

Table des matières

Table des matières.....	1
Objectif	2
Etape 0 : Choix de la problématique au niveau mondial	2
Etape 1 : ETL avec Talend et DatawareHouse avec Oracle	2
Etape 3 : OLAP avec Pentaho (Mondrian)	6
Etape 4 : Reporting avec BIRT	11
Conclusion et perspectives.....	12

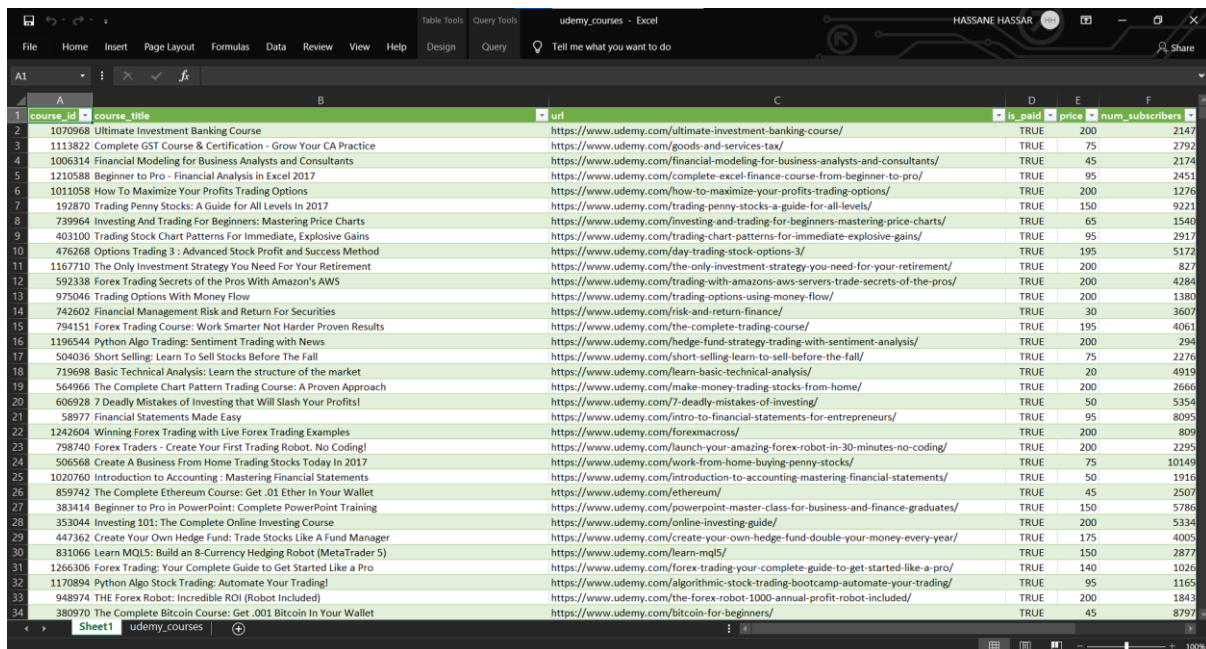
Objectif

Ce projet BI vise à mettre bout-a-bout les différentes notions vues dans cet élément : ETL, Dataware House, OLAP, et Reporting (Chaine décisionnelle).

Etape 0 : Choix de la problématique au niveau mondial

Tout d'abord, comme demande, nous avons choisi un **dataset** sous forme d'un fichier **CSV** depuis le site **Kaggle** qui s'agit d'un portal open data international. Notre problématique est claire d'étude autour des **cours UdeMy**.

Dataset URL: <https://www.kaggle.com/pranabpandit/udemy-courses>



course_id	course_title	url	is_paid	price	num_subscribers
1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-banking-course/	TRUE	200	2347
1111822	Complete GST Course & Certification - Grow Your CA Practice	https://www.udemy.com/goods-and-services-tax/	TRUE	75	2792
1006314	Financial Modeling for Business Analysts and Consultants	https://www.udemy.com/financial-modeling-for-business-analysts-and-consultants/	TRUE	45	2176
1210588	Beginner to Pro - Financial Analysis in Excel 2017	https://www.udemy.com/complete-excel-finance-course-from-beginner-to-pro/	TRUE	95	2451
1011058	How To Maximize Your Profits Trading Options	https://www.udemy.com/how-to-maximize-your-profits-trading-options/	TRUE	200	1276
192870	Trading Penny Stocks: A Guide for All Levels In 2017	https://www.udemy.com/trading-penny-stocks-a-guide-for-all-levels/	TRUE	150	9221
739964	Investing And Trading For Beginners: Mastering Price Charts	https://www.udemy.com/investing-and-trading-for-beginners-mastering-price-charts/	TRUE	65	1540
403100	Trading Stock Chart Patterns For Immediate, Explosive Gains	https://www.udemy.com/trading-stock-chart-patterns-for-immediate-explosive-gains/	TRUE	95	2917
476268	Options Trading 3 : Advanced Stock Profit and Success Method	https://www.udemy.com/how-to-trade-stock-options-3/	TRUE	195	5172
1167710	The Only Investment Strategy You Need For Your Retirement	https://www.udemy.com/the-only-investment-strategy-you-need-for-your-retirement/	TRUE	200	827
592338	Forex Trading Secrets of the Pros With Amazon's AWS	https://www.udemy.com/trading-with-amazons-aws-servers-trade-secrets-of-the-pros/	TRUE	200	4284
975046	Trading Options With Money Flow	https://www.udemy.com/trading-options-using-money-flow/	TRUE	200	1380
742602	Financial Management Risk and Return For Securities	https://www.udemy.com/risk-and-return-finance/	TRUE	30	3607
794151	Forex Trading Course: Work Smarter Not Harder Proven Results	https://www.udemy.com/the-complete-trading-course/	TRUE	195	4061
1196544	Python Algo Trading: Sentiment Trading with News	https://www.udemy.com/hedge-fund-strategy-trading-with-sentiment-analysis/	TRUE	200	294
504036	Short Selling: Learn To Sell Stocks Before The Fall	https://www.udemy.com/short-selling-learn-to-sell-before-the-fall/	TRUE	75	2276
719698	Basic Technical Analysis: Learn the structure of the market	https://www.udemy.com/learn-basic-technical-analysis/	TRUE	20	4919
564966	The Complete Chart Pattern Trading Course: A Proven Approach	https://www.udemy.com/make-money-trading-stocks-from-home/	TRUE	200	2666
608928	7 Deadly Mistakes of Investing that Will Slash Your Profits!	https://www.udemy.com/7-deadly-mistakes-of-investing/	TRUE	50	5354
58977	Financial Statements Made Easy	https://www.udemy.com/intro-to-financial-statements-for-entrepreneurs/	TRUE	95	8095
1242604	Winning Forex Trading with Live Forex Trading Examples	https://www.udemy.com/forexmacross/	TRUE	200	809
798740	Forex Traders - Create Your First Trading Robot. No Coding!	https://www.udemy.com/launch-your-amazing-forex-robot-in-30-minutes-no-coding/	TRUE	200	2295
506568	Create A Business From Home Trading Stocks Today In 2017	https://www.udemy.com/work-from-home-buying-penny-stocks/	TRUE	75	10149
1020760	Introduction to Accounting : Mastering Financial Statements	https://www.udemy.com/introduction-to-accounting-mastering-financial-statements/	TRUE	50	1916
859742	The Complete Ethereum Course: Get .01 Ether In Your Wallet	https://www.udemy.com/ethereum/	TRUE	45	2507
383414	Beginner to Pro in PowerPoint: Complete PowerPoint Training	https://www.udemy.com/powerpoint-master-class-for-business-and-finance-graduates/	TRUE	150	5786
353044	Investing 101: The Complete Online Investing Course	https://www.udemy.com/online-investing-guide/	TRUE	200	5334
447362	Create Your Own Hedge Fund: Trade Stocks Like A Fund Manager	https://www.udemy.com/create-your-own-hedge-fund-double-your-money-every-year/	TRUE	175	4005
831066	Learn MQL5: Build an 8-Currency Hedging Robot (MetaTrader 5)	https://www.udemy.com/learn-mql5/	TRUE	150	2877
1266306	Forex Trading: Your Complete Guide to Get Started Like a Pro	https://www.udemy.com/forex-trading-your-complete-guide-to-get-started-like-a-pro/	TRUE	140	1026
1170894	Python Algo Stock Trading: Automate Your Trading!	https://www.udemy.com/algo-trading-stock-trading-bootcamp-automate-your-trading/	TRUE	95	1165
948974	THE Forex Robot: Incredible ROI (Robot Included)	https://www.udemy.com/the-forex-robot-1000-annual-profit-robot-included/	TRUE	200	1843
380970	The Complete Bitcoin Course: Get .001 Bitcoin In Your Wallet	https://www.udemy.com/bitcoin-for-beginners/	TRUE	45	8797

