

Etapa 1: Download da ferramenta Talend Open Studio for Big Data

Acessar o link <https://www.talend.com/download/> e baixar a versão FREE

Requisitos para instalação:

Memória: mínimo 3 GB e recomendado 4 GB

Disco: mínimo 3 GB e recomendado 10 GB

Clicar em variáveis de ambiente e adicione uma nova variável de sistema com o nome JAVA_HOME.

Informe a pasta onde está localizado a JRE ou JDK do sistema operacional.

Ex: C:\Program Files\Java\jre1.8.0_201

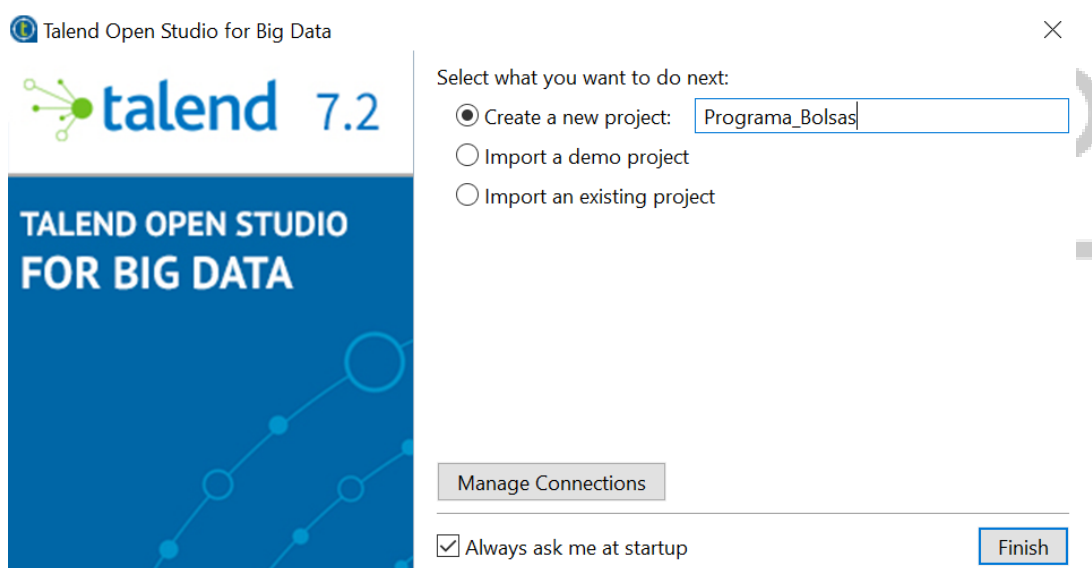
Após edite a variável Path e adicione %JAVA_HOME%\bin

