



Hackaton #CodeImpot

## Equipe Performance

2 avril 2016

# Composition de l'équipe



 DIRECTION GÉNÉRALE DES  
FINANCES PUBLIQUES

- Olivier Blanc



 gouv.fr

- Michel Blancard



 Polyconseil

- Henri Jouhaud



 Polyconseil

- Pierre Marseille



 Polyconseil

- Alexandre Lopes Meda



 Polyconseil

- Ulysse Pineau



 Polyconseil

- Lorenzo Randé



 manymore

- Stéphane Rathgeber

# Objectif x 1 000

Ou comment diminuer le temps  
de calcul de l'impôt  
de plus d'**1 mois** à **X minutes**

*Temps de calcul de l'impôt extrapolé à partir de « 1h/100 000 foyers en législation 2013, 1h45/100 000 foyers en législation 2014 »  
([http://jms.insee.fr/files/documents/2015/S20\\_3\\_ACTE\\_V2\\_AEBERHARDT\\_JMS2015.PDF](http://jms.insee.fr/files/documents/2015/S20_3_ACTE_V2_AEBERHARDT_JMS2015.PDF))*

## Jeux de données

**Vectorisation**

**Calcul distribué**

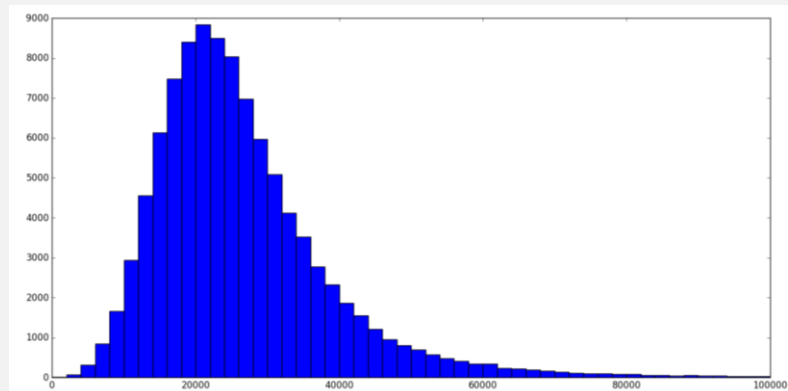
**Optimisation du graph**

## Sélections de paramètres (>10% population)

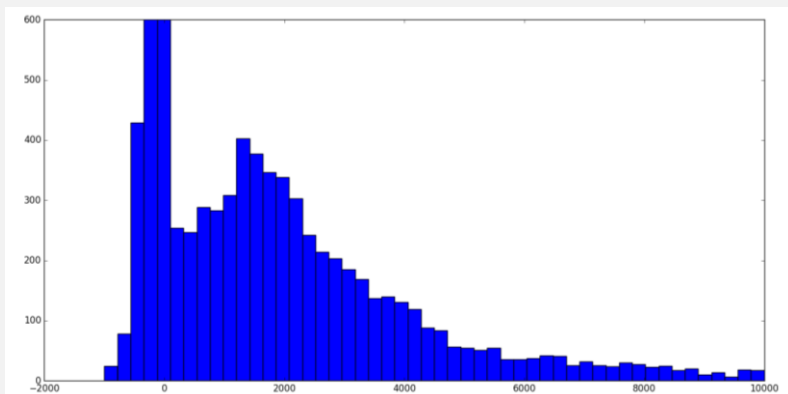
- Situation de famille : marié, célibataire, veuf, divorcé, pacsé
- Enfants mineurs à charge
- Enfants majeurs à charge
- Salaires, pensions de retraite, autres revenus
- Revenus d'actions
- Crédit d'impôt
- Revenus fonciers
- CSG déductible
- Dons aux associations

Génération aléatoire  
de foyers basé sur  
un modèle prenant  
en entrée les  
valeurs moyennes

