Informatica PowerCenter

Lesson 6- Workflow and Session Log

Lesson Objectives

- In this lesson you will learn about:
  - Workflow Variables
  - Workflow Schedule
  - Scheduler Options
  - Editing a Workflow
  - Deleting a Workflow
  - Session Logs
  - Setting Tracing Levels
  - Update Strategy Transformation

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

6.1. Workflow Variable

Description

- Workflow variables are used to reference and record information.
  - For Example, we can use variable in a decision task to determine whether the previous task executed/ran properly.
  - If it did, you can run the next task and if not, you can stop the workflow.
- Types of Workflow variables
  - Predefined workflow variables
    - Task-specific
    - Built-in
  - User-defined workflow variables

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

6.1. Workflow Variable
Predefined workflow variables

- The Workflow Manager provides predefined workflow variables for tasks within a workflow.
- Each workflow contains a set of predefined variables that you use to evaluate workflow and task conditions.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Types of predefined variables

- Task-specific variables: The Workflow Manager provides a set of task-specific variables for each task in the workflow.
  - Use task-specific variables in a link condition to control the path the Integration Service takes when running the workflow.
  - The Workflow Manager lists task-specific variables under the task name in the Expression Editor. Eg: Condition, End Time, Errorcode etc

Use workflow variables when you configure the following types of tasks:

Decision tasks. Decision tasks determine how the Integration Service runs a workflow. For example, use the Status variable to run a second session only if the first session completes successfully.

Links. Links connect each workflow task. Use workflow variables in links to create branches in the workflow. For example, after a Decision task, you can create one link to follow when the decision condition evaluates to true, and another link to follow when the decision condition evaluates to false.

Timer tasks. Timer tasks specify when the Integration Service begins to run the next task in the workflow. Use a user-defined date/time variable to specify the time the Integration Service starts to run the next task.

Use the following keywords to write expressions for user-defined and predefined workflow variables:
AND
OR
NOT
TRUE
FALSE
NULL
SYSDATE

Types of predefined variables - Task Specific

Specific Variables	Description	Task Types	Datatype
Condition	Evaluation result of decision condition expression. If the task fails, the Workflow Manager keeps the condition set to null. Sample syntax: $Dec_TaskStatus.Condition = <TRUE | FALSE | NULL | any integer>	Decision	Integer
EndTime	Date and time the associated task ended. Precision is to the second. Sample syntax: $s_item_summary.EndTime > TO_DATE('11/10/2004 08:13:25')	All tasks	Date/Time
ErrorCode	Last error code for the associated task. If there is no error, the Integration Service sets ErrorCode to 0 when the task completes. Sample syntax: $s_item_summary.ErrorCode = 24013	All tasks	Integer
PrevTaskStatus	Status of the previous task in the workflow that the Integration Service ran. Statuses include: ABORTED FAILED SUCCEEDED STOPPED Use these key words when writing expressions to evaluate the status of the previous task. Sample syntax: $Dec_TaskStatus.PrevTaskStatus = FAILED	All tasks	Integer

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Note: For a complete of Task Specific Predefined workflow variables, please refer to the documentation.
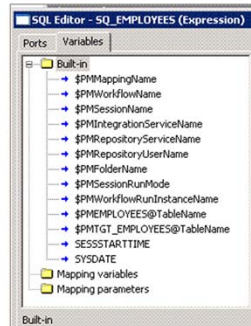
Figure 6.1

$PMSessionName : This variable can be used in mappings or mapplets.

$PMWorkflowName: This variable can be used in a mapping, a mapplet, or in workflow tasks such as Decision tasks, and links. You can also use $PMWorkflowName in input fields that accept mapping or workflow variables.

Note: For complete details of Built-in variables, please refer to the documentation.

