



10 Sep, 2024

CDI Implementation Strategy, Deployment and Best Practices

- Bala Kaleeswaran, Solution Architect, IPS
- Pavan Kumar Soma Prabhu, Principal Services Solutions Architect, IPS
- Giri Chintarlapalli, Sr Principal Consultant, IPS

Where data & AI come to The LIFE logo, where the letters L, I, F, and E are rendered in a colorful, multi-colored font.

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 [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.

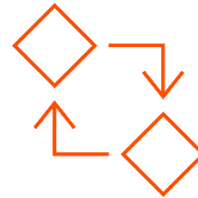
Feature Rich Success Portal



**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.

SQL ELT

Need for ELT

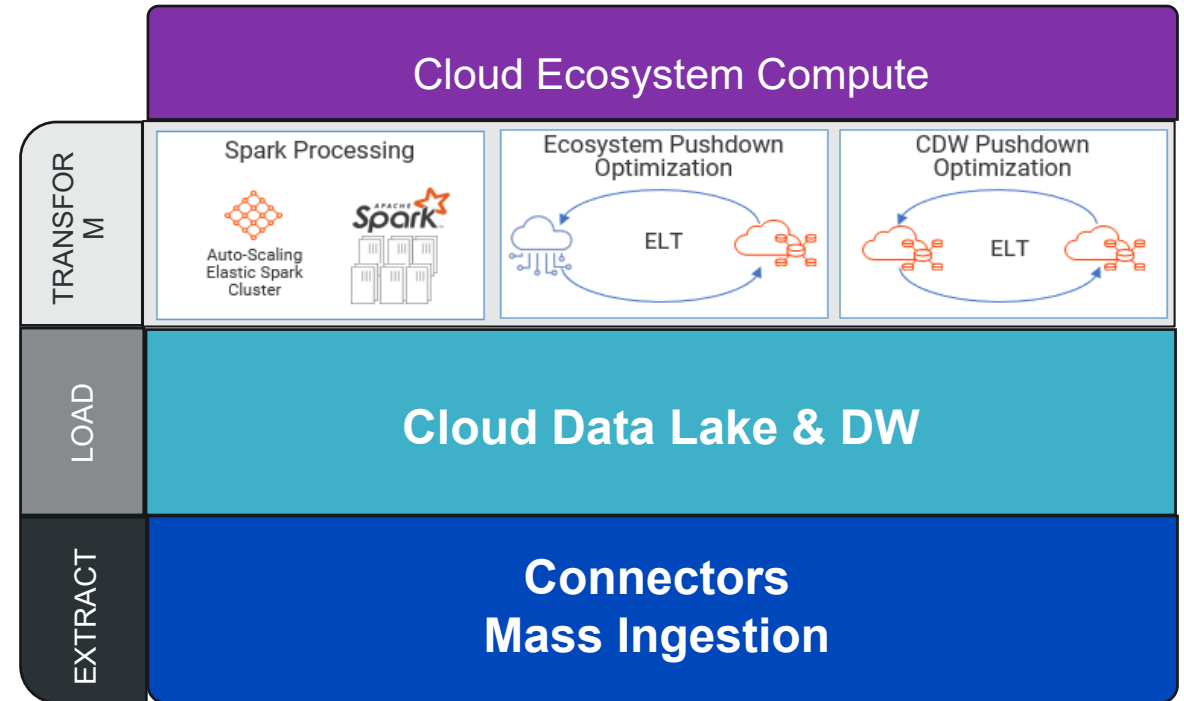
- Faster Data Ingestion
- Raw data retention
- Large amounts of data, processing

Solution

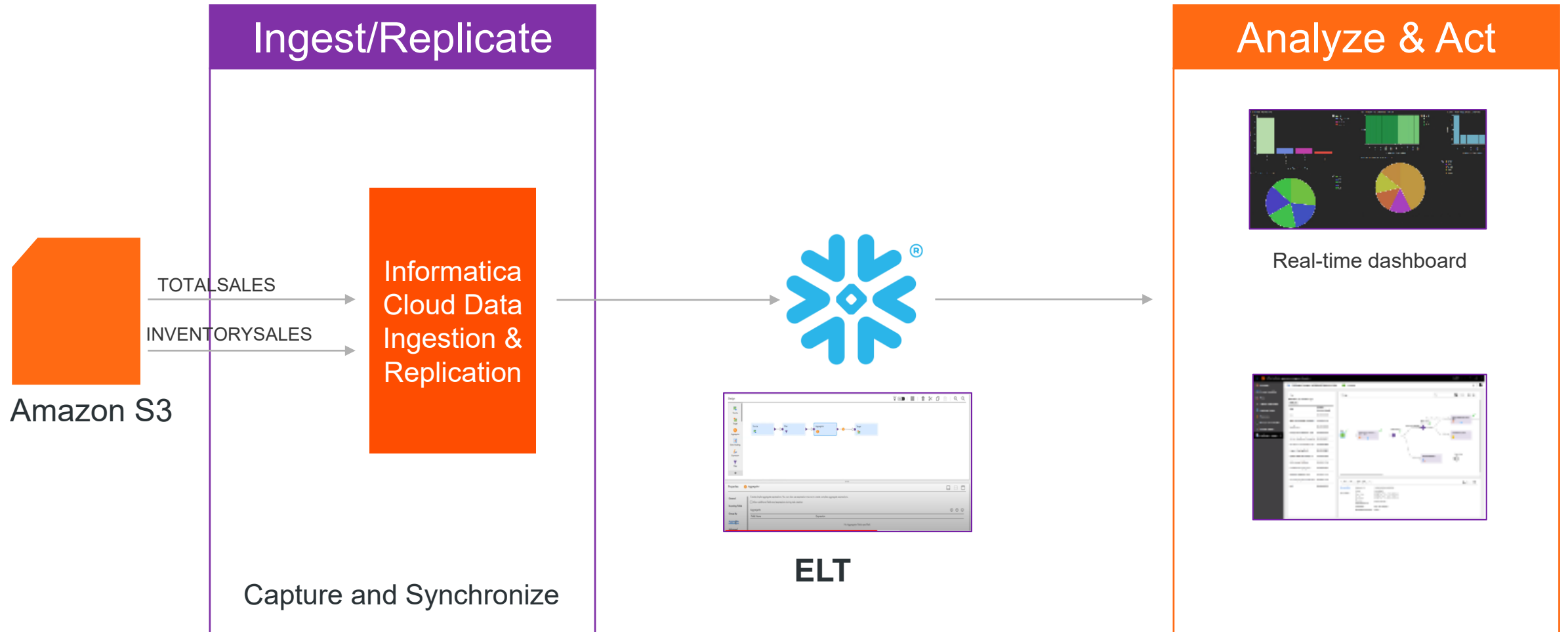
- Choice of SQL ELT experience
- Cost and performance optimized

