# **AYI ACADEMY: TALEND**

# Material práctico



Módulo 1: Hello Talend

Módulo 2: Best Practices

# Hola Talend!!!

En esta parte explicaremos los aspectos típicos de como crear un Job de tipo DI (Data Integration) en Talend Studio.

Y también veremos como ejecutar Jobs, revisar los logs y como obtener mas ayuda en los componentes que usamos.

Después del siguiente ejercicio, usted será capaz de hacer lo siguiente: crear un proyecto

- Crear un Job.
- Agregar, conectar y configurar componentes.
- Ejecutar un job y visualizar los resultados en consola (ver los logs).
- Obtener ayuda.
- Rediseñar un job siguiendo las mejores prácticas.

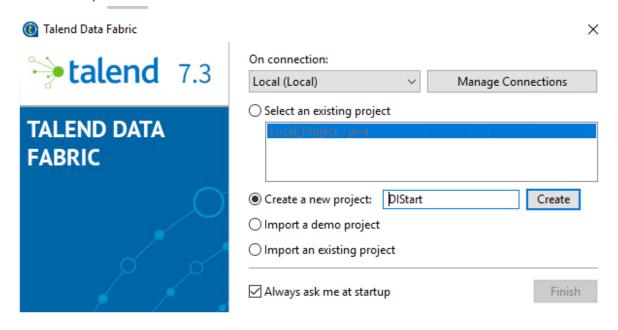
• Generar y visualizar la documentación con HTML.

### Creando un proyecto:

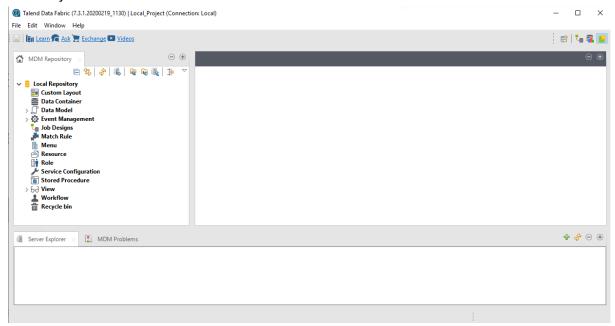
1) Para empezar abrir el Talend Studio se abrirá un ventana que permitirá



- 2) abrir un proyecto existente o crear uno nuevo
- 3) En este caso empezamos de cero por ende vamos a crear un proyecto vacio (nuevo)
  - a) Para la parte que dice "On connection" seleccionamos "Local (Local)" esto significa que el proyecto se va a almacenar de forma local en tu computadora.
  - b) Luego, seleccionamos "Create a new project" y a la derecha ponemos el nombre de nuestro proyecto que va a ser "DIStart"
  - c) Click en "Create" y esperamos hasta que el proyecto aparezca en la lista. Al ser el único proyecto es va aparecer seleccionado por defecto.
  - d) Por último hacemos click en el botón de "Finish"



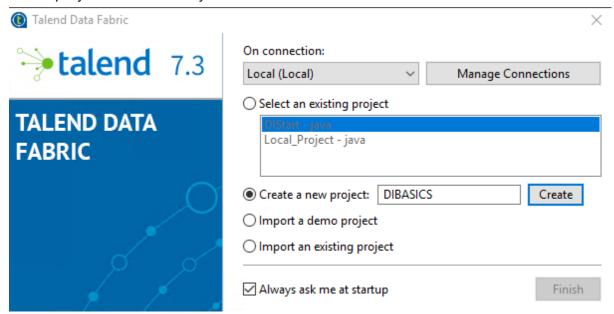
Como podrás ver, la interfaz principal de Talend Studio se lanza con la perspectiva MDM. Ahora tu proyecto está vacío, entonces solo podrás ver el repositorio. Dicho repositorio aloja todas las cosas que vayas a crear en tu proyecto como por ejemplo: Jobs, conexiones a bases de datos, metadatos y más. Nota: por defecto el repositorio está alojado de forma local.



4) Cambiar de proyecto o de espacio de trabajo (workspace) En el menú archivo, hacer click sobre "Switch Project or Workspace". Talend se cerrará y luego automáticamente se volverá a abrir permitiendo que puedas seleccionar otro proyecto. Durante este curso tu proyecto será "DIBASICS"

### Configurando un proyecto

 Seguir los mismos pasos que se describieron con anterioridad para crear el proyecto "DIBASICS" y hacer click en el botón Finish



2) Cambiar a la perspectiva de integration

Talend Studio te permite crear varios tipos de aplicaciones, como por ejemplo integración de datos, Big Data, aplicaciones de integración, manejo de datos maestros y Data Quality.

Dependiendo del tipo de aplicación, puedes trabajar con una o más perspectivas, la perspectiva seleccionada siempre va a estar activada.

Ahora nosotros deseamos trabajar con Data Integration entonces seleccionaremos la perspectiva "Intégration" si es que no está ya seleccionada.



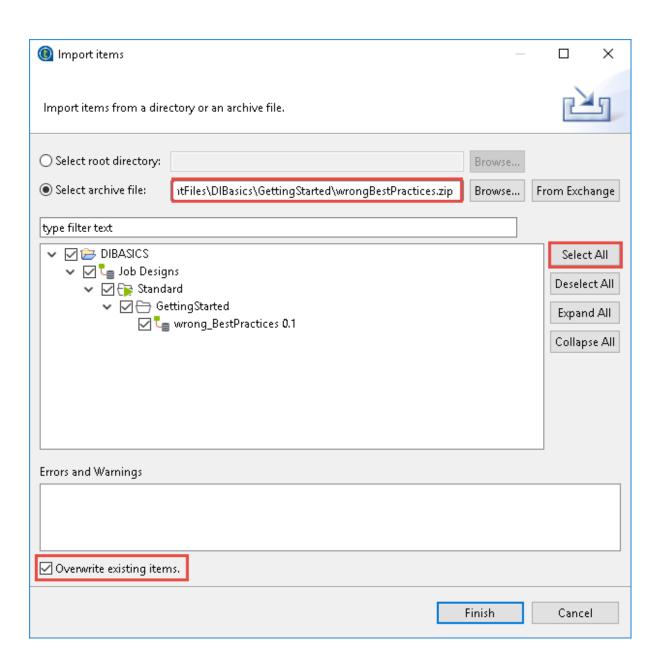
3) Cargar un archivo.

