

Business Intelligence Solution Providers Specialized in creating talent resource pool

## IBM Data Stage Lab Guide#15 Sort Value and Range Map

## **Description:**

BISP is committed to provide BEST learning material to the beginners and advance learners. In the same series, we have prepared a list of beginner's guide and FAQs for IBM Data Stage. We have built complete financial Data Model and various data transformation techniques. Download many such learning documents, student guide, Lab Guide and Hands-on practice materials. This document guides you through Sort Value and Range Map. Join our professional training to learn from Experts.

History:

Version 0.1 0.1 Description Change Initial Draft Review#1 Author Varun Khare Amit Sharma Publish Date 12th Aug 2012 18<sup>th</sup> Aug 2012

www.bispsolutions.com

www.bisptrainings.com 1

## **Introduction Range Map**

The Write Range Map stage takes an input data set produced by sorting a file and writes it to a file in a form usable by the range partitioning method, basically this is debugging stage. The Write Range Map stage is a Development/Debug stage. It allows you to write data to a range map.

The Write Range Map stage takes an input flat file produced by sorting a flat file and writes it to a file in a form usable by the range partitioning method. The range partitioning method uses the sorted flat file to determine partition boundaries. The stage can have a single input link. It can only run in sequential mode.

A typical use for the Write Range Map stage would be in a job which used the flat file to sort stage, the Sort stage to sort it and the Write Range Map stage to write the range map which can then be used with the range partitioning method to write the original flat file to a file set

Devel	opment/Debug	
🔀 Column Generator	🛃 Head	
💿 Peek	👸 Row Generator	
🤹 Sample	🚰 Tail	
脉 Write Range Map		

## Steps to Load Data Into File Set DATASTAGE

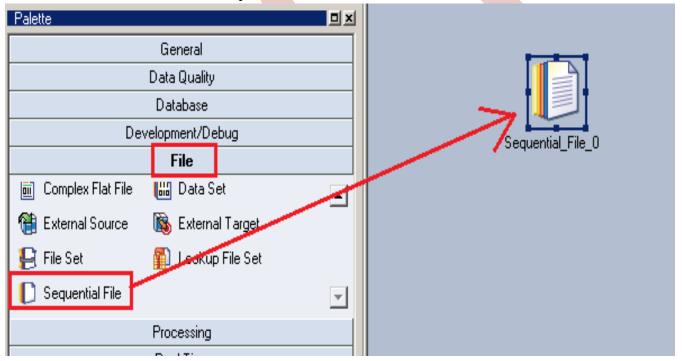
Step-1 :- In Windows Click Designer Client of Datastage



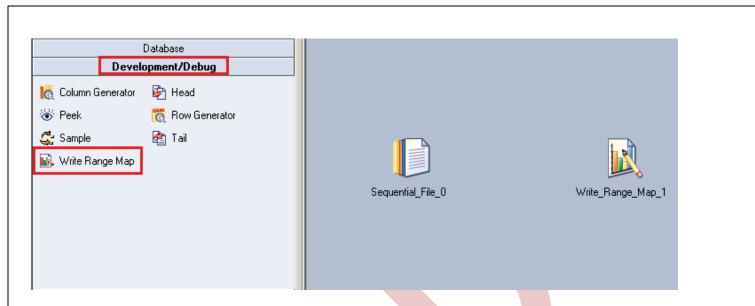
Step-2:- New Window opens, then click Jobs ,then click Parallel Job because we are using parallel jobs.

💸 WebSphere DataStage Designer -	
WebSphere DataStage and Qua	
Eile View Repository Import Export	
	≦ / A_ (≠   ≫ @   ♥   ®   ●   ₩ ₩   # <i>™ /</i>   ₿   ⊕ ⊖   ≫ @ @
Repository 🔲 🗵	X New
ANALYZERF Open quick find ANALYZERPROJECT Data Elements IMS Databases (C IMS Viewsets (PS Jobs Machine Profiles	Most Recent Date Quality Houtines Houtines Stage Types Assistants
Palette  Favorites Favorites This group is initially empty and is intended to contain shortcuts to commonly used repository items.	Jobs
Shortcuts can be created by dragging items from the Repository Tree or by using the Customize option from the menu.	OK Cancel Help

