

OpenStreetMap dans l'Enseignement Supérieur et la Recherche

Trois exemples d'utilisation d'OpenStreetMap dans l'ESR

Sylvain Bouveret Anthony Hombiat

Soirée du groupe local OpenStreetMap

Coop-Infolab, lundi 16 janvier 2017





Introduction

Préliminaires techniques

Le modèle de données

L'ontologie OSM

OSM à l'Ensimag

Contexte pédagogique

Infrastructure technique

Contenu détaillé du TP

Retours sur le TP

OSM et (big) dataviz

Défi technique

Sous le capot

Import OSM

Suite

OSM, les attributs et le web sémantique



Introduction

Préliminaires techniques

- Le modèle de données
- L'ontologie OSM

OSM à l'Ensimag

- Contexte pédagogique
- Infrastructure technique
- Contenu détaillé du TP
- Retours sur le TP

OSM et (big) dataviz

- Défi technique
- Sous le capot
- Import OSM
- Suite

OSM, les attributs et le web sémantique



- ▶ **Données** : OpenStreetMap n'est pas une carte, c'est une **base de données**.
 - ▶ La « carte » en ligne n'est fournie qu'à des fins de visualisation
 - ▶ Pas de vue satellite ni aérienne (ce n'est pas l'objectif)
- ▶ **Ouvert** : les données sont utilisables sans restriction (ODbL).
- ▶ **Collaboratif** :
 - ▶ nombreuses contributions individuelles
 - ▶ plus on est de contributeurs, plus on rit, et accessoirement plus on complète la couverture mondiale



État courant de la base (15 janvier 2017)

Nb utilisateurs	3 378 964
Nb de points GPS versés	5 526 744 423
Nb de nœuds	3 689 727 068
Nb de chemins	387 495 900
Nb de relations	4 712 062

Big data ?



Ouvert \Rightarrow on peut récupérer facilement des données...

- ▶ OSM API v0.6 <http://api.openstreetmap.org/api/0.6/way/23000671>
 - ▶ des requêtes HTTP GET, PUT, POST, ou DELETE
 - ▶ des données renvoyées sous forme XML (facilement exploitable)
http://wiki.openstreetmap.org/wiki/OSM_Protocol_Version_0.6/DTD
- ▶ Overpass API http://wiki.openstreetmap.org/wiki/Overpass_API
 - ▶ API lecture seule optimisée
 - ▶ un langage de requête puissant avec des critères de filtrage par location, type d'objets, attributs,...



Sans passer par des API, on peut aussi télécharger une image de la base de données...

- ▶ <http://download.geofabrik.de/osm/> : fichiers OSM (format XML) par zone géographique, ville, pays...
- ▶ <http://download.cloudmade.com/> : Garmin Map files, Shape files, TomTom POI, Adobe Illustrator,..., par zone géographique, ville,...
- ▶ <http://planet.osm.org/> : Fichiers OSM + changesets.

- ▶ **OSM** files: [planet.osm](http://planet.osm.org/) pour la planète entière (actuellement + de 54Go compressé, + de 600Go décompressé), ou extraits régionaux.
- ▶ Diffs : changesets pour des mises-à-jour régulières.



OpenStreetMap est un projet **contributif**

↪ *Volunteered Geographic Information*

Données conventionnelles	Données VGI
Qualité des données ✓	Qualité aléatoire ✗
Qualité garantie ✓	Pas de garantie sur la qualité ✗
Processus de production lourd ✗	Processus de production léger ✓
Réactivité faible ✗	Rapidité, fraîcheur ✓
Restrictions (usage) ✗	Libre, ouvert ✓

- ▶ S'il est important d'avoir des garanties sur la qualité (ou quelqu'un qui en répond) ↪ **données conventionnelles**
- ▶ Si la fraîcheur et la liberté d'utilisation est plus importante (e.g. **gestion de crise**), **données VGI**.