El archivo wrongBestPractices.zip está alojado en la ruta "C:\StudentFiles\DIBasics\GettingStarted". Para importar la metadata y ponerla en el repositorio tenemos que hacer lo siguiente:.

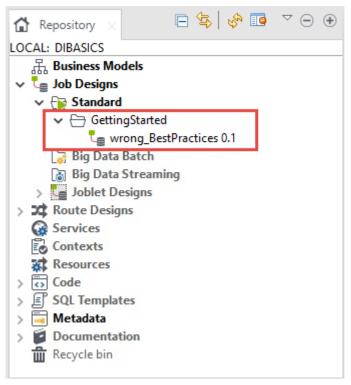
a) En la barra "Quick Access" en el área de trabajo hacer clic en la carpeta con una flecha arriba del icono.



- b) En la ventana de "Import items", seleccionar el botón "Select archive file " luego hacer clic en el botón de "Browse".
- c) Navegar a la ruta "C:\StudentFiles\DIBasics\GettingStarted" y abrir el archivo "wrongBestPractices.zip".
- d) Clic en el botón "Select All".
- e) Asegurarse de que "Overwrite existing items" esté seleccionado y luego hacer clic en finalizar



El job importado será alojado en la carpeta GettingStarted dentro de **Repository > Job Designs > Standard** 

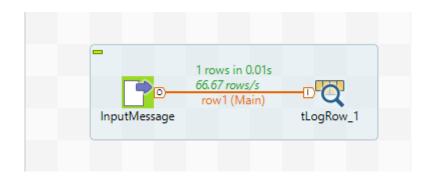


Este yo fue diseñado con documentación incorrecta. luego vamos a corregirlo usando las mejores prácticas de Talend Studio.

Nota: alojar a Jobs en una carpeta en particular puede ayudarte a estar más organizado. Para este ejercicio, la carpeta está en un lugar por defecto dónde se crea y se almacenan los jobs.

Hasta este punto has aprendido a como iniciar TalendStudio, cómo abrir un proyecto existente e importar un Job. ¡Ahora estás preparado para crear tu primer Job!

Creando tu primer Job: En este ejercicio vos vas a crear un job que lea un String desde un archivo y lo mande a la consola. Al final el Job quedará de esta forma.



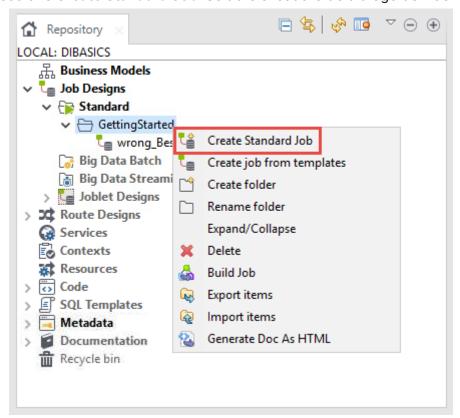
Por defecto, los standard Jobs se crean y almacenan en **Repository** > **Job Designs** > **Standard**. Para agrupar todos los trabajos relacionados con una tarea específica, coloquelos en una carpeta dedicada en esta ubicación.

Durante este ejercicio, se almacenarán todos los trabajos en la carpeta GettingStarted.

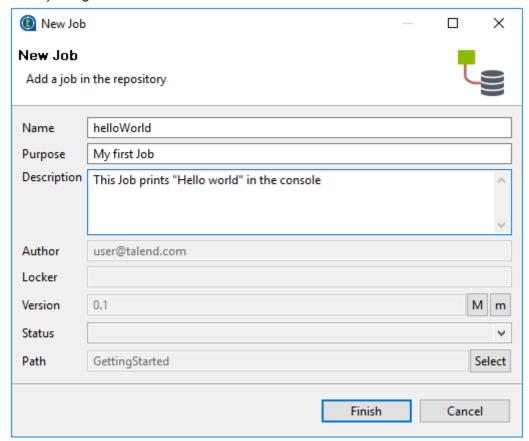
1) Cree un standard Jobs.

En el Repositorio, expanda **Job Designs** > **Standard** y haga clic con el botón derecho en **GettingStarted**.

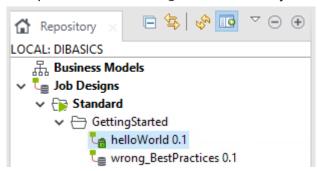
Seleccione Create Standard Job. Se abre el cuadro de diálogo de Nuevo trabajo.



- 2) Defina las propiedades del trabajo en el cuadro de diálogo Nuevo trabajo.
  - a) En el cuadro de texto Nombre, ingrese helloWorld.
  - b) Rellene los cuadros de texto Propósito y Descripción.
  - c) Deje los valores predeterminados para los demás parámetros.
  - d) Haga clic en Finalizar.



e) Podrás ver que cambió el "Designer", el Palette y el resto de las vistas



### Agregando componentes al job

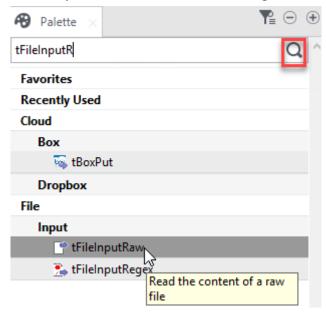
Para este Job vas necesitar un componente que lea el String "Hello Word!" desde un archivo y otro componente que muestre en pantalla la salida de la consola. Para hacer esto, tenés que usar los componentes tFileInputRaw and tLogRow.

tFileInputRaw: Lee todos los datos de un archivo en crudo y los manda a través de una sola salida para que luego se ha recibido y procesado por otro componente.

tLogRow: Muestra los datos o los resultados de una ejecución por consola, esto quiere decir que nos permite monitorear los datos que son procesados a través de la interfaz.