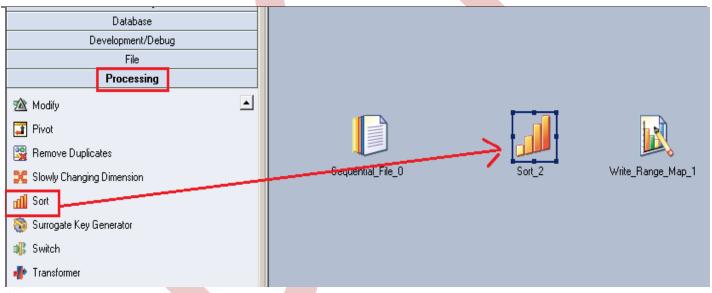
**Step-3:-** First go on File Palette then scroll the cursor and choose Sequential File and drag it to Parallel pane and rename it. This file is used for input data.



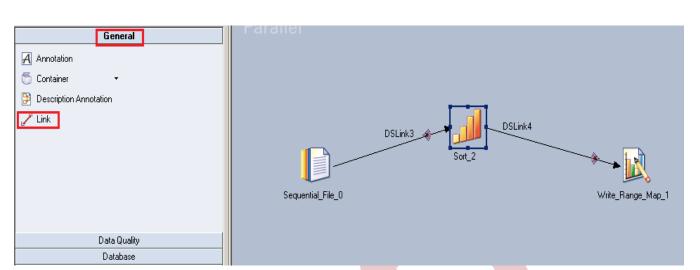
**Step-4 :-** Choose write range map from development/debug tab this is used for finding range of input file.



**Step-5:-** Choose an sort stage from processing palette. This sort process is used when we want to sort any number of values according to ascending or descending order than we apply this on a particular column and output is based on key column in which we apply sorting we get result.



**Step-6:-** Now make an connection with these stages that's why we have to use this link stage for connecting these stages.



Step-7:- Double click on Input file and browse an input file that can be .txt, .csv any test file.

Paralle	Sequential_File_0 - Sequential File	
T Grune	Stage Output	
	Output name: DSLink6	Colu <u>m</u> ns <u>V</u> iew Data
	<u>G</u> eneral <u>Properties</u> Format <u>C</u> olumns Advanced	
equential_File_0	Source         Image: Source         Image	File:       Switch to multiline editor         D:\Creditcard\SRC_ACCOUNT_OFFICER       Switch to multiline editor         Information:       Insert job parameter         Type: Pathname       Browse for file         Name of a file that data will be read from.       Image: Comparison of the second se
•		OK Cancel Help

**Step-8:-** Go to Format tab and Remove double Quote because we don't need them then Click on OK. Note:- If you use date as a input than specify here which type of format you used

Sequential_File_7 - Sequential File Stage Output	
Output name: DSLink11	Colu <u>m</u> ns ⊻iew Data
Properties:	[No property selected]
Quote = double Type defaults General C String C Decimal Decimal Time Time Timestamp	Available properties to add:    Available properties to add:
	Load Defaults >
	OK Cancel <u>H</u> elp

Step-9:- Here I'm using this date format so we specify here than we able to load the date from flat file.

🚺 Sequential_File_7 - Sequential File	
Stage Output	
Output name: DSLink11	Colu <u>m</u> ns <u>V</u> iew Data
Properties: Delimiter = comma Type defaults General Decimal Numeric Format string = %dd-%mm-%yyy Time Timestamp	Format string         %dd-%mm-%yyyy         The string format of a date. By default, the format of the date string is %yyyy-%mm-%dd.         Available properties to add:
	Load Defaults >
	OK Cancel <u>H</u> elp

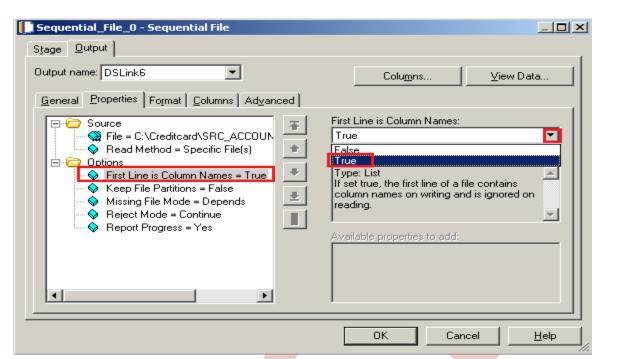
**Step-10:-** Go to Columns tab and enter same column names as declared in Input file and put data type, length then click on View Data. This will show your Input Data.

