



Folha de Exercícios n.º PL05	
Unidade Curricular:	Aplicações Informáticas na Biomedicina
Ano Letivo:	2018/2019 - 1.º Semestre

1 Definição, Criação e Execução de Jobs no Talend

O ficheiro disponibilizado juntamente com esta ficha prática laboratorial, nomeadamente “mental_health.csv”, contém dados reais de um estudo realizado em 2014 que avalia as atitudes dos trabalhadores em empresas de Tecnologias da Informação (TIs) relativamente à saúde mental, bem como a frequência de transtornos mentais em ambientes de trabalho.

O *dataset* é constituído por 27 diferentes colunas, nomeadamente:

- *timestamp*;
- *age*;
- *gender*;
- *country*;
- *state* (*If you live in the United States, which state or territory do you live in?*);
- *self_employed* (*Are you self-employed?*);
- *family_history* (*Do you have a family history of mental illness?*);
- *treatment* (*Have you sought treatment for a mental health condition?*);
- *work_interfere* (*If you have a mental health condition, do you feel that it interferes with your work?*);
- *no_employees* (*How many employees does your company or organization have?*);
- *remote_work* (*Do you work remotely (outside of an office) at least 50% of the time?*);
- *tech_company* (*Is your employer primarily a tech company/organization?*);
- *benefits* (*Does your employer provide mental health benefits?*);
- *care_options* (*Do you know the options for mental health care your employer provides?*);
- *wellness_program* (*Has your employer ever discussed mental health as part of an employee wellness program?*);
- *seek_help* (*Does your employer provide resources to learn more about mental health issues and how to seek help?*);
- *anonymity* (*Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?*);
- *leave* (*How easy is it for you to take medical leave for a mental health condition?*);
- *mental_health_consequence* (*Do you think that discussing a mental health issue with your employer would have negative consequences?*);
- *phys_health_consequence* (*Do you think that discussing a physical health issue with your employer would have negative consequences?*);
- *coworkers* (*Would you be willing to discuss a mental health issue with your coworkers?*);
- *supervisor* (*Would you be willing to discuss a mental health issue with your direct supervisor(s)?*);
- *mental_health_interview* (*Would you bring up a mental health issue with a potential employer in an interview?*);

-
- *phys_health_interview* (Would you bring up a physical health issue with a potential employer in an interview?);
 - *mental_vs_physical* (Do you feel that your employer takes mental health as seriously as physical health?);
 - *obs_consequence* (Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?);
 - *comments* (Any additional notes or comments).

Tendo em conta o seu enquadramento teórico e prático sobre o processo de ETL e o Talend, o objetivo principal a ser alcançado com a resolução desta ficha prática laboratorial é a definição, a criação e a execução de diversas *jobs* com as componentes disponibilizadas na mesma.

Com esta ficha prática laboratorial, pretende-se que:

1. Defina três diferentes *jobs* e justique a utilidade de cada *job* com uma contextualização.
2. Crie as *jobs* definidas no Talend, recorrendo às componentes adequadas. No entanto, deverá incluir, no mínimo, os seguintes elementos:
 - (a) *File Components*: *tFileInputDelimited*, *tFileOutputExcel*;
 - (b) *Processing Components*: *tMap*, *tFilterRow*, *tSortRow*;
 - (c) *Databases Components*: *tDBConnection*, *tDBOutput*;
 - (d) Filtros de rejeição.
3. Corra cada *job* definida e criada.

A resolução desta ficha prática laboratorial deverá ser enviada por e-mail para os docentes da unidade curricular - marisa@di.uminho.pt e jmac@di.uminho.pt - até ao dia 13 de novembro de 2018. Um responsável por grupo deverá enviar num mail com título “Ficha5_AIB_GRUPOX”, substituindo “X” pelo número do grupo de trabalho, a pasta zipada do projeto com as *jobs* criadas, bem como um documento de apoio no formato .pdf com a contextualização e a explicação de cada *job* criada (com figuras).