- ▶ Libre, gratuite (*i.e.* compatible avec l'idée même de recherche) ⇒ reproductible et dans un certain sens, vérifiable
- ▶ Extrêmement complète et à jour par endroits
- ▶ Gros volumes de données exploitables

Source de données inestimable pour la recherche en géomatique, en géographie, en informatique, en sociologie, économie,...

Mais aussi une mine de sujets de recherches : traitement efficace de gros volumes de données, évaluation de la qualité des données, analyse des pratiques de contribution volontaire...



Introduction

Préliminaires techniques

Le modèle de données

L'ontologie OSM

OSM à l'Ensimag

Contexte pédagogique

Infrastructure technique

Contenu détaillé du TP

Retours sur le TP

OSM et (big) dataviz

Défi technique

Sous le capot

Import OSM

Suite

OSM, les attributs et le web sémantique



- ▶ **Éléments** : primitives du modèle

Trois types d'éléments

1. **nœud (node)** : brique de base.

- Point géographique: latitude & longitude (WGS84)
→ Point d'intérêt, partie d'un chemin



2. **chemin (way)** : suite ordonnée de nœuds

chemin ouvert → entités linéaires (routes, voies ferrées...)

chemin fermé → surfaces (bâtiment,...)



3. **relation** : groupe de primitives avec rôles associés

Met en relation des nœuds, des chemins, et potentiellement d'autres relations pour former des entités complexes (ex : école, université...)



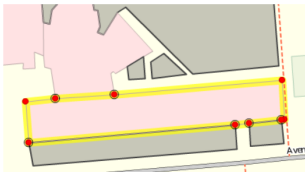
Chaque entité OSM (nœud, chemin, relation) possède :

- ▶ un **identifiant numérique** : OSM ID
- ▶ un ensemble d'attributs génériques :
 - ▶ **uid, user** : identifiant et nom contributeur
 - ▶ **timestamp** : heure de la dernière modification
 - ▶ **visible** : si cet attribut est faux, l'objet n'est pas affiché sur la carte
 - ▶ **version** : nombre de modifications de l'entité
 - ▶ **changeset** : le dernier groupe de modifications au cours duquel l'objet a été créé ou mis à jour
- ▶ un ensemble d'**attributs**, ou **tags** (paires clefs-valeurs, e.g. `highway=residential`) : utilisation non restreintes, mais gouvernée par un processus communautaire agile → définit l'**ontologie** de base d'OSM.



Way: Ensimag - D (23000671)

[Download XML](#) | [View history](#) | [Edit way](#)



Edited at: Sun, 10 Mar 2013 09:00:14 +0000

Edited by: [Liberal](#)

Version: 11

In changeset: [15311804](#)

Comment: added details

Tags:

- amenity = university
- building = yes
- contact:website = <http://ensimag.grenoble-inp.fr/>
- name = Ensimag - D

Nodes:

- 344548301 (part of way [Ensimag \(200061448\)](#))
- 247961085
- 247961086
- 2148742026 (part of way [204894090](#))
- 2148742020 (part of way [204894090](#))
- 2148742029 (part of way [204894089](#))
- 247961087 (part of ways [204894089](#) and [204894089](#))
- 247961132
- 344548275 (part of way [Ensimag \(200061448\)](#))
- 344548301 (part of way [Ensimag \(200061448\)](#))

Part of: Relation [Ensimag \(2776018\)](#)



Relation: Ensimag (2776018)

[Download XML](#) | [View history](#)



Edited at: Thu, 21 Feb 2013 15:05:04 +0000

Edited by: [Theneurasthenicrat](#)

Version: 1

In changeset: [15113270](#)

Comment: Local modification for Ensimag Buildings, St-Martin-d'Hères, France, and add of a relation for these buildings to characterize Ensimag.