A versão que usaremos é a 7.2

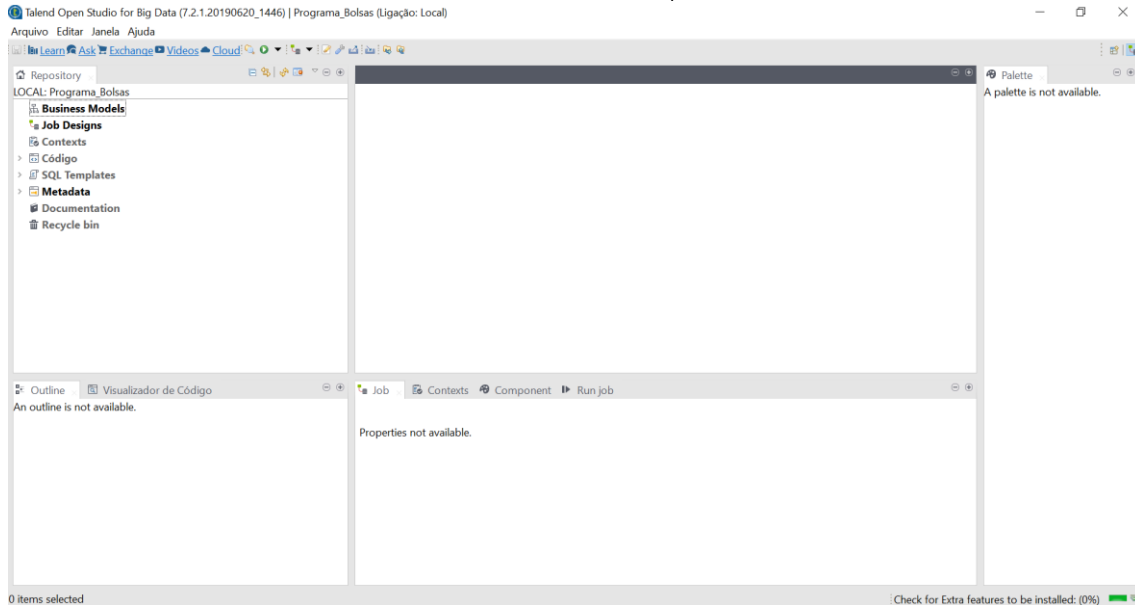
Etapa 2: Instalação

Descompacte o arquivo e inicie o executável.

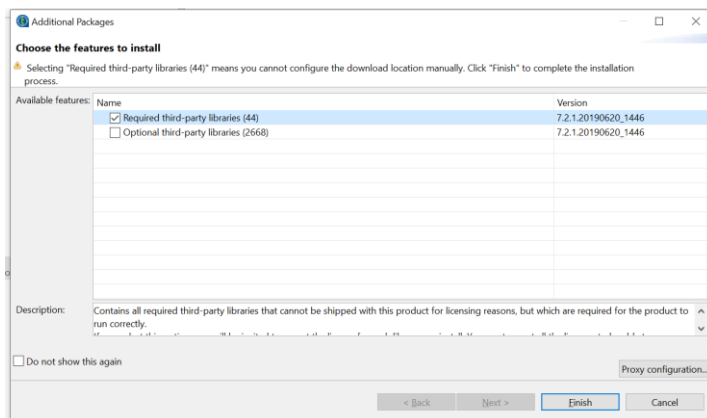
Assim que surgir a mensagem abaixo, clique em “Criar um novo Projeto”, defina o nome do Projeto e clique em Criar. O nome do mesmo não pode conter espaços.



Após finalizar a instalação ele apresenta a interface como abaixo:



Sugestão, instale as bibliotecas definidas como “required”



Etapa 3: Criando o primeiro projeto

Vamos descobrir o Talend Studio e sua interface, criando um projeto e uma tarefa de demonstração simples.

Dentro do Talend Studio

A interface do Talend Studio inclui vários painéis chamados visualizações.

O Repositório do Projeto (**Repository**) lista todos os itens do projeto, como **Jobs** (programas ETL java), Contextos (**Contexts**), **Código**, **SQL Templates**, metadados (**Metadata**) e documentação (**Documentation**) do projeto.

O Job Designer é a visualização principal do Studio em que os componentes são usados para criar tarefas ETL.

A **Palette** lista todos os componentes disponíveis, organizados em pastas.

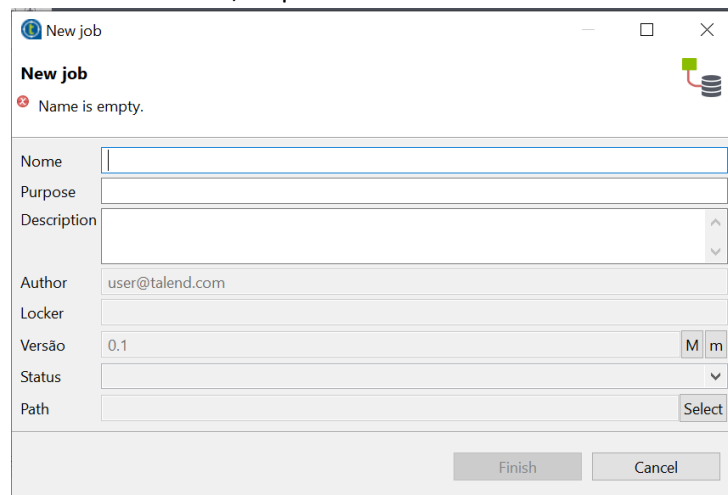
A **Component Area** exibe todos os parâmetros para configurar um componente. Observe que as informações exibidas nesta área dependem do que está selecionado no Job Designer.

