April 21st 2020

Informatica Cloud Mass Ingestion

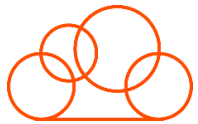
Informatica

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 to view on our INFASupport YouTube channel and Success Portal. The link will be emailed as well.
- Please take time to complete the post-webinar survey and provide your feedback and suggestions for upcoming topics.

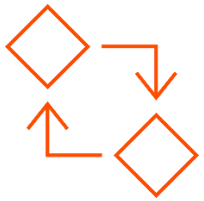
© Informatica. Proprietary and Confidential.

https://success.informatica.com



Bootstrap product trial experience

Enriched Onboarding experience



FREE Product Learning Paths and weekly Expert sessions

Informatica Concierge with Chatbot integrations

Tailored training and content recommendations

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. Proprietary and Confidential.

Agenda

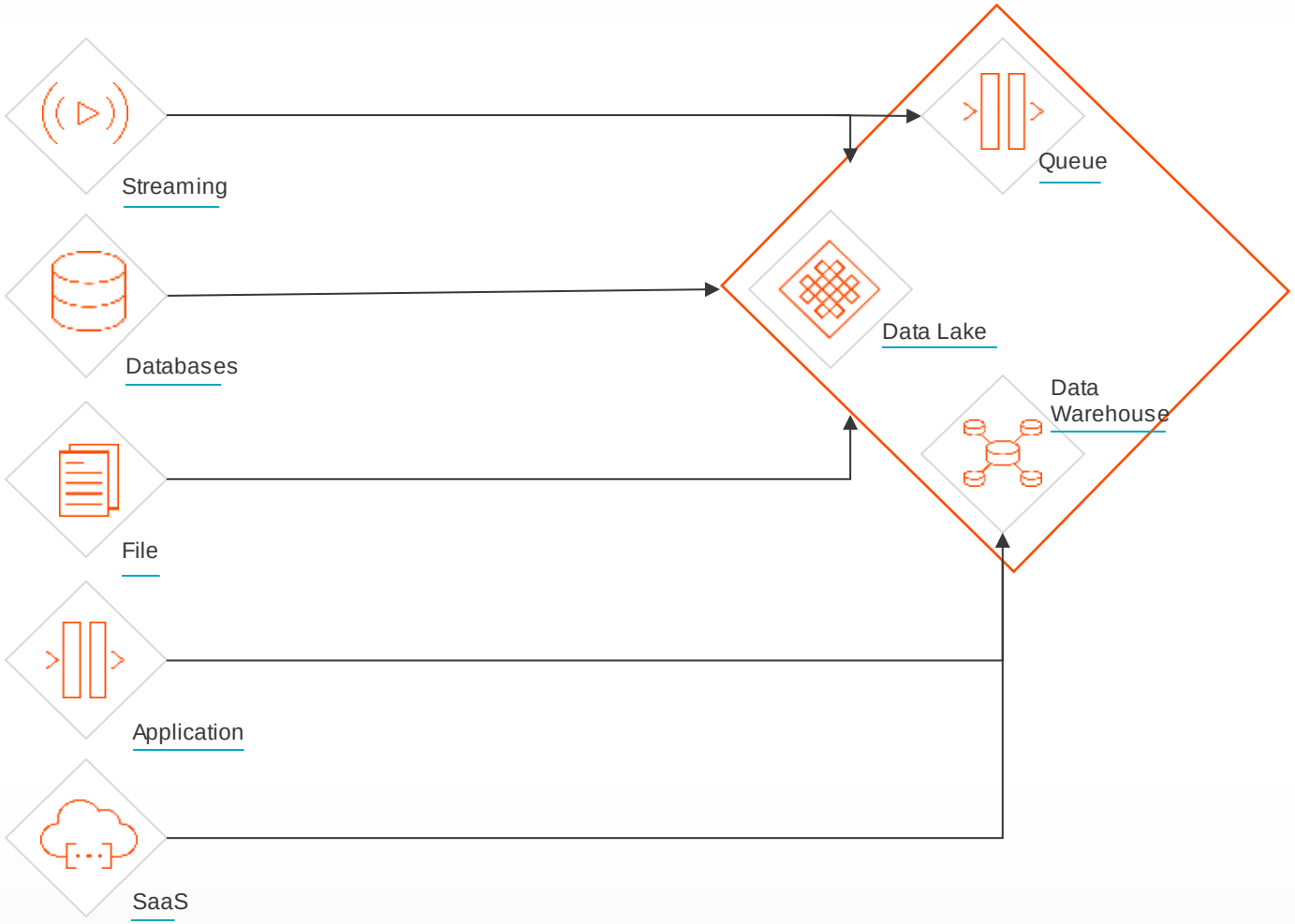
1. Ingestion Patterns
2. Ref. Architecture
3. Mass Ingestion
4. Streaming Ingestion
5. File Ingestion
6. DB Ingestion
7. Q&A

Does this sound familiar?

- *I have my data lake in cloud & I need to ingest data from variety of sources. How do I do that?*
- *I have standardized on Kafka as my enterprise messaging system. How do I get data onto Kafka from variety of streaming and batch systems?*
- *I have data in Kafka and need to get to my cloud data lake. Can someone help me?*
- *I have to capture change data from on-prem Oracle and move it to cloud data lake. How do I do that?*
- *I have large files on remote servers that need to be loaded to my data lake. How do I do that?*

© Informatica. Proprietary and Confidential.