Une fois le jeu de données choisi, nous avons comme chaîne de décision : **Talend, Oracle, Pentaho (Mondrian) et Birt**. Dans les étapes suivantes, nous allons présenter et/ou décrire brièvement le travail effectué.

Etape 1 : ETL avec Talend et DatawareHouse avec Oracle

Nous avons merge les deux étapes de la chaine décisionnelle dans une seule étape. Car, d'abord, elles sont étroitement liées et pour des raisons de rédactions et compréhension. Outre, même dans le cadre du TP nous l'avons les vus ensemble.

Nous sommes déjà familiarisés avec **Talend** puisque, nous l'avons manipulé dans les manipulations (TPS). Juste à titre indicatif et sorte de rappel :

« **Talend Open Studio** » est un ensemble de produits open source pour le développement, test, déploiement et administration des projets d'intégration de données et d'applications. Talend fournit une plateforme unifiée qui rend la gestion et l'intégration des données et applications plus facile, en fournissant un environnement unifié pour la gestion de tout leur cycle de vie. Il existe plusieurs solutions offertes par Talend :

- **Big Data** : Environnement qui facilite la gestion des données volumineuses.

- **Data Intégration** : Ensemble d'outils pour l'intégration de données pour accéder, transformer et intégrer les données à partir d'un système en temps réel pour remplir les besoins d'intégration des données.

- **Data Quality** : Permet d'assurer le profiling et monitoring des données pour identifier des anomalies et assurer la qualité des données.

- **ESB** : Permet la création, la connexion, la médiation et la gestion des services et leurs interactions.

Pour les besoins de notre mini-projet, nous utilisons « **Talend Data Integration** » pour la transformation des données et leur intégration.

S'agissant d'une synthèse, nous ne présenterons pas toutes les captures d'écran détaillées mais plutôt les principales étapes suivies.

Phase 1 : Préparation des sources de donnée

Etape 1 : Définition des sources de données dans le Référentiel pour pouvoir générer leurs schémas.

New Delimited File

File - Step 4 of 4

Add a Schema on repository
Define the Schema

Name: UdemyCourses

Comment:

Schema

Click to update schema preview

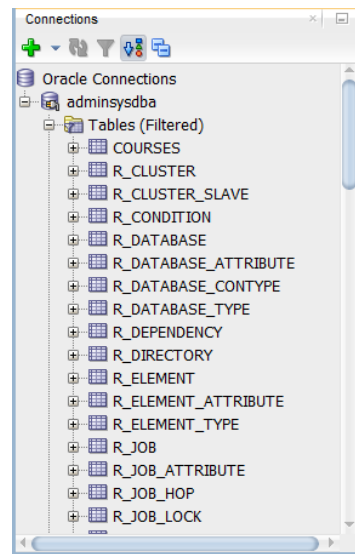
Guess

Description of the Schema

Column	K...	Type	✓ N..	Date Pattern (Ctr...	Length	Precision	Default	Comment
course_id	✓	Integer	✓		7	0		
course_title		String	✓		62	0		
url		String	✓		82	0		
is_paid		String	✓		75	0		
price		String	✓		92	0		
num_subscribers		String	✓		5	0		
num_reviews		Integer	✓		4	0		

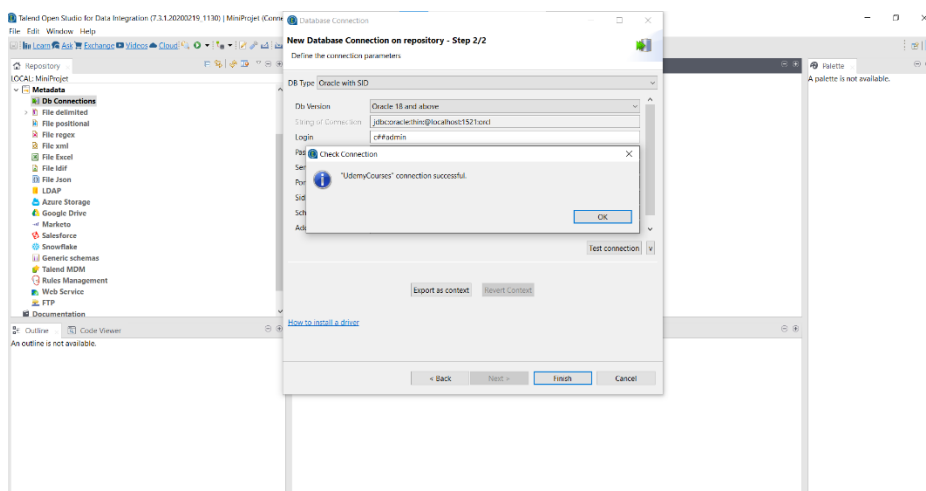
< Back Next > Finish Cancel

Etape 2 : Création de la base de données **adminsysdba** sous ORACLE contenant une table appelée **Courses** avec comme structure les mêmes colonnes que celle de notre source de données : **udemy_courses.csv**

