

# ¿Qué ofrece Autentia Real Business Solutions S.L?

Somos su empresa de **Soporte a Desarrollo Informático**.  
 Ese apoyo que siempre quiso tener...

## 1. Desarrollo de componentes y proyectos a medida



## 2. Auditoría de código y recomendaciones de mejora

## 3. Arranque de proyectos basados en nuevas tecnologías

1. Definición de frameworks corporativos.
2. Transferencia de conocimiento de nuevas arquitecturas.
3. Soporte al arranque de proyectos.
4. Auditoría preventiva periódica de calidad.
5. Revisión previa a la certificación de proyectos.
6. Extensión de capacidad de equipos de calidad.
7. Identificación de problemas en producción.



## 4. Cursos de formación (impartidos por desarrolladores en activo)

Spring MVC, JSF-PrimeFaces /RichFaces,  
 HTML5, CSS3, JavaScript-jQuery

Gestor portales (Liferay)  
 Gestor de contenidos (Alfresco)  
 Aplicaciones híbridas

Tareas programadas (Quartz)  
 Gestor documental (Alfresco)  
 Inversión de control (Spring)

Control de autenticación y  
 acceso (Spring Security)  
 UDDI  
 Web Services  
 Rest Services  
 Social SSO  
 SSO (Cas)

JPA-Hibernate, MyBatis  
 Motor de búsqueda empresarial (Solr)  
 ETL (Talend)

Dirección de Proyectos Informáticos.  
 Metodologías ágiles  
 Patrones de diseño  
 TDD

BPM (jBPM o Bonita)  
 Generación de informes (JasperReport)  
 ESB (Open ESB)

**AdictosAlTrabajo**

Terrakas 1x04  
¡Estreno 6 de septiembre!  
terrakas.com



**autentia**  
Soporte a desarrollo informático  
Hosting patrocinado por  
**enredados**

Entra en Adictos a través de  

E-mail

Contraseña

Entrar

Deseo registrarme  
Olvidé mi contraseña

[Inicio](#) [Quiénes somos](#) [Formación](#) [Comparador de salarios](#) [Nuestro libro](#) [Más](#)

» Estás en: [Inicio](#) [Tutoriales](#) Lectura y tratamiento de ficheros XML con Talend



Daniel Casanova Frutos

Consultor tecnológico de desarrollo de proyectos informáticos.

Ingeniero Técnico En Informática De Sistemas por la Universidad Alfonso X El Sabio.

Puedes encontrarme en [Autentia](#): Ofrecemos servicios de soporte a desarrollo, factoría y formación

Somos expertos en Java/J2EE

[Ver todos los tutoriales del autor](#)



**Fecha de publicación del tutorial: 2012-09-04**

Tutorial visitado 4 veces [Descargar en PDF](#)

## Lectura y tratamiento de ficheros XML con Talend.

### 0. Índice de contenidos.

- 1. Introducción.
- 2. Entorno.
- 3. Lectura y tratamiento de ficheros XML.
- 4. Referencias.
- 5. Conclusiones.

### 1. Introducción

En este tutorial veremos como la herramienta **Talend** nos permite de manera sencilla extraer, manipular y cargar la información de ficheros **XML**.

Talend, como ya vimos en el tutorial anterior 'Lectura y tratamiento de ficheros Excel con Talend (I): nociones básicas.', es una herramienta de diseño **ETL (Extract, Transform, Load)**, que permite manipular la información de distintas fuentes de datos.

En este tutorial nos centramos en el tratamiento de información sobre ficheros del tipo **XML** (eXtensible Markup Language).

### 2. Entorno.

El tutorial está escrito usando el siguiente entorno:

- Hardware: Portátil MacBook Pro 15' (2.2 GHz Intel Core i7 Duo, 8GB DDR3 SDRAM).
- Sistema Operativo: Mac OS X Lion 10.7.3 (11D50d)
- Talend Open Studio V5.1.1 r84309

### 3. Lectura y tratamiento de ficheros XML

Como ya sabemos XML es un Lenguaje de Etiquetado Extensible simple que permite el intercambio de una gran cantidad de datos.

