

Guía Migración Datastage versión 7.5.x a Information Server 8.5

Revision History

Date	Version	Description	Author
19/08/11	1.0	Integración de documento v1.	Alan Sumano
23/08/11	2.1	Actualización general	Alberto Ortiz
12/10/11	2.2	Actualización general	Alan Sumano
09/11/11	2.3	Actualización general	Alan Sumano
28/11/11	2.4	Correcciones ortográficas y de formato. Agregar tabla de contenido.	Alberto Ortiz

Contenido

1. Introducción.....	3
2. Alcance y actualización.....	3
3. Descripción del proceso de migración.....	3
3.1. Análisis del proyecto actual.....	4
3.2. Ejecutar respaldos de componentes.....	4
3.3. Definiciones de la topología a utilizar.....	6
3.4. Instalación 7	
3.5. Configuración 9	
3.6. Reestablecer los proyectos desde la versión 7.5.....	9
3.7. Recompilar los procesos de carga.....	10
3.8. Explotar el nuevo ambiente DataStage.....	11
3.9. Migración Server Edition a Enterprise Edition.....	11
3.10. Transferencia de conocimiento.....	13

1. Introducción

El objetivo de este documento es enlistar las actividades y componentes necesarios para migrar los ambientes y proyectos de DataStage de la versión 7.5.x a la plataforma Information Server de IBM, en la versión 8.5.

2. Alcance y actualización

Este documento funciona como compilación de documentación, mejores prácticas y recomendaciones en proyectos similares provistas por el equipo técnico de IM y SWG Services, sin embargo no sustituye a la documentación oficial de los productos ni debe ser tomado como entregable o parte de algún Statement of Work.

Ya que la información provista es constantemente actualizada en los sitios públicos correspondientes de IBM, debe verificarse la actualización y cambios correspondientes, los cuales podrían afectar los pasos recomendados en esta guía.

La implementación y ejecución son recomendados y deben adecuarse a la plataforma final del cliente.

Esta guía sólo aplica a las versiones enlistadas en este documento y puede requerir la integración de diferentes componentes o “tipos” de licenciamiento de alguno de ellos.

Utilice la documentación actualizada y oficial disponible en:

<http://www-01.ibm.com/software/data/infosphere/info-server/overview/>

3. Descripción del proceso de migración

El proceso de migración consiste, de manera general, en documentar, respaldar, transportar y poner en operación los proyectos, procesos de carga, extracción y transformación de datos desarrollados en la versión 7.5.x de DataStage a una nueva instalación de Information Server 8.5 u 8.7.

Este proceso debe seguir una serie de pasos recomendados y el uso de algunas utilerías o componentes específicamente desarrollados para ello.

Parte fundamental del proceso de migración es la planeación y diseño de la migración. En estas etapas es en donde este documento es utilizado como referencia, y en las etapas posteriores será utilizado como punto de control o de aseguramiento de calidad, como complemento a la documentación provista en el sitio oficial y a los artefactos desarrollados por el equipo de trabajo. Este documento no funciona como un entregable ni como un alcance de ningún plan de trabajo.

El proceso considera las siguientes etapas:

- Análisis del proyecto actual.
- Ejecutar respaldos de componentes
- Definiciones de la topología a utilizar
- Instalación
- Configuración
- Reestablecer los proyectos desde la versión 7.5
- Recompilar los procesos de carga
- Explotar el nuevo ambiente de DataStage
- Migración Server Edition a Enterprise Edition.
- Transferencia de conocimiento

3.1. Análisis del proyecto actual.

El objetivo de ésta actividad es contar con un inventario que tenga la descripción y propósito de:

- Proyectos Datastage.
- Usuarios (OS y aplicativos).
- Parámetros y variables de sistema utilizados en los jobs.
- Jobs o procesos de carga.
- Carpetas o file systems utilizados, así como su definición y capacidad.
- Conectores utilizados (tipo de stage, versión, manejador, entre otros).
- Archivos temporales, incluyendo su ubicación y formato, utilizados como almacenamiento, de referencia, etc.
- Programas o rutinas que ejecuten los jobs de carga, por ejemplo programas en Java, C++.
- El calendario de ejecución de los procesos, así como sus dependencias y orden, si es que son secuencias.

El resultado de éste inventario puede servir de valor de entrada para la definición de actividades y control del proceso de migración.

3.2. Ejecutar respaldos de componentes

Se deben respaldar todos los componentes del inventario:

- Filesystems, archivos temporales o finales, programas. Éste respaldo es físico.
- Usuarios, parámetros y variables. Deben ser capturados, incluyendo si es que aplica, los grupos a los que pertenecen de forma en que se puedan editar o reconfigurar.

