

Data Stage FAQs

INTRODUCTION TO DATA STAGE, CONCEPTS, ROUTINES, FAQs, INTERVIEW QUESTIONS, SQL QUERIES, UNIX FAQs

WEDNESDAY, FEBRUARY 10, 2010

About Change Capture Stage (IMPORTANT)

Change Capture Stage

The Change Capture Stage is a processing stage. The **stage** compares two data sets and makes a record of the differences. An example before and after data set are given in Parallel Job Developer's Guide. Follow this link for a list of steps you must take when deploying a Change Capture stage in your job.

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 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 to see if one is an edited copy of the other.

The stage assumes that the incoming data is hash-partitioned and sorted in ascending order (this is done automatically if (auto) is selected on the partitioning tab). 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.

The Change Capture stage is very similar to the Difference stage.

The stage editor has three pages:

Stage page. This is always present and is used to specify general information about the stage.

Inputs page. This is where you specify details about the data set having its duplicates removed.

Outputs page. This is where you specify details about the processed data being output from the stage.

The **General tab** allows you to specify an optional description of the stage.

The Properties tab lets you specify what the stage does. The Advanced tab allows you to specify how the stage executes. The Link Ordering tab allows you to specify which input link carries the before data set and which the after data set.

Blog Archive

▼ 2010 (22)

▼ February (22)

[Real time Problems in datastage with Solutions](#)

[ETL Fundamentals](#)

[DataStage Parallel Processing Architecture Overview...](#)

[datastage faqs](#)

[About AWK Command in Unix](#)

[The 15 Most Important UNIX commands](#)

[UNIX FAQs](#)

[DATASTAGE ARCHITECTURE DIAGRAM](#)

[Can one retrieve only the Nth row from a table?](#)

[Difference between TRUNCATE, DELETE and DROP comma...](#)

[Important SQL Queries](#)

[How to release a lock held by jobs?](#)

[4 Ways of Executing DS Job](#)

[Aggregator Stage](#)

[About Change Capture Stage \(IMPORTANT\)](#)

[Concepts Of DWH](#)

[Data Warehousing Basic Questions](#)

[Data Stage Components](#)

[How to Make Money Online Learn Secrets for FREE ?](#)

[Datastage interview questions from major companies...](#)

[DataStage Terms and Concepts](#)

[Data Stage Interview Questions](#)

No comments:

Post a Comment

Enter your comment...

Comment as: Google Account ▼

Publish

Preview

[Newer Post](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)