

Unit objectives

After completing this unit, you should be able to:

- Sort data using in-stage sorts and Sort stage
- Combine data using Aggregator stage
- Combine data Remove Duplicates stage

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

2

Notes:

Sort Stage

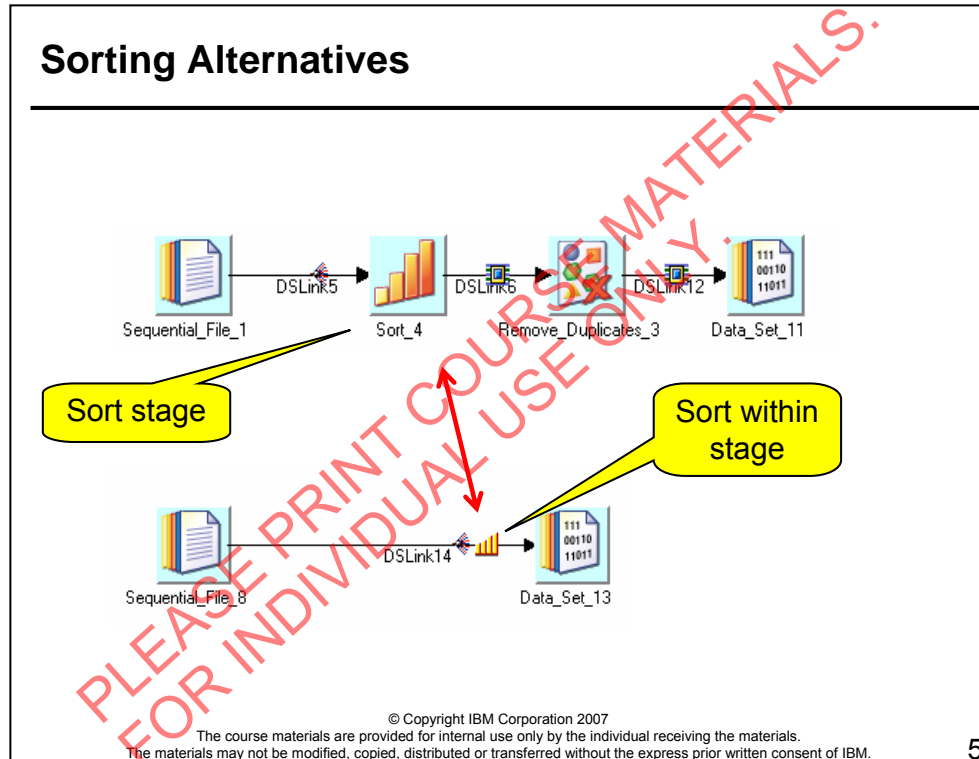
© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

3

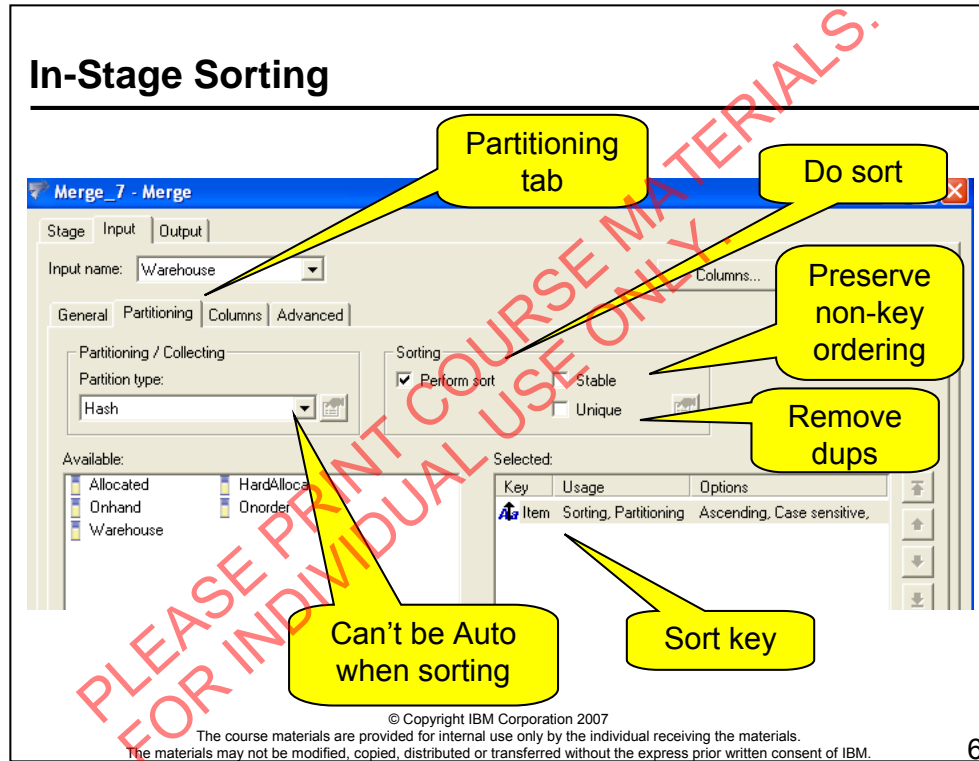
Sorting Data

- Uses
 - Some stages require sorted input
 - Join, merge stages require sorted input
 - Some stages use less memory with sorted input
 - E.g., Aggregator
- Sorts can be done:
 - Within stages
 - On input link Partitioning tab, set partitioning to anything other than Auto
 - In a separate Sort stage
 - Makes sort more visible on diagram
 - Has more options

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

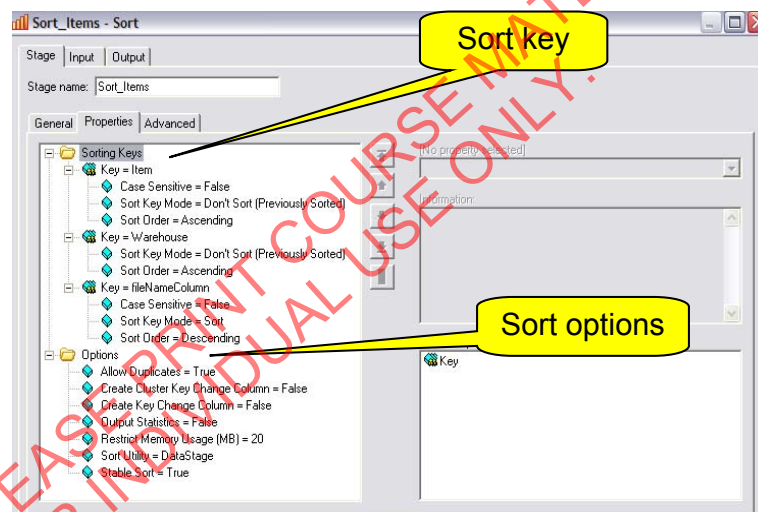


Which should you use? The first is more explicit and has more options. For example, the first has properties for specifying the amount of memory allocated.



Stable will preserve the original ordering of records within each key group. If set to false no prior ordering of records is guaranteed to be preserved by the sorting operation.

Sort Stage



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

7

Sort keys

- Add one or more keys
- Specify sort mode for each key
 - Sort: Sort by this key
 - Don't sort (previously sorted):
 - Assumes the data has already been sorted on this key
 - Continue sorting by any secondary keys
- Specify sort order: ascending / descending
- Specify case sensitive or not

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

8

Sort Options

- Sort Utility
 - DataStage – the default
 - Unix: Don't use. Slower than DataStage sort utility
- Stable
- Allow duplicates
- Memory usage
 - Sorting takes advantage of the available memory for increased performance
 - Uses disk if necessary
 - Increasing amount of memory can improve performance
- Create key change column
 - Add a column with a value of 1 / 0
 - 1 indicates that the key value has changed
 - 0 means that the key value hasn't changed
 - Useful for processing groups of rows in a Transformer

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Partitioning V. Sorting Keys

- Partitioning keys are often different than Sorting keys
 - Keyed partitioning (e.g., Hash) is used to group related records into the same partition
 - Sort keys are used to establish order within each partition
- Example:
 - Partition on HouseHoldID, sort on HouseHoldID, EntryDate
 - Partitioning on HouseHoldID ensures that the same ID will not be spread across multiple partitions
 - Sorting orders the records with the same ID by entry date
 - Useful for deciding which of a group of duplicate records with the same ID should be retained

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Aggregator Stage

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

11

Aggregator Stage

Purpose: Perform data aggregations

Specify:

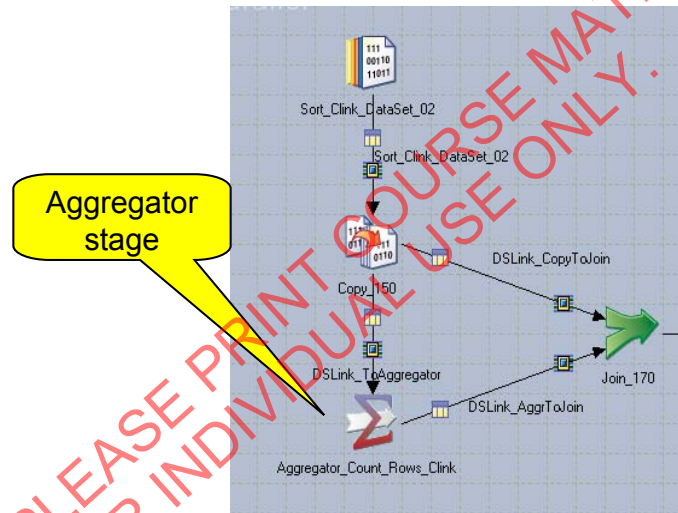
- One or more key columns that define the aggregation units (or groups)
- Columns to be aggregated
- Aggregation functions include, among many others:
 - count (nulls/non-nulls)
 - Sum
 - Max / Min / Range
- The grouping method (*hash table* or *pre-sort*) is a performance issue

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

12

Job with Aggregator Stage



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

13

Aggregation Types

- Count rows
 - Count rows in each group
 - Put result in a specified output column
- Calculation
 - Select column
 - Put result of calculation in a specified output column
 - Calculations include:
 - Sum
 - Count
 - Min, max
 - Mean
 - Missing value count
 - Non-missing value count
 - Percent coefficient of variation

© Copyright IBM Corporation 2007

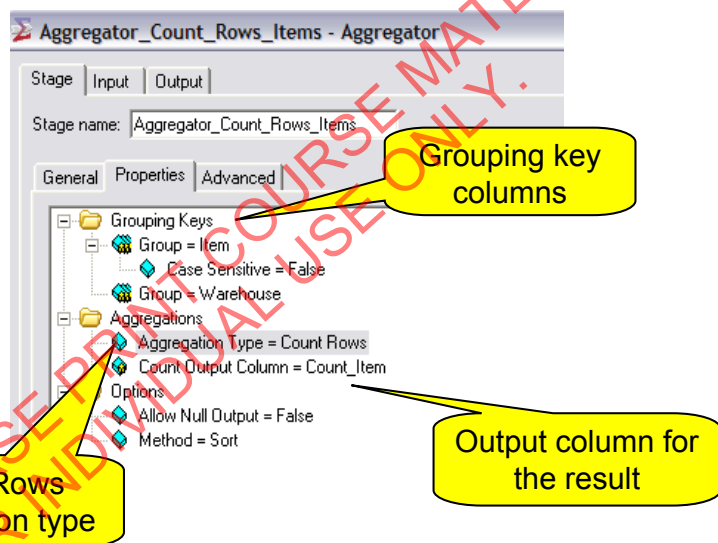
The course materials are provided for internal use only by the individual receiving the materials.

The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

14

There are two basic aggregation types: Count rows, calculation. The former counts the number of rows in each group. With the latter type, you select an input column that you want to perform calculations on. Then you select the calculations to perform on that input column and the output columns to put the results in.

Count Rows Aggregator Properties

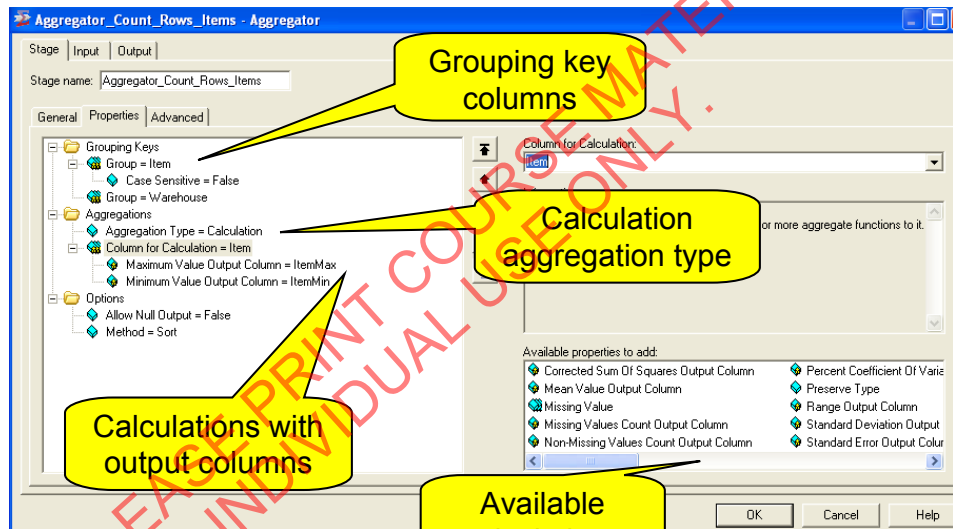


© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

15

Calculation Type Aggregator Properties



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

16

Grouping Methods

- Hash (default)
 - Calculations are made for all groups and stored in memory
 - Hash table structure (hence the name)
 - Results are written out after all input has been processed
 - Input does not need to be sorted
 - Useful when the number of unique groups is small
 - Running tally for each group's aggregations needs to fit into memory
- Sort
 - Requires the input data to be sorted by grouping keys
 - Does not perform the sort! Expects the sort
 - Only a single aggregation group is kept in memory
 - When a new group is seen, the current group is written out
 - Can handle unlimited numbers of groups

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

17

Remove Duplicates Stage

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

18

Removing Duplicates

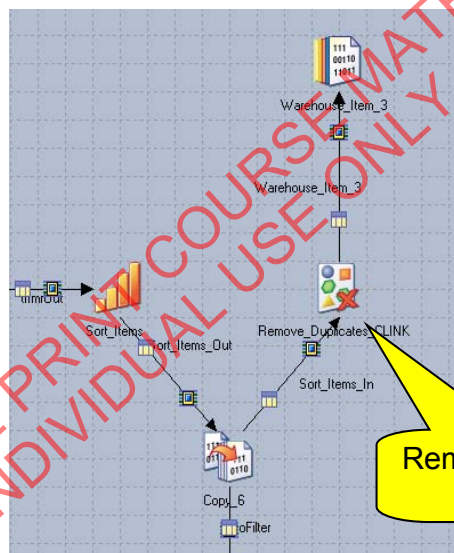
- Can be done by Sort stage
 - Use unique option
 - No choice on which duplicate to keep
 - Stable sort always retains the first row in the group
 - Non-stable sort is indeterminate

OR

- Remove Duplicates stage
 - Has more sophisticated ways to remove duplicates
 - Can choose to retain first or last

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Remove Duplicates Stage Job

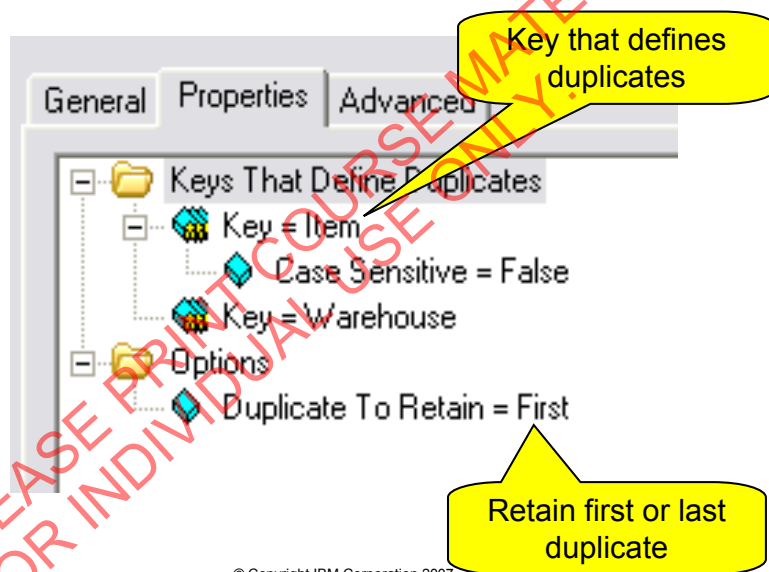


© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

20

Remove Duplicates Stage Properties



© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Checkpoint

1. What stage is used to perform calculations of column values grouped in specified ways?
2. In what two ways can sorts be performed?
3. What is a stable sort?

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

22

Notes:

Write down your answers here:

1.

2.

