

Transcripcions en espanyol i català

Creación de Jobs

«Juan Vidal Gil
Àngels Rius Gavidia»

[Información visual: Sobre la interfaz del *software* Spoon, que se muestra a pantalla completa, la voz en *off* del profesor acompaña las diferentes acciones que va desarrollando.]

Juan Vidal Gil:

Hola, mi nombre es Juan Vidal y soy profesor colaborador de la UOC en el área de Informática, Telecomunicaciones y Multimedia. En este vídeo vamos a ver cómo generar un job en Pentaho Data Integration. Arrancamos, como siempre, la herramienta Spoon para generar este job.

Un job es un proceso que se compone de un conjunto de transformaciones generando la típica cadena de ejecución ETL en la que encadeno un conjunto de tareas. Podríamos decir que un job es un nivel superior a una transformación. Los niveles que tendríamos en un ETL generado por Pentaho serían, en primer lugar, los diferentes pasos que componen una transformación tal como la que tenemos ahora mismo en pantalla.

[Información visual: Varios elementos de una transformación se unen en una cadena.]

Pero, por otro lado, tendríamos las diferentes transformaciones que componen un job, teniendo esta jerarquía de niveles. Por tanto, job, transformación y paso.

Si nos fijamos, tenemos en pantalla dos transformaciones: la transformación “217” y la transformación “214”, que vimos en anteriores vídeos. Y nuestro objetivo podría ser generar un primer job para unir estas dos transformaciones, una ejecución que encadene ambas transformaciones.

Para ello, entramos en el menú, nuevo job.

[Información visual: “File”, “New”, “Job”. En la ventana de transformaciones, se crea una nueva pestaña.]

Si os fijáis, el menú de la izquierda, al generar un nuevo job, ha cambiado. Los pasos que tengo aquí a disposición –o los elementos que tengo a disposición– son los típicos de una cadena de ejecución, mientras que, cuando trabajo a nivel transformación, son los típicos pasos para transformar datos.

Vuelvo al menú de jobs. Mi objetivo es crear un pequeño proceso ETL para unir ambas transformaciones.

Los diferentes elementos de un job los encuentro en las diferentes carpetas que tengo a la izquierda. La carpeta “General” tiene los pasos para iniciar, finalizar e incluir jobs dentro de una transformación, así como dar valor a las variables.

[Información visual: La carpeta “General” incluye los elementos “Start”, “Dummy”, “Job”, “Transformation”, “Set variables” y “Success”.]

Tendríamos, en la carpeta “Mail”, elementos típicos de notificación de mensajes y de errores en un job. La carpeta “File Management” incluye los pasos habituales de gestión de ficheros, borrados, copiados, comprimir, descomprimir. Tendríamos también la carpeta de “Conditions” para poder incluir condiciones de ejecución en un job. Y así podríamos seguir con el resto de carpetas habituales o útiles para generar un job.

[Información visual: El resto de las carpetas son: “Scripting”, “Bulk loading”, “Big Data”, “XML”, “Utility”, “Repository”, “File transfer” y “File encryption”.]

Vamos, por tanto, a generar este primer job. Tenemos que arrastrar el elemento de inicio a la pantalla. Este elemento nos va a permitir generar un punto de arranque en el job.

[Información visual: Arrastra el elemento “Start” del menú lateral izquierdo a la pantalla. Sobre “Start”, botón derecho del ratón, “Edit”. Se abre un cuadro de diálogo llamado “Job Scheduling” con varias opciones de configuración. “Ok”.]

Podemos ver cómo realizar su configuración, que básicamente consiste en poder planificar o no planificar esta ejecución. Nos va a servir de inicio. Y lo siguiente es incluir las transformaciones que componen este job. Incluyo una primera transformación que debo unir desde mi elemento de inicio.

[Información visual: Arrastra el elemento “Transformation” del menú lateral izquierdo a la pantalla. Coloca el cursor sobre el elemento “Start” en la pantalla y selecciona el icono de unión. Una flecha une el elemento “Start” con el elemento “Transformation”.]

Parametrizo esta transformación. Lo que pretendo hacer en esta transformación es incluir la transformación “214”. Por tanto, lo que hago es poner un nombre a esta transformación. Dentro del job, “214”. E incluyo la transformación a la que quiero hacer referencia en este punto. La buscaría, la seleccionaría y la incluiría.

