

DataStage es una herramienta ETL que permite crear y mantener fácil y rápidamente almacenes de datos procedentes de sistemas de Aplicaciones Empresariales incluyendo SAP, Siebel, Oracle y PeopleSoft CRM y de otros sistemas relacionados al negocio.

Datastage soporta la extracción, integración y transformación de altos volúmenes de datos desde estructuras simples hasta muy complejas. Entre funciones de DataStage, la más destacada es el diseño de tareas que extraen, integran, agregan, cargan y transforman los datos para el Data Warehouse o Data Mart. Las tareas son compiladas para crear ejecutables que son calendarizados por el Director y ejecutadas por el Servidor.

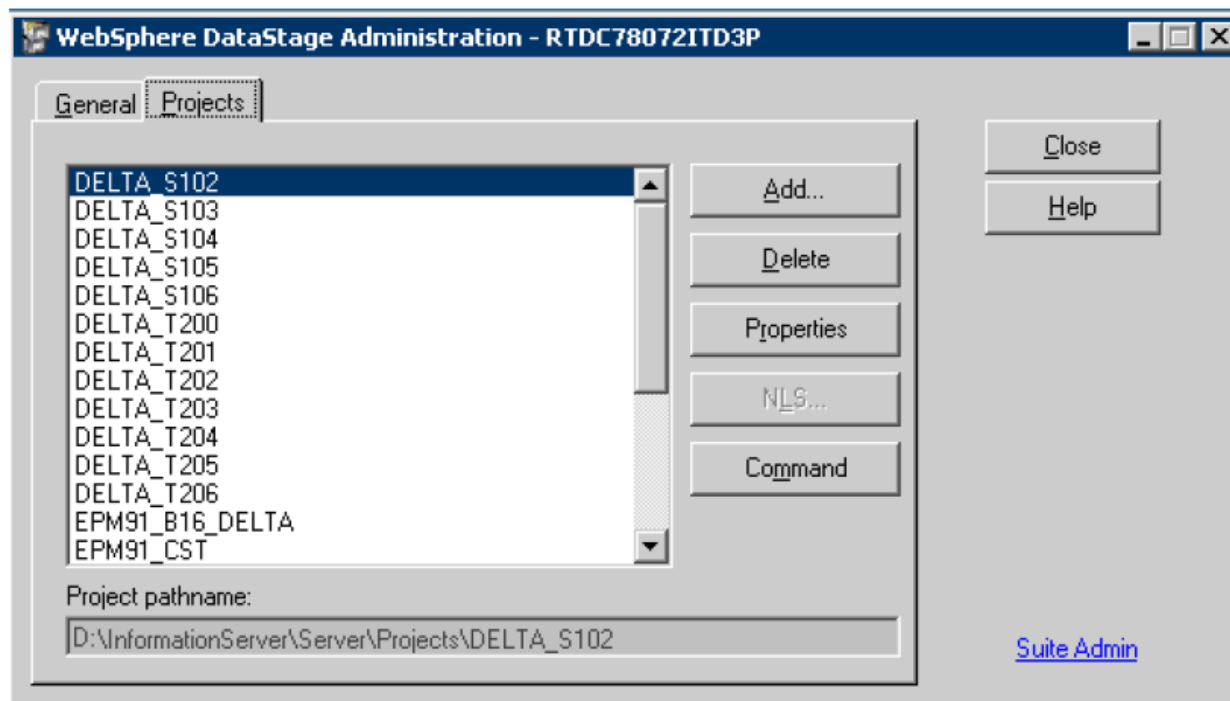
Historia Datastage

Antes el producto se llamaba Ardent DataStage, luego adquirido por Ascential y en 2005 lo compró IBM. Desde entonces su nombre oficial es IBM WebSphere Datastage.