- Proyectos y jobs de carga Datastage. A través de la capacidad de exportar, disponible en el cliente **DataStage Manager Client** o a través de la línea de comando DS.
- Al realizar el export, se obtendrá un documento de texto con extensión **.dsx**, el cual puede contener todo el proyecto de DS o un sólo job o componente. Al hacer el export, se debe ser cuidadoso para seleccionar las opciones de incluir todos los jobs del proyecto y confirmar que estarán todos los items dependientes. El export puede hacerse de forma selectiva en varios grupos o archivos **.dsx** lo único que debe asegurarse es que existan todos los jobs e items que deseen llevarse o sean necesarios en el nuevo proyecto.
- Archivos de configuración. Dependiendo del OS en el que se encuentre DataStage instalado, se contará con los siguientes archivos de configuración, los cuales requieren ser respaldados:
 - Unix. Settings files
 - Windows. DSParam files
- Debe asegurarse de que existe también el respaldo de los siguientes archivos:
 - dsenv
 - .odbc.ini
 - uvodbc.config
- Variables de ambiente. Utilice dsadmin command para capturar las variables de ambiente completas.
- En el caso de hashed files, requiere un procedimiento específico, el cual puede revisar en la referencia "Migrating hashed files" descrita abajo.

Referencias:

Move job dependency objects

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/movingjobdependencies.html>

Exporting projects from the DataStage Manager client

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/exportinglegacyprojects.html>

Saving DataStage settings files

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/index.jsp?topic=/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/savingconfigfiles.html>

Migrating hashed files

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/movingjobdependencies.html>

3.3. Definiciones de la topología a utilizar

La topología a utilizar para la versión 8.5 debe ser revisada y entendida previo a la instalación. Ésta topología definirá, entre otras cosas, en donde residirán y como interactuarán los componentes de la nueva versión.

Es común que durante los ejercicios de dimensionamiento se defina ésta topología, la siguiente figura muestra dos posibilidades para distribuir los componentes: two-tier y three-tier.

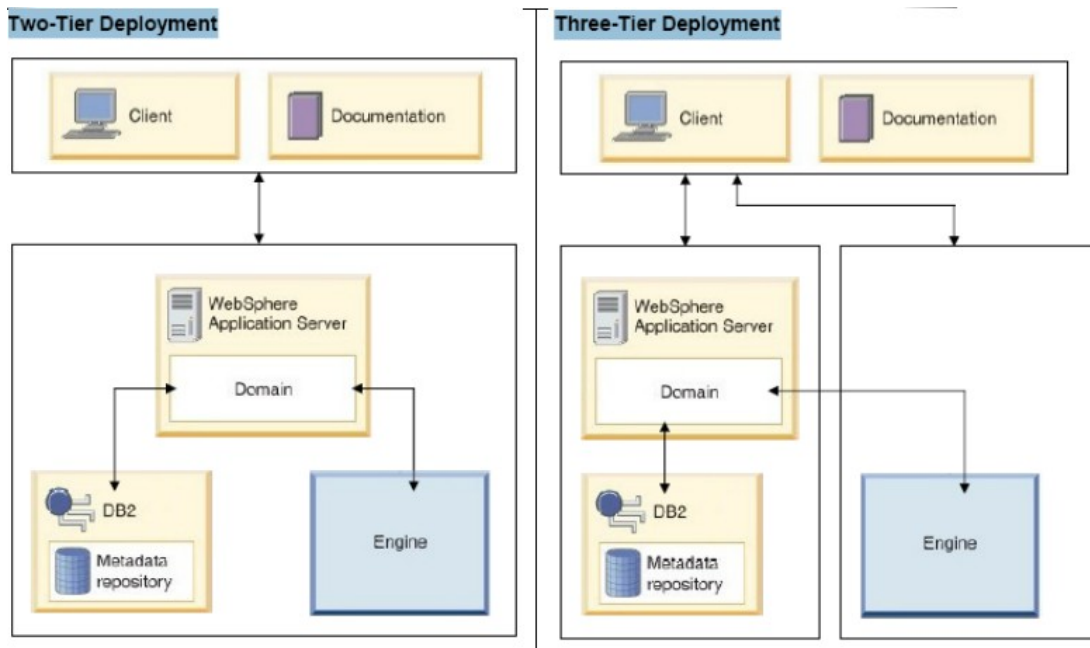


Figura 1. Ejemplos de despliegue basado en capas requeridas por IBM Information Server.

