Nov 14, 2023

Mastering Data Analysis with IDMC Log Analyzer

- Ananya Choudhury, Principal Consultant, IPS

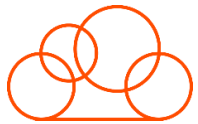
Informatica

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 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.

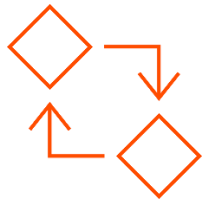
© Informatica. Proprietary and Confidential.

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

© Informatica. Proprietary and Confidential.

Informatica

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

© Informatica. Proprietary and Confidential.

Informatica

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

Log Analyzer Utility

Introduction

The Log Analyzer Utility provides user a simplified option to search for data coming from activity logs and taskflows logs based in tasks/taskflows ran in IDMC.

This utility automates the log extraction process and making it accessible in a database for consumption, where technical leaders, developers and functional administrators can quickly identify trends in the operation of the IDMC platform.

This utility also captures the metering usage data for CDI Tasks and stores them in centralized database.

The Log Analyzer improves overall efficiency in the identification of Bottle necks, operational trends, incidents, audit trails and immediate attention points within the customer's integration solution.

© Informatica. Proprietary and Confidential.

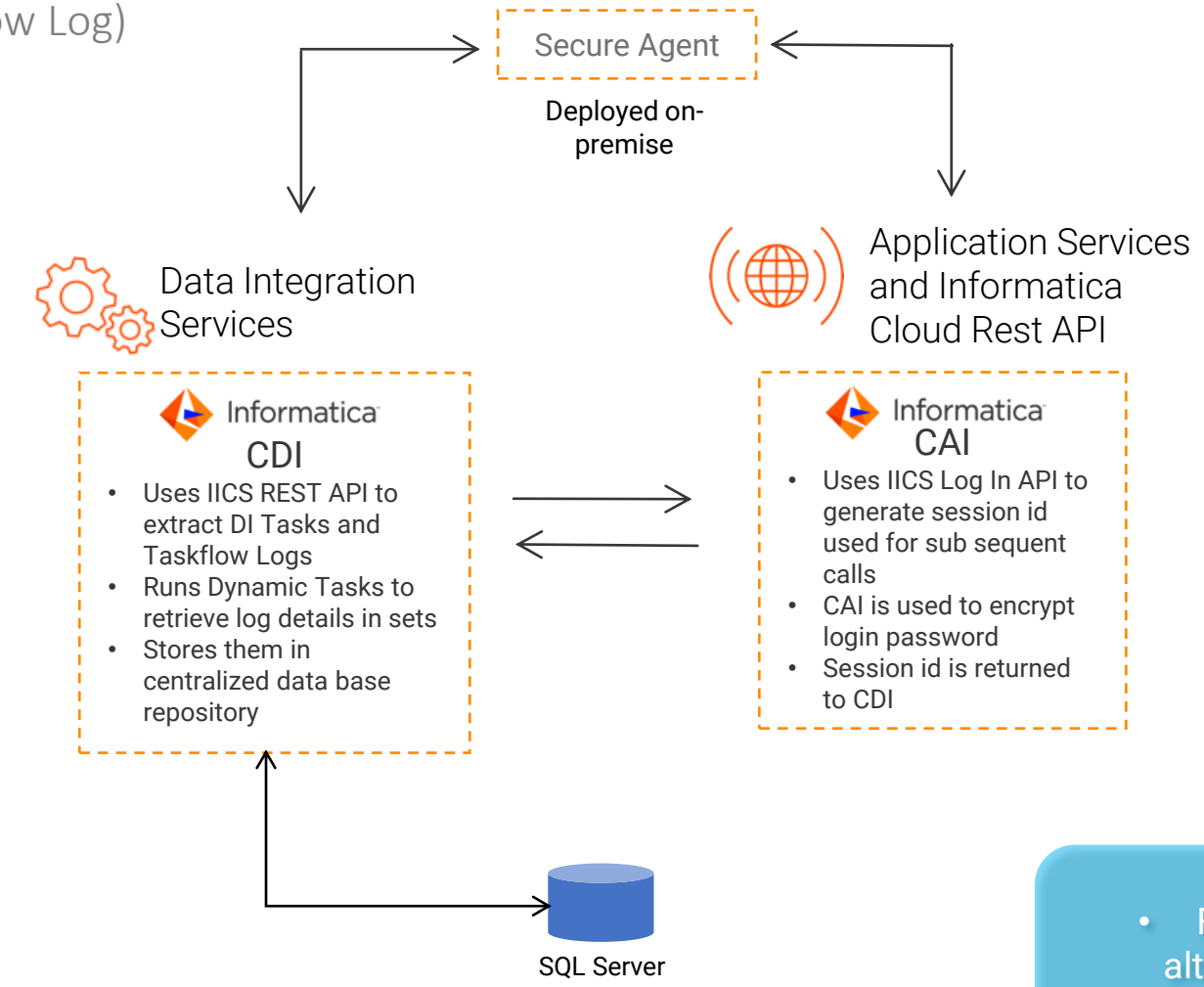
Key Features of Log Analyzer

- Integrations with Out-of-the-box IDMC APIs to expand the capabilities of IDMC in Performance Analysis areas
- Automated extraction of activity and taskflow logs, to get information for completed jobs from the Monitor service in any IDMC Org.
- Dumping of activity and taskflow logs in a centralized database to enable users to search for information related to transactions.
- Storing CDI Task metering data usage, enables the user to understand job-level details for the tasks and IPU meter for a specified date range.

© Informatica. Proprietary and Confidential.