Checkpoint solutions

1. Aggregator stage
2. Using the Sort stage. In-stage sorts.
3. Stable sort preserves the order of non-key values.

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

23

Unit summary

Having completed this unit, you should be able to:

- Sort data using in-stage sorts and Sort stage
- Combine data using Aggregator stage
- Combine data Remove Duplicates stage

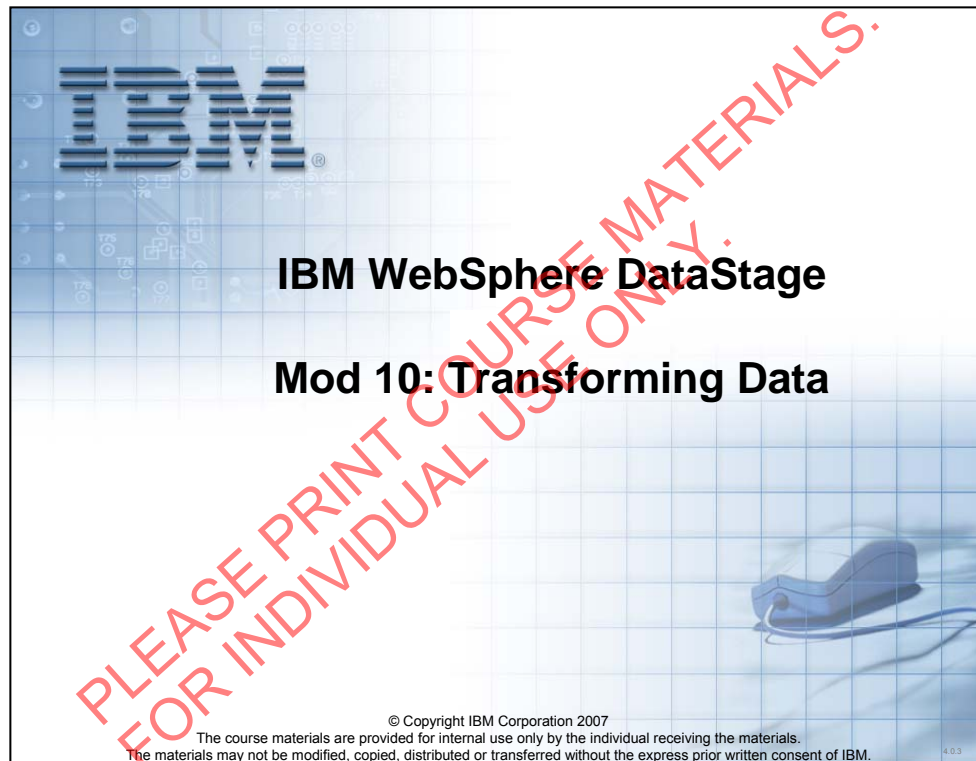
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

24

Notes:



Unit objectives

After completing this unit, you should be able to:

- Use the Transformer stage in parallel jobs
- Define constraints
- Define derivations
- Use stage variables
- Create a parameter set and use its parameters in constraints and derivations

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

26

Notes:

Transformer Stage

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

27

Transformer Stage

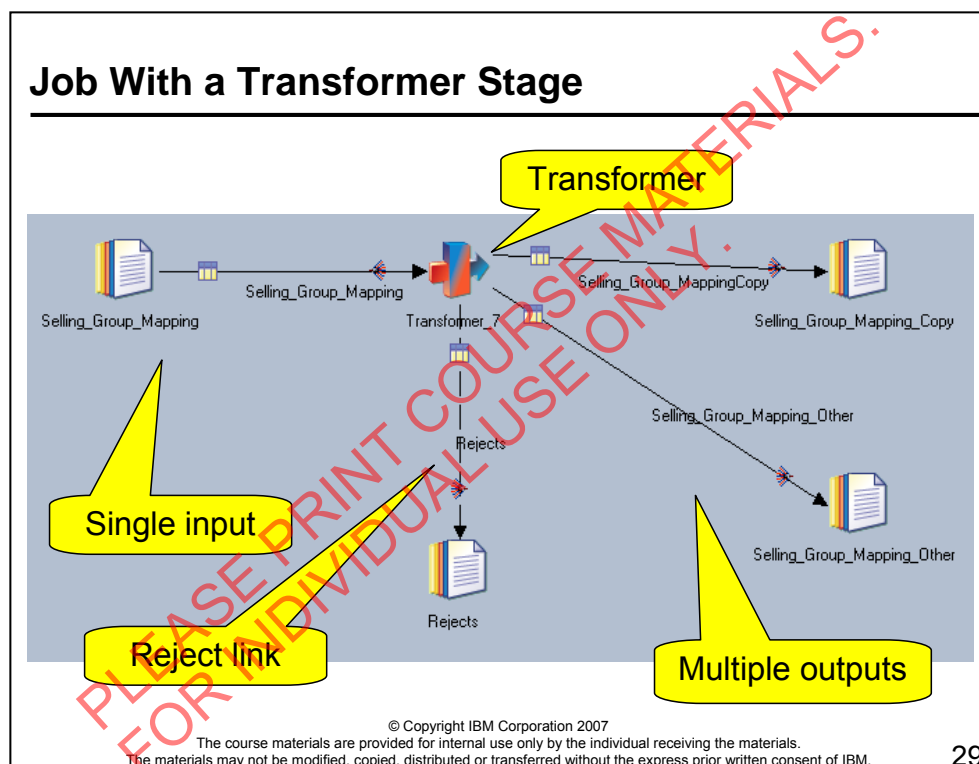
- Column mappings
- Derivations
 - Written in Basic
 - Final compiled code is C++ generated object code
- Constraints
 - Filter data
 - Direct data down different output links
 - For different processing or storage
- Expressions for constraints and derivations can reference
 - Input columns
 - Job parameters
 - Functions
 - System variables and constants
 - Stage variables
 - External routines

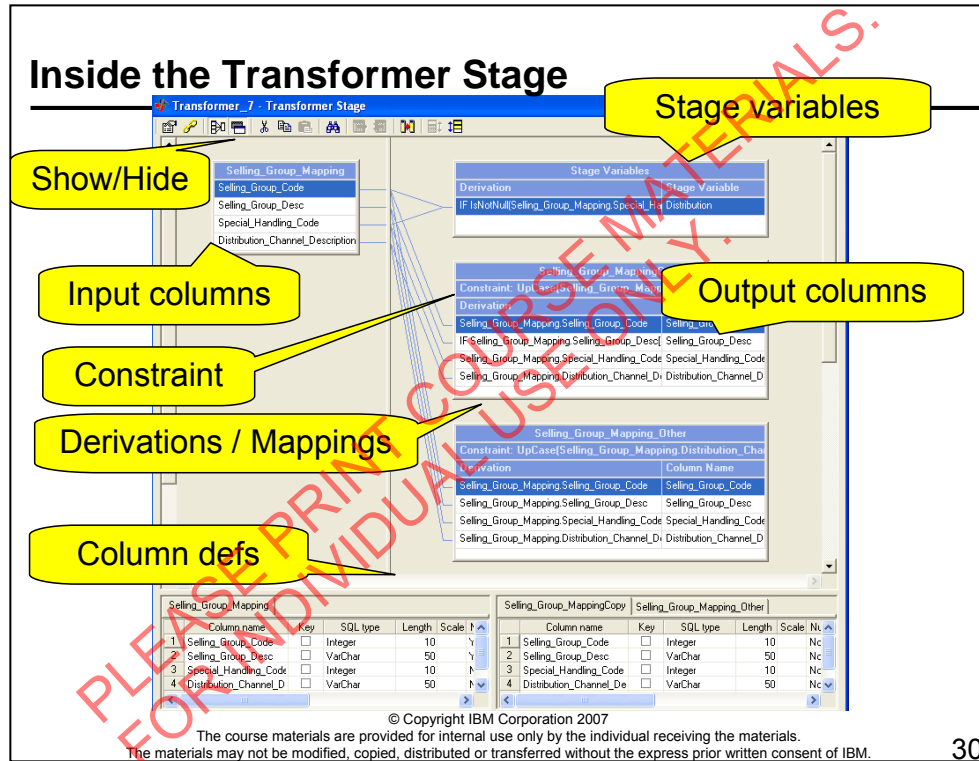
© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

28

In this module, we'll discuss two stages that can be used to perform derivations to implement business logic. Our main focus is on the Transformer stage. We'll also briefly discuss the Modify stage.





30

Link naming conventions are important because they identify appropriate links in the stage properties screen shown above.

Four quadrants:

- Incoming data link (one only)

- Outgoing links (can have multiple)

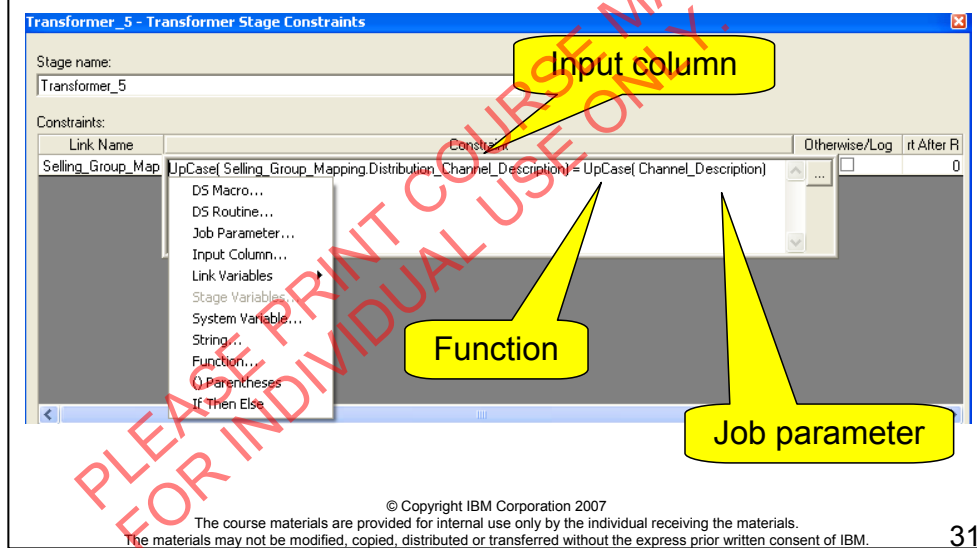
- Meta data for all incoming links

- Meta data for all outgoing links – may have multiple tabs if there are multiple outgoing links

Note the constraints bar – if you double-click on any you will get screen for defining constraints for all outgoing links.

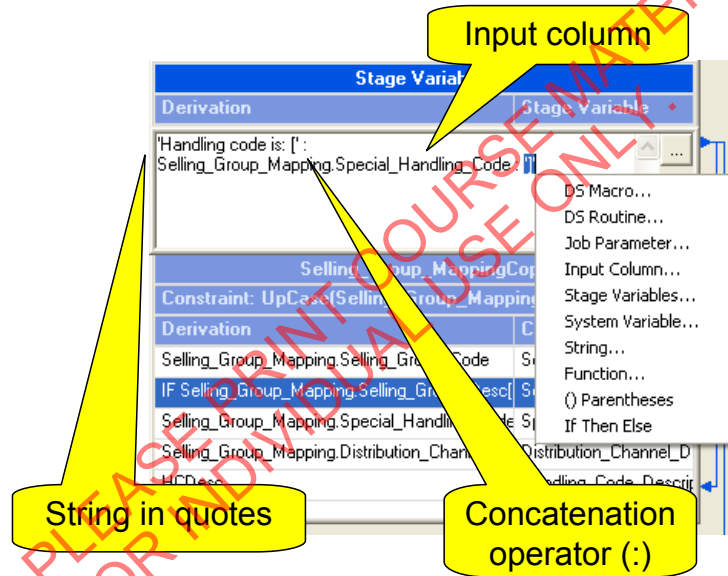
Double-click to the left of an output column or stage variable to define its derivation.

Defining a Constraint



Click the Constraints icon at the top of the Transformer (yellow chain) to open the Constraints window. Select items from the menu or type in items to build the expression.

Defining a Derivation



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

32

IF THEN ELSE Derivation

- Use IF THEN ELSE to conditionally derive a value
- Format:
 - IF <condition> THEN <expression1> ELSE <expression2>
 - If the condition evaluates to true then the result of expression1 will be copied to the target column or stage variable
 - If the condition evaluates to false then the result of expression2 will be copied to the target column or stage variable
- Example:
 - Suppose the source column is named In.OrderID and the target column is named Out.OrderID
 - Replace In.OrderID values of 3000 by 4000
 - IF In.OrderID = 3000 THEN 4000 ELSE Out.OrderID

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

33

String Functions and Operators

- Substring operator
 - Format: "String" [loc, length]
 - Example:
 - Suppose In.Description contains the string "Orange Juice"
 - InDescription[8,5] → "Juice"
- UpCase(<string>) / DownCase(<string>)
 - Example: UpCase(In.Description) → "ORANGE JUICE"
- Len(<string>)
 - Example: Len(In.Description) → 12

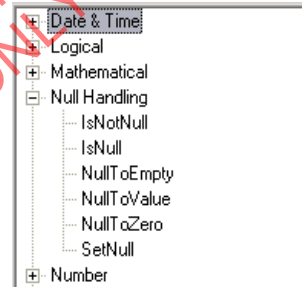
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

34

Checking for NULLs

- Nulls can be introduced into the data flow from lookups
 - Mismatches (lookup failures) can produce nulls
- Can be handled in constraints, derivations, stage variables, or a combination of these
- NULL functions
 - Testing for NULL
 - IsNull(<column>)
 - IsNotNull(<column>)
 - Replace NULL with a value
 - NullToValue(<column>, <value>)
 - Set to NULL: SetNull()
 - Example: IF In.Col = 5 THEN
SetNull() ELSE In.Col



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

35

If you perform a lookup from a lookup stage and choose the continue option for a failed lookup, you have the possibility of nulls entering your data flow.

Important: Be sure to use the NULL functions in stage variable derivations rather than output column derivations. Reference the stage variable in column derivations rather than performing the NULL function in the column derivation. The reason for this is that records containing NULLs are rejected BEFORE column derivations are performed (but not before stage variable derivations are performed).

Transformer Functions

- Date & Time
- Logical
- Null Handling
- Number
- String
- Type Conversion

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

36

Transformer Execution Order

- Derivations in stage variables
- Constraints are executed before derivations
- Column derivations in earlier links are executed before later links
- Derivations in higher columns are executed before lower columns

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

37

Transformer Stage Variables

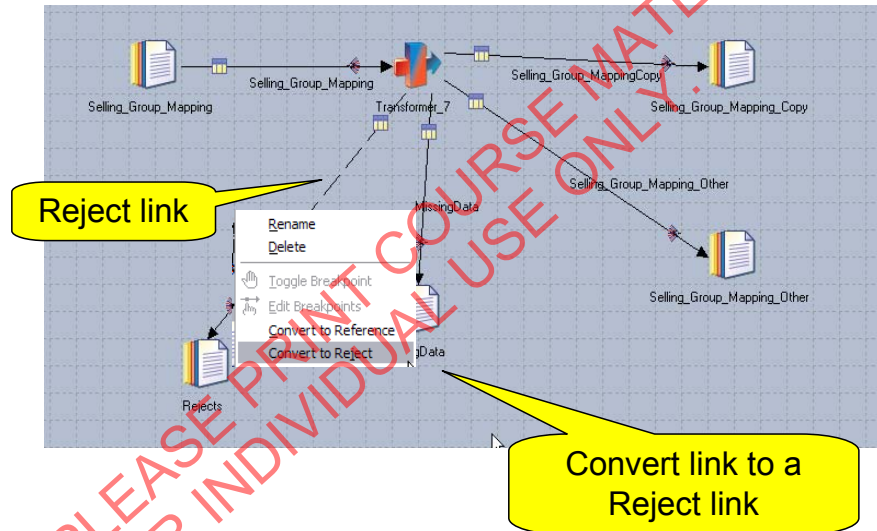
- Derivations execute in order from top to bottom
 - Later stage variables can reference earlier stage variables
 - Earlier stage variables can reference later stage variables
 - These variables will contain a value derived from the previous row that came into the Transformer
- Multi-purpose
 - Counters
 - Store values from previously read rows to make comparisons with the currently read row
 - Store derived values to be used in multiple target field derivations
 - Can be used to control execution of constraints

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

38

Transformer Reject Links



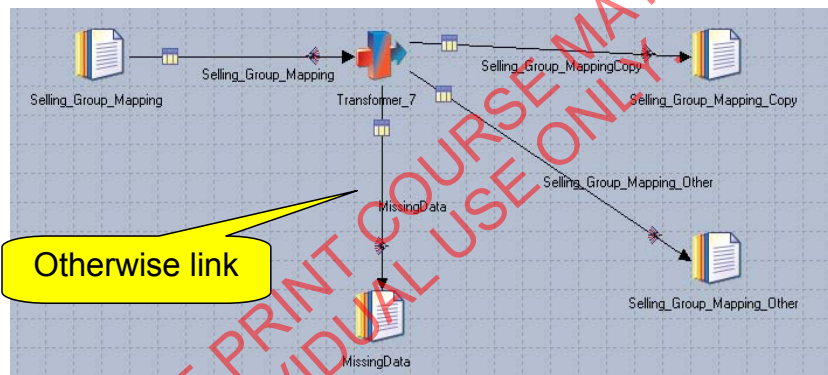
© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

39

Rows are sent to the Reject link if a nullable column is used in a calculation or derivation and that column has a NULL value. It is important to test for and replace NULLs in any derivations that reference nullable input columns. Position these tests for and replaces of NULLs within stage variable derivations.

Otherwise Link



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

40

The otherwise link captures what doesn't go down any previous links. It needs to come after the other links in the link ordering. It's defined on the Constraints tab, as shown in the next slide.