A menudo es tedioso el proceso de parseo de este tipo de ficheros, por lo que Talend nos ofrece la posibilidad de realizar dicha labor de manera sencilla, sin necesidad de conocer el API de distintas librerías, ni tener que hacer peripecias para obtener los nodos del fichero así como la información que representan.

Para manipular ficheros xml, Talend nos ofrece el componente de entrada de ficheros **tFileInputXml**.

En este tutorial veremos un sencillo de ejemplo de manipulación de un fichero XML, así como el uso completo de dicho componente.

Partimos de un fichero xml con el siguiente contenido:

```

1  <?xml version="1.0" encoding="ISO-8859-1" ?>
2  <root>
3    <cabecera>Listado de productos</cabecera>
4    <producto>
5      <nombre>Pantalla</nombre>
6      <descripcion>Pantalla LED para ordenador</descripcion>
7      <precio>
8        <moneda>EUR</moneda>
9        <valor>100</valor>

```

### Catálogo de servicios Autentia



### Síguenos a través de:



### Últimas Noticias

- » Estreno Terrakas 1x04: "Terraka por un día"
- » Nuevos cursos de gestión de la configuración en IOS y Android
- » La regla del Boy Scout y la Oxidación del Software
- » Autentia conquista los Alpes
- » Orientación a objetos y la importancia del "Tell, Don't Ask"

[Histórico de noticias](#)

### Últimos Tutoriales

- » Despliegando una aplicación en Cloud Foundry con Maven
- » Primeros pasos en Android (II)
- » Eclipse Juno, la versión 4.2 de Eclipse
- » Double Opt-In y autologin con el soporte de Spring MVC y Spring Security.
- » Posicionamiento de componentes en HTML con el soporte de CSS.

Impulsores Comunidad ¿Ayuda?

0 personas han traído clicks a esta página

sin clicks + + + + + + + +

powered by [karmacacy](#)

```

10     </precio>
11   </producto>
12   <producto>
13     <nombre>Teclado</nombre>
14     <descripcion>Teclado para ordenador por USB</descripcion>
15     <precio>
16       <moneda>EUR</moneda>
17       <valor>23</valor>
18     </precio>
19   </producto>
20   <producto>
21     <nombre>Ratón</nombre>
22     <descripcion>Ratón ordenador portatil</descripcion>
23     <precio>
24       <moneda>EUR</moneda>
25       <valor>10</valor>
26     </precio>
27   </producto>
28 </root>
29

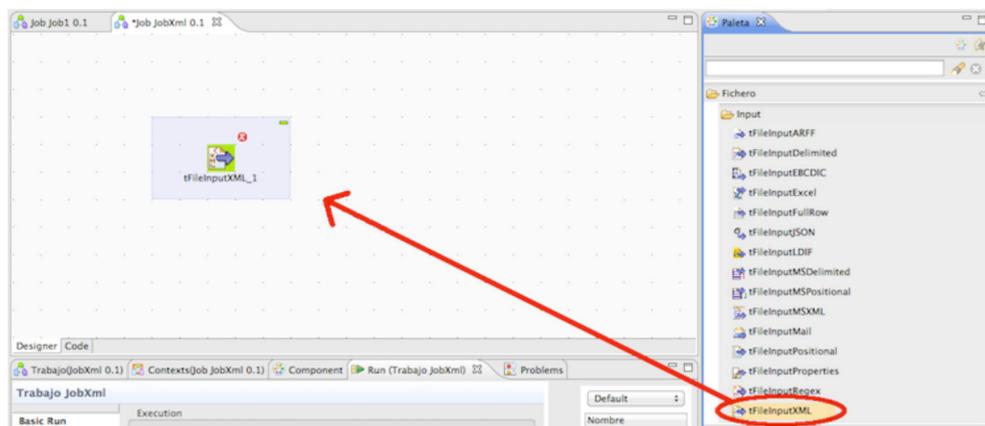
```

Como vemos, es una representación sencilla de una lista de productos, así como la descripción de los mismos, su precio y la moneda.