Log Analyzer

(Activity Log, Taskflow Log)
High level Flow

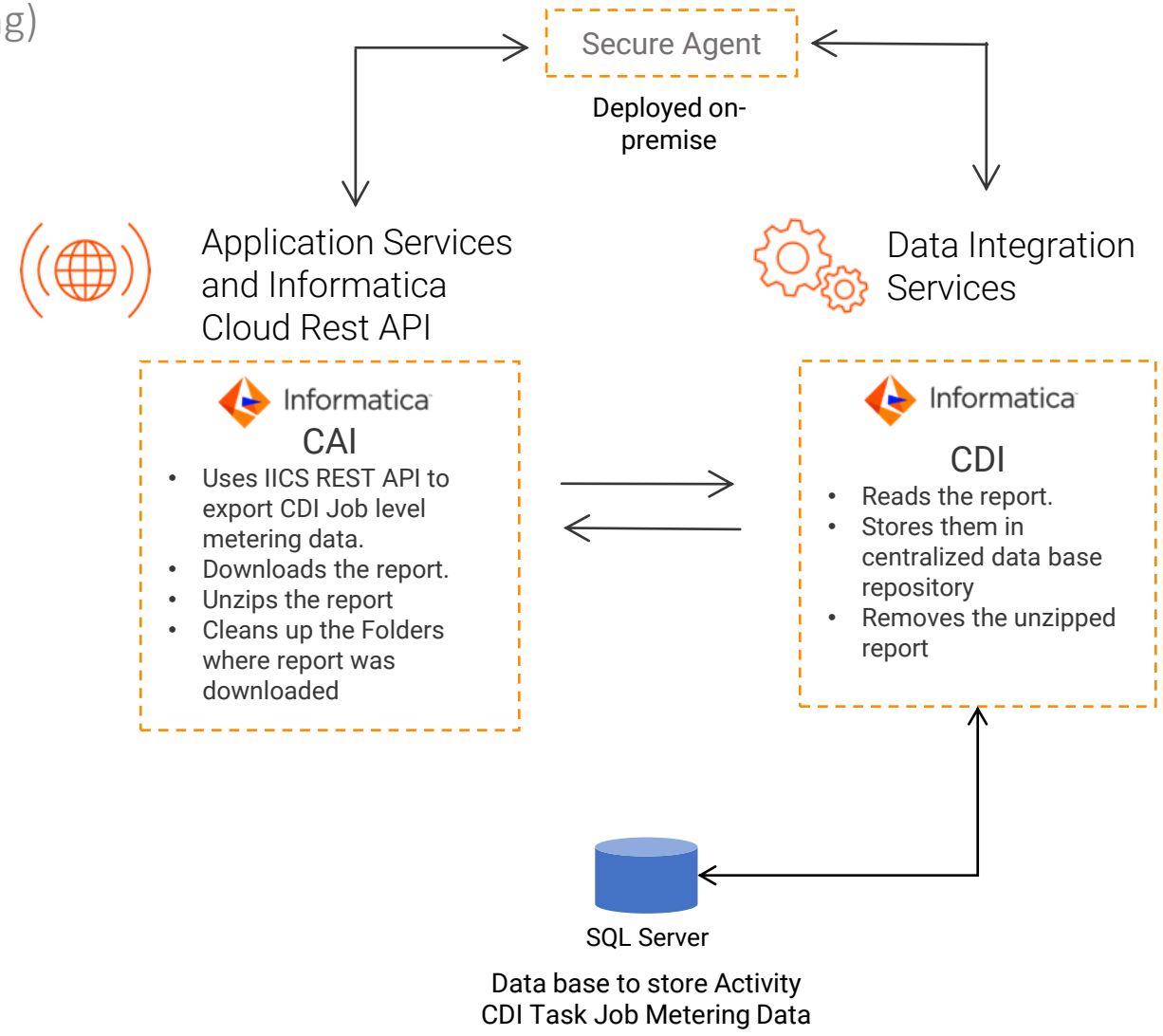


- For Password Encryption CAI is used. As an alternative Azure Key Vault can be used to store password.

© Informatica. Proprietary and Confidential.

Log Analyzer

(IPU Job Metering)
High level Flow



© Informatica. Proprietary and Confidential.

Components

- ***Activity Log* automated processing*:***

Automates the handling, analysis, and extraction of insights from activity log data.

Through CDI assets like mappings, mapping tasks and connectors, activity logs are collected, processed and written to a database without the need for manual intervention.

This saves time and effort while enabling user to effectively track and monitor integration tasks activities, detect anomalies, and identify patterns and trends.

- ***Taskflow Log* automated processing *:***

Automates the handling, analysis, and extraction of insights from Taskflow log data.

Through CDI assets like mappings, mapping tasks and connectors, taskflow logs are collected, processed and written to a database without the need for manual intervention.

This saves time and effort while enabling user to effectively track and monitor taskflow and subtasks execution details , detect anomalies, and identify patterns and trends.

- **CDI Tasks Metering automated processing *:***

Automates the handling, analysis, and extraction of insights from CDI IPU Metering usage data.

Through CAI, CDI assets like processes, service connector, app connections, mappings, mapping tasks and connectors, metering usage data are collected, processed and written to a database without the need for manual intervention.

This saves time and effort while enabling user to effectively track and monitor consumed metering usage per task details , helping identify patterns and trends.

© Informatica. Proprietary and Confidential.

Backend Infrastructure

- ***Login API:***

This resource provides a login functionality for Informatica Intelligent Cloud Services. When logging in, the response includes a session ID and base URL that are necessary for subsequent REST API calls.
Reference : https://docs.informatica.com/integration-cloud/b2b-gateway/current-version/rest-api-reference/platform-rest-api-version-3-resources/login.html

- ***ActivityLog API :***

This resource is used to request log information for completed jobs.
Reference: https://docs.informatica.com/integration-cloud/b2b-gateway/current-version/rest-api-reference/platform-rest-api-version-2-resources/activitylog.html

- ***TaskflowStatus API :***

This resource provides a convenient way to retrieve the status of a taskflow and associated subtasks.
Reference: https://docs.informatica.com/integration-cloud/data-integration/current-version/taskflows/taskflows/monitoring-taskflow-status-with-the-status-resource.html