Defining an Otherwise Link

Transformer_7 - Transformer Stage Constraints

Stage name:
Transformer_7

Constraints:

| Link Name | Constraint | Otherwise/Log | Abort After Rows |
|-----------------------------|--|-------------------------------------|------------------|
| Selling_Group_MappingCopy | UpCase(Selling_Group_Mapping.Distribution_Channel_Description) = UpCase(C | <input type="checkbox"/> | 0 |
| Selling_Group_Mapping_Other | UpCase(Selling_Group_Mapping.Distribution_Channel_Description) <> UpCase(C | <input type="checkbox"/> | 0 |
| MissingData | | <input checked="" type="checkbox"/> | 50 |

Check to create
otherwise
condition

Can specify abort
condition

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

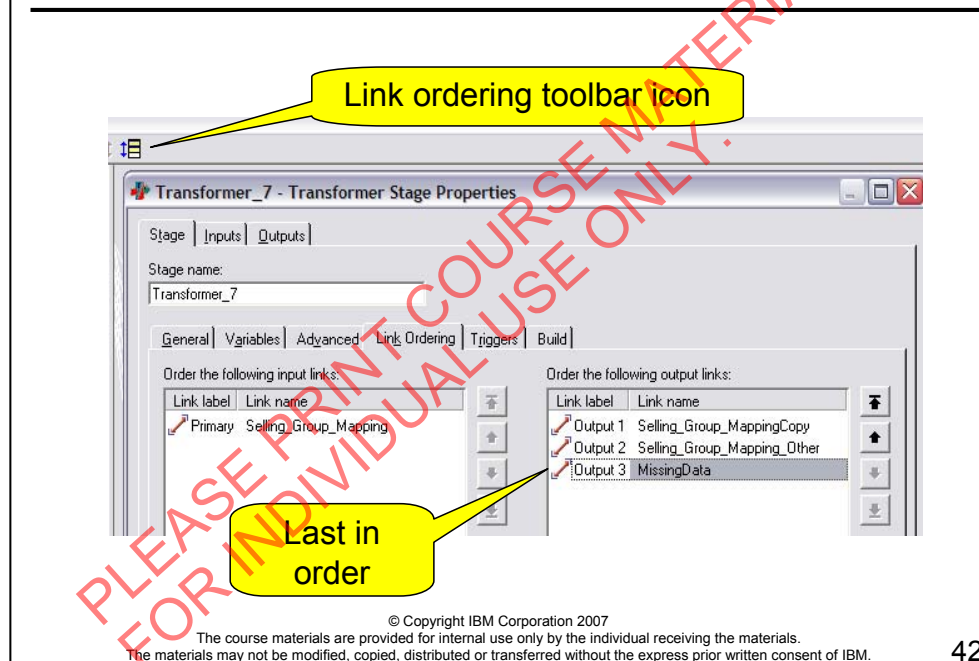
41

The Otherwise link must come after the other links carrying the main constraints. You can change the ordering of links, as shown on the next slide.

A warning is put in the log if any rows go down the otherwise link.

You can specify an abort condition for any output link. The abort occurs after the specified number of rows occurs in one of the partitions. It is not based on the total number of rows, but on the number of rows in a single partition.

Specifying Link Ordering



42

The Link Ordering icon is at the last icon on the toolbar.

Parameter Sets

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

43

Parameter Sets

- Store a collection of parameters in a named object
- One or more values files can be named and specified
 - A values file stores values for specified parameters
 - Values are picked up at runtime
- Parameter Sets can be added to the job parameters specified on the Parameters tab in the job properties

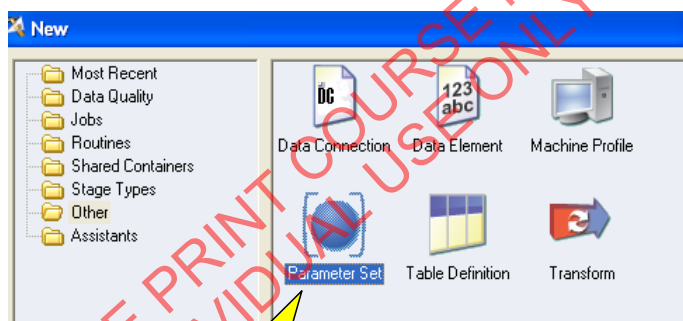
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

44

Creating a New Parameter Set



New Parameter Set

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

45

Click New and then select the Other folder.

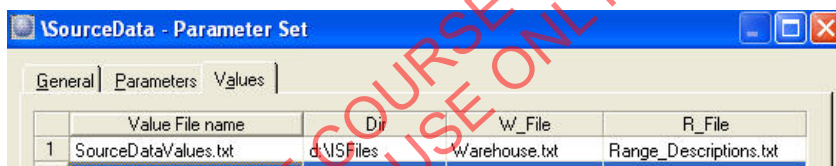
Parameters Tab

Specify parameters

Parameter set name is specified on General tab

| Parameter name | Prompt | Type | Default Value | Help Text |
|----------------|--------|--------|------------------------|-----------|
| Dir | Dir | String | d:\\$Files | |
| W_File | W_File | String | Warehouse.txt | |
| R_File | R_File | String | Range_Descriptions.txt | |

Values Tab



| | Value File name | Dir | W_File | R_File |
|---|----------------------|-----------|---------------|------------------------|
| 1 | SourceDataValues.txt | d:\SFiles | Warehouse.txt | Range_Descriptions.txt |

Values file name

Values for
parameters

© Copyright IBM Corporation 2007

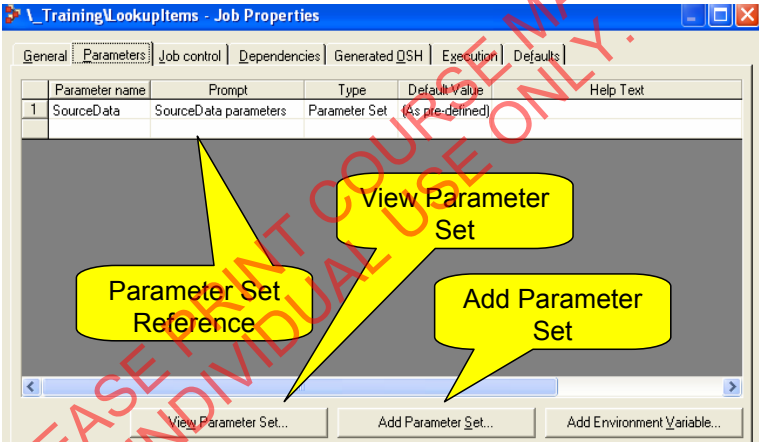
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

47

Parameter Set values files are located in the ...\\ParameterSets\\Project\\ directory of the Information Server install directory.

You can specify more than one value file with different sets of values.

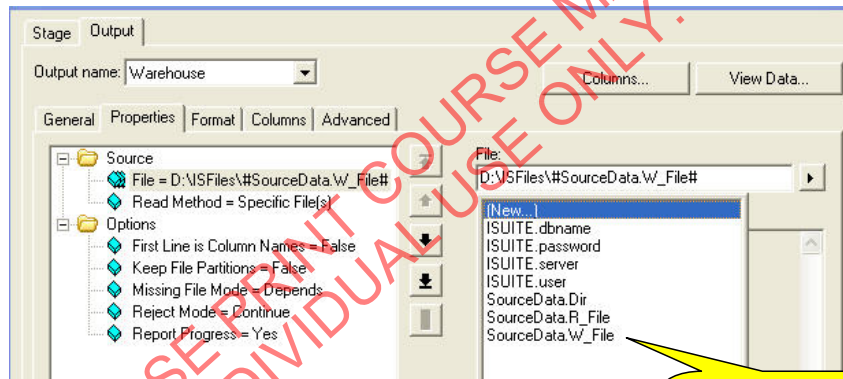
Adding a Parameter Set to Job Properties



PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Using Parameter Set Parameters



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

49

Notice that Parameter Set parameters are qualified by the name of the Parameter Set.

Checkpoint

1. What occurs first? Derivations or constraints?
2. Can stage variables be referenced in constraints?
3. Where should you test for NULLS within a Transformer?
Stage Variable derivations or output column derivations?

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

50

Notes:

Write down your answers here:

1.

2.

Checkpoint solutions

1. Constraints
2. Yes
3. Stage variable derivations. Reference the stage variables in the output column derivations.

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

51

Unit summary

Having completed this unit, you should be able to:

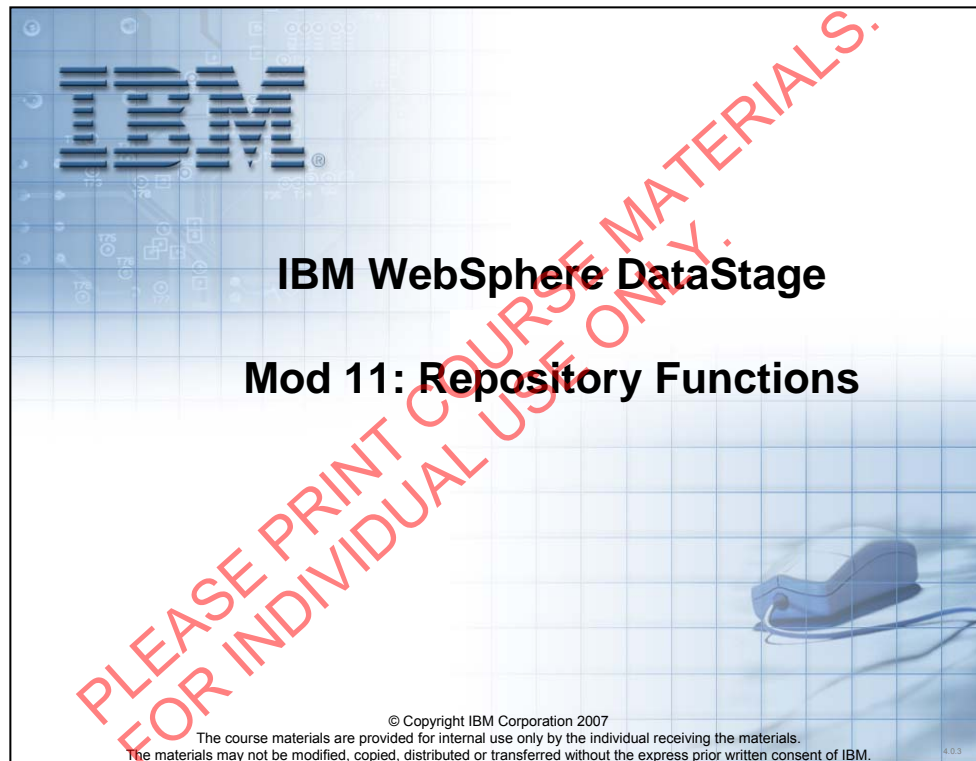
- Use the Transformer stage in parallel jobs
- Define constraints
- Define derivations
- Use stage variables
- Create a parameter set and use its parameters in constraints and derivations

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

52

Notes:



Unit objectives

After completing this unit, you should be able to:

- Perform a simple Find
- Perform an Advanced Find
- Perform an impact analysis
- Compare the differences between two Table Definitions
- Compare the differences between two jobs

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

54

Notes:

Searching the Repository

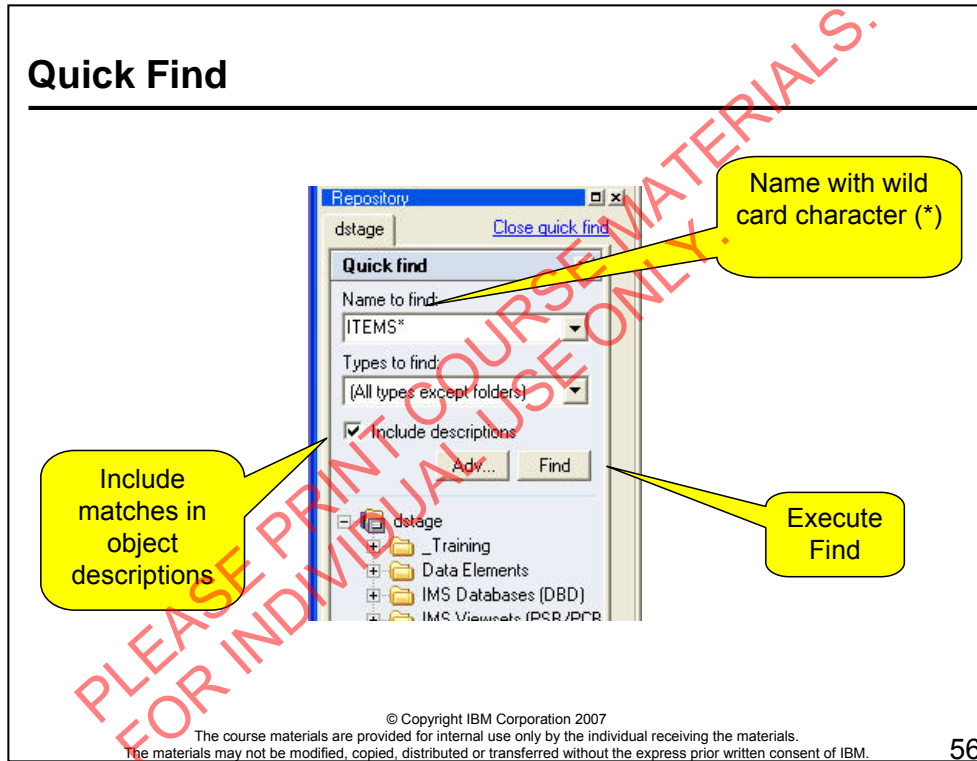
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

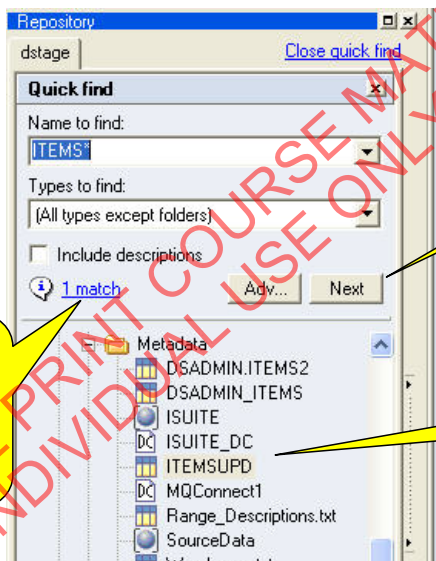
55

Quick Find



If the Include descriptions box is checked, the text in Short descriptions and Long descriptions will be searched.

Found Results



Number found; Click to open Advanced Find window

Highlight next item

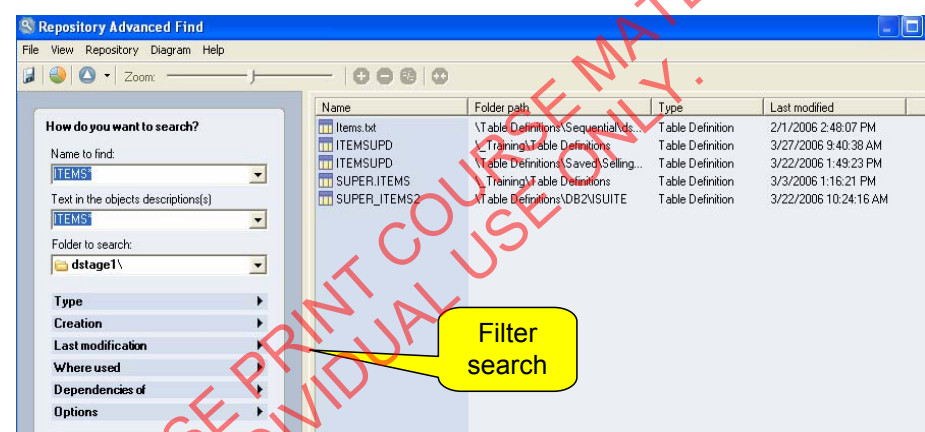
Found item

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

57

Advanced Find Window



PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Advanced Find Filtering Options

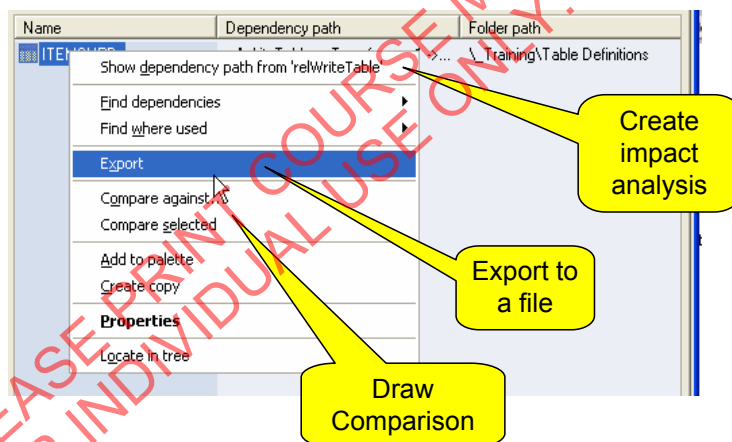
- Type: type of object
 - Job, Table Definition, etc.
- Creation: range of dates
 - E.g., Up to a week ago
- Last modification: range of dates
 - E.g., Up to a week ago
- Where used: objects that use specified objects
 - E.g., a job that uses a specified Table Definition
- Dependencies of: objects that are dependencies of objects
 - E.g., a Table Definition that is referenced in a specified job
- Options
 - Case sensitivity
 - Search within last result set