Benefits

- ✓ Simple and Intuitive GUI
- ✓ Native CDW compute
- ✓ Ecosystem context aware
- ✓ Additional functions specific to CDW



Demo – Data Flow



Supported Sources (August 2024)

Target	Supported sources
Snowflake Data Cloud	<ul style="list-style-type: none">• Snowflake Data Cloud• Amazon S3• Microsoft Azure Data Lake Storage Gen2
Databricks	<ul style="list-style-type: none">• Databricks• Amazon S3• Microsoft Azure Data Lake Storage Gen2
Google BigQuery	<ul style="list-style-type: none">• Google BigQuery• Amazon S3• Google Cloud Storage
Amazon Redshift	<ul style="list-style-type: none">• Amazon Redshift• Amazon S3

Rules and Guidelines - Subset

General Guidelines

- Input parameters
- Precision matching

Data Types

- No Time and TIMESTAMPTZ data types.
- No array

Functions

- Dateadd function - DATEADD('Year', <Value>, <Field name>)
- TO_NUMBER(TO_BOOLEAN(bool_field))

Target Operations

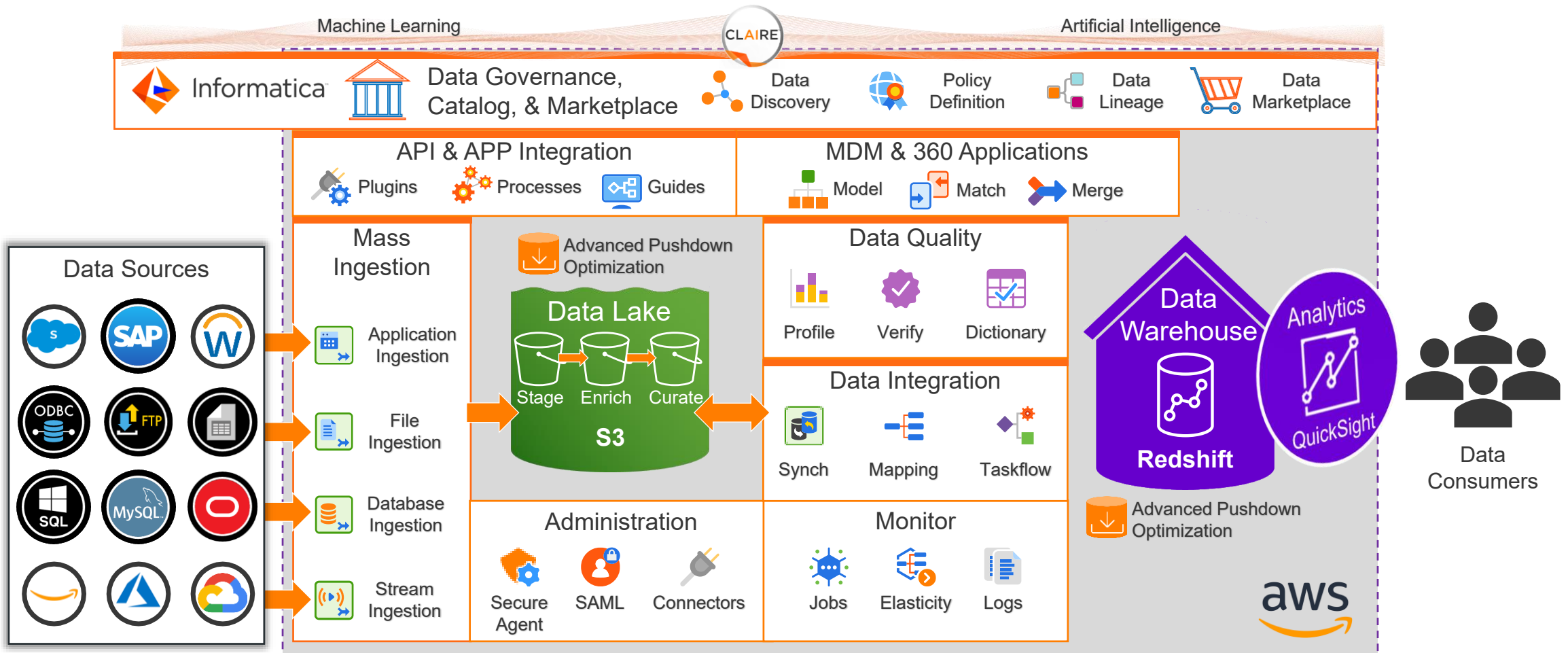
- Values in Warehouse and Role
- Upsert considerations