- ***Metering Data API :***

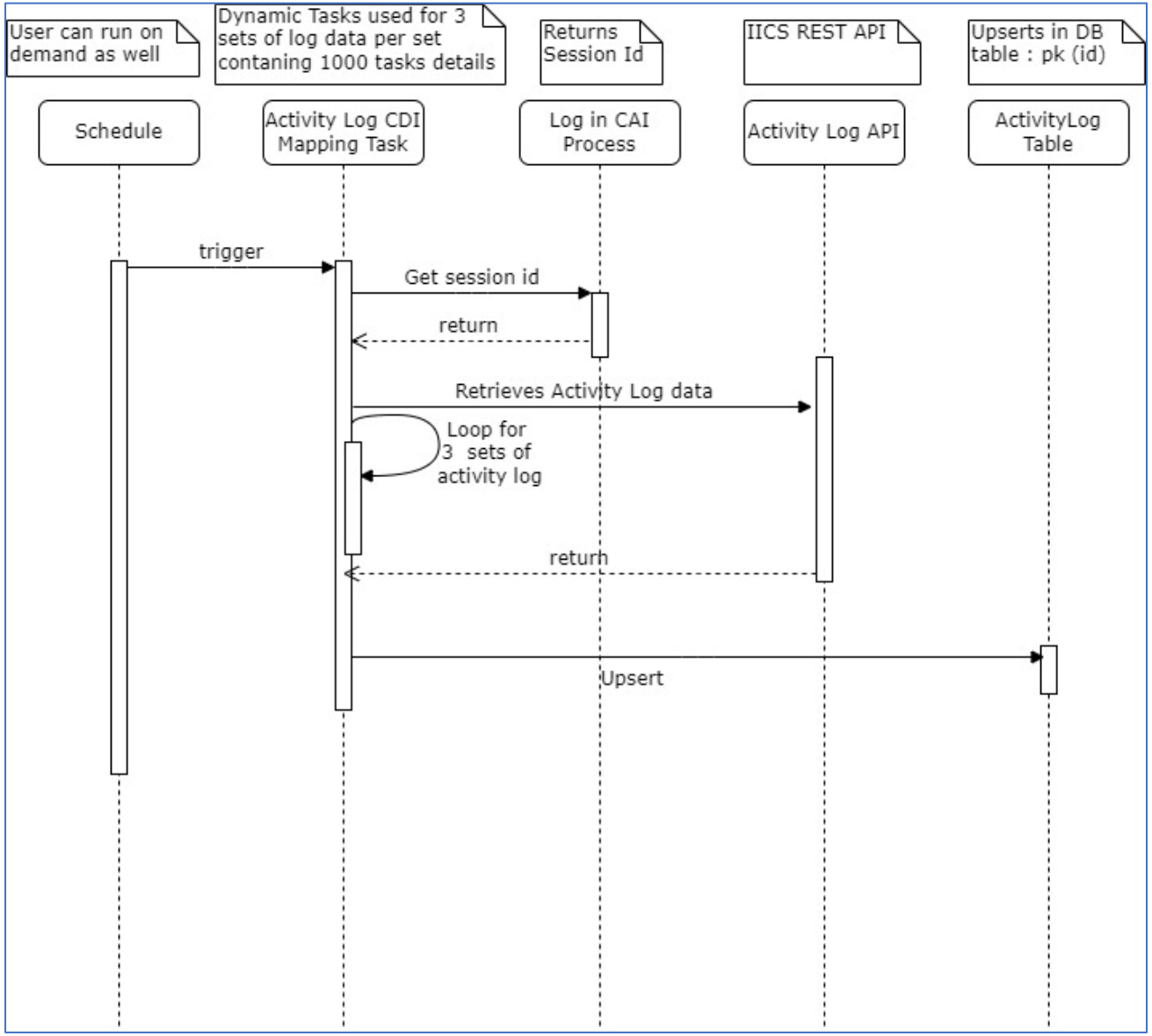
This resource provides a convenient way to retrieve the job metering usage data.
Reference: https://docs.informatica.com/integration-cloud/data-integration/current-version/rest-api-reference/platform-rest-api-version-3-resources/metering-data/sending-an-export-job-request-for-job-level-metering-data.html