iene	eral <u>P</u> roperties Fo <u>r</u> mat <u>Column</u> s	_ ∏Ad⊻	anced	Colu <u>m</u> r			<u>(</u> iew Data.	
	Column name	Key	SQL type	Extended	Length	Scale	Nullable	cri
1	ACCOUNT_OFFICER_CD		Integer		10		No	
2	CREATED_BY		VarChar	Unicode	255		No	
3	CREATION_DATE		VarChar	Unicode	255		No	
4	DEFINITION_LANGUAGE		VarChar	Unicode	255		No	
5	ENABLED_FLAG		VarChar	Unicode	255		No	
6	LAST_MODIFIED_BY		VarChar	Unicode	255		No	
7	LAST_MODIFIED_DATE		VarChar	Unicode	255		No	
8	LEAF_ONLY_FLAG		VarChar	Unicode	255		No	
9	ACCOUNT_OFFICER_DISPLAY		Integer		10		No	
•				- 1				
•								Ľ

**Step-11:-** After Clicking on View Data if your data shows like this than our Data is valid and if some error comes then we have to rectify them and check our column names.

ACCOUNT_OF	FICER_CD	CREATED	BY	CREATION	DATE		<b>_</b>						
52223		-		11-JUN-10	07.19	.36.0000			Colum	ns (	V	iew Data	
52535				11-JUN-10	07.19	.36.0000							
53110				11-JUN-10	07.19	.36.0000		anced					
53458				11-JUN-10	07.19	9.36.0000		SQL type	Extended	Longth	Coole	Nullable	lori
53961				11-JUN-10	07.19	9.36.0000		Integer	Extended	Length 10	SCale	No	Cil
54254				11-JUN-10	07.19	9.36.0000		VarChar	Unicode	255		No	-
54608				11-JUN-10	07.19	.36.0000		VarChar	Unicode	255		No	-
54706				11-JUN-10	07.19	.36.0000		VarChar	Unicode	255		No	-
55304				11-JUN-10	07.19	.36.0000		VarChar	Unicode	255		No	
55362				11-JUN-10	07.19	.36.0000		VarChar	Unicode	255		No	
55675				11-JUN-10	07.19	.36.0000	-	VarChar	Unicode	255		No	
56174				11-JUN-10	07.19	.36.0000		VarChar	Unicode	255		No	
56339				11-JUN-10	07.19	.36.0000		Integer		10		No	
56619				11-JUN-10	07.19	.36.0000							
57060				11-JUN-10	07.19	.36.0000			-	_			•
57619				11-JUN-10	07.19	.36.0000	L., .	L					-
58010				11-3170-10	07 19	36 0000	-			Save		Load.	

**Step-12:-** After choosing file, select 'True' from First line in column Names dropdown for removing first line from Input file.





ACCOUNT\_OFFICER\_CD, CREATED\_BY, CREATION\_DATE, DEFINITION\_LANGUAGE, ENABLED\_FLAG, LAST\_MODIFIED\_BY, LAST\_MODIFIED\_DATE, LEAF\_ONLY\_FLAG, ACCOUNT\_OFFICER\_DISPLAY\_CD,

HCCOONT_OFFIC	CEK_CD,CKEA	ATED_DT,CREATION	DATE, DEFINITION	N_LANGUMGE,	,ENADLEU_FLAG,LA	SI_MODIFIED_DI,L
52223,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 52223.36,
52535,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 52534.89,
53110,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 53110.24,
53458,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 53457.6.,
53961.	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 53960.7.
54254,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 54253.88,
54608,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 54607.99,
54706,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 54705.91,
55304,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 55303.81,
55362,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 55362.01,
55675,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 55674.97,
56174,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 56173.79,
56339,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 56338.61,
56619,	11-JUN-10	07.19.36.000000	PM, US, Y,		07.19.36.000000	
57060,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 57059.61,
57619,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 57619.17,
58010,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 58010.18,
58235,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 58235.35,
58616,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 58616.48,
58900,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 58899.61,
59154,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 59153.7,,
59567,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 59566.99,
60254,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 60254.29,
60740,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 60740.15,
60951,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 60951.47,
61134,	11-JUN-10	07.19.36.000000	PM, US, Y,	11-JUN-10	07.19.36.000000	PM, Y, 61133.62,

**Step-14:-** Now double click on sort stage and see here there are some field showing like key field here we have to specify in which column we want to apply sorting so we select here and we also specify in which order we want our target data.

