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.
Success Portal

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.
Agenda

1. Data warehousing enhancements
2. Usability
3. Operational Insights
4. Connectivity
5. Serverless Offering
6. Demo
Blueprint for Cloud Data Warehouse/Data Lake

1. **Data Catalog & Data Governance**
   - Discovery
   - Lineage
   - Glossary

2. **Data Ingestion**
   - On-Premises
     - Mainframe
     - Application Servers
     - Databases
     - Documents
     - Data Warehouse
   - SaaS
     - ERP
     - DRM
   - Streaming
     - IoT
     - Machine Data
     - Apps
     - Log files
     - Social
     - Mobile

3. **Data Integration & Quality**
   - Cloud Data Lake
     - Landing Zone
     - Data Enrichment
     - Enterprise Zone
   - Cloud Storage
   - Spark Processing

4. **Data Provisioning**
   - Cloud Data Warehouse
   - Data Integration
   - Data Provisioning

5. **Data Provisioning**
   - Real-time Analytics
   - Enterprise Analytics
   - Line of Business / Self-Service Analytics

6. **Stream Processing**
   - Stream Storage

Business User
Data Analyst
Line of Business
Data Engineer
Data Scientist

© Informatica. Proprietary and Confidential.
Cloud Lakehouse Data Management Pillars

DATA INTEGRATION
- Codeless integration
- Mass ingestion for files, DB and streaming
- Push down optimization
- Serverless and elastic scaling
- Spark-based processing in the cloud
- Broad connectivity
- Stream processing
- MLOps

DATA QUALITY
- Data profiling
- Data quality rules
- Dictionaries to manage values lists
- Cleansing, standardization, parsing, verification and deduplication/consolidation processes
- Integrated into data integration
- Data quality analytics

METADATA MANAGEMENT
- Data discovery
- End-to-end lineage
- Metadata – technical, business, operational, usage
- Connect and scan metadata – databases (DW, DL), apps, ETL, BI tools and others
- Common metadata foundation

AI-DRIVEN AUTOMATION

CLOUD NATIVE
- Multi cloud
- API driven
- Microservices
- Containerization
- Serverless architecture
- Minimal install and setup
- Auto-upgrades
- Usage-based pricing
- Trust certifications

© Informatica. Proprietary and Confidential.
# Cloud Lakehouse Data Management Pillars

## Data Integration
- Codeless integration
- Mass ingestion for files, DB and streaming
- Push down optimization
- Serverless and elastic scaling
- Spark-based processing in the cloud
- Broad connectivity
- Stream processing
- MLOps

## Data Quality
- Data profiling
- Data quality rules
- Dictionaries to manage values lists
- Cleansing, standardization, parsing, verification and deduplication/consolidation processes
- Integrated into data integration
- Data quality analytics

## Metadata Management
- Data discovery
- End-to-end lineage
- Metadata – technical, business, operational, usage
- Connect and scan metadata – databases (DW, DL), apps, ETL, BI tools and others
- Common metadata foundation

## AI-Driven Automation

## Cloud Native
- Multi cloud
- API driven
- Microservices
- Containerization
- Serverless architecture
- Minimal install and setup
- Auto-upgrades
- Usage-based pricing
- Trust certifications
Data Warehousing

1. Transformations
2. Parameterization
3. Taskflow
Transformation Enhancements

- Support for Transaction Control Transformation.
- Support for dynamic file name creation for targets.
- Union transformation can now support more than 2 input groups.
- Support for additional Lookup properties for cached-lookup.
Transformation Enhancements

• System variables can be used in creating expressions.
• Join types extended for multi-object sources.
• Target update SQL override support for relational targets like Oracle, mySQL, SQL Server and ODBC.
• Support for target load order (Flow run order) in the mapping designer.
## Transformation Enhancements