Estos pasos van a mostrar dos formas diferente de agregar componentes a tu Job

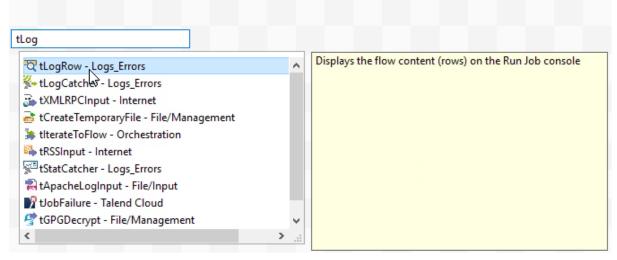
- 1) Agregar el componente tFileInputRaw usando el "Palette"
  - a) En el palette, localizar el cuadro de búsqueda
  - b) Escribir *tFileInputRaw*. presionar enter y te aparecerán los componentes que coincidan con tu búsqueda.
  - c) Dentro de la lista seleccionar tFileInputRaw
  - d) Arrastre el componente dentro del área Designer



Cuando el componente este agregado, el Job se verá así:

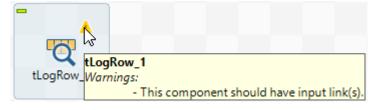


- 2) Agregando el componente tLogRow desde adentro del Designer.
  - a) Click en cualquier lado dentro del Designer.
  - Empezar a escribir tLogRow. Mientras vas escribiendo, los componentes que vayan coincidiendo empezaran a aparecer. Nótese que un pequeña descripción es mostrada al lado de cada nombre de componente a medida que lo seleccione con el mouse
  - c) hacer doble click en **tLogRow** y el componente aparecerá en el Designer.



3) Inspeccione los Warnings

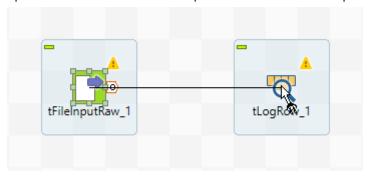
Nótese que ambos componentes están mostrando un pequeño icono de advertencia. Si con el mouse te pones arriba de ese icono, podrás visualizar el mensaje de advertencia. En este caso dice el componente debería tener la entrada y salida enlazados



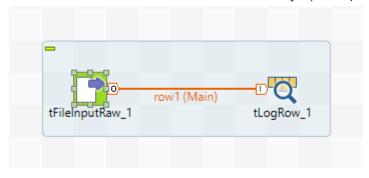
# Conectando componentes

Para poder enviar datos desde tFileInputRaw hasta tLogRow, los componentes deben estar conectados.

- 1) Haga clic con el botón derecho en el medio del componente de entrada, tFileInputRaw. El componente se resalta en un marco.
- 2) Seleccione **Row** > **Main**. Aparece un link de conexión entre el componente de entrada y el puntero del mouse. Ahora puede vincular el componente de salida.

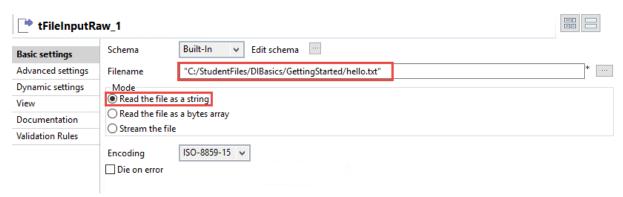


3) Haga clic en el medio del componente de salida, tLogRow, para crear el vínculo. Las flechas en el enlace muestran la dirección del flujo (de izquierda a derecha).

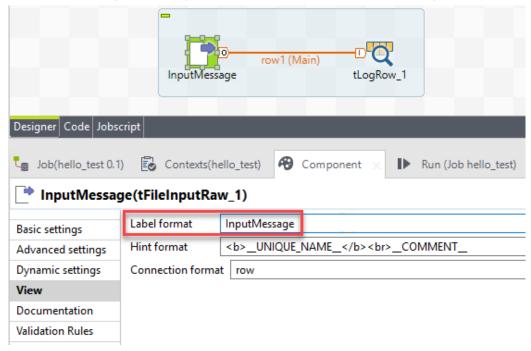


### Configurando componentes

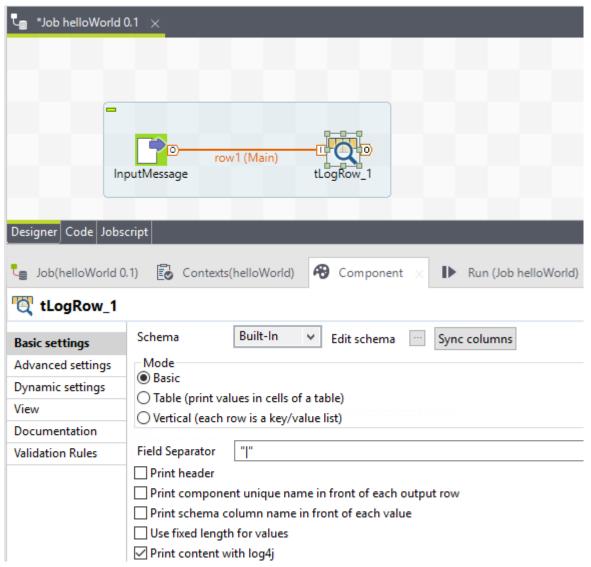
- Configure el componente tFileInputRaw para leer un archivo de texto.
  Puede configurar componentes usando la vista de componentes debajo del
  Designer. Según el componente que utilice, la vista Componente muestra
  diferentes propiedades.
  - a) Haga doble clic en el componente tFileInputRaw. Se abre la vista de componentes.
  - b) Para el campo Filename, haga clic en [...] para seleccionar la ruta del archivo "C:/StudentFiles/DIBasics/GettingStarted/hello.txt"
  - c) Seleccione el botón de Read the file as a string



- 2) Cambie el nombre del componente tFileInputRaw.
  - a) Click en la pestaña de View
  - Reemplace el texto en el cuadro de texto "Label format" con InputMessage. Esto actualiza el nombre del componente en Designer.



