Me with Myself...

Feel the way I feel...!

| Home | Technically Me. ♛ | Research and Hacks | Mannu..... Finally a Computer Engineer | Disclaimer : Don't sue me :) |

Wednesday, April 9, 2014

Datastage – Change Capture Stage

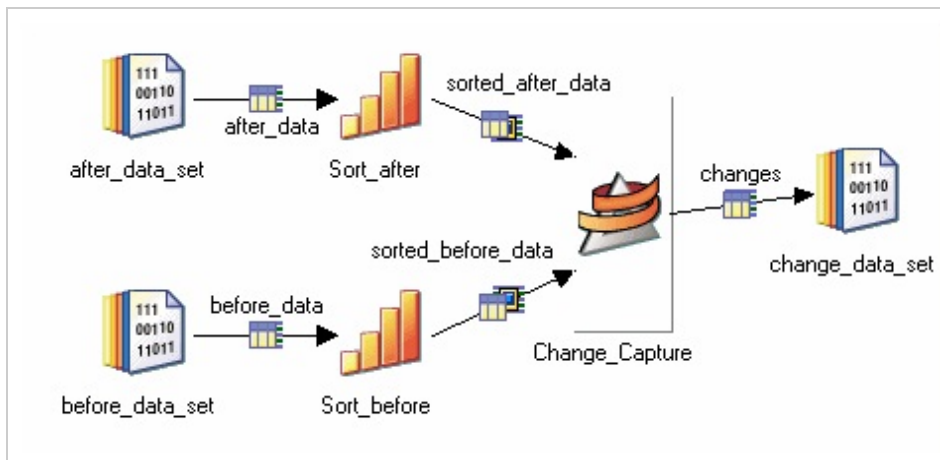**What is Change Capture Stage?**

The Change Capture Stage is a processing stage. The stage compares two data sets and makes a record of the differences.

The Change Capture stage takes two input data sets, denoted before and after, and outputs a single data set whose records represent the changes made to the before data set to obtain the after data set. The stage produces a change data set, whose table definition is transferred from the after data set's table definition with the addition of one column: a change code with values encoding the four actions: insert, delete, copy, and edit. The preserve-partitioning flag is set on the change data set.

The compare is based on a set a set of key columns, rows from the two data sets are assumed to be copies of one another if they have the same values in these key columns. You can also optionally specify change values. If two rows have identical key columns, you can compare the value columns in the rows to see if one is an edited copy of the other.

The stage assumes that the incoming data is key-partitioned and sorted in ascending order. The columns the data is hashed on should be the key columns used for the data compare. You can achieve the sorting and partitioning using the Sort stage or by using the built-in sorting and partitioning abilities of the Change Capture stage.

You can use the companion Change Apply stage to combine the changes from the Change Capture stage with the original before data set to reproduce the after data set.

To use a Change Capture stage:

In the Stage page Properties Tab:
- o Specify the key column. You can repeat this property to specify a composite key. Before and after rows are considered to be the same if they have the same value in the key column or columns.
- o Optionally specify one or more Value columns. This enables you to determine if an after row is an edited version of a before row.
    - ▪ (You can also set the Change Mode property to have DataStage treat all columns not defined as keys treated as values, or all columns not defined as values treated as keys.)
- o Specify whether the stage will output the changed row or drop it. You can specify this individually for each type of change (copy, delete, edit, or insert).

In the Stage page Link Ordering Tab:
- o specify which of the two links carries the before data set and which carries the after data set.
- o If the two incoming data sets aren't already key partitioned on the key columns and sorted, set DataStage to do this on the Input Page Partitioning Tab.

In the Output Page Mapping Tab:
- o specify how the change data columns are mapped onto the output link columns.

Change Capture stage: NLS Locale tab

This appears if you have NLS enabled on your system. It lets you view the current default collate convention, and select a different one for this stage if required. You can also use a job parameter to specify the locale, or browse for a file that defines custom collate rules. The collate convention defines the order in which characters are collated. The Change Capture stage uses this when it is determining the sort order for key columns. Select a locale from the list, or click the arrow button next to the list to use a job parameter or browse for a collate file

Change Capture stage: Input page

Index

Archive

The Input page allows you to specify details about the incoming data sets. The Change Capture expects two incoming data sets: a before data set and an after data set.
The General tab allows you to specify an optional description of the input link. The Partitioning tab allows you to specify how incoming data is partitioned before being compared. The Columns tab specifies the column definitions of incoming data. The Advanced tab allows you to change the default buffering settings for the input link.

Change Capture stage: Output page
The Output page allows you to specify details about data output from the Change Capture stage. The Change Capture stage can have only one output link.
The General tab allows you to specify an optional description of the output link. The Columns tab specifies the column definitions of the data. The Mapping tab allows you to specify the relationship between the columns being input to the Change Capture stage and the Output columns. The Advanced tab allows you to change the default buffering settings for the output link.

