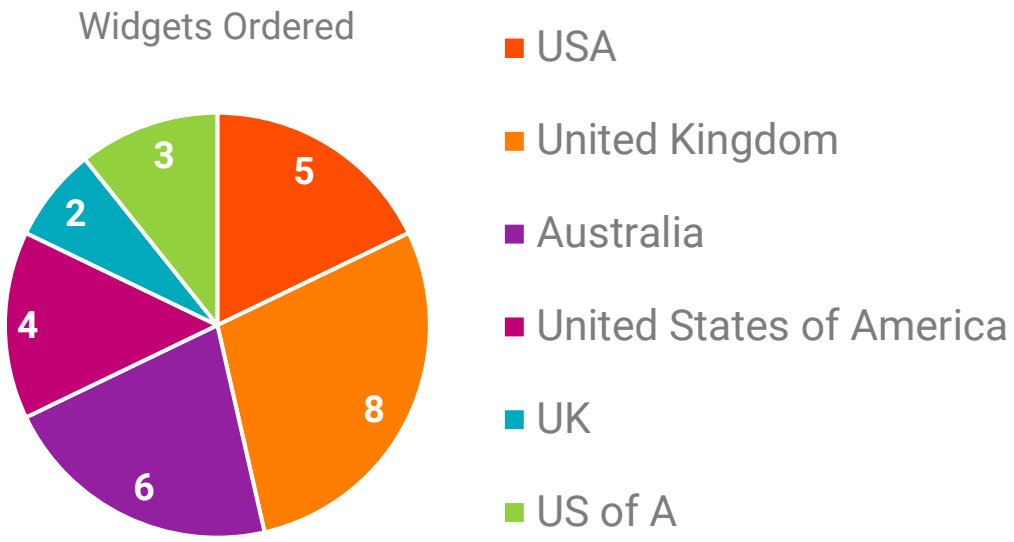
Enrich

AddressL1:	1008 Avenue of the Americas
AddressL2:	Suite 7
City:	New York
State:	NY
Zip Code:	10018-5402
Longitude:	40.7325525
Latitude:	-74.004970

Informatica

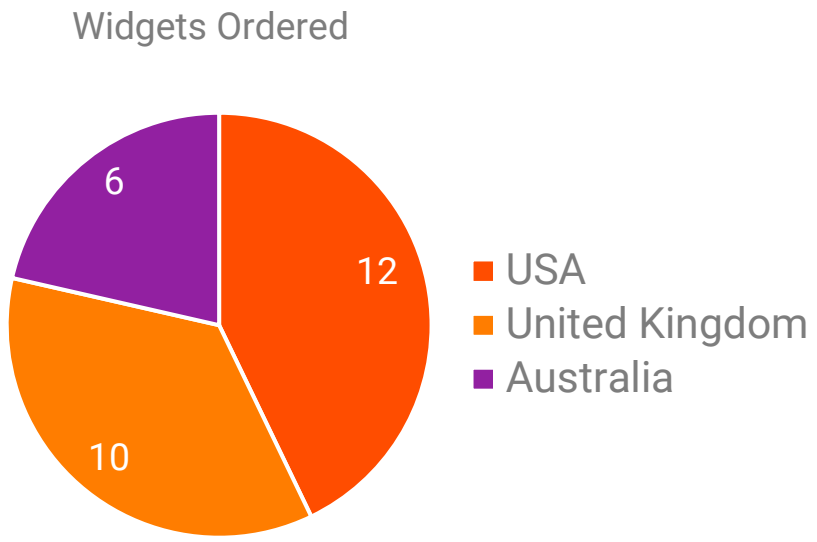
Why Data Quality for Business Intelligence?

Without Standardization



6 customers, United Kingdom is largest, Australia is second

With Standardization



3 customers, USA is largest, United Kingdom is second

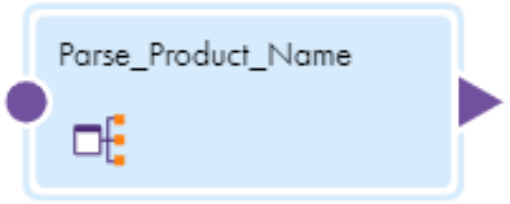
Informatica

Why Data Quality for Artificial Intelligence?