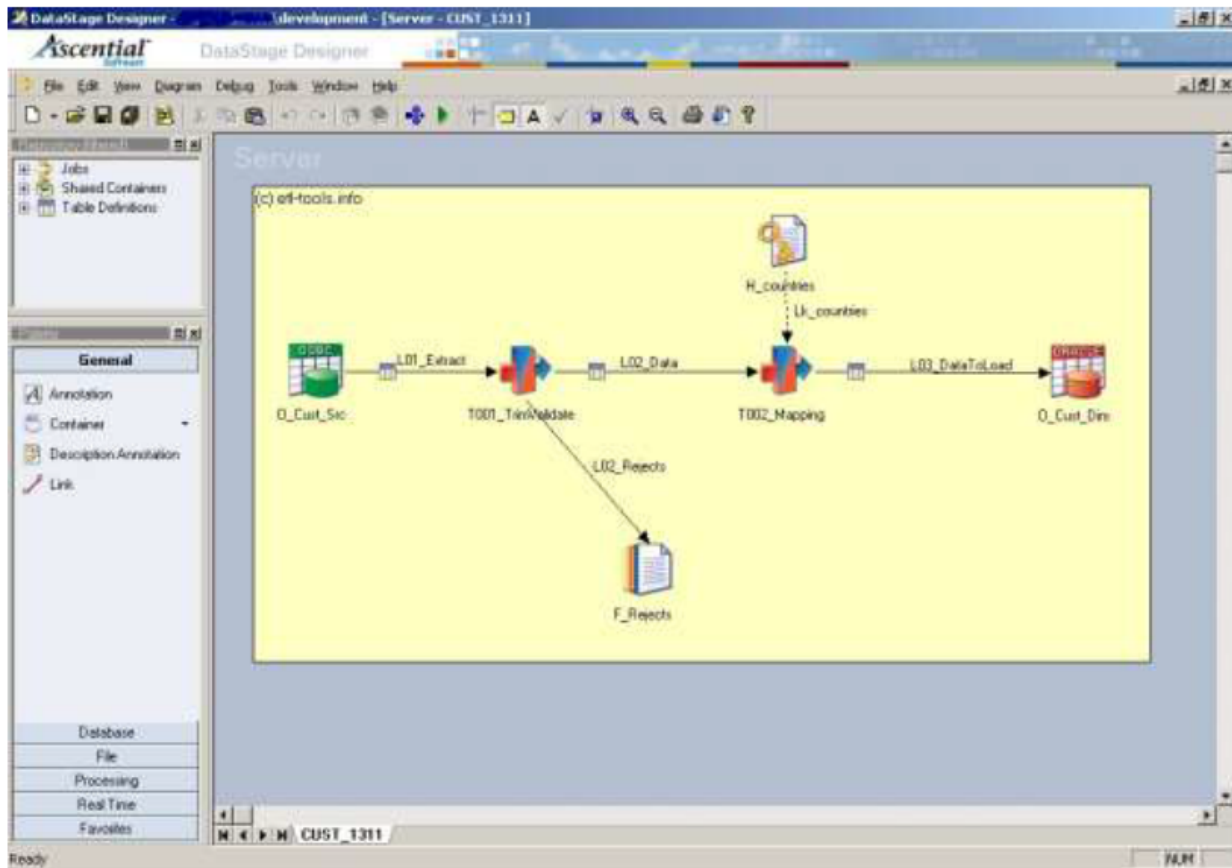
Componentes de DataStage

Hay cuatro componentes principales de la herramienta Datastage:

1. .Administrator - Interfaz de usuario usada para configurar proyectos de Datastage y usuarios. Gestiona administración de proyectos de DataStage en ambientes de desarrollo y producción.



2. Designer - usada para crear, diseñar y compilar tareas Datastage (también permite testear y ejecutar). Mayormente usado por los desarrolladores.