- 3) Verifique las propiedades de tLogRow Hay varias opciones disponibles para formatear la forma en que se imprime la salida. El modo Básico muestra la salida sin formatear. Cuando los datos a mostrar son más complejos y contienen múltiples filas, es mejor usar el modo Tabla o Vertical, en cuyo caso el separador de campo le permite separar los campos en cada fila.
  - a) Haga doble clic en tLogRow\_1 para abrir la vista de componentes.
  - b) Para Mode, seleccione Basic.

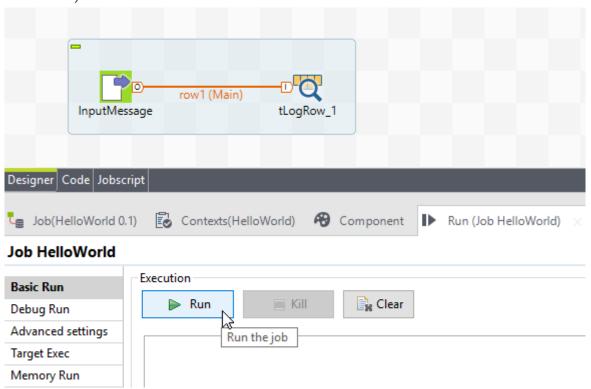


#### 4) Guardar el Job

Observe que en la pestaña del Designer, hay un asterisco (\*)delante del nombre del Job, lo que significa que el Job tiene cambios sin guardar. Para guardar el Job, en la barra de herramientas de acceso rápido, a la izquierda, haga clic en el botón Guardar o también puede hacerlo con **CTRL+S** 

## Ejecutando un Job

- 1) Abre la vista de Run del Job HelloWorld.
  - a) Localiza y abre la vista Run que está debajo del Designer
  - b) Click en run



2) Análisis de la salida del Job

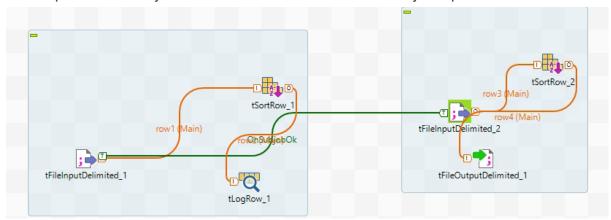
Talend Studio guarda y luego compila el Job. Esto puede tomar unos momentos. Luego se ejecuta el Job y el mensaje "¡Hola mundo!" se muestra en la consola en la vista Run. El trabajo termina con [exit code = 0], lo que significa que se ejecutó con éxito.



#### Diseñando un Job con uso de buenas prácticas

En este ejercicio, aprenderás las mejores prácticas para diseñar Jobs de Talend Studio. esta vez se comienza desde un trabajo existente que no se diseñó correctamente. El propósito de este job es leer datos de un archivo de entrada que contiene información sobre los clientes (nombre y apellido, número de cliente, calle, ciudad y estado) y ordenarlos, primero por estado y luego por apellido.

El Job que rediseñará y documentará de acuerdo con las mejores prácticas se ve así:



Por lo general, hay tres fases en el desarrollo de un proyecto: planificación, desarrollo y documentación. Talend aboga por utilizar las mejores prácticas durante cada fase. El enfoque de este ejercicio es aplicar las mejores prácticas durante la fase de desarrollo que afecta la documentación resultante: convenciones de nomenclatura, anotación de trabajo, configuración de componentes y otras prácticas relacionadas con la documentación.

La documentación del trabajo se puede generar cuando sea necesario y se mantiene automáticamente. La documentación es extensible y acepta notas y comentarios personalizados.

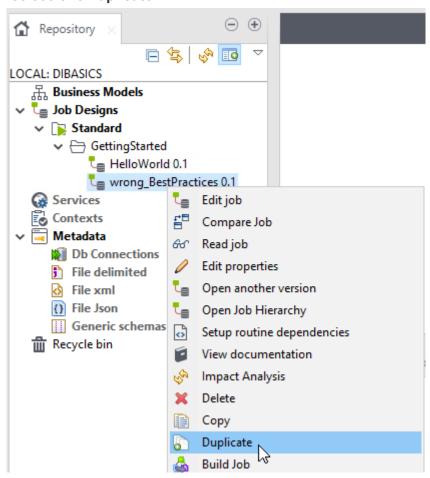
# Rediseño de un trabajo utilizando las mejores prácticas

Para crear un diseño de trabajo limpio, aplique las siguientes pautas:

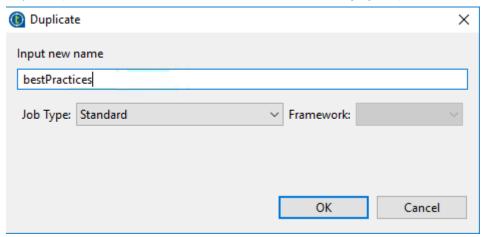
- Configure las propiedades básicas del Job: asigne un nombre apropiado al Job y complete el propósito y la descripción
- Identificar y nombrar los subJobs
- Organice el diseño del trabajo e identifique el componente iniciable
- Organizar el diseño del subJob
- 1) Duplicar Job

Antes de corregir el diseño del Job, duplique el Job **wrong\_BestPractices** y asigne el nombre de *bestPractices* al duplicado.