A **Run Job** dispara a execução de um Job Talend e exibe seus resultados na janela do console.

O Talend Studio ajuda a criar programas ETL usando uma interface gráfica do usuário. Esses programas são chamados de Data Integration (DI) Jobs.

Criando um Job

1. Verifique se a perspectiva de integração está selecionada. (Janela->Perspectiva->Integration)
2. No Repositório do Projeto, clique com o botão direito do mouse em **Job Designs**.
3. Para abrir o assistente **New Job**, clique em **Create Job**.



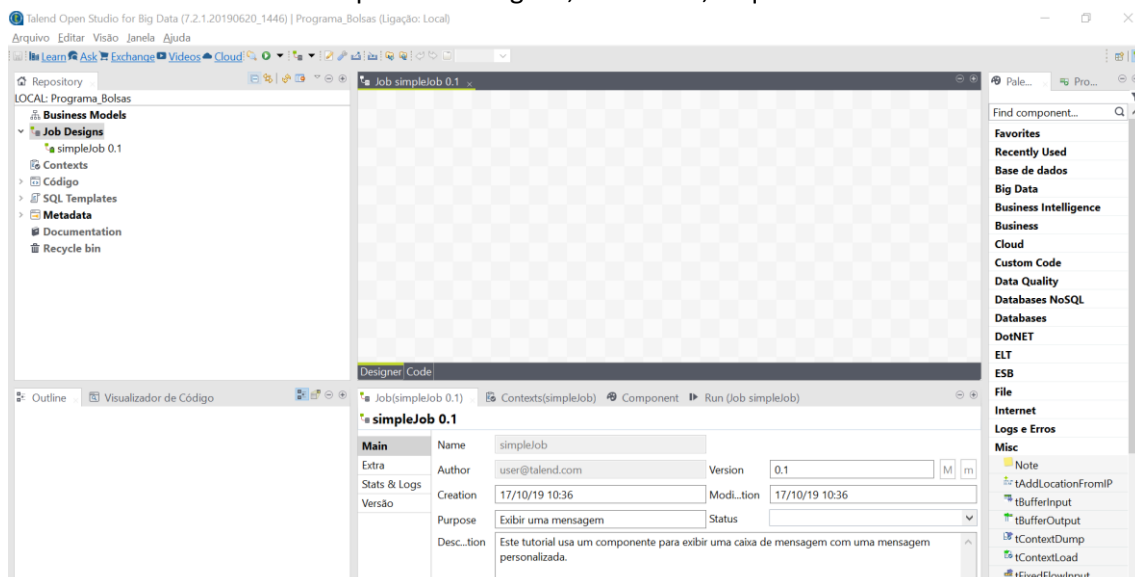
The 'New job' dialog box is shown. It has a title bar 'New job' and a close button. Below the title bar, it says 'New job' and 'Name is empty.' with a red error icon. There are input fields for 'Nome', 'Purpose', and 'Description'. The 'Author' field is pre-filled with 'user@talend.com'. There are also fields for 'Locker', 'Versão' (0.1), 'Status', and 'Path'. At the bottom, there are 'Finish' and 'Cancel' buttons.

4. No campo Nome do assistente New Job, preencha o nome do Trabalho como simpleJob.
5. No campo Finalidade (Purpose), digite: Exibir uma mensagem.
6. No campo Descrição (Description), insira: Este tutorial usa um componente para exibir uma caixa de mensagem com uma mensagem personalizada.
7. Para fechar o assistente e criar o trabalho, clique em **Finish**.

O Job Design abre um Job vazio. O Talend Studio oferece uma biblioteca abrangente de mais de 900 componentes para integração de dados.