- ***SqlServer Database:***

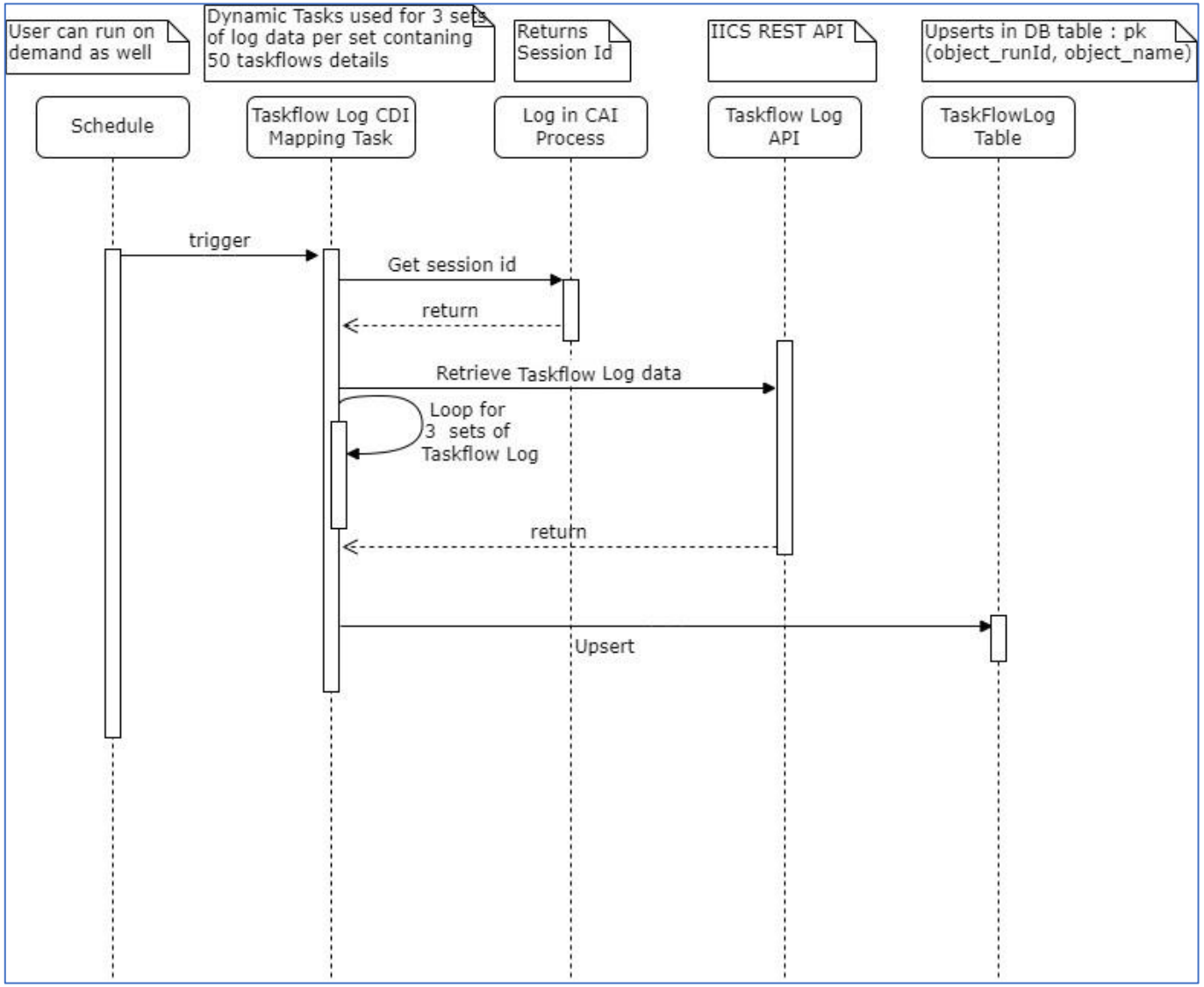
The SQL Server Database serves as the target system for storing the resulting data of the Activity and Taskflow logs, job metering usage data in the log analyzer.
This relational database contains only the necessary attributes required for efficient log analysis and processing.
It provides a reliable and scalable solution for storing and managing the log data generated by the system, ensuring optimal performance and facilitating effective analysis within the context of user operations.

© Informatica. Proprietary and Confidential.