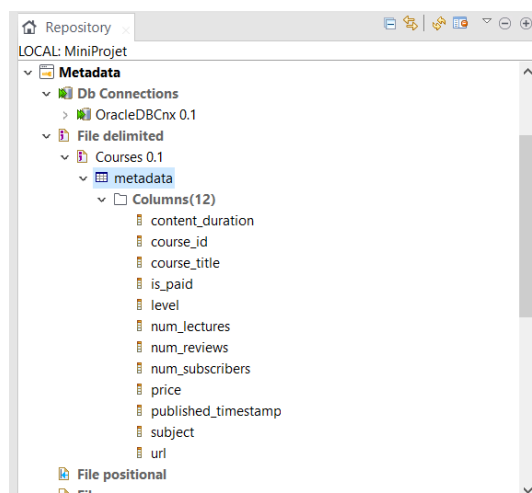


Oracle BD est un SGBDR comme MySQL qui utilise le **SQL** pour l'interrogation de ces données.

Etape 3 : Configuration de la connexion avec la BD créée dans l'étape précédente.

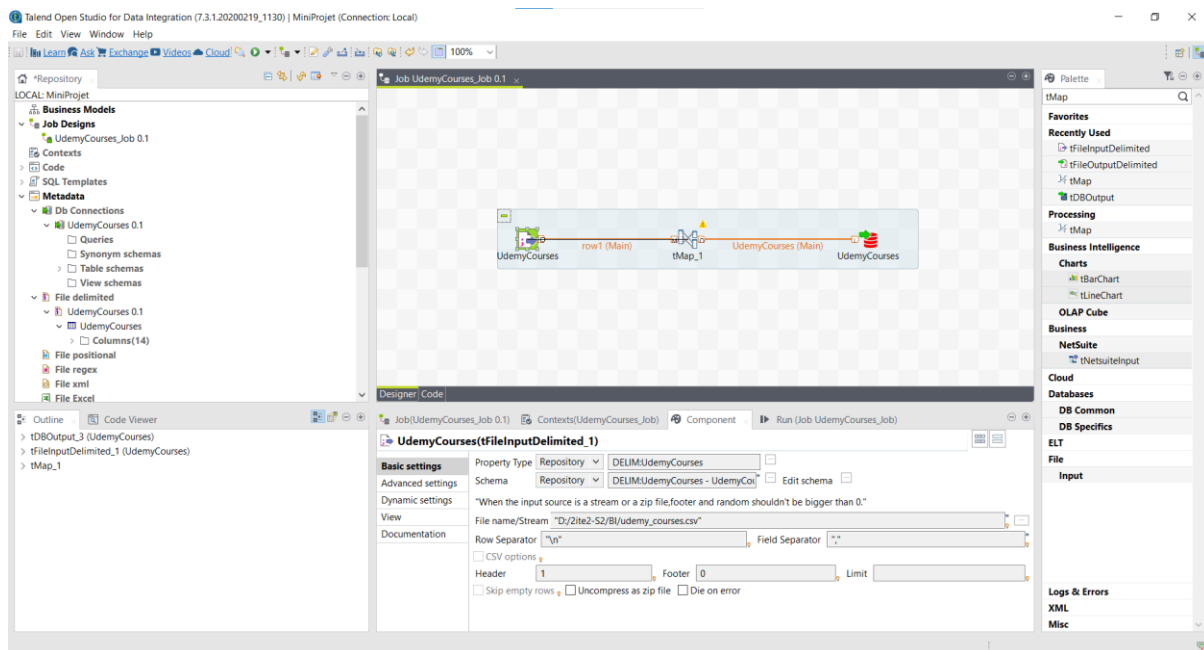


Etape 4 : Récupération du schéma

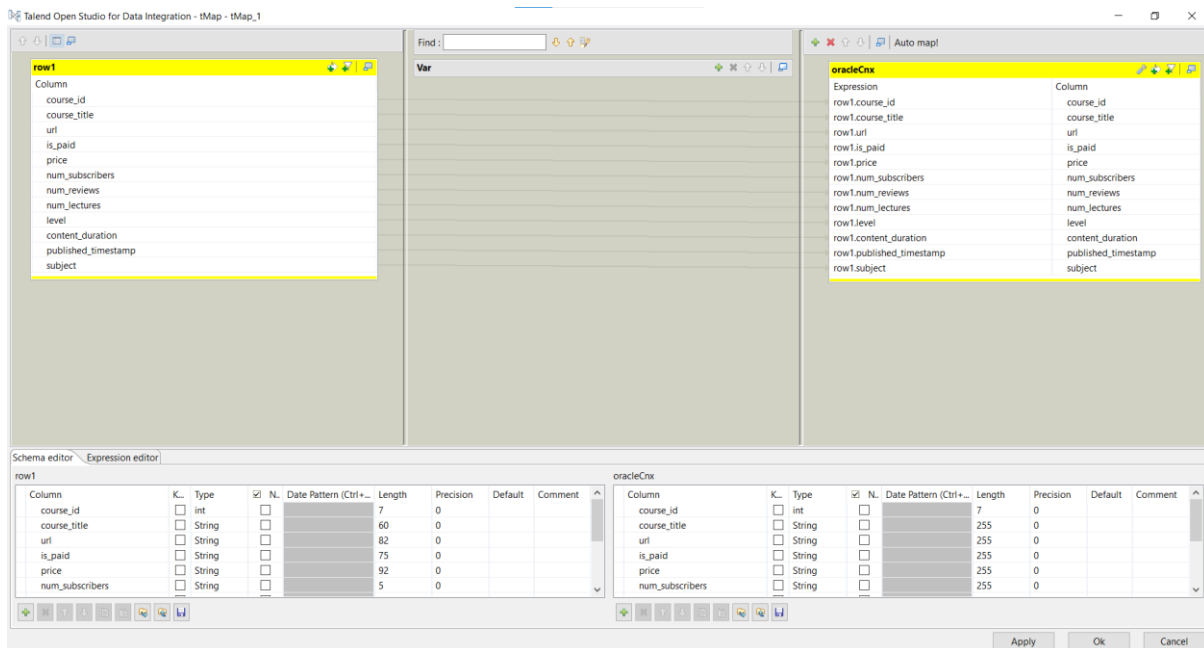


Phase 2 : Tri de documents

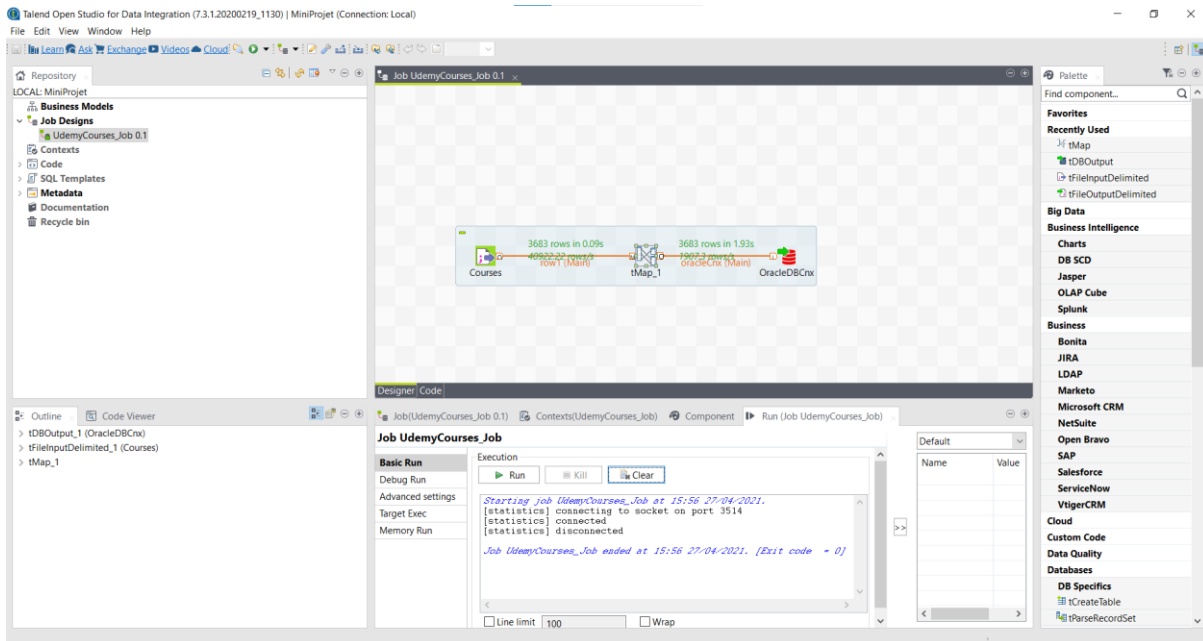