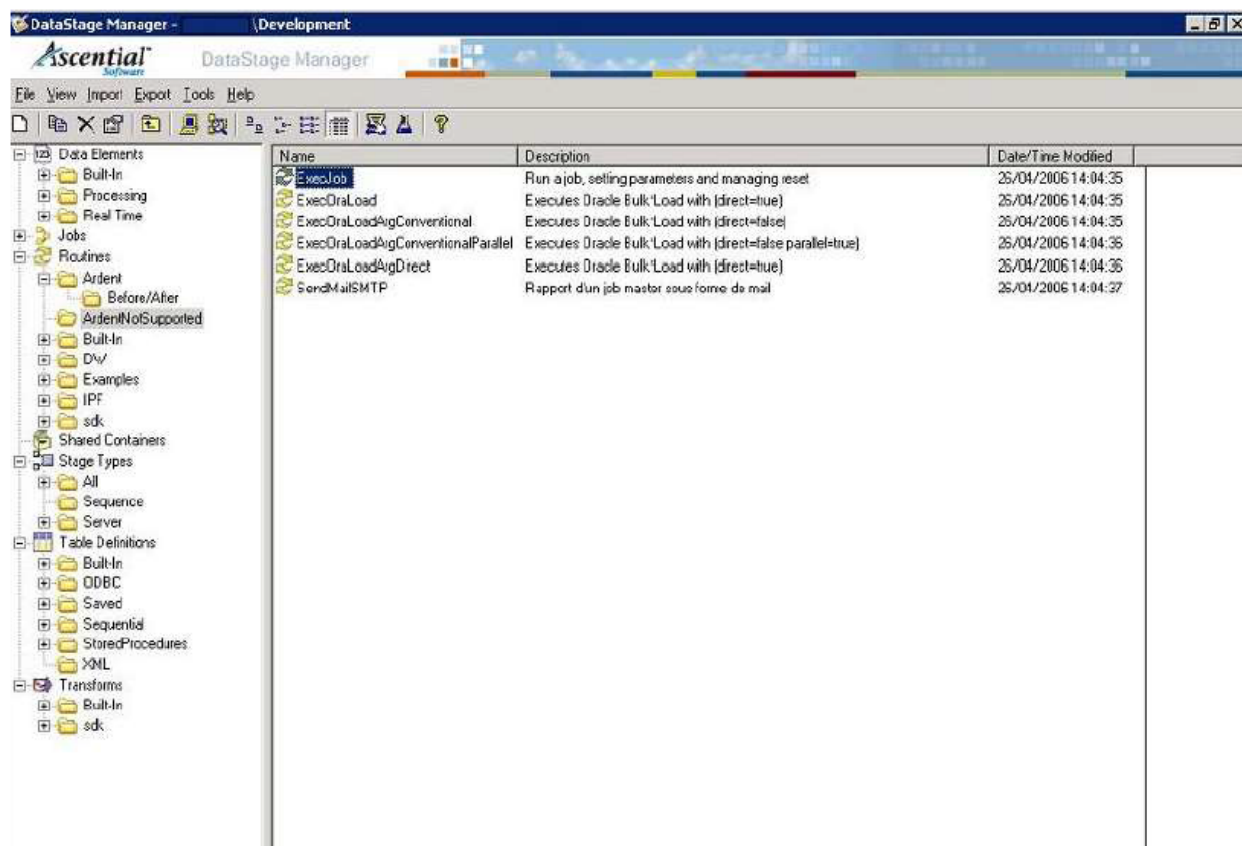


3. Director - para validar, calendarizar, testear, ejecutar y monitorizar jobs Datastage. Usado por los operadores y los testers.

>Occurred	>On date	Type	Event
9:16:49 PM	8/25/2003	Control	Starting Job test4x. (...)
9:16:49 PM	8/25/2003	Info	Environment variable settings: (...)
9:16:50 PM	8/25/2003	Info	Parallel job initiated
9:16:51 PM	8/25/2003	Info	main_program: DataStage XE Parallel Extender V6.0.1 (...)
9:16:52 PM	8/25/2003	Info	main_program: orchgeneral loaded (...)
9:16:52 PM	8/25/2003	Info	main_program: APT configuration file: /dsp/Ascential/DataStage/Configurations/default.ap (...)
9:17:00 PM	8/25/2003	Warning	main_program: Warning: the value of the PWD environment variable (/) does not appear to be a sy...
9:17:01 PM	8/25/2003	Warning	Transformer_11: When checking operator: On output data set 0: Data set [Virtual dataset; output of
9:17:01 PM	8/25/2003	Warning	Aggregator_24: When checking operator: When binding output interface field "sum_book_rev_neg
9:17:01 PM	8/25/2003	Warning	Aggregator_25: When checking operator: When binding output interface field "sum_book_rev_posi
9:17:01 PM	8/25/2003	Warning	dellsmithNegativeSum.negativeSum: When checking operator: When validating export schema: At
9:17:01 PM	8/25/2003	Warning	dellsmithNegativeSum.negativeSum: When checking operator: When validating export schema: At
9:17:01 PM	8/25/2003	Warning	dellsmithNegativeSum.negativeSum: When checking operator: Sequential operator cannot preservi
9:17:01 PM	8/25/2003	Warning	dellsmithPositiveSum.positiveSum: When checking operator: When validating export schema: At fie
9:17:01 PM	8/25/2003	Warning	dellsmithPositiveSum.positiveSum: When checking operator: When validating export schema: At fie
9:17:05 PM	8/25/2003	Warning	dellsmithPositiveSum.positiveSum: When checking operator: Sequential operator cannot preservi
9:17:05 PM	8/25/2003	Info	ismth2003072.AllRecords.0 Progress: 10 percent (...)
9:17:05 PM	8/25/2003	Info	ismth2003072.AllRecords.0 Import complete. 1500 records imported successfully. 0 rejected.
9:17:05 PM	8/25/2003	Info	dellsmithNegativeSum.negativeSum.0: Export complete. 1 records exported successfully. 0 rejected.
9:17:05 PM	8/25/2003	Info	dellsmithPositiveSum.positiveSum.0: Export complete. 1 records exported successfully. 0 rejected.
9:17:05 PM	8/25/2003	Info	main_program: Step execution finished with status = OK.
9:17:05 PM	8/25/2003	Info	main_program: Startup Time 0:13 Production Run Time 0:00
9:17:07 PM	8/25/2003	Info	Parallel job reports successful completion
9:17:07 PM	8/25/2003	Control	Finished Job test4x.