Expression Validations

- Function with an expression that returns the String data type and the output port is Integer
- Ensure that the string argument for the Date or Time function is valid

Enterprise Cloud Data Management Solution for AWS

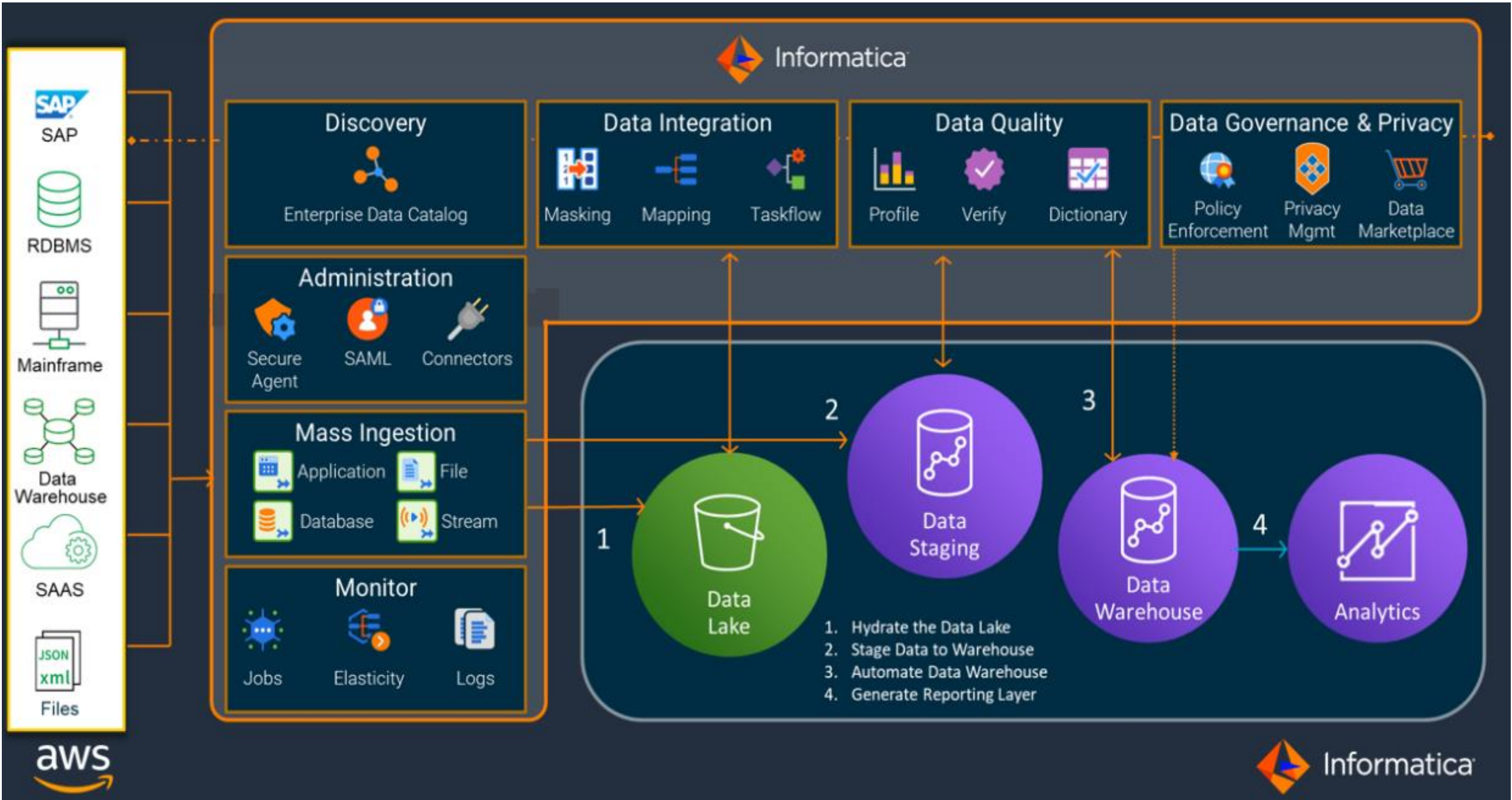
Informatica's Intelligent Data Management Cloud



Lakehouse Management – Ingestion & Integration

Ingestion Challenges	How Cloud Data Ingestion helps
Speed	Ingestion job in minutes
Complexity	Single, unified Cloud native
Scalability	Ingest terabytes of any data
Flexibility	Track, capture & update changed data