[Información visual: Sobre el elemento “Transformation”, botón derecho, selecciona “Edit”. Se abre un cuadro de diálogo con los detalles de la transformación. En la parte superior del cuadro de diálogo, “Name of job entry”, escribe “214”. A continuación, en “Transformation specification”, selecciona el botón del extremo derecho, selecciona el archivo “IN_214.ktr”, “Open”, “Ok”.]

De esta manera, he incluido una transformación en el job.

Si quiero incluir una segunda transformación, operaría de igual modo. Añado una transformación. La uno con la anterior, de forma similar a cómo uníamos los pasos de una transformación. E incluyo la segunda transformación. En este caso, “217”. Selecciono e incluyo la transformación.

[Información visual: Arrastra un segundo elemento “Transformation” del menú lateral izquierdo a la pantalla. Coloca el cursor sobre la transformación “214” y selecciona el icono de unión. Una flecha une el elemento “214” con el nuevo elemento “Transformation”. Sobre este elemento, botón derecho, “Edit”. Se abre el mismo cuadro de diálogo anterior con los detalles de la transformación. En la parte superior del cuadro de diálogo, “Name of job entry”, escribe “217”. A continuación, en “Transformation

specification”, selecciona el botón del extremo derecho, selecciona el archivo “IN_217.ktr”, “Open”, “Ok”.]

He encadenado dos transformaciones. Podría añadir diferentes elementos que incluyeran condiciones, gestión de ficheros, etcétera. En este caso, simplemente, voy a añadir el elemento de finalización, indicando el caso de finalización correcta, al que uno desde la última transformación incluida.

[Información visual: Arrastra el elemento “Success” del menú lateral izquierdo a la pantalla. Coloca el cursor sobre la transformación “217” y selecciona el icono de unión. Una flecha une el elemento “217” con el elemento “Success”.]

Sería un job muy sencillo en el que solamente incluyo dos transformaciones.

Cuando salvo este job, veo ahora que me va a salvar el job con otro tipo de extensión. La extensión típica de un job es .kjb. Y la extensión típica de una transformación sería .ktr.

[Información visual: “File”, “Save as”. Salva el archivo con el nombre “job_214_217” y como tipo “Kettle jobs”.]

Puedo ver, luego, a nivel de archivos, como tienen diferentes extensiones. Aquí es un job de Kettle. Salvo. Tengo el job salvado. Puedo ver en ejecución cómo se ejecuta un job y cómo puedo ver los resultados de esa ejecución.

[Información visual: Comprueba la ejecución a través del icono “Play” en el menú horizontal superior. Se abre un cuadro de diálogo. Selecciona el botón “Launch” en la parte inferior.]

Me aparece algo similar a la ventana de resultados. Aquí lo que voy viendo es la ejecución, cómo va desencadenando las diferentes transformaciones, cómo las va arrancando y cómo me va notificando la finalización del mismo. Tengo algo similar a la ventana de resultados, pero aquí estoy trabajando ya a un nivel de job. Voy viendo el resultado de cada una de las transformaciones que componen el job. Tendría, también, una ventana de “Log” para ver la ejecución a un nivel más detallado. También sería posible.

[Información visual: La ventana de ejecución, “Execution Results”, cuenta con cuatro pestañas en la parte superior: “History”, “Logging”, “Job metrics” y “Metrics”.]

De este modo, he ejecutado un primer job en Pentaho. Puedo acceder a un job más completo.

[Información visual: Muestra el ejemplo de un job más complejo, formado por más elementos.]

Incluiría un conjunto mayor de transformaciones, que podría ser un job para cargar un Data Warehouse en el cual estoy incluyendo dimensiones, estoy incluyendo hechos y diferentes tablas adicionales, tablas auxiliares. Puedo ver que aquí ya tengo todo un conjunto de transformaciones más amplio, que las tengo clasificadas de acuerdo con lo que estoy cargando en cada uno de los puntos, llegando, al final, a un punto de finalización. El job podría contener, también, elementos de condiciones, podría tener

elementos de gestión de ficheros, notificación, etcétera, pero esto es un ejemplo de un primer job de Pentaho en el cual estamos cargando el típico Data Warehouse con dimensiones, hechos, tablas auxiliares, etcétera.

