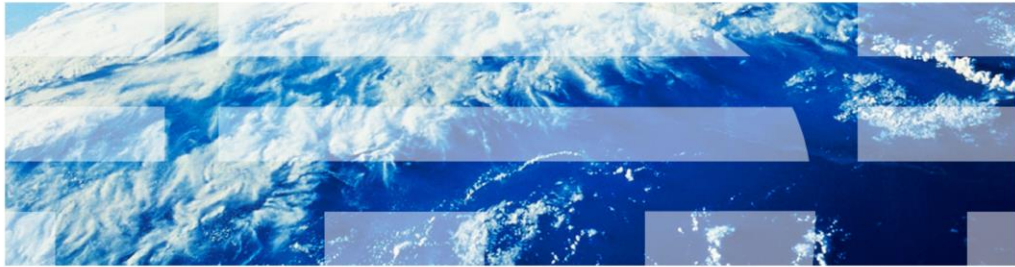


# InfoSphere Information Server DataStage V8.1 – 9.1

Manually copy a dataset from one server to another



© 2013 IBM Corporation

This presentation will discuss how to manually copy a dataset from one server to another.

## Objectives

- Create a simple DataStage® parallel job to create a dataset
- Detail files associated to dataset that are created and locations
- Detail which data files and configuration file need to be copied to destination server
- Detail location of where these files need to be copied on destination server
- Detail configuration changes required on destination server to process this dataset
- Create a simple job to read this dataset on destination server
- Why use this method to copy a dataset?
- Restrictions that apply to above method

The objectives of this presentation are to document the process by which you can copy a dataset created on one DataStage server to another DataStage server. The presentation will go through the steps to create a very simple parallel DataStage job that will create an output dataset file using a two node configuration file on a DataStage server which is referred to as the source server. The presentation will also cover which files are created on the source server and locations of these files after the job has run. Next, it will explain which files need to be copied on the destination server and their locations and what configuration changes that are required. The presentation will also go through steps to create a very simple parallel job to read this copied dataset on the destination. Finally, the presentation will discuss why you may want to use this method to copy a dataset and what restrictions apply to this method.

## Create a simple parallel job to create a dataset

The screenshot displays the configuration for a 'Row\_Generator\_0 - Row Generator' stage in DataStage. The job graph shows 'Row\_Generator\_0' connected to 'Data\_Set\_3' via a 'DSLink2' stage, with a rate of '20 rows/sec' and '10 rows/sec'.

The 'Row\_Generator\_0 - Row Generator' configuration window shows the following settings:

- Stage: Output
- Stage name: Row\_Generator\_0
- General tab selected
- Execution mode: Parallel
- Combinability mode: (Auto)
- Preserve partitioning: Default (Propagate)

The 'Columns' tab shows a single column definition:

Column name	Key	SQL type	Extended	Length	Scale	Nullable
col1	<input type="checkbox"/>	VarChar	<input type="checkbox"/>		10	No

The 'Target' window shows the file path: `/home/dsadm/Federico/ds_file/datasetfile.ds`.

The '\Jobs\create\_ds - Job Properties' window shows the 'Parameters' tab with the following configuration:

Parameter name	Prompt	Type	Default Value	Help Text
\$APT_CONFIG_FILE	Configuration file	Pathname	/opt/IBM/InformationServer/Server/Configurations/test.apt	The Parallel job configuration file.

3

Manually copy a dataset from one server to another

© 2013 IBM Corporation

The first step in this presentation is to create a very simple job using a row generator to write to a dataset file as demonstrated on this slide. For the purpose of this presentation, a single column definition is used and the row generator is set to run in parallel and use a two node configuration file called test.apt. The configuration file details are covered in the later slides. The dataset descriptor file, called datasetfile.ds, is created in the location displayed on this slide.

## Configuration file used

- \$APT\_CONFIG=/opt/IBM/InformationServer/Server/Configurations/test.apt

```
{
  node "node1"
  {
    fastname "gbaix85485"
    pools ""
    resource disk "/tmp/Datasets" {pools ""}
    resource scratchdisk "/tmp/Scratch" {pools ""}
  }
  node "node2"
  {
    fastname "gbaix85485"
    pools ""
    resource disk "/tmp/Datasets" {pools ""}
    resource scratchdisk "/tmp/Scratch" {pools ""}
  }
}
```

The previous job uses the APT\_CONFIG file to create a two node dataset which creates two data files in the /tmp/Datasets directory on a node called gbaix85485, which is referred to as the source server.

## Dataset files created and locations

- Job creates multiple files

```
$ uname -a
AIX gbaix85485 1 7 00C922AC4C00
$ pwd
/tmp/Datasets
$ ls -l
-rw-rw-r-- 1 dsadm  dstage   131072 28 May 14:32
datasetfile.ds.dsadm.gbaix85485.0000.0000.0000.88008e.51a4bfe2.0000.7929b062
-rw-rw-r-- 1 dsadm  dstage   131072 28 May 14:32
datasetfile.ds.dsadm.gbaix85485.0000.0001.0000.88008e.51a4bfe2.0001.3c93059f

$ pwd
/home/dsadm/Federico/ds_file
$ ls -l
-rw-rw-r-- 1 dsadm  dstage    2061 28 May 14:32 datasetfile.ds
```

Once the job has been run, using the two node APT\_CONFIG file demonstrated in the previous slide, the example files displayed on this slide are created in the locations displayed. The dataset raw data files are created in the /tmp/Datasets directory. Notice that two files are created since you used a two node configuration file. The dataset descriptor file called datasetfile.ds is created in the /home/dsadm/Federico/ds\_file directory as defined in the output dataset stage properties.