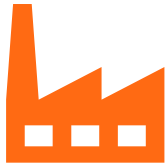
Data Integration Challenges	Cloud Data Integration to the rescue!
Flexibility & Security	Flexible and Secure
Scalable/ Varying patterns	Pattern aware and Scalable
Adapt	Intelligent and Adaptive
Performance	Optimized performance for Data Processing



IDMC – CI/CD integration

Where data & AI come to 

CI/CD Integration Solutions



Asset management utility
– CLI



IDMC Rest API

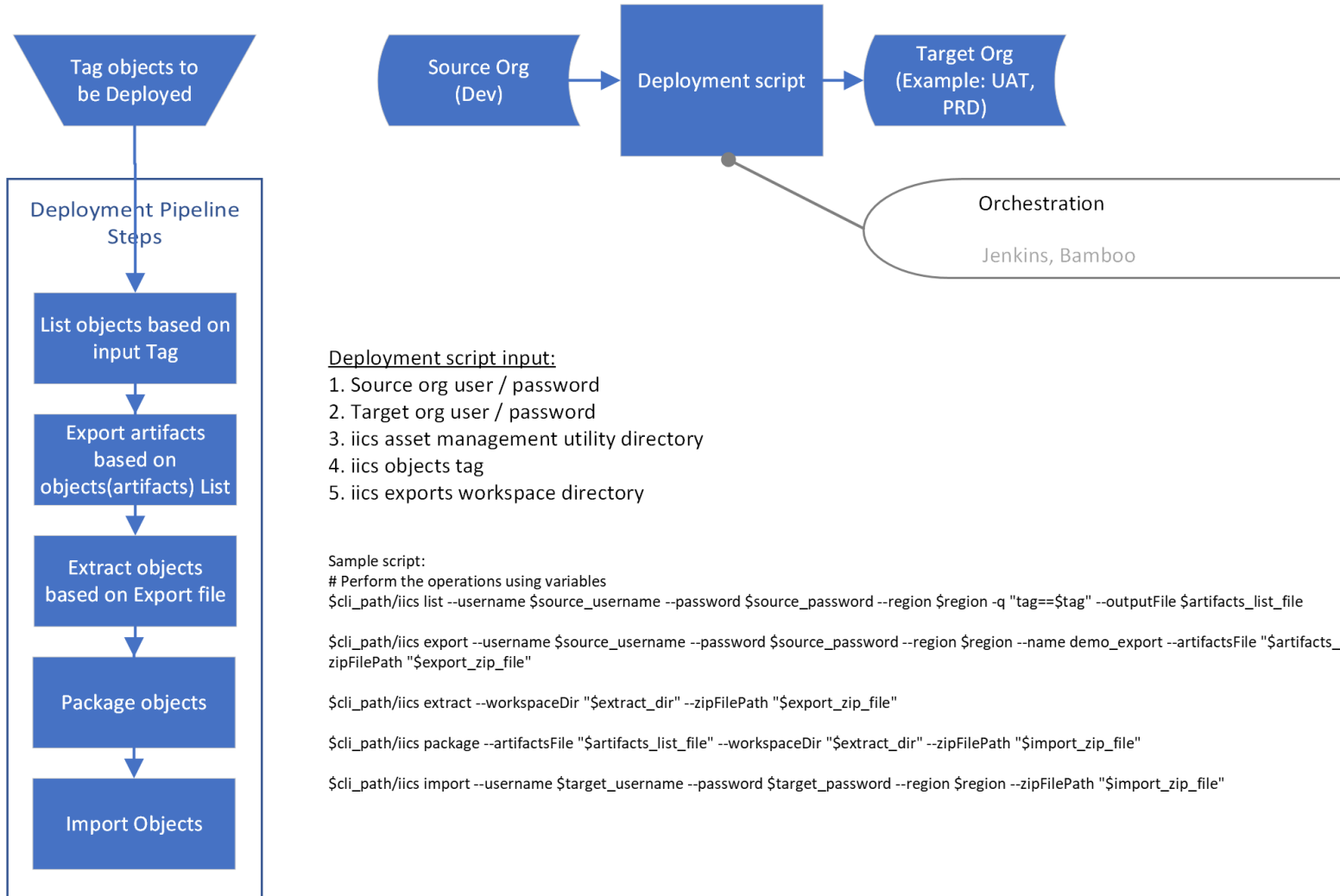


Assurance service

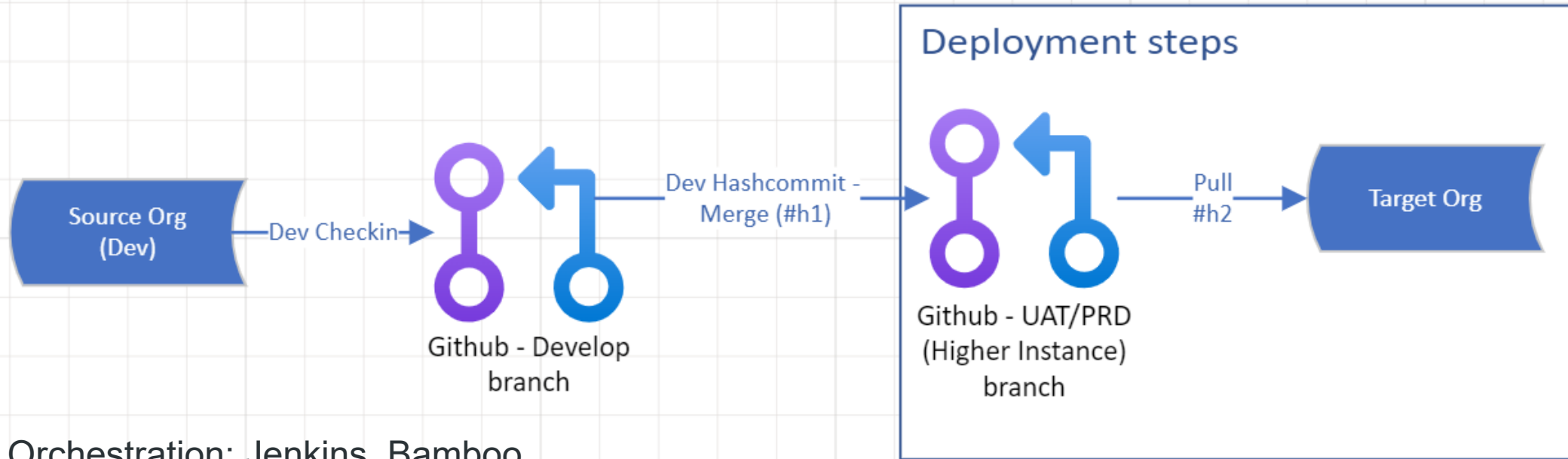


Version control
integration

CI/CD Process flow – CLI and Rest API



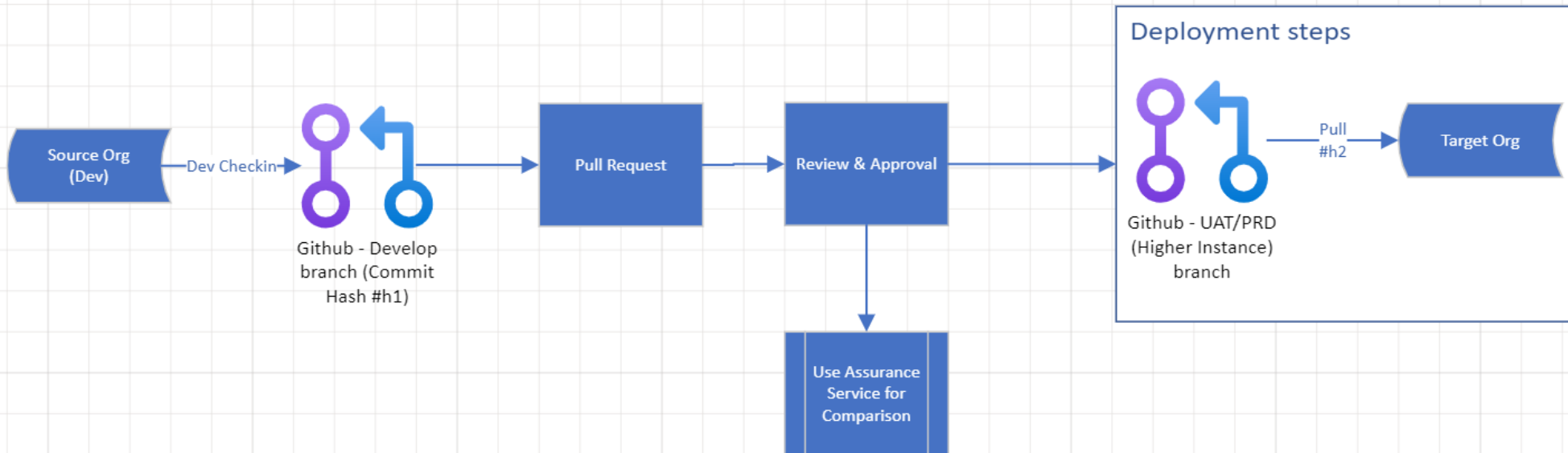
CI/CD Process flow – Source control integration



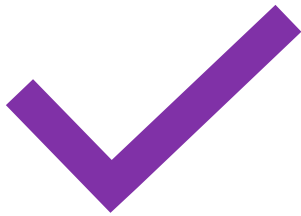
Orchestration: Jenkins, Bamboo,
Source control workflow actions

1. Clone repository
2. Checkout and Fetch
3. Merge (hash => target branch)
4. Git Push
5. Pull from Git to Informatica target org

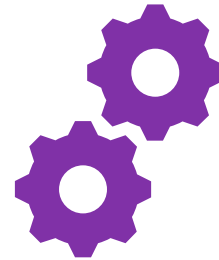
CI/CD Process – Source code Integration – Pull request based



Pre-requisites



Enable
GitRepoConnect
service



Enable source control



Enable source control
user profile

Demonstration

- Asset deployment utility
 - Orchestration through bash script
 - Recommendation: Orchestration with Jenkins, Bamboo
- Source code integration – Github – Automatic merge
 - Orchestration with python
 - Recommendations: Use source control workflows, Jenkins, Bamboo