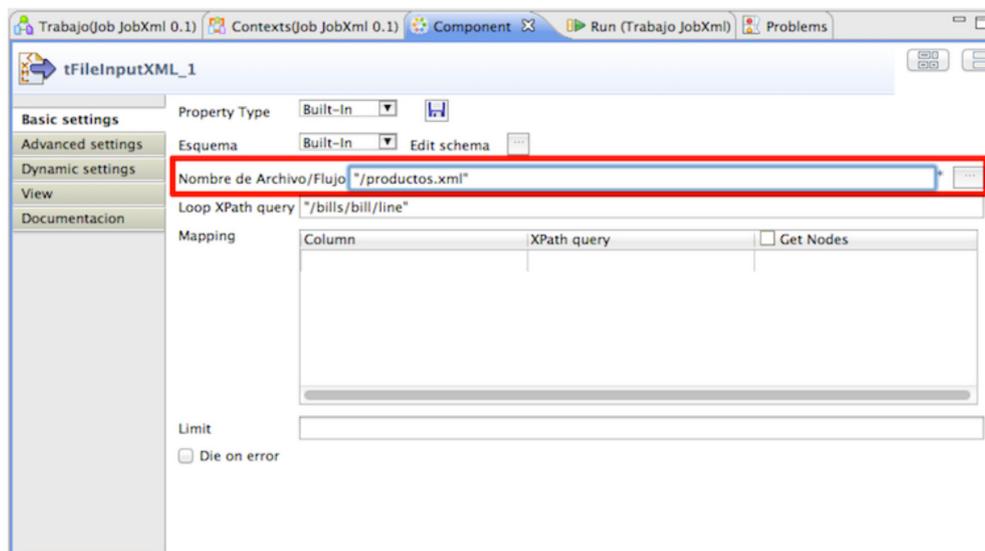
Lo primero de todo es crear un nuevo trabajo en el entorno de Talend. Para ello, como ya hemos visto con anterioridad, realizamos click derecho sobre **'Job Design'** para crear el trabajo, le damos un nombre y pulsamos en finalizar.

Ya tenemos listo nuestro entorno de desarrollo en Talend.

Seleccionamos de la paleta de componentes el componente **tFileInputXml** y lo arrastramos a nuestro diagrama de trabajo:



Lo primero es configurar nuestro componente para que conozca la ruta física donde está el fichero xml. Para ello configuramos la propiedad **'Nombre del Archivo/Flujo'**, en nuestro caso el fichero se llama *productos.xml*:



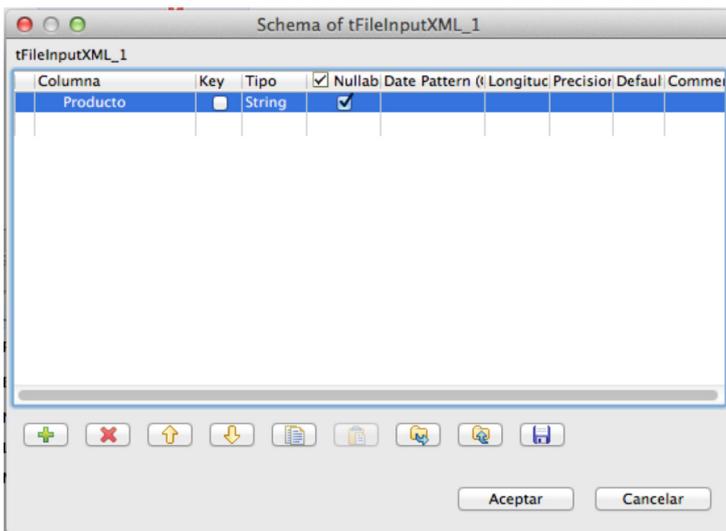
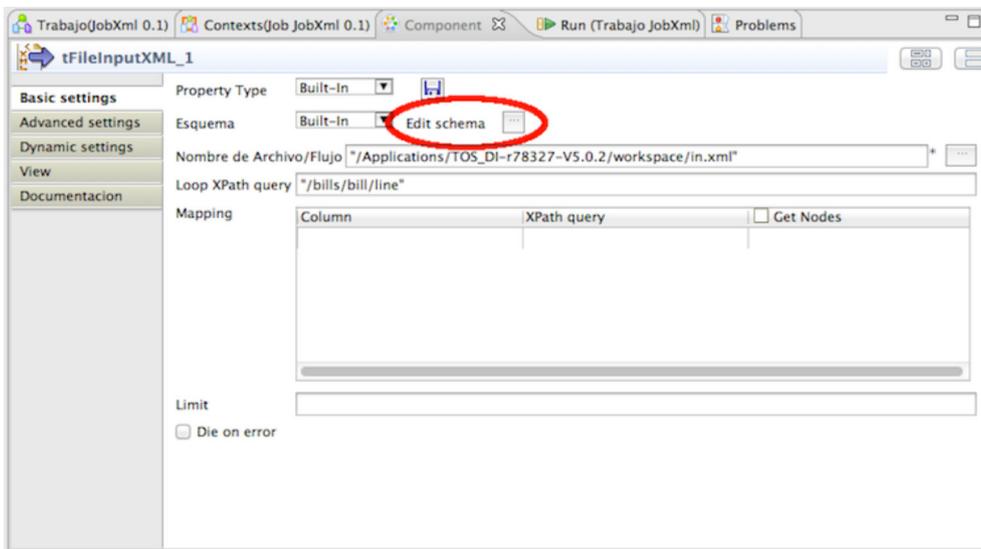
A continuación editamos su esquema mediante la opción **'Edit schema'**, y creamos una columna *'Producto'* que representará la información completa de cada nodo *Producto*:

## Últimos Tutoriales del Autor

- » Lectura y tratamiento de ficheros Excel con Talend (I): nociones básicas.
- » JMeter. Uso de funciones.
- » JMeter. Gestión de usuarios
- » JMeter y JSF. Extracción del parámetro ViewState
- » Instalación de Glassfish 2.1

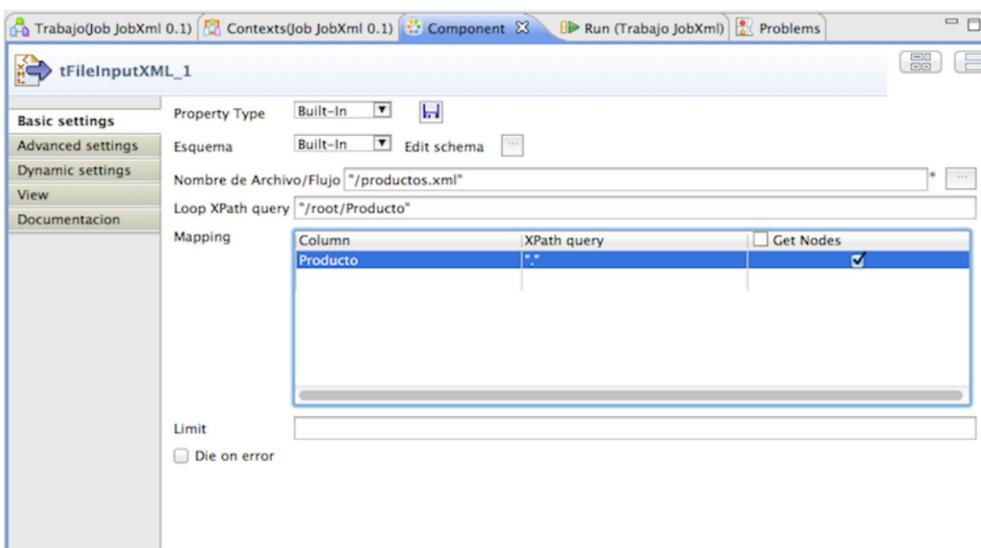
## Últimas ofertas de empleo

- 2011-09-08  
Comercial - Ventas - MADRID.
- 2011-09-03  
Comercial - Ventas - VALENCIA.
- 2011-08-19  
Comercial - Compras - ALICANTE.
- 2011-07-12  
Otras Sin catalogar - MADRID.
- 2011-07-06  
Otras Sin catalogar - LUGO.