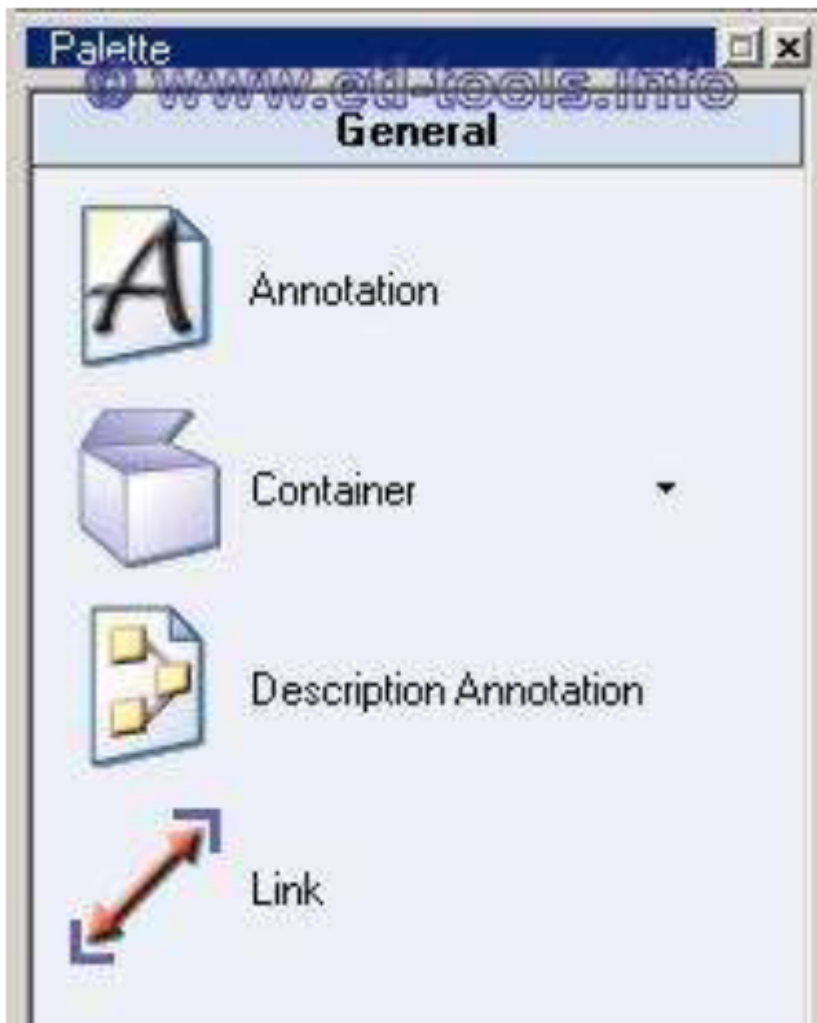
Log for job: test4x 25 entries (filtered)