IDMC – Best practice

Where data & AI come to **LIFE**



CDI Implementation Guidelines and Best Practices

Pavan Kumar Soma Prabhu
Principal Services Solutions Architect , IPS

Where data & AI come to 

Agenda

- Migration from OnPrem to Cloud
- Mapping Guidelines
- Naming Conventions
- Performance Tuning
- Testing Mechanisms
- Secure Agent Sizing Guidelines
- Error Handling
- CI/CD
- Job Monitoring

Migration from OnPrem to Cloud

- **Architecture**

- Core Services → Micro Services

- **Service:**

- Integration Service → Data Integration Server micro service

- **Server**

- On Prem Server → On Prem/Cloud Hosted Secure Agent
- Repository DB → Informatica cloud hosted metadata repository
Note: Lesser network hops if cloud hosted is suggested

- **Assets**

- PC Mappings → IDMC CDI Mappings
- PC Sessions → IDMC CDI Mapping Tasks
- PC Workflows → IDMC CDI Taskflows
- PC Job Execution Script → IDMC Scripts
- PC Parameter Files → IDMC Parameter Files

- Informatica provides Cloud Data Integration for PowerCenter service, for easier asset migration from PowerCenter to IDMC-CDI

Mapping Guidelines

- Understand Requirements
- Data Mapping and Quality
- Optimize transformations and Performance
- Error Handling and Security
- Testing and Deployment

Naming Conventions

- Descriptive names
- Avoid special characters
- Follow Informatica Naming Conventions
- Prefix or Suffix for clarity
- Consistency across assets and objects
- Avoid reserved words
- Use business naming standards

IDMC Naming Standards reference: <https://knowledge.informatica.com/s/article/DOC-15669>

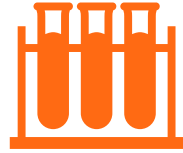
Performance Tuning

- Optimize Source and Target
- Transformation optimization
- Mapping Design optimization
- Mapping task and taskflow optimization
- Secure Agent HA and optimizations