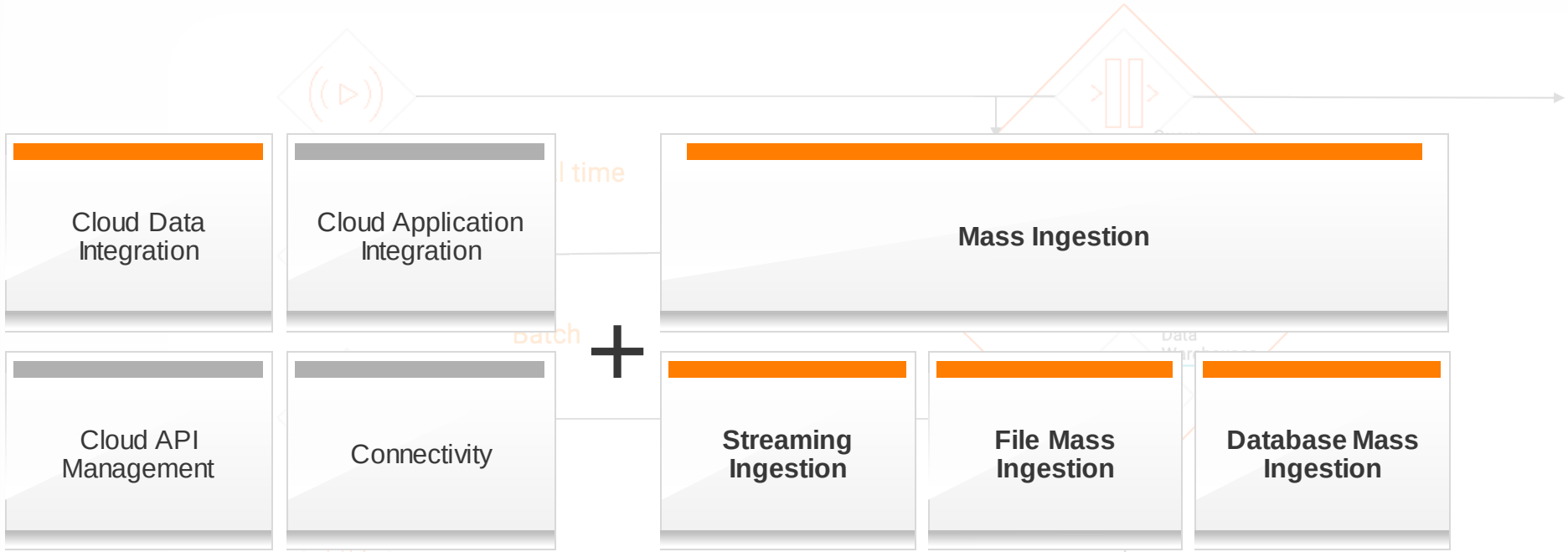
Ingestion Patterns – Batch & Real-time



© Informatica. Proprietary and Confidential.

Informatica

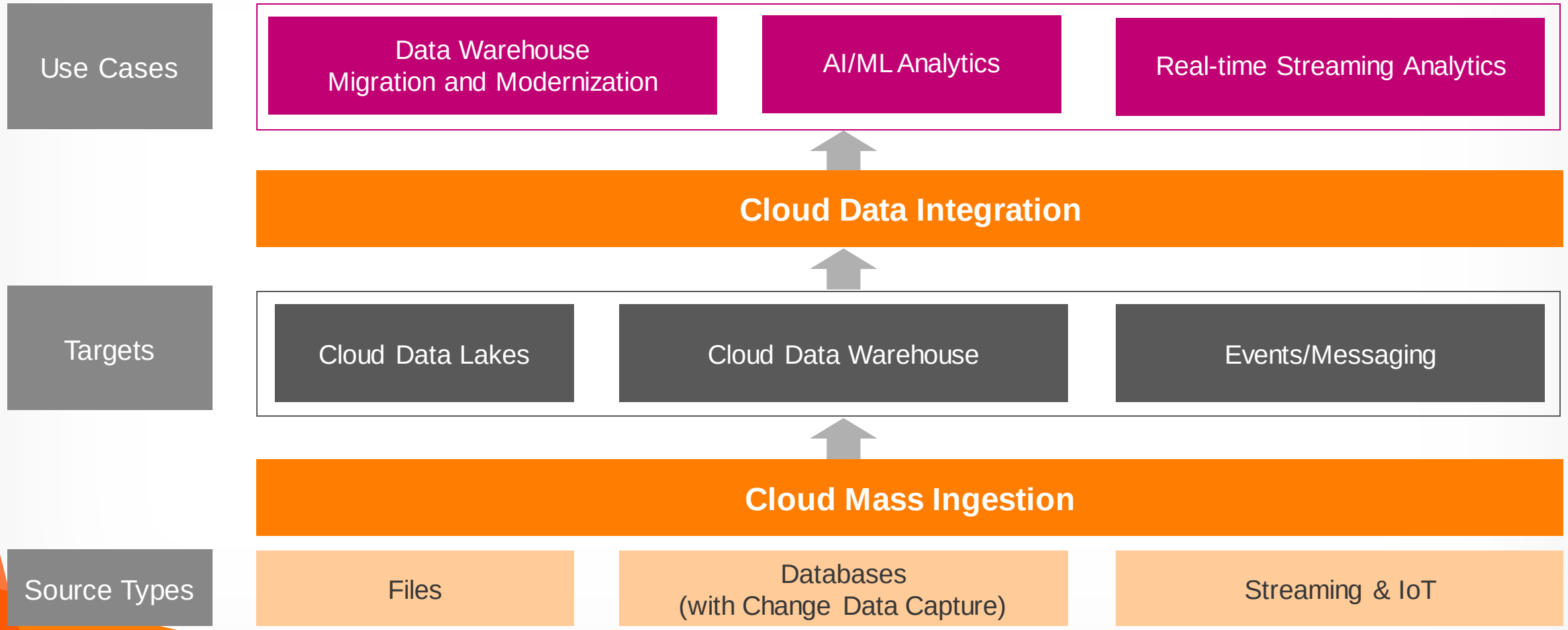
Mass Ingestion Options



© Informatica. Proprietary and Confidential.

Data Lake/Datawarehouse Reference Architecture

- Ingestion feeds into integration..



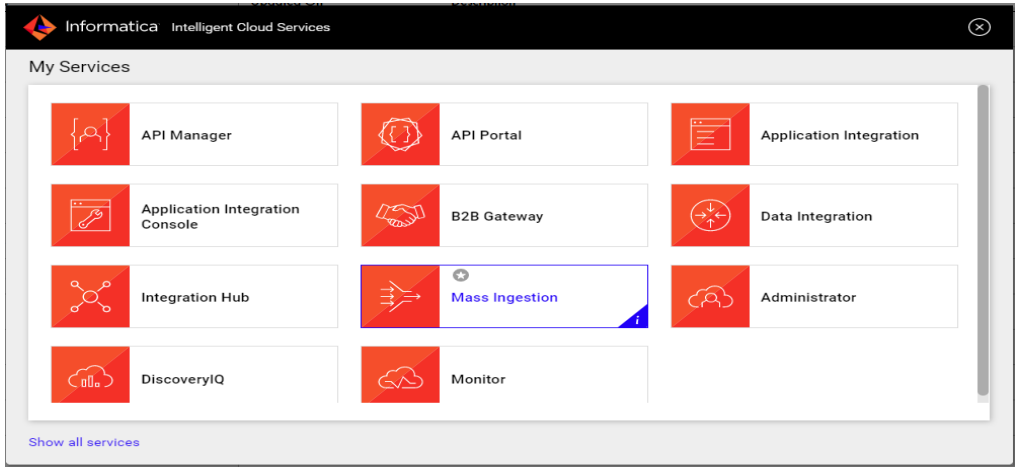
© Informatica. Proprietary and Confidential.

