Ingest and Replicate Applications Data in Minutes

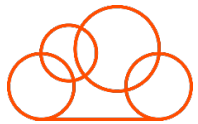
- Om Verma, Senior Product Manager

Informatica

Housekeeping Tips

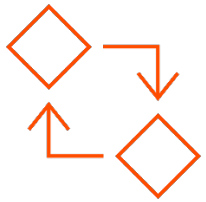
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via 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 INFASupport YouTube channel and 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.

Informatica

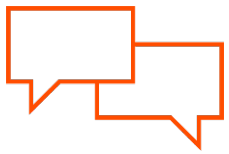
Agenda

1. CDW/DL Challenges and DI Market Drivers
2. CDW/DL Reference Architecture
3. Why Informatica 'Cloud Mass Ingestion' and Use Cases?
4. Cloud Mass Ingestion – 'Applications' – Deep Dive
5. Demo
6. Summary and Call to Action

Informatica

Data Integration Market Drivers

Unprecedented Growth in Data Diversity & Volume, Emergence of Fusion Data Teams, and Need to Accelerate Data Operationalization

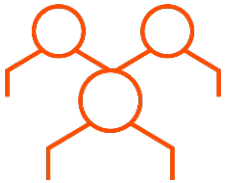


Data Diversity
(mobile, social, IoT)

46 billion connected devices

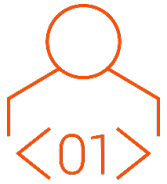
AIML projects fail to deliver

Only 21% AI initiatives in production



Chronic shortage of data engineers

50% annual growth in open data engineering positions



Data Engineering Democratization

500 million business data users

Explosion in Data Volume

64.2 zettabytes of data per year

Informatica

Businesses struggling with point solutions, complex architecture, lack of resources, runaway costs

Complexity

of organizations don't have a complete architecture to manage end-to-end set of data activities

- 75% of point products don't integrate and interoperate
- Data practitioners spend over 80% of their time preparing data instead of analyzing the data
- ~50% of organizations challenged by data quality

Resource Constraints

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

- Difficulty finding specialized skills fast
- Lack of automation impacting the ability to scale
- Lack of self-service access for non-tech users delaying rapid innovation

Cost Overruns

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

- Difficulty predicting compute costs
- Lack of visibility and control of users and usage
- Increasing data transfer costs

Informatica

Challenges at Various Stages of Data Management

How do I discover the data I need to bring into the cloud?

Discover

How do I get disparate, real-time data to the cloud?

Ingest and Replicate

How do I keep track and govern data across cloud and on-premises?

Govern

How do I ensure the data I am bringing into the cloud is high quality and usable?

Cleanse

How do I get a 360 view of my customer, partner and supplier data?

Master

How can I protect and secure my high-risk data across the cloud environment(s)?

Secure

Informatica

Challenges at Various Stages of Data Management

How do I discover the data I need to bring into the cloud?

Discover

How do I get disparate, real-time data to the cloud?

Ingest and Replicate

How do I keep track and govern data across cloud and on-premises?

Govern

How do I ensure the data I am bringing into the cloud is high quality and usable?

Cleanse

How do I get a 360 view of my customer, partner and supplier data?

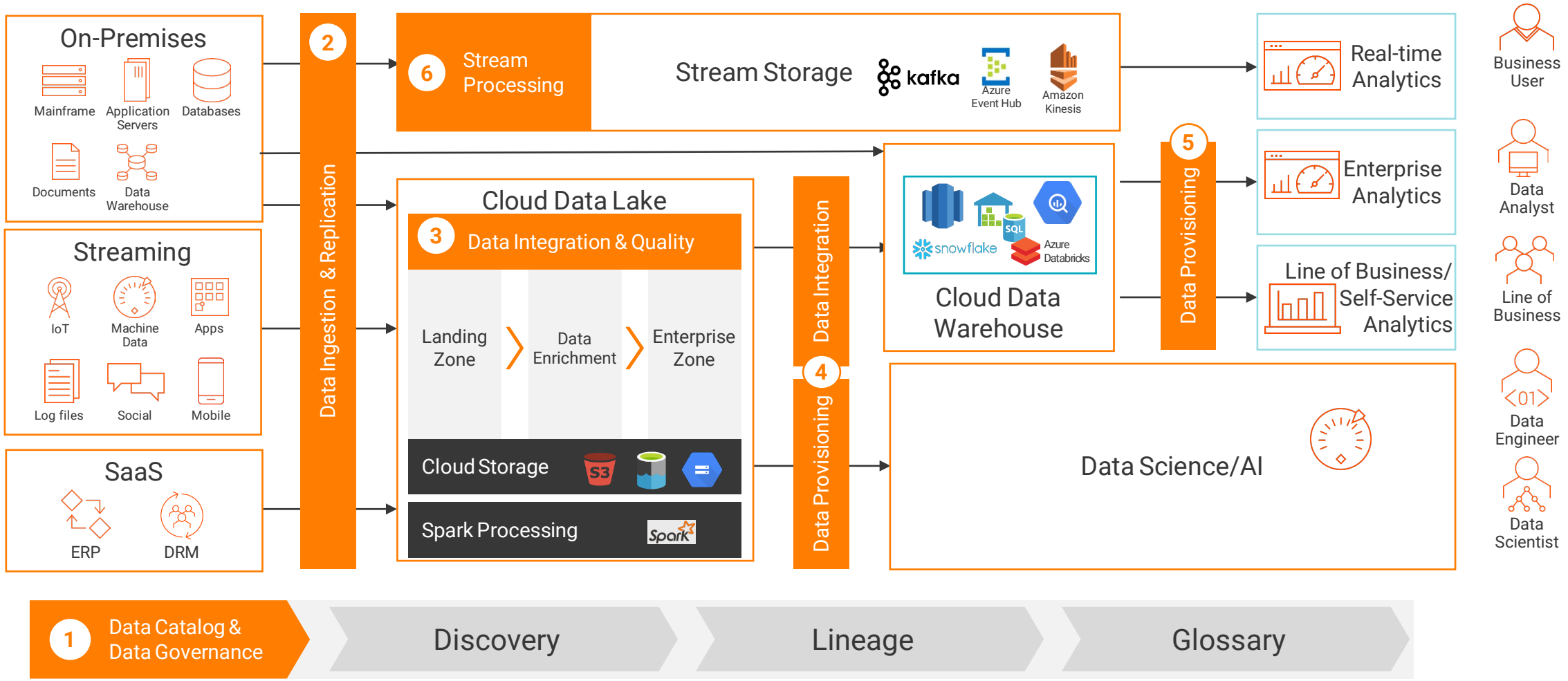
Master

How can I protect and secure my high-risk data across the cloud environment(s)?