Change Capture stage: Mapping tab
For the Change Capture stage the Mapping tab allows you to specify how the output columns are derived, that is, what input columns map onto them and which column carries the change code data.
The left pane shows the columns from the before/after data sets plus the change code column. These are read only and cannot be modified on this tab.
The right pane shows the output columns for each link. This has a Derivations field where you can specify how the column is derived. You can fill it in by dragging input columns over, or by using the Auto-match facility. By default the data set columns are mapped automatically. You need to ensure that there is an output column to carry the change code and that this is mapped to the Change_code column.
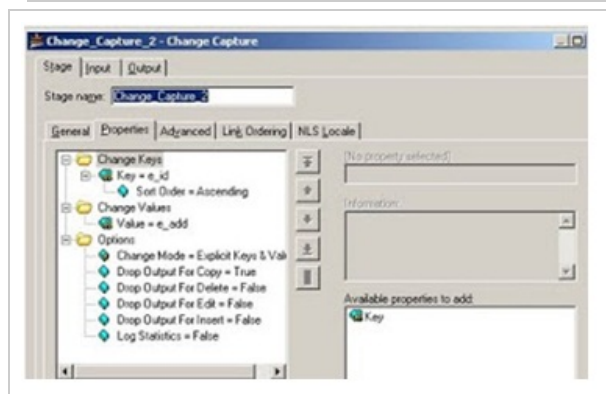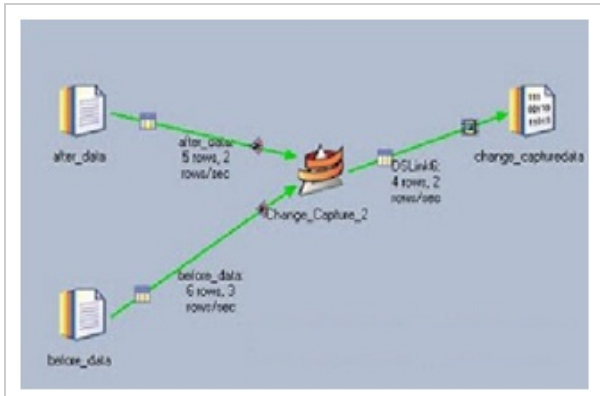
Examples:
Change_before_data
e_id,e_name,e_add
  11,kim,syd
  22,jim,mel
  33,pim,perth
  44,lim,canb
  55,pom,adeliade
  66,shila,bris
============================
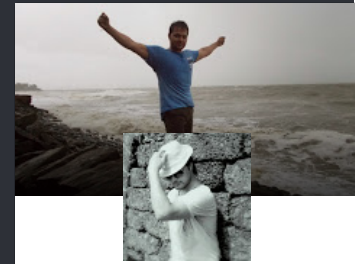**Change_after_datae**
id,e_name,e_add
  11,kim,syd
  22,jim,canb
  33,pim,syd
  44,lim,canb
  55,pom,perth

In Change Capture Stage Select Key column as e_idand value column as e_add.Because e_add is the column we

are going to update.

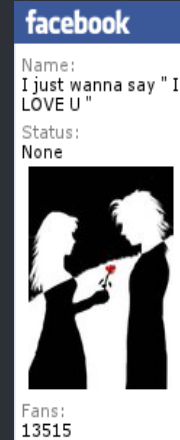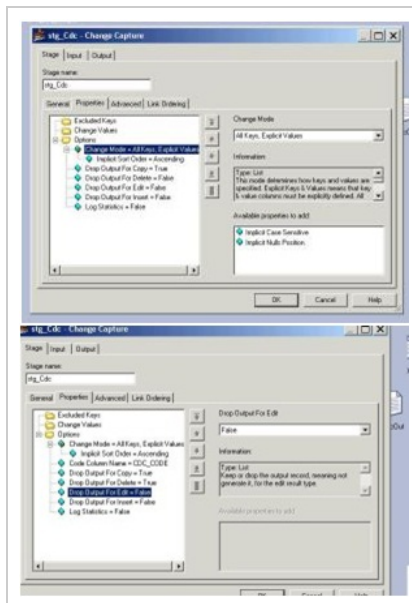
Configuring CDC stage to get only new records
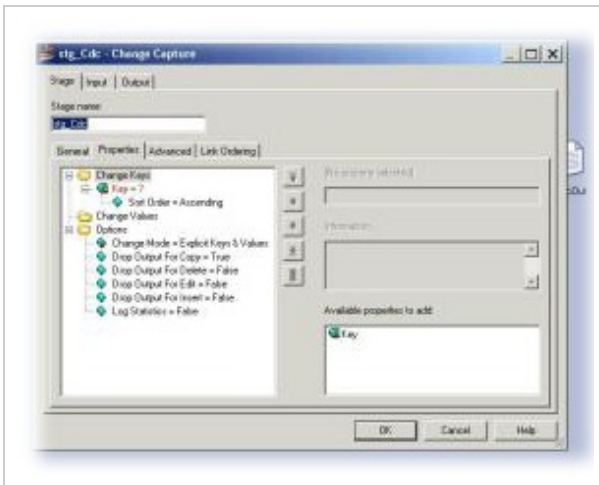
In the above CDC configuration the main option we use is Change Mode.
In this scenario we are using **All Keys,Explicit Values**.
All Keys, Explicit Values:In this mode we don't need to specify key columns.