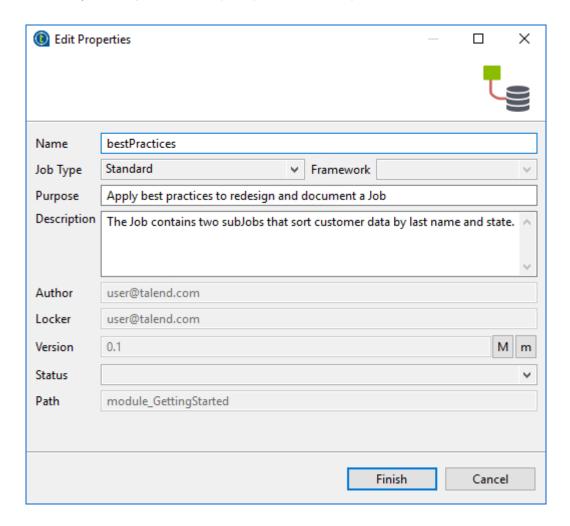
 a) En el Repositorio, expanda Job Designs > Standard > GettingStarted, luego haga clic con el botón derecho en wrong\_BestPractices y seleccione Duplicate.



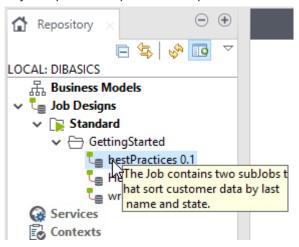
b) Asigne el nombre bestPractices al nuevo trabajo y haga clic en OK.



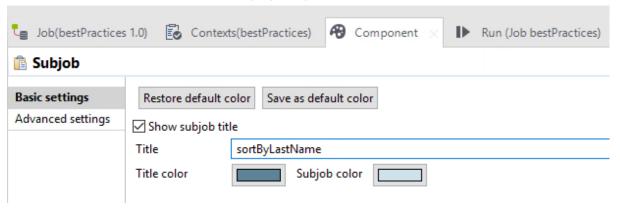
- 2) Configure las propiedades básicas del Job
  - a) En el Repositorio, expanda Job Designs > Standard > GettingStarted, luego haga clic con el botón derecho en bestPractices 0.1 y seleccione Editar propiedades.
  - b) En la ventana Editar propiedades, complete los cuadros de texto Propósito y Descripción con algo significativo. Haga clic en Finalizar.



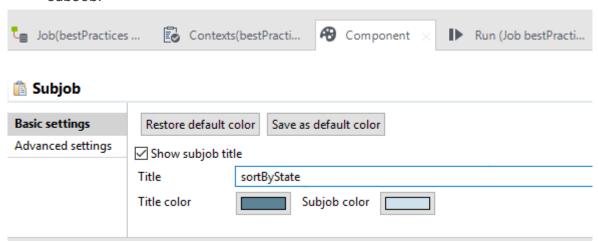
 c) Localiza el Job en el repositorio. Pase el puntero del mouse sobre él y observe que la descripción que ingresó aparece en un cuadro de texto.
 Esto es muy útil para comprender lo que hace un Job sin tener que abrirlo.



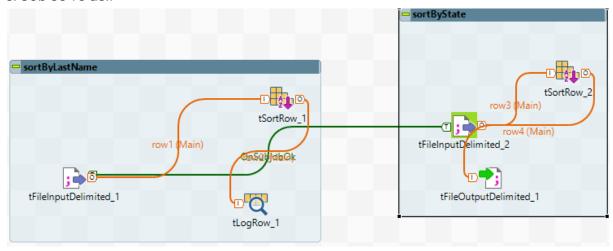
- 3) Identifique y nombre los subJobs.
  - Tómese un momento para observar cómo el diseño del Job le impide comprender lo que se supone que debe hacer este Job y por dónde empezar.
  - b) Para identificar subJobs, observe los dos grupos de componentes en marcos azules. Cada grupo representa un subJobs.



- 4) Repita los pasos anteriores para nombrar el segundo trabajo secundario.
  - a) Localiza el subJob a la derecha y haga clic en cualquier parte de su marco azul para seleccionarlo.
  - b) En la vista de Component, seleccione la casilla Show subJob title. Para Título, ingrese sortByState. El nombre aparece en la parte superior del marco del subJob.



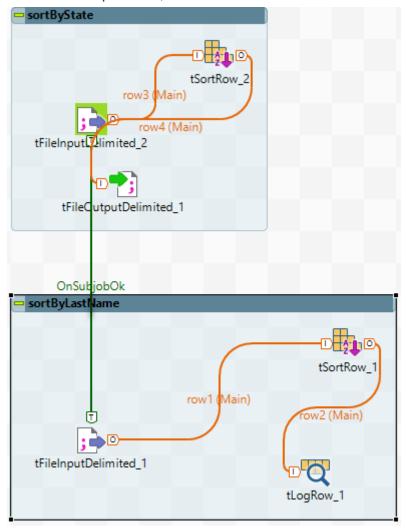
#### el Job se ve así:



#### 5) Organice el diseño del Job.

Para seguir las mejores prácticas de diseño de Talend Studio, coloque los elementos en el Designer comenzando en la esquina superior izquierda y extendiéndose hasta la esquina inferior derecha, ya que el flujo de datos generalmente procede de izquierda a derecha y de arriba hacia abajo. Para alinearse con esta directriz, coloque los subtrabajos de arriba a abajo, siguiendo el orden de ejecución.

- a) Para reorganizar el diseño del Job, busque el componente tFileInputDelimited\_2 y observe que es el único que aparece en un marco verde. Este componente se denomina componente iniciable porque inicia el procesamiento del flujo de datos.
- b) Coloque el subJob que contiene el componente iniciable en la parte superior. Cuando finaliza esta operación, el diseño del Job se ve así.