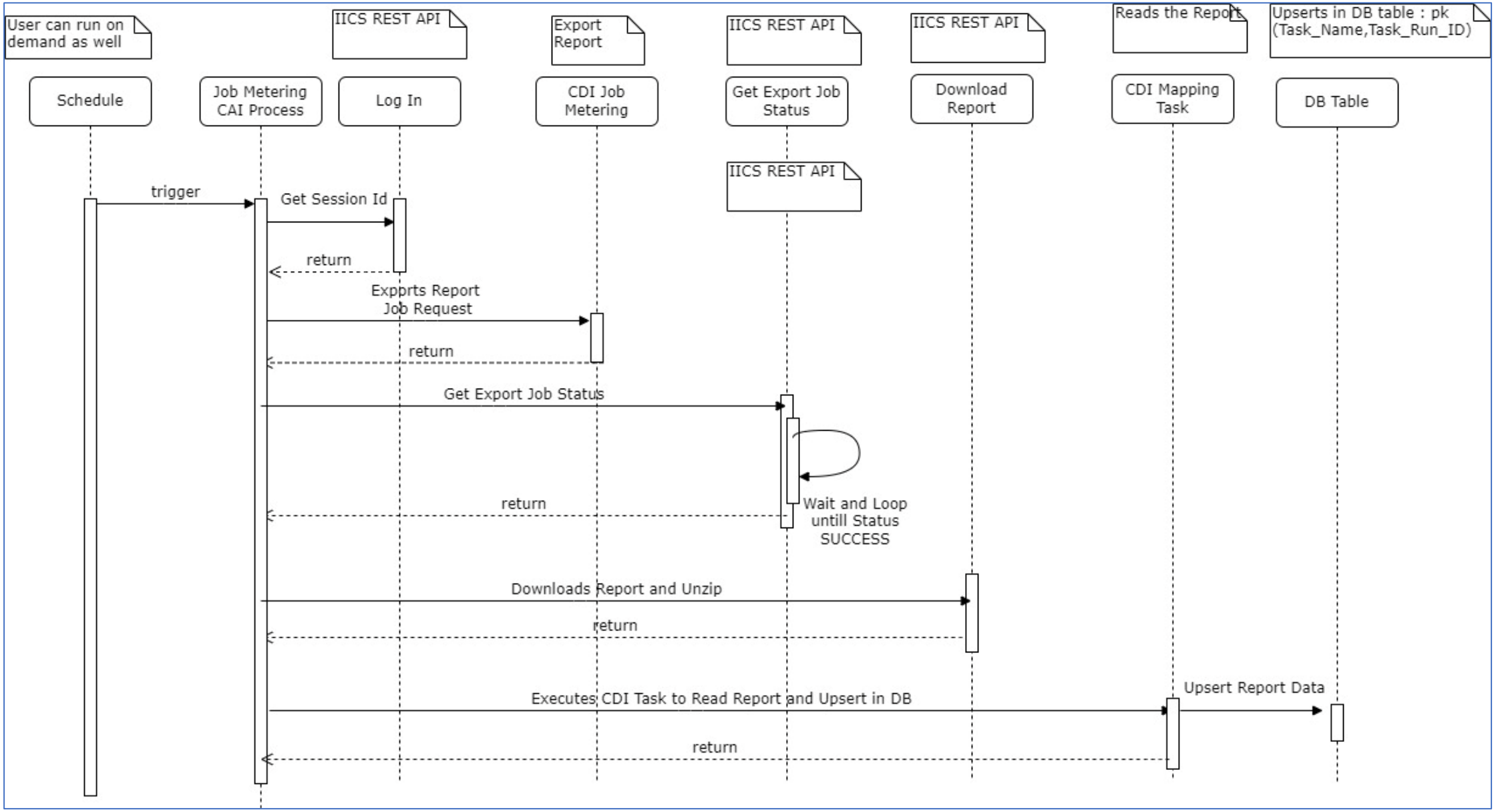
High level Sequence Flow - Activity Log



High level Sequence Flow - Taskflow Log



High level Sequence Flow – Job Metering Log



DB Table Details

- ***Table Specifications:***


ActivityLogTable_Spec.docx


TaskFlowLogTable_Spec.docx


CDI_JobMetering_Spec.docx

- ***Primary Keys and Relationship:***

The following table shows the attributes with which an association can be established between both tables to obtain a unique record from both.