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

59

Using the Found Results



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

60

Impact Analysis

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

61

Performing an Impact Analysis

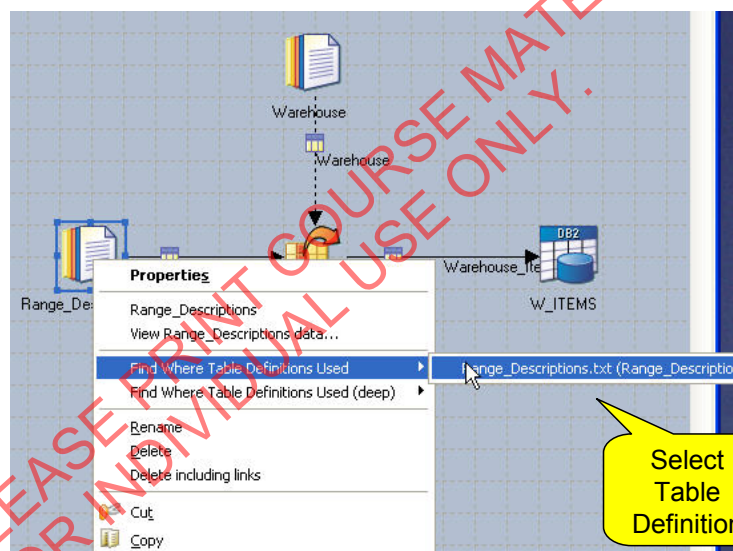
- Find where Table Definitions are used
 - Right-click over a stage or Table Definition
 - Select “Find where Table Definitions Used” or
 - Select “Find where Table Definitions Used (deep)”
 - Deep includes additional object types
 - Displays a list of the objects using the Table Definition
- Find object dependencies
 - Select “Find dependencies” or
 - Select “Find dependencies (deep)”
 - Displays list of objects dependent on the one selected
- Graphical functionality
 - Display the dependency path
 - Collapse selected objects
 - Move the graphical object
 - “Birds-eye” view

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

62

Initiating an Impact Analysis from a Stage



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

63

Displaying the Dependencies Graphically

Impact Analysis - Results Viewer

6 Jobs

LookupItemDescri...
LookupItemDescri...
LookupItemDescri...
LookupItems
LookupItems_1
LookupItems_2

Depends on

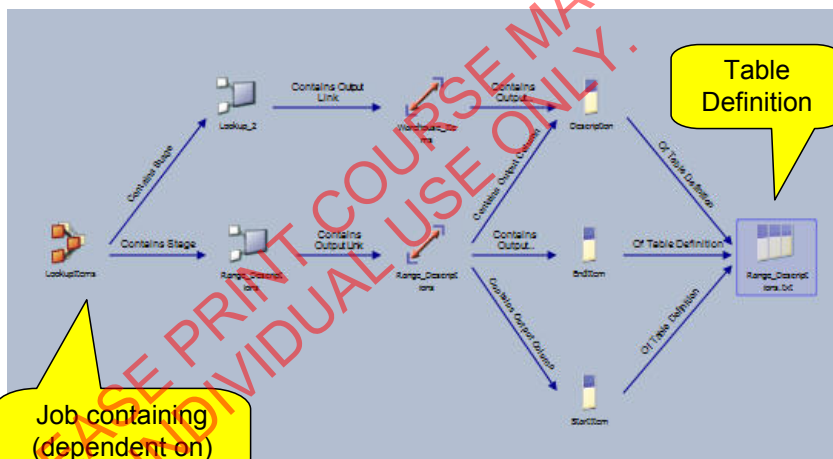
Range_Descriptions.txt

Birds-Eye view

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

64

Displaying the Dependency Path



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

65

Generating an HTML Report

WebSphere

DataStage and QualityStage

Advanced Find Results

Report generated on Monday, August 14, 2006, at 2:24:00 PM
From project datage on server HAWKDEMO
DataStage server version 8.0

Find Criteria

Name matches: *
Contained within folder: \
Types to include: (All types except folder)
Where used:

- _TrainingMetadataRange_Descriptions.txt

Case insensitive: Yes

Results

| Name | Dependency path | Folder path | Type |
|------------------------|--|----------------|--------------|
| LookopItemDescriptions | <ul style="list-style-type: none">• LookopItemDescriptions -> Lookop_2 -> Warehouse_Items -> Description -> Range_Descriptions.txt• LookopItemDescriptions -> Range_Descriptions -> Range_Descriptions -> Description -> Description -> Range_Descriptions.txt | _TrainingJobs | Parallel Job |

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Job and Table Difference Reports

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

67

Finding the Difference Between Two Jobs

- Example: Job1 is saved as Job2. Changes are made to Job2. What changes have been made?
 - Here Job1 may be a production job. Job2 is a copy of the production job after enhancements or other changes have been made to it

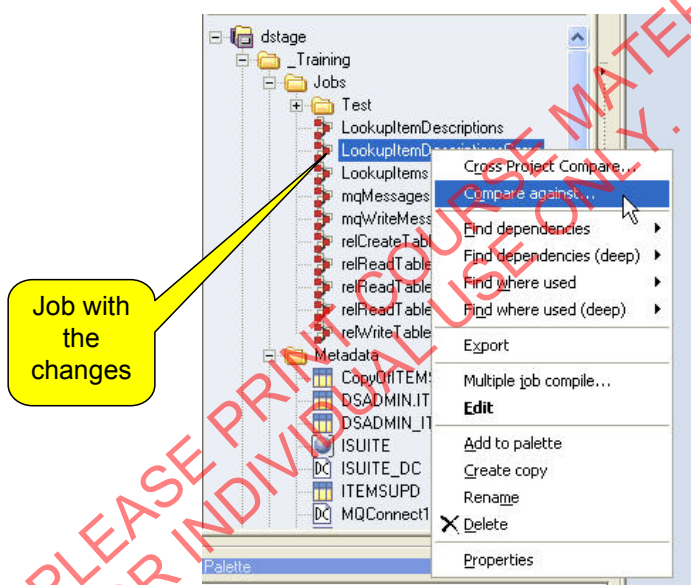
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

68

Initiating the Comparison



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Comparison Results

Comparison Results

Comparing LookupItemDescriptionsComp against LookupItemDescriptions

- Job Properties (1 change)
 - Property **Name** was **changed** from LookupItemDescriptions to LookupItemDescriptionsComp
- Stages (6 Changes)
 - Range_Descriptions (2 Changes)
 - Outputs (2 Changes)
 - Range_Descriptions (2 Changes)
 - Properties (1 change)
 - Property First Line is Column Names was changed from FirstLineColumn
 - Column Changes (1 Change)
 - StartItem (1 change)
 - Property Precision was changed from 50 to 111
 - W_ITEMS (2 Changes)
 - Inputs (2 Changes)
 - Warehouse_Items (1 Change)
 - Warehouse_Items was Removed**
 - W_ITEMS (1 Change)
 - W_ITEMS was Added**
 - Lookup_2 (2 Changes)
 - Outputs (2 Changes)
 - Warehouse_Items (1 Change)
 - Warehouse_Items was Removed**
 - W_ITEMS (1 Change)
 - W_ITEMS was Added**

© Copyright IBM Corporation 2007

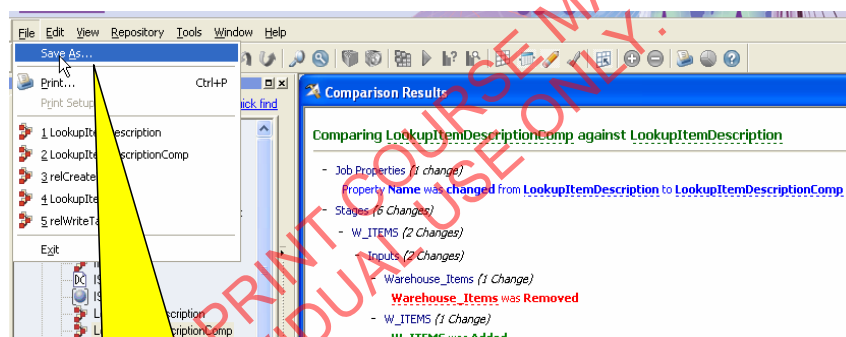
Click stage and
link references to
highlight in open
jobs

Click underlined item
to open stage editor

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

70

Saving to an HTML File



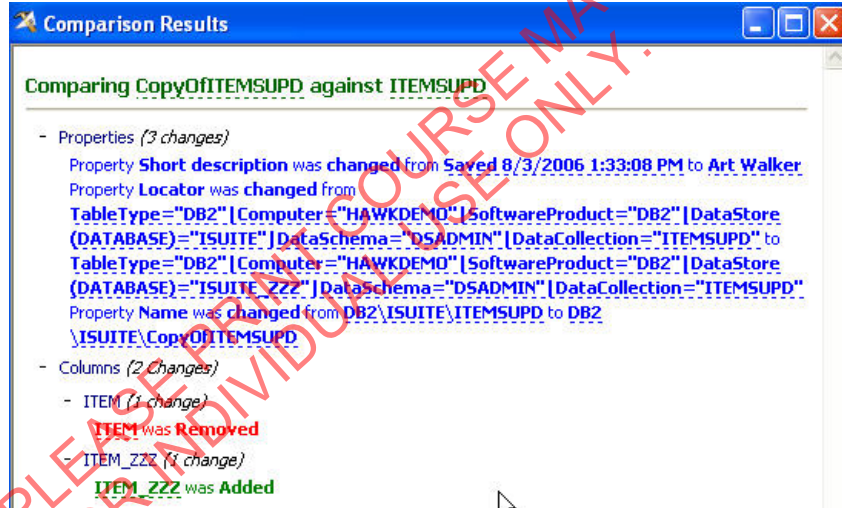
© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

71

Comparing Table Definitions

- Same procedure as when comparing jobs



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Checkpoint

1. You can compare the differences between what two kinds of objects?
2. What “wild card” characters can be used in a Find?
3. You have a job whose name begins with “abc”. You can’t remember the rest of the name or where the job is located. What would be the fastest way to export the job to a file?
4. Name three filters you can use in a Advanced Find?

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

73

Notes:

Write down your answers here:

1.

2.

Checkpoint solutions

1. Jobs. Table Definitions.
2. Asterisk (*). It stands for any zero or more characters.
3. Do a Find for objects matching "abc*". Filter by type job. Locate the job in the result set, click the right mouse button over it, and then click Export.
4. Type of object, creation date range, last modified date range, where used, dependencies of, other options including case sensitivity and search within last result set.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

74

Unit summary

Having completed this unit, you should be able to:

- Perform a simple Find
- Perform an Advanced Find
- Perform an impact analysis
- Compare the differences between two Table Definitions
- Compare the differences between two jobs

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

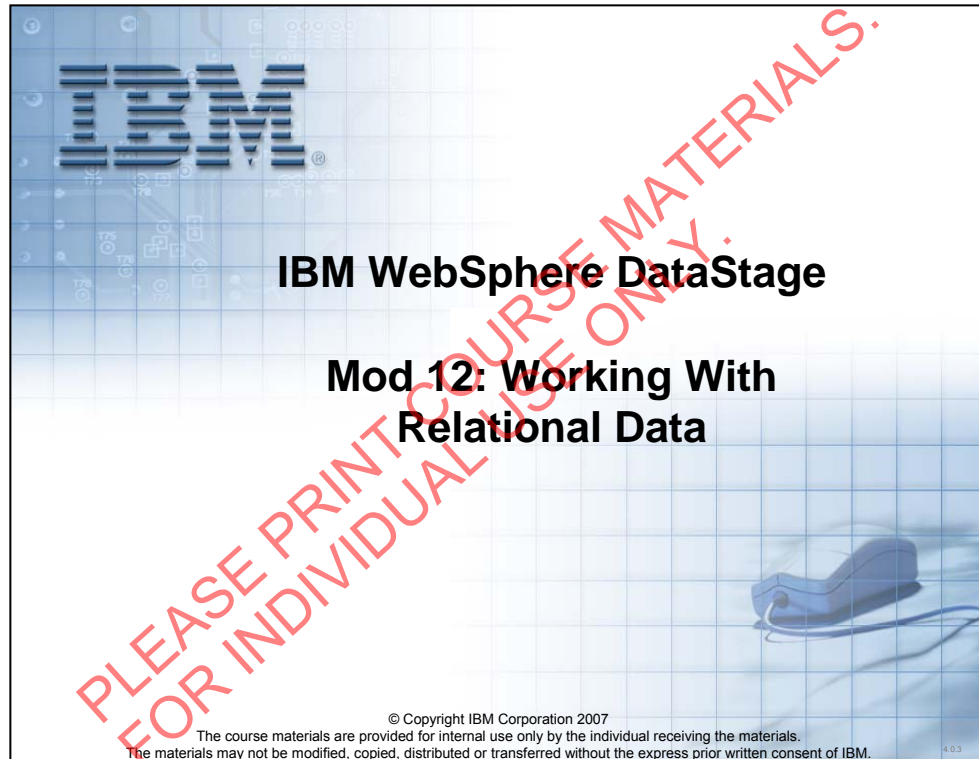
75

Notes:

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

76



Unit objectives

After completing this unit, you should be able to:

- Import Table Definitions for relational tables
- Create Data Connections
- Use Connector stages in a job
- Use SQL Builder to define SQL Select statements
- Use SQL Builder to define SQL Insert and Update statements
- Use the DB2 Enterprise stage

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

78

Notes:

Working with Relational Data

- Importing relational data
 - Import using ODBC or orchdbutil
 - orchdbutil is preferred, in order to get correct type conversions
- Data Connection objects
 - Store database connection information in a named object
- Stages available to access relational data
 - Connector stages
 - Parallel support
 - Most functionality
 - Consistent GUI and functionality across all relational types
 - Enterprise stages
 - Parallel support
 - Plug-in stages
 - Functionality ported from DataStage Server Jobs
 - Selecting data
 - Build SELECT statements using SQL Builder
- Writing data
 - Build INSERT, UPDATE, DELETE statements using SQL Builder

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

79

Our primary focus is on the Connector stages, but the DB2 Enterprise stage is also discussed.

Importing Table Definitions

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

80

Importing Table Definitions

- Can import using ODBC or using Orchestra schema definitions
 - Orchestra schema imports are better because the data types are more accurate
- Import>Table Definitions>Orchestra Schema Definitions
- Import>Table Definitions>ODBC Table Definitions

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

81

The orchdbutil GUI is limited to one table at a time. However, this utility is also available as a command-line utility that can be scripted to import a large number of Table Definitions.

Orchestrate Schema Import

1 of 7

Import from:

- ☐ File on Local system
- ☐ File on Server: 134.168.53.13
- ☒ Database table (via orchdbutil)

Enter the name of the database table whose schema is to be imported: SGM_01

For owner:

DBMS type: Database name: Database server:

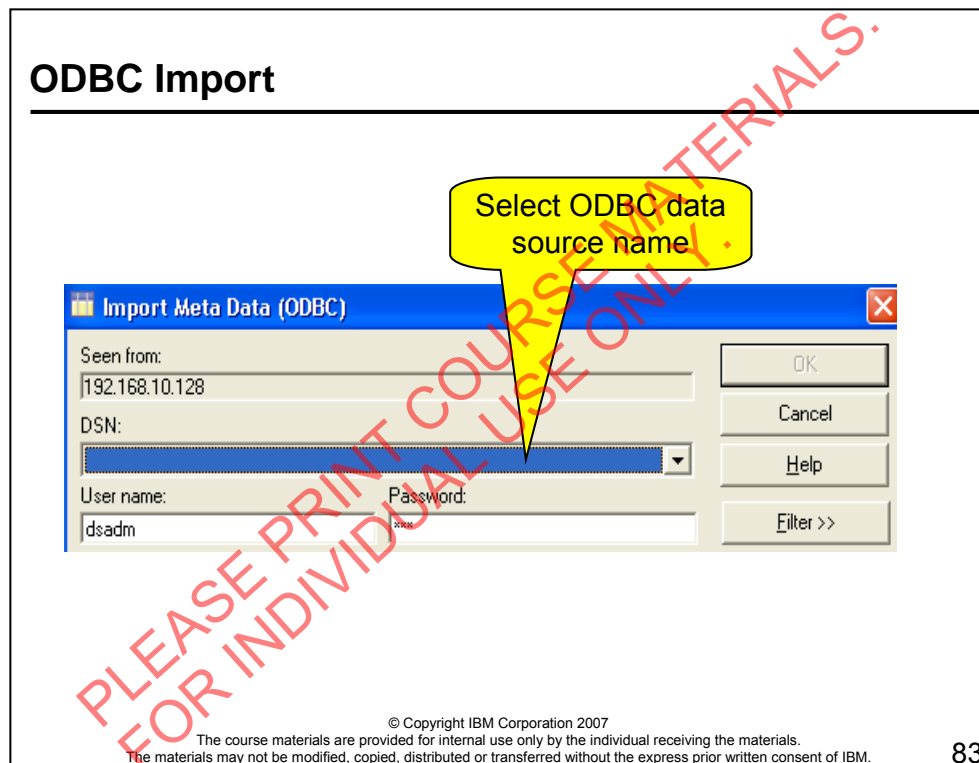
User name: Password:

☐ Use own DB options

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

82



The ODBC data source that accesses the database containing the tables to be imported must have been previously defined.