Hand Coding Data Quality is Time Consuming, Error Prone, Hard to Maintain, and Does Not Scale

Parse Product Description

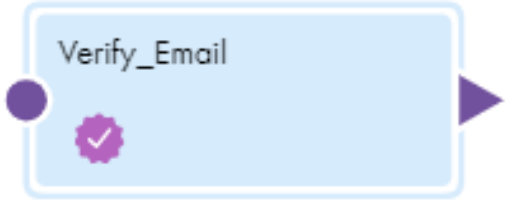
my_string = 'Names: Widget, Cog




```
# split the string at ','
step_2 = step_1.split(',')
# strip leading and trailing edge spaces of each item of the list
step_3 = [name.strip() for name in step_2]
# do all the above operations in one go
one_go = [name.strip() for name in my_string.split(':')[1].split(',')]
for idx, item in enumerate([step_0, step_1, step_2, step_3]):
print("Step {}: {}".format(idx, item))
print("Final result in one go: {}".format(one_go))
```

Verify Email

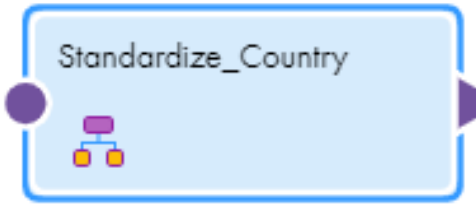
NOT(REGEX(



c ,'([a-zA-Z0-
3}\\.[a-
[a-zA-Z0-9\\-
0-9]{1,3})'))

Standardize Country

IIF(OR(UPPER(Country)="UNITED
ST
UP
A")
UK
BR



)="
Kingdom",IIF(UPPER(Country)="OZ",
"Australia",Country)))

Standardized	Variantion_1	Variation_2	Variation_3
USA	United States	US	America
United Kingdom	Great Britan	UK	

Informatica

The Informatica Intelligent Data Management Cloud (IDMC) Platform

Informatica

Informatica Solves Your Data Management Challenges

Cost Overruns

- Up to 70% reduction in ETL processing time

Resource Constraints

- New integration job setup time Reduced by 40% to 55%

Complexity

- Cloud Data Quality reduced the time dedicated to quality assurance tasks by up to 35% each week

All statistics from Nucleus Research: ROI Guidebook for Informatica, August 2020

Informatica

Mature, Secure, Multi-Cloud Platform

Connects to Everything

7M+
Integrations/day
200%+ growth YoY

~300%
Growth of
API volume

95M+
Continuous cloud
security checks

1B+
Jobs and processes
per month

2x
Volume of data
every 6 months

10K+
Metadata-aware
connectors

SOC 3 · AICPA SOC 2 (Formerly SAS 70 Reports) · HIPAA Health Insurance Portability & Accountability Act · ISO/IEC 27001:2013 Aligned Program · CSA cloud security alliance® · Privacy Shield Framework · salesforce appexchange · FedRAMP · CYBER ESSENTIALS

aws · Google Cloud · Microsoft Azure · snowflake · databricks

servicenow · salesforce · workday · Adobe · SAP · ORACLE NETSUITE

Informatica

A Leader in Five Gartner Magic Quadrant Reports

Informatica is placed furthest in completeness of vision and highest in ability to execute in ALL of these Magic Quadrant Reports

Enterprise Integration Platform as a Service

Sep 2020

Eric Thoo, et al.,
21 Sep 2020

Data Integration Tools

Aug 2020

Ehtisham Zaidi, et al.,
18 Aug 2020

Data Quality Solutions

Jul 2020

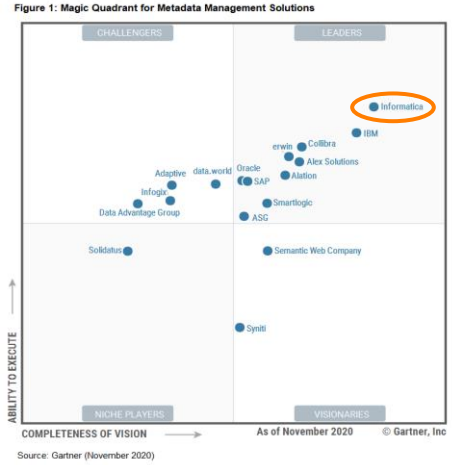
Melody Chien, et al.,
27 July 2020

Master Data Management Solutions

Jan 2021

Simon Walker, et al.,
27 Jan 2021

Metadata Management Solutions



Nov 2020

Guido De Simoni, et al.,
11 Nov 2020

Informatica

Two Ways to Take us for a FREE Test-Drive NOW

Cloud Data Integration on Azure & AWS

www.informatica.com/azurefree

www.informatica.com/awsfree

Process up to 500M rows of data per month free-of-charge

- **Cloud Data Integration Service** automates, high performance data integration at scale
- **Cloud Mass Ingestion Service** streamlines building and running of complex integrations with high-performance ingestion