## Files to be copied to the destination server

- Files to be copied to destination server
  - \$APT\_CONFIG\_FILE  
/opt/IBM/InformationServer/Server/Configurations/test.apt
  - Dataset descriptor file  
/home/dsadm/Federico/ds\_file/datasetfile.ds
  - Dataset raw data files  
/tmp/Datasets/datasetfile.ds.dsadm.gbaix85485.0000.0000.0000.88008e.51a4bfe2.0000.7929b062  
/tmp/Datasets/datasetfile.ds.dsadm.gbaix85485.0000.0001.0000.88008e.51a4bfe2.0001.3c93059f

Copy the files displayed on this slide from the source server, which for the purpose of this presentation is gbaix85485, to the destination server, which, for the purpose of this presentation, is gbaix85487. Copy the test.apt file that was used previously, the dataset descriptor file, datasetfile.ds, and the two dataset raw data files that were created using the two node configuration file to the destination server. The configuration and the dataset descriptor files can be copied to any location on the destination server and the DataStage job design modified to point to these new locations. However, the dataset raw data files need to be copied to the same directory structure on the destination server, hence, these directories will need to be created if they do not already exist. If you want to use existing directories, create soft links to the directories that you want. Ensure that the correct ownership and permissions are set on the files and directories.

## Configuration changes on the destination server (1 of 2)

- Edit copied configuration file to update server name

```
$ uname -a
```

```
AIX gbaix85487 1 7 00C922AC4C00
```

```
{
  node "node1"
  {
    fastname "gbaix85487"
    pools ""
    resource disk "/tmp/Datasets" {pools ""}
    resource scratchdisk "/tmp/Scratch" {pools ""}
  }
  node "node2"
  {
    fastname "gbaix85487"
    pools ""
    resource disk "/tmp/Datasets" {pools ""}
    resource scratchdisk "/tmp/Scratch" {pools ""}
  }
}
```

7

Manually copy a dataset from one server to another

© 2013 IBM Corporation

Once the files have been copied and directories created on the destination server, edit the copied configuration file and update the fastname to reflect the new destination server name, in this example, gbaix85487. As an additional note, it is also possible to use a configuration file that contains the same set or superset of nodes, but not a subset. Keep the nodes the same to make things simpler.

## Configuration changes on the destination server (2 of 2)

- \$APT\_CONFIG\_FILE location on destination server  
/home/dsadm/Federico/APT\_CONFIG\_LOC/test.apt
- Dataset descriptor file location on destination server  
/home/dsadm/Federico/input\_ds\_file/datasetfile.ds

As previously mentioned, the configuration file and the dataset descriptor file can be copied to a different location on the destination server and does not need to have the same location as on the source server. In this example, the locations of the files have changed so the job reading this dataset needs to be modified to reflect the changes on the location. These details are used on the following slides.



## Create simple job to read the dataset

The screenshot displays the configuration for a DataStage job to read a dataset. It includes a job graph, a dataset configuration window, a stage configuration window, and a job properties window.

**Data\_Set\_0 - Data Set Configuration:**

Column name	Key	SQL type	Extended	Length	Scale	Nullable
col1	<input type="checkbox"/>	VarChar		10		

**DSLink2 Stage Configuration:**

Source: File = /home/dsadm/Federico/input\_ds\_file/ds...  
 Options: Missing Columns Mode = Ignore  
 File: /home/dsadm/Federico/input\_ds\_file/datasetfile.ds

**Job Properties:**

Parameter name	Prompt	Type	Default Value	Help Text
\$APT_CONFIG_FILE	Configuration file	Pathname	/home/dsadm/Federico/APT_CONFIG_LOC/test.apc	The Parallel job configuration file.

9

Manually copy a dataset from one server to another

© 2013 IBM Corporation

Once all the files are in their locations as displayed in the previous slides and you have verified that all the file ownership and permissions are correct, the next step is to create a simple job to read the dataset. A simple dataset input file and a peek stage suffices to test. Ensure that you point to the correct APT\_CONFIG file and dataset file locations. Once all the necessary changes have been completed, compile and run the job and verify that the data has been read correctly.

## Why use this method to copy a dataset?

- Testing DataStage jobs on a different DataStage server
  - Required to use the same input dataset file
- Assist IBM support to replicate issues that may be directly related to specific input dataset data

There are a couple of instances where it may be required to use a copy of a dataset from another server rather than recreating the dataset on the second server. One example is where testing needs to be done where the data extraction on the new system has not been enabled, hence, the same dataset files cannot be created. In this case, it is necessary to use an existing dataset from another similar server environment to run some initial testing.

This method can also be very useful when raising support cases where a test case requires an input dataset to replicate the issue.

## Restrictions

- Dataset files
  - Both servers must be same OS platform and major release
- Target server must be same DataStage release or higher than source

This method of manually copying the dataset files from one server to another does have some restrictions. First, the source server operating system where the dataset has been created must be the same as that of the target server. Ideally, you should copy the datasets between the same releases of DataStage. However, you should also be able to read older datasets with newer versions of DataStage.

## Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, AIX, DataStage, and InfoSphere are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)" at <http://www.ibm.com/legal/copytrade.shtml>

Other company, product, or service names may be trademarks or service marks of others.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2013. All rights reserved.