

Definição de Jobs no Talend

Grupo 10

César Silva
PG41842

Filipe Cunha
A83099

Hugo Gião
PG41073

José Alves
A82885

Tiago Ramires
PG41101

19 de Novembro de 2019



Job 1

Através da criação deste job pretendemos saber quais as pessoas que se encontram a efetuar tratamento e que tenham historial na família. Para isso, a informação contida no ficheiro csv disponibilizado foi mapeada para uma base de dados *MySql*. Após o mapeamento foi aplicado um filtro relacional onde eram verificados os campos **family_history** e **treatment** que selecionava esses mesmos pacientes. De seguida é efectuada uma ordenação por idade, para facilitar a visualização e enviado o resultado para um ficheiro *Excel*.

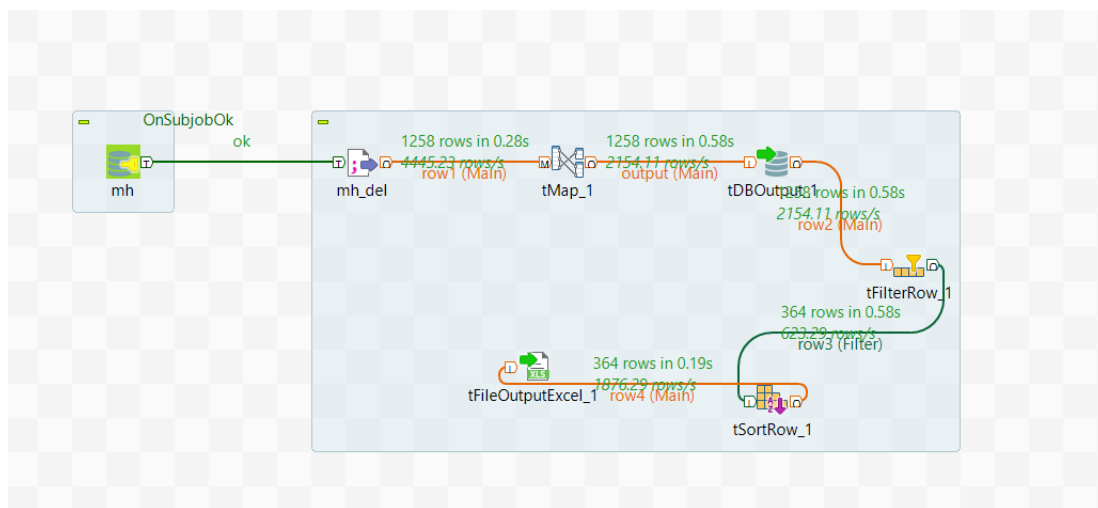


Figura 1: Job1 no *Talend*

Job 2

Neste Job carregamos a informação contida num ficheiro csv e mapeamos essa informação numa base de dados. Filtramos pela coluna os pacientes que estão em tratamento, separando os outputs. A informação que não passou pelo filtro é armazenada num ficheiro *Excel*, enquanto a que passou, é-lhe posteriormente aplicado outro filtro onde verificamos se esses pacientes já comentaram a sua situação com algum companheiro de trabalho através do campo *coworker*. A essa informação é aplicado o mesmo conceito anteriormente mencionado, onde os outputs são divididos. Dessa divisão são criados, por fim, mais dois ficheiros *Excel*, um com a informação positiva do filtro anteriormente aplicado e outro com a informação dos pacientes que já falaram tanto com os companheiros de trabalhos como os seus supervisores.