<u>Vindow H</u> elp Parallel	Stage     Input     Output       Stage name:     Sort_2
USLink6	General       Properties       Adyanced       NLS Locale         Image: Soting Keys       Soting Keys       Image: Soting Keys       Image: Soting Keys         Image: Soti Key Mode = Soting       Image: Soti Order = Ascending       Image: Soti Order = Ascending       Image: Soti Order = Ascending         Image: Soti Order = Ascending       Image: Soti Order = Ascending       Image: Soti Order = Ascending       Image: Soti Order = Ascending         Image: Soti Order = Ascending       Image: Soti Order = Ascending       Image: Soti Order = Ascending       Image: Soti Order = Ascending         Image: Soti Order = Ascending       Image: Soti Order       Image: Soti Order       Image
	OK Cancel <u>H</u> elp

**Step-15:-** Now click to input and choose hash key partition because these are key based stage so we have to choose hash and check on perform sort for sorting.

· · ·	3
📶 Sort_2 - Sort	
Stage Input	
Input name: DSLink6	Colu <u>m</u> ns
General Partitioning Columns Advanced	
Partitioning / Collecting Partition type: Hash	Sorting
Available:	Selected:
ACCOUNT_OFFICER_DISPLAY_CD CREATED_BY CREATION_DATE DEFINITION_LANGUAGE ENABLED_FLAG LAST_MODIFIED_E LAST_MODIFIED_DATE	Kev Usage
LEAF_ONLY_FLAG	
	OK Cancel <u>H</u> elp

Step-16:- Click on Input and see all columns are available or not

	name: DSLink6				Ca	olu <u>m</u> ns.		
ene	eral P <u>a</u> rtitioning					<b>a</b> 1		
	Column name	Key	SQL type	Extended			Nullable	Description
1	ACCOUNT_OFFI(		Integer		10		No	
2	CREATED_BY		VarChar	Unicode	255		No	
3	CREATION_DAT		VarChar	Unicode	255		No	
4	DEFINITION_LAN		VarChar	Unicode	255		No	
5	ENABLED_FLAG		VarChar	Unicode	255		No	
6	LAST_MODIFIED		VarChar	Unicode	255		No	
7	LAST_MODIFIED		VarChar	Unicode	255		No	
В	LEAF_ONLY_FLA		VarChar	Unicode	255		No	
9	ACCOUNT_OFFI(		Integer		10		No	
L								
					1			
								•
							Save	Load

**Step-17** Now go to output tab than select all links and drag them into Output Link for sending data which satisfies filter condition into output links..

Sort_2 - Sort	
Output name: DSLink7	Colu <u>m</u> ns
General Mapping Columns Advanced	
<u> </u>	Drag all the columns here
Columns	DSLink7
Expression Column N	Derivation Col
DSLink6.ACCOUNT_CACCOUNT	DSLink6.ACCOUNT_OFFICER_ACC
DSLink6.CREATED_B CREATED	DSLink6.CREATED_BY CRE
DSLink6.CREATION_I CREATION	DSLink6.CREATION_DATE CRE
DSLink6.DEFINITION_ DEFINITIC	DSLink6.DEFINITION_LANGU/ DEF
DSLink6.ENABLED_FI ENABLED	DSLink6.ENABLED_FLAG EN/
DSLink6.LAST_MODIF LAST_M0	DSLink6.LAST_MODIFIED_BY_LAS
	Find Auto-Match
	OK Cancel Help

**Step-18:-** Now see there are multiple options available there now select key what we use in sort stage same key use and also declare output file set file. A typical use for the Write Range Map stage would be in a job which used the flat file to sort stage, the Sort stage to sort it and the Write Range Map stage to write the range map which can then be used with the range partitioning method to write the original flat file to a file set.