Testing Mechanisms



Unit
Testing



System
Integration
Testing



User
Acceptance
Testing



Smoke
Testing



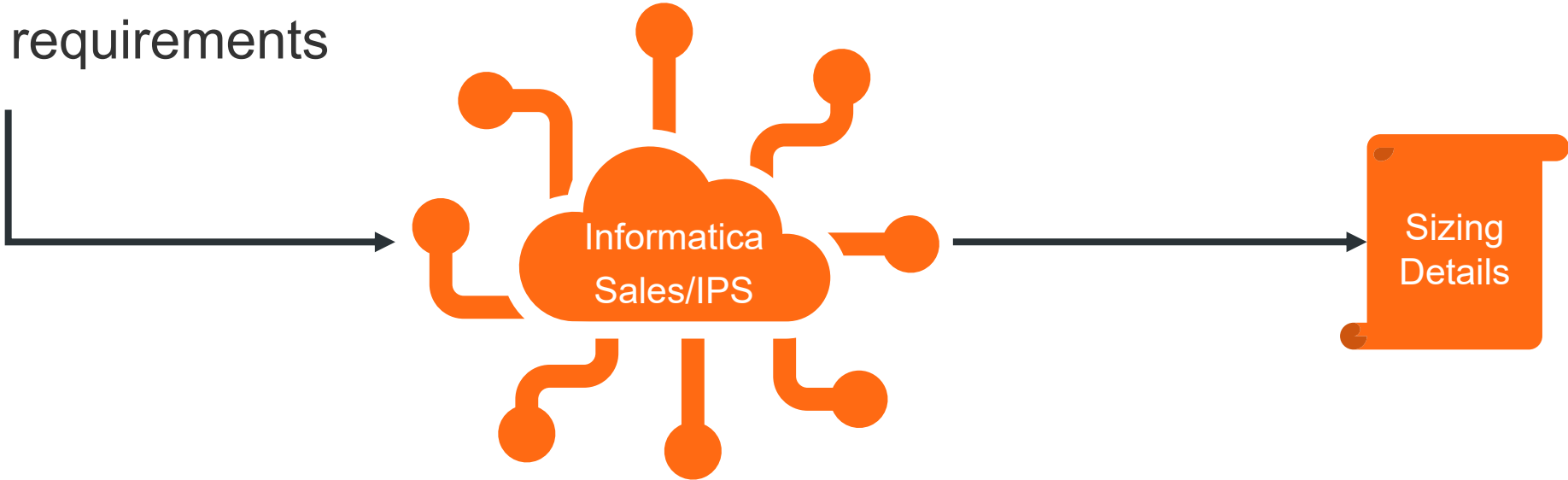
Regression
Testing



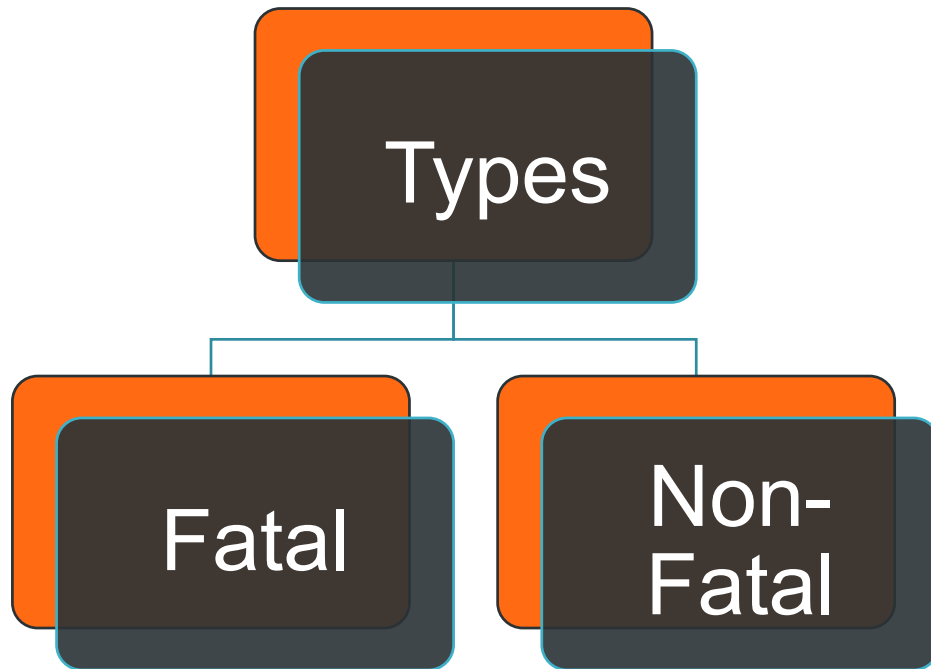
Performance
Testing

Secure Agent Sizing Guidelines

- Type of Workload
- Data Volume
- Concurrency
- Performance requirements



Error Handling



- **Mapping**

- ERROR()
- ABORT()

- **Taskflows**

- On Error
- On Warning
- Fail taskflow on completion
- Custom Error Handling

- **Mapping Tasks**













- Stop on Errors
- Override Tracing
- On Stored Procedure Error
- On Pre/Post-Session Command Task Error
- Error Log Type
- Error Log File Directory
- Error Log File Name
- Log Row Data
- Log Source Row Data
- Data Column Delimiter

CI/CD

- What is CICD:

Continuous Integration and Continuous Deployment is a set of practices and tools designed to improve the development, deployment, and delivery process of software applications

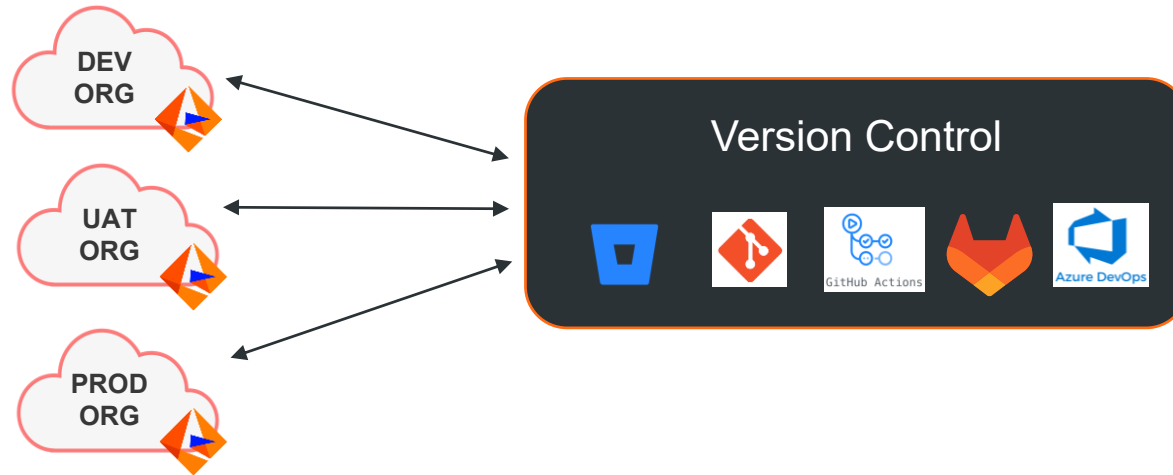
- Source Control Repositories Supported:

Repository	Self-hosted	Cloud-hosted
 Atlassian Bitbucket		
 GitHub <small>GitHub Actions</small>		
 GitLab		N/A
 Microsoft Azure DevOps <small>Azure DevOps</small>	N/A	
 Generic Repo		N/A

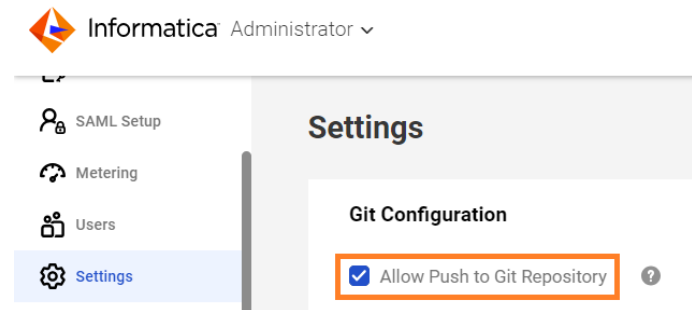
Note: Please refer to [source control](#) for latest support

CI/CD (Contd..)