We need only newly inserted or edited record and hence we set Drop output for insert = FALSE,
Drop Output for Edit = FALSE and the remaining all three are set to FALSE because we don't want copy, edit or delete records.
Note:Always make sure the link ordering is corrected otherwise you will get wrong results

We need to configure Code Column Name in the options tab:
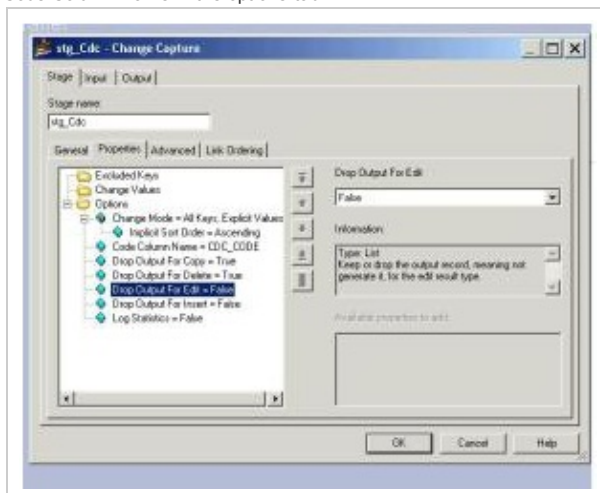
Copy code
Allows you to specify an alternative value for the code that indicates the after record is a copy of thebefore record.
By default this code is 0.
Deleted code
Allows you to specify an alternative value for the code that indicates that a record in the before set hasbeen deleted
from the after set. By default this code is 2.
Edit code
Allows you to specify an alternative value for the code that indicates the after record is an edited versionof the
before record. By default this code is 3.
Insert Code
Allows you to specify an alternative value for the code that indicates a new record has been inserted inthe after set
that did not exist in the before set. By default this code is 1.

Usage of Change Capture stage:
We have used the concept of Change Capture in our Project for the purpose of Data validation.
We are comparing the data from both the Environments PROD and QA. These validation jobs provide us the
differences in data from both the environments.
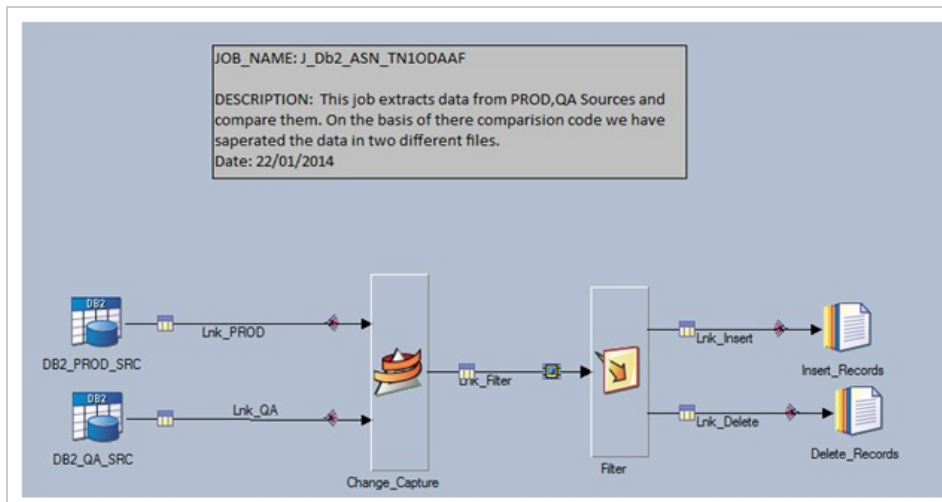
Posted by manmohan mishra at 10:16 AM

Labels: Datastage   M B t F P   g+1   Recommend this on Google

Reactions: ☐ funny (0)   ☐ interesting (0)   ☐ cool (0)

No comments:

Post a Comment

```
Enter your comment…
```

Comment as: Google Accoui ▼

[Publish]   [Preview]

Links to this post

Create a Link

Newer Post   Home   Older Post

Subscribe to: Post Comments (Atom)

Google   Gigaom -   Google News: The biggest missed opportunity in media right now   powered by Google™

Mannnu. Awesome Inc. template. Powered by Blogger.