4. Manager - interfaz de usuario usada para visualizar y editar el contenido del repositorio.

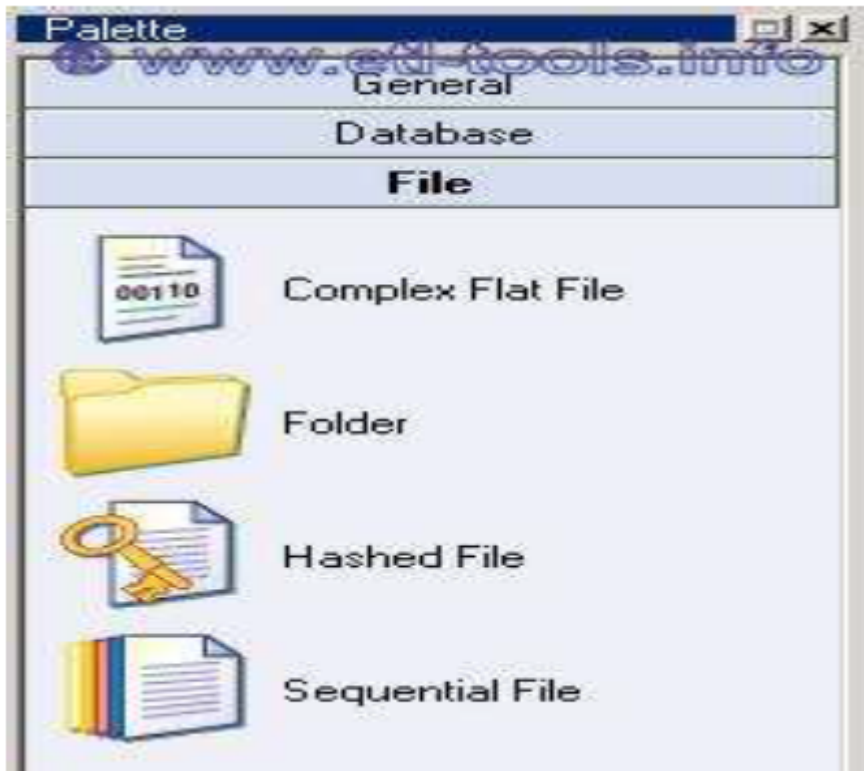


Diseño de las tareas - Datastage palette

1. Datastage Server Palette - Los stages generales



2. Datastage Server Palette - Los stages de fichero



3. Datastage Server Palette - Los Stages de Base de Datos



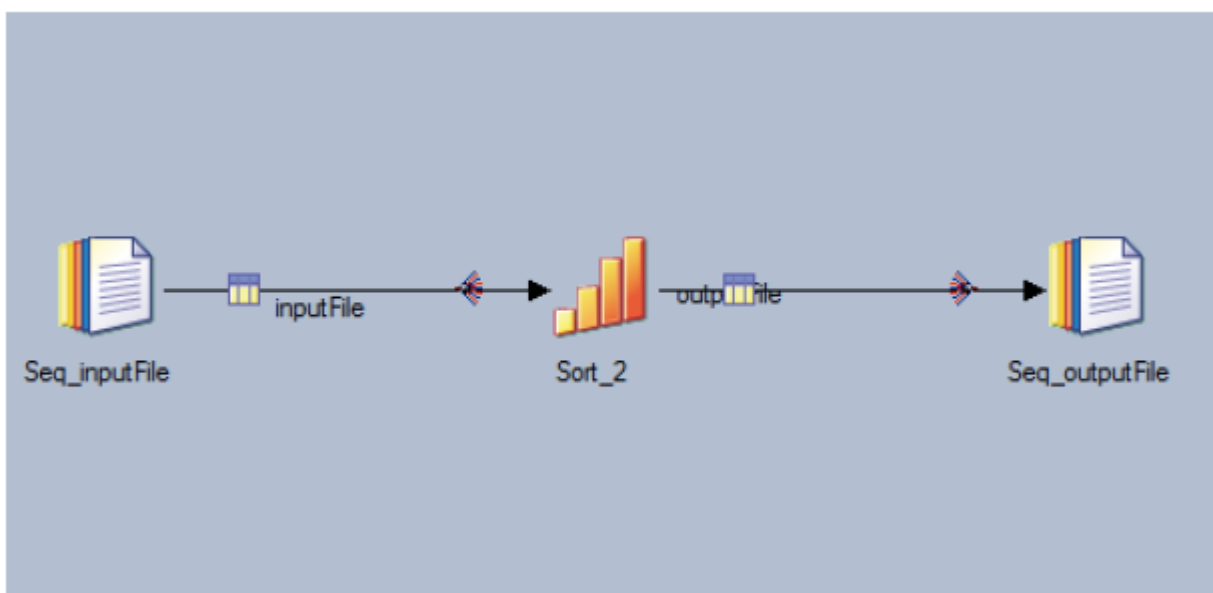
4. Datastage Server Palette - Los Stages de Transformar y Filtrar



5. Datastage Server Palette - Elementos de Secuencias



6. Ejemplo de Job que carga datos de un archivo de texto, a otro archivo de texto



Fuente: http://etl-tools.info/es/datastage/curso_datastage.htm

Julio Pari (IT Architect IBM)



Si te ha interesado este artículo y deseas un apoyo o asesoría en algún requerimiento, envíame un mensaje a: (info@juliopari.com) o sino a través de LinkedIn:
<https://www.linkedin.com/in/juliopari/>
