

## Projet BI

Thématique : Les infrastructures en France

### Introduction

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.

### Organisation et contraintes

Le projet est organisé en binômes. Sa thématique est centrée sur les infrastructures en France et doit donc s'y rapporter d'une façon ou d'une autre. Par infrastructure, on entend les hôpitaux, écoles, universités, administrations publiques, transports, bibliothèques...

Les contraintes du projet sont les suivantes :

- Le projet doit être centré sur une problématique claire d'étude (que vous devez définir)
- Le projet doit être construit à partir de plusieurs tables de données extraites depuis le site ParisData, data.gouv, ou tout autre portail OpenData
- Une partie des données doit être issue d'une capture Web
  - A travers une API
  - A travers d'autres sources de données
  - A travers de l'extraction de pages HTML
    - <http://www.garethjames.net/a-guide-to-web-scraping-tools/>
- Une attention particulière doit être portée à la dimension Localisation qui doit être la plus fine possible
- Le projet doit viser à
  - la modélisation conceptuelle et logique (normalisée et dénormalisée) de 3 schémas en étoile
  - la modélisation conceptuelle et logique du schéma en constellation agrégeant les 3 schémas en étoile précédents (on doit avoir au minimum 2 dimensions partagées entre paires de faits)
  - la construction des 3 schémas en étoile + 3 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

### 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/>

Date de rendu = 19 mars

Envoi par mail : [laure.soulier@lip6.fr](mailto:laure.soulier@lip6.fr)

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