

A LIRE AVANT DE RENSEIGNER LE DOCUMENT

Ce document servira à identifier les besoins en collecte et analyse de données pour la création d'indicateurs et à connaître les pratiques en analyse de données dans le cadre de Hubble. Cette synthèse permettra la création du modèle conceptuel pour la plateforme Hubble.

En fonction du travail de collecte et/ou d'analyse que vous avez fait, vous renseignerez uniquement les parties qui concernent votre travail. Nous vous demandons d'être le plus précis possible dans vos descriptions.

=====

Date de rédaction : 6/7/2016

Nom du rédacteur du document : Christophe Després, Pierre Jacoboni
Spécialités : Informatique

Cas d'étude Hubble : PACES

Scénario hubble : 2 et 3

Personnes impliquées pour la collecte et l'analyse : Christophe Després, Pierre Jacoboni

Période de la collecte : printemps 2016

Période de l'analyse :

=====

Dispositif d'apprentissage (Etude de cas de Hubble)

Type de dispositif : Première année Paces à Grenoble

Finalité de l'apprentissage : Préparation aux concours de fin de première année de médecine, odontologie, pharmacie et maïeutique.

Utilisation du dispositif et fonctionnalités : Cours, QCM, questions en ligne

Contexte de production de données :

(Au besoin indiquer les différents moments de la production (savoir si des données ont été produites sur plusieurs années))

Décrire la problématique posée et les objectifs de l'analyse :

Production des données :

Décrire le processus de production des données brutes :

Liste des variables initialement recueillies : Nom et Description

Plateformes/outils utilisés:

Points forts	Points faibles

Stockage des données:

Plateformes/outils utilisés:

Points forts de ces plateformes	Points faibles

Description des pré-traitements:

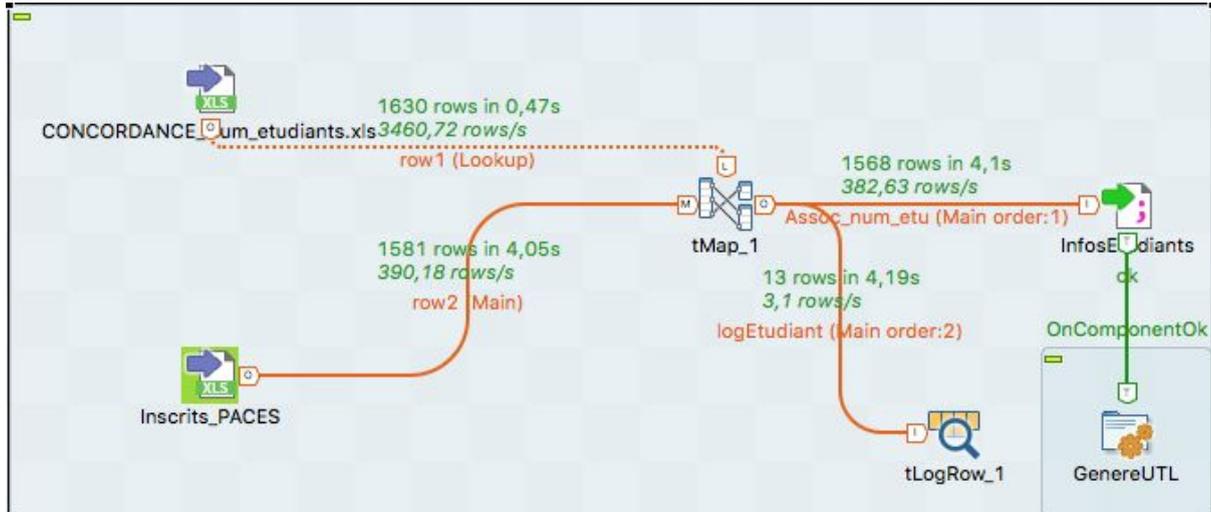
Objectifs des pré-traitements : A partir de l'arborescence de fichiers Excel et CSV de l'année 2011/2012 concernant les résultats aux QCM, les résultats aux concours, les questions en ligne et les informations des étudiants (origine, type de bac, etc.), construire des fichiers CSV synthétisant ces informations afin de les importer dans la plateforme UTL.

Décrire le processus de pré-traitement :

Les différentes étapes de pré-traitement présentées ci-dessous sont réalisées par des jobs dans la plateforme Talend Open Studio for Data Integration. Chaque job se termine par une étape appelée "GenereUTL" qui génère les fichiers modèle et données à importer dans UTL.