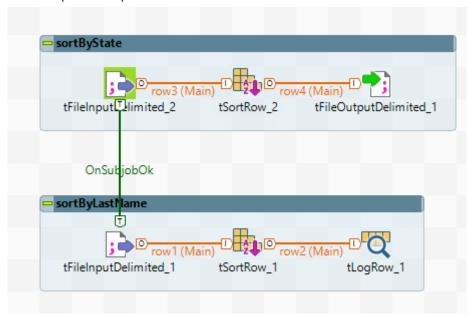


#### 6) Organice el diseño del subJob

Tenga en cuenta la cadena ETL (Extraction Extracción, Transformation Transformación y Load Carga) y el hecho de que el flujo de datos procede de izquierda a derecha.

a) Identificar los componentes que corresponden a cada operación de la cadena. Tal como lo indican los nombres de los tipos de componentes,

- tFileInputDelimited lee datos de un archivo y tSortRow se usa para ordenar (por lo tanto transformar) los datos, mientras que tFileOutputDelimited y tLogRow se usan para la salida.
- b) Siguiendo estas pautas, reorganice el diseño del subtrabajo **sortByState** y repita los pasos para el segundo subJob. Confirme que sus resultados coinciden con los de la captura de pantalla.



#### Documentando el Job

En este ejercicio cubre las mejores prácticas importantes relacionadas con la fase de desarrollo en lo que respecta a la documentación y el mantenimiento básico del Job. Se introducen varias pautas con respecto a las propiedades del Job: denominación y etiquetado de componentes, uso de comentarios, descripciones detalladas y anotación del trabajo.

Durante la fase de documentación, después de proporcionar estos elementos informativos, puede generar y visualizar la documentación HTML de los Job. Al igual que con cualquier proyecto, una buena documentación puede ser fundamental para ayudar a otras personas a comprender y utilizar los Talend Jobs.

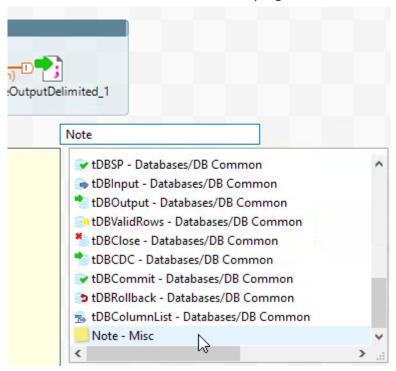
Este ejercicio utiliza el job bestPractices como ejemplo.

#### Anotación de Job

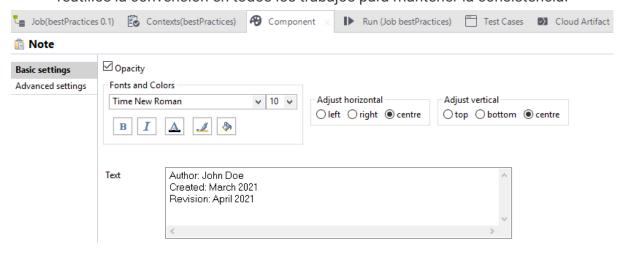
Puede insertar fácilmente notas en un job. Son visibles para todos los desarrolladores que muestran el Trabajo en sus Diseñadores.

 Agregando notas en un job
 Abra el job bestPractices 0.1 si no está abierto. Para crear más espacio en Designer, utilice el botón Contraer para cerrar el subJob sortByLastName.

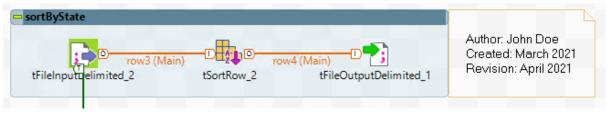
Para anotar el Job, haga clic en cualquier parte del Designer y comience a ingresar Note. Seleccione Note en la lista desplegable.



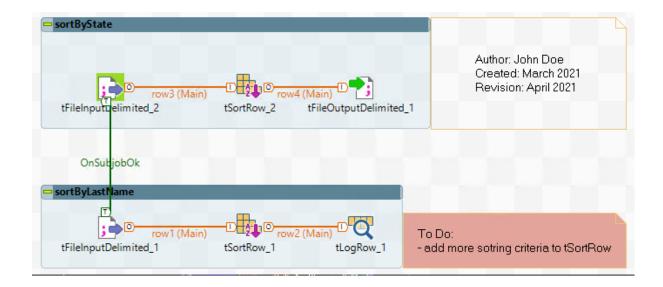
2) Dando formato a la nota Haga doble clic en Note y, en la vista Componente, introduzca texto en el campo. Considere usar una convención de alto nivel que incluya configuraciones de formato e información básica (por ejemplo, nombre del autor, fecha de revisión) y reutilice la convención en todos los trabajos para mantener la consistencia.



3) Dimensiones y posición el tamaño de la nota. Coloque la nota a la derecha del subJob sortByState. Cambie el tamaño y la posición según sea necesario para acomodar el texto que ingresó.



4) Añade una lista de "cosas por hacer". Coloque otra nota debajo de los componentes para documentar los elementos pendientes que quedan por completar. La incorporación de un color de relleno diferente ayuda a diferenciar el componente.



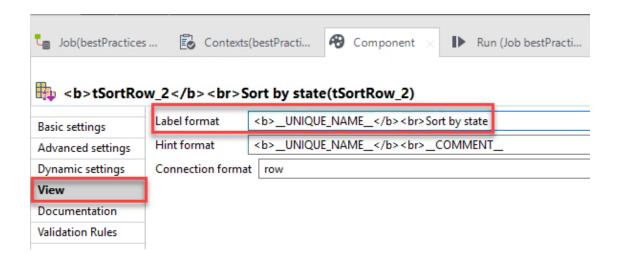
### Documentar un componente

Este ejercicio presenta varias prácticas recomendadas de nomenclatura y documentación de componentes.

 Proporcione etiquetas significativas para los componentes.
 Haga doble clic en el componente tSortRow\_2 para abrir la vista de componentes. Haga clic en la pestaña View y observe el valor predeterminado para el Label format.

