

Nuts & Bolts of DataStage

Something about DataStage, DataStage Administration, Job Designing, Developing, DataStage troubleshooting, DataStage Installation & Configuration, ETL, DataWareHousing, DB2, Teradata, Oracle and Scripting.

[Home](#)[Interview Questions](#)[DataStage Scenarios](#)[Series Posts](#)[E-Books](#)[About Me !!](#)**Total Pageviews**~~1~~ ~~4~~ ~~5~~ ~~4~~ ~~6~~ ~~2~~ ~~0~~**Try Me**

Must Reads

[DataSet in DataStage](#)[Issuing commands to a Queue Manager \(runmqsc\)](#)[Hash Files in DataStage](#)[XMeta DB : Datastage Repository](#)[InfoSphere DataStage Jobstatus returned Codes from dsjob](#)[Conductor Node in Datastage](#)[Schema File in Datastage](#)[Sort stage to remove duplicate](#)[14 Good design tips in Datastage](#)[Datastage Coding Checklist](#)**Monday, May 26, 2014**

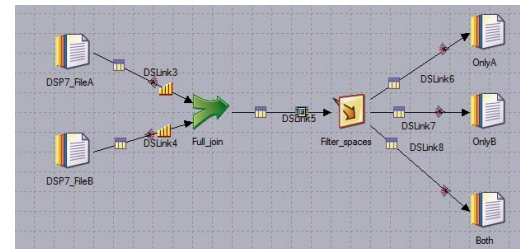
DataStage Scenario - Design8 - job1

DataStage Scenario Problem --> [DataStage Scenario - Problem8](#)

Solution Design :

a) Job Design :

Below is the design which can achieve the output as we needed. Here, we are reading 2 seq file as a input, then data is passing through a Join and Filter stage to achieve the output.



b) Join Stage Properties : For the required output, we have to join the both input file on "Col1". Here we are using FULL OUTER JOIN.

Get daily dose of Tech Food

Email address...

tech foodies

DataStage4You



111 have us in circles

[View all](#)

Blog Archive

▼ 2014 (103)

- October (7)
- September (9)
- August (5)
- July (12)
- June (10)

▼ May (13)

[Surrogate Key Generator - Create/Update State File...](#)

[Surrogate Key Generator - Create State File](#)

[Some Oracle SQL should known by Developer - 1](#)

[Interview Questions : DataWareHouse - Part6](#)

[DataStage Scenario - Design8 - job1](#)

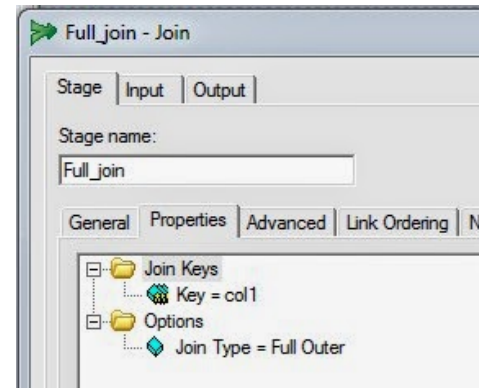
[Data Warehouse Testing Checklist](#)

[DataStage Scenario - Design7- job1](#)

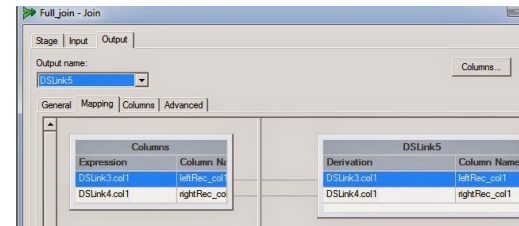
[DataStage Scenario - Design6 - job1](#)

[What ETL is not?](#)

[Framework \(usually followed \) in ETL Testing](#)



FULL OUTER JOIN will generate the 2 columns LEFT n RIGHT. Map the both columns to Join Output.



b) Filter Stage Properties : Now, In Filter Stage, We will filter the data as per requirement.

Here

FileA = Left and FileB = Right

So...

Data Only in **FileA** = **where Right_Col1 = 0 or "" or NULL**

Data Only in **FileB** = **where Left_Col1 = 0 or "" or NULL**

Data in both **FileA and FileB** = where **(Right_Col1 <> 0 or "" or NULL) and (Left_Col1 <> 0 or "" or NULL)**

Note :-

Left and Right Col will be = 0 or "" or NULL if-

NULL => Data will be NULL if the Source is DB. (But here we are using seq file)

In case of Sequence File as Source

"" => Data will be "" if the column data type is VarChar.

0 => Data will be 0 if the column data type is Integer.