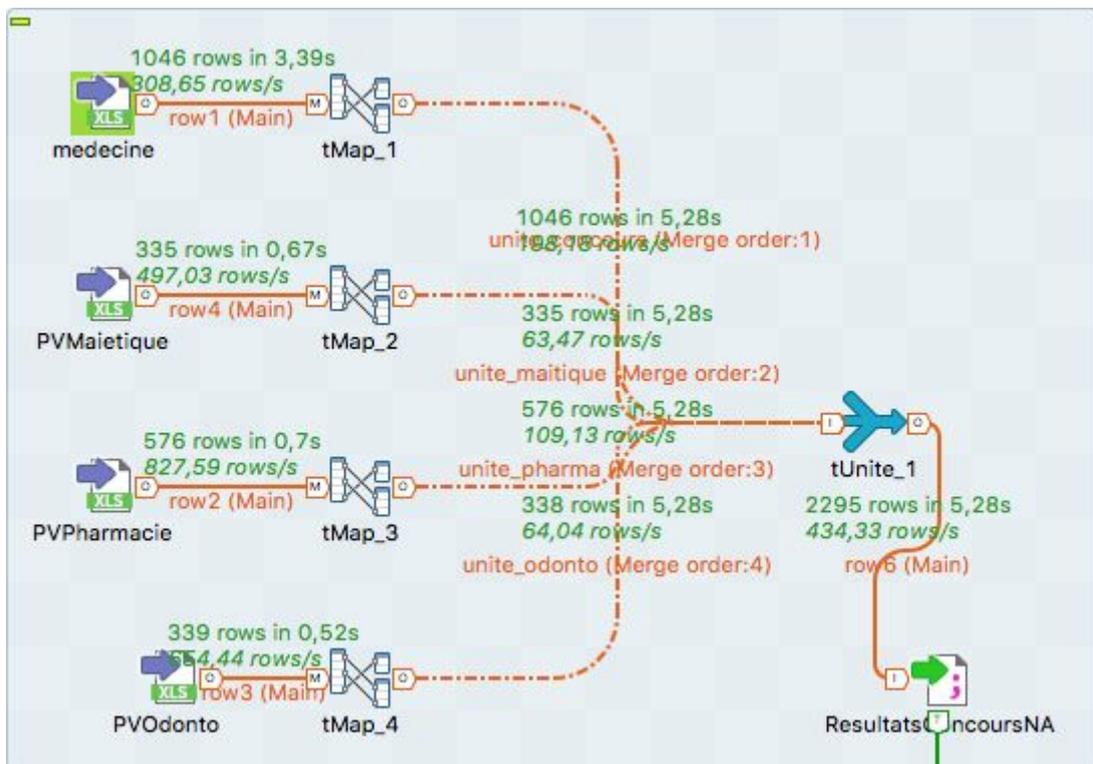
- Anonymer les informations sur les étudiants : Le fichier Inscrit_Paces.xls contient les informations concernant les étudiants inscrit (informations administratives, situation

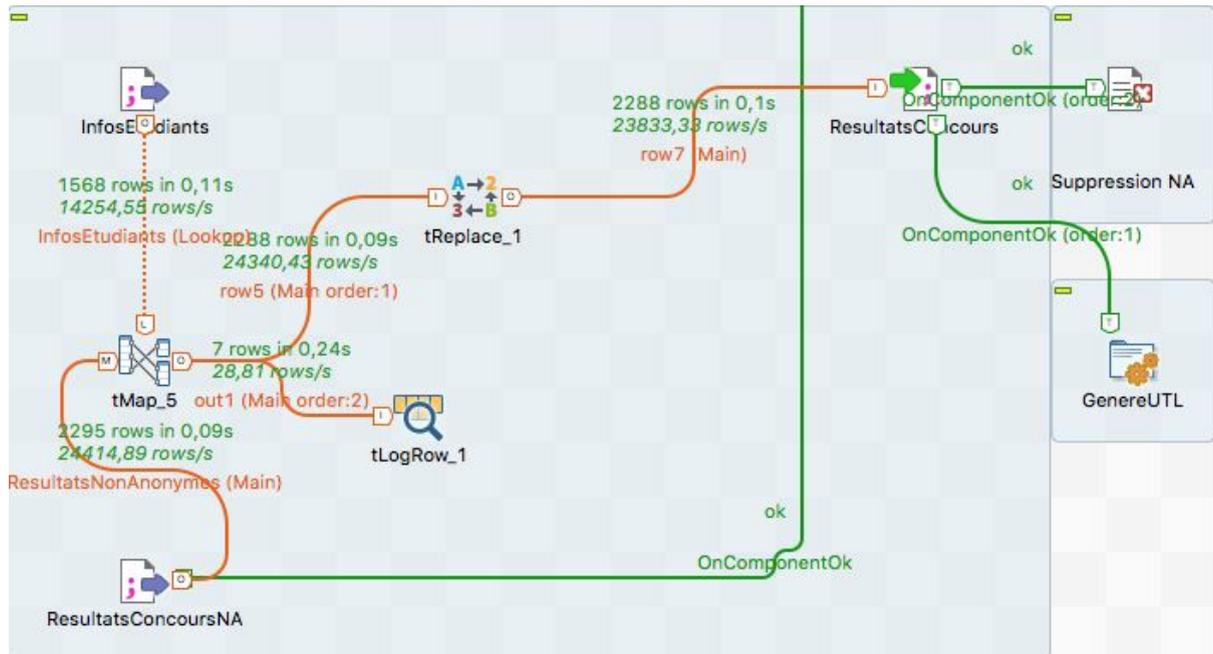
familiale, bac, etc.) référencées par un identifiant unique : le numéro de l'étudiant. Afin d'anonymiser on substitue ce numéro par un numéro d'anonymat issu du fichier CONCORDANCE_Num_etudiants.xls Ce numéro d'anonymat est l'identifiant unique utilisé dans les autres fichiers (résultats aux QCM, concours, etc.). Le job InfosEtudiant réalise cette étape d'anonymisation.