Table Name	Primary Key Field Name
ActivityLog	Id
TaskFlowLog	object_runId, object_name
CDI_JobMetering	Task_Name, Task_Run_ID

ActivityLog Table	TaskFlowLog Table	CDI_JobMetering Table
objectname	Object_name	Task_Name
runId	object_runId	Task_Run_ID

© Informatica. Proprietary and Confidential.

Use cases

- ***Identification of candidates for performance improvement :***

This use case aims to identify mapping tasks that can be optimized for improved performance.

```
SELECT * FROM dbo.ActivityLog
WHERE  successsourcerows/duration_in_secs < 50000
AND duration_in_secs > 0
AND successsourcerows > 50000
AND duration_in_minutes > 5
AND state = 1
ORDER BY duration_in_minutes DESC
```

This query identifies mapping tasks with low performance reading data at the source, specifically focusing on tasks that have an execution rate of 50,000 rows per second or below on successfully completed tasks, and a total duration above 5 minutes.
Example of results:

objectname	Start Time utc	Duration in secs	Duration In minutes	Success Source rows	Success Target rows
mt_SUMM_HPP_HPSTG_DIR_FREQSEVR_CALC	2023-06-13 03:52:48.0000000	7509	125.15	78935332	9355678
mt_HPP_HPSTG_DIR_FREQSEVR_CALC_50_Load	2023-06-13 07:07:25.0000000	630	10.5	9355678	341189

© Informatica. Proprietary and Confidential.

Informatica

Use cases

- ***Compare execution patterns between two secure agents :***

By examining the behavior and performance of the secure agents, it becomes possible to identify any variations in their execution patterns.

```
SELECT agentId, COUNT(agentId) AS objects_executed FROM dbo.TaskFlowLog TF
INNER JOIN dbo.ActivityLog AL
ON TF.object_runId = AL.runId
AND TF.object_name = AL.objectname
AND TF.object_runtime_env = AL.runtimeEnvironmentId
WHERE agentId IS NOT NULL
AND AL.state = 1
AND TF.taskflow_start_time_est > '2023-06-15'
AND TF.taskflow_end_time_est < '2023-06-16'
GROUP BY agentId
```

The results for this example show skewed performance over the input date period, where one agent executed only approx. 32% of the jobs compared to the other agent.

agentId	Tasks executed
01BX0F08000000000004	27
01BX0F08000000000005	58

© Informatica. Proprietary and Confidential.

Use cases

- ***Continued - Compare execution patterns between two secure agents :***

The following database query is used to identify at a more detailed level which tasks and the number of times they were executed on each secure agent:

```
SELECT  objectid, COUNT(objectid) AS object_executions, objectname, agentId FROM dbo.TaskFlowLog TF
INNER JOIN dbo.ActivityLog AL
ON TF.object_runId = AL.runId
AND TF.object_name = AL.objectname
AND TF.object_runtime_env = AL.runtimeEnvironmentId
WHERE agentId IS NOT NULL
AND state = 1
AND TF.taskflow_start_time_est > '2023-06-15'
AND TF.taskflow_end_time_est < '2023-06-16'
GROUP BY objectId, agentId, objectname
ORDER BY objectId, object_executions
```

Example Result:

objectid	objectname	No. of executions	agentId
01BX0F0Z0000000001FF	mt_UPDATE_PROCESS_CONTROL	8	01BX0F08000000000004
01BX0F0Z0000000001FF	mt_UPDATE_PROCESS_CONTROL	12	01BX0F08000000000005
01BX0F0Z0000000001J4	mt_API_Log_Analytics_Message	12	01BX0F08000000000004
01BX0F0Z0000000001J4	mt_API_Log_Analytics_Message	34	01BX0F08000000000005
01BX0F0Z0000000001OY	mt_MFX_SRC_UW_SENDPOSTER	3	01BX0F08000000000004
01BX0F0Z0000000001OY	mt_MFX_SRC_UW_SENDPOSTER	6	01BX0F08000000000005

© Informatica. Proprietary and Confidential.