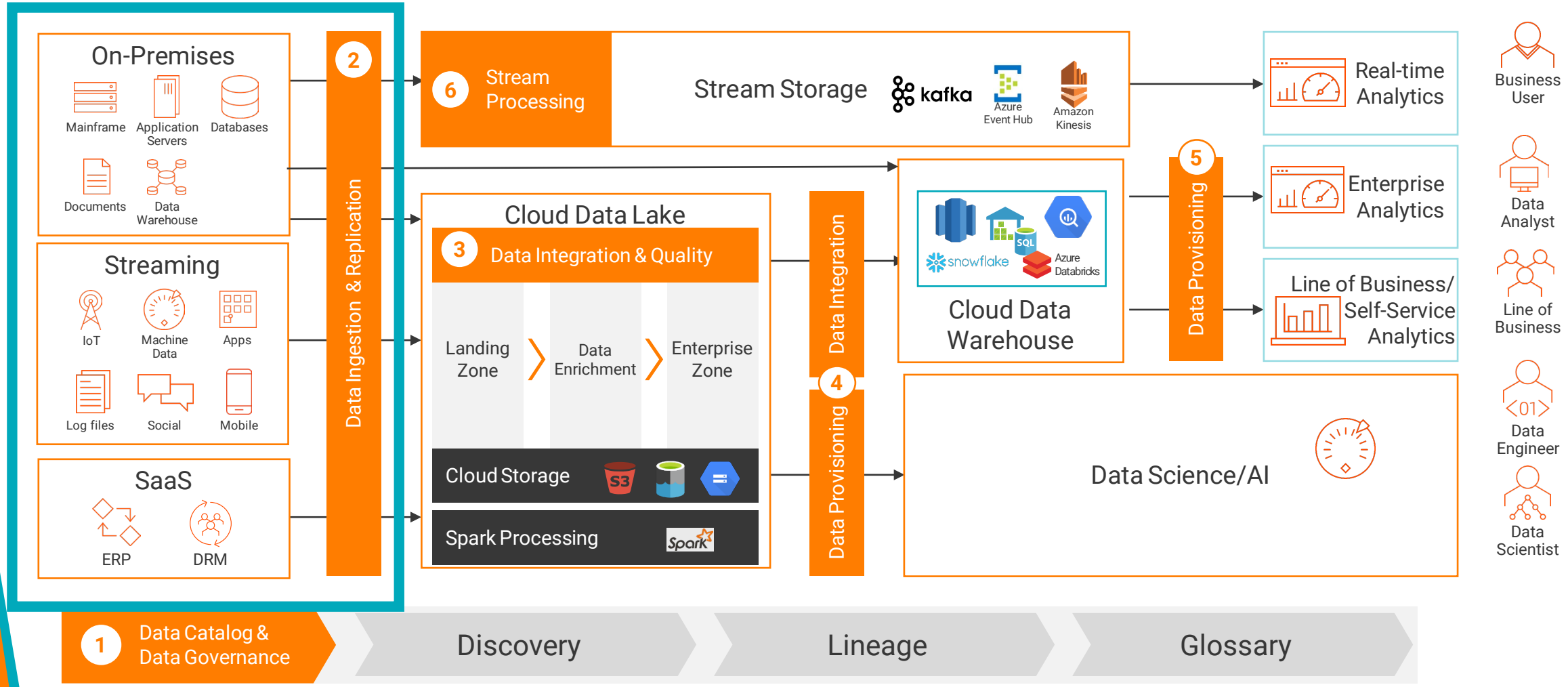
Secure

Informatica

Informatica Data Management Cloud - Reference Architecture for CDW/DL



Informatica Data Management Cloud - Reference Architecture for CDW/DL

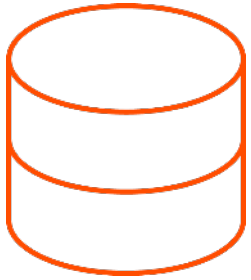


Existing Data Ingestion Solutions Requires Hand-Coding And Don’t Support All Ingestion Patterns

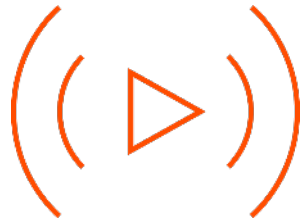
Requires Hand-Coding

```
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 = new LiveStream(sc.getConf)
    ls.relay(JP.sparkConfToJson(sc.getConf))
    ls.relay(JP.hadoopConfToJson(sc.hadoopConfiguration))
    val lis = new Listener(ls,"TAG")
    sc.addSparkListener(lis)
sqlContext.sparkSession.experimental.extraPreprocessing = new TaggingRules().rules
    val accs = List()
    ls.relay(JP.sparkAppDetailsToJson(sc.getConf, accs))
    lis.accumulators = accs
    import sqlContext.implicits._
    import sqlContext.implicits._
    import org.apache.spark.sql.functions.{stddev_samp, var_samp}
    val icast = caster("MM/DD/YYYY HH24:MI:SS")
    val acast = adapterCaster()
    val jcast = JavaCaster()

    try {
    Tuple2(sqlContext.sql(Params.resolve("DROP TABLE IF EXISTS
`default`.`w8314020283544407639__src1_lookuptests_newmap`", "MM/DD/YYYY HH24:MI:SS")),
sqlContext.sql(Params.resolve("CREATE TABLE
`default`.`w8314020283544407639__src1_lookuptests_newmap` (`col0` STRING, `col1` DECIMAL(18, 2),
`col2` STRING) ROW FORMAT LOCATION
'hdfs://cdh52.vm.com:8020//tmp/sess8314020283544407639//W8314020283544407639_COPY_src1_L
ookupTests_newMap' TBLPROPERTIES ('.columns.types'='string,decimal(18,2),string',
'pwx.mapping.file.path'='./src1_MAPPING_1204153139656796_1204153864551238.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__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
`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap`.`col0` as a0,
`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap`.`col1` as a1,
`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap`.`col2` as a2
FROM `default`.`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap`",
"MM/DD/YYYY HH24:MI:SS"))).tag("SRC_testPassiveEquiJoinLkp_lkp_source").itoDF);
    val v1 = v0(0);
    val v2 = v0(2);
    val v3 = v0.groupBy(v1, v2).agg(v1, v2, last(v0(1), false)).itoDF("m");
    val v4 = updatePartitions(asBlock(sqlContext.sql(Params.resolve("SELECT
`w8314020283544407639__src1_lookuptests_newmap`.`col0` as a0,
`w8314020283544407639__src1_lookuptests_newmap`.`col1` as a1,
`w8314020283544407639__src1_lookuptests_newmap`.`col2` as a2 FROM
`default`.`w8314020283544407639__src1_lookuptests_newmap`", "MM/DD/YYYY
HH24:MI:SS"))).tag("SRC_src1").itoDF("d"), v3);
    val v5 = v4.join(v3, v3(0).<=>(v4(1)).&&(v3(1).<=>(v4(2))), "left_outer").itoDF;
    asBlock(sqlContext.sql(Params.resolve("INSERT OVERWRITE TABLE
`default`.`w8314020283544407639__testpassiveequijoinlkptgt_lookuptests_newmap` SELECT tbl0.c0 as
a0, tbl0.c1 as a1 FROM tbl0", "MM/DD/YYYY HH24:MI:SS")), v5.iselect(v5(0),
v5(5)).itoDF("TGT_").tag("TGT_testPassiveEquiJoinLkpTgt").itoDF("c").createOrReplaceTempView("tbl0")
);
    } finally {
    sqlContext.sql(Params.resolve("DROP TABLE IF EXISTS
`default`.`w8314020283544407639__testpassiveequijoinlkptgt_lookuptests_newmap`", "MM/DD/YYYY
HH24:MI:SS"));
    sqlContext.sql(Params.resolve("DROP TABLE IF EXISTS
`default`.`w8314020283544407639__testpassiveequijoinlkp_lkp_source_lookuptests_newmap`",
"MM/DD/YYYY HH24:MI:SS"));
```

Doesn’t Support All Ingestion Patterns

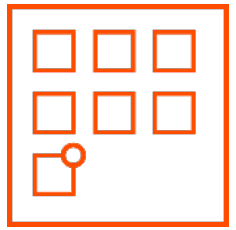


Database



Streaming

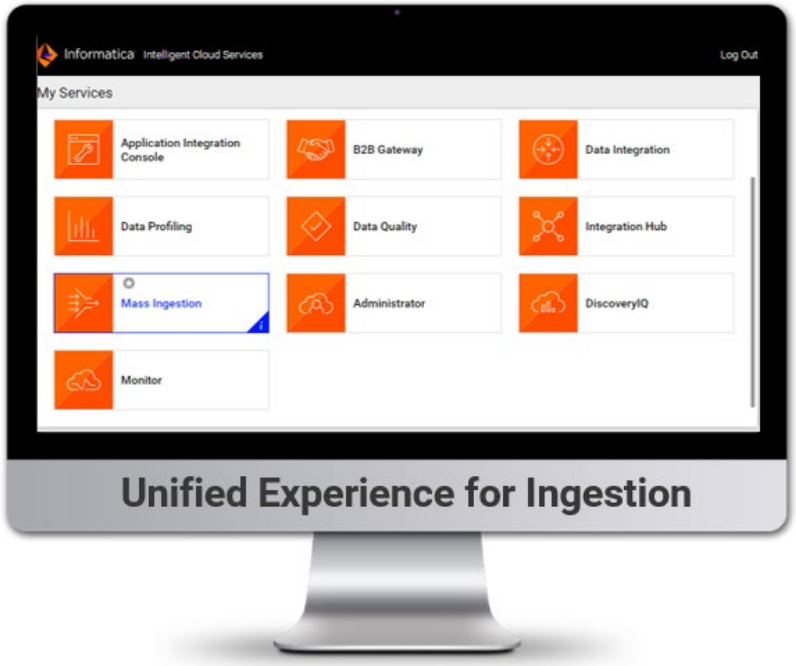
Files



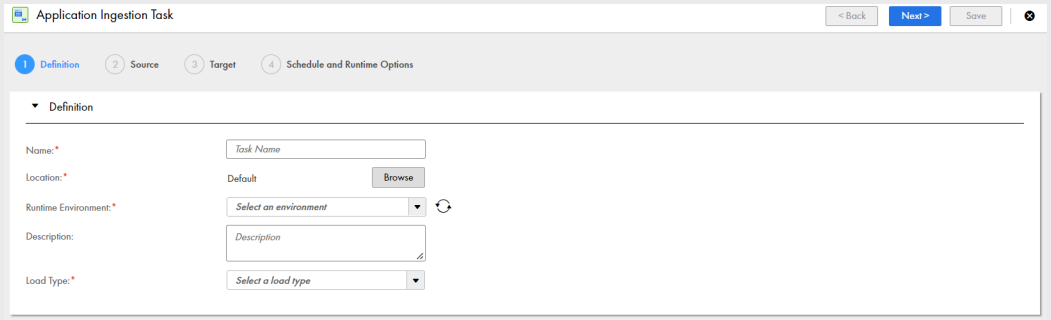
Applications

Informatica

Informatica's Data Ingestion and Replication Solution – Cloud Mass Ingestion



✓ Step-by-step wizard for designing and creating an ingestion task



✓ Deployment, scheduling, real-time monitoring and lifecycle management

Ingest in Real-Time

Real-Time Monitoring

✓ Versatile out-of-the-box connectivity to sources and targets

Databases & CDC

Streaming Sources

Files

Applications

Informatica

Cloud Mass Ingestion (CMI)

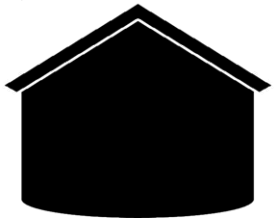
Use Case Patterns

Data Lake Ingestion

- Mass ingestion of application, on-premises database content into a cloud or on-premises data lake
- Mass ingestion of files into cloud and on-premises data lakes
- Streaming and IoT data ingestion into a data lake

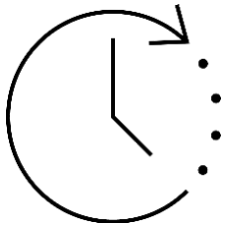
DB/DWH Replication/Data Warehouse Modernization

- Mass ingestion of on-premises database, data warehouse, applications and mainframe content into a cloud data warehouse (Snowflake, Synapse etc.)
- Synchronize ingested data with Change Data Capture (CDC) and applying



Real Time Analytics

- Log files and clickstream ingestion
- CDC ingestion
- IoT data ingestion
- App change data ingestion



Informatica

Cloud Mass Ingestion

Services

Choose the asset type you want to create.

Application Ingestion Task

Database Ingestion Task

File Ingestion Task

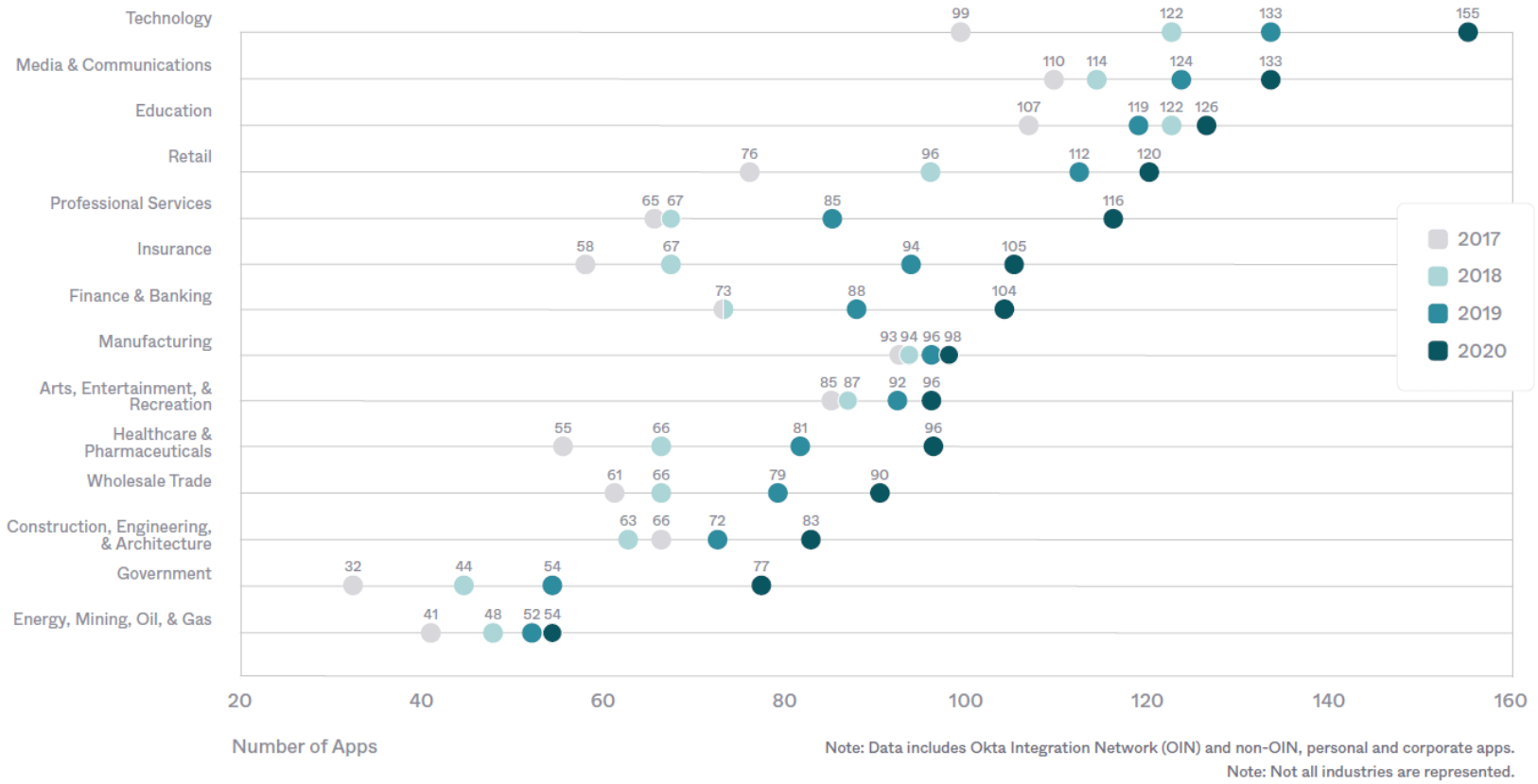
Streaming Ingestion Task

Cancel

Mass Ingestion Applications New!

Informatica

Average Number of Apps per Customer, by Industry



Note: Data includes Okta Integration Network (OIN) and non-OIN, personal and corporate apps.
Note: Not all industries are represented.

Source: https://www.okta.com/sites/default/files/2021-03/Businesses-at-Work-2021.pdf

Informatica

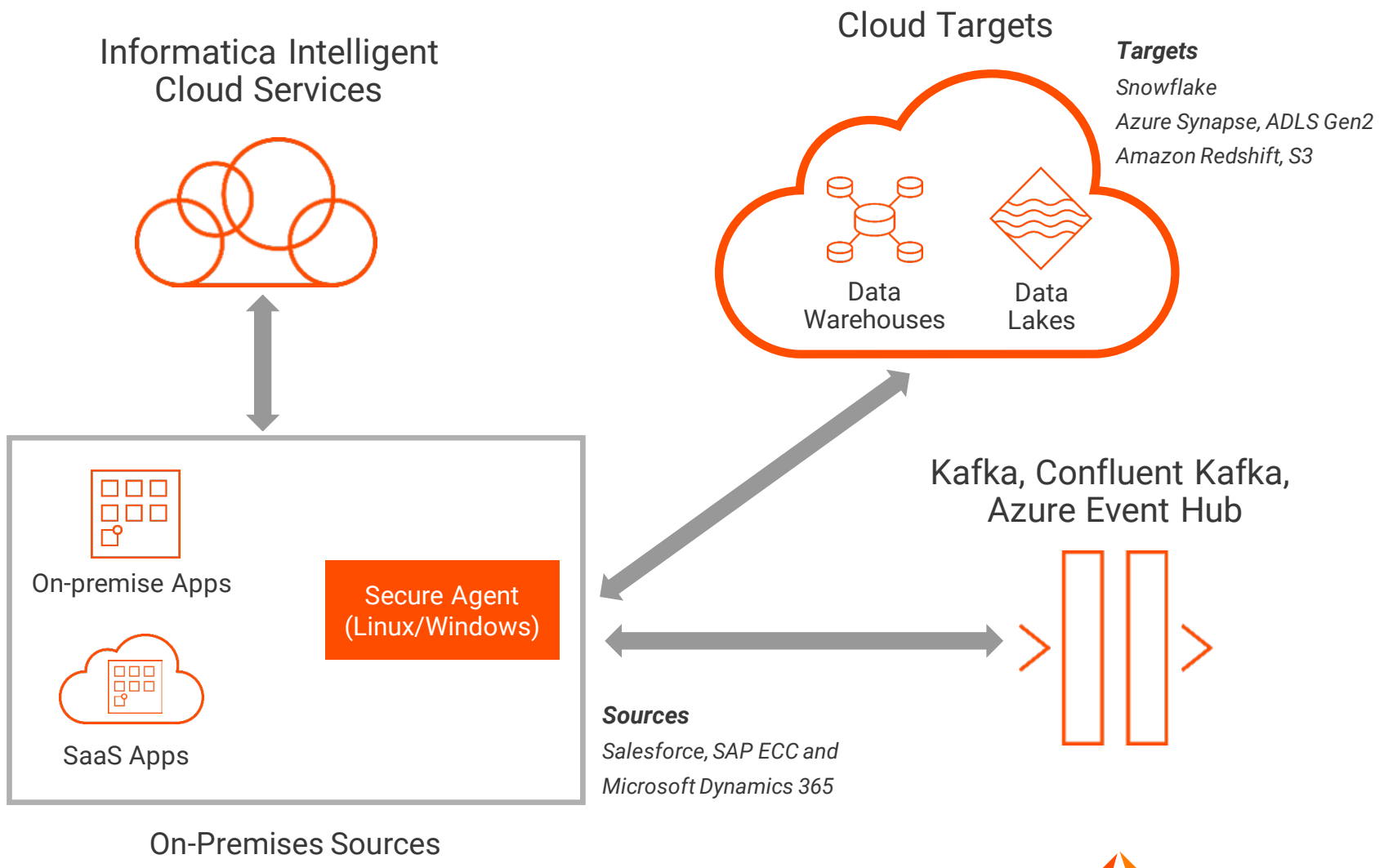
Cloud Mass Ingestion - Applications

Provides Application ingestion capabilities as part of IICS Mass Ingestion service

Ingest data from SaaS and on-prem applications like Salesforce, SAP ECC and Dynamics 365. Also supports schema drift for supported applications

Real-time monitoring of ingestion jobs with lifecycle management and alerting in case of issues

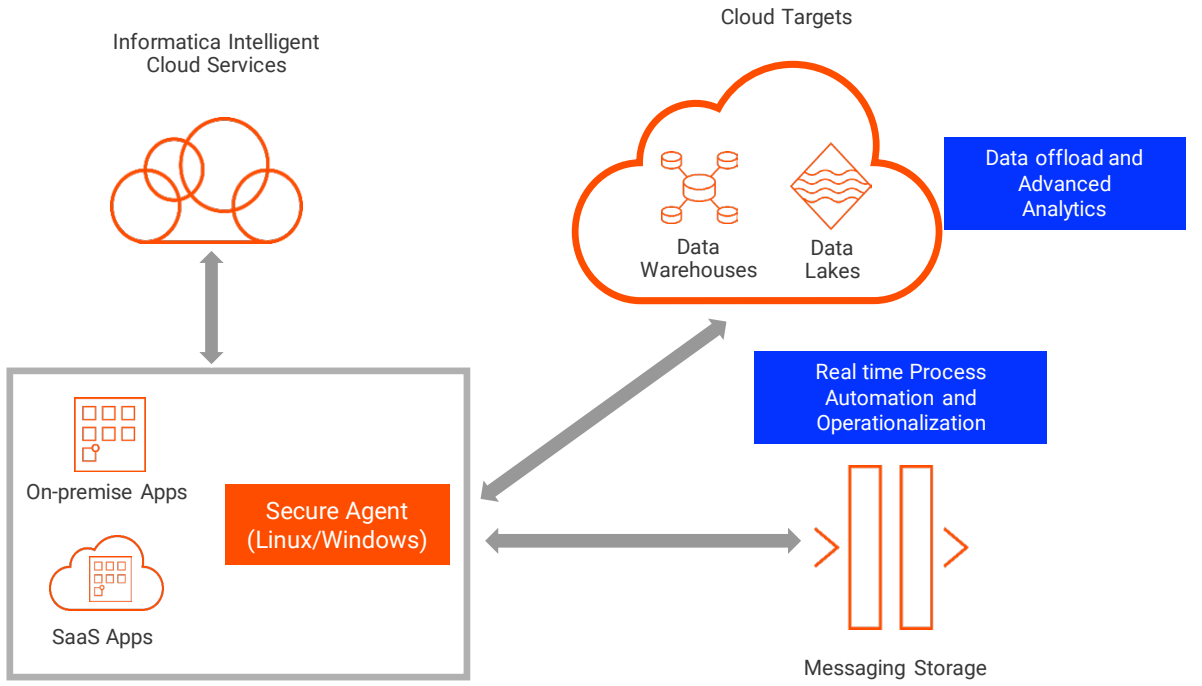
Orchestrate Application data ingestion in **hybrid / cloud** as **managed** and **secure** service



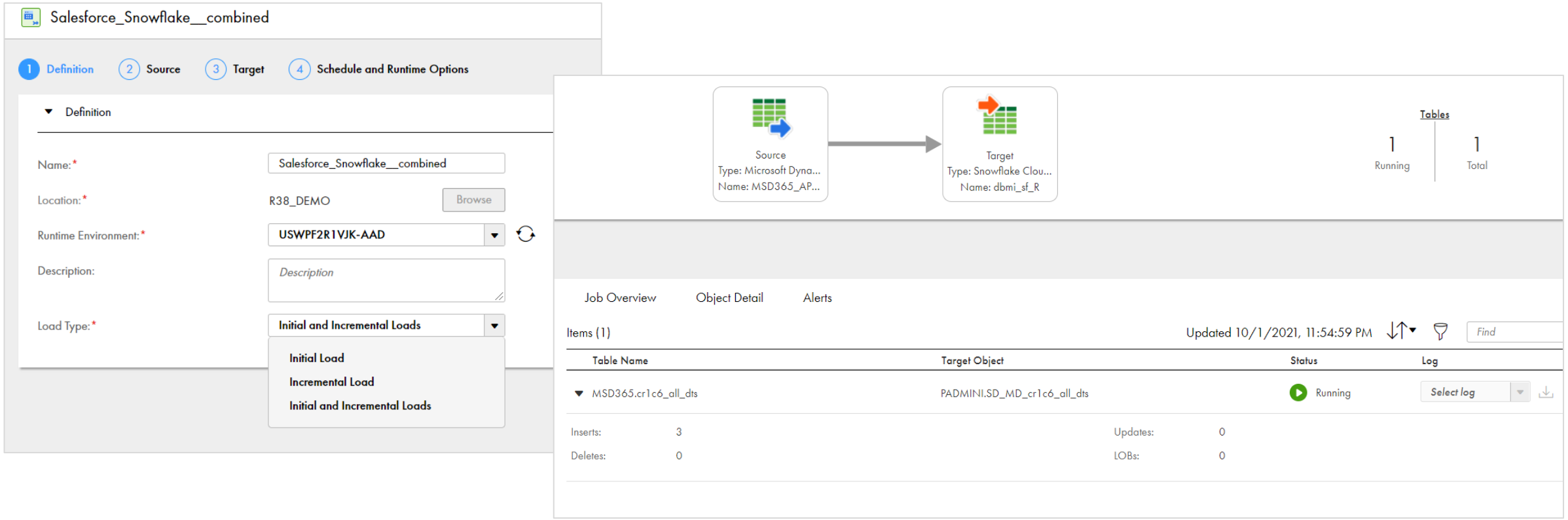
Mass Ingestion Applications – Key Use Cases

- Analytics offload from SaaS application & DW consolidation
- Advanced Analytics on Application Data
  - Initial & change data ingestion, apply changes, schema drift, near real time ingestion
- Real time process automation & action operationalization (for App modernization)
  - CDC data ingestion onto Messaging Systems
  - Integration with CAI and CIH for downstream consumption*
  - Task flow Integration for downstream processing*

**roadmap*



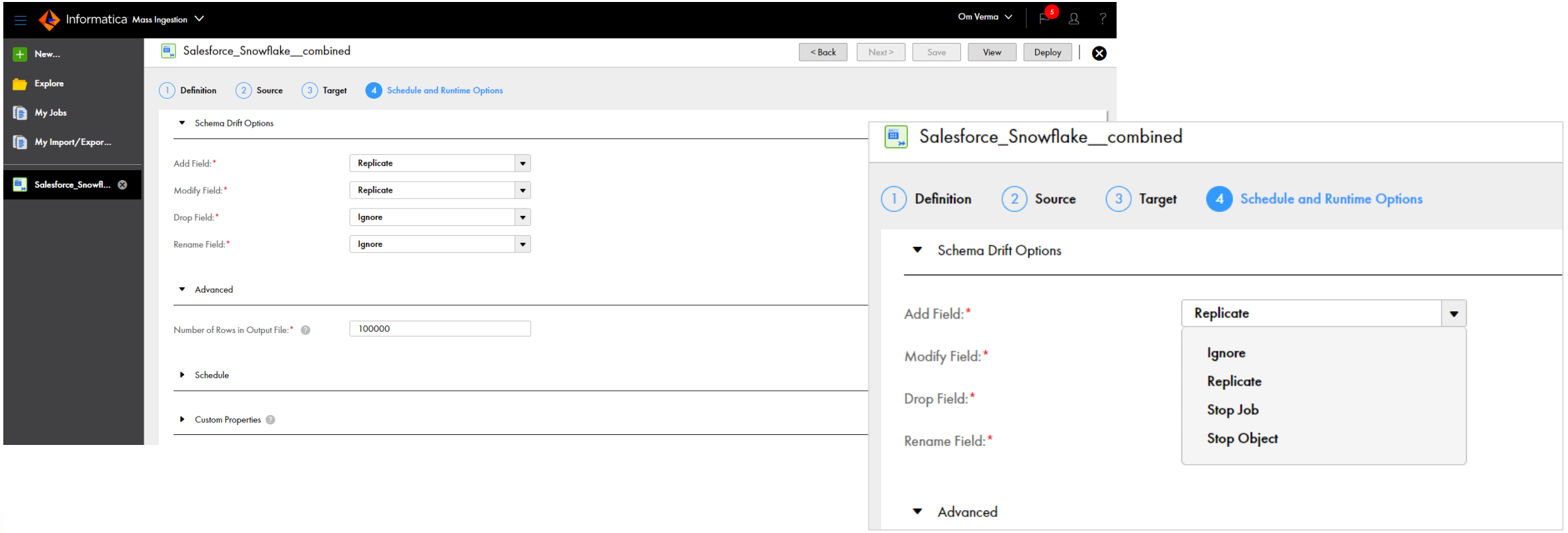
Efficient Capture and Ingestion of CDC Data



Efficient change data collection from source system and applies the changes onto the target

Informatica

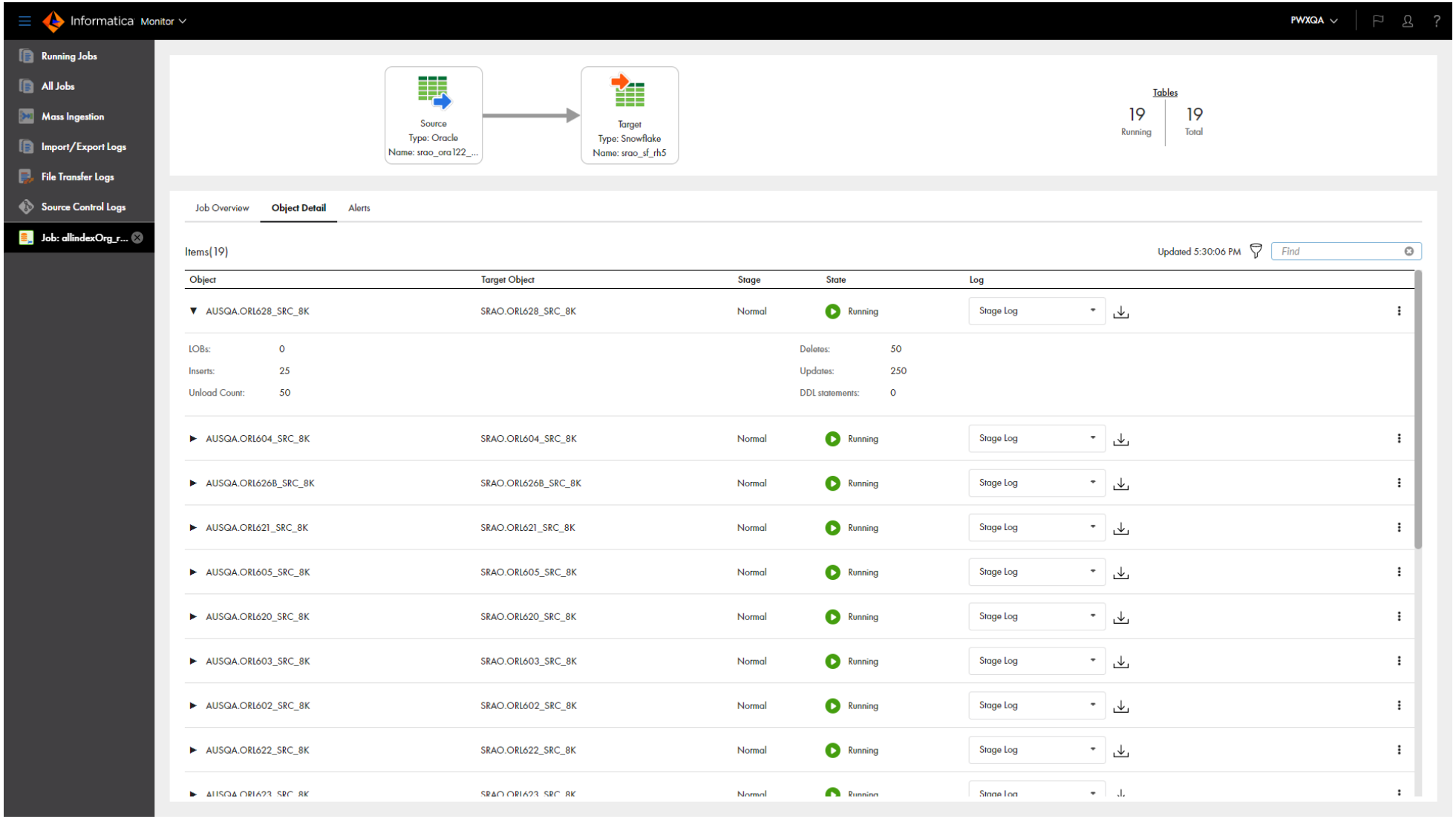
Automatically Addresses Schema Drift at the Source



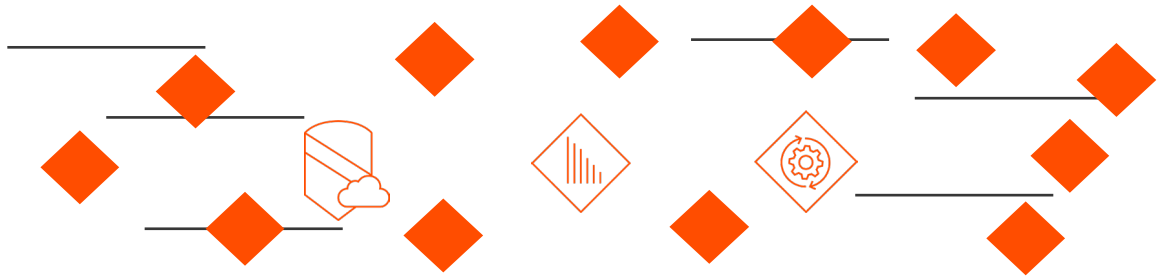
Schema drift recognizes changes on the source application schema and automatically processes and applies the changes

Informatica

Real time monitoring



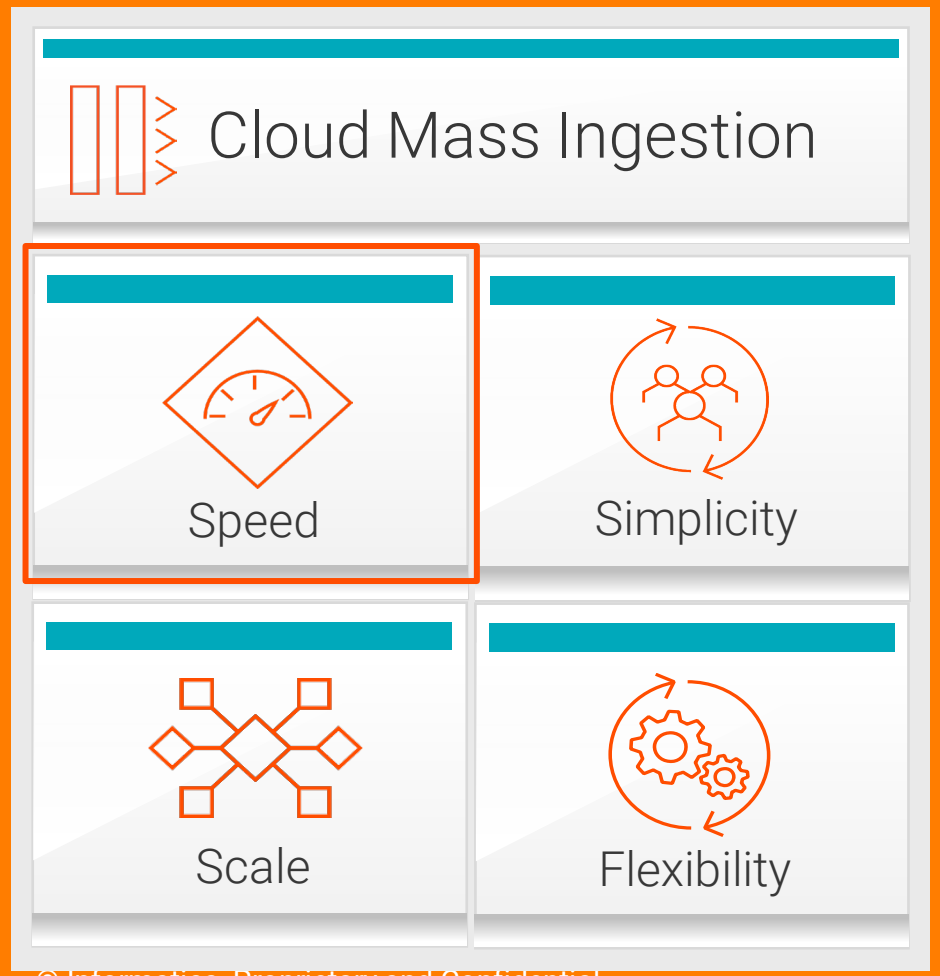
Out Of The Box Native Connectivity



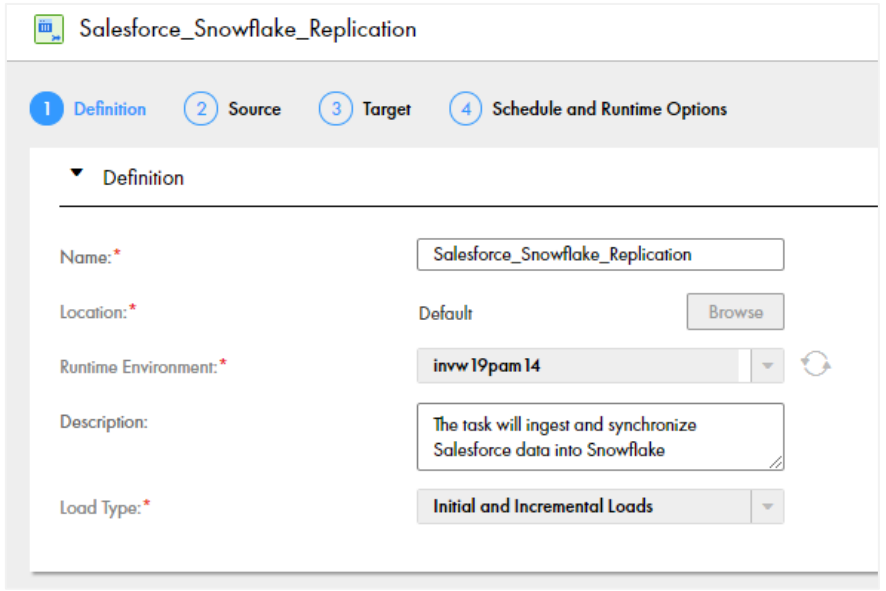
Secure Scalable Maintenance free

and many more…

Why Informatica Cloud Mass Ingestion?



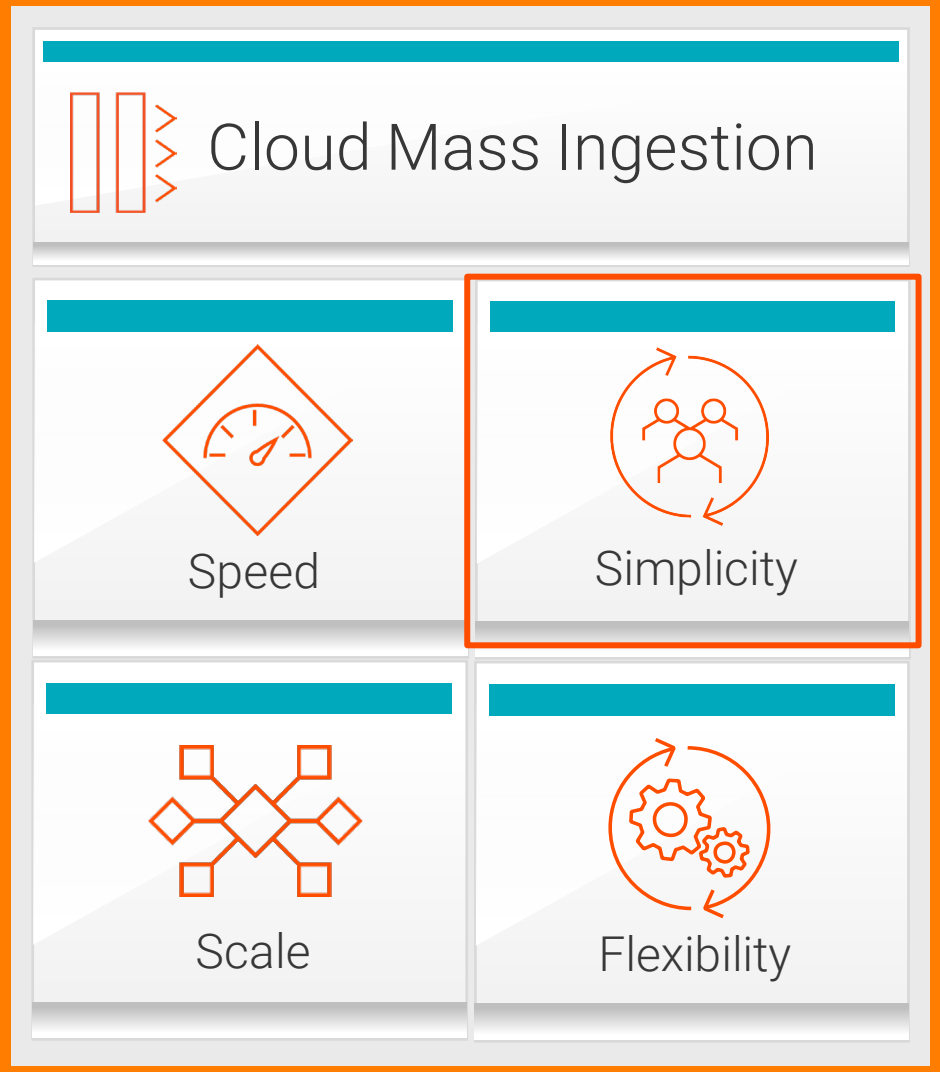
Build Data Ingestion Jobs in Minutes



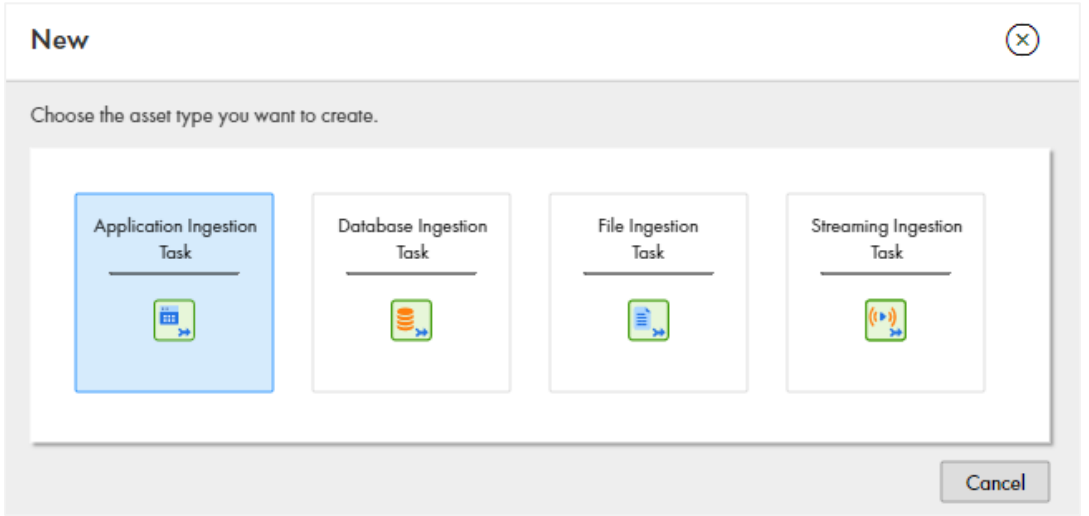
Build data ingestion jobs in minutes with a simple, easy to use 4-step wizard-based experience

Informatica

Why Informatica Cloud Mass Ingestion?



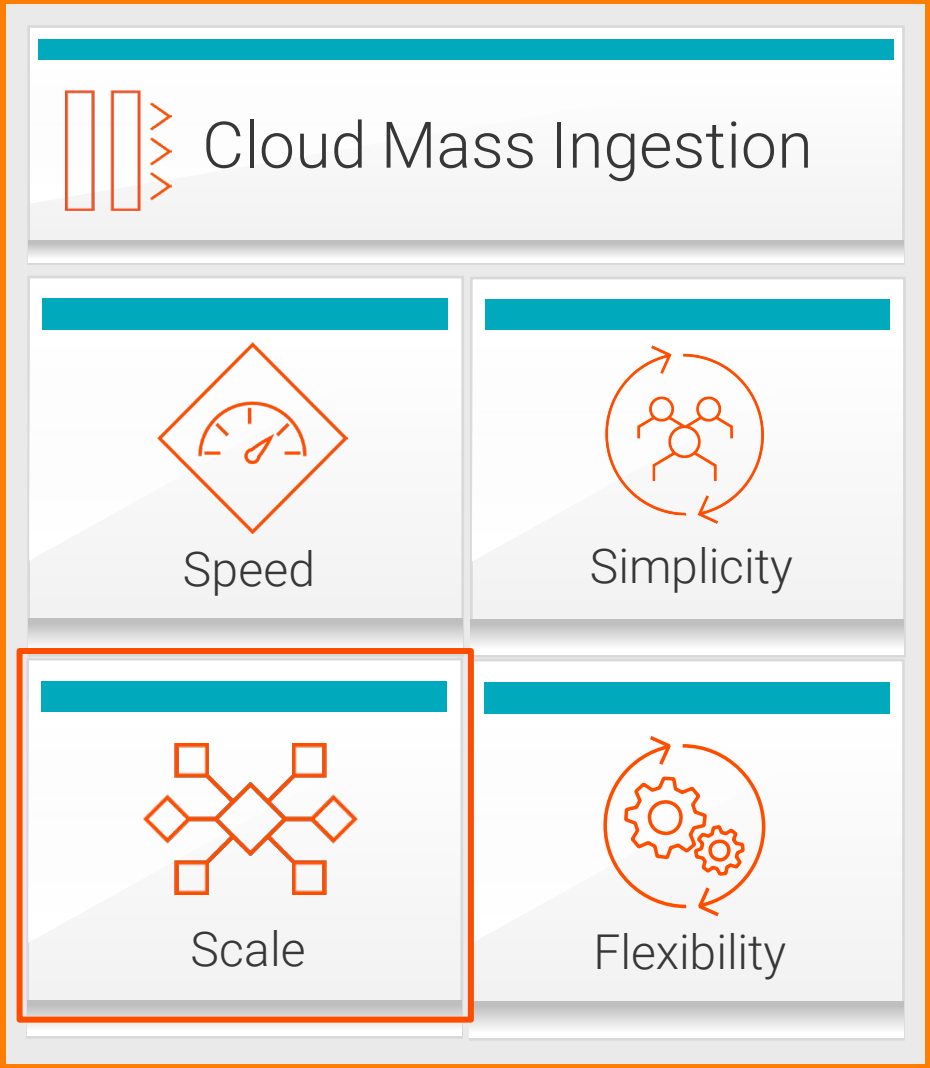
Simplify your Data Ingestion Tasks



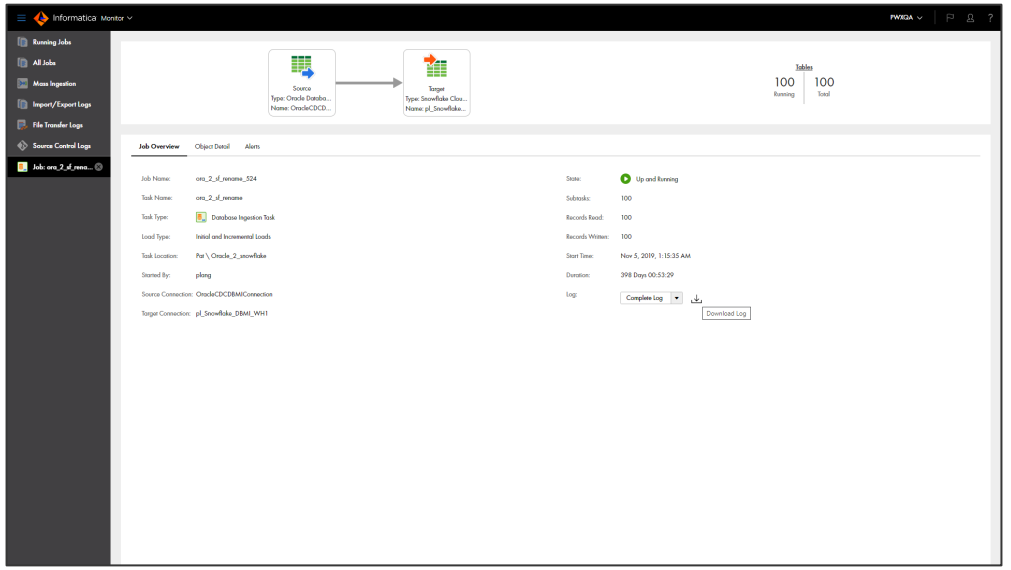
Simplify data ingestion with a single, unified cloud-native data ingestion solution with out-of-the-box connectivity

Informatica

Why Informatica Cloud Mass Ingestion?