Es importante referir que si se desea utilizar la misma plataforma de HW que hoy en día mantiene los servicios y componentes de Datastage 7.5.x, ésta debe ser validada previamente, ya que la arquitectura de los productos a través de éstas versiones ha evolucionado y por lo tanto los requerimientos de HW y SWG serán diferentes, incluyendo niveles de OS, compiladores, etcétera. . Además se debe confirmar que cualquier equipo utilizado será capaz de procesar el volumen de datos en la ventana de tiempo definida, soportando todas las transformaciones y operaciones requeridos.

Si aún no ha corrido el ejercicio de sizing, asegúrese de ejecutarlo y de que éste incluye la proyección de crecimiento en el volumen de datos a manejar, el cual no sólo se limita a los valores transaccionales o actualizaciones, si no también al crecimiento y actualización de catálogos, archivos de referencia o búsqueda, incluso considere si los repositorios destino estarán almacenando grandes canti-

dades de información histórica y en algún punto del proceso son considerados fuentes de datos, aunque sea de forma indirecta.

Incluya en el dimensionamiento toda la información pertinente a otras actividades que espera ejecutar, incluso si no pertenecen al propósito específico de Datastage. Por ejemplo si considera en el futuro realizar actividades de limpieza de datos, perfilado o alguna integración o despliegue de servicios web. Esto ayudará a que la plataforma propuesta esté preparada para integrar los diversos componentes.

Éste documento parte de la premisa de que la definición de topología ya ha sido definida y aprobada. En caso de que se tengan dudas de las definiciones disponibles revisar el Information Center para Information Server o contactar al representante IBM.

Referencias:

InfoSphere Information Server information roadmap
<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/index.jsp>

3.4. Instalación

La instalación de Information Server es una actividad que requerirá de planeación, control y seguimiento. IBM provee diferentes checklists y guías oficiales de instalación. Esta documentación proveerá los pasos a seguir para una instalación exitosa con base en las definiciones específicas del proyecto.

Referencias:

InfoSphere Information Server information roadmap
<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/index.jsp>

Installing IBM InfoSphere Information Server software
http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.install.doc/topics/cont_iisinfsv_install.html

Como mejores prácticas y recomendaciones a seguir, enlistamos los puntos de referencia con los que los usuarios se encuentran al realizar una instalación.

- Topología. Debe estar claro como será desplegado el software y contar con todos los elementos de HW requeridos o definidos.
- Conectividad. La conectividad en los equipos de la plataforma debe ser asegurada. Deben validarse especialmente si existen firewalls o controles de acceso agregar puertos.
- Usuarios y permisos. Asegurarse de que se cuenta con la lista completa de usuarios con los permisos y roles correspondientes que indica el manual

- Pre-requisitos de Software Confirmar todos los niveles de servicio: Service pack, APAR, releases, Patch, valores de kernel, etcétera. Para todos los componentes:
 - Sistema Operativo.
 - Manejadores de Base de Datos
 - Runtime o compiladores
 - Information Server
 - Websphere Application Server
 - Conectores o Packs (SAP, SIEBEL, etc)
- Asegúrese que todos son compatibles y soportados, ¡La falta de soporte o compatibilidad en alguno de estos puede impedir la instalación y además dejar fuera de soporte su solución!
- Cada Sistema Operativo tendrá diferentes pre-requisitos. Asegúrelos todos Incluso las plataformas Windows requerirán especial atención en este punto. No desestime éste segmento y cúbralo a cabalidad, agilizará la instalación y soporte.
- Confirme si los productos corren a 32bit o 64 bit. Pueden tener diferentes requerimientos o no ser compatibles.
- Ejecutar todos los comandos que solicita el proceso de instalación. En algunas ocasiones se le solicitará ejecutar comandos que modifiquen algunos permisos de instalación o la ejecuten con ciertos atributos. Esto es muy importante para algunas actividades posteriores como la credencialización y asignación de usuarios administradores.
- Si tiene que reiniciar por cualquier motivo la instalación, asegúrese de ejecutar la limpieza correspondiente y de repetir todos los pasos de ambientación indicados (Como el mencionado arriba).
- Confirme que la instalación que está haciendo es la adecuada para su o sus proyectos:
 - ¿Ha validado si su proyecto requerirá Alta Disponibilidad? ¿Ha seleccionado alguna configuración en específico? (Cluster, Activo Pasivo, etc)
 - Si utilizará procesamiento en paralelo requerirá de configuraciones especiales de disco para maximizar los beneficios de esta capacidad. ¿Ha validado esta alternativa?
 - ¿Forma parte de su proceso de instalación y configuración ésta actividad?