## Exemple : distribution des salaires

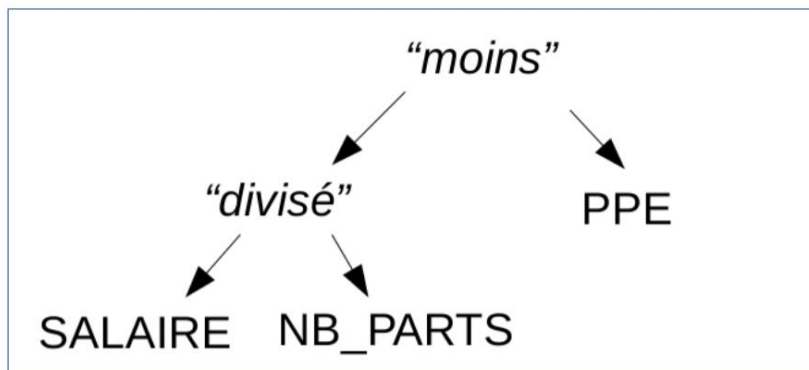


## Distribution de l'IR pour 10000 foyers générés



# Vectorisation

$$\text{IMPOT} = (\text{SALAIRE} / \text{NB\_PARTS}) - \text{PPE}$$



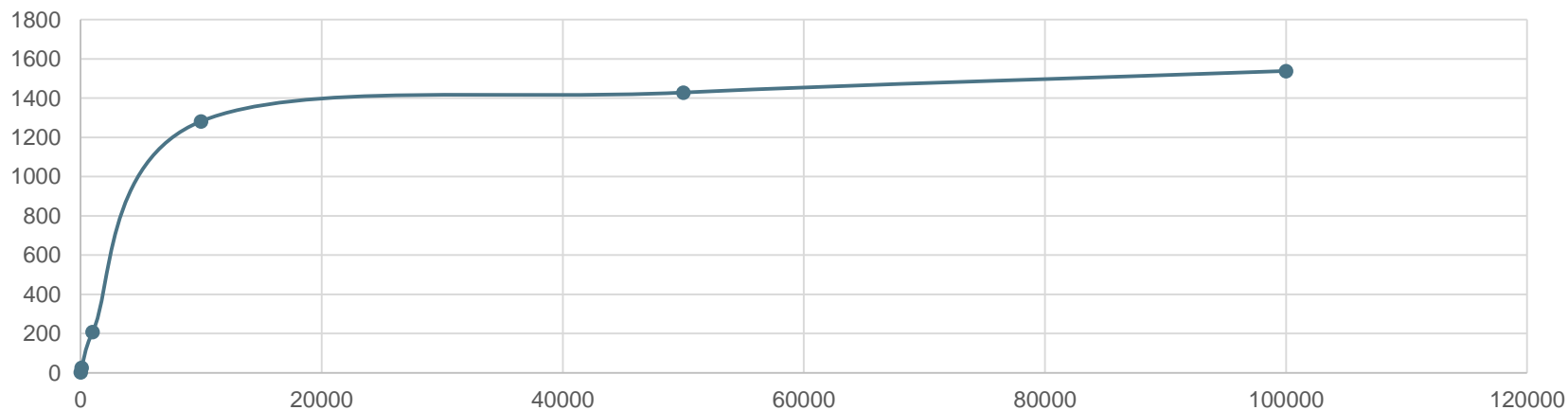
Temps de parcours de l'arbre : 400ms

Temps de calcul : 0.4ms

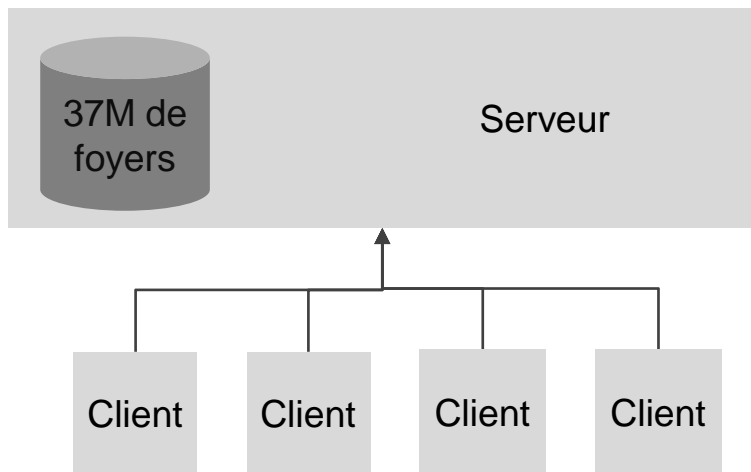
1 parcours + 1 calcul : **400,4 ms**

1 parcours + 10.000 calculs : **4,4 s**

Vitesse de calcul (foyers/seconde) en fonction du nombre de foyers dans le lot



# Calcul distribué

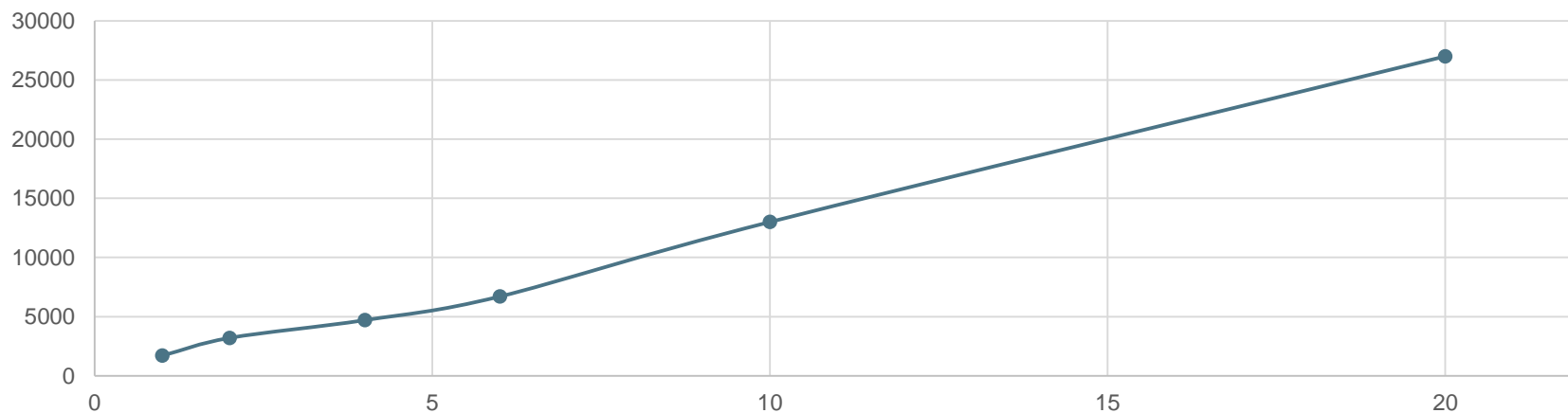


La vitesse de calcul agrégée au niveau serveur est limitée par :

- La vitesse de la base de donnée centrale
- La vitesse du réseau entre le serveur et les clients