Etape 1 : Création d'un nouveau Job avec comme chaîne d'exécution contient trois composants : **tFileInputDelimited**, **tMap** et **tOracleDB** comme output.



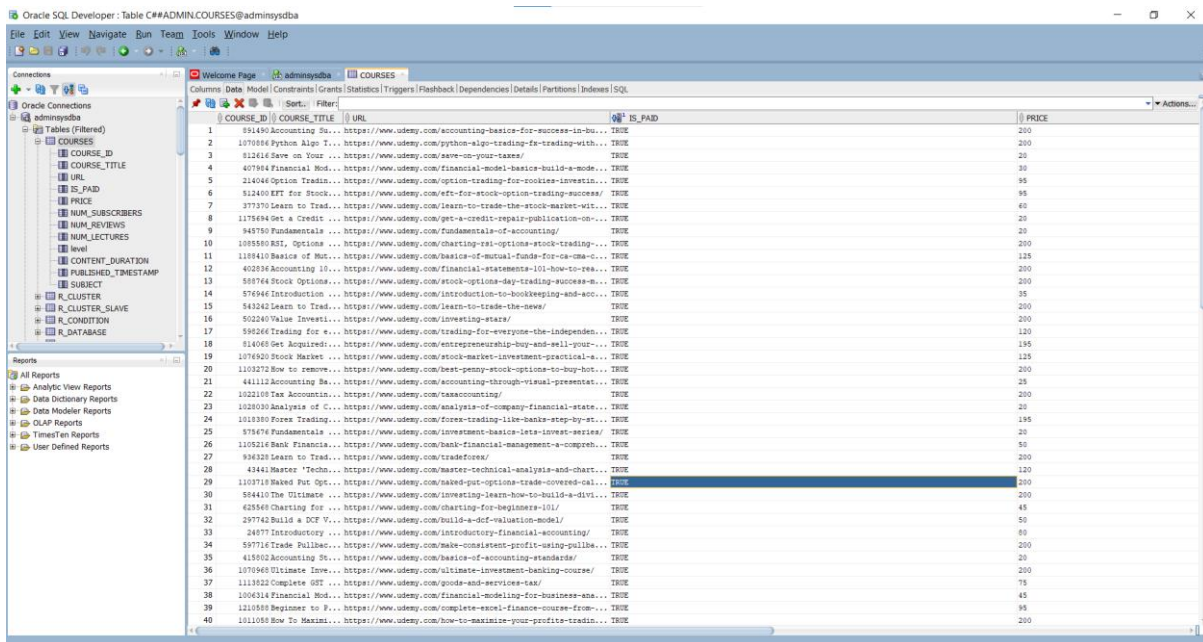
Etape 2 : Configuration du tMap



Etape 3 : Exécution du Job



Etape 4 : Vérification du chargement des données au niveau du DW(Oracle)



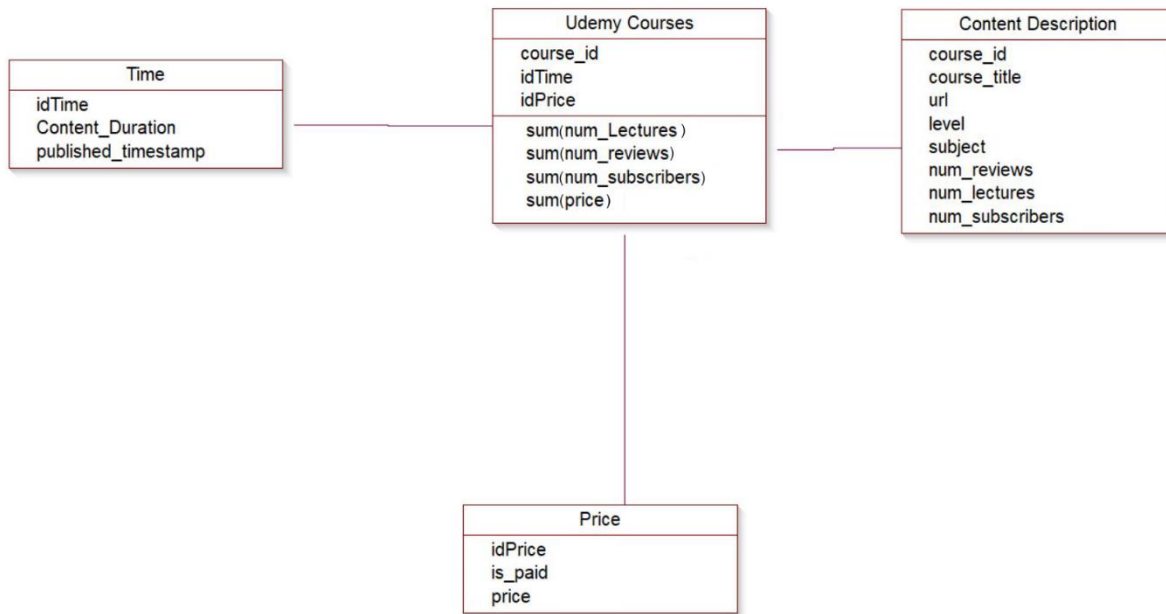
Comme conclusion, nous arriverons à intégrer et transformer les données et de les charger au niveau du DW. L'étape suivante concerne la phase OLAP.