Parallel	Write_Range_Map_1 - Write Range Map           Stage         Input	
	Input name: DSLink4 Columns  General Properties Partitioning Columns Advanced  File Update Mode = Create  Key = ?  Key = ?  Information:	
Sequ	Available properties to add:	× •
E_Range_Map_1	OK Cancel	

**Step-19:-** Now select an key value as select in sort stage and select range map file for sending data into target file

🙀 Write_Range_Map_1 - Write Range Mag	p		<u> </u>
Stage Input			
Input name: DSLink4		Colu <u>m</u> ns	
<u>G</u> eneral <u>Properties</u> P <u>a</u> rtitioning <u>C</u> olumns	Ad <u>v</u> anced		
Options File Update Mode = Create  Key = ACCOUNT_OFFICER_CD  Sort Order = Ascending  Range Map File = ?		Key: ACCOUNT_OFFICER_CD ACCOUNT_OFFICER_DISPLAY_CD CREATED_BY CREATION_DATE DEFINITION_LANGUAGE ENABLED_FLAG LAST_MODIFIED_BY LAST_MODIFIED_DATE LEAF_ONLY_FLAG Case Sensitive Nulls Position Sort as EBCDIC	
		OK Cancel	<u>H</u> elp

**Step-20:-** As I provide path for target file we have to specify target file name than we able to load data into target file.

Write_Range_Map_1 - Write Range Map Stage Input Input name: DSLink4	
General Properties Partitioning Columns Advance © Options File Update Mode = Create © Soft Order = Ascending © Range Map File = ? •	Range Map File: C:\real\first.txt
	OK Cancel <u>H</u> elp

**Step-21:-** See here there are two option available here create or overwrite that means if create new file than we choose create otherwise if we have existing file and we want to overload than we use this file.

Input name: DSLink4	Columns  File Update Mode:  Create  Create  Verwrite  Type: List  Create will cause an error if the file already exists; Overwrite will an overwrite existing file.  Available properties to add:

Step-22:- Click on column tab and see all the columns are showing or not.

	ame: DSLink4	-	[	C	Colu <u>m</u> ns	1			
ie	ral Properties P <u>a</u> rtitioning	Colun	ns Ad <u>v</u> anced	ı —					
	Column name	Key	SQL type	Extended	Length	Scale	Nullable	·escriptio	
1	ACCOUNT_OFFICER_CD		Integer		10		No		
			VarChar	Unicode	255		No		
_	CREATION_DATE		VarChar	Unicode	255		No		
	DEFINITION_LANGUAGE		VarChar	Unicode	255		No		
	ENABLED_FLAG		VarChar	Unicode	255		No		
	LAST_MODIFIED_BY		VarChar	Unicode	255		No		
	LAST_MODIFIED_DATE		VarChar	Unicode	255		No		
	LEAF_ONLY_FLAG		VarChar	Unicode	255		No		
	ACCOUNT_OFFICER_DIS		Integer		10		No		
1									
ĺ								•	
)	🖀 🕨 🕼 🕼 🖩	•		ی 🥥 😂					
	اماله								
arallel									
				<b>.</b> ]	201114				
		DSL	ak2 . 🔭		SLink4				
		DSLi	nk3		OSLink4			_	
		DSLi	<u> </u>		DSLink4			&	
		DSLi	<u> </u>	ort_2	OSLink4				
		DSLi	<u> </u>		DSLink4				
		DSLi	<u> </u>		DSLink4				ange_Map_1

**Step-24:-** If this is shown that means your compilation is done otherwise it shows error.Now click on run button or (CTRL+F5).

🌂 Compile Job - load1	X
Compilation Status	
Job successfully compiled with no errors.	<u>S</u> how Error
	More
	<u>C</u> lose
	<u>H</u> elp

**Step-24:-** Now click on Run tab and see here there are some options available here means see here there's are warning so we have select no limit of warning like this. and we can validate before run the job.

8   B   I   I   I   I   I   B   S   C   C   S   S   C   S   C   S   S				
arallel				
🗯 load1 - Job Run Options				
Limits General				
Rows  Rows  Rows:  1000  Rows:	Warnings ● <u>No limit</u> ● <u>A</u> bort job after: Warnings: <u>50</u>			
Bun				

Step-25:- After click on run wait for a while than it shows GREEN line that means your tranformation is successfully done otherwise if it shows RED Line that means not Done and BLUE Line means Under Process.