- When you enable source control for the organization, select an empty repository.
- Use different organizations for development, testing, staging, and production

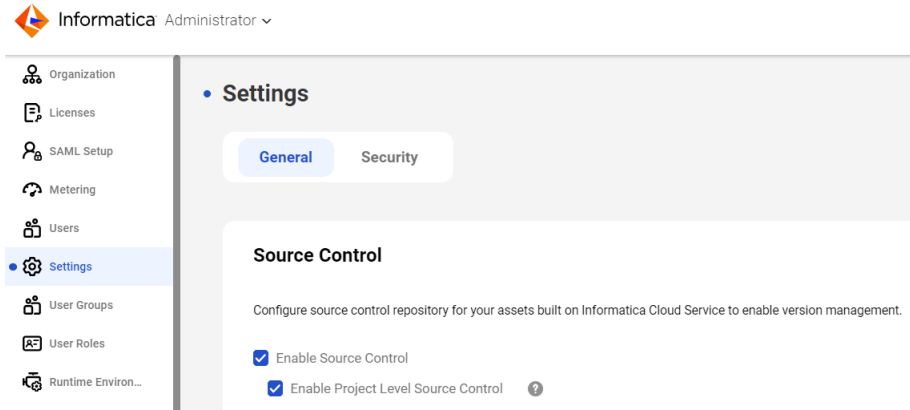


- Configure development organizations with read/write access to the source control repository, and configure non-development organizations with read-only access to the source control repository



CI/CD (Contd..)

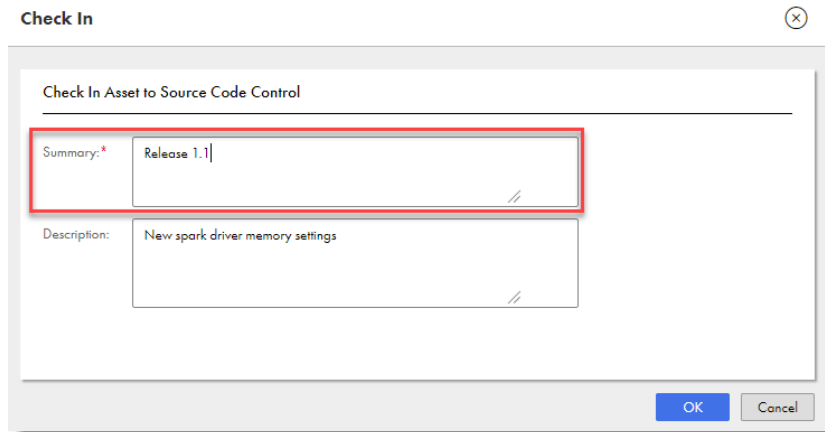
- Enabling project level repositories can help in parallel development and collaboration across teams in the organization



- Do not share source control credentials among multiple Informatica Intelligent Cloud Services users
- Create connections and runtime environments before you pull assets from the repository
- Ensure that reusable assets such as mappings and components are present in the repository before you use them

CI/CD (Contd..)

- When you rename or move an asset, check out the asset's first-level dependent assets and include them in the same check-in
- Enter comments when you check in assets, preferably release numbers with changes to the assets as part the release.



Check In

Check In Asset to Source Code Control

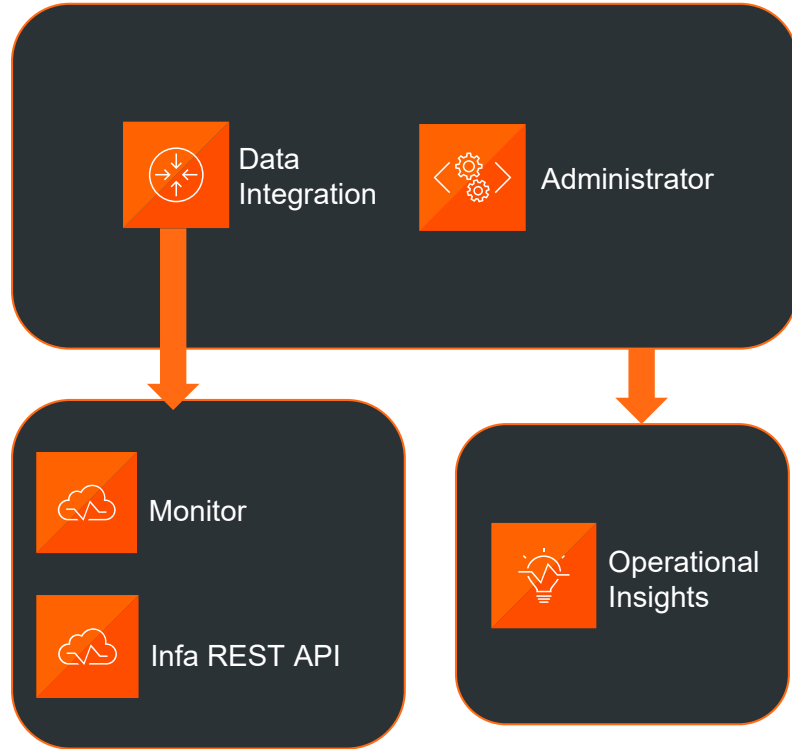
Summary:* Release 1.1

Description: New spark driver memory settings

OK Cancel

- When you check in multiple assets at one time, limit the number of assets to 1000 or fewer

Job Monitoring



- **Features:**

- Real-Time Monitoring
- Alerts and Notifications
- Detailed Job Logs
- Performance Metrics
- Historical Data and Trends

- **Benefits:**

- Enhanced Reliability
- Improved Troubleshooting
- Optimized Performance

Thank You

Where data & AI come to