Etape 3 : OLAP avec Pentaho (Mondrian)

Pentaho est un logiciel de business intelligence qui fournit l'intégration de données, des services OLAP, des rapports, des tableaux de bord d'information, l'exploration de données et des capacités d'extraction, de transformation et de chargement. Dans le cadre de ce mini projet, on va utiliser **Mondrian** pour la création des **Cube OLAP**.

Etape 0 : Modélisation conceptuelle et logique de schémas en étoile

Nous avons modélisé notre schéma en étoile suivant trois dimensions : Time, Content Description et Price.



Etape 1 : Installation de Pentaho Server et Schema Workbench

This PC > Local Disk (C:) > Pentaho > server > pentaho-server

Name	Date modified	Type	Size
third-party-tools	4/28/2021 12:12 PM	File folder	
tomcat	4/28/2021 12:15 PM	File folder	
Encr	9/7/2020 5:51 PM	Windows Batch File	2 KB
encr	9/7/2020 5:51 PM	Shell Script	2 KB
import-export	9/7/2020 5:51 PM	Windows Batch File	2 KB
import-export	9/7/2020 5:51 PM	Shell Script	2 KB
pentaho_start_user_console	9/7/2020 5:55 PM	Icon	287 KB
pentaho_stop_user_console	9/7/2020 5:55 PM	Icon	288 KB
pentaho_user_console	9/7/2020 5:55 PM	Icon	287 KB
promptuser	5/11/2021 10:13 AM	JavaScript File	1 KB
promptuser	4/28/2021 3:53 PM	Shell Script	1 KB
set-pentaho-env	9/7/2020 8:02 PM	Windows Batch File	7 KB
set-pentaho-env	9/7/2020 8:02 PM	Shell Script	6 KB
start-pentaho	9/7/2020 8:02 PM	Windows Batch File	3 KB
start-pentaho	9/7/2020 8:02 PM	Shell Script	3 KB
start-pentaho-debug	9/7/2020 8:02 PM	Windows Batch File	3 KB
start-pentaho-debug	9/7/2020 8:02 PM	Shell Script	3 KB
stop-pentaho	9/7/2020 5:51 PM	Windows Batch File	2 KB
stop-pentaho	9/7/2020 5:51 PM	Shell Script	2 KB

(x86)

)

Manage schema-workbench

Share View Application Tools

<< New Volume (D:) > Zite2-S2 > BI > MiniProjet > psw-ce-3.14.0.0-12 > schema-workbench >

Search schema-workbench

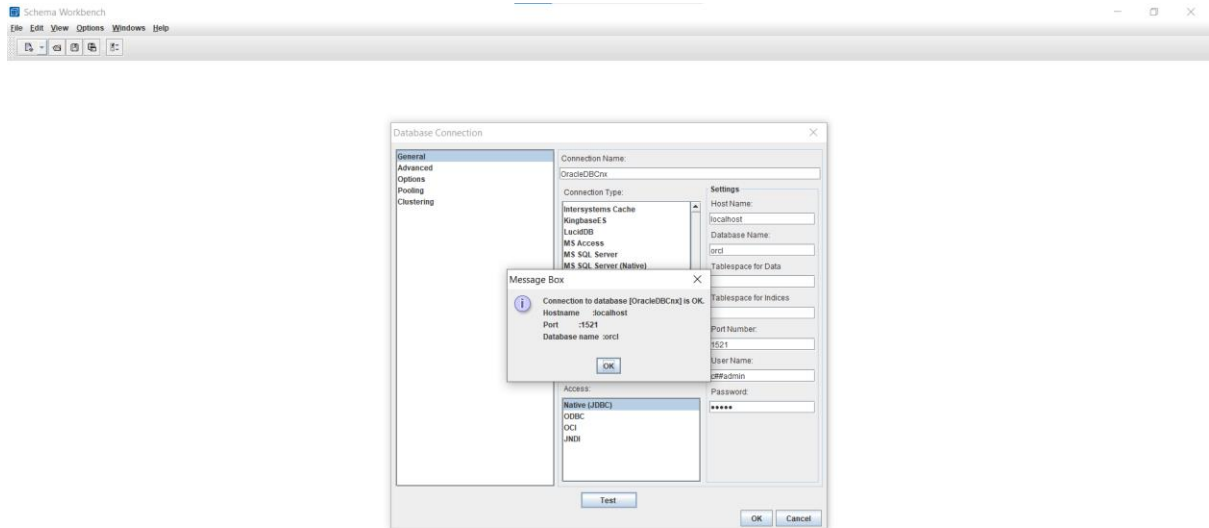
Name	Date modified	Type	Size
demo	5/16/2017 6:49 PM	File folder	
drivers	4/28/2021 6:17 PM	File folder	
lib	5/16/2017 6:49 PM	File folder	
plugins	5/16/2017 6:48 PM	File folder	
Schema_Workbench.app	5/16/2017 6:48 PM	File folder	
cpappend	5/16/2017 6:48 PM	Windows Batch File	1 KB
LICENSE	5/16/2017 6:48 PM	Firefox HTML Doc...	13 KB
log4j	5/16/2017 6:48 PM	XML Document	4 KB
mondrian.properties	5/16/2017 6:48 PM	PROPERTIES File	42 KB
PentahoSchemaWorkbench_OSS_Licenses	5/16/2017 5:58 PM	Firefox HTML Doc...	950 KB
readme	5/16/2017 6:48 PM	Text Document	1 KB
set-pentaho-env	5/16/2017 6:48 PM	Windows Batch File	5 KB
set-pentaho-env	5/16/2017 6:48 PM	Shell Script	4 KB
workbench	5/16/2017 6:48 PM	Windows Batch File	1 KB
workbench.command	5/16/2017 6:48 PM	COMMAND File	1 KB
workbench	5/16/2017 6:48 PM	Shell Script	2 KB