Adicionando o Componente tMsgBox

1. Para adicionar o componente tMsgBox, na Palette, clique em **Misc**

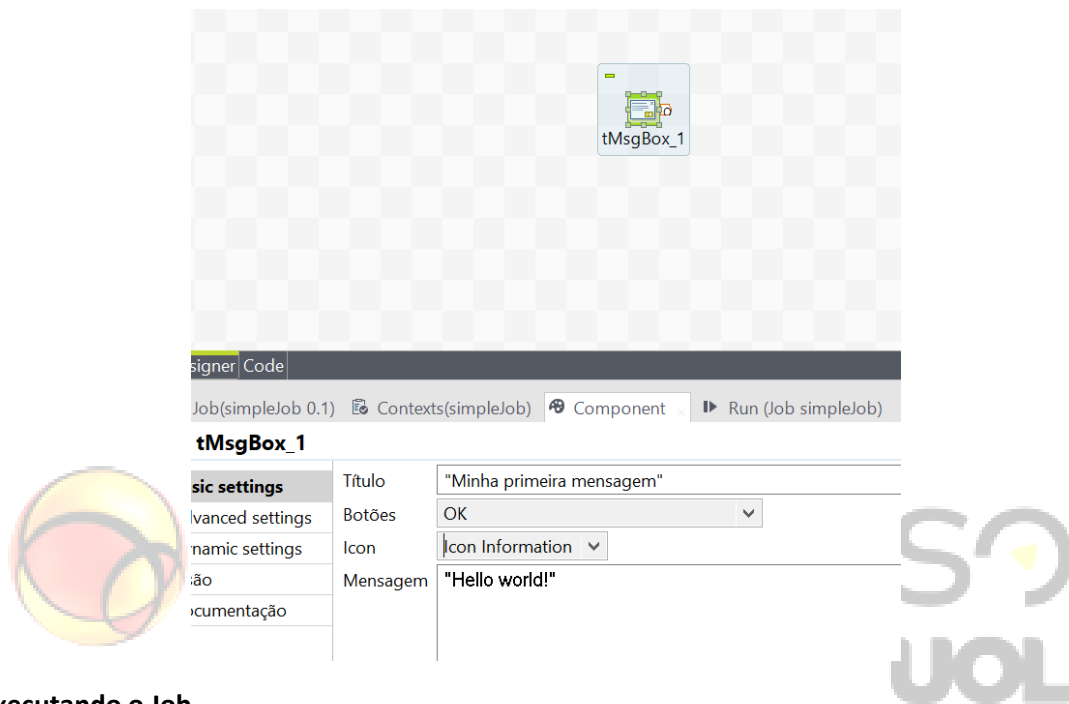


The Talend Studio interface is shown. The 'Repository' pane on the left shows the 'Job Designs' folder expanded, with 'simpleJob 0.1' selected. The 'Job simpleJob 0.1' is open in the 'Designer' view. The 'Main' tab is active, showing the job details. The 'Palette' on the right shows the 'Misc' category selected, with the 'tMsgBox' component highlighted. The 'Job simpleJob 0.1' details are shown in the bottom pane, including the 'Name', 'Author', 'Version', 'Creation', 'Modification', 'Purpose', and 'Description'.

2. Clique em tMsgBox e arraste para o Job Designer. Este componente simples orientado a teste exibe uma caixa de mensagem.

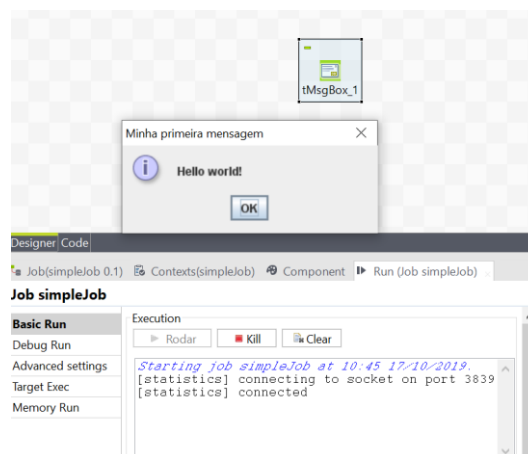
Configurando o componente tMsgBox_1

1. No Job Designer, clique em **tMsgBox_1**
2. Para definir as configurações básicas do componente, na visualização Componentes, clique no **Component**.
3. Altere os campos 'Título' e 'Mensagem', altere o ícone no menu suspenso



Executando o Job

Para executar, na aba Run (Job simpleJob), clique em Rodar. A caixa de mensagem é exibida.