ETL Testing : Trends & Challenges

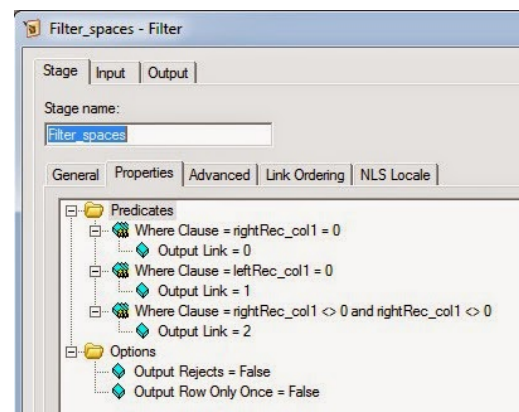
ETL Testing : Approach

DataStage Scenario - Problem19

- ▶ April (10)
- ▶ March (9)
- ▶ February (16)
- ▶ January (12)
- ▶ 2013 (167)
- ▶ 2012 (175)
- ▶ 2011 (8)

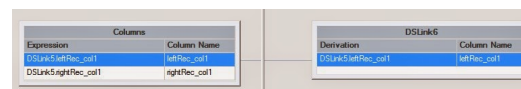
Tags Cloud

&PH& 421 Administration advantage Agents aggregator Answers application architecture ASB attribute authorities backup basic binary block books Buffer certification change channel checkpoint cleanup clear client Code column Column Generator commands compiler Concept Conceptual conductor Configuration container copy counter create Crontab Data database DataSet DataStage DataWareHouse DB2 DBMS deadlock debug delete deploy design develop difference dimension Dimensional director Documentation dsenv dsjob DSpam DSRPC dump duplicate encrypt engine environment Errors ETL exception execution export fact factless FAQ file FileSet filter free ftp fun function fundamentals granularity Guest hadoop handling hash head hide horizontal Host huge hyperlink import increase index Information input install Interview issue istool Java jdbc Job join keys leaders Link Linux list listener load local locks Logging Logical Login logs lookup macro mail maintenance managers memory merge message queue Metadata Model modify Monitor MQ MQSC multiple names NLS node notes notification odbc odbc.ini operator Optimizing Oracle orchadmin ORLogging orphan OS osh output package Parallel Parallelism parameter partition password peek performance Perl phantom Physical pivot player port Practices problem process profile programming Project purge Putty Questions read registry reject release remove report Resource Restart Roles



Now, we will map the columns to output.

For "Data available only in FileA" :- Assign Left to Output.



For "Data available only in FileB" :- Assign Right to Output.



For "Data available in Both " :- Assign Right or Left to Output.

For More ---> [VISIT THIS LINK](#)

By ETL DataStage at 23:19 0 Comments

Labels: Code, DataSet, DataStage, design, develop, filter, function, input, Job, join, output, problem, run, scenario, Seq File, sequential

[routine](#) [row generator](#) [rows](#) [RTLogging](#) [run sample](#) [SCD scenario](#) [scheduler](#) [Schema](#) [score](#) [Scratch](#) [Script](#) [section](#) [Seq File](#) [sequence](#) [Server](#) [Service](#) [session](#) [Setting](#) [Share](#) [Shell](#) [shell scripting](#) [shortcuts](#) [show](#) [slowly](#) [snowflake](#) [solution](#) [sort](#) [source](#) [space](#) [SQL](#) [SSH](#) [stages](#) [Standards](#) [Star](#) [Start](#) [statistics](#) [status](#) [Stop](#) [storage](#) [surrogate](#) [switch](#) [system](#) [table](#) [tail](#) [target](#) [temporary](#) [teradata](#) [time](#) [tips](#) [tool](#) [trace](#) [transformation](#) [transformer](#) [trigger](#) [Troubleshoot](#) [tuning](#) [Tutorial](#) [type](#) [unique](#) [Unix](#) [User](#) [Utility](#) [UV](#) [uvodbc.config](#) [variables](#) [version](#) [videos](#) [view](#) [Vincent](#) [McBurney](#) [Virtual](#) [warnings](#) [WAS](#) [websphere](#) [windows](#) [write](#) [Write Range Map](#) [XMETA](#) [xml](#) [z/OS](#)

0 Comments DataStage4You

Login

Sort by Best

Share Favorite



Start the discussion...

Be the first to comment.

Subscribe

Add Disqus to your site

Privacy

Newer Post

Home

Older Post

Subscribe to: Post Comments (Atom)

Disclaimer

The postings on this site are my own and don't necessarily represent IBM's or other companies positions, strategies or opinions. All content provided on this blog is for informational purposes only. The owner of this blog makes no representations as to the accuracy or completeness of any information on this site or found by following any link on this site. The owner will not be liable for any errors or omissions in this information nor for the availability of this information. The owner will not be liable for any losses, injuries, or damages from the display or use of his information. //-----

Did you find this Blog helpful ?? Let me know @ www.facebook.com/datastage4you. Ethereal template. Powered by [Blogger](#).