(C)

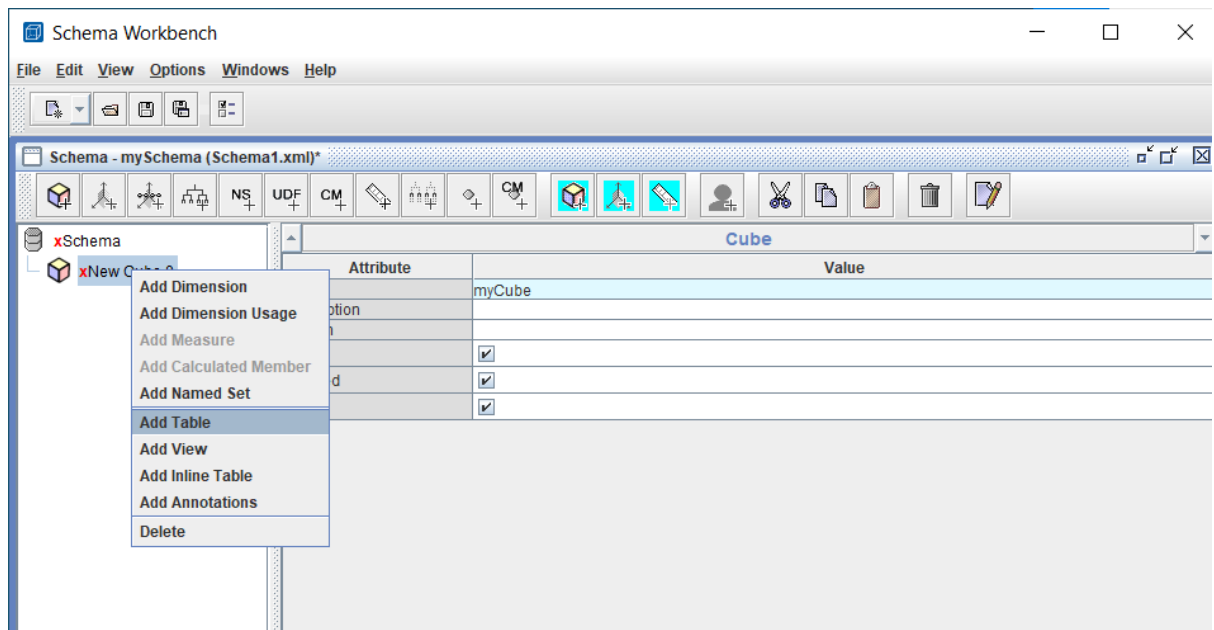
me (D)

Etape 2: Cube OLAP avec Schema Workbench

1. La connexion avec la BD a réussi



2. Creation du schema



3. Fichier Schema.xml

```

1 <Schema name="mySchema">
2   <Cube name="myCube" visible="true" cache="true" enabled="true">
3     <Table name="UDEMYCOURSES">
4     </Table>
5     <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Courses">
6       <Hierarchy name="identification" visible="false" hasAll="true" allMemberName="all Courses" primaryKey="pk_course">
7         <Level name="id" visible="true" column="COURSE_ID" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
8         </Level>
9         <Level name="title" visible="true" column="COURSE_TITLE" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
10        </Level>
11        <Level name="url" visible="true" column="URL" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
12        </Level>
13        <Level name="numLectures" visible="true" column="NUM_LECTURES" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
14        </Level>
15        <Level name="numReviews" visible="true" column="NUM_REVIEWS" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
16        </Level>
17        <Level name="numSubscribers" visible="true" column="NUM_SUBSCRIBERS" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
18        </Level>
19      </Hierarchy>
20    </Dimension>
21    <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Subject">
22      <Hierarchy name="SubjectHierarchy" visible="false" hasAll="true" allMemberName="all subjects" primaryKey="pk_subject">
23        <Level name="subject" visible="true" column="SUBJECT" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
24        </Level>
25      </Hierarchy>
26    </Dimension>
27    <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Level">
28      <Hierarchy name="LevelHierarchy" visible="false" hasAll="true" allMemberName="all levels" primaryKey="pk_level">
29        <Level name="level" visible="true" column="LEVEL" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
30        </Level>
31      </Hierarchy>
32    </Dimension>
33    <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Time">
34      <Hierarchy name="TimeHierarchy" visible="false" hasAll="true" allMemberName="all time" primaryKey="pk_time">
35        <Level name="Duration" visible="true" column="CONVENT_DURATION" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
36        </Level>
37        <Level name="Year" visible="true" column="PUBLISHED_TIMESTAMP" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
38        </Level>
39      </Hierarchy>
40    </Dimension>
41    <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Price">
42      <Hierarchy name="PriceHierarchy" visible="false" hasAll="true" primaryKey="pk_price">
43        <Level name="paid" visible="true" column="IS_PAID" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
44        </Level>
45      </Hierarchy>
46    </Dimension>
47    <Measure name="totLectures" column="NUM_LECTURES" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
48    </Measure>
49    <Measure name="totReviews" column="NUM_REVIEWS" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
50    </Measure>
51    <Measure name="totSubscribers" column="NUM_SUBSCRIBERS" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
52    </Measure>
53  </Cube>
54 </Schema>

```

4. Publication du schema

Schema Workbench

File Edit View Options Windows Help

Schema - mySchema (Schema.xml)

Attribute Value

Attribute	Value
name	myCube
description	
caption	
cache	<input checked="" type="checkbox"/>
enabled	<input checked="" type="checkbox"/>
visible	<input checked="" type="checkbox"/>