Caso 2: Ler um arquivo CSV e inserir no Banco de dados

A partir de agora, vamos criar um processo de integração. Vamos ler o arquivo CSV FUNC.CSV e inserir na tabela FUNCIONARIO do MySQL.

- 1º) Vamos criar a tabela. Se ela já existir, vamos apagar o conteúdo dela

```
CREATE TABLE FUNCIONARIO (  
    CodFun INTEGER NOT NULL,  
    Nome VARCHAR(50) NOT NULL,  
    Nascimento DATE NOT NULL,  
    Contratacao DATE NOT NULL,  
    PRIMARY KEY (CodFun)  
);  
  
delete from FUNCIONARIO;
```

2º) Criar um novo Job

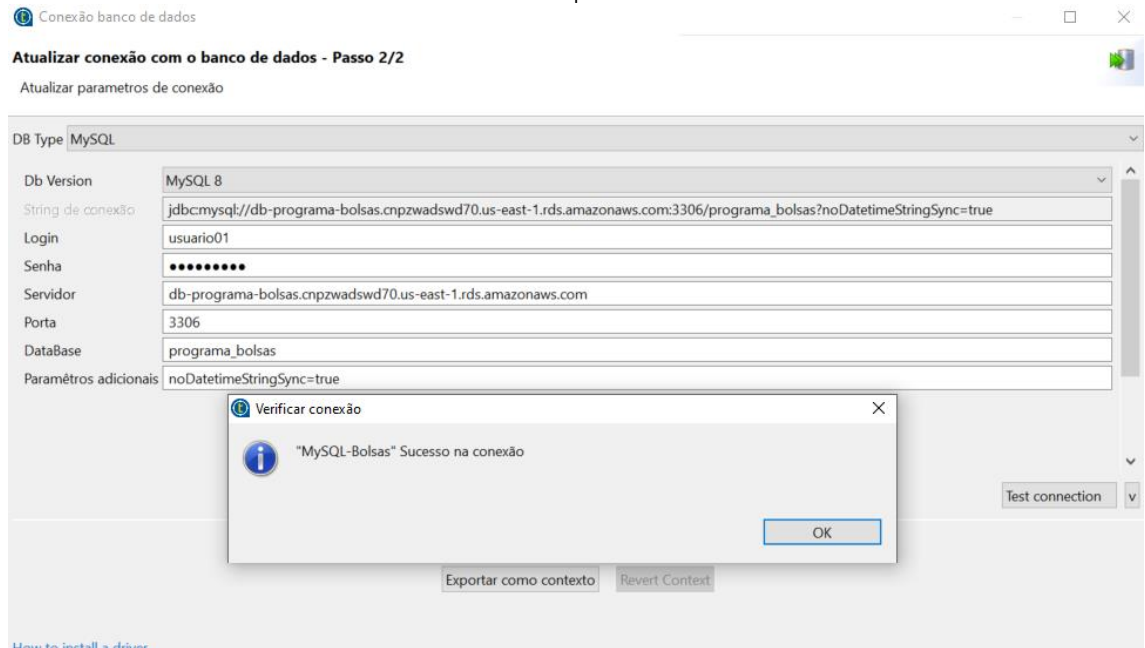
1. Verifique se a perspectiva de Integration está selecionada.
2. No Repositório de Projetos, clique com o botão direito do mouse em **Job Designs** e clique em **Create Job**.
3. No campo Nome do assistente New Job, preencha o nome do Trabalho como readCSVFile.
4. É uma boa prática adicionar um objetivo e uma descrição a um trabalho. Em seguida, clique em Concluir para criar seu trabalho.