User-defined Workflow variables

- You create user-defined workflow variables when you create/edit a workflow
- These variables are used in tasks within that workflow
- You can edit and delete user-defined workflow variables
- Use user-defined variables when you need to make a workflow decision based on criteria you specify.
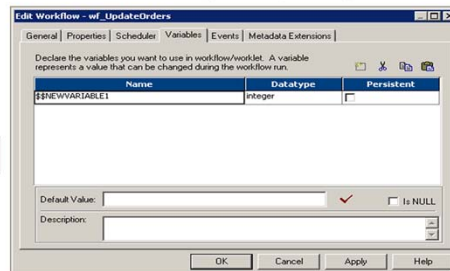
Figure 6.2

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved   8

While creating a user defined work flow variable verify that the name of a user defined workflow variable does not have a $ since single $ is reserved for system variables. Use $$ instead.

6.1. Workflow Variable
Example

- For example, you create a workflow to load data to an orders database nightly. You also need to load a subset of this data to headquarters periodically, every tenth time you update the local orders database. Create separate sessions to update the local database and the one at headquarters. The workflow looks like as the picture below:
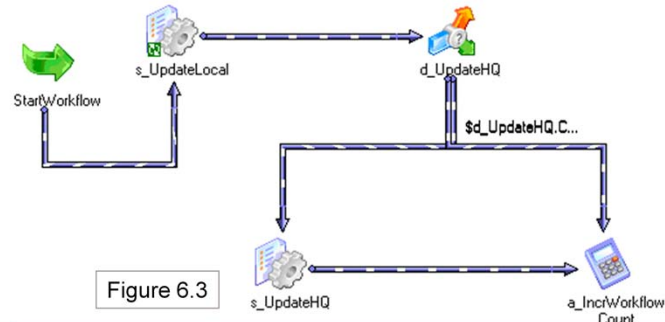
Figure 6.3

Use a user-defined variable to determine when to run the session that updates the orders database at headquarters.

To configure user-defined workflow variables, set up the workflow as follows:

1. Create a persistent workflow variable, $$WorkflowCount, to represent the number of times the workflow has run.

2. Add a Start task and both sessions to the workflow.

3. Place a Decision task after the session that updates the local orders database. Set up the decision condition to check to see if the number of workflow runs is evenly divisible by 10. Use the modulus (MOD) function to do this.

4. Create an Assignment task to increment the $$WorkflowCount variable by one.

5. Link the Decision task to the session that updates the database at headquarters when the decision condition evaluates to true. Link it to the Assignment task when the decision condition evaluates to false. When you configure workflow variables using conditions, the session that updates the local database runs every time the workflow runs. The session that updates the database at headquarters runs every 10th time the workflow runs.

Demo

- Create a user defined Workflow variable.

6.2. Workflow Schedule
Description

- A Workflow schedule defines when the Integration service should start the data transformation and load process
- By default a Workflow runs on demand
- A Workflow schedule is defined using a Workflow Scheduler
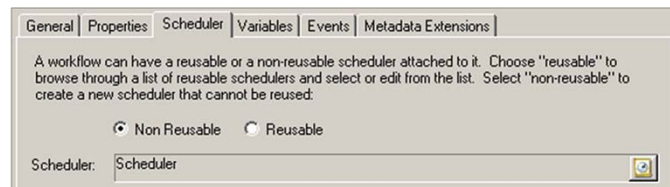- A Workflow schedule can be reusable or non-reusable

General | Properties | Scheduler | Variables | Events | Metadata Extensions |

A workflow can have a reusable or a non-reusable scheduler attached to it. Choose "reusable" to browse through a list of reusable schedulers and select or edit from the list. Select "non-reusable" to create a new scheduler that cannot be reused:

⦿ Non Reusable   ⦾ Reusable

Scheduler: Scheduler

Figure 6.4

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved   11

Every Workflow created in the Workflow Manager, runs on a schedule. The Integration service runs a scheduled Workflow unless the prior run of the Workflow fails. When a Workflow fails, the Integration service removes the Workflow from the schedule. A Workflow must be rescheduled manually in the Workflow if a scheduled Workflow fails. By default, a Workflow runs on demand.