Connectors Stages

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

84

Connector Stage Types

- ODBC
 - Conforms to ODBC 3.5 standard and is Level 3 compliant
 - Certified with Oracle, DB2 UDB, SQL Server, and others
 - DataDirect drivers
- DB2 UDB
 - 8.1 and 8.2
- WebSphere MQ
 - WSMQ 5.3 and 6.0 for Client / Server
 - WSMB 5.0
- Teradata

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

85

Connector Stage Features

- Common stage editor
- Convenient drop-down lists to choose properties
- Job parameters can be inserted into any property
- Required properties are identified with a visual indicator
- Warning indicator for properties requiring attention
- Metadata retrieval
- Integrated SQL Builder
- Parallel support
 - Read: parallel connections to the server and modified SQL queries for each connection
 - Write: parallel connections to the server
- Transaction isolation level support
 - Read Uncommitted
 - Read Committed
 - Repeatable Read
 - Serializable
- Before / After commands
 - Executed once before or after the job runs

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

86

Metadata Retrieval

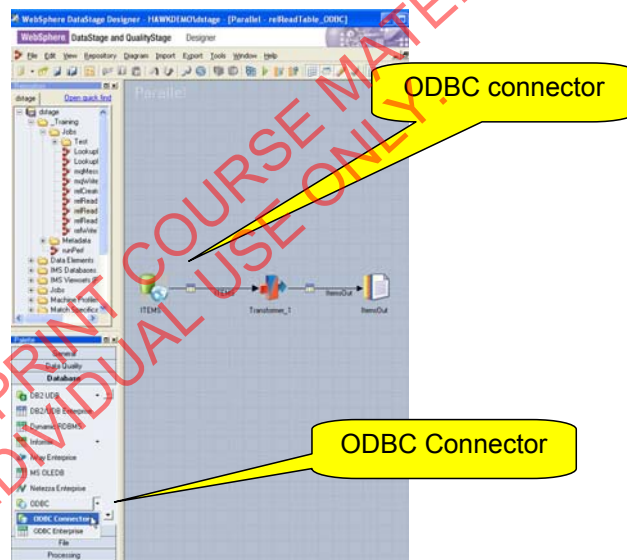
- List data sources
 - Can list both user and system DSNs
- List users defined for the given database
- Individual table or view metadata
 - Column data type, size, precision, etc.
- Database information
 - Name, vendor name, version number, etc.
- List of database schemas
- List tables that match the filter criteria
- List related tables
 - Those related by foreign key
 - Those whose foreign keys point to the primary key in the given table

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

87

Connector Stages



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.

The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

88

All the connector stages have a similar GUI and similar functionality. Here, the ODBC stage will be used for illustration.

Stage Editor

Click to display stage properties

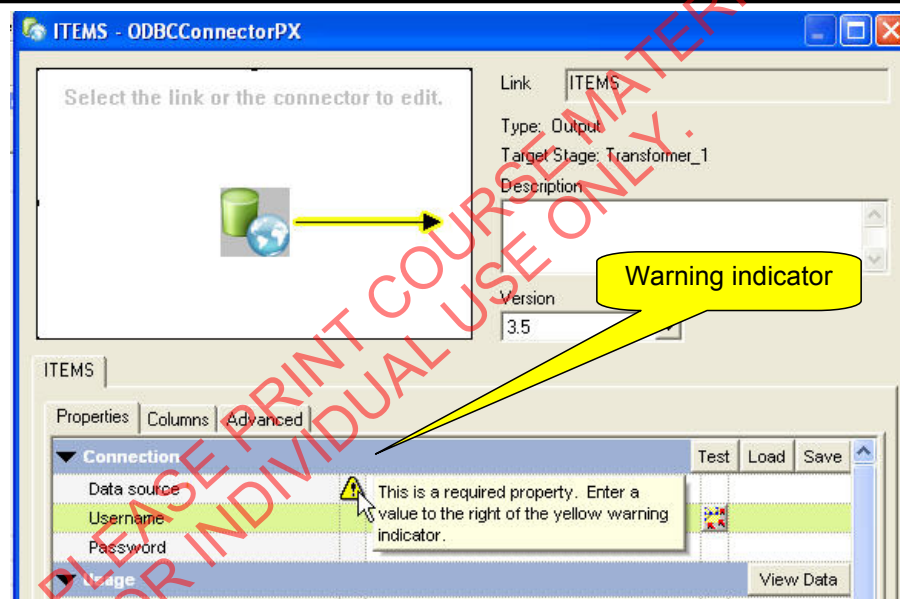
Click to display link properties and columns

Navigator panel

Properties

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Property Warning Indicators



© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

90

Auto-Generated SQL

Select the link or the connector to edit.

Link: ITEMS

Type: Output

Target stage: Transformer_1

Description:

Variant: 3.5

ITEMS

Properties | Columns | Advanced

| Connection | | Test | Load | Save |
|-------------|--------|------|------|------|
| Data source | ISUITE | | | |
| Username | | | | |
| Password | | | | |

View Data

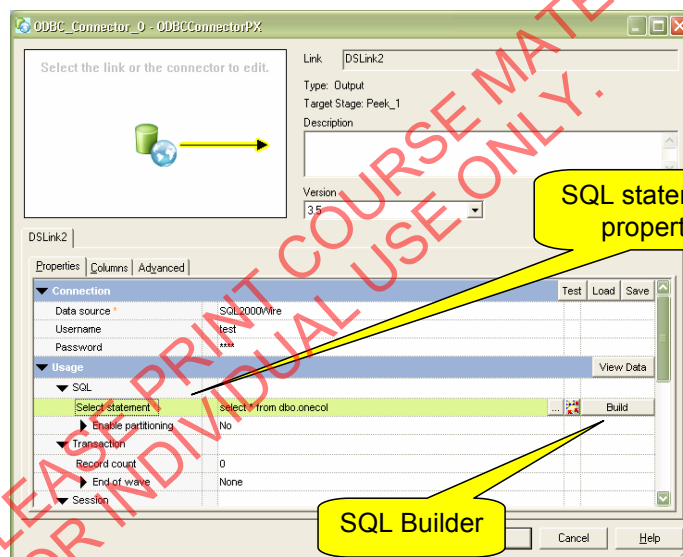
| Usage | |
|---------------------------|-------|
| Generate SQL | Yes |
| Table name | ITEMS |
| Enable quoted identifiers | No |

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

91

Invoking SQL Builder



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Connector Stage Properties

Connection information

SQL

Transaction and session management

Before/After the SQL

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

DB2 Enterprise Stage

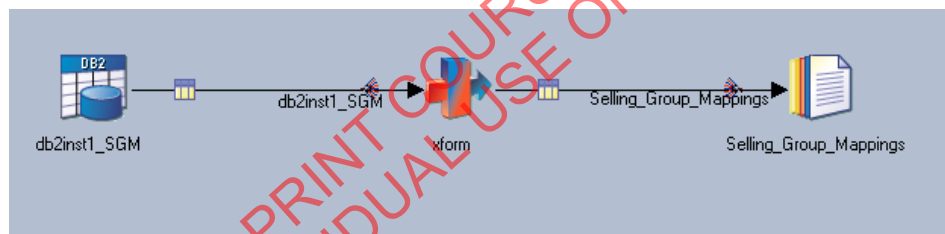
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

94

DB2 Enterprise Source Stage

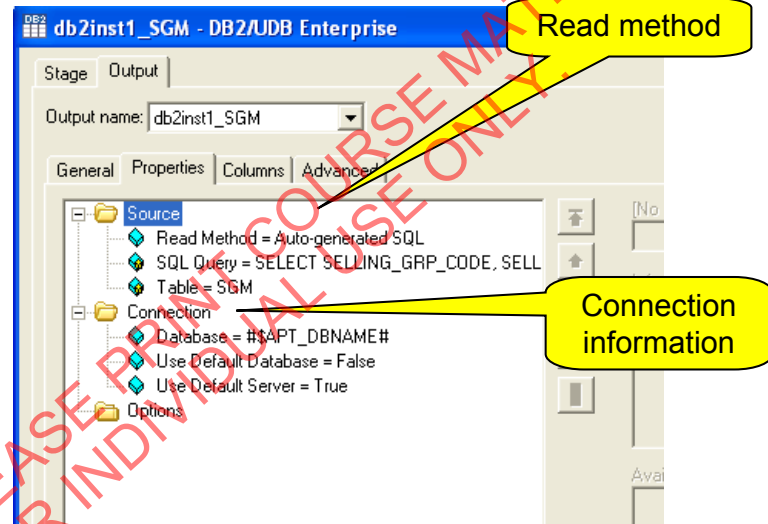


PLEASE PRINT * COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

95

DB2 Enterprise Source Stage Properties

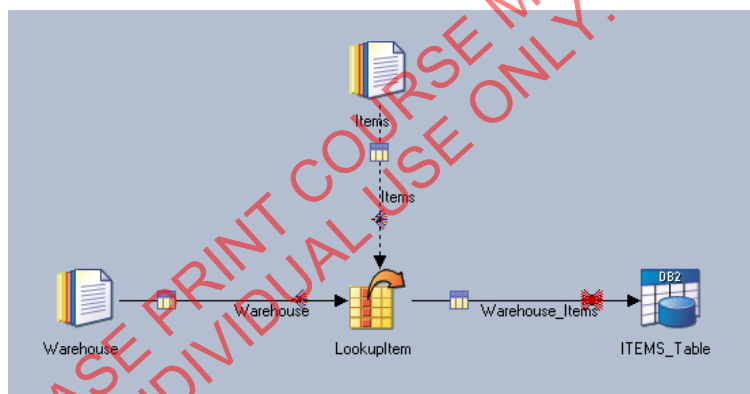


© Copyright IBM Corporation 2007
 The course materials are provided for internal use only by the individual receiving the materials.
 The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

96

Read method includes: Auto-generated SQL, User-defined SQL, SQL Builder generated SQL, and Table. The Table method reads from the table without using an SQL query.

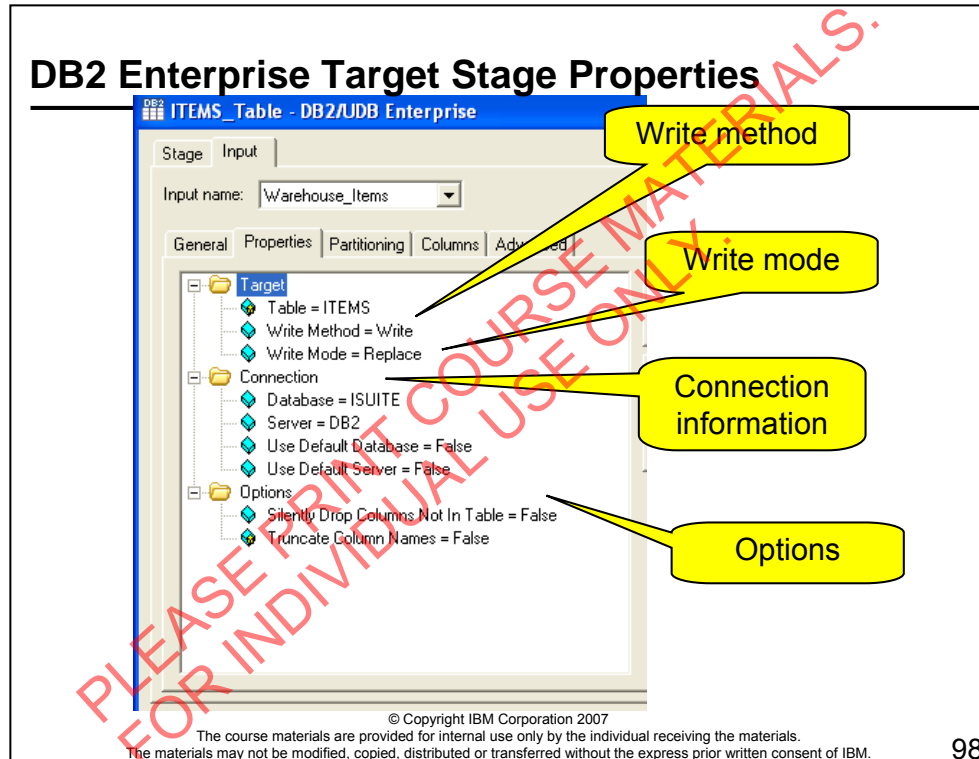
DB2 Enterprise Target Stage



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

97



Write method includes: Write to INSERT; Load uses fast DB2 loader technology; Upsert uses an INSERT or an UPDATE SQL statement. Delete Rows uses a DELETE SQL statement.

Write mode includes: Append to existing table; Create new table; Replace by dropping table and creating new; Truncate table records only.

Building a Query Using SQL Builder

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

99

SQL Builder

- Uses the Table Definition
 - Be sure the Locator tab information is specified fully and correctly
 - Schema and table names are based on Locator tab information
- Drag Table Definitions to SQL Builder canvas
- Drag columns from Table Definition to Select columns table

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

100

Table Definition Locator Tab

Locator tab

Table schema name

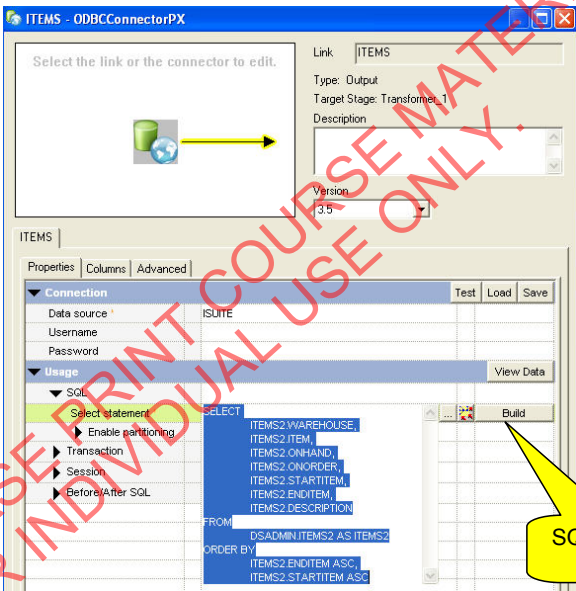
Table name

TableType: DB2
Computer: hawk_demo
SoftwareProduct: DB2
DATABASE: ISUITE
DataSchema: SUPER
TABLE: ITEMS
Definition imported through Data Connection : none.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

101

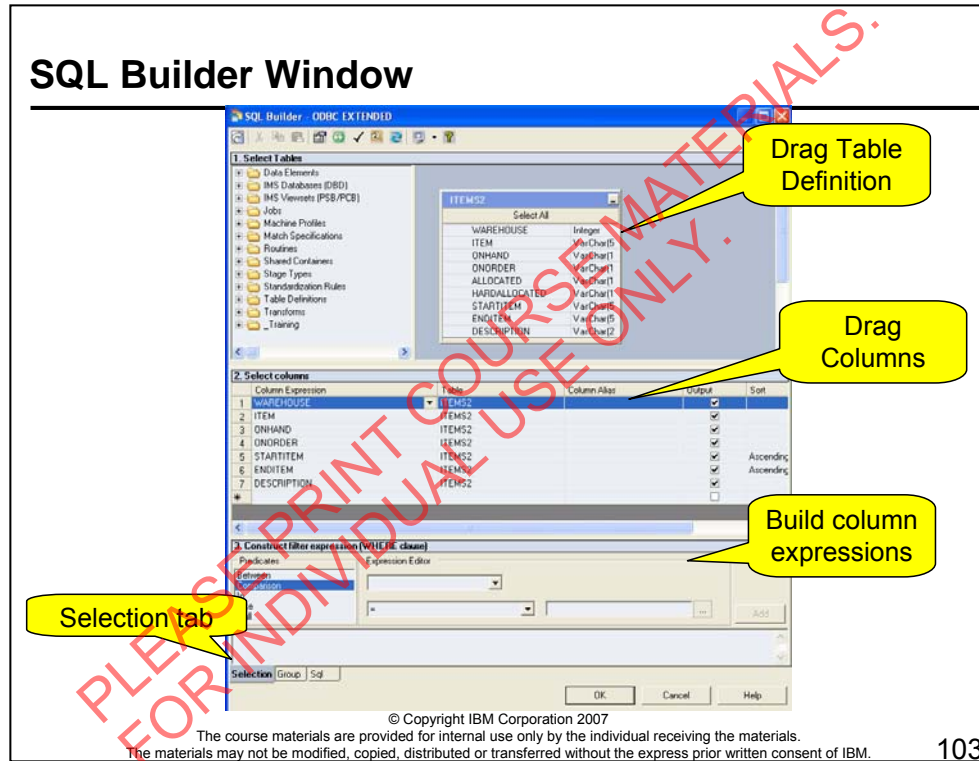
SQL Builder Read Method



SQL Builder for Select

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.



103

You build the query on the Selection tab, which is the first window you see when you open SQL Builder. Begin by dragging the Table Definitions for the tables to be read. Be sure the information on the Locator tab of the Table Definition is full and correct. In particular, be sure the table name and schema are specified. Otherwise the SQL statement that is corrected will not contain this information.