3º) Criar os Metadados de um arquivo CSV

1. Expandir o menu **Metadata**. Em **File delimited**, clicar com o botão direito e escolher **criar arquivo delimitado**.
2. Definir um Nome para o arquivo de Funcionários e clicar em **Next**
3. Na aba configurações do arquivo, localize o arquivo FUNC.CSV e escolha o Format **WINDOWS**. Após isso clique em **Next**
4. Neste novo passo defina o separador do arquivo como **virgula** e o **Escape Char Setting** como CSV. Clique em **Refresh Preview** para pré-visualizar as configurações definidas.
5. Finalize as configurações clicando em **Finish**

4º) Criar os Metadados do MySQL de destino


1. Expandir o menu **Metadata**. Em **Db Connections**, clicar com o botão direito e escolher **criar conexão**.
2. Definir um Nome para o Banco de dados da Biblioteca, eu defini MySQL-Bolsas. Defina o objetivo e a descrição para sua documentação ficar clara. Por fim clicar em **Next** para ir ao próximo passo
3. Definir as configurações da conexão. Depois teste para ter certeza que está correto, assim como na imagem abaixo:

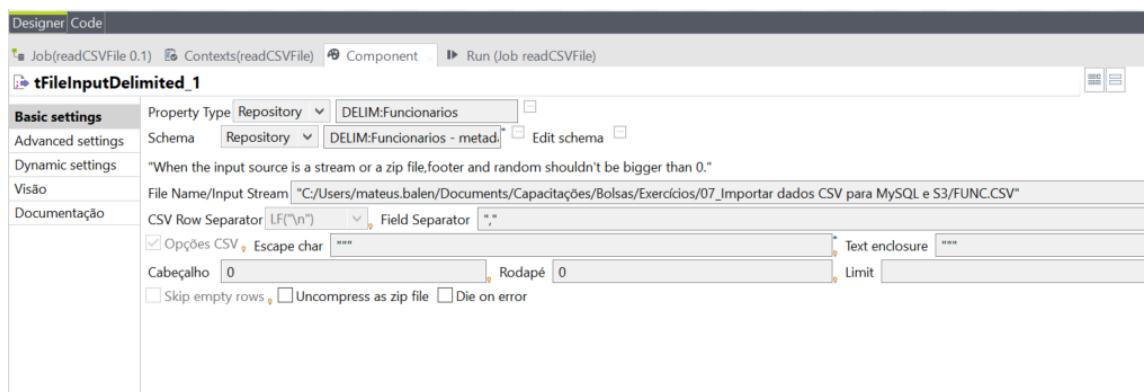


4. Finalize as configurações clicando em **Finish**

5º) Adicionar o componente tFileInputDelimited

6º) Configurar o componente tFileInputDelimited



1. Clique em **tFileInputDelimited_1**
2. Para definir as configurações básicas do componente, na visualização Componente, clique no **Component**.
3. A primeira coisa a ser feita é a escolha do Metadado criado para o Arquivo CSV. Então em **Property Type** escolha **Repository**. Após escolha clicando no ícone  escolha o metadado **Funcionarios 0.1**
4. As configurações devem ficar como na imagem abaixo