El siguiente paso es configurar el componente para que conozca cual es el nodo a partir del cual debe leer, y que nodo representa cada Producto. Para navegar por los distintos nodos del fichero xml, el componente utiliza **Xpath** (<http://www.w3schools.com/xpath>). Xpath (Xml path language) se trata de un lenguaje que permite construir sentencias que recorren y procesan un documento xml. Podemos ver tutoriales muy interesantes ya publicados en <http://www.adictosaltrabajo.com>

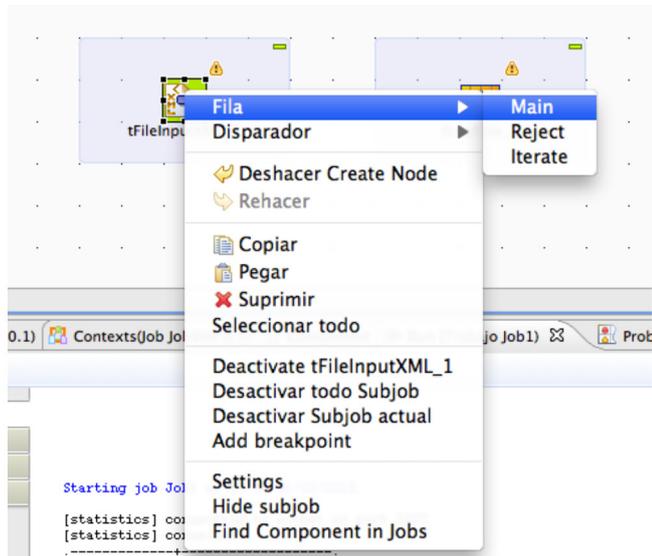
Aplicamos la siguiente configuración:



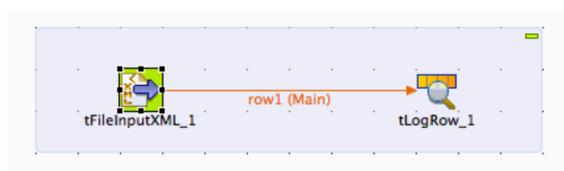
- **Loop XPath query:** Nodo padre a partir del cual leer la información del xml aplicando Xpath.
- **Mapping:** Se define cada columna del flujo de salida y su correspondencia con Xpath. Aparte podemos marcar el check 'Get Nodes' para que arrastre la información de todos los nodos hijos o de niveles inferiores.

Por ejemplo hemos configurado que la columna del flujo de salida 'Producto' serán todos los nodos (expresión Xpath '.') a partir del nodo definido en 'Loop XPath query', es decir a partir del nodo root/Producto.

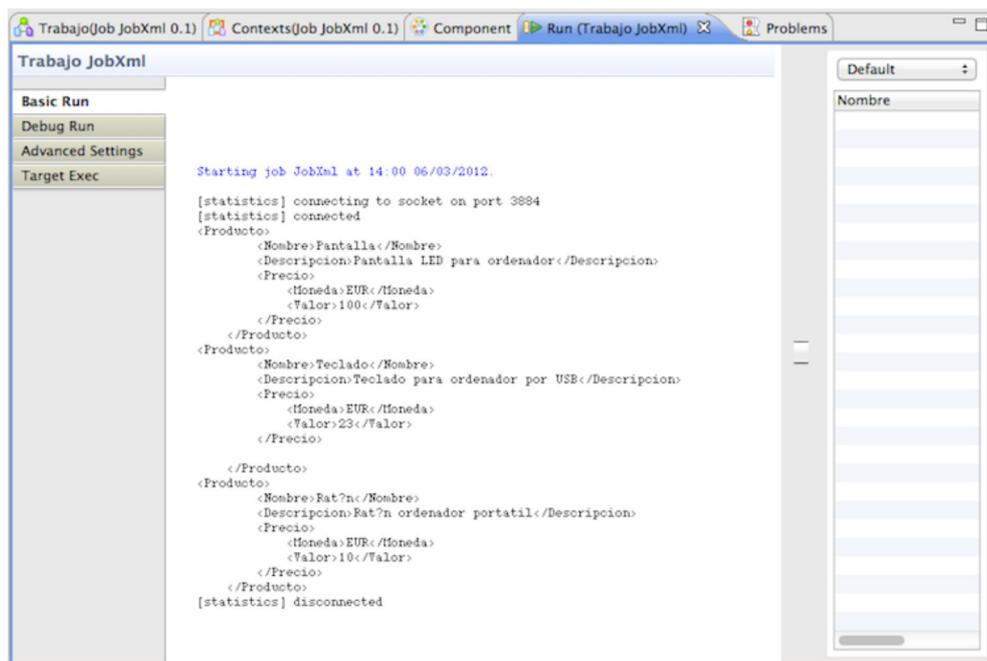
Añadimos un componente de log 'tLogRow' y dirigimos el flujo principal del componente tFileInputXML hacia el componente tLogRow:



Resultando el esquema:

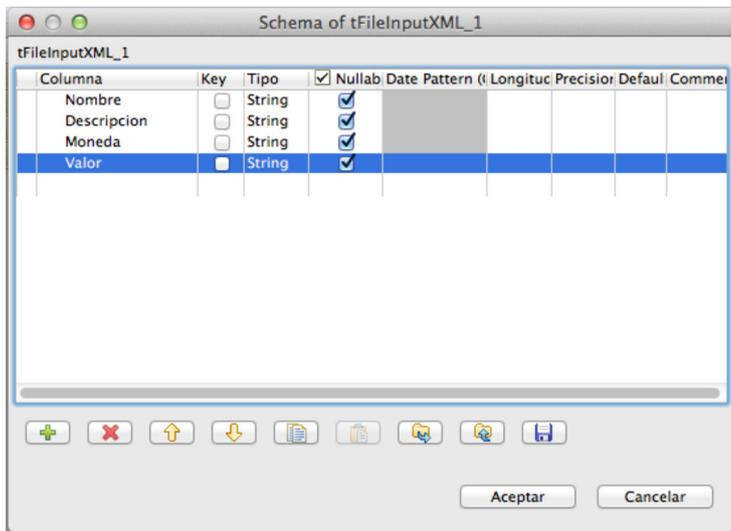


Al ejecutar el trabajo mediante el botón 'Play' obtenemos el siguiente resultado:

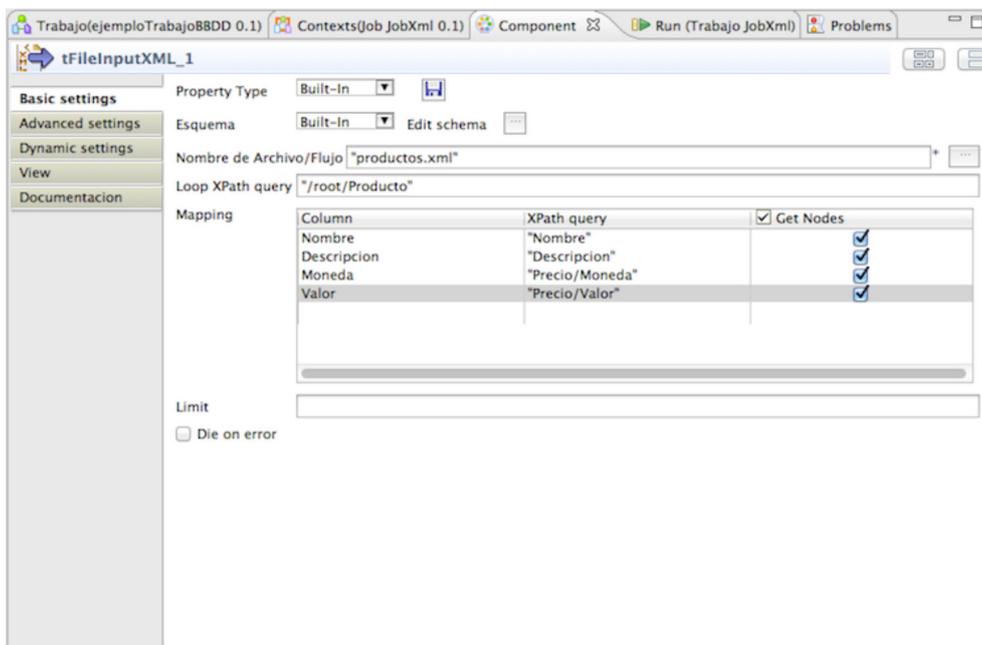


Como podemos ver nos ha devuelto la información de cada nodo Producto del fichero xml.