Creating a Calculated Column

Columns dragged from table

Select function

Select Expression Editor

Open a second Expression Editor

Functions

Select function

Function parameters

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

104

Here a new calculated column is being defined. The Expression Editor is selected from the column drop-down list. The UPPER function is selected. To define the parameter value the UPPER function is applied to, a second Expression Editor is opened. Here the SUBSTR function is selected to so select 20 characters from the beginning of the DESCRIPTION column.

Constructing a WHERE Clause

Select predicate

Job parameter

Add condition to clause

Add second job parameter

Sorting the Data

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

| | Column Expression | Table | Column Alias | Output | Sort | Sort Order |
|---|-------------------|-------------|--------------|-------------------------------------|-----------|------------|
| 1 | WAREHOUSE | ITEMS_ALIAS | | <input checked="" type="checkbox"/> | Ascending | 2 |
| 2 | ITEM | ITEMS_ALIAS | | <input checked="" type="checkbox"/> | Ascending | 1 |
| 3 | ONHAND | ITEMS_ALIAS | | <input checked="" type="checkbox"/> | | |
| 4 | ONORDER | ITEMS_ALIAS | | <input checked="" type="checkbox"/> | | |

3. Construct filter expression (WHERE clause)

Second column
to sort by

Sort Ascending
/ Descending

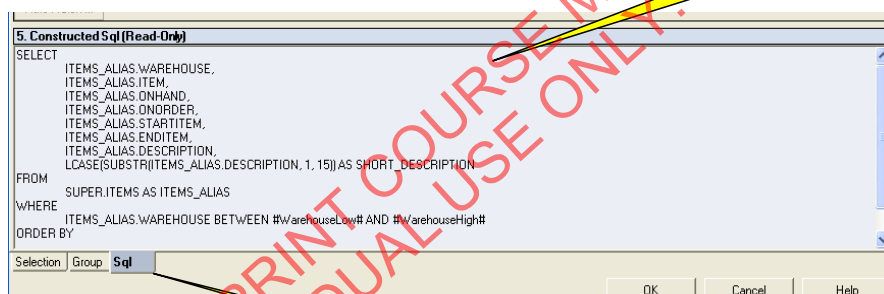
First column to
sort by

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

106

Viewing the Generated SQL



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

107

Building SQL to Write to a Table

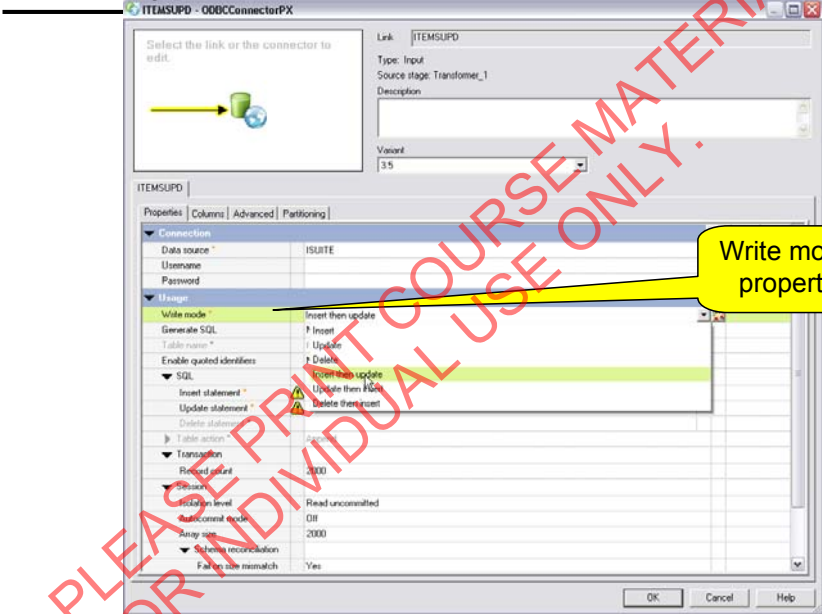
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

108

Select Type of Write



© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Invoking SQL Builder

ITEMSUPD | 3.5

Properties | Columns | Advanced | Partitioning

▼ Connection

Data source * | ISUITE

Username

Password

Test | Load | Save

▼ Usage

Write mode * | Insert then update

View Data

▼ SQL

Insert statement | INSERT INTO DSADMIN.ITEMSUPD (WAREHOUS... | Build

Update statement | UPDATE DSADMIN.ITEMSUPD SET WAREHOU...

Delete statement

Table action * | Append

Transaction

Session

Before/After SQL

Specify INSERT

Specify UPDATE

SQL Builder

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

110

SQL Builder INSERT

Drag columns to load

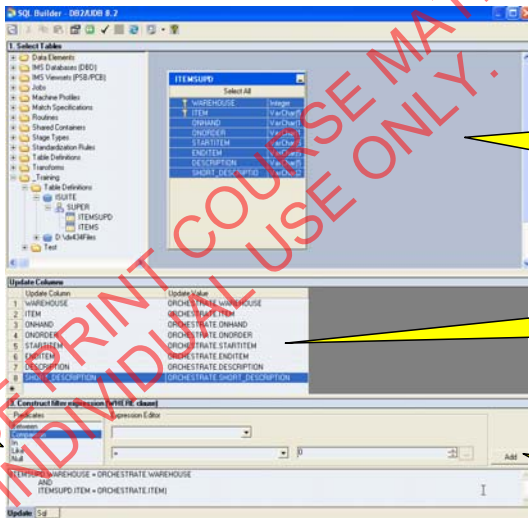
Drag Table Definition

Specify values

PLEASE PRINT COURSE MATERIALS ONLY.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

SQL Builder Update



Drag columns to load

WHERE clause

Drag Table Definition

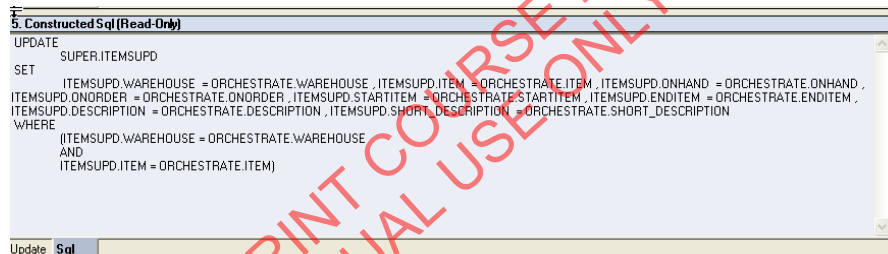
Specify values

Add condition to clause

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials. The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Viewing the SQL



```
5. Constructed Sql (Read-Only)
UPDATE
  SUPER.ITEMSUPD
SET
  ITEMSUPD.WAREHOUSE = ORCHESTRATE.WAREHOUSE , ITEMSUPD.ITEM = ORCHESTRATE.ITEM , ITEMSUPD.ONHAND = ORCHESTRATE.ONHAND ,
  ITEMSUPD.ONORDER = ORCHESTRATE.ONORDER , ITEMSUPD.STARTITEM = ORCHESTRATE.STARTITEM , ITEMSUPD.ENDITEM = ORCHESTRATE.ENDITEM ,
  ITEMSUPD.DESCRPTION = ORCHESTRATE.DESCRPTION , ITEMSUPD.SHORT_DESCRIPTION = ORCHESTRATE.SHORT_DESCRIPTION
WHERE
  (ITEMSUPD.WAREHOUSE = ORCHESTRATE.WAREHOUSE
  AND
  ITEMSUPD.ITEM = ORCHESTRATE.ITEM)
```

Update **Sql**

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

113

Data Connections

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

114

Data Connection

- Stores database parameters and values in a named object in the Repository
 - Optionally can create a parameter set with the parameters and values from with the Data Connection object
- Associated with a stage type
- Property values can be specified in a job stage of the given type by loading the Data Connection into the stage

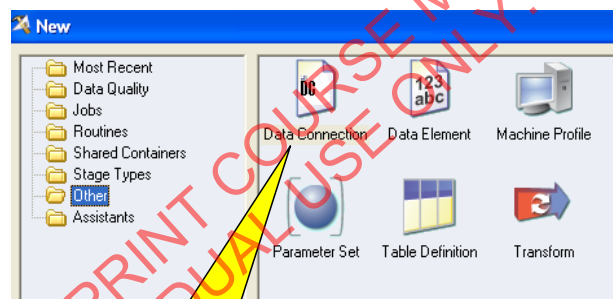
PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

115

Creating a New Data Connection Object



New Data
Connection

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

116

You can optionally save the parameters and values specified in a job stage into a new Data Connection. This provides another method for creating a Data Connection.

Select the Stage Type

dc _TrainingVSUITE - DataConnection

General Parameters

Connect using Stage Type:
DB2/UDB data access (PxDBC)

Connection parameters

| | Parameter name | Value |
|---|--|----------------|
| 1 | <input checked="" type="checkbox"/> server | #SUITE.server# |
| 2 | <input checked="" type="checkbox"/> dbname | #SUITE.dbname# |
| 3 | <input checked="" type="checkbox"/> user | #SUITE.user# |
| 4 | <input checked="" type="checkbox"/> password | |

Using Parameter Set:
JSUITE [Create Parameter Set](#)

Select the stage type

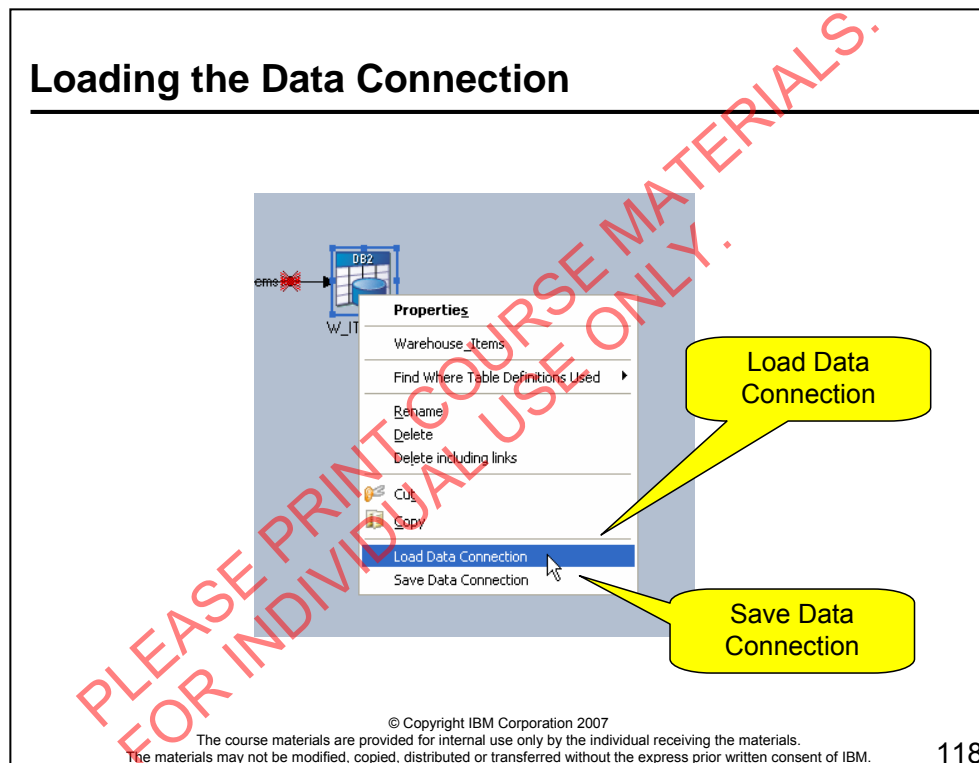
Optionally create a parameter set

Specify parameter values

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

117



Another way of loading the Data Connection is to drag and drop it onto the stage.

Checkpoint

1. Which of the following types stages provide the most functionality in DataStage parallel jobs? ODBC Plug-in, ODBC Connector, ODBC Enterprise?
2. What are three ways of building SQL statements in Connector stages?
3. Which of the following statements can be specified in Connector stages? Select, Insert, Update, Upsert, Create Table.
4. What are two ways of loading Data Connection metadata into a database stage?

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

119

Notes:

Write down your answers here:

1.

2.

Checkpoint solutions

1. Connector stages.
2. Manually. Using SQL Builder. Have the Connector stage generate the SQL.
3. All of them.
4. Click the right mouse button over the stage and click "Load Data Connection." Drag the Data Connection from the Repository and drop it on the stage.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

120

Unit summary

Having completed this unit, you should be able to:

- Import Table Definitions for relational tables
- Create Data Connections
- Use Connector stages in a job
- Use SQL Builder to define SQL Select statements
- Use SQL Builder to define SQL Insert and Update statements
- Use the DB2 Enterprise stage

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

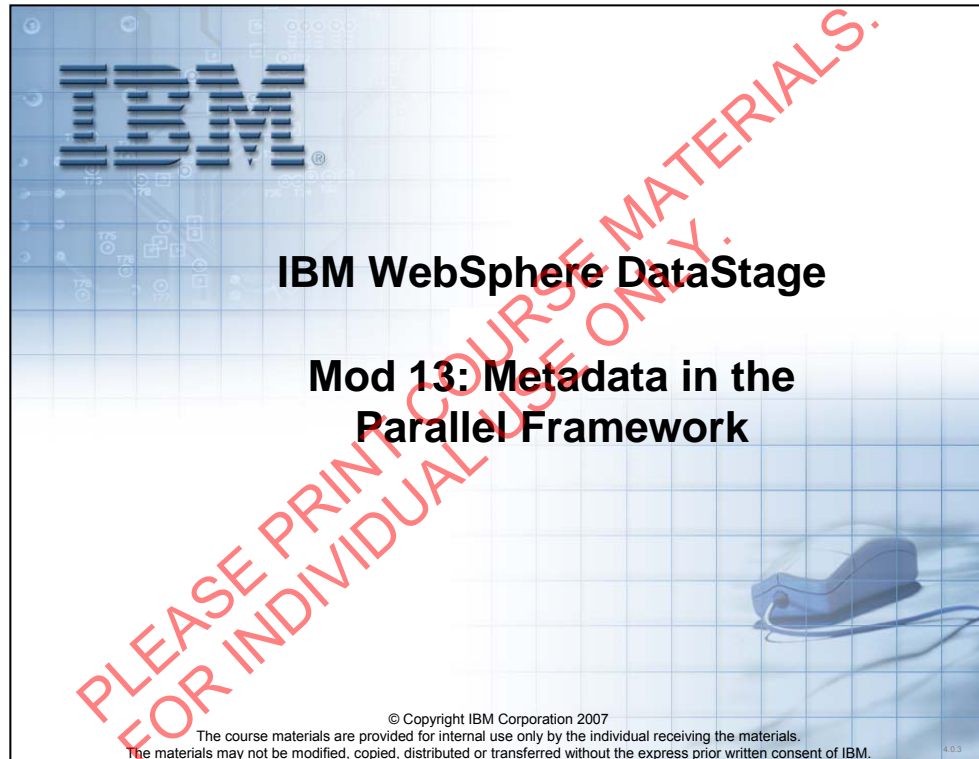
121

Notes:

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

122



Unit objectives

After completing this unit, you should be able to:

- Explain schemas
- Create schemas
- Explain Runtime Column Propagation (RCP)
- Turn RCP on and off
- Build a job that reads data from a sequential file using a schema
- Build a shared container

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

124

Notes:

Schema

- Alternative way to specifying column definitions and record formats
 - Similar to a Table Definition
- Written in a plain text file
- Can be imported as a Table Definition
- Can be created from a Table Definition
- Can be used in place of a Table Definition in a Sequential file stage
 - Requires RCP
 - Schema file path can be parameterized
 - Enables a single job to process files with different column definitions

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.

The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

125

The format of each line describing a column is:

column_name:[nullability]datatype;

column_name. This is the name that identifies the column. Names must start with a letter or an underscore (_), and can contain only alphanumeric or underscore characters. The name is not case sensitive. The name can be of any length.

nullability. You can optionally specify whether a column is allowed to contain a null value, or whether this would be viewed as invalid. If the column can be null, insert the word 'nullable'. By default columns are not nullable.

You can also include 'nullable' at record level to specify that all columns are nullable, then override the setting for individual columns by specifying 'not nullable'. For example:

```
record nullable (
    name:not nullable string[255];
    value1:int32;
    date:date)
```

datatype. This is the data type of the column.

Creating a Schema

- Using a text editor
 - Follow correct syntax for definitions
 - Not recommended
- Import from an existing data set or file set
 - On DataStage Manager import > Table Definitions > Orchestrate Schema Definitions
 - Select checkbox for a file with .fs or .ds
- Import from a database table
- Create from a Table Definition
 - Click Parallel on Layout tab

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

126

Another good way of capturing a schema is to set \$OSH_PRINT_SCHEMAS and copy entries from the DataStage Director log.