- Como cualquier aplicación, el DS se verá afectado positiva o negativamente por la disposición del almacenamiento, esto tomará especial relevancia en casos en donde se decida virtualizar el ambiente y en especial manera para aquellos en que se esté planeando el aprovechamiento de las capacidades especiales de paralelismo y particionamiento.

Es recomendable tener estos puntos cerrados antes de la instalación y configuración, ya que cambios posteriores pueden tener un impacto considerable en los niveles de servicio, y pueden requerir ventanas de mantenimiento a la plataforma, además del costo del tiempo y los recursos destinados para ello.

IBM Information Server 8.5 tiene ahora soporte para alta disponibilidad en Windows, lo que no sucedía con versiones anteriores y la documentación del procedimiento es muy detallada.

Referencias:

New installation road map (high availability active-passive configuration)

http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.ha.install.doc/topics/wsisinst_pln_newinstallationharoadmap.html

3.5. Configuración

Una vez instalada la plataforma, se requieren actividades específicas de configuración. La documentación proveerá la guía para terminar la definición y puesta a punto, entre otros, de:

- Usuarios, credencialización y roles
- Alta disponibilidad
- Paralelismo y particionamiento
- Conectores a fuentes de datos y destinos

Referencias:

Configuring the software after installation

http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.install.doc/topics/cont_iisinfsv_config.html

3.6. Reestablecer los proyectos desde la versión 7.5

El objetivo en esta etapa es preparar el ambiente para poder recrear los procesos de carga que están pendientes de ser importados. Para esta actividad es preciso retomar los archivos de configuración previamente respaldados.

Los pasos a seguir son:

- Importar los objetos (en formato .dsx). Utilizando la utilidad de import en el **DataStage Designer Client**
- Actualice los nuevos archivos de configuración usando como base los archivos que respaldó previamente. Recuerde que tiene también los archivos
 - dsenv
 - .odbc.ini
 - uvodbc.config

Nota: éstos archivos no deben ser sobrescritos, sólo utilizados como guía para crear las nuevas entradas correspondientes.

- Reintegre al proyecto, las carpetas y archivos que forman parte de la definición de los jobs. Copie a la estructura de carpetas los archivos temporales o finales que respaldó. Tenga especial atención en los usuarios y permisos, haciendo las modificaciones correspondientes si hubo cambios. Además verifique que los jobs o procesos puedan leer y si es necesario, modificar los archivos. También es probable que la estructura de directorios se haya modificado, sin embargo recuerde que puede utilizar como referencias relativas la carpeta DSHOME.
- Revisión hashed files. Si está usando hashed files en su proyecto versión 7.5.x debe seguir el proceso de restauración definido, por favor valide en la parte de referencias de ésta sección.

Referencias:

Importing DataStage projects

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/importingdatastageprojects.html>

Restoring contents of DataStage 7.5.3 or earlier settings files

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/restoringcontentsofconfigurationfiles.html>

Restoring DataStage job dependencies

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/restoringjobdependencies.html>

Restoring hashed files

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/restoringhashedfiles.html>

3.7. Recompilar los procesos de carga

Una vez que ha homologado los ambientes y que ha importado todos los objetos. Debe recompilar los jobs. Ésta actividad puede hacerse individualmente o de forma masiva.

Referencias:***Recompiling jobs***

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.productization.iisinfsv.migrate.doc/topics/compilingjobs.html>

3.8. Explotar el nuevo ambiente DataStage

Una vez realizada la migración, puede obtener ventajas inmediatas de su nueva plataforma:

- Reporteo. Revise las capacidades de reporte, le serán útiles para documentar adecuadamente todos los jobs y proyectos. También será útil para actualizaciones y mantenimientos. Asegurese de conocer los templates disponibles, le ahorarán tiempo y esfuerzo.
- Administración de sesiones. Gestione las sesiones desde la web console de Information Server.
- De un paso adelante en gobernabilidad. Obtenga beneficios de contar con metadatos integrados y homologados.6. Migración Server Edition a Enterprise Edition.
- Procesamiento en paralelo (si ses que actualizó a la versión EE).
- Incorporar Blueprint Director para el diseño de arquitecturas para proyectos de información.
- Nuevos Stages y Conectores Server Edition y enterprise Edition.
- Integración SOA.
- Extensión a PACKS y Otras capacidades (DataQuality, Integración de cambios en Tiempo real).
- Utilizar la nueva consola de operaciones en la versión 8.7.
- Utulizar el nuevo debugger por pasos en versión 8.7.