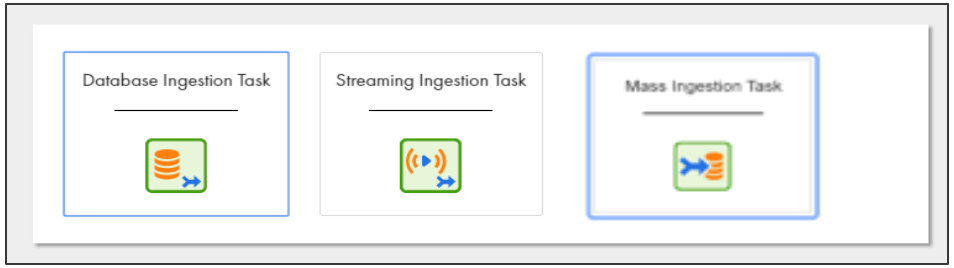
Informatica

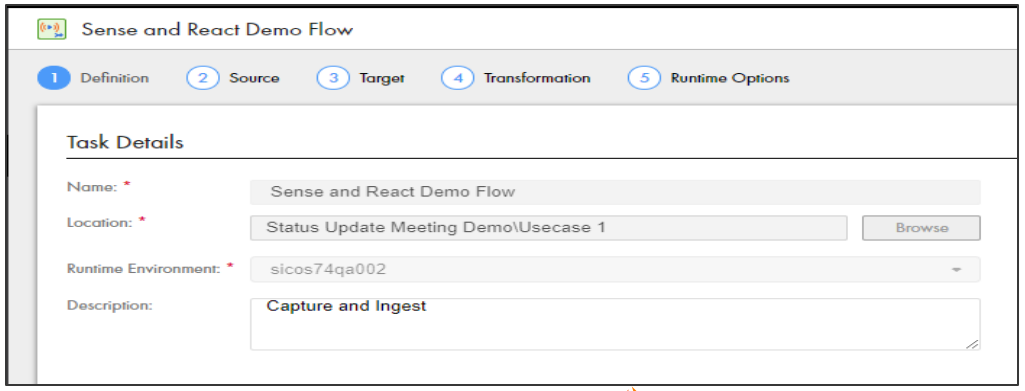
Cloud Mass Ingestion Service –Overview

- Cloud native service for all ingestion uses cases
  - **File**
  - **Database** (initial and incremental - **CDC**)
  - **Streaming** & **IoT**

- Unified user experience for ingestion
  - Common wizard experience for designing
  - Deployment & scheduling
  - Real time monitoring experience







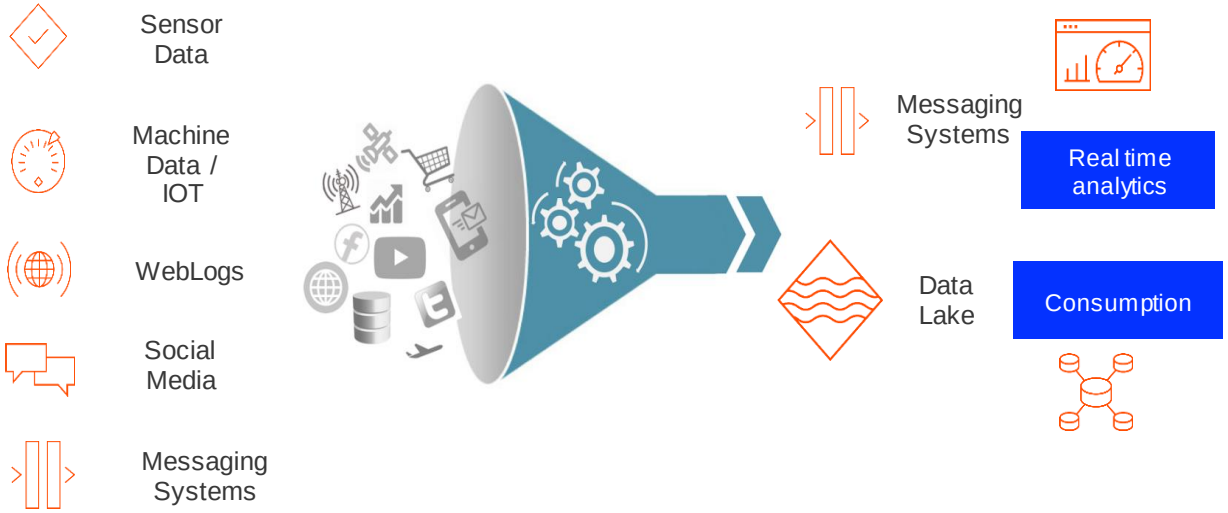
© Informatica. Proprietary and Confidential.