Importing a Schema

Import Orchestrate Schema

Import location
Select the location of the file or database table containing the schema definition you wish to import

Import from:

- ☐ File on Local system
- ☒ File on Server: 169.254.146.250
- ☐ Database table (via orchdoutll)

Select the name of the file to be searched for schema definitions:
/opt/AdvTrain/dm1/interval.schema

☐ File is data set or file set

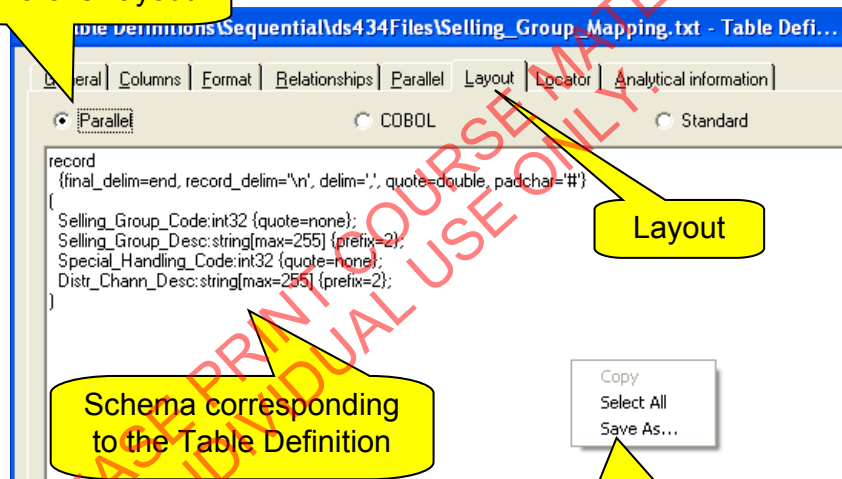
< Back Next >

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

127

Creating a Schema From a Table Definition

Parallel layout

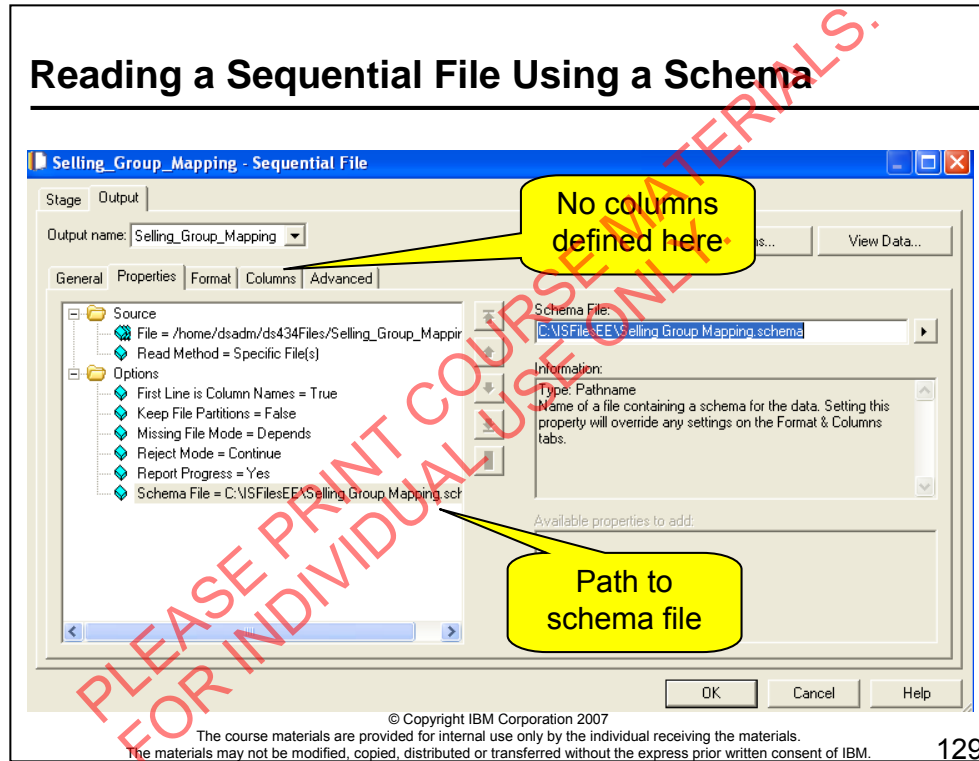


Save schema

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

128



Schemas can only be used when Runtime Column Propagation is turned on in the stage. This is discussed later in this module.

Runtime Column Propagation (RCP)

- When RCP is turned on:
 - Columns of data can flow through a stage without being explicitly defined in the stage
 - Target columns in a stage need not have any columns explicitly mapped to them
 - No column mapping enforcement at design time
 - Input columns are mapped to unmapped columns by name
- How implicit columns get into a job
 - Read a file using a schema in a Sequential File stage
 - Read a database table using "Select *"
 - Explicitly define as an output column in a stage earlier in the flow
- Benefits of RCP
 - Job flexibility
 - Job can process input with different layouts
 - Ability to create reusable components in shared containers
 - Component logic can apply to a single named column
 - All other columns flow through untouched

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

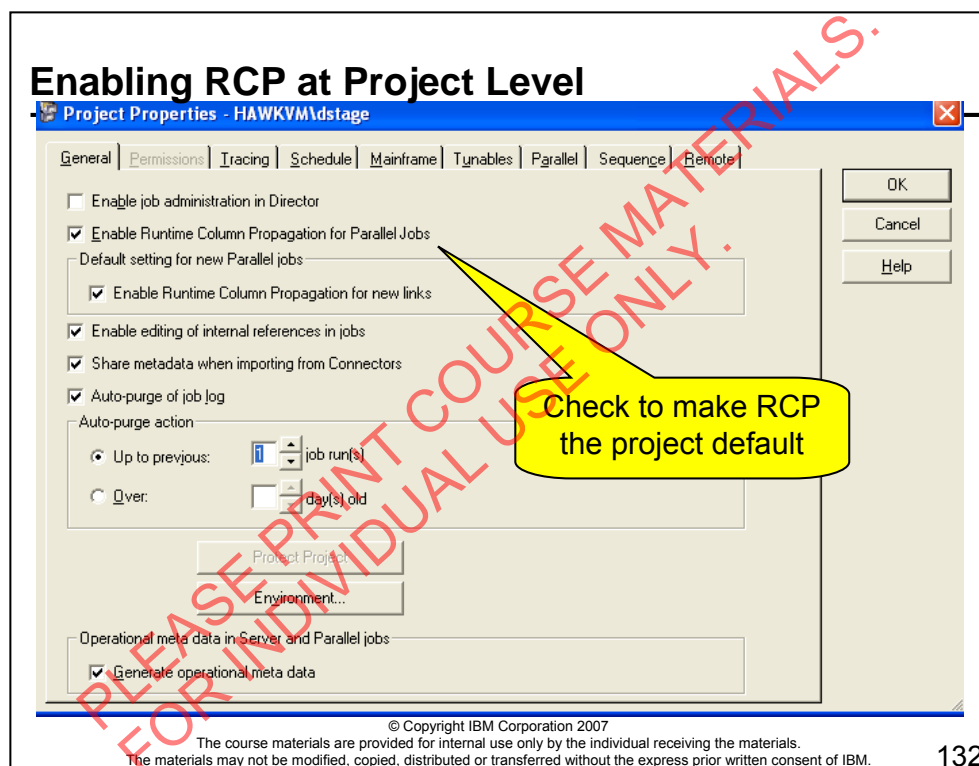
Enabling Runtime Column Propagation (RCP)

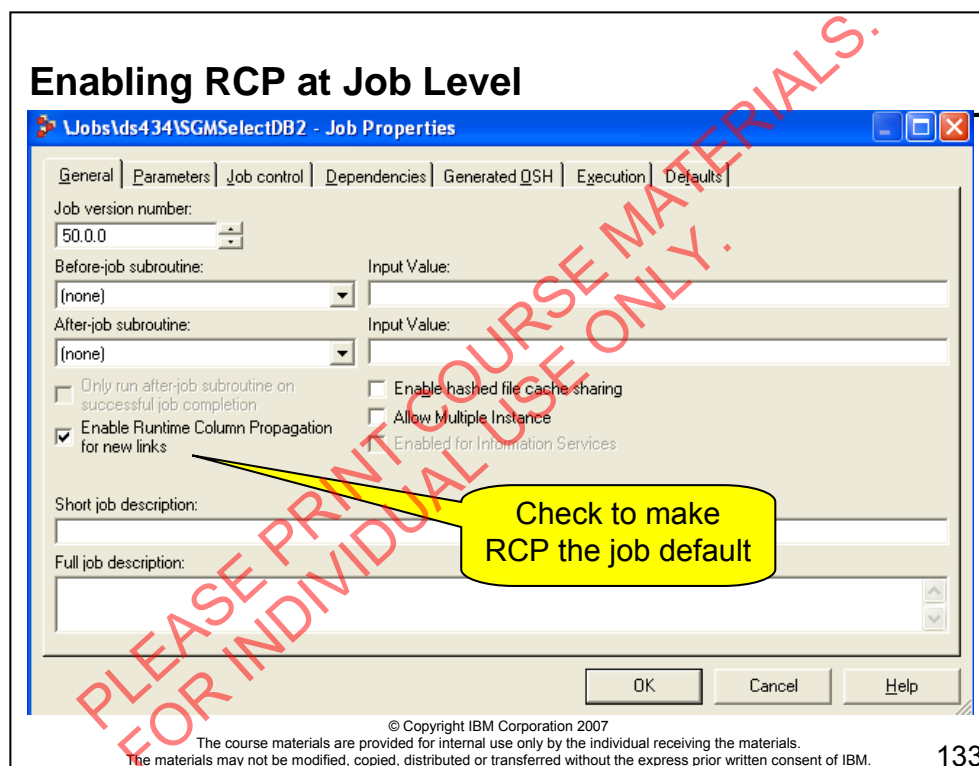
- Project level
 - DataStage Administrator Parallel tab
- Job level
 - Job properties General tab
- Stage level
 - Link Output Column tab
- Settings at a lower level override settings at a higher level
 - E.g., disable at the project level, but enable for a given job
 - E.g., enable at the job level, but disable a given stage

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

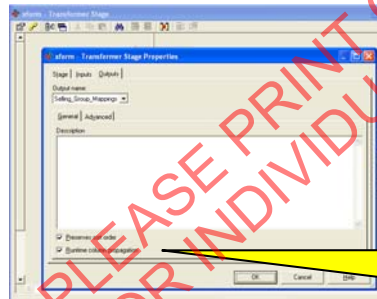
131



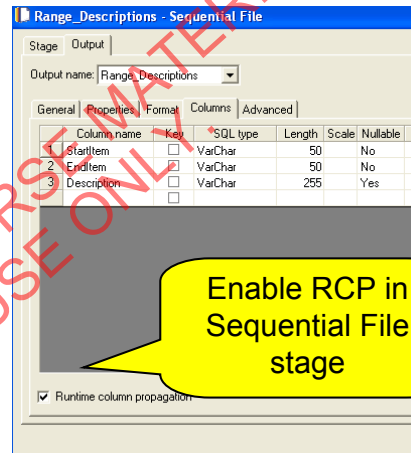


Enabling RCP at Stage Level

- Sequential File stage
 - Output Columns tab
- Transformer
 - Open Stage Properties
 - Stage Properties Output tab



Check to enable RCP in Transformer



Enable RCP in Sequential File stage

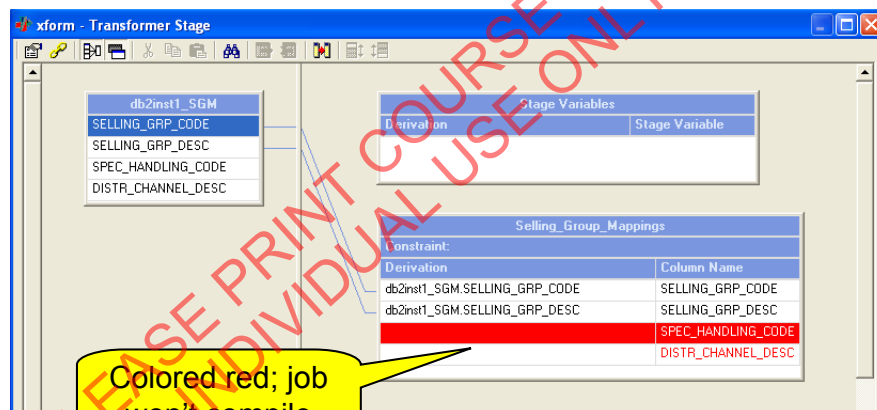
© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

134

When RCP is Disabled

- DataStage Designer enforces Stage Input to Output column mappings.



© Copyright IBM Corporation 2007

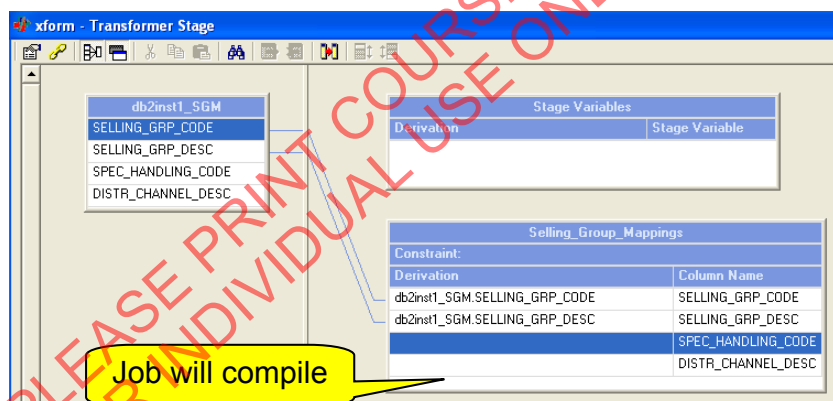
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

135

Modify operators can add or change columns in a data flow.

When RCP is Enabled

- DataStage does not enforce mapping rules
- Runtime error if no incoming columns match unmapped target column names



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

136

Shared Containers

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

137

Shared Containers

- Encapsulate job design components into a stored container
- Provide reusable job design components
- Example
 - Apply stored Transformer business logic

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

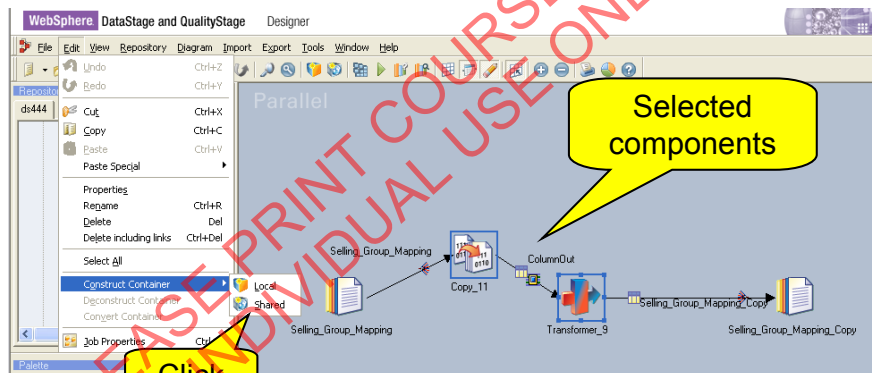
© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

138

Creating a Shared Container

- Select stages from an existing job
- Click Edit>Construct Container>Shared

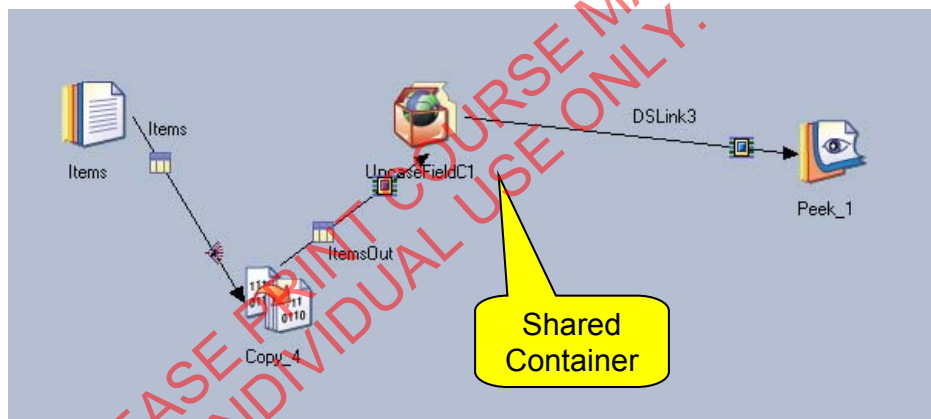


© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

139

Using a Shared Container in a Job



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

140

Mapping Input / Output Links to the Container

UpcaseFieldC1 - Shared Container Stage

Stage **Inputs** Outputs

Input name:
ItemsOut

General Columns Advanced

Link mapping
Map to container link:
Selling_Group_Mapping

Description

Select container link to map input link to

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

141

Checkpoint

1. What are two benefits of RCP?
2. What can you use to encapsulate stages and links in a job to make them reusable?

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

142

Notes:

Write down your answers here:

- 1.
- 2.

Checkpoint solutions

1. Job flexibility. Ability to create reusable components.
2. Shared containers

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

143

Unit summary

Having completed this unit, you should be able to:

- Explain schemas
- Create schemas
- Explain Runtime Column Propagation (RCP)
- Turn RCP on and off
- Build a job that reads data from a sequential file using a schema
- Build a shared container

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

144

Notes:



Unit objectives

After completing this unit, you should be able to:

- Use the DataStage Job Sequencer to build a job that controls a sequence of jobs
- Use Sequencer links and stages to control the sequence a set of jobs run in
- Use Sequencer triggers and stages to control the conditions under which jobs run
- Pass information in job parameters from the master controlling job to the controlled jobs
- Define user variables
- Enable restart
- Handle errors and exceptions

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.