7º) Adicionar o componente tDBOutput

8º) Configurar o componente tDBOutput

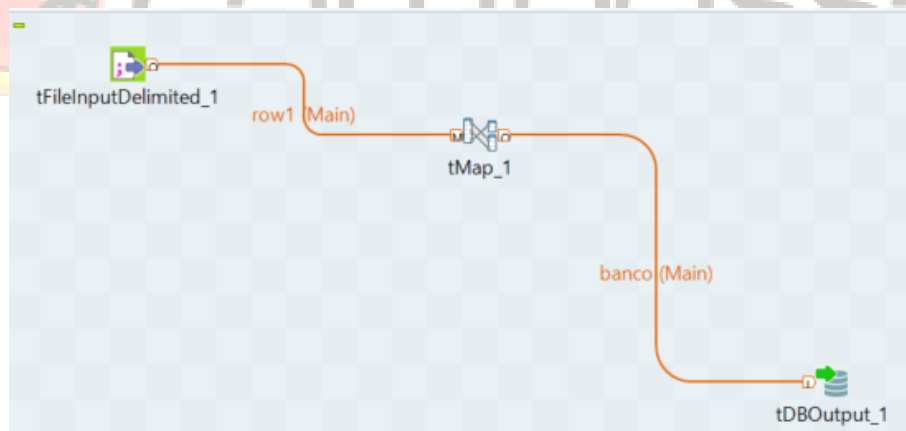
1. Clique em **tDBOutput_1**
2. Para definir as configurações básicas do componente, na visualização Componente, clique no **Component**


3. A primeira coisa a ser feita é a escolha do Tipo do Banco de Dados. Então escolha MySQL no campo **Database** e clique em **Apply**.
4. Agora precisa-se escolher o Metadado criado para o MySQL. Então em **Property Type** escolha **Repository**. Após escolha clicando no ícone  escolha o metadado **MySQL-Bolsas 0.1**
5. Vamos escolher a tabela que iremos utilizar, então digite o nome da tabela entre as aspas do campo **Tabela**
6. Vamos agora escolher as ações que queremos realizar na tabela. Neste exemplo vamos definir em **Ação para o dado** a opção Inserir ou Atualizar.
7. Vamos clicar no campo **Sync columns** para atualizar a lista de colunas
8. Agora para que ocorra a atualização será necessário definir o campo chave, caso este não esteja definido marque o campo **CodFun** como **key** ao clicar em  ao lado do campo **Sync columns**

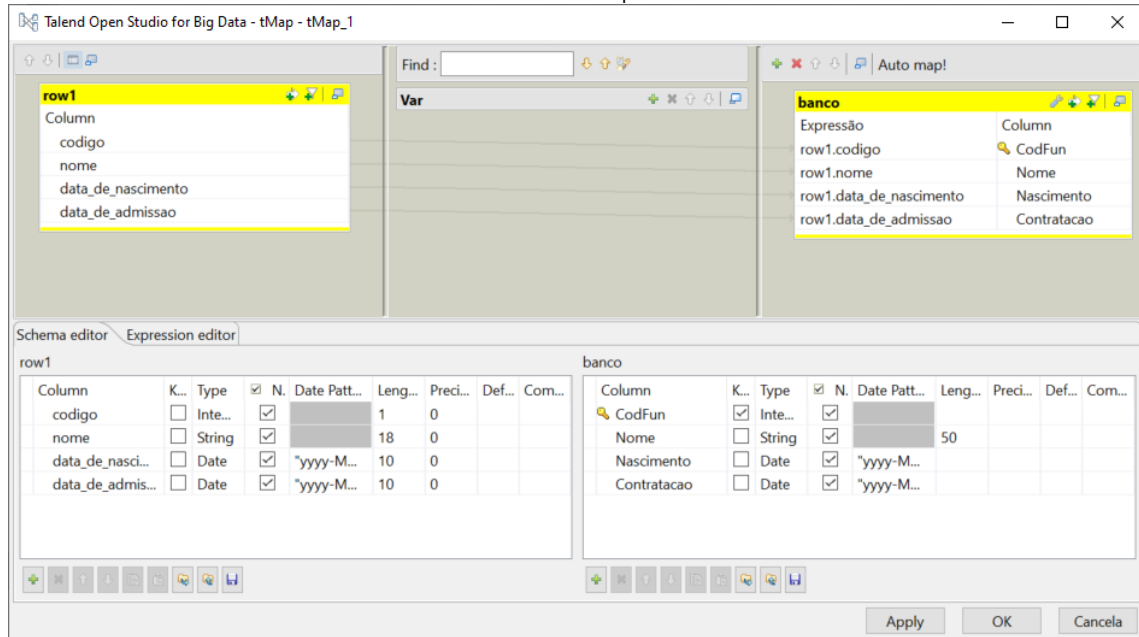
9º) Agora precisa-se relacionar a origem (arquivo) e o destino (Mysql). Então adicione o componente tMap