On constate que 13 étudiants n'ont pas de Numéro d'anonymat (Absents de CONCORDANCE_Num_etudiants.xlsx).

- Regrouper et anonymer les notes aux concours :
 - Partie 1 : regrouper toutes les notes aux concours dans un seul fichier dont les colonnes sont : Code_etu ; Note ; Resultat ; Classement ; Matière par exemple : 21113191 ; 3.883 ; AJ ; '799/1046 ; Résultat concours Med.
 - Partie 2 : anonymer les résultats

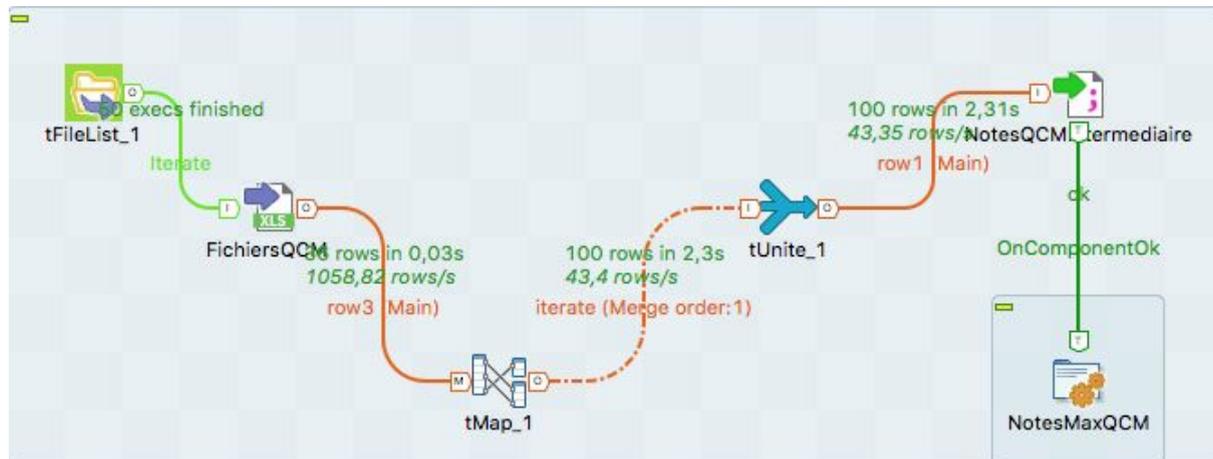


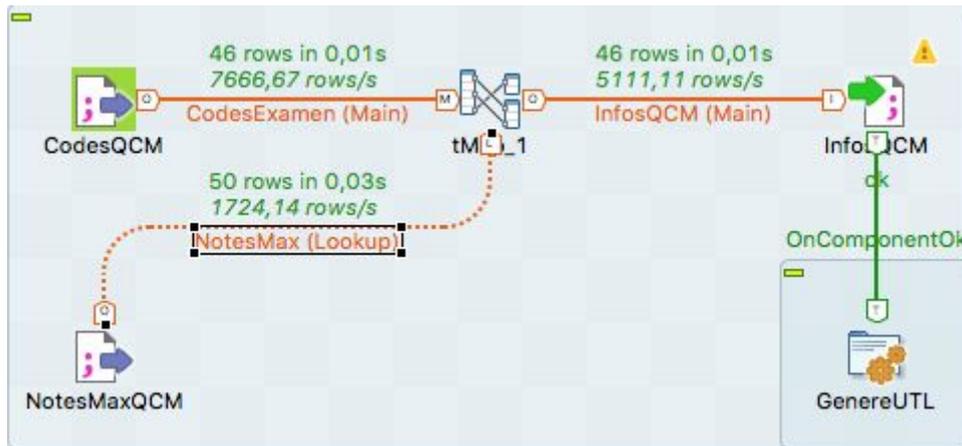


- Construire la table des notes maximales aux QCM : Les QCM sont organisés par semestre et par cycle (environ 48 QCM). Les QCM ont des notes maximales différentes. Afin de les normaliser dans l'étape suivante on construit une table contenant les notes maximales.

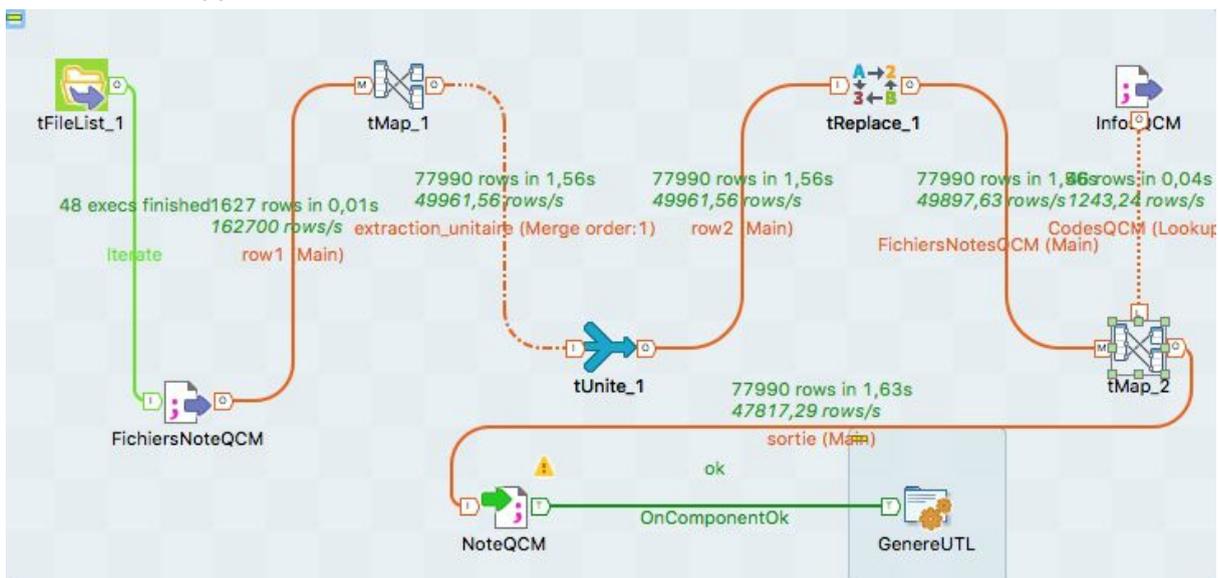
```

Libelle;NoteMax
UE1_BIOCHIMIE(1);22
UE1_BIOCHIMIE(1);22
UE3-1_BIOPHYSIQUE(1);30
UE3-1_BIOPHYSIQUE(1);30
UE2_HBDD(1);45
UE4_MATHEMATIQUE(1);16
UE4-BIOCELLULAIRE(1);44
UE1-BIOCHIMIE(2);20
  
```





- Normaliser les notes de QCM : Les résultats aux QCM sont répartis dans plusieurs fichiers (environ 48). On regroupe et normalise ces résultats dans un seul même fichier appelé NoteQCM.



- Job principal : Ce job centralise et ordonne l'exécution des job présentés ci-dessus puis se connecte à la base Exist d'UTL pour transférer automatiquement les modèles et données.

Plateformes/outils utilisés:

Talend Open Studio for Data Integration

Points forts	Points faibles
Réutilisabilité Job exportable (génération automatique de code java)	Investissement pour la prise en main

Description des analyses : (Faire une description de chacune des analyses conduites)

Liste des variables : Nom et Description

Liste des méthodes mise en œuvre :

Mode opératoire technique, logiciels utilisés

Résultats obtenus:

Points forts des analyses	Points faibles des analyses

Description des données produites au cours du traitement

Objectif de la création de ces nouvelles données :

Mode de calcul de ces variables :

Description des nouvelles variables : Nom et description

Nom	description

Description des Itérations

Pourquoi le processus d'analyse a été itéré ?

Points forts des itérations	Points faibles des itérations