<table>
<thead>
<tr>
<th><strong>DATA INTEGRATION</strong></th>
<th><strong>DATA QUALITY</strong></th>
<th><strong>STRUCTURE PARSER</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transaction Control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Target TX - Dynamic target file names, Target Update</strong></td>
<td><strong>Parser</strong></td>
<td><strong>Support for hierarchical output types – JSON, XML, Avro, Parquet</strong></td>
</tr>
<tr>
<td><strong>SQL override</strong></td>
<td><strong>De-duplicate</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Union Transformation to support more than 2 groups.</strong></td>
<td><strong>Already Available:</strong></td>
<td><strong>Support for Real time source (Kafka)</strong></td>
</tr>
<tr>
<td><strong>Persistent cache Lookup</strong></td>
<td><strong>Rule Specification</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Extend join types in source transformation for related objects</strong></td>
<td><strong>Verifier</strong></td>
<td><strong>Support pass through fields</strong></td>
</tr>
<tr>
<td><strong>System variables in Expression Transformation</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additional licenses are required to use Data Quality and Structure Parser transformations.
Parameterization
Parameterization Enhancements

- Parameter files can now be accessed from Cloud storages like S3/ADLS.
- Generate parameter file template easily.
- Support for pre/post SQL parameterization, Lookup SQL override.
- Request message parameterization for Workday/Oracle Financials.
- Override Source/target File Directory, Session log file name in mapping tasks.

```plaintext
#USE_SECTIONS
$$oracleConn=Oracle(SK)
$$city=SF
[Global]
$$if_conn=FF ja_con
$$st=CA
[Default].[Accounts].[April]
$$QParam=SELECT * from con.ACCOUNT where city=LAX
$$city=LAX
$$tab=accounts.csv
$$oracleConn=Oracle_Src
$$state=$st
```
Task flows
Support for file ingestion tasks and file watch step

• New Ingestion step to include File mass ingestion tasks in task flow.

• File-watch step – Based on a file listener task and wait for file arrival mid-stream in the task flow.
Usability
Mid-stream data preview

- In the mapping designer, you can now preview the output of each transformation as you build the mapping.
- The mapping pipeline up-to the point of preview needs to be valid (not the entire mapping).
- Data preview is a ‘mini-job’ that is run using a secure agent and can be monitored.
- By default first 100 rows are fetched. This can be configured.
- The preview results can be downloaded.
- Supported only for CDI mappings in R34.
Session log Enhancements

- Session log file now includes
- Task name and IICS FRS ID.
- Agent Group name and Id
- Agent name and Id.
- Log file name when downloaded contains the mapping task name (not just logX anymore!)
- Timezone information
- Transformation names for Sources and Targets.
Operational Insights -
(Cloud Data Integration Jobs)
Operational Insights for Cloud Services

- Infrastructure monitoring – Secure agent health, resource consumption and ability to set alerts.

- Cloud Data Integration jobs monitoring - Aggregated analytics and visual dashboards with job history for last 30 days.

- Cloud Application Integration – Preview in Spring 2020 release

- Availability - In all PODs with the Spring release.
  - Included with the Data Integration/Cloud Application Integration subscription.
  - Please reach out to your customer success manager or raise a support case if the service is not enabled on any of your Informatica orgs.
Jobs heat map and scheduled jobs

- Quickly identify the peak time with the jobs heat map.
- Heat map helps you analyze the impact on your integration jobs and optimize resources and schedules.
- View the number of upcoming scheduled jobs for the next 24 hours.
Historical job runs

• Visualize the previous runs of any integration job to spot anomalies in the task run.

• Historical job runs available for the last 30 days.

• Export the data to csv for further analytics and reporting needs.
Project level and Secure Agent view of activity

- Dashboards that provide project/folder level overview of the data integration jobs.
- Quickly identify business units that have job failures and need attention.
- Compare the load on your secure agents to optimize resource and scheduling of the integration jobs.
Connectivity Highlights
Big Push for Pushdown Optimization
Ecosystem PDO for AWS

- Push mapping logic to AWS to leverage COPY command
- To load data from S3 to Redshift
  - Using NATIVE connectors
  - Basic transformations between S3 and Redshift: Filter and Expression
- Roadmap – More transformations and connectivity!
Big Push for Pushdown Optimization
Snowflake Enhancements (via ODBC)

- Sequence Generator
- Unconnected Lookups
- Multiple Targets/Target Instances
- Insert/Update/Upsert with targets
- Source Query
- Additional Functions, such as MD5
- Create Target
Azure – R34 Enhancements

• Azure Synapse (SQL DW V3)
  - ADLS Gen2 as temp storage for Polybase
  - Mass Ingestion from ADLS Gen2 to Synapse
  - Unconnected Lookup
  - Logging enhancements
  - Source/Target overrides with Parameter Files
  - Partial Parameterization (database, schema, and table) support in PRE/POST SQL, SQL/LOOKUP Override
  - Performance enhancements