10º) Configurar o componente tMap

1. Clicar no canto do componente **tFileInputDelimited_1** e arrastá-lo até o componente **tMap_1**
2. Clicar no canto do componente **tMap_1** arrastá-lo até o componente **tDBOutput_1**
3. Definir um nome para a saída dos componentes. As ligações devem ficar mais ou menos assim:



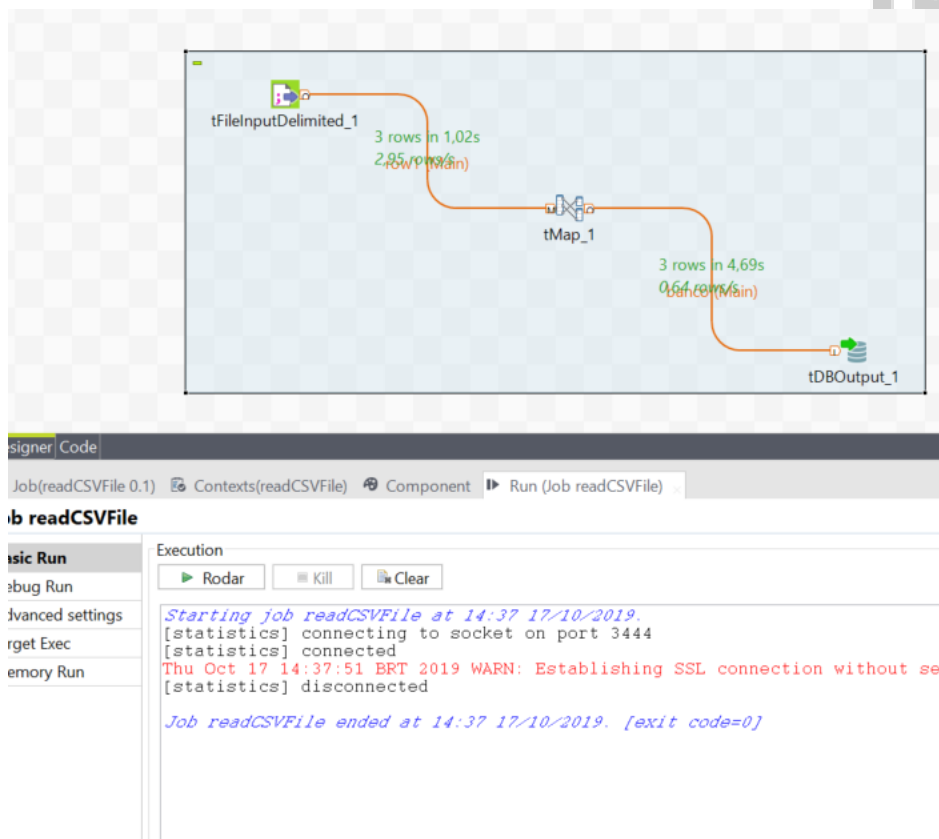
4. Depois ao clicar no componente **tMap_1**, clicar em Map Editor . No editor arraste cada campo para o campo do arquivo à esquerda para o correspondente campo de destino à direita. O mapeamento deve ficar como na figura abaixo:



5. Clique em **OK** e o mapeamento estará finalizado

11º) Executar o Job criado.

1. Ao clicar na área do Job ou em qualquer componente, escolha a aba **Run** (Job readCSVFile)
2. Nesta janela clique no botão **Rodar**
3. A execução deverá retornar como a imagem abaixo:



4. Para ter certeza de que o processo está funcionando. Altere os dados do arquivo CSV, execute novamente o Job e consulte a tabela no banco para ter certeza que tudo funcionou.

Mais informações em: <https://www.talend.com/resources/discovering-talend-studio/>