Figure 6.5

The schedule settings defined for a Workflow can be edited in the scheduler if a reusable scheduler is created. The Integration service reschedules the Workflow according to the new settings.

Non-reusable scheduler settings can only be used in the Workflow that they are defined in.

In case of a reusable scheduler, the same set of schedule settings can be used for any Workflow in a folder.

The Workflow Manager marks a Workflow as "invalid" if the scheduler associated with a Workflow is deleted.

6.3. Scheduler Options
Brief

- Run Options
  - Run On Server Initialization
  - Run On Demand
  - Run Continuously
- Schedule Options
  - Run Once
  - Run Every
  - Customized Repeat
- Start Options
  - Start Date
  - Start Time
- End Options
  - End On
  - End After
  - Forever

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved   13

A Workflow can be scheduled to run continuously, repeat at a given time or interval, or start manually.

The various options for the schedule settings are:

Run Options

> Run On Server Initialization - The Integration service runs the Workflow as soon as the service is initialized. The Integration service then starts the next run of the Workflow according to the settings in Schedule Options

> Run On Demand - The Integration service runs the Workflow when the Workflow is started manually

> Run Continuously - The Integration service runs the Workflow as soon as the service initializes. The Integration service then starts the next run of the Workflow as soon as it finishes the previous run

Schedule Options

> Run Once - The Integration service runs the Workflow once, as scheduled in the scheduler

> Run Every - The Integration service runs the Workflow at regular intervals, as configured

> Customized Repeat - The Integration service runs the Workflow on the dates and times specified in the Repeat dialog box

Demo

- Create a Scheduler and Schedule a Workflow
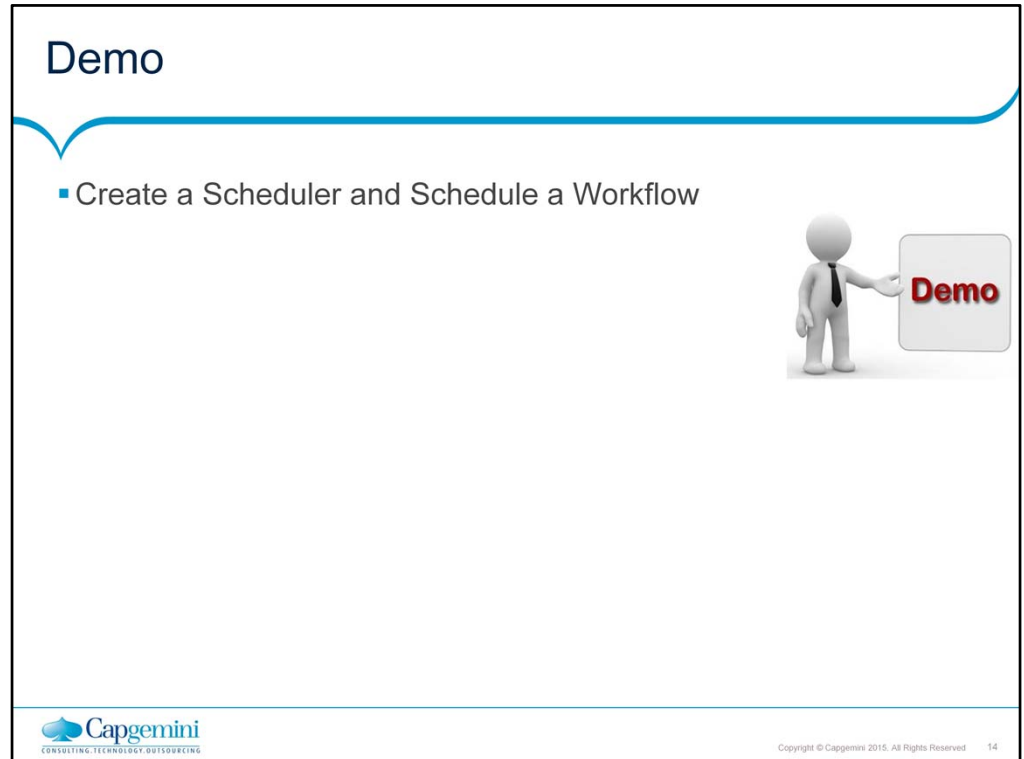
Steps to create a reusable scheduler:

Select Schedulers from Workflows menu.

Click New to add a new scheduler.

In the General tab, enter a name for the scheduler.

Configure the scheduler settings in the Scheduler tab.

Click on OK.

Steps to schedule a Workflow:

In the Workflow Designer, open the Workflow.

Choose Workflows | Edit.

In the Scheduler tab, choose Non-reusable if you want to create a non-reusable set of schedule settings for the Workflow. Choose Reusable if an existing reusable scheduler for the Workflow has to be selected.

Note: If there is no reusable scheduler in the folder, it has to be created before one chooses Reusable. The Workflow Manager displays a warning message if there is no existing reusable scheduler.

Click the right side of the Scheduler field to select the reusable scheduler.

Click on OK.

6.4. Editing a Workflow
Brief

- A Workflow can be edited to modify various properties
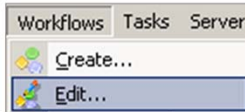- The repository updates the Workflow information when the Workflow is saved
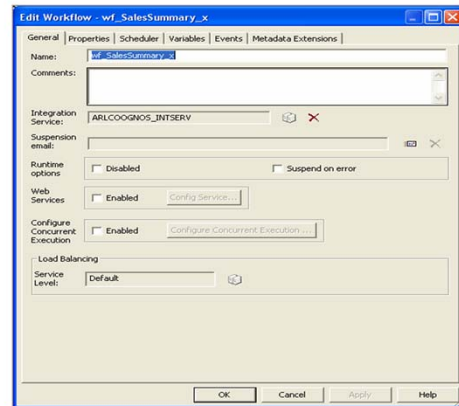
Figure 6.6

Figure 6.7

To edit a Worklflow the menu option Workflows | Edit is used. When a Workflow is edited, the repository updates the Workflow information when the Workflow is saved. If a running Workflow is edited, the Integration service uses the updated information the next time the Workflow runs.

6.5. Deleting a Workflow
Explanation

- A Workflow that is no longer in use can be deleted
- If a Workflow that is running is deleted, then the Integration service aborts the Workflow
- A Workflow can be deleted from
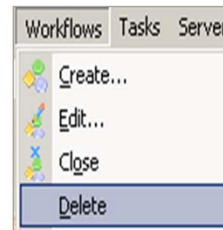 - The Navigator window
  - Workflow Designer workspace

Workflows	Tasks	Server
Create...		
Edit...		
Close		
Delete		

Figure 6.8

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved   16

When a Workflow is deleted, all non-reusable Tasks and reusable Task instances associated with the Workflow are deleted. Reusable Tasks used in the Workflow remain in the folder when the Workflow is deleted.

If a Workflow that is running is deleted, then the Integration service aborts the Workflow. If a Workflow that is scheduled to run is deleted, the Integration service removes the Workflow from the schedule.

A Workflow can be deleted in the Navigator window, or the Workflow currently displayed in the Workflow Designer workspace can be deleted.

> To delete a Workflow from the Navigator window, open the folder, select the Workflow and press the Delete key
>
> To delete a Workflow currently displayed in the Workflow Designer workspace, choose Workflows-Delete

6.6. pmcmd command
Explanation

- pmcmd is a program you use to communicate with the Integration Service.
- It is used to perform some of the tasks that can also be performed in the Workflow Manager, such as starting and stopping workflows and sessions.
- It can be used in the following modes
  - Command line mode
  - Interactive mode

Command line mode: You invoke and exit pmcmd each time you issue a command. You can write scripts to schedule workflows with the command line syntax. Each command you write in command line mode must include connection information to the Integration Service.