\_\_UNIQUE\_NAME\_\_ es una macro generada automáticamente por Talend Studio. Asigna un nombre único a los componentes cuando se crean. Sin embargo, este nombre puede no ser intuitivo. Para que la etiqueta sea más significativa, pon el nombre en negrita y agrega una breve descripción:

<b>\_\_UNIQUE\_NAME\_\_</b><br>Sort by state



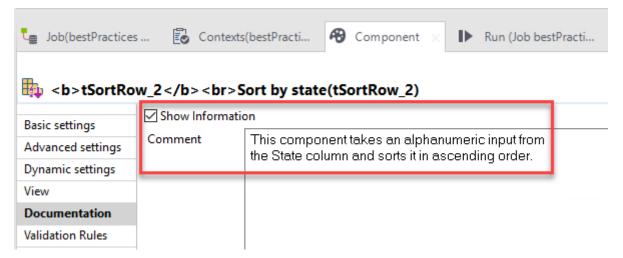
Aunque el formato de texto en el cuadro de texto Formato de etiqueta parece menos legible, dirija su atención al Diseñador. Puede ver claramente el icono del componente con su nombre en negrita y un resumen de su propósito. Toda esta información puede ser útil, especialmente para los nuevos desarrolladores u otras partes interesadas que trabajan en su Job.



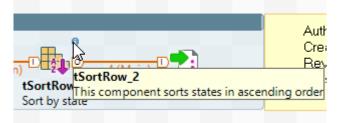
- 2) Agregue una sugerencia a un componente.
  - a) Para personalizar el **Hint format**, reemplace **\_COMMENT\_** con información útil, como: Este componente ordena los estados en orden ascendente



 b) Haga clic en la pestaña Documentation. Seleccione la casilla Show Information e ingrese información adicional en el cuadro de texto Comment.



3) Pase el puntero del mouse sobre el ícono de información en el componente tSortRow\_2 y observe que se muestra el texto de sugerencia.



- 4) Si el tiempo lo permite, repita estos pasos para documentar otros componentes.
- 5) Guarde el trabajo antes de seguir adelante.

### Ver y generar documentación HTML

Talend Studio le permite generar documentación detallada de Job en formato HTML. La vista previa de la documentación le permite comprobar rápidamente cómo se ve el HTML desde el espacio de trabajo de Studio. Puede utilizar esta función en cualquier momento del ciclo de desarrollo.

Vea la documentación HTML en Studio.
 En el Repositorio, expanda Job Designs > Standard > GettingStarted, luego haga clic con el botón derecho en bestPractices 0.1 y seleccione View documentation.
 Se abre una nueva pestaña con una vista previa de la documentación HTML.



Jobs

Generated by Talend Data Fabric

| Project Name | DIBASICS        | GENERATION DATE            | Feb 8, 2021, 3:11:44 PM |
|--------------|-----------------|----------------------------|-------------------------|
| AUTHOR       | user@talend.com | Talend Data Fabric VERSION | 7.3.1.20200219_1130     |

#### **Summary**

Project Description
Description
Preview Picture
Settings
Context List
Component List
Components Description

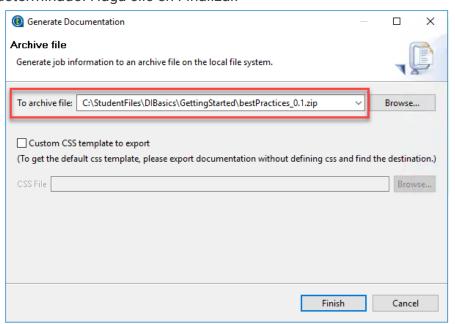
2) Examine la lista de componentes. Haga clic en el hipervínculo Component List y, a continuación, seleccione tSortRow\_2 en la lista de componentes del trabajo. A medida que se desplaza hacia abajo, observe los parámetros detallados de los componentes. Hacia el final de la lista, puede ver el comentario y su valor asociado, que coincide con el comentario que agregó anteriormente.

| ⊗ bestPractices_0.1.html ×             |   |  |  |
|--|---|--|--|
| Sort on disk                           | false   |  |  |
| Temp data directory path               | "C:/Users/Student/workspace/temp"                         |  |  |
| Create temp data directory if does not | true  |  |  |
| exist                                  |   |  |  |
| Buffer size of external sort           | 1000000   |  |  |
| Label format                           | <b>UNIQUE_NAME</b><br>Sort by state                       |  |  |
| Hint format                            | <b>_UNIQUE_NAME</b><br>This component sorts states in     |  |  |
| Time format                            | ascending order   |  |  |
| Connection format                      | row   |  |  |
| Show Information                       | true  |  |  |
| Comment                                | This component takes an alphanumeric input from the State |  |  |
| Comment                                | column and sorts it in ascending order.                   |  |  |
| Use an existing validation rule        | false   |  |  |
| Validation Rule Type                   |   |  |  |

Ahora que la vista previa muestra la información que desea ver en las páginas de documentación, puede generar la documentación HTML real.

3) Genere una salida HTML para la documentación del trabajo. En el repositorio, expanda Job Designs > Standard > GettingStarted, luego haga clic con el botón derecho en bestPractices 0.1 y seleccione Generate Doc As HTML. Examine para seleccionar la ruta

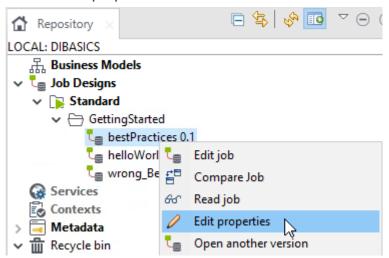
**C:\StudentFiles\DIBasics\GettingStarted** y mantenga el nombre de archivo predeterminado. Haga clic en Finalizar.



4) Extraiga el archivo, luego ubique y abra **bestPractices\_0.1.html** Tenga en cuenta que se incluyen todos sus comentarios, así como los ajustes de configuración de los componentes y otra variedad de información.