Tags:

- `amenity = university`
- `name = Ensimag`
- `type = site`
- `uri = http://ensimag.grenoble-inp.fr/`
- `wikipedia = fr:Ensimag`

Members:

- [Way Ensimag - H \(31391089\)](#)
- [Way Ensimag - E \(200061446\)](#)
- [Way Ensimag \(200061448\)](#)
- [Way Ensimag - D \(23000671\)](#)



Introduction

Préliminaires techniques

Le modèle de données

L'ontologie OSM

OSM à l'Ensimag

Contexte pédagogique

Infrastructure technique

Contenu détaillé du TP

Retours sur le TP

OSM et (big) dataviz

Défi technique

Sous le capot

Import OSM

Suite

OSM, les attributs et le web sémantique





- ▶ Dans OSM, des paires clefs-valeurs sont utilisées pour décrire les éléments (nœuds, chemins et relations)
- ▶ Attributs classifiants (exemple : `type = 'rivière'`)
- ▶ Autres attributs (exemple : `nom = 'Isère'`)

					
<ul style="list-style-type: none">• <code>highway=motorway</code>• <code>lanes=4</code>• <code>surface=asphalt</code>		<ul style="list-style-type: none">• <code>highway=secondary</code>• <code>surface=asphalt</code>• <code>lanes=2</code>		<ul style="list-style-type: none">• <code>highway=track</code>• <code>name=<track_name></code> (if applicable)• <code>tracktype=grade3</code>• <code>surface=sand</code>	

http://wiki.openstreetmap.org/wiki/Tagging_samples/out_of_town

- ▶ Les clefs et les valeurs sont des champs texte libre
- ▶ Un ensemble de conventions permettent de s'accorder sur les bonnes manières d'étiqueter les éléments les plus communs.



- ▶ des attributs pour les objets du monde réel
 - ▶ `building = church, hotel, school, university...`
 - ▶ `highway = motorway, primary, secondary...`
 - ▶ ...
- ▶ des attributs pour les objets immatériels
 - ▶ `boundary = administrative, national_park...`
 - ▶ ...
- ▶ des attributs communs (non classifiants)
 - ▶ `name = *`
 - ▶ `source = *`
- ▶ Quelques ressources :
 - ▶ **Map Features** (http://wiki.openstreetmap.org/wiki/Map_Features)
 - ▶ **Taginfo** (<http://taginfo.openstreetmap.org/>)



Way: Ensimag - D (23000671)

- amenity = university
- building = yes
- contact:website = <http://ensimag.grenoble-inp.fr/>
- name = Ensimag - D

Way: Ensimag - H (31391089)

- amenity = university
- building = yes
- layer = 1
- name = Ensimag - H
- source = cadastre-dgi-fr source

Relation: Ensimag (2776018)

- amenity = university
- name = Ensimag
- type = site
- url = <http://ensimag.grenoble-inp.fr/>
- wikipedia = fr:Ensimag



Introduction

Préliminaires techniques

Le modèle de données

L'ontologie OSM

OSM à l'Ensimag

Contexte pédagogique

Infrastructure technique

Contenu détaillé du TP

Retours sur le TP

OSM et (big) dataviz

Défi technique

Sous le capot

Import OSM

Suite

OSM, les attributs et le web sémantique

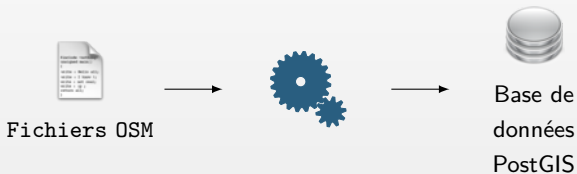


Deux formations :

- ▶ Ensimag 3^{ème} année (Bac+5) ingénieurs, filière Ingénierie des Systèmes d'Information : **Introduction aux Systèmes d'Information Géographique** (4h30 de cours + 3h++ de TP)
- ▶ Master spécialisé BigData (Ensimag / Grenoble École de Management) : **Introduction aux Systèmes d'Information Géographique** (3h de cours + 3h++ de TP) – étudiants d'origines diverses



- ▶ Objectifs du TP : interrogation et visualisation de données géographiques
- ▶ Infrastructure technologique : Un serveur de base de données Postgres / PostGIS, client Postgres, Java pour la partie carto (avec une bibliothèque maison).
- ▶ Deux parties :
 1. Interrogation directe de la base de données (client Postgres)
 2. On embarque tout ça dans une application Java pour visualiser
- ▶ Données : on travaille avec la région Rhône-Alpes (données fin 2013) ~~~> environ 12Go





Plusieurs outils et schémas possibles...