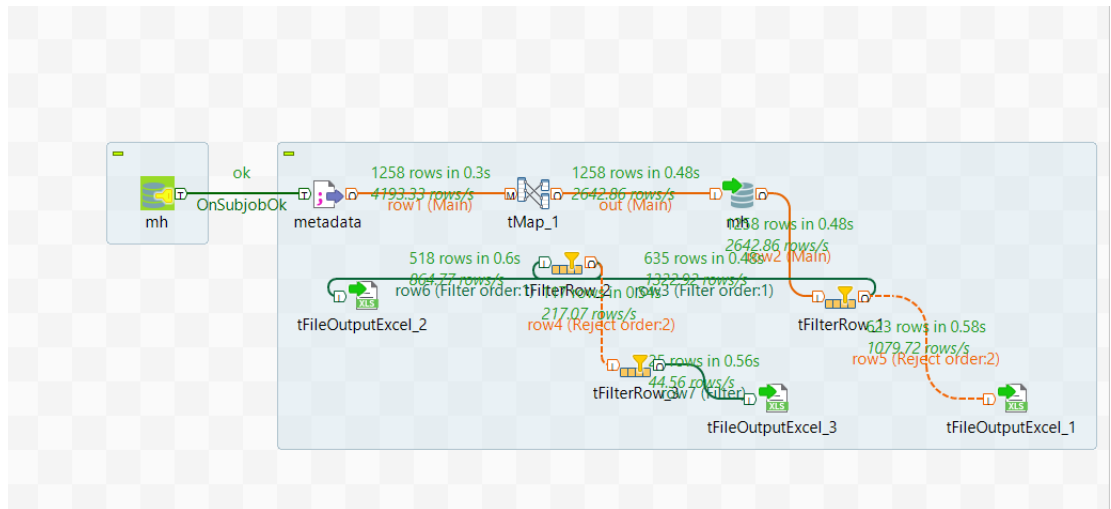


Figura 2: Job2 no *Talend*

Job 3

Este job tem como objectivo a normalização dos dados e processamento dos mesmos para facilitar o desenvolvimento de um *DW* a partir do dataset "mental_health".

Começou-se por converter a coluna com as idades dos pacientes para a respectiva data de nascimento, de seguida para normalizar a coluna dos géneros dos pacientes converteram-se todos campos que contivessem a string "male" para "M" e "female" para "F", mantendo os "M" e "F" já existentes, todos os outros foram considerados como "O" (Outros) pois para futura extração de informação do dataset achou-se melhor tratar estes como um tipo visto que masculino e feminino representam grande parte do *dataset*.

Para preparar os dados para a criação de um *DW*, separou-se todas as unidades temporais da coluna timestamp, permitindo assim mais tarde a criação de uma dimensão *Data* e uma dimensão *Time*. De seguida os resultados foram ordenados por data e finalmente foram substituídos todos os campos que continham "more than" e "less than" por » e « respectivamente para facilitar futuro processamento. Por último exportamos os dados para um ficheiro *Excel* para confirmação dos dados

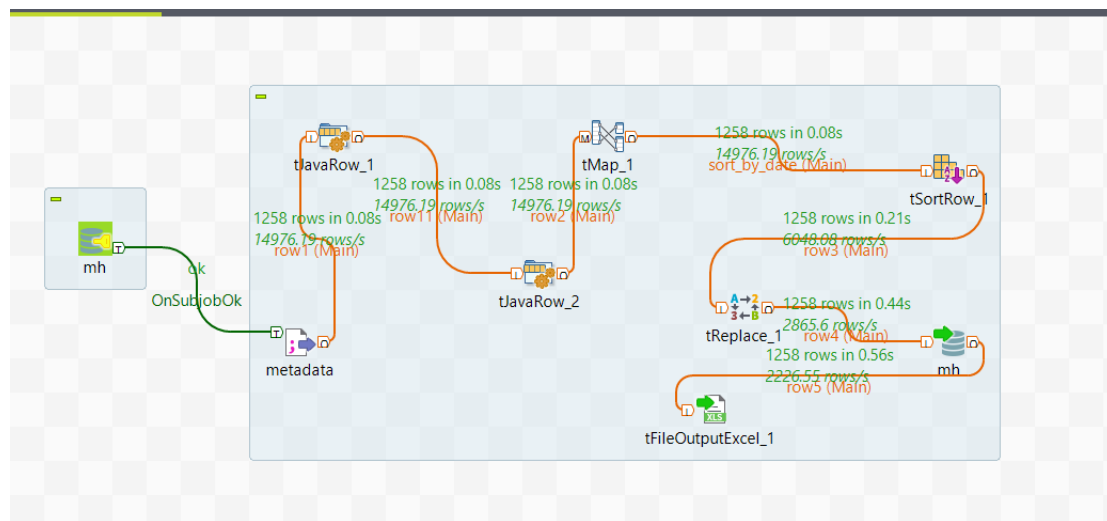


Figura 3: Job3 no *Talend*