

# Projet DataWarehouse – Thématique : Analyse mondiale

Le projet BI vise à mettre bout-à-bout les différentes notions vues dans la première moitié de l’enseignement : ETL, Datawarehouse, OLAP, Reporting. Il servira de base à un deuxième projet lié à la deuxième partie du cours (Data Mining), qui sera présenté plus tard.

Le projet doit analyser une problématique de votre choix au niveau mondial (santé, éducation, économie, politique, militaire, ....).

## 1 Organisation et contraintes

Les contraintes du projet sont les suivantes :

- Le projet est organisé en binôme (un ou deux groupes de 3 sont tolérés, avec une exigence plus conséquente).
- Le projet doit être centré sur une problématique claire d’étude (que vous devez définir)
- Le projet doit être construit à partir de la base de données relationnelle "world" <http://downloads.mysql.com/docs/world.sql.gz>
- Une partie des données doit être issue d’une capture Web (API, Open Data, extraction pages HTML (<http://www.garethjames.net/a-guide-to-web-scraping-tools/>)). Vous pouvez aussi exploiter des portal open data des organisations internationales :
  - <http://www.who.int/gho/database/en/>
  - <https://data.world/datasets/>
  - <https://datacatalog.worldbank.org/dataset/global-economic-monitor>
- ....
- Une attention particulière doit être portée à la dimension Localisation (et éventuellement Temps) qui doit être le plus fine possible

Au niveau du travail à effectuer, il est demandé :

- Une modélisation conceptuelle et logique (normalisée et dénormalisée) de schémas en étoile. Il sera apprécié d’avoir éventuellement plusieurs schémas et/ou un schéma en constellation agrégeant plusieurs schémas en étoile
- La construction du/des schémas en étoile + cubes OLAP associés
- Une analyse des données sous forme de graphiques simples (générés à partir de la sortie des cubes OLAP + outil de visualisation) doit être proposée
- Une visualisation cartographique doit être proposée

## 2 Le rapport

Le rapport doit inclure (entre autres) :

- La problématique ciblée (contexte et motivation)
- Les données à disposition (sources et processus d’extraction)
- La modélisation des données (schémas en étoile + en constellation)
- La démarche de développement de construction des cubes OLAP
- Les différentes analyses et visualisation

Référence : <http://jplusplus.github.io/guide-du-datajournalisme/>

## 3 Deadlines

Date de rendu = 25 mars 2019

Envoi par mail : [laure.soulier@lip6.fr](mailto:laure.soulier@lip6.fr) & [Clara.Gainon-de-Forsan-de-Gabriac@lip6.fr](mailto:Clara.Gainon-de-Forsan-de-Gabriac@lip6.fr)

Sujet du mail : [DAC-BI] PROJET BI 2019 nom1 nom2