Publish

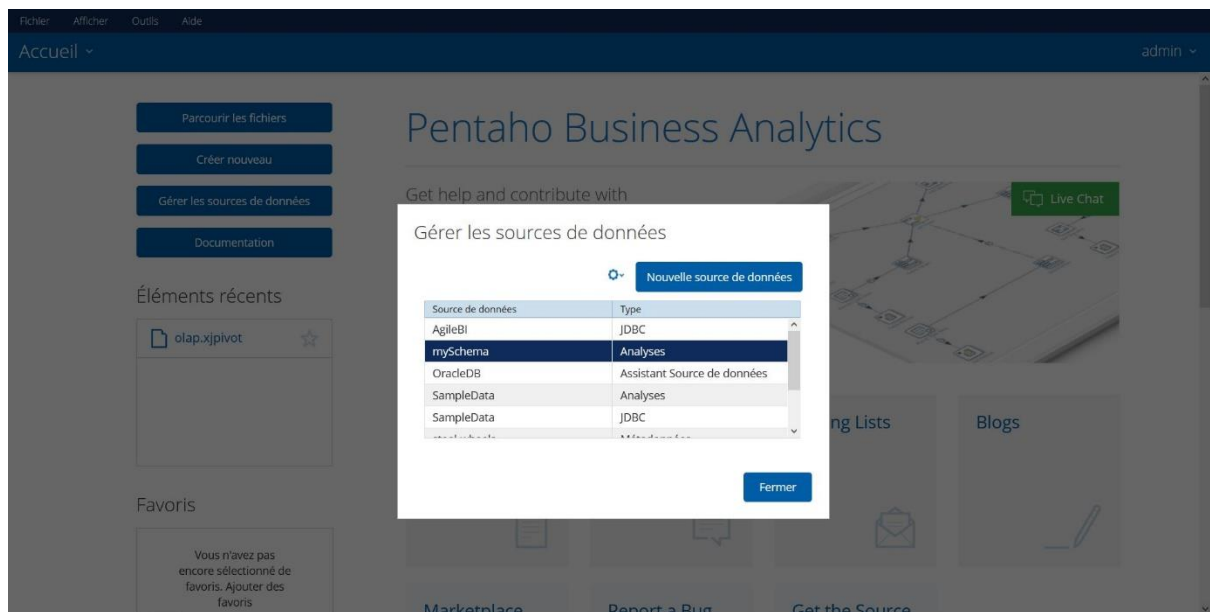
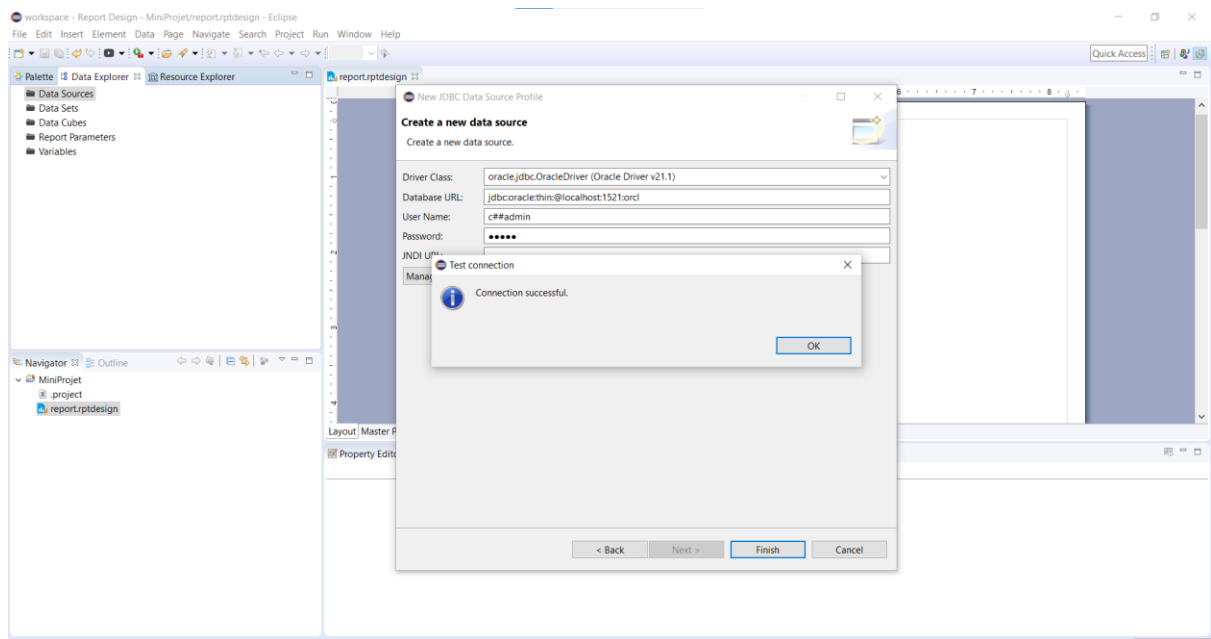
Publish Successful

OK

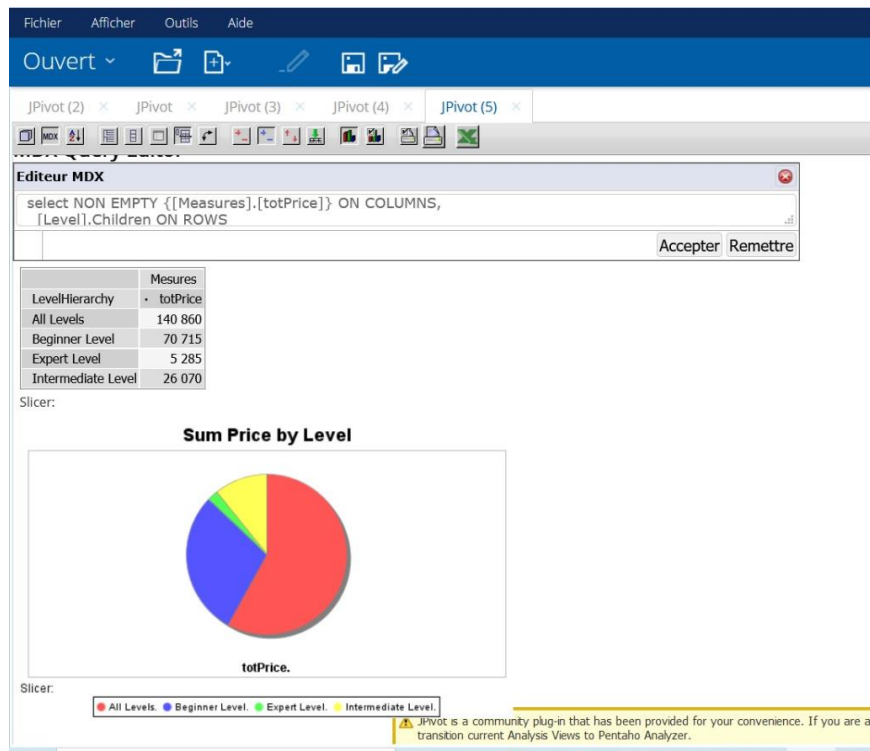
Database - (Oracle)