*Free allowance is equivalent to up to 25 Informatica Processing Unit (IPUs) per month

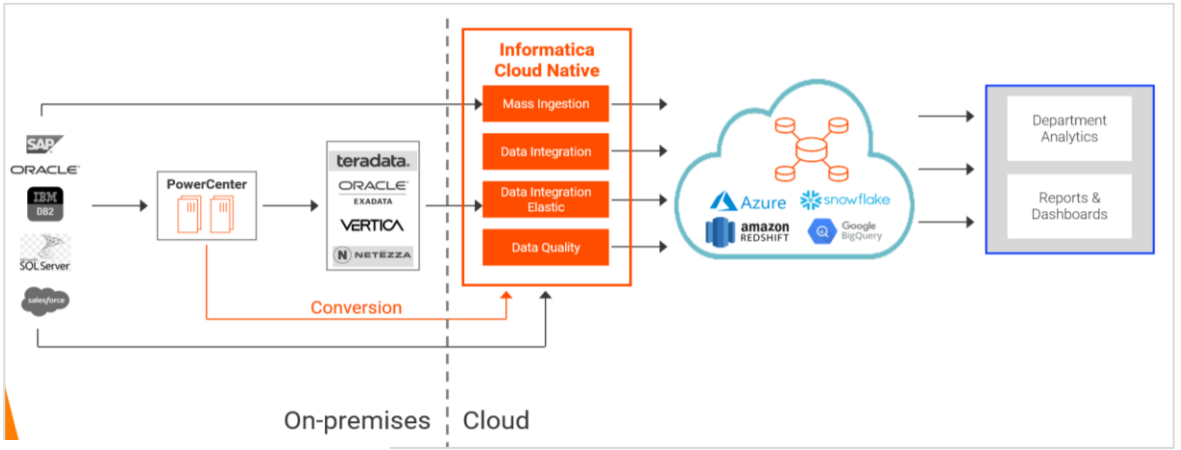
Full product 30-day trial

www.informatica.com/trials

Informatica

Next Session: The Best Way to Modernize from PowerCenter to the Cloud

1. How can we leverage what we've already built in PowerCenter?
2. What if we want to start using a new cloud data warehouse?
3. Does Informatica have a solution for modernizing PowerCenter?
4. Are there any incentives for PowerCenter customers?



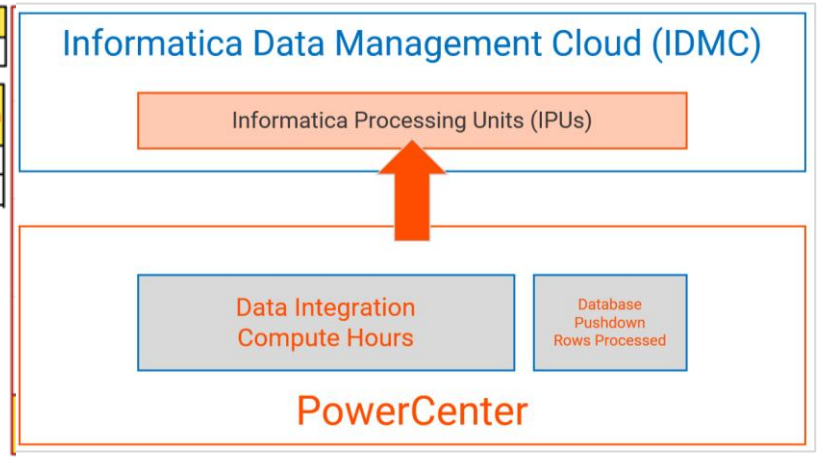
WORKFLOW

Total Workflows	Total Worklets	Total Sessions	Total Flags Scanned
67	13	363	2026

Fully Automated Today	This Many will be Fully Automated by			Possible for Manual Conversion
	October, 2020	Q1, 2021	Q2, 2021	
	6	37	1	23
	9%	64%	66%	34%

Possible for Manual Conversion
Needs Review
23

COMPLEXITY	% of Total	Workflows
Low	39%	26
Medium	51%	34
High	7%	5
Very High	3%	2
Grand Total	100%	67



Informatica

Thank You