Interactive Mode: You establish and maintain an active connection to the Integration Service. This lets you issue a series of commands.

6.6. pmcmd command
Running in Command Line Mode

- Command line mode invokes and exits pmcmd each time you issue a command.
- Command line mode is useful if you want to run pmcmd commands through batch files, scripts, or other programs.

Figure 6.9

By default, the PowerCenter installer installs pmcmd in the \server\bin directory.

When you run pmcmd in command line mode, you enter connection information such as domain name, Integration

To run pmcmd commands in command line mode:

> At the command prompt, switch to the directory where the pmcmd executable is located.

> Enter pmcmd followed by the command name and its required options and arguments:

>> pmcmd command_name [-option1] argument_1 [-option2] argument_2...

Service name, user name and password in each command. For example, to start the workflow "wf_SalesAvg" in

folder "SalesEast," use the following syntax:

D:\Informatica\powercenter8.5.1\server\bin>pmcmd startworkflow -sv INTIM0045 -d Informatica8 -u INF1 -p INF1 -f INF1FLD wf_SalesAvg

6.6. pmcmd command
Running in Interactive Mode

- pmcmd is used in interactive mode to start and stop workflows and sessions without writing a script.
- In this mode, the connection information such as domain name, Integration Service name, username, and password is entered.
- The subsequent commands are run without entering the connection information for each command.


```
C:\WINDOWS\system32\cmd.exe - pmcmd

D:\Informatica\PowerCenter8.5.1\server\bin>pmcmd

Informatica(r) PMCMD, version [8.5.1], build [56.1207], Windows 32-bit
Copyright (c) Informatica Corporation 1994 - 2007
All Rights Reserved.

Invoked at Fri Jul 01 11:06:43 2011

pmcmd> connect -sv INTBU0041 -d Informatica8 -u vandana -p vandana
Connected to Integration Service: [INTBU0041].
pmcmd> setfolder vandana
pmcmd> startworkflow wf_m_employees
pmcmd> _
```

Figure 6.10

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

To run pmcmd commands in interactive mode:

1. At the command prompt, switch to the directory where the pmcmd executable is located.

By default, the PowerCenter installer installs pmcmd in the \server\bin directory.

2. At the command prompt, type pmcmd.

This starts pmcmd in interactive mode and displays the pmcmd> prompt. You do not have to type pmcmd before each command in interactive mode.

3. Enter connection information for the domain and Integration Service. For example:

   connect -sv MyIntService -d MyDomain -u seller3 -p jackson

4. Type a command and its options and arguments in the following format:

   command_name [-option1] argument_1 [-option2] argument_2...

   pmcmd runs the command and displays the prompt again.

5. Type exit to end an interactive session.

For example, the following commands invoke the interactive mode, establish a connection to Integration Service

"MyIntService," and start workflows "wf_SalesAvg" and "wf_SalesTotal" in folder "SalesEast":

pmcmd

pmcmd> connect -sv INTIM0045 -d Informatica8 -u INF1 -p INF1

pmcmd> setfolder INF1

pmcmd> startworkflow wf_SalesAvg

pmcmd> startworkflow wf_SalesTotal

6.6. pmcmd command
Return Codes

- In command line mode, pmcmd indicates the success or failure of a command with a return code.
- Return code "0" indicates that the command succeeded.
- Any other return code indicates that the command failed.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

6.6. pmcmd command
Session Log Messages

- The Integration service precedes each message in the log file with a thread identification and then a code and number
- The code defines a group of messages for a specific process
- The number defines a specific message
- The message can provide general information or it can be an error message
- The level of detail in a session log depends on the tracing level set

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING   Copyright © Capgemini 2015. All Rights Reserved   21

The Session Log contains informational, warning, and error messages from the session processes.

The session log includes a load summary that reports the number of rows inserted, updated, deleted, and rejected for each target as of the last commit point. The Integration service reports the load summary for each session by default.

6.6. pmcmd command
Configuring Session Logs

- Properties that can be configured for a Session Log are:
  - Location
    - $PMSessionLogDir server variable is mapped to a local directory on the Integration service
    - The default name for the session log is mapping name.log
    - It can be changed
  - Archive
  - Used to specify the number of runs of the Session Log to save
  - Default is one
  - Tracing levels
    - Define the amount of detail in the Session Log

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Session Logs options are configured in the session properties. The following information can be configured for a session log:

Location - The directory where the session log should be created. By default, the Integration service creates the session log in the directory configured for the $PMSessionLogDir server variable. A different directory can be entered, but if the directory does not exist or is not local to the Integration service that runs the session, the session fails

Name - The session log can be named or the default name can be accepted. The default name for the session log is s_mapping name.log

Archive - By default, the Integration service does not archive session logs. It creates one session log for each session and overwrites the existing log with the latest session log. We can specify the number of log files to be created by alterating the properties- Save session log for these runs, in Task properties-Config objects-log options

Tracing levels - Setting a tracing level for the session can control the type of information the Integration service includes in the session log. By default, the Integration service uses tracing levels configured in the mapping

The name and location of the session log can be configured on the Properties tab of the session properties in the Workflow Manager.

6.7. Setting Tracing Levels
Description

Tracing Level	Description
None	**Integration service uses the tracing level set in the mapping.**
Terse	**Integration service logs initialization information as well as error messages and notification of rejected data.**
Normal	• Integration service logs initialization and status information • Errors encountered • Skipped rows due to transformation row errors • Summarizes session results, but not at the level of individual rows

The amount of detail in the session log depends on the tracing level set. Tracing levels can be defined for each transformation or for the entire session. By default, the Integration service uses tracing levels configured in the mapping.

Setting a tracing level for the session overrides the tracing levels configured for each transformation in the mapping. If a normal tracing level or higher is selected, the Integration service writes row errors into the session log, including the transformation in which the error occurred and complete row data.

6.7. Setting Tracing Levels
Description (contd..)

Tracing Level	Description
Verbose Initialization	▪ Additional initialization details ▪ Name of index and data files used ▪ Detailed transformation statistics
Verbose Data	**In addition to verbose initialization tracing,** **Integration service logs each row that passes into the** **mapping**

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Generally the tracing level is kept at Normal. For debugging purpose, the tracing level is set to Verbose Data to get a row by row data loading status.

6.7. Setting Tracing Levels
Viewing Session Logs

- Session Logs
  - Can be viewed through the Workflow monitor
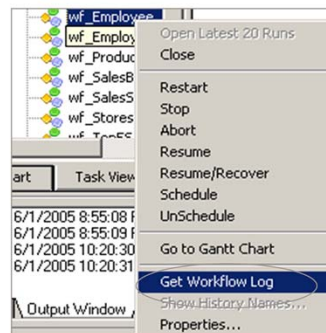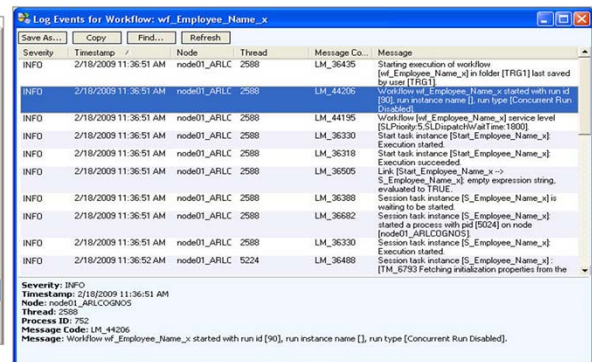 - Are temporary text files that can be opened with any text editor

Figure 6.11

Figure 6.12

Session logs are text files that can be opened with any text editor. The Integration service saves session logs in the directory specified in the Session Log File Directory field in the session properties.

Session logs can be viewed through the Workflow Monitor. The Workflow Monitor creates a temporary file that stores the session log. A session log file can be viewed even if the session fails.

The Integration service generates the session log based on the Integration service code page. The language of the session log can be specified based on the locale of the machine hosting the Integration service .

6.8. Update Strategy Transformation
Description

- Used to specify how each individual row will be used to update target tables (insert, update, delete, reject)
- Active transformation
- Commonly used in combination with Lookup transformation
- Update strategy can be used at two levels
  - Within a session
  - Within a mapping

While designing the data warehouse, often there might be a requirement to decide what type of information to store in targets. As part of the target table design, we need to determine whether to maintain all the historic data or just the most recent changes.

For example, there might be a target table, T_CUSTOMERS, that contains customer data. When a customer address changes, there might be a need to save the original address in the table, instead of updating that portion of the customer row. In this case, a new row needs to be created containing the updated address. The original row may be required to be preserved with the old customer address. This illustrates how historical information might be stored in a target table. However, if the T_CUSTOMERS table has to be a snapshot of current customer data, the existing customer row can be updated. The original address might be lost in this case. The model chosen constitutes the update strategy, how to handle changes to existing rows.

Update strategy can be used at two different levels:

Within a session - When a session is configured, the Integration service can be instructed to either treat all rows in the same way (for example, treat all rows as inserts), or use instructions coded into the session mapping to flag rows for different database operations.

Within a mapping - Within a mapping, the Update Strategy Transformation is used to flag rows for insert, delete, update, or reject.

6.8. Update Strategy Transformation
Setting the Update Strategy

- One of the following methods can be used to define the update strategy
  - Add an Update Strategy Transformation to the mapping to control how rows are flagged for insert, update, delete, or reject within a mapping
  - Define an Update Strategy when a session is configured.

To control how rows are flagged for insert, update, delete, or reject within a mapping, add an Update Strategy transformation to the mapping. The update strategy can be also be defined when a session is configured. All rows can be flagged for insert, delete, or update. Alternatively the data driven option can be selected, where the Integration service follows instructions coded into Update Strategy Transformations within the session mapping.

6.8. Update Strategy Transformation
Constants for Update Strategy Transformation

- The following constants can be used in the Update Strategy expression while creating an Update Strategy Transformation in the mapping

Operation	Constant	Numeric Value
Insert	DD_INSERT	0
Update	DD_UPDATE	1
Delete	DD_DELETE	2
Reject	DD_REJECT	3

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved   28

DD_INSERT - Flags records for insertion in an update strategy expression.
DD_INSERT is equivalent to the integer literal 0

DD_UPDATE - Flags records for update in an update strategy expression.
DD_UPDATE is equivalent to the integer literal 1

DD_DELETE - Flags records for deletion in an update strategy expression.
DD_DELETE is equivalent to the integer literal 2

DD_REJECT - Flags records for rejection in an update strategy expression.
DD_REJECT is equivalent to the integer literal 3. DD_REJECT is generally used to filter or validate data. If you flag a record as reject, the Integration service skips the record and writes it to the session reject file

Example:

The following expression marks items with an ID number of 1001 for deletion, and all

other items for insertion:

IIF(ITEM_ID = 1001, DD_DELETE, DD_INSERT)

Demo

- Create an Update Strategy Transformation
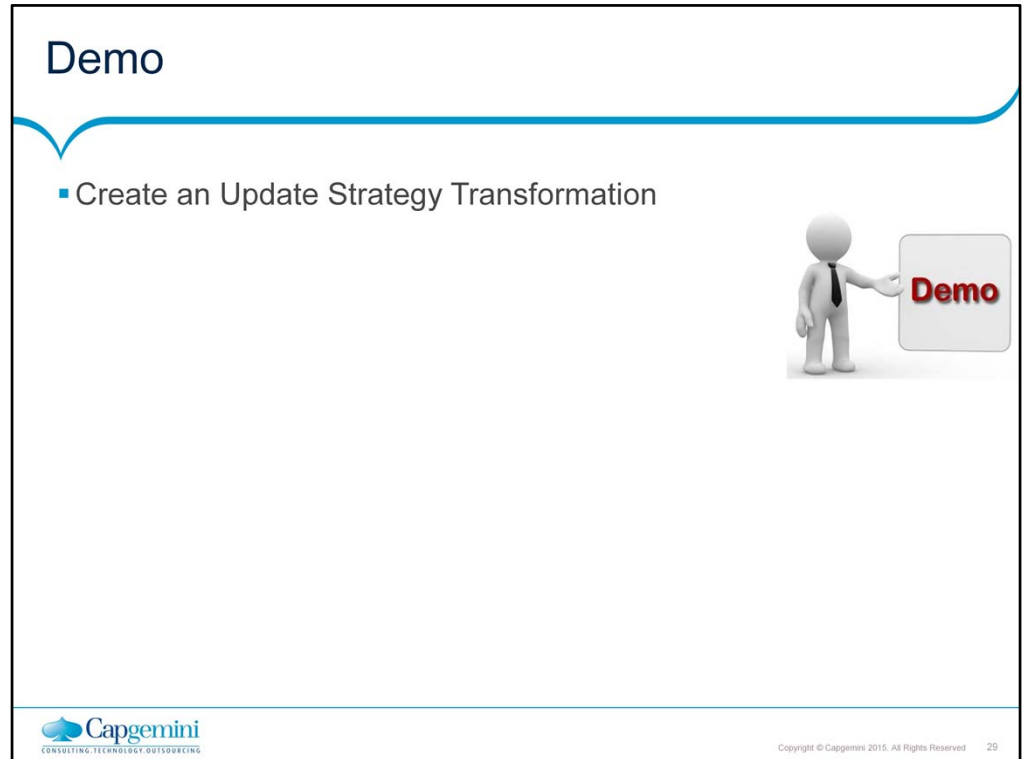
Steps to create an Update Strategy Transformation:

In the Mapping Designer, add an Update Strategy Transformation to a mapping.

Choose Layout | Link Columns.

Click and drag all ports from another transformation representing data to be passed through the Update Strategy Transformation. In the Update Strategy Transformation, the Designer creates a copy of each port which is clicked and dragged. The Designer also connects the new port to the original port. Each port in the Update Strategy Transformation is a combination input/output port.

Normally, all of the columns destined for a particular target would be selected. After they pass through the Update Strategy Transformation, this information is flagged for update, insert, delete, or reject.

Open the Update Strategy Transformation and rename it. The naming convention for Update Strategy Transformations is UPD_TransformationName.

Click the Properties tab.

Click the button in the Update Strategy Expression field. The Expression Editor appears.

Enter an update strategy expression to flag rows as inserts, deletes, updates, or rejects.

Validate the expression and click OK.

Click OK to save the changes.

Connect the ports in the Update Strategy Transformation to another transformation or a target instance.

Choose Repository | Save.

Summary

- After completing this lesson you now know:
  - To schedule a Workflow
  - To edit and deleted a Workflow
  - View and configure session logs
  - Set tracing levels
  - Configure an Update Strategy Transformation

Summary

Review Question

- Question 1: The default name for the session log is _____

- Question 2: In case of a _____ scheduler, the same set of schedule settings can be used for Workflows in the folder

- Question 3: The _____ tracing level summarizes session results, but not at the level of individual rows

- Question 4: Update Strategy Transformation is a passive transformation.
- True/False