Measure totSubscribers is invalid

5. Connexion de Mondrian au BI server



6. Exemple d'un Vue OLAP et requête MDX avec Pentaho Mondrian server.

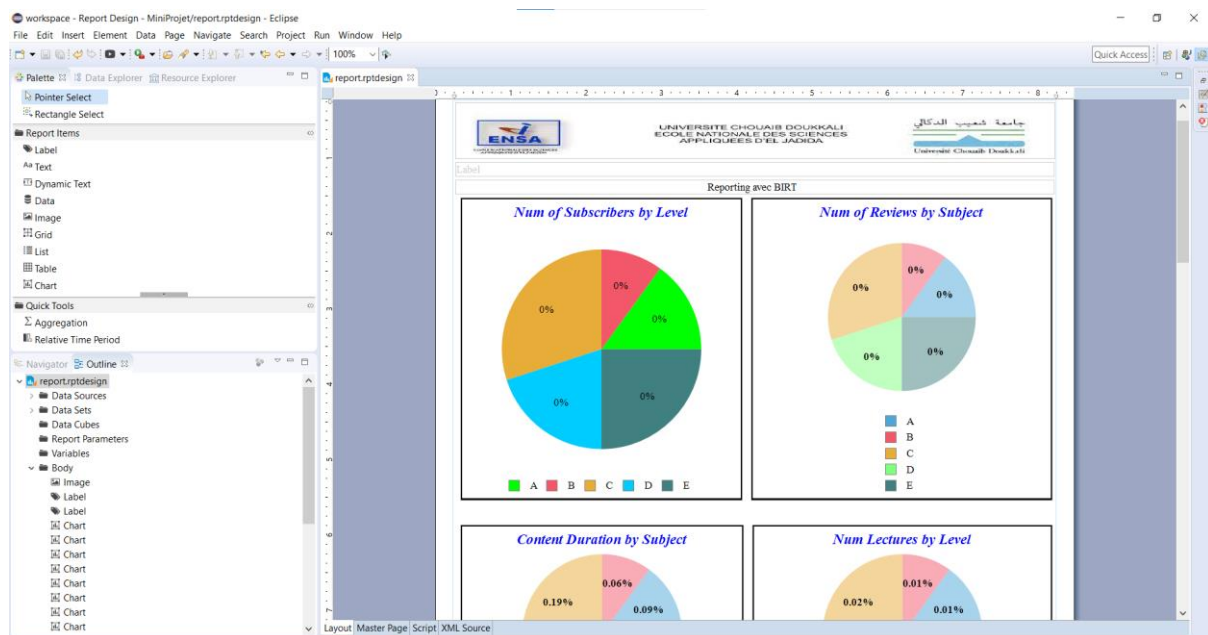


Etape 4 : Reporting avec BIRT

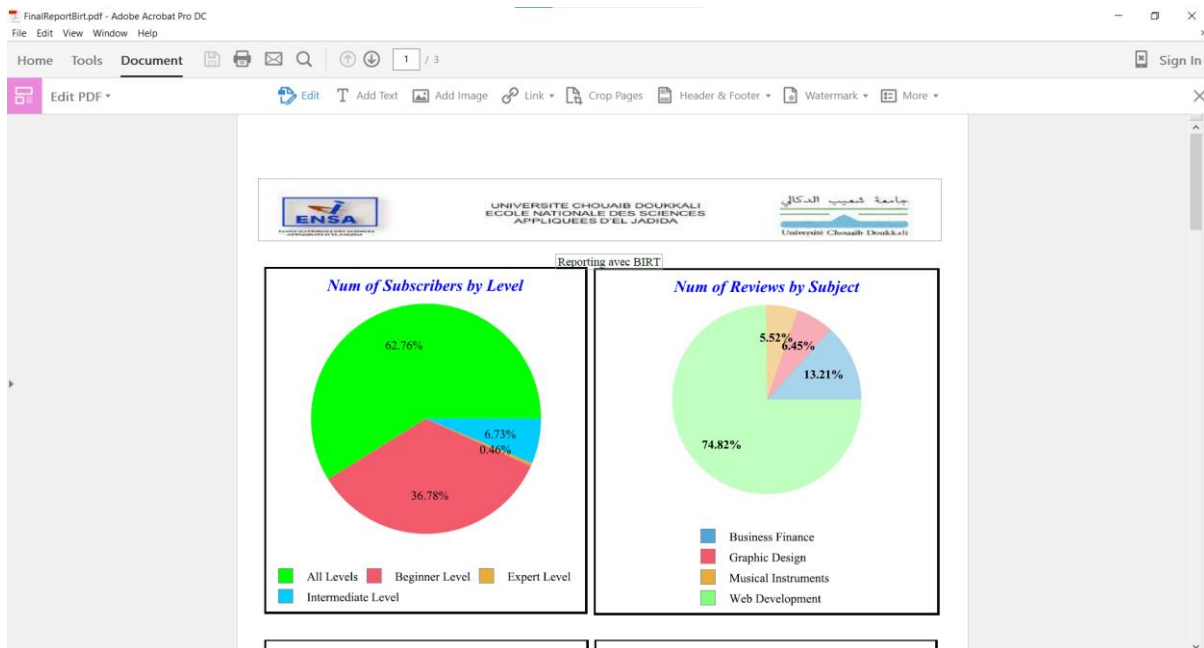
Eclipse BIRT est un système de Reporting open source basée sur la plateforme Eclipse qui s'intègre à votre application Java/Java EE pour produire des rapports convaincants.

En effet, **Eclipse BIRT** est très intuitive à maîtriser, nous avons commencé d'abord par la création de la connexion avec la **BD Oracle** et puis personnaliser notre rapport selon nos besoins.

Voici notre tableau de bord sous **Eclipse BIRT**



Le rapport sous format PDF



La phase de **restitution** constitue la dernière phase concerne la communication des résultats. Pour faciliter l'accès à l'information, pour tous les utilisateurs selon leurs profils métiers et afin d'extraire les éléments de décision pour dynamiser la réactivité globale dans l'entreprise.

« **Une image vaut mille mots** » c'est dans cet optique qu'on a exploite les outils de la visualisations (**Data visualisation**) : La Data Visualisation est une technique d'exploration et d'analyse des données numériques à l'aide de graphiques.

Ici, nous avons utilisé principalement le type **Pie Chart** : Un graphique à secteurs montre un nombre statique et la façon dont les catégories représentent une partie d'un tout - la composition de quelque chose. Un graphique à secteurs représente les nombres en **pourcentages** et la somme totale de tous les segments doit être égale à **100%**. C'est le plus adapte vu le jeu de données qu'on dispose.

N.B : L'idée derrière ces graphes c'est d'aider les décideurs pour mieux comprendre le comportement des clients et également l'amélioration future de la qualité des cours Udemy.

Conclusion et perspectives

Nous sommes très fiers d'avoir appliqué les différentes notions vues dans cet élément de BI : ETL, DatawareHouse, OLAP et Reporting, à travers un mini-projet concret.