▶ **Osm2pgsql**, schéma osm2pgsql

- ▶ ✓ géométries complexes (ex : régions, pays) reconstruites algorithmiquement
- ▶ ✓ expose explicitement les entités les plus courantes (routes, attributs...) ~> efficace pour des requêtes spécifiques.
- ▶ ✗ schéma peu clair (mixe des modes de représentations différents)
- ▶ ✗ relations clairsemées

▶ **Osmosis**, schéma pgsnapshot

- ▶ ✓ schéma clair
- ▶ ✓ plus proche du modèle OSM
- ▶ ✓ géométries PostGIS pour les données géo, et colonnes HStore pour les paires clefs-valeurs
- ▶ ✗ données non prétraitées ~> embêtant pour les limites administratives
- ▶ ✗ tous les attributs dans la même colonne ~> peut être très lent



Quelques requêtes de base...

Q.1 Nombre de contributeurs pour la région Rhône-Alpes ?

Q.2 Coordonnées géographiques de la mairie de Grenoble ?

Q.3 Quels sont tous les types de routes que l'on trouve dans la base, et combien de routes y a-t-il par type ? Classez par ordre décroissant sur le nombre de routes.



Des fonctions de mesure...

Q.4 Quels sont tous les types de routes que l'on trouve dans la base, et combien de kilomètres de routes y a-t-il par type ? Classez par ordre décroissant sur la longueur de routes.

Q.5 Quelle est l'aire totale de l'Ensimag en m^2 ?

Analyse spatiale (croisement avec une base IGN)

Q.6 Ensemble des quartiers de Grenoble, avec, pour chaque quartier, le nombre d'écoles ("amenity"=>"school") qu'il contient, le tout ordonné par nombre d'écoles décroissant.



Requêtes hard-core...

Q.7 Quelle municipalité (ville, village) constitue le centre géographique de la région ?

Q.8 Déserts de population. Vous voulez partir en vacances tranquillement. Existe-t-il un endroit en région Rhône-Alpes éloigné de plus de 10 kilomètres d'un bâtiment ?



D'autres idées de requêtes ?



Objectif de cette partie : embarquer des requêtes dans une application Java pour construire des cartes.

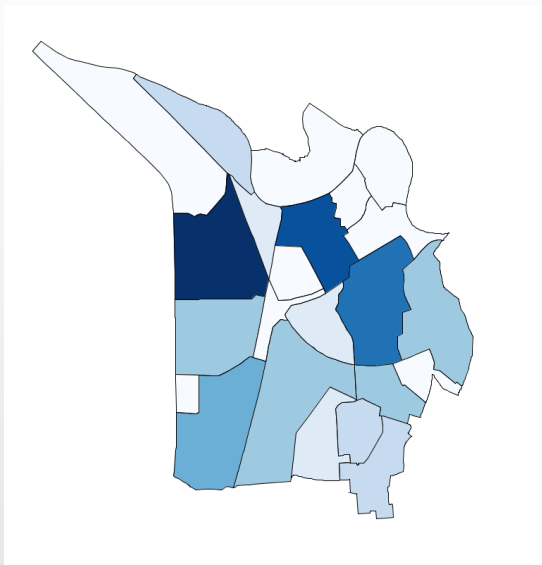
- ▶ Première question : afficher l'ensemble des bâtiments et routes de Grenoble.
- ▶ Suite : Libre