• ADLS Gen2
  - File Port support
  - Azure Gov Cloud
  - User Authenticated Proxy
  - Source/Target overrides with Parameter Files
  - New Data types for Parquet files (Date/Datetime/Decimal)
  - Performance Enhancements
  - New Connector for CDI-E service (tech preview)

• CDM Folders (Tech Preview)
  - Support for CDM Schema v0.9
  - Integration with new MSFT CDM Folders SDK
  - Support for Unicode Characters

• Blob Storage (Blob V3)
  - Source/Target overrides with Parameter Files
Google – R34 Enhancements

• **Google BigQuery**
  - CDC as target support for Google BigQuery GBQV2
  - Merge support for Upsert operation for GBQ V2
  - GBQ V2 – cached and uncached lookup support
  - GBQ V2 – SQL override support

• **Google Cloud storage V2**
  - Folder and Subfolder support for Google Storage connector.
Serverless Runtime for CDI – Preview in R34
Serverless Offering (New!)

- Web browser (Build & manage)
- Outbound connection HTTPS: 443

**Metadata**

- Compute
  - Auto-Scaling
  - Multiple AWS regions
  - Resiliency and HA
  - Tenant Isolation

- DMZ

- Microservices

- Informatica VPC

**Customer VPC**

- Data
  - Amazon Redshift
  - Amazon RDS

- Data Warehouse

- ERP

- Databases

- VPN

- Pluggable micro-service based engines
- Auto upgrade
- High Availability with Zero downtime
- Clustering - Agent groups
- Secure - HIPAA, SOC2, PCI

© Informatica. Proprietary and Confidential.
Informatica Serverless in IICS R34 (Preview)

• Which IICS services supports serverless feature?
  - CDI and CDI-elastic only

• Which Cloud ecosystem is supported by serverless feature?
  - AWS only

• What sources and targets will be supported?
  - For CDI-elastic: S3, Redshift, JDBC.
  - For CDI: RedShift, RDS, Dynamo DB. Also any supported on-premise data source to which there is a connection from the subnet used in serverless config.

Interested in the Preview? - Please reach out to your customer success manager or account manager!
Connectivity Highlights
Big Push for Pushdown Optimization

Ecosystem PDO for AWS

• Push mapping logic to AWS to leverage COPY command
• To load data from S3 to Redshift
  - Using **NATIVE** connectors
  - Basic transformations between S3 and Redshift: Filter and Expression
  - Roadmap – More transformations and connectivity!
Big Push for Pushdown Optimization

**Snowflake Enhancements (via ODBC)**

- Sequence Generator
- Unconnected Lookups
- Multiple Targets/Target Instances
- Insert/Update/Upsert with targets
- Source Query
- Additional Functions, such as MD5
- Create Target
Azure – R34 Enhancements

• **Azure Synapse (SQL DW V3)**
  - ADLS Gen2 as temp storage for Polybase
  - Mass Ingestion from ADLS Gen2 to Synapse
  - Unconnected Lookup
  - Logging enhancements
  - Source/Target overrides with Parameter Files
  - Partial Parameterization (database, schema, and table) support in PRE/POST SQL, SQL/LOOKUP Override
  - Performance enhancements

• **ADLS Gen2**
  - File Port support
  - Azure Gov Cloud
  - User Authenticated Proxy
  - Source/Target overrides with Parameter Files
  - New Data types for Parquet files (Date/Datetime/Decimal)
  - Performance Enhancements
  - New Connector for CDI-E service (tech preview)

• **CDM Folders (Tech Preview)**
  - Support for CDM Schema v0.9
  - Integration with new MSFT CDM Folders SDK
  - Support for Unicode Characters

• **Blob Storage (Blob V3)**
  - Source/Target overrides with Parameter Files
Google – R34 Enhancements

- **Google BigQuery**
  - CDC as target support for Google BigQuery GBQV2
  - Merge support for Upsert operation for GBQ V2
  - GBQ V2 – cached and uncached lookup support
  - GBQ V2 – SQL override support

- **Google Cloud storage V2**
  - Folder and Subfolder support for Google Storage connector.
Demo
Q & A
Thank You