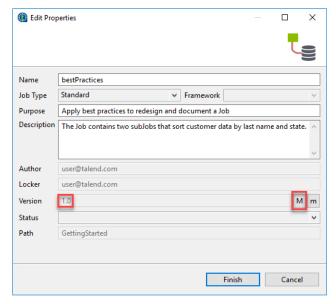
Creación de una nueva versión de un trabajo Cierre el trabajo de **bestPractices 0.1** para editar las propiedades del trabajo y crear una nueva versión.

Cree una nueva versión del trabajo.
 En el Repositorio, expanda Job Designs > Standard > GettingStarted, luego haga clic con el botón derecho en bestPractices 0.1 y seleccione Edit properties. Se abre la ventana Editar propiedades.



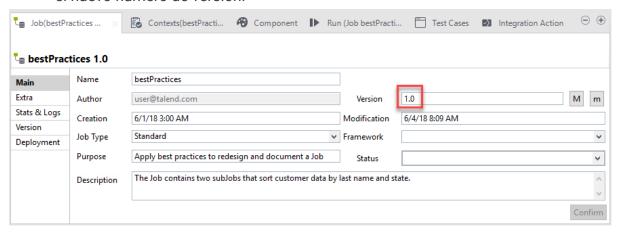
A la derecha de Versión, haga clic en M (abreviatura de versión principal). Observe que el valor de la versión cambia de 0.1 a 1.0. Estos dos botones se utilizan para ajustar los números de la versión principal y secundaria.

Haga clic en Finalizar.

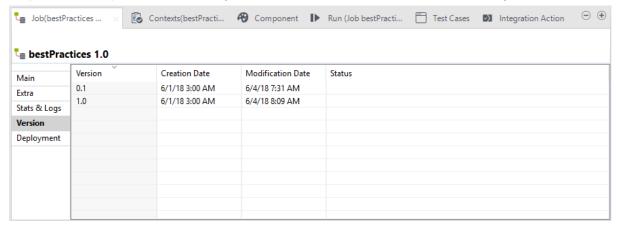


En el Repositorio, observe que el Trabajo se muestra con la nueva versión: **bestPractices 1.0**.

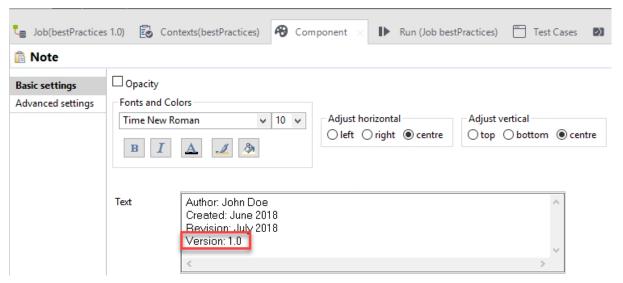
Examine la nueva información del trabajo.
 Haga doble clic para abrir el trabajo de bestPractices 1.0. En la vista Job, observe el nuevo número de versión.



Haga clic en la pestaña Versión y observe el historial de revisión del trabajo.

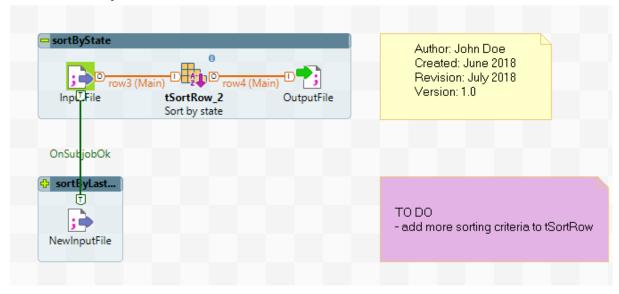


3) Modificar el Job Haga clic en la nota a la derecha del subJob sortByState. En la vista de componentes, modifique el contenido del cuadro de texto para agregar una línea con el número de versión del Job.



Confirme que la nota se ha modificado en consecuencia en el Diseñador. Aplique las mejores prácticas para cambiar el nombre de los componentes como se muestra en la captura de pantalla.

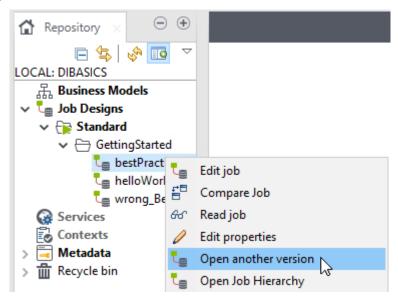
#### Guarde el trabajo.



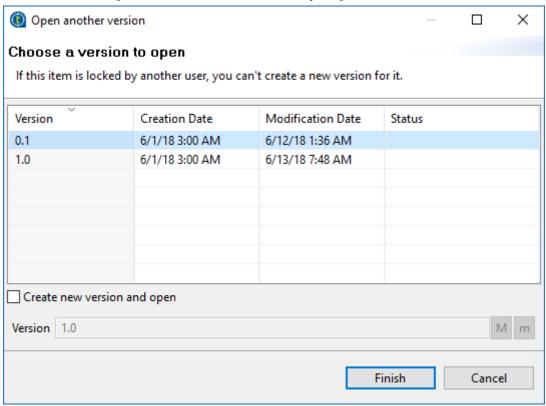
4) Abra otra versión del trabajo.

A veces es útil consultar una versión anterior de un trabajo para identificar los cambios realizados entre los dos.

En el Repositorio, expanda **Job Designs > Standard > GettingStarted**, luego haga clic con el botón derecho en **bestPractices 1.0** y seleccione **Open another version**.



En el cuadro de diálogo, seleccione la versión 0.1 y haga clic en Finalizar.



La versión 0.1 del Trabajo se abre en el Diseñador. Tómese un momento para observar cómo estaban los elementos que modificó recientemente en esta versión anterior del

Trabajo. Si es necesario, haga doble clic en el trabajo bestPractices 1.0 en el repositorio para abrirlo y ver los cambios.