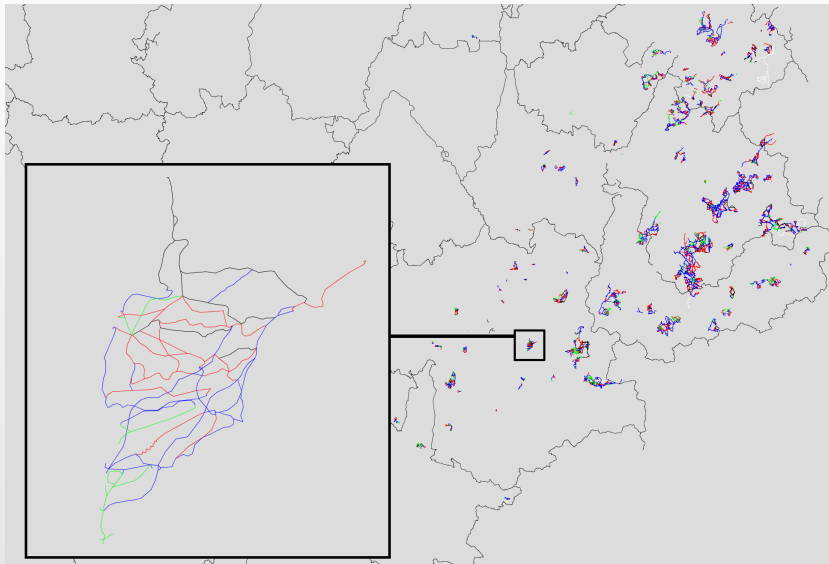


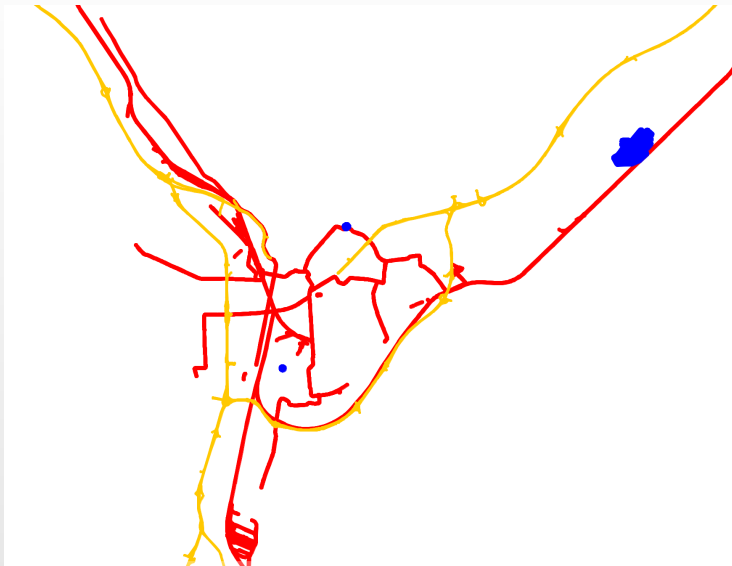
ÉCHELLE D'ÉCRANS EN QUARTIER
BERRIAT ST BRUNO : 12
CENTRE VILLE : 32
ESPIONNAGE-BASTARDÈRE : 8
RONDEAU-LIBÉRATION : 7
BARDY CLAUDE : 6
MONTMARTRE : 11
ALPES-ALGERIA : 5
ABEYLI-JOURNAUX : 5
VALLENDREUX : 4
JEAN MARC : 4
VALLENDREUX2 : 3
CATHOLIC GR : 3
CENTRE CARRÉ : 2
TERREIRE : 1
SAINT-JULIEN : 1
NOTRE-DAME : 1
MUTUALITE : 1
NESTRA-SIBIC : 1
LE VÉRIE : 1
POUR ANGLE : 1
CHAMPAGNET : 1

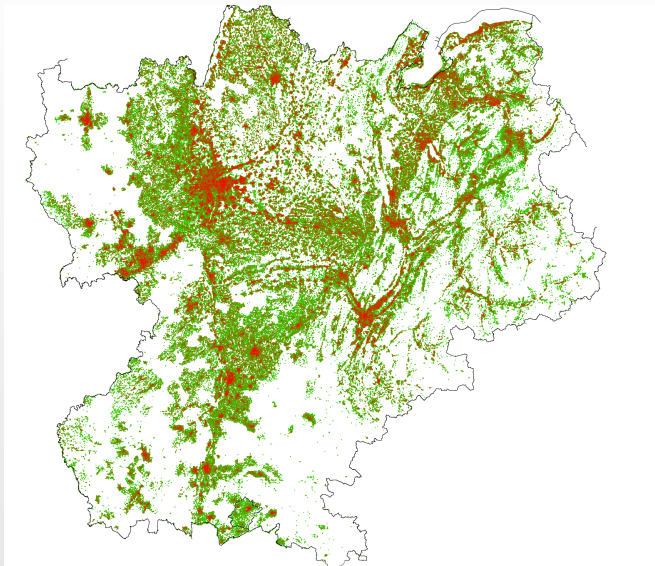


Coordonnées curseur : (5.713080182668527, 45.769146167059263)











- ▶ ✓ Apprentissage des notions fondamentales des SIG (requêtes spatiales, systèmes de référence, interrogation par attributs)
- ▶ ✓ Données réelles \rightsquigarrow intérêt des étudiants
- ▶ ✓ Liberté de choix du sujet (2^{ème} partie) \rightsquigarrow intérêt des étudiants
- ▶ ✓ Volume de données conséquents \rightsquigarrow réflexion sur l'efficacité des requêtes
- ▶ ✓ effet de bord : la VGI, les données libres
- ▶ ✗ Problèmes de performance



Introduction

Préliminaires techniques

Le modèle de données

L'ontologie OSM

OSM à l'Ensimag

Contexte pédagogique

Infrastructure technique

Contenu détaillé du TP

Retours sur le TP

OSM et (big) dataviz

Défi technique

Sous le capot

Import OSM

Suite

OSM, les attributs et le web sémantique



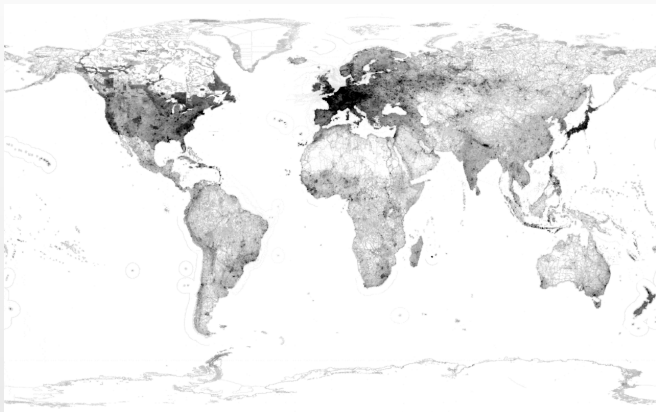
Dans la version de la base sur laquelle nous avons travaillé :

- ▶ 2,863,152,191 nœuds (planète complète) \rightsquigarrow 23.8 Go
- ▶ 144,036,346 nœuds (building) \rightsquigarrow 1.1 Go

Comment explorer de manière continue ce jeu de données de manière interactive ?

- ▶ Milliards de nœuds \rightsquigarrow ne tient pas en RAM
- ▶ Lecture exhaustive du fichier \rightsquigarrow plusieurs (dizaines de) secondes

Travail de recherche de Renaud Blanch (LIG, Univ. Grenoble-Alpes).





On renverse le point de vue :

- ▶ Chaque point du jeu de données \mapsto son pixel sur l'écran
 - ▶ 3 milliards++ de points à considérer pour les vues larges ✗
 - ▶ lecture exhaustive du fichier ✗
- ▶ Chaque pixel \mapsto nombre de points du jeu de données
 - ▶ 1 million de pixels à considérer (vue 1000×1000) ✓
 - ▶ OK si les opérations par pixel sont efficaces...