Step-26:- Now Go to SQL Console and connect with the same login credentials as I mentioned in oracle enterprise stage src\_creditcard/password so you can choose your own username and password. Before loading, remember to check structure of this file is available and then you should be able to load the data. then simple query:

彦 🍛   🙉 🛞	) 🚄   🔠	y 🔒 🛛 🗗	۶ ► 🔲 🐳 🚱 🕹 🕢
>0ccurred	>On date	Туре	Event
2:46:04 PM	3/7/2013	Info	Parallel job initiated
🛛 🌌 2:46:04 PM	3/7/2013	Info	Parallel job default NLS map ASCL_MS1252, default locale OFF
🛛 🌄 2:46:05 PM	3/7/2013	Info	main_program: IBM WebSphere DataStage Enterprise Edition 8.0.1.4458 ()
🛛 🌄 2:46:05 PM	3/7/2013	Info	main_program: orchgeneral: loaded ()
🛛 🌌 2:46:05 PM	3/7/2013	Info	main_program: APT_configuration file: C:/IBM/InformationServer/Server/Confi
🛄 🛄 2:46:05 PM	3/7/2013	Warning	Write_Range_Map_21: When checking operator: A sequential operator cann
🛄 😃 2:46:05 PM	3/7/2013	Warning	Sequential_File_0,0: Delimiter for field "empno" not found; input: <empty>, at</empty>
🛄 😃 2:46:05 PM	3/7/2013	Warning	Sequential_File_0,0: Import warning at record 14.
🛄 😃 2:46:05 PM	3/7/2013	Warning	Sequential_File_0,0: Import unsuccessful at record 14.
🛛 🏧 2:46:05 PM	3/7/2013	Info	Sequential_File_0,0: Import complete; 14 records imported successfully, 1 reje
🛛 🏧 2:46:05 PM	3/7/2013	Info	main_program: Step execution finished with status = OK.
🛛 🏧 2:46:05 PM	3/7/2013	Info	main_program: Startup time, 0:01; production run time, 0:00.
🛛 🏧 2:46:05 PM	3/7/2013	Info	Parallel job reports successful completion
2:46:05 PM	3/7/2013	Control	Finished Job write.
🛛 🚳 4:44:02 PM	3/7/2013	Control	Starting Job write.
🛛 🚳 4:44:03 PM	3/7/2013	Info	Environment variable settings: ()
🛛 🚳 4:44:03 PM	3/7/2013	Info	Parallel job initiated
🛛 🐼 4:44:03 PM	3/7/2013	Info	Parallel job default NLS map ASCL_MS1252, default locale OFF
🛛 🚳 4:44:04 PM	3/7/2013	Info	main_program: IBM WebSphere DataStage Enterprise Edition 8.0.1.4458 ()
🛛 🐼 4:44:04 PM	3/7/2013	Info	main_program: orchgeneral: loaded ()
🛛 🌌 4:44:04 PM	3/7/2013	Info	main_program: APT configuration file: C:/IBM/InformationServer/Server/Confi
🛄 😃 4:44:06 PM	3/7/2013	Warning	Write_Range_Map_21: When checking operator: A sequential operator cann
🛄 🛄 4:44:06 PM	3/7/2013	Warning	Sequential_File_0,0: Delimiter for field "empno" not found; input: <empty>, at</empty>
🖳 🛄 4:44:06 PM	3/7/2013	Warning	Sequential_File_0,0: Import warning at record 14.
🛄 😃 4:44:06 PM	3/7/2013	Warning	Sequential_File_0,0: Import unsuccessful at record 14.
🏼 🌌 4:44:06 PM	3/7/2013	Info	Sequential_File_0,0: Import complete; 14 records imported successfully, 1 reje
🏼 🌌 4:44:06 PM	3/7/2013	Info	main_program: Step execution finished with status = OK.
🏼 🌌 4:44:06 PM	3/7/2013	Info	main_program: Startup time, 0:02; production run time, 0:00.
4:44:06 PM	3/7/2013	Info	Parallel job reports successful completion
4:44:06 PM	3/7/2013	Control	Finished Job write.
Log for job: write			100 entries i