The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

146

Notes:

What is a Job Sequence?

- A master controlling job that controls the execution of a set of subordinate jobs
- Passes values to the subordinate job parameters
- Controls the order of execution (links)
- Specifies conditions under which the subordinate jobs get executed (triggers)
- Specifies complex flow of control
 - Loops
 - All / Some
 - Wait for file
- Perform system activities
 - Email
 - Execute system commands and executables
- Can include Restart checkpoints

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

147

Basics for Creating a New Job Sequence

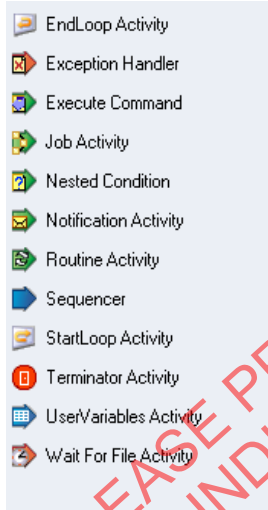
- Open a new job sequence
 - Specify whether its restartable
- Add stages
 - Stages to execute jobs
 - Stages to execute system commands and executables
 - Special purpose stages
- Add links
 - Specify the order in which jobs are to be executed
- Specify triggers
 - Triggers specify the condition under which control passes across a link
- Specify error handling
- Enable / Disable restart checkpoints

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

148

Job Sequencer Stages



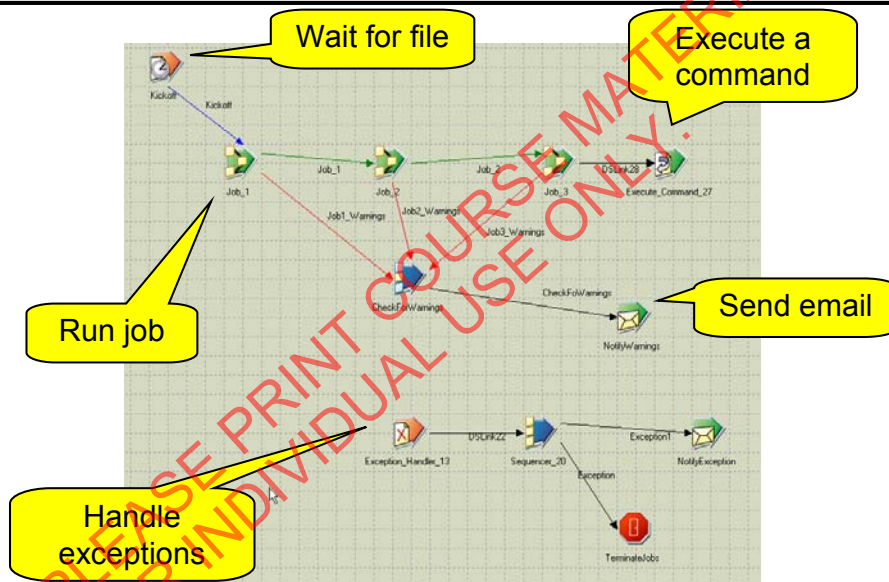
- Run stages
 - Job Activity: Run a job
 - Execute Command: Run a system command
 - Notification Activity: Send an email
- Flow control stages
 - Sequencer: Go if All / Some
 - Wait for File: Go when file exists / doesn't exist
 - StartLoop / EndLoop
 - Nested Condition: Go if condition satisfied
- Error handling
 - Exception Handler
 - Terminator
- Variables
 - User Variables

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

149

Example



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

150

Sequence Properties

General | Parameters | Job control | Dependencies

Job version number:
50.0.0

☐ Allow Multiple Instance

Compilation options

- ☒ Add checkpoints so sequence is restartable on failure
- ☒ Log warnings after activities that finish with status other than OK
- ☒ Automatically handle activities that fail
- ☒ Log report messages after each job run

Exception stage to handle aborts

Restart

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

151

Job Activity Stage Properties

The screenshot shows the 'Warehouse_01 - Job Activity' dialog box with three tabs: 'General', 'Job', and 'Triggers'. The 'Job' tab is selected. The 'Job name' dropdown is set to 'Warehouse_01'. The 'Execution action' dropdown is set to 'Reset if required, then run'. The 'Parameters' table lists three parameters: '\$APT_CO...' with value '\$APT_CONFIG_FILE', '\$APT_DU...' with value '\$APT_DUMP_SCORE', and 'FileDate' with value 'FileDate'. The 'Type' is 'Pathname' and the 'Prompt' is 'Configuration file'. Annotations with yellow callouts point to the 'Job name' dropdown (labeled 'Job to be executed'), the 'Execution action' dropdown (labeled 'Execution mode'), and the 'Parameters' table (labeled 'Job parameters to be passed').

Warehouse_01 - Job Activity

General Job Triggers

Job name: Warehouse_01

Execution action: Reset if required, then run Do not checkpoint run.

Parameters

| Name | Value Expression |
|-------------|-------------------|
| \$APT_CO... | \$APT_CONFIG_FILE |
| \$APT_DU... | \$APT_DUMP_SCORE |
| FileDate | FileDate |

Type: Pathname
Prompt: Configuration file

Insert Parameter Value
Clear
Clear All
Set to Default

Job to be executed

Execution mode

Job parameters to be passed

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Job Activity Trigger

The screenshot shows the 'Triggers' tab in the DataStage Job configuration window. It features a table of triggers and a list of trigger types.

| Name | Expression Type | Expression |
|---------------|-------------------------|--|
| Job3_Warnings | Warning - (Conditional) | "Execution finished with warnings" |
| Ok | Custom - (Conditional) | Job_3.\$JobStatus = DSJS.RUNOK Or Job_3.\$JobStatus = DSJS.RUNWARN |

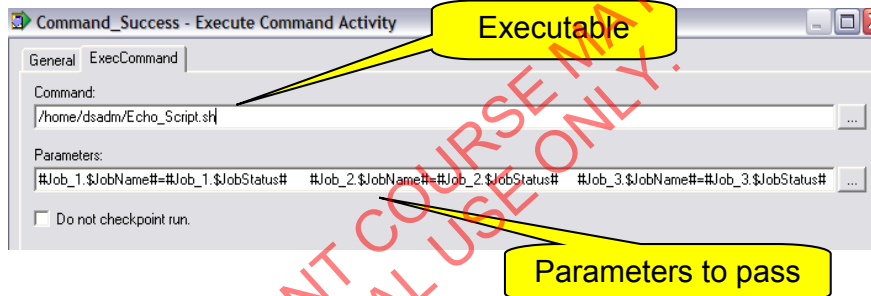
Annotations on the screenshot:

- Output link names:** Points to the 'Job' tab and the 'Job' column in the table.
- List of trigger types:** Points to the dropdown menu showing options like 'Unconditional', 'Otherwise', 'OK - (Conditional)', 'Failed - (Conditional)', 'Warning - (Conditional)', 'UserStatus - (Conditional)', and 'Custom - (Conditional)'.
- Build custom trigger expressions:** Points to the 'Expression' column and the 'Job' column in the table.

© Copyright IBM Corporation 2007
 The course materials are provided for internal use only by the individual receiving the materials.
 The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

153

Execute Command Stage



- Execute system commands, shell scripts, and other executables
- Use e.g. to drop or rename database tables

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

154

Notification Activity Stage

General | Notification

SMTP Mail server name:
NA-MSG-01

Senders email address:
jwilliams23@us.ibm.com

Recipients email address:
jwilliams23@us.ibm.com

Email subject:
Jobs executed successfully: #Job_1.\$JobName#, #Job_2.\$JobName#, #Job_3.\$JobName#

Attachments
/home/dsadm/ds324Files/Customers.txt

Email body:
All jobs ran successfully

☒ Include job status in email
☐ Do not checkpoint run.

Include job status info in email body

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

155

User Variables Stage

User Variables stage

Variable

Expression defining the value for the variable

| Name | Expression |
|------------------|--------------------------------------|
| varMessagePrefix | DSJobName : "" : DSJobStartDate : "" |

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

Referencing the User Variable

seqJob1 - Job Activity

General Job Triggers

Job name:
seqJob1

Execution action:
Reset if required, then run ☐ Do not checkpoint run.

Parameters

| Name | Value Expression |
|------------------|---------------------------|
| NumRecs | RecCount1 |
| \$APT_DUMP_SCORE | \$APT_DUMP_SCORE |
| PeekHeading | UserVars.varMessagePrefix |

Type: String
Prompt: NumRecs

Insert Parameter
Clear
Clear All
Set to Default
All to Default

Variable

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

157

Flow of Control Stages

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

158

Wait for File Stage

Kickoff - Wait for file Activity

General | **Wait For File** | Triggers

Filename: /home/dsadm/StartRun

☐ Wait for file to appear

☒ Wait for file to disappear

Timeout length (hh:mm:ss): 00:00:00

☒ Do not timeout

☐ Do not checkpoint run.

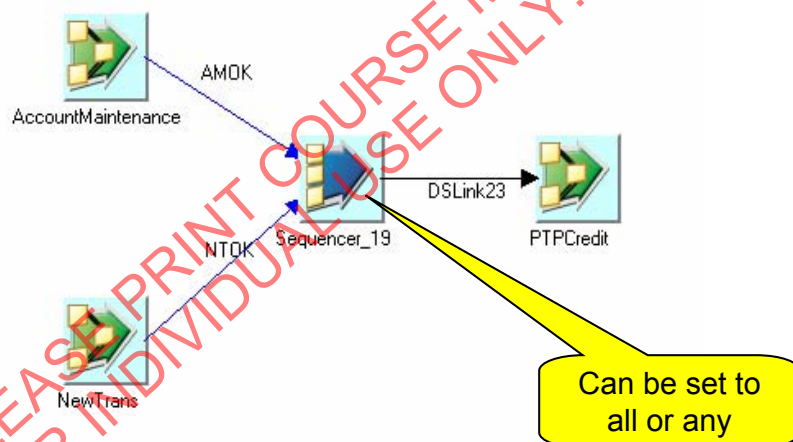
© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

159

Sequencer Stage

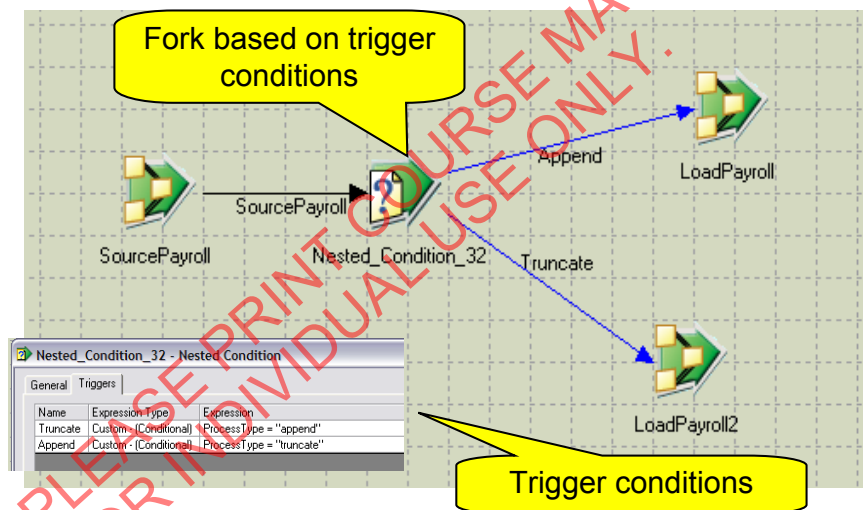
- Sequence multiple jobs using the Sequence stage



© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

160

Nested Condition Stage

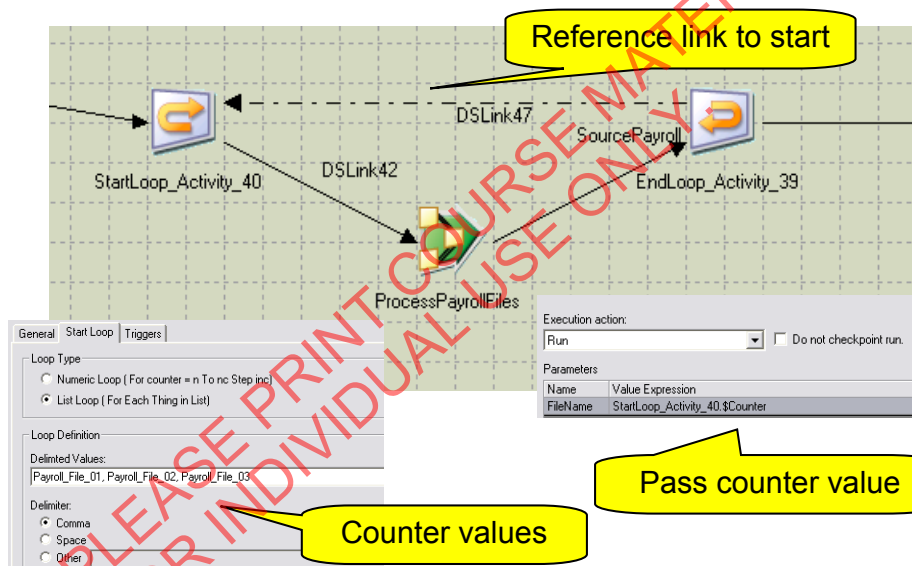


© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

161

Loop Stages



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

162

Error Handling

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

163

Handling Activities that Fail

General | Parameters | Job control | Dependencies

Category: ds434_New Job version num: 50.0.0

☐ Allow Multiple Instance

Compilation options

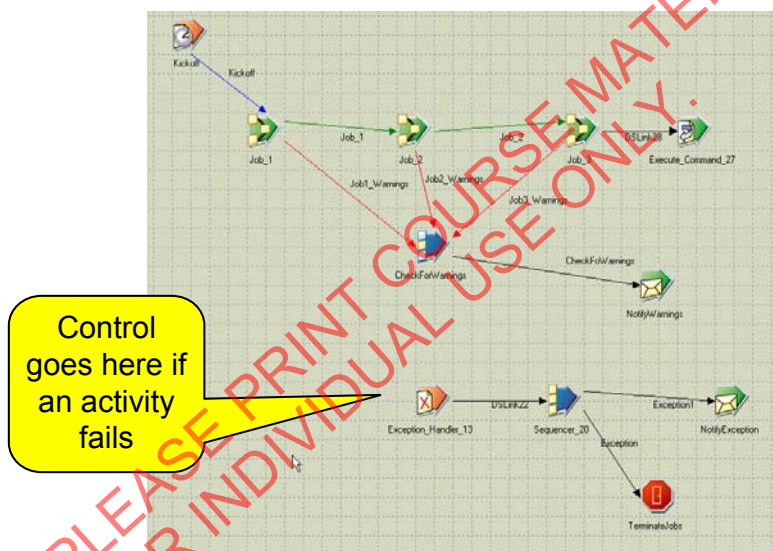
- ☒ Add checkpoints so sequence is restartable on failure
- ☒ Log warnings after activities that finish with status other than OK
- ☒ Automatically handle activities that fail
- ☒ Log report messages after each job run

Pass control to
Exception stage when
an activity fails

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

164

Exception Handler Stage



© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

165

Restart

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

166

Enable Restart

General | Parameters | Job control | Dependencies

Category: ds434_New Job version num: 50.0.0

☐ Allow Multiple Instance

Compilation options

☒ Add checkpoints so sequence is restartable on failure

☒ Log warnings after activities that finish with status other than OK

☒ Automatically handle activities that fail

☒ Log report messages after each job run

Enable checkpoints to be added

© Copyright IBM Corporation 2007
 The course materials are provided for internal use only by the individual receiving the materials.
 The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

167

If a Sequence fails, and it is set to "Add check points so sequence is restartable on failure", then when the Sequence is re-run, activities that completed successfully in the prior run are skipped over (unless the "Do not checkpoint run" option was set for an activity).

Disable Checkpoint at a Stage

ProcessPayrollFiles - Job Activity

General Job Triggers

Job name:
SequencesJob_A

Execution action:
Run ☐ Do not checkpoint run

Parameters

| Name | Value Expression |
|----------|---------------------------------|
| FileName | StartLoop_Activity_40.\$Counter |

Type: String
Prompt: FileName

Insert Parameter Value
Clear
Clear All
Set to Default
All to Default

Don't checkpoint this activity

© Copyright IBM Corporation 2007
The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

168

Checkpoint

1. Which stage is used to run jobs in a job sequence?
2. Does the Exception Handler stage support an input link?

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

169

Checkpoint solutions

1. Job Activity stage
2. No, control is automatically passed to the stage when an exception occurs.

PLEASE PRINT COURSE MATERIALS.
FOR INDIVIDUAL USE ONLY.

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

170

Unit summary

Having completed this unit, you should be able to:

- Use the DataStage Job Sequencer to build a job that controls a sequence of jobs
- Use Sequencer links and stages to control the sequence a set of jobs run in
- Use Sequencer triggers and stages to control the conditions under which jobs run
- Pass information in job parameters from the master controlling job to the controlled jobs
- Define user variables
- Enable restart
- Handle errors and exceptions

© Copyright IBM Corporation 2007

The course materials are provided for internal use only by the individual receiving the materials.
The materials may not be modified, copied, distributed or transferred without the express prior written consent of IBM.

171

Notes: