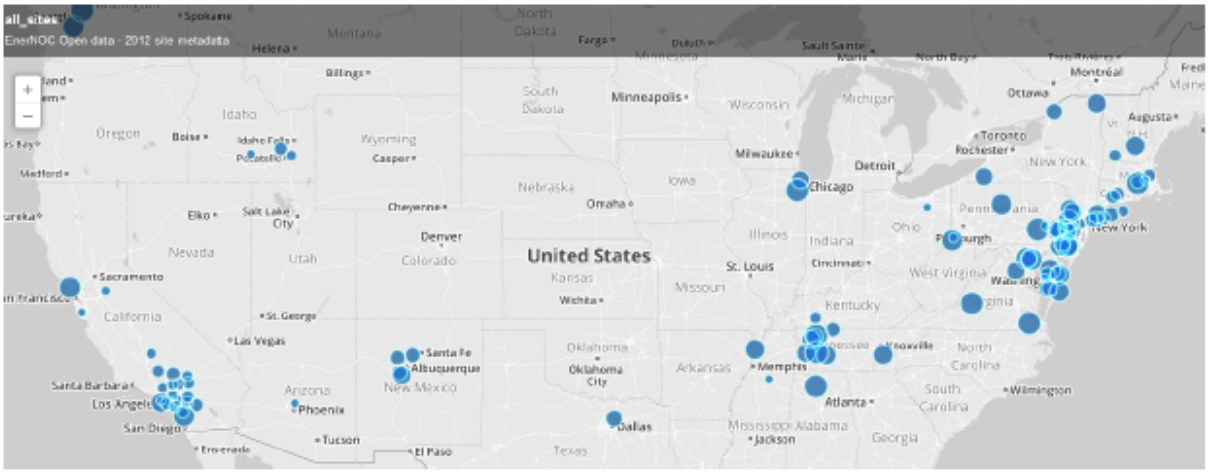
# Big Data Project with NoSQL DB

Data to be analyzed:

**ENERNOC electricity consumption data**

Anonymized 5-minute energy usage data for 100 commercial/industrial sites for 2012. Metadata for each site is provided including ID, industry, square footage, lat/lng and timezone. Notice that because latitude/longitude are randomized by a small amount, the exact location of a site on a map may not appear correct (e.g. the point might be over a body of water, etc.)

Can be downloaded from:
https://open-enernoc-data.s3.amazonaws.com/anon/index.html




For your information: a time series of electricity consumption is called a “Load Curve” (noted ”LD” in the following).

Objectives of the project:

- **Data integration :** explain how the data has been recovered from the EnerNOC website and how it has been integrated into MongoDB (or any chosen NoSQL DB). you can alsø choose to load the data directly into a file system such as HDFS and use Hive for querying
- **Data Modeling:** What data model and representation model should you use in your NoSQL database? Why?
- **Query: define and run the following queries with your chosen NoSQL DB**
    - Try some simple queries (5 simples queries): SELECT queries to explore your data,

ORDER BY to sort data, etc. You can use what we have seen in the MongoDB Lab for example

- Calculate the sum LD for the 100 sites (timestamp interval: 5 minutes)
- Calculate the average LD by sector of activity (timestamp interval: 5 minutes)
- Calculate the total LD for the 100 sites (timestamp interval: a week)
- Calculate the average LD by sector of activity (timestamp interval: a week)

- **Bonus (to go further: Tableau Dashboard**

(http://www.tableausoftware.com/academic/students)**):** Imagine and develop an informative dashboard on this dataset (summary table, projection on a map of average electricity consumption or energy intensity, compare a customer to the population or its sector of activity, etc.).

**Evaluation Criteria:**

- **Report:** clearly explain the objectives, the difficulties, justify the solutions of your approach and present the code developed.

**deadLine : June, 17th 2018**

**To do:**

- This project can be done alone or in a group of 2-3 people
- Submit your report on moodle