Informatica

Mass Ingestion Streaming

Common Use Cases

- Lake Ingestion
  - Kafka data ingestion onto Cloud data lake
  - JMS Data (from traditional systems) ingestion onto Cloud Data Lake for batch analytics
- Accelerate Kafka adoption
  - IoT data ingestion onto Kafka (with simple filtering)
  - Weblogs and Clickstream ingestion onto Kafka for real-time analytics

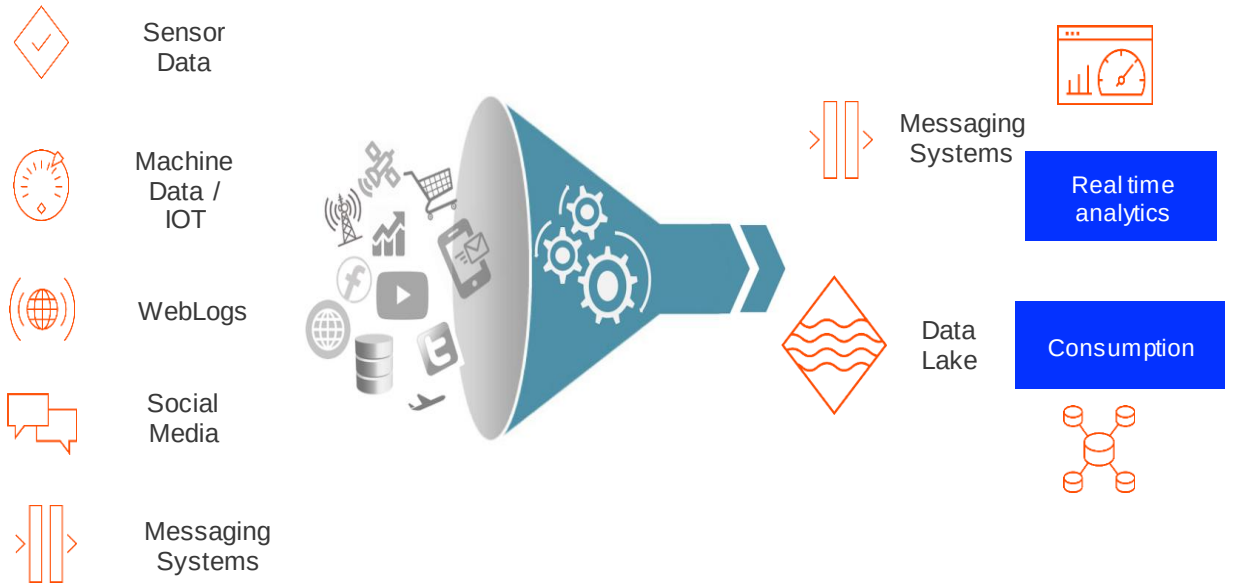


© Informatica. Proprietary and Confidential.

Informatica

Main Capabilities

- Sources – Kafka, MQTT, Tail Log File, JMS, Amazon Kinesis
- Targets – Kafka, Amazon S3, Amazon Kinesis, Amazon Firehose, Azure EventHub, Azure ADLS Gen2
- Transformations
  - ➢ Filter
  - ➢ Segregator
  - ➢ Combiner
  - ➢ Python



© Informatica. Proprietary and Confidential.

Informatica

Benefits

1. Single ingestion solution for all patterns
   - Save time and money
2. Wizard driven experience for ingestion
   - Increase business agility
3. Enable business the ingest streaming data for their usage
   - Faster decision making
4. Edge transformations for cleansing data
   - Increased trust in data assets
5. Connectivity to streaming sources & targets
   - No need to hand code
6. Real time monitoring and alerting
   - Faster troubleshooting

© Informatica. Proprietary and Confidential.

Mass Ingestion Streaming Demo

© Informatica. Proprietary and Confidential.

Informatica

Mass Ingestion Databases

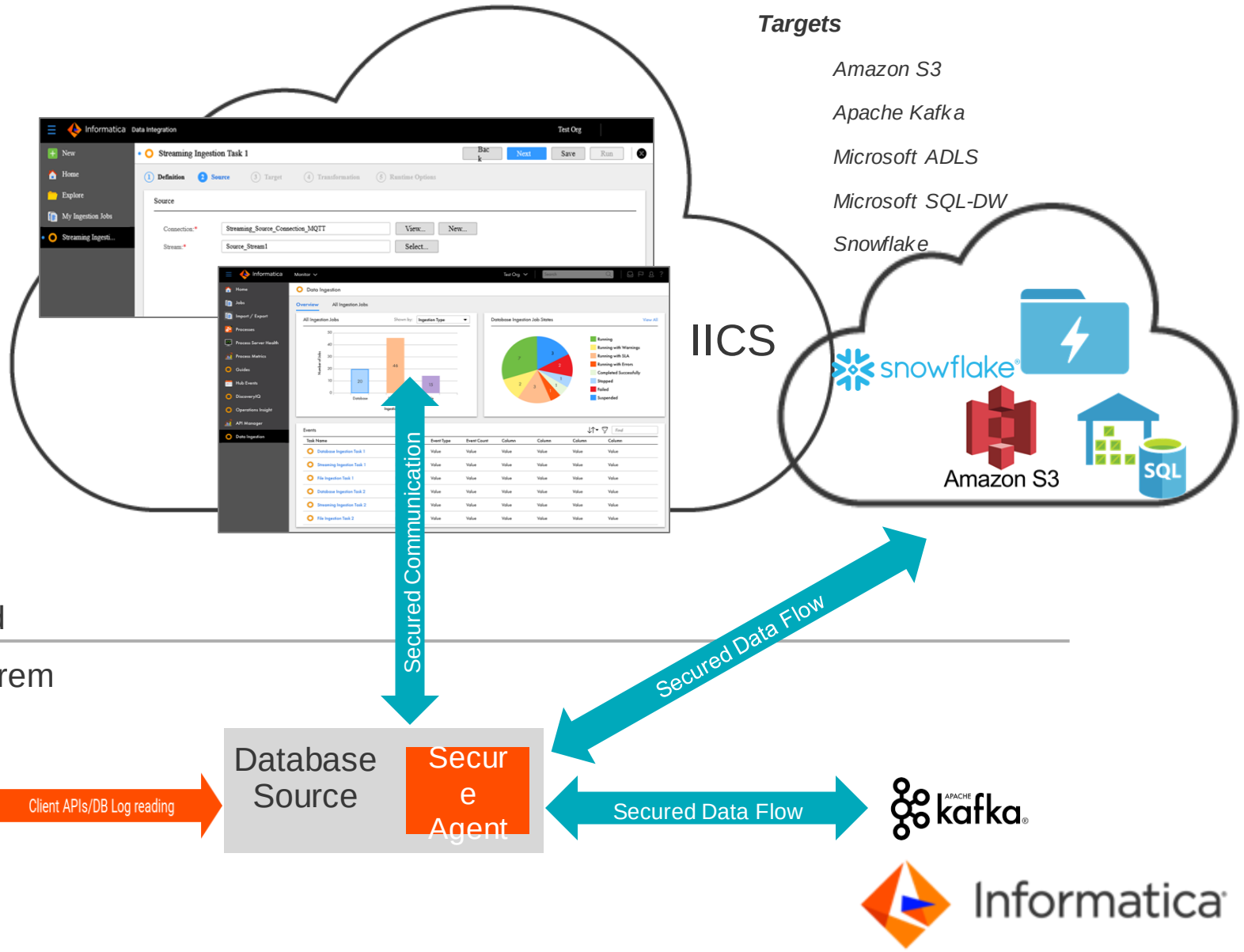
Cloud Mass Ingestion Databases

Provides Database ingestion capabilities as part of IICS Mass Ingestion service

Ingest relational database data from Oracle, SQL-Server & MySQL. Also supporting Schema Drift on CDC supported Databases

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

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



© Informatica. Proprietary and Confidential.

Common Use Cases

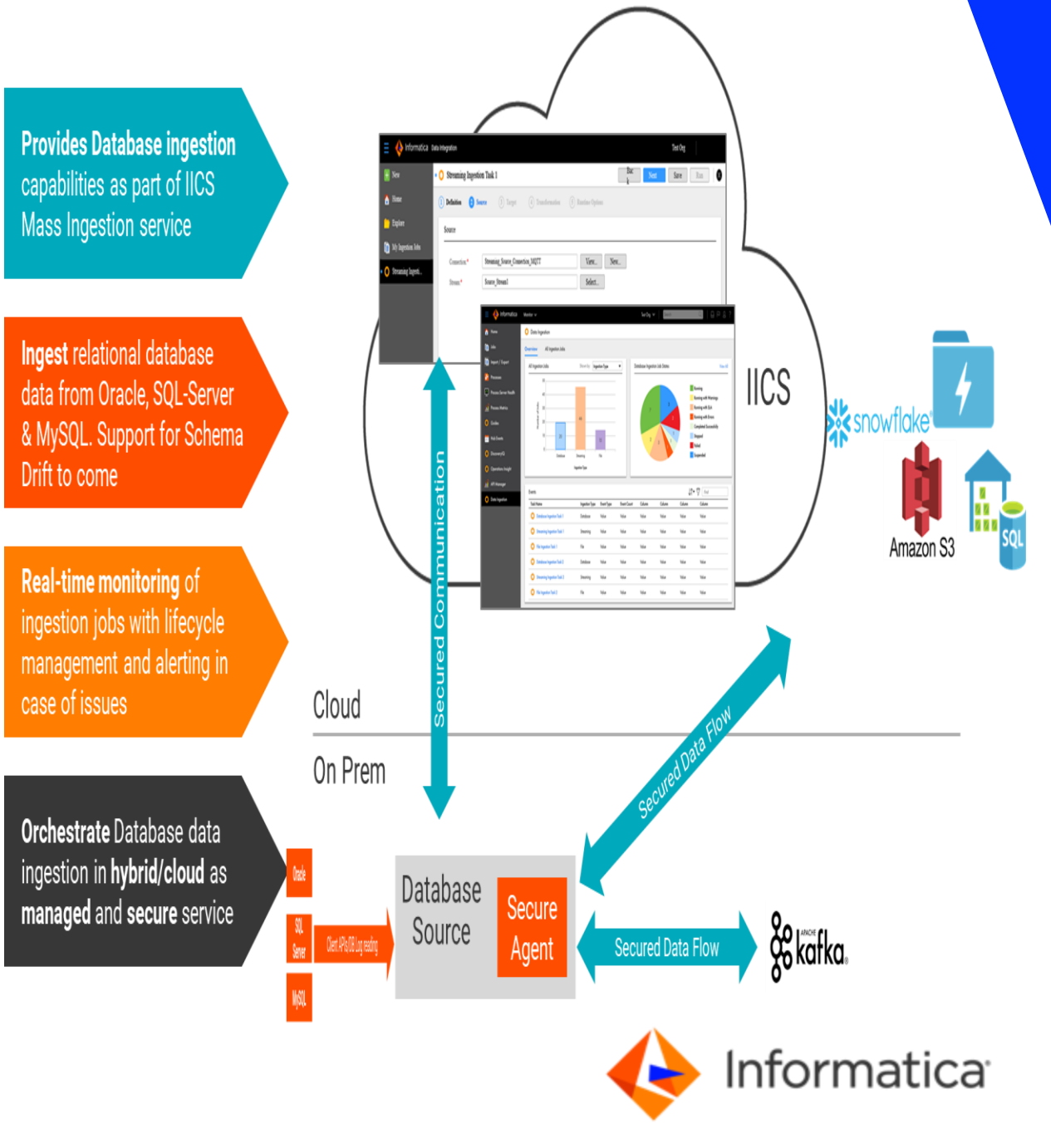
- **Data Lake Ingestion**

  Ingestion of Database content onto Cloud Data Lake

- **Database Migration**

  Database Migration from On-Premises to an Alternate location (and /or type)

  Ingestion of Database content onto Cloud RDBS



© Informatica. Proprietary and Confidential.

Supported Database Ingestion Modes

- Initial Load
  - Lift & Shift of source (selected) database table content to an available target.
- Incremental Load (Continuous)
  - Extraction from (selected) database table logged data to an available target. Will auto react to any detected Schema drift making necessary modifications to target data store.
- Initial & Incremental Load (Continuous)
  - Lift & shift of source (selected) database table content to an available target, once table copy complete will automatically switch to Incremental load by continuing to monitor (same selected) database table logged data to the target. Will auto react to any detected Schema drift making necessary modifications to target data store.

Notes:

- *Continuous, meaning that Ingestion job once deployed and started, will run indefinitely*
- *Incremental (CDC) processing uses database logged data and is not reliant on database tables having any time stamp columns for data identification.*

© Informatica. Proprietary and Confidential.

Mass Ingestion Database Demo

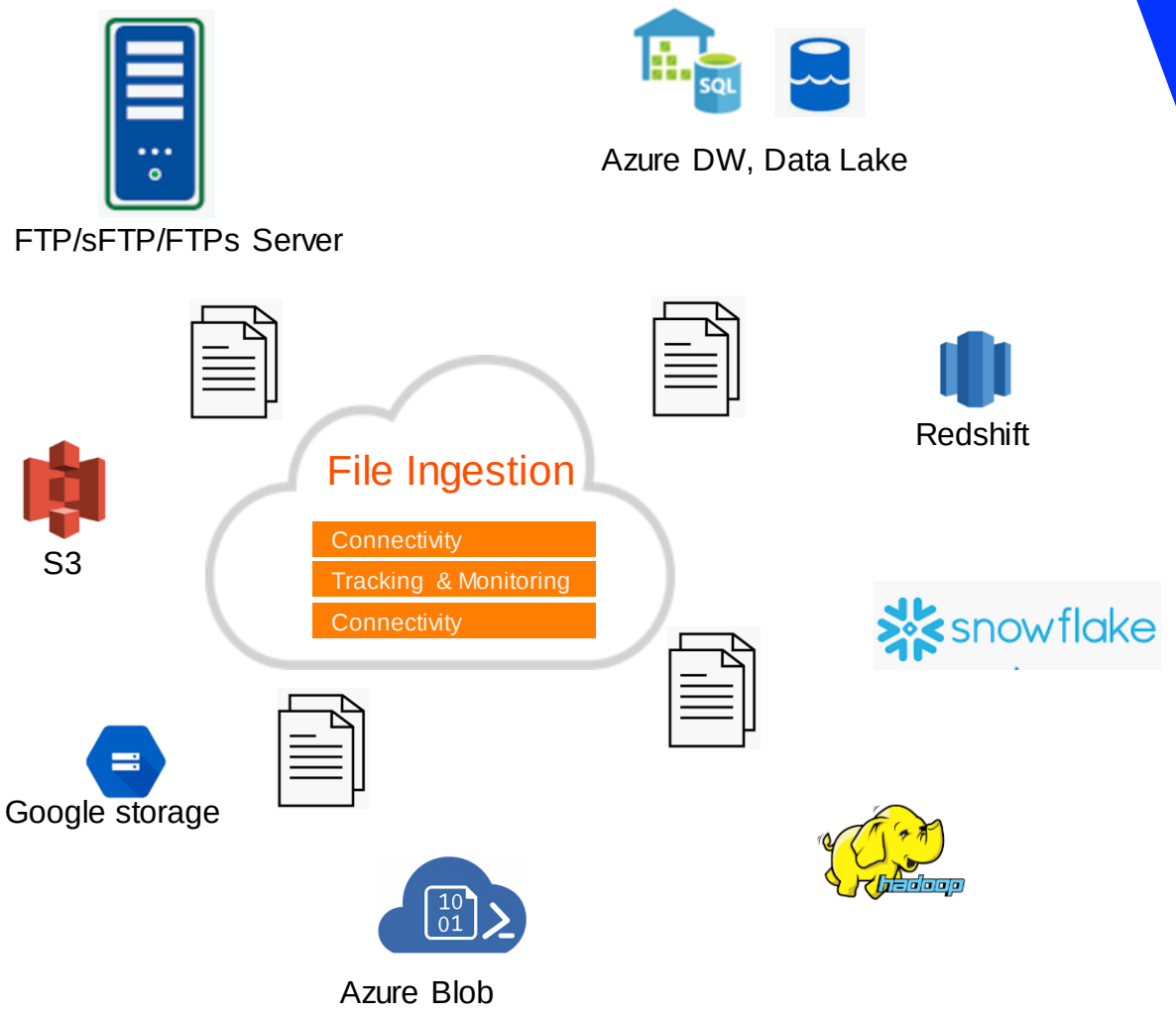
© Informatica. Proprietary and Confidential.

Informatica

Mass Ingestion Files

Common Use Cases

- ➢Lake Ingestion
- ✓ Ingest data that arrive as files to Data lakes and repositories
- ✓ Transfer files from remote FTP/SFTP/FTPS servers
- ✓ No data manipulation is needed. Focus on ingestion.
- ✓ Any file size, any file types

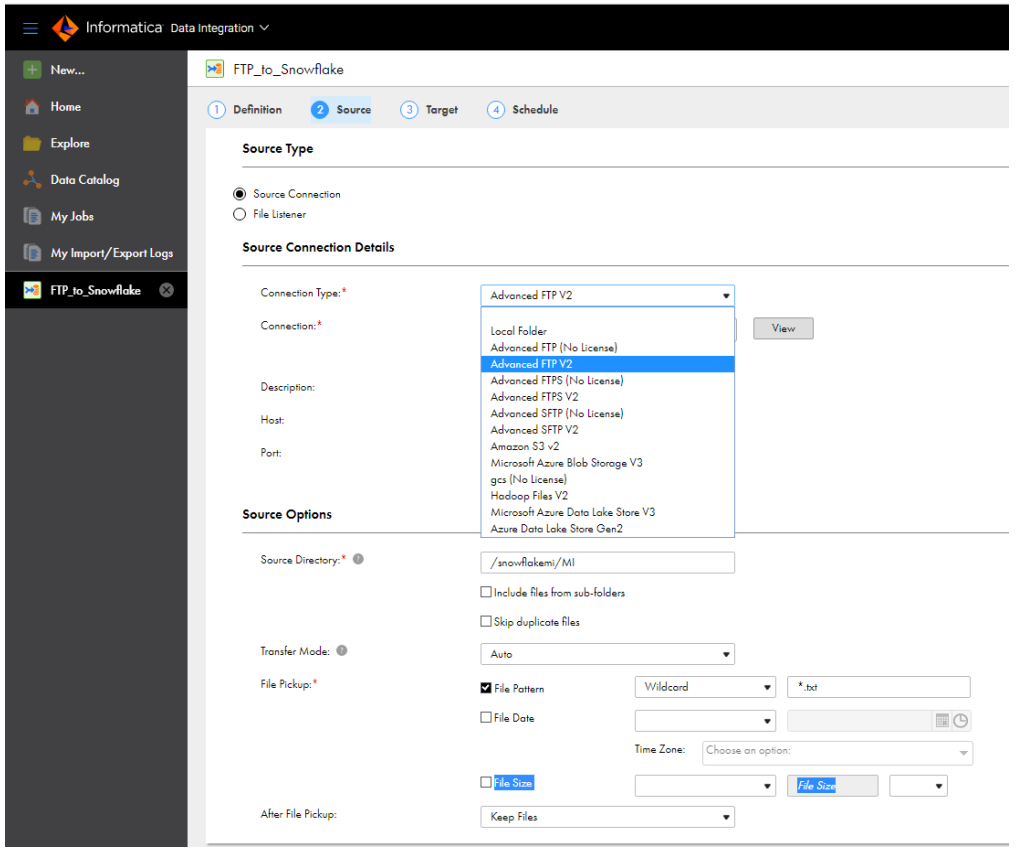


© Informatica. Proprietary and Confidential.

Informatica

Main Capabilities

- Simple, wizard-based task definition
- Wide list of supported sources/targets
- Advanced connectors for handling FTP/SFTP/FTPs
- Filter files by file name pattern, file size, file date

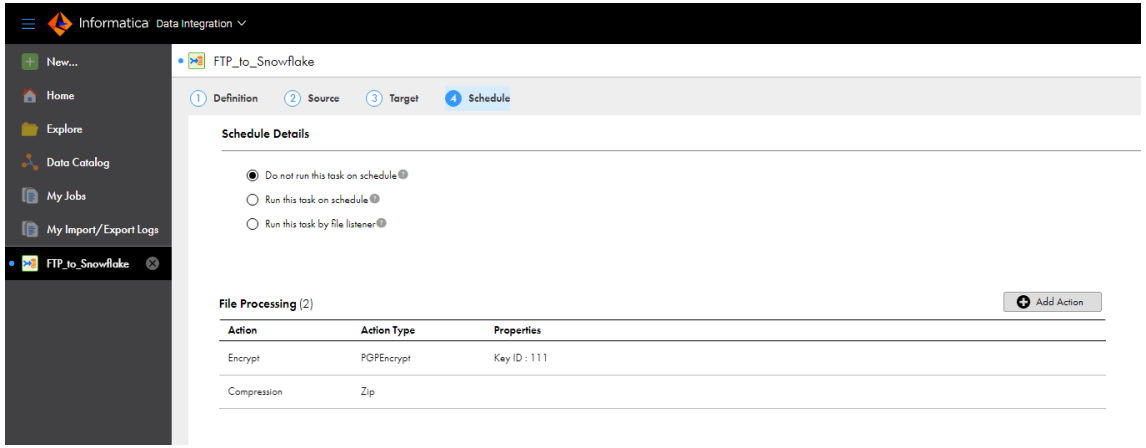


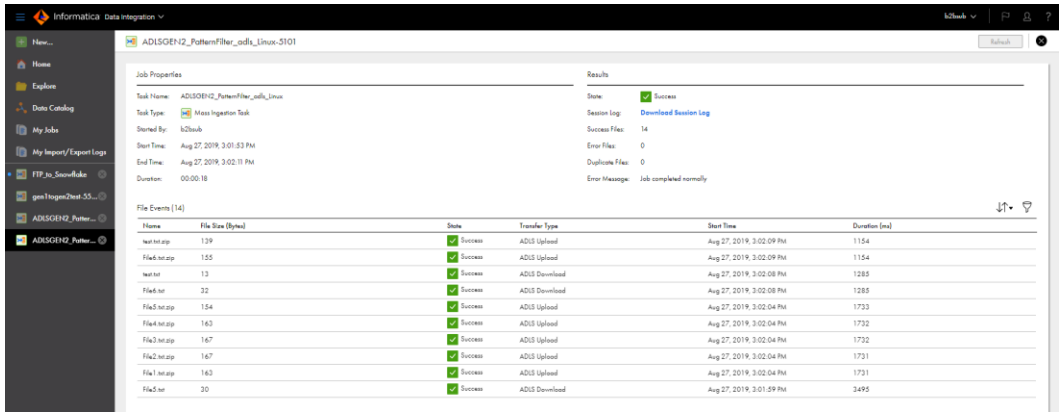
© Informatica. Proprietary and Confidential.

Informatica™

Main capabilities

- API, schedule or file event triggered
- File actions :
  - Compress/decompress (Zip, Gzip ,Tar)
  - Encrypt/decrypt (PGP)
  - Virus scan
- Highly scalable, any file type
- Tracking and monitoring - Job and file level





© Informatica. Proprietary and Confidential.

Informatica

File Mass Ingestion

Supported sources/targets

- Supported sources:
  - Advanced FTP/sFTP/FTPs
  - Amazon S3
  - Azure Blob
  - Azure Data Lake (GEN1, GEN2)
  - Google Storage
  - HDFS file
- Supported Targets:
  - Advanced FTP/sFTP/FTPs
  - Amazon S3 , Redshift
  - Azure Blob, Data warehouse, Data Lake (GEN1, GEN2)
  - Google Cloud Storage
  - Google Big Query
  - HDFS file
  - Snowflake

© Informatica. Proprietary and Confidential.

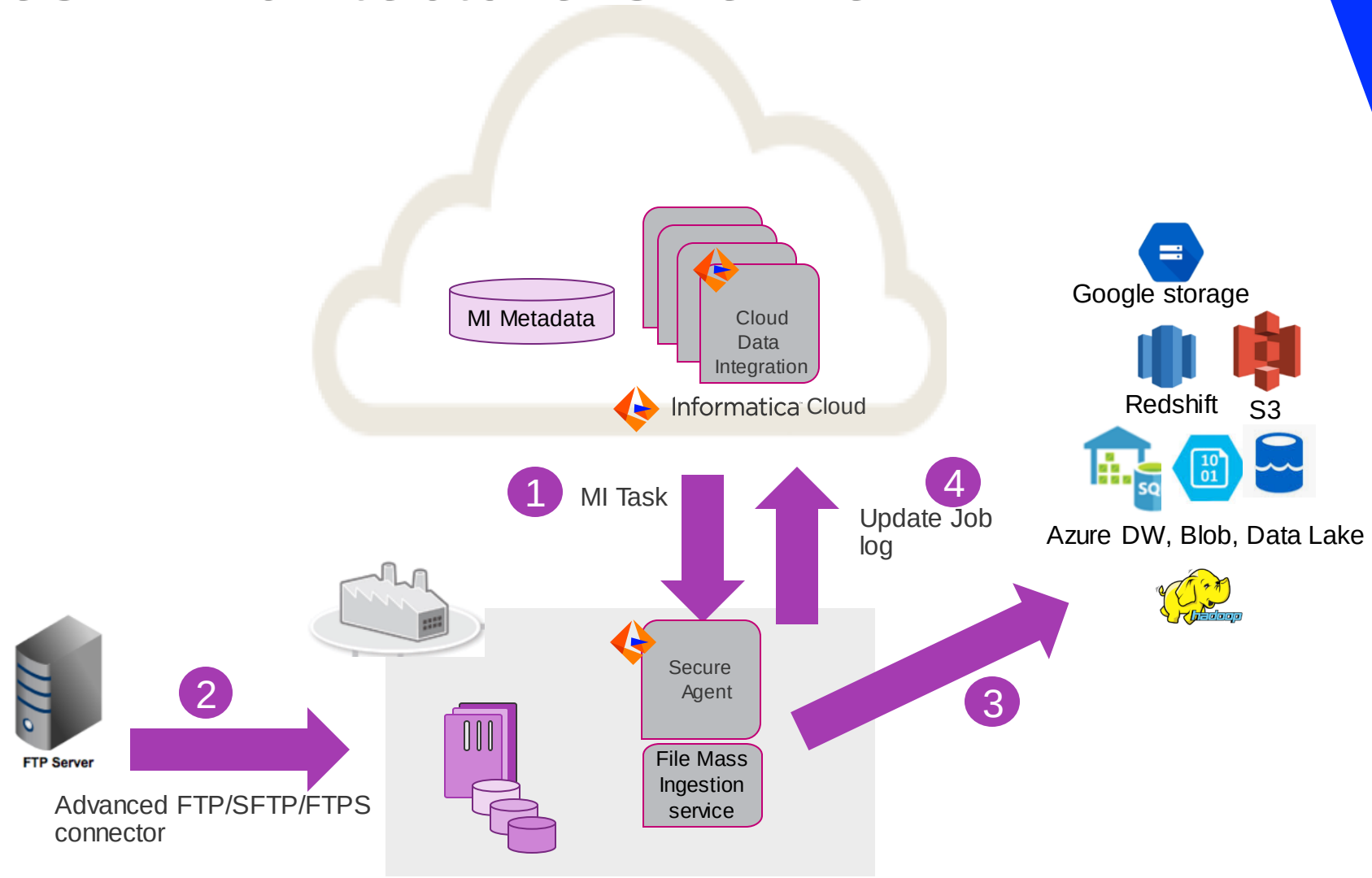
Mass Ingestion Files – Architecture Overview

Provides **file transfer capabilities** for exchanging files between on premise and Cloud repositories, using standard protocols

Transfer **any file type** with a **high performance and scalability**

Job and file level **tracking and monitoring**

Orchestrate File transfer and ingestion in **hybrid/cloud** as **managed** and **secure** service



© Informatica. Proprietary and Confidential.

Informatica

Mass Ingestion Files Demo

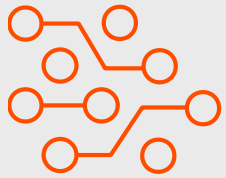
© Informatica. Proprietary and Confidential.

Informatica

Summary

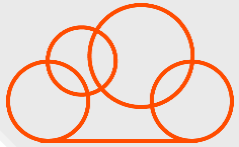
Cloud native ingestion

- IICS service for mass ingestion
- Orchestration for ingestion from variety of patterns



Connectivity

- On-prem Database & CDC
- On-prem & cloud files
- IoT & Streaming
- Cloud data lakes, Datawarehouse and messaging hub



Wizard Driven Design

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

Real-time Monitoring

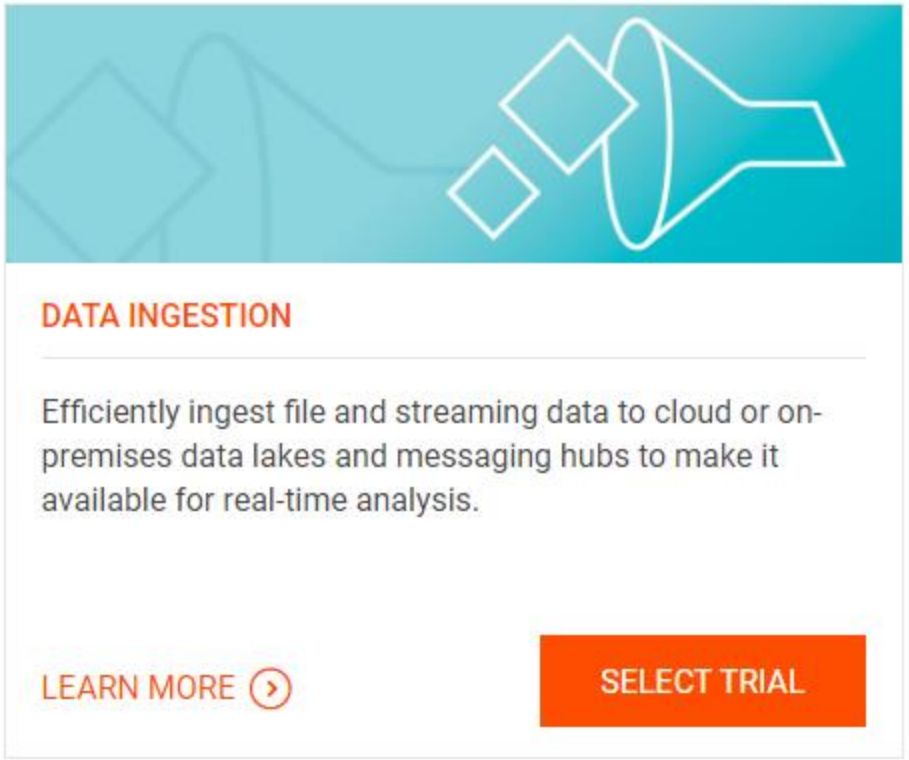
- Pictorial view of the ingestion job
- Real time flow visualization
- Lifecycle management

© Informatica. Proprietary and Confidential.

Informatica

FREE TRIAL – Cloud Mass Ingestion Service!

- Register TODAY here: https://www.informatica.com/trials.html



© Informatica. Proprietary and Confidential.

Learn more..

- **Ingestion home :** https://www.informatica.com/products/cloud-integration/ingestion-at-scale.html
- **Product Documentation:** Homepage
- **Resources:**
  - Mass Ingestion Streaming - Getting Started Video & Getting Started Guide

© Informatica. Proprietary and Confidential.

© Informatica. Proprietary and Confidential.

Thank You