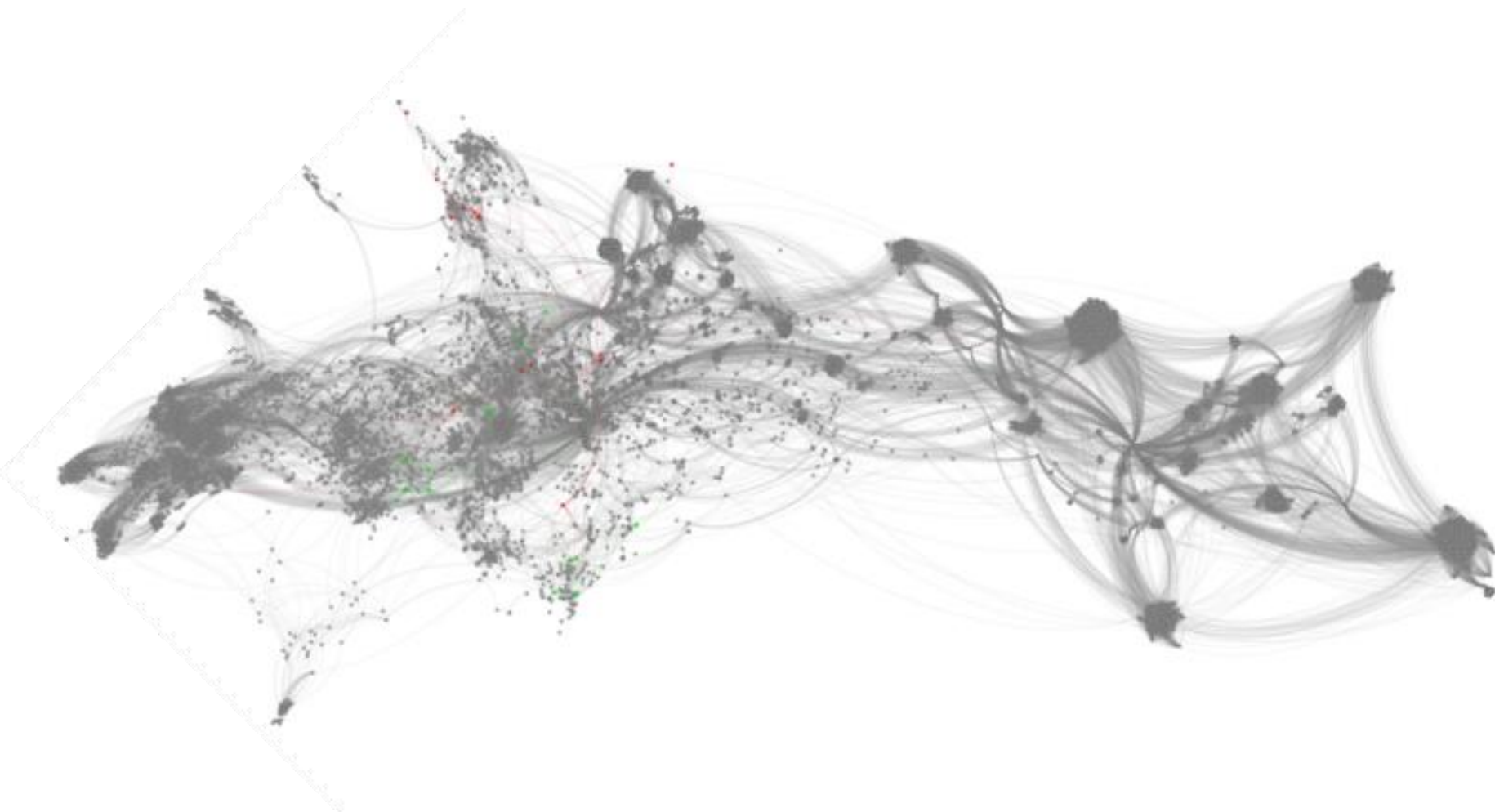
Chaque client peut calculer individuellement ~1500 foyers/seconde

Vitesse de calcul (foyers/seconde) en fonction du nombre de clients



Utilisation du cloud Amazon EC2 pour les tests. Serveur sur c3.4xlarge et clients sur t2.medium

# Optimisation du graph: avant





## Optimisation du graph: après



# Objectif x 1 000 dépassé !

Ou comment diminuer le temps  
de calcul de l'impôt  
de plus d'**1 mois à 23 minutes**

(Soit 20s pour 500 000 foyers)

*Temps de calcul de l'impôt extrapolé à partir de « 1h/100 000 foyers en législation 2013, 1h45/100 000 foyers en législation 2014 »  
([http://jms.insee.fr/files/documents/2015/S20\\_3\\_ACTE\\_V2\\_AEBERHARDT\\_JMS2015.PDF](http://jms.insee.fr/files/documents/2015/S20_3_ACTE_V2_AEBERHARDT_JMS2015.PDF))*