Use cases

- ***Collect metrics about failures or bottle necks :***

Collecting metrics about failures or bottlenecks in Mapping Tasks and Taskflows is crucial for identifying and resolving performance issues in data integration processes.
By monitoring and measuring various metrics such as execution time, resource utilization, data throughput, and error rates, customer's Technical leaders can gain insights into the health and efficiency of their mapping tasks and Taskflows.

The following database query returns a count of taskflows and their associated objects in a specified date range and sorts them by status:

```
SELECT
        taskflow_status,
        COUNT(taskflow_status) AS taskflow_count,
        object_status,
        COUNT(object_status) as object_count
FROM dbo.TaskFlowLog
WHERE taskflow_start_time_est BETWEEN '2023-06-16 00:00:00' AND '2023-06-20 23:00:00'
GROUP BY taskflow_status, object_status
ORDER BY taskflow_status, object_status
```

Example Result:

taskflow_status	taskflow_count	object_status	object_count
FAILED	60	COMPLETED	60
FAILED	14	FAILED	14
FAILED	41	SUCCESS	41
RUNNING	2	COMPLETED	2
RUNNING	8	SUBMITTED	8
RUNNING	4	SUCCESS	4
SUCCESS	825	COMPLETED	825
SUCCESS	351	SUCCESS	351
SUSPENDED	1	FAILED	1
SUSPENDED	2	SUCCESS	2

© Informatica. Proprietary and Confidential.

Use cases

➢ ***Continued : Collect metrics about failures or bottle necks :***

To identify Taskflows successfully completed entirely, that is, with all child objects completed and without errors, we have the following query:

```
SELECT
   TF.taskflow_runId,
   TF.taskflow_name,
   TF.taskflow_location,
   TF.taskflow_status,
   TF.taskflow_start_time_est,
   TF.taskflow_end_time_est,
   ABS(TF.object_runId) AS object_runId,
   TF.object_name,
   TF.object_type,
   AL.objectId,
   TF.object_location,
   TF.object_status,
   TF.object_start_time_est,
   TF.object_end_time_est,
   ABS(TF.object_duration_in_seconds) AS total_duration_in_seconds,
   ABS(TF.object_duration_in_minutes) AS total_duration_in_minutes,
   AL.agentId,
   ISNULL(AL.successsourcerows, 0) AS successsourcerows,
   ISNULL(AL.successtargetrows, 0) AS successtargetrows
FROM dbo.TaskFlowLog TF
LEFT JOIN dbo.ActivityLog AL
ON TF.object_runId = AL.runid
AND TF.object_name = AL.objectname
WHERE taskflow_status = 'SUCCESS'
AND (object_status = 'SUCCESS'
OR object_status='COMPLETED')
AND taskflow_start_time_est BETWEEN '2023-06-17 00:00:00' AND '2023-06-20 23:00:00'
ORDER BY TF.object_start_time_est, TF.taskflow_start_time_est, AL.duration_in_secs DESC
```

taskflow_name	taskflow_status	object_name	object_type	object_status	total_duration_in_seconds	total_duration_in_minutes	Success Source rows	Success Target rows
TF_POLICY_BATCH_INITIALIZATION	SUCCESS	mt_ST_POLICY_PREVIOUS_BATCHID_CLEANUP	MTT	COMPLETED	26	0.43	12	12
TF_POLICY_BATCH_INITIALIZATION	SUCCESS	mt_BATCH_HDR_DTL_ARCH	MTT	COMPLETED	17	0.28	73	73
TF_POLICY_BATCH_INITIALIZATION	SUCCESS	mt_SS_POLICY_PREVIOUS_BATCHID_CLEANUP	MTT	COMPLETED	16	0.27	6	6
TF_POLICY_BATCH_INITIALIZATION	SUCCESS	cmd_newline_removal	command	SUCCESS	0	0	0	0

An example of the results of this query can be seen in the table above (Some columns removed for simplification):

© Informatica. Proprietary and Confidential.

Demo

Q&A

Thank You