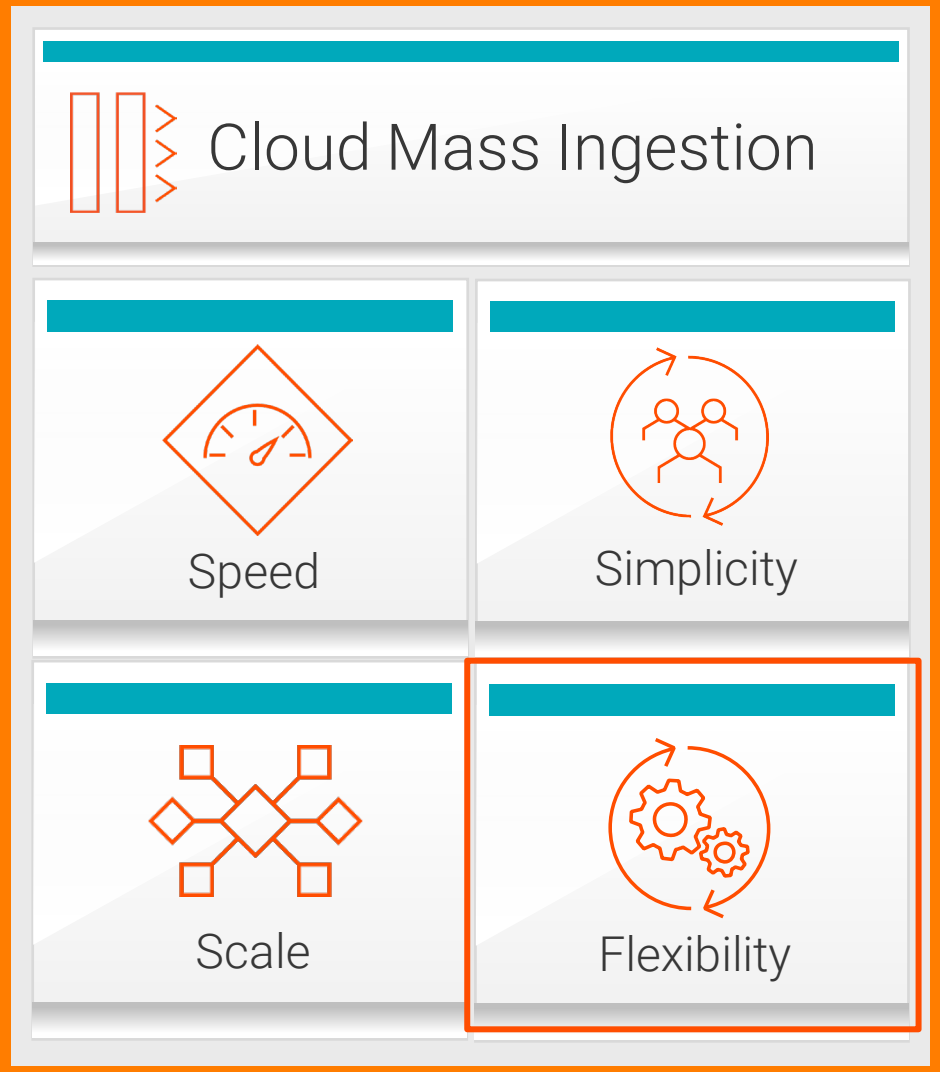
Data Ingestion at Scale



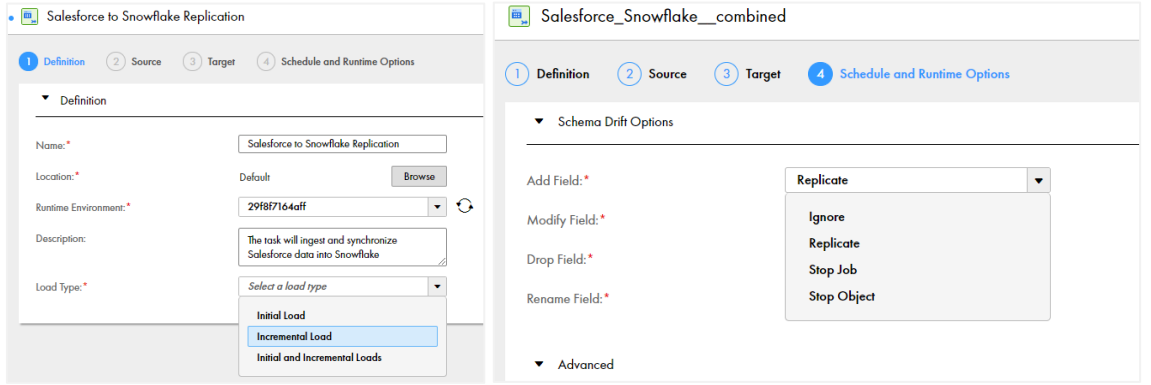
Ingest terabytes of any data,
any pattern, at any latency at scale
in real-time and batch with no data limit

Informatica

Why Informatica Cloud Mass Ingestion?



Flexibility to Track, Capture and Update Changes



Track, capture, and update changed data in real-time with automatic schema drift support to accelerate application replication and synchronization use cases

Informatica

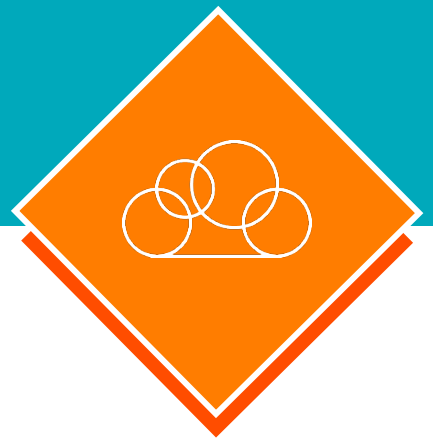
Summary

Cloud Native Ingestion

- Unified service for ingestion from various sources
- Orchestration for ingestion from variety of patterns
- Support for CDC and Schema Drift

Connectivity

- On-prem database and CDC
- On-prem and cloud files
- IoT and streaming
- Cloud data lakes, data warehouse and messaging hub



Wizard-driven Design

- Simple, easy-to-use wizard
- Edge transformations
- Intent-driven ingestion

Real-time Monitoring

- Pictorial view of the ingestion job
- Real-time flow visualization
- Lifecycle management

Informatica

Free 30-day Trial

Free 30-Day Trial: Cloud Mass Ingestion

Ingest any data at scale to make it immediately available for real-time processing, database replication, and application synchronization. Use an automated, wizard-based approach to efficiently ingest databases, applications, files, and streaming data at scale into a cloud or on-premises data lakes or data warehouses.

In this trial, you can:

- Transfer any size or type of file with high performance and scalability
- Collect, filter, and combine data from streaming sources such as IoT endpoints and messaging systems
- Ingest data at scale from common relational databases, Saas, and on-prem applications and propagate the data into a cloud data warehouse, cloud data lake, and message hub
- Track, capture, and update changed data in real-time with automatic schema drift support

https://www.informatica.com/trials/data-ingestion.html

Thanks