Esperamos os haya sido de utilidad este vídeo. Muchas gracias por vuestra atención. Un saludo.

Creació de Jobs

«Juan Vidal Gil
Àngels Rius Gavidia»

[Informació visual: Sobre la interfície del *software* Spoon, que es mostra a pantalla completa, la veu en *off* del professor acompanya les diferents accions que duu a terme.]

Juan Vidal Gil:

Hola, em dic Juan Vidal i soc professor col·laborador de la UOC a l'àrea d'Informàtica, Telecomunicacions i Multimèdia. En aquest vídeo veurem com generar un job a Pentaho Data Integration. Com sempre, arranquem l'eina Spoon per generar aquest job.

Un job és un procés que es compon d'un conjunt de transformacions i que genera la típica cadena d'execució ETL en què encadenen un conjunt de tasques. Podríem dir que un job és un nivell superior a una transformació. Els nivells que tindríem en un ETL generat per Pentaho serien, en primer lloc, els diferents passos que formen una transformació tal com la que tenim ara mateix a la pantalla.

[Informació visual: Diversos elements d'una transformació s'uneixen en una cadena.]

Però, d'altra banda, tindríem les diferents transformacions que formen un job, amb aquesta jerarquia de nivells. Per tant, job, transformació i pas.

Si ens hi fixem, tenim a la pantalla dues transformacions: la transformació "217" i la transformació "214", que hem vist en vídeos anteriors. I el nostre objectiu podria ser generar un primer job per unir aquestes dues transformacions, una execució que encadeni ambdues transformacions.

Per fer-ho, entrem al menú, nou job.

[Informació visual: "File", "New", "Job". A la finestra de transformacions, es crea una nova pestanya.]

Si us hi fixeu, quan generem un job nou, el menú de l'esquerra canvia. Els passos que tinc aquí a disposició —o els elements que tinc a disposició— són els típics d'una cadena d'execució, mentre que, quan treballo amb transformacions, són els passos típics per transformar dades.

Torno al menú de jobs. El meu objectiu és crear un petit procés ETL per unir ambdues transformacions.

Els diferents elements d'un job els trobo a les diferents carpetes que tinc a l'esquerra. La carpeta "General" té els passos per iniciar, finalitzar i incloure jobs dins d'una transformació, així com donar valor a les variables.

[Informació visual: La carpeta "General" inclou els elements "Start", "Dummy", "Job", "Transformation", "Set variables" i "Success".]

A la carpeta "Mail", tindríem elements típics de notificació de missatges i d'errors en un job. La carpeta "File Management" inclou els passos habituals de gestió de fitxers,

esborrats, copiats, comprimir, descomprimir. També tindríem la carpeta de “Conditions” per incloure condicions d’execució en un job. I així podríem seguir amb la resta de carpetes habituals o útils per generar un job.

[Informació visual: La resta de carpetes són: “Scripting”, “Bulk loading”, “Big Data”, “XML”, “Utility”, “Repository”, “File transfer” i “File encryption”.]

Per tant, generem aquest primer job. Hem d’arrossegar l’element d’inici a la pantalla. Aquest element ens permetrà generar un punt d’arrancada al job.

[Informació visual: Arrossega l’element “Start” del menú lateral esquerre a la pantalla. Sobre “Start”, botó dret del ratolí, “Edit”. S’obre un quadre de diàleg que es diu “Job Scheduling” i que inclou diverses opcions de configuració. “Ok”.]

Podem veure com realitzar-ne la configuració, que bàsicament consisteix a planificar o no aquesta execució. Ens servirà d’inici. El següent és incloure les transformacions que formen aquest job. Incloc una primera transformació que he d’unir des del meu element d’inici.

[Informació visual: Arrossega l’element “Transformation” del menú lateral esquerre a la pantalla. Col·loca el cursor sobre l’element “Start” a la pantalla i selecciona la icona d’unió. Una fletxa uneix l’element “Start” amb l’element “Transformation”.]

Parametrizo aquesta transformació. El que pretenc fer en aquesta transformació és incloure la transformació “214”. Per tant, poso un nom a aquesta transformació. Dins del job, “214”. Incloc la transformació a què vull fer referència en aquest punt. La buscaria, la seleccionaria i la inclouria.

