Definição de Jobs no Talend
Grupo 10

César Silva
PG41842

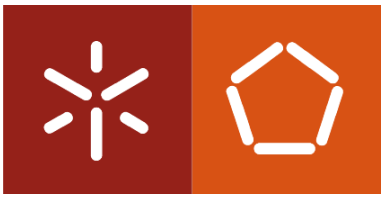
Hugo Gião
PG41073

Filipe Cunha
A83099

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 familia. 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*.

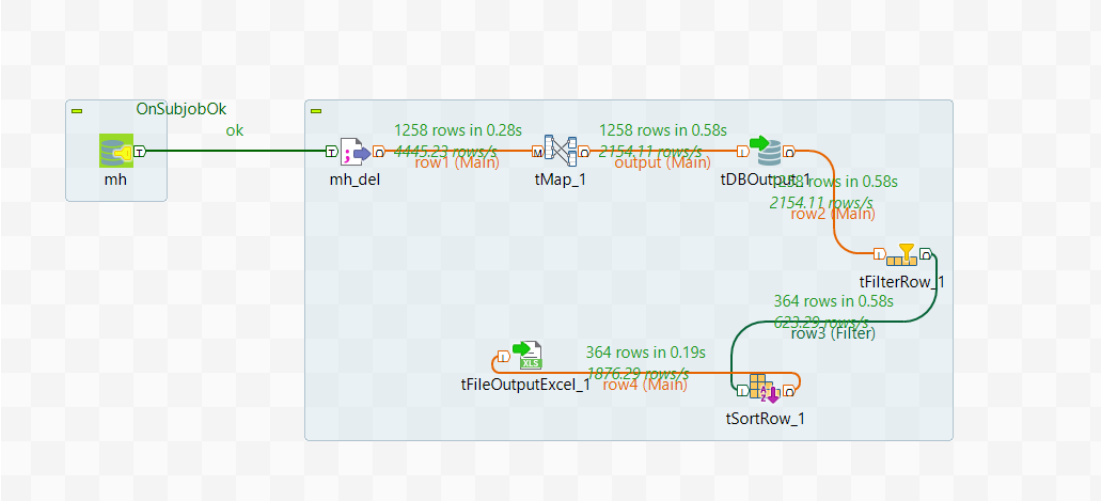


Figura 1: Pacientes em tratamento

Job 2

Neste Job carregamos a informação contida num ficheiro csv e mapeamos essa informação numa base de dados. Filtramos pela coluna de tratamento e escrevemos num ficheiro csv as linhas dos clientes que não tenham obtido tratamento. Procedemos por filtrar as linhas dos clientes que tenham obtido tratamento pela coluna coworkers, e guardada a informação dos pacientes que passam no filtro num ficheiro csv. A lista dos rejeites e posteriormente filtrada para remover os pacientes que não tenham partilhado informação com o seu supervisor e esta informação e guardada num ficheiro csv.

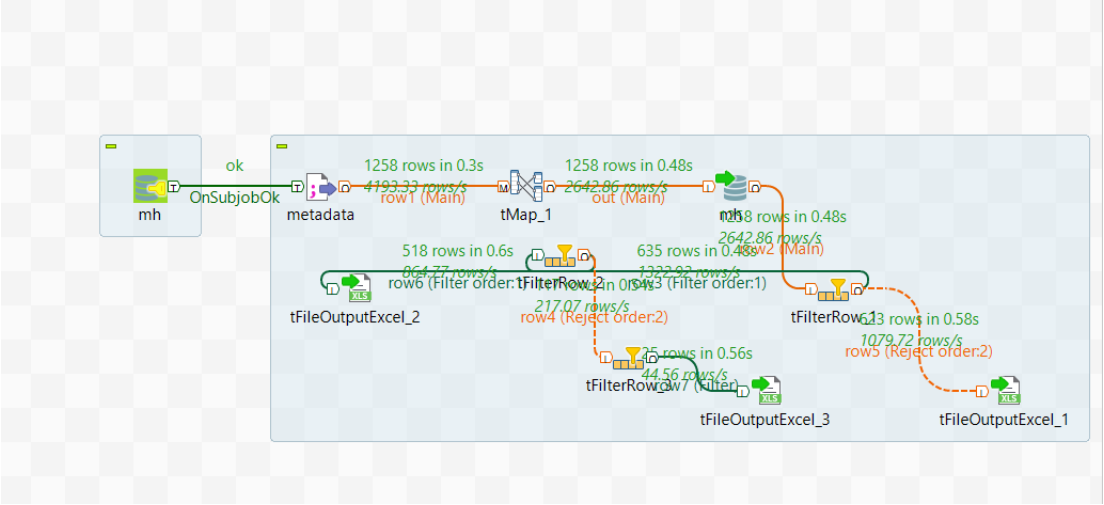


Figura 2: Caption

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, utilizando os métodos *TalendDate*, de seguida para normalizar a coluna dos géneros dos pacientes decidiu-se, converter 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* (dependendo da granularidade). Após a criação de estas colunas foi aplicado um *sort* ascendente a nível das mesmas obtendo o resultado ordenado por data, de seguida foi aplicado um replace para substituir todas as *strings* "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

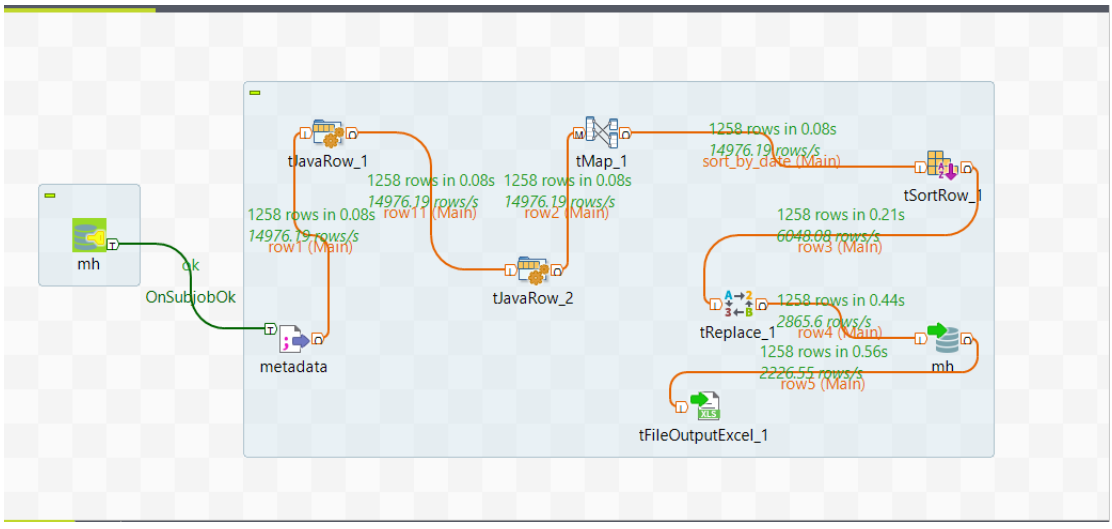


Figura 3: Caption