Las anteriores son algunas de las capacidades que puede utilizar.

3.9. Migración Server Edition a Enterprise Edition.

Una vez que la plataforma 8.5 está funcionando,el siguiente paso será obtener los mejores beneficios que el procesamiento en paralelo y particionamiento proveen. Es deseable que los procesos de carga estén en versión Parallel.

Los jobs desarrollados en versión Server no son migrados a través de una utilería a la versión Enterprise o Parallel. Esto es debido a que son interpretados y compilados en diferentes tecnologías, por lo tanto deben ser desarrollados nuevamente usando como base el flujo y reglas previamente creados para el job versión 7.5.

La plataforma 8.5 enterprise Edition permitirá la ejecución, incluso incorporación en secuencias, de jobs server y enterprise, pudiendo los dos tipos afectar a los mismos repositorios y archivos de referencia, excepto aquellos que son de uso exclusivo de los procesos paralelo.

Esto permitirá que contemos con un ambiente heterogéneo en el que algunos jobs sean migrados a Enterprise y otros permanezcan en Server.

El proceso ideal para la migración de estos jobs considera como entrada:

- El análisis de procesos de carga de masa crítica.
- El análisis de procesos de carga con mas áreas de oportunidad, basados en tiempos de respuesta y procesamiento.

Estas dos actividades nos darán el listado de aquellos jobs que la organización desea mantener como prioritarios debido a su aportación al negocio y aquellos que presentan problemas de desempeño, tal vez afectando otros procesos de carga dependientes o comprometiendo las ventanas de tiempo para el procesamiento y entrega, generando retrasos en la publicación de información de otras soluciones, como las de Inteligencia de negocios, entre otras.

Esto es basado en que comúnmente, aunque debe validarse caso por caso, el principio de Pareto aplica para los procesos de carga en diferentes áreas críticas:

- 20% de los procesos de carga tienen el 80% de la información crítica a procesar.
- 20% de los procesos de carga “mueven” el 80% del volumen total de la información.

El proceso de migración puede concentrarse entonces en ese 20% de jobs detectados.

Adicionalmente, estos procesos pueden ser optimizados durante el proceso de rediseño, incorporando nuevos stages, por ejemplo los que ordenan la información (Sort) o los que soportan la administración de datos en catálogos y dimensiones, entre otros.

A continuación se describe en general el objetivo de cada fase del proyecto:

Planeación Estratégica y Análisis: Planear las actividades para el diseño, desarrollo, pruebas y despliegue de los procesos.

Diseño: Se especifica a mayor detalle el ambiente de datos, ambiente de acceso de datos, esquemas de extracción, procesos de mantenimiento y el detalle general del ambiente técnico y mejores prácticas. Se hace un reporte de las interfaces de entrada y salida de los jobs, se analizan los mapeos existentes y se desarrolla el plan maestro de pruebas.

Construcción: Desarrollo de los jobs seleccionados en modo paralelo, siguiendo mejores prácticas especificadas en la fase de Diseño (nomenclatura, particionamiento, desempeño, escalabilidad y parametrización). Pruebas unitarias de los jobs y preparación del procedimiento para pruebas integrales y de sistema.

Pruebas: Esta fase asegura que el los procesos migrados cumplen todas las especificaciones funcionales. Se establece el ambiente para pruebas integrales y se procesa el conjunto de datos de prueba con parámetros de ambiente productivo. Se realizan las pruebas de acuerdo al plan y se aplican los cambios necesarios.

Preparación para Producción: Se asegura que el ambiente del cliente está listo para mover el desarrollo a producción. Se validan los procesos de migración corriendo las secuencias de 7.5 y de 8.5 para verificar que los resultados son consistentes. Se realiza la transferencia de conocimientos del diseño de los jobs y su operación.

3.10. Transferencia de conocimiento

La transferencia de conocimientos no sólo comprende compartir con el cliente la información y metodología usada durante la migración, sino que se debe planear la capacitación de los desarrolladores del cliente con cursos formales de la versión 8.5 de DataStage.

Referencias:

DataStage Integration Developer Training Path:

<http://www-304.ibm.com/jct03001c/services/learning/ites.wss/us/en?pageType=page&c=a0010059>

InfoSphere Training Paths:

<http://www-304.ibm.com/jct03001c/services/learning/ites.wss/us/en?pageType=page&c=a0000430>