[Informació visual: Sobre l’element “Transformation”, botó dret, selecciona “Edit”. S’obre un quadre de diàleg amb els detalls de la transformació. A la part superior del quadre de diàleg, “Name of job entry”, escriu “214”. Tot seguit, a “Transformation specification”, selecciona el botó de l’extrem dret, selecciona l’arxiu “IN_214.ktr”, “Open”, “Ok”.]

D’aquesta manera, he inclòs una transformació al job.

Si vull incloure una segona transformació, operaria de la mateixa manera. Afegeixo una transformació. La uneixo amb l’anterior, de forma semblant a com uníem els passos d’una transformació. Incloc la segona transformació. En aquest cas, “217”. Selecciono i incloc la transformació.

[Informació visual: Arrossega un segon element “Transformation” del menú lateral esquerre a la pantalla. Col·loca el cursor sobre la transformació “214” i selecciona la icona d’unió. Una fletxa uneix l’element “214” amb el nou element “Transformation”. Sobre aquest element, botó dret, “Edit”. S’obre el mateix quadre de diàleg anterior amb els detalls de la transformació. A la part superior del quadre de diàleg, “Name of job entry”, escriu “217”. Tot seguit, a “Transformation specification”, selecciona el botó de l’extrem dret, selecciona l’arxiu “IN_217.ktr”, “Open”, “Ok”.]

He encadenat dues transformacions. Podria afegir diferents elements que incloguessin condicions, gestió de fitxers, etcètera. En aquest cas, simplement, afegiré l’element de finalització, indicant el cas de finalització correcta, i l’uniré a l’última transformació inclosa.

[Informació visual: Arrossega l'element "Success" del menú lateral esquerre a la pantalla. Col·loca el cursor sobre la transformació "217" i selecciona la icona d'unió. Una fletxa uneix l'element "217" amb l'element "Success".]

Seria un job molt senzill en el qual només incloc dues transformacions.

Quan deso aquest job, veig que em desarà el job amb un altre tipus d'extensió. L'extensió típica d'un job és .kjb. I l'extensió típica d'una transformació seria .ktr.

[Informació visual: "File", "Save as". Desa l'arxiu amb el nom "job_214_217" i com a tipus "Kettle jobs".]

Puc veure, pel que fa als arxius, que tenen extensions diferents. Aquí és un job de Kettle. Deso. Tinc el job desat. Puc veure en execució com s'executa un job i com puc veure els resultats d'aquesta execució.

[Informació visual: Comprova l'execució a través de la icona "Play" del menú horitzontal superior. S'obre un quadre de diàleg. Selecciona el botó "Launch" de la part inferior.]

M'apareix quelcom semblant a la finestra de resultats. Aquí el que veig és l'execució, com va desencadenant les diferents transformacions, com les va arrancant i com em va notificant la finalització del job. Tinc quelcom semblant a la finestra de resultats, però aquí ja treballo amb el job. Vaig veient el resultat de cadascuna de les transformacions que formen el job. També tindria una finestra de "Log" per veure l'execució de forma més detallada. També seria possible.

[Informació visual: La finestra d'execució, "Execution Results", disposa de quatre pestanyes a la part superior: "History", "Logging", "Job metrics" i "Metrics".]

D'aquesta manera, he executat un primer job a Pentaho. Puc accedir a un job més complet.

[Informació visual: Mostra l'exemple d'un job més complex, format per més elements.]

Inclouria un conjunt més gran de transformacions, que podria ser un job per carregar un Data Warehouse en el qual incloc dimensions, fets i diferents taules addicionals, taules auxiliars. Puc veure que aquí ja tinc tot un conjunt de transformacions més ampli. Les tinc classificades d'acord al que carrego en cadascun dels punts. I, al final, arribo a un punt de finalització. El job podria contenir, també, elements de condicions, podria tenir elements de gestió de fitxers, notificació, etcètera, però això és un exemple d'un primer job de Pentaho en què carreguem el típic Data Warehouse amb dimensions, fets, taules auxiliars, etcètera.

Esperem que aquest vídeo us hagi estat útil. Moltes gràcies per la vostra atenció. Una salutació.