Ahora mediante **Xpath** obtendremos la información de los nodos que forman un producto en distintas columnas del flujo de salida. Para ello editamos el esquema del componente añadiendo las columnas correspondientes:



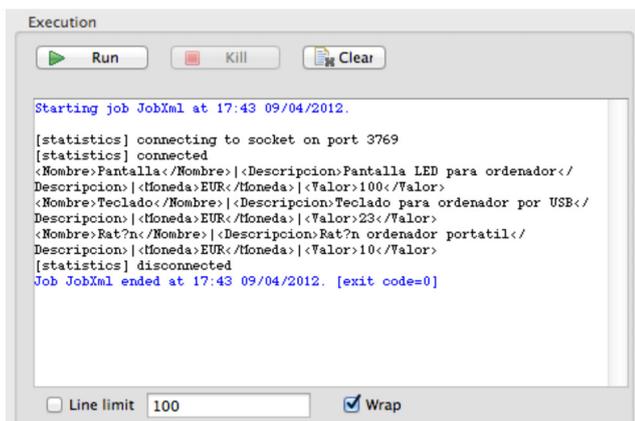
En las preferencias del componente establecemos las expresiones Xpath adecuadas para recuperar la información de cada nodo y reflejarla en cada columna del flujo de salida:



Analizando las expresiones Xpath de cada columna:

- **Nombre:** Nodo nombre a partir del nodo padre definido `/root/Producto`
- **Descripción:** Nodo nombre a partir del nodo padre definido `/root/Producto`
- **Moneda:** Obtener el nodo 'Moneda' que es hijo del nodo 'Precio' (devolverá la información del nodo `/root/Producto/Moneda/Valor`)
- **Valor:** Obtener el nodo 'Valor' que es hijo del nodo 'Precio' (devolverá la información del nodo `/root/Producto/Precio/Valor`)

A continuación al ejecutar el trabajo podemos observar que obtenemos la información adecuada de cada uno de los nodos definidos en el componente:

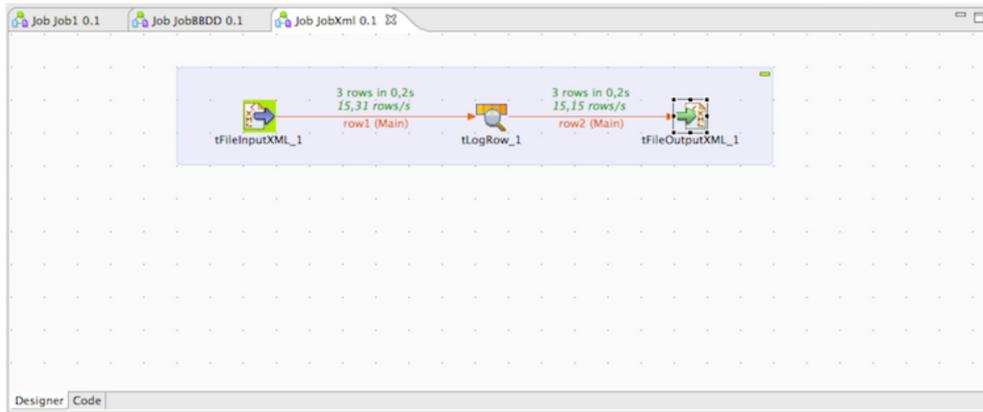


Talend nos permite volcar el flujo de información proveniente de un fichero a otro fichero con la misma u otra tipología que el fichero origen.