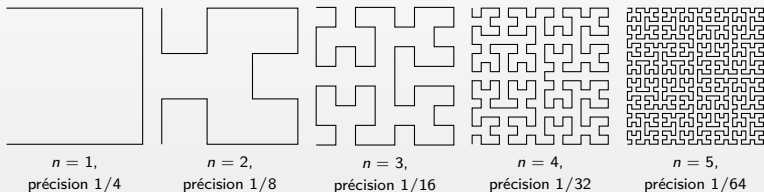
OK si les opérations par pixel sont efficaces...

Comment trouver le nombre de points sous chaque pixel sans lire tout le fichier ?

Idée : On ordonne les points de manière à ce que chaque pixel représente un **intervalle**, et on ne lit que les bornes de l'intervalle ($\Theta(\log(n))$) au lieu de $\Theta(n)$).



Pour passer d'une vue 2D à un ordre linéaire (1D) en **préservant la localité**, on utilise une fonction de hachage de **Hilbert** :





- ▶ Python/Cython (compilation du code Python en C avec typage statique)
- ▶ Numpy/Scipy (structures de données efficaces)
- ▶ OpenCL (calcul parallèle du hash de Hilbert)
- ▶ OpenMP (recherche dans le tableau trié)
- ▶ OpenGL (affichage)
- ▶ fichiers en mémoire pour les données
- ▶ Disques SSD



Chaîne de traitement des données :

planet.osm → buildings.dat, hashes de Hilbert ordonnés

- ▶ **1^{ère} étape** : téléchargement du jeu de données planet

```
wget -c ftp://ftp.splinel.de/pub/openstreetmap/pbf/planet-latest.osm.pbf
```

- ▶ **2^{ème} étape** traitement du jeu de données avec Osmosis

```
osmosis -tf accept-ways buildings=* -read-pbf-fast file=planet-latest.osm.pbf
```

Attention au stockage des fichiers temporaires (sature le disque).



Chaîne de traitement des données :

planet.osm → buildings.dat, hashes de Hilbert ordonnés

- ▶ **3^{ème} étape** : analyse du fichier SQL COPY. Script Python :
 - ▶ lecture ligne par ligne du fichier SQL
 - ▶ récupération de la géométrie de chaque bâtiment (shapely)
 - ▶ calcul du centroïde (shapely)
 - ▶ calcul du hash de hilbert
 - ▶ écriture ligne par ligne d'un fichier texte
- ▶ **4^{ème} étape** : tri du fichier haché avec la commande sort Unix
- ▶ **5^{ème} étape** : Encodage binaire du fichier (utilisation memmap numpy)

```
>>> data = np.memmap('planet.dat', mode='w+', dtype=np.float64, shape=(10,))
```

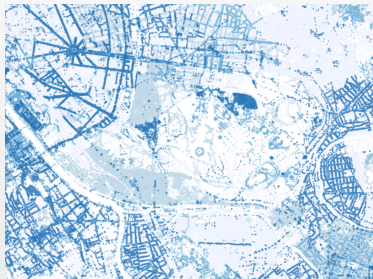
Temps de traitement, environ 1 semaine...



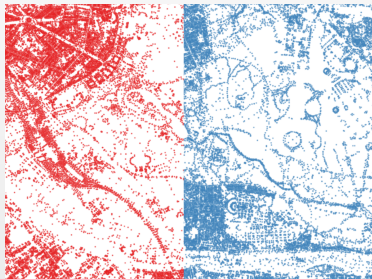
- ▶ Architecture client / serveur pour meilleure intégration
- ▶ Espace + temps
- ▶ Attributs
- ▶ Affichage multi-couches



Évolution des données OSM (osmium)



Evolution nœuds sur Paris (2012 → 2015). Tom Cornebize (Ensimag).



Evolution nœuds sur Paris (2012 vs 2015). Tom Cornebize (Ensimag).



Introduction

Préliminaires techniques

Le modèle de données

L'ontologie OSM

OSM à l'Ensimag

Contexte pédagogique

Infrastructure technique

Contenu détaillé du TP

Retours sur le TP

OSM et (big) dataviz

Défi technique

Sous le capot

Import OSM

Suite

OSM, les attributs et le web sémantique

Présentation par Anthony Hombiat, LIG STeamer.