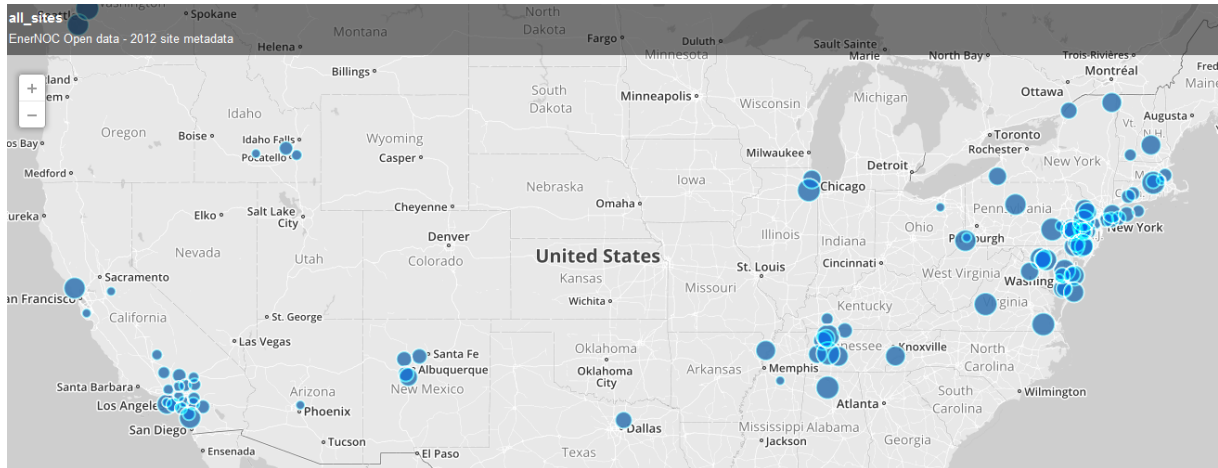


Projet Mastère SIO 2018 : Hadoop

Données utilisées : données ENERNOC – 100 Load Curves – sampling : 5-minutes – period : 2012

<https://open-enernoc-data.s3.amazonaws.com/anon/index.html>



Pour information, une chronique de consommation électrique est appelée « Load Curve » en anglais et « Courbe de Charge » (notée CdC) en français.

Objectifs :

- **Intégration des données** : expliquer comment les données ont été récupérées sur le site d'EnerNOC et comment elles ont été intégrées dans une base HIVE
- **Modélisation de données** : Quel modèle et mode de représentation des données pour votre base HIVE ? Pourquoi ? Quels verrous voyez-vous si vous deviez gérer les mêmes données pour 10 millions de sites ?
- **Requêtage + fonction utilisateur** : Réaliser les requêtes suivantes avec Hive/Spark
 - Calculer la CdC totale des 100 sites (pas de temps 5 minutes)
 - Calculer la CdC moyenne par secteur d'activité (pas de temps 5 minutes)
 - Calculer la CdC totale des 100 sites (pas de temps hebdomadaire)
 - Calculer la CdC moyenne par secteur d'activité (pas de temps hebdomadaire)
 - Classer les secteurs d'activité par intensité énergétique (ratio énergie/surface) : au total, puis par saison (hiver, printemps, été, automne)
 - Faire la même chose en développant une fonction utilisateur qui calcule directement l'intensité énergétique
 - Trouver la journée la plus énergivore par site
- **Dashboard Tableau** (<http://www.tableausoftware.com/academic/students>) : Imaginer et réaliser un tableau de bord informatif sur ce jeu de données (tableau de synthèse, projection sur une carte des consommations moyennes ou de l'intensité énergétique, comparer un client par rapport à la population ou à son secteur d'activité, etc.). Racontez-nous l'histoire des données (voir : storytelling de Tableau Software).

- **Croisement de données** : récupérer la météo (températures US 2012) + explorer la thermosensibilité par secteur d'activité (corrélation entre consommation électrique et température). Ci-dessous des exemples de sites fournissant des données météo :
<ftp://ftp.ncdc.noaa.gov/pub/data/noaa>
<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/ish-format-document.pdf>

Critères d'évaluation :

- ✓ Rapport : Expliquer clairement la démarche, les difficultés, les solutions apportées + code
 - ✓ Présentation 20/30 minutes : Faire une synthèse du travail réalisé + présenter le DashBoard réalisé.
- ➔ Formez 5 groupes (4 de 2 personnes et 1 de 3 personnes) et envoyez le nom des membres par e-mail à : sio.bigdata@gmail.com