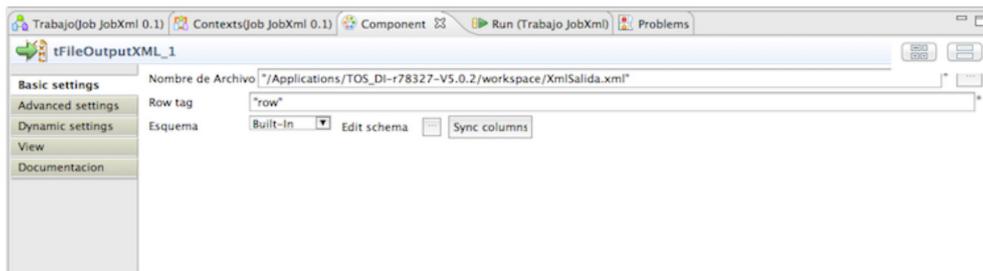
En nuestro ejemplo utilizamos el componente **tFileOutputXml**, el cual tiene un funcionamiento similar al componente visto en

tutoriales anteriores **tFileOutputExcel**.

Ponemos dicho componente en nuestra zona de trabajo y dirigimos el flujo de salida del componente tLogRow hacia el nuevo componente:



Configuramos el componente tFileOutputXML:



Básicamente indicamos el nombre y ruta del fichero de salida. Por otro lado podemos ver que dicho componente permite editar su esquema de salida.

Por otro lado si pulsamos en las preferencias sobre la pestaña '**Advanced Settings**' en las propiedades del componente, podemos observar que el componente tFileOutputXml ofrece propiedades interesantes, como por ejemplo para crear el directorio de la ruta física de destino sino existe, o como dividir la información en distintos ficheros Xml.

Trás ejecutar el trabajo vemos como se crea de manera adecuada el fichero **XmlSalida.xml**

En resumen Talend nos permite obtener y manipular de manera sencilla el contenido de un fichero XML. En este ejemplo hemos visto de manera detallada el uso de ciertos componentes para el tratamiento de información proveniente de ficheros XML; aunque no son los únicos. La lista completa:

#### XML

	tAdvancedFileOutputXML	Creates an XML structure from rows broken down into fields
	tDTDValidator	Validates the XML input file against a DTD file
	tEDIFACTtoXML	Generates a XML file from an EDIFACT file
	tExtractXMLField	Parses input row fields as tFileInputXML
	tFileInputXML	Extracts rows broken down into fields from an XML structured file
	tFileOutputXML	Creates an XML structure from rows broken down into fields
	tWriteXMLField	Creates an XML structure from rows broken down into fields
	tXMLMap	Allows joins, column or row filtering, transformations, and multiple outputs
	tXSDValidator	Validates XML with XSD
	tXSLT	Transforms XML with XSL

En posteriores tutoriales veremos de manera más detallada el funcionamiento de cada uno de ellos.

Estos son sólo algunos ejemplos para manejar la información un tipo de fichero en concreto, aunque Talend nos provee de una infinidad de componentes para manipular la información desde un fichero cualquiera sea su tipología.

#### 4. Referencias.

- <http://www.talendforge.org/components/>

#### 5. Conclusiones.

Como podemos ver la herramienta **TALEND** nos permite de manera sencilla tratar la información que contienen ficheros XML. Seguiremos aconsejando esta herramienta debido a su licencia libre y a su facilidad de uso.

Espero os haya servido de ayuda. Un saludo.

Daniel Casanova

[dcasanova@autentia.com](mailto:dcasanova@autentia.com)

### A continuación puedes evaluarlo:

[Regístrate para evaluarlo](#)

### Por favor, vota +1 o compártelo si te pareció interesante

Share |

0

Animate y coméntanos lo que pienses sobre este **TUTORIAL**:



» [Regístrate](#) y accede a esta y otras ventajas «



Esta obra está licenciada bajo [licencia Creative Commons de Reconocimiento-No comercial-Sin obras derivadas 2.5](#)

Copyright 2003-2012 © All Rights Reserved | [Texto legal y condiciones de uso](#) | [Banners](#) | [Powered by Autentia](#) | [Contacto](#)

W3C XHTML 1.0

W3C CSS

XML RSS

XML RDF