



BIGDATA ET DONNÉES SEO

Vincent Heuschling
@vhe74

AGENDA

- ▶ SEO ?
- ▶ TRAITER et ANALYSER
- ▶ MESURER
- ▶ OUTILS
- ▶ EVOLUTION DES MOYENS



SEO ?

Optimiser le trafic d'un site web issu des moteurs de recherche.

► Définir quels éléments du site sont à forte valeur et les promouvoir vers les moteurs de recherches

► Mesurer et étudier le positionnement du site sur des recherches vis à vis de sa concurrence

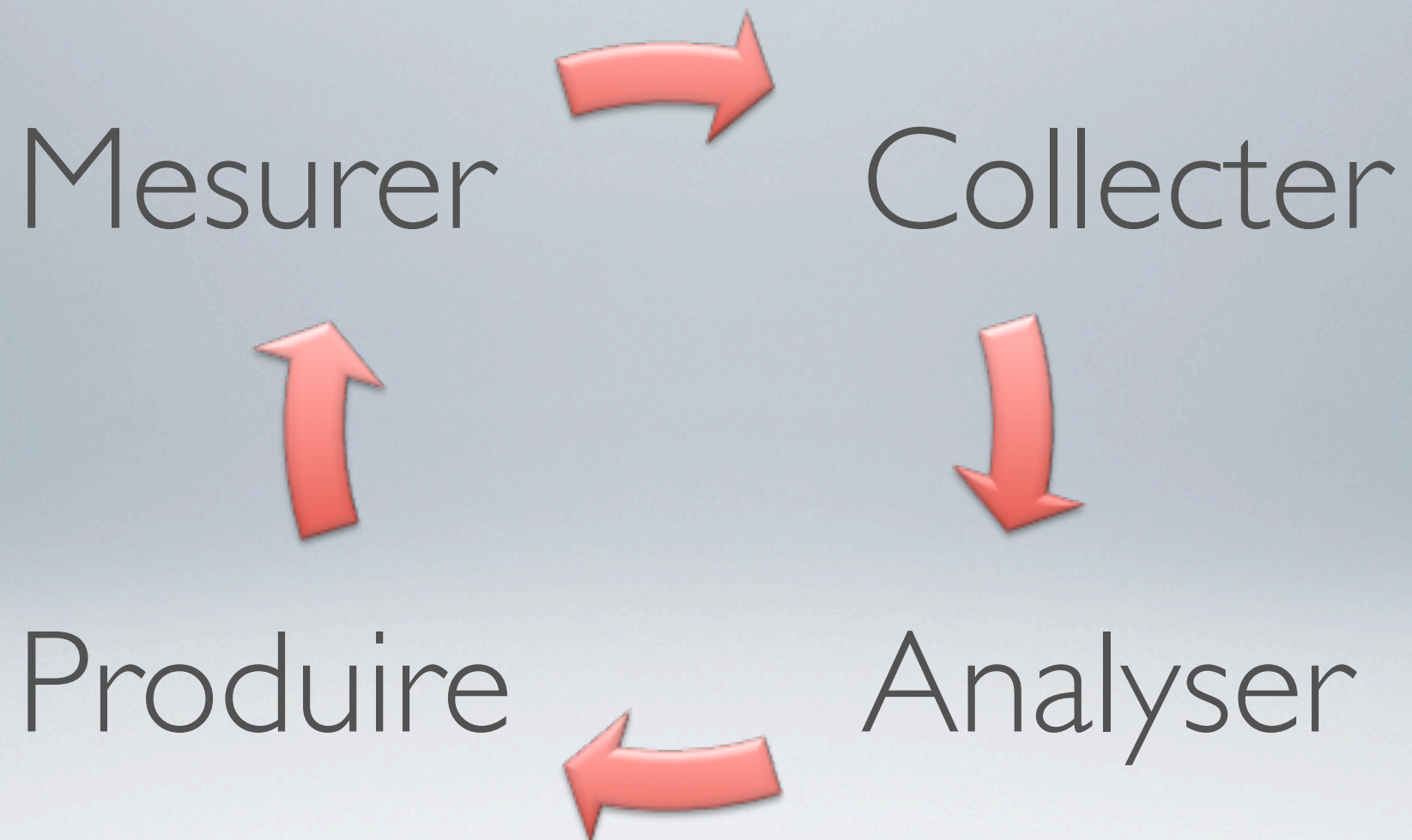


ENJEUX

- ▶ 100 millions de requêtes par mois
- ▶ Patrimoine informationnel très riche
- ▶ Business model basé sur la pub
- ▶ Marché ultra concurrentiel
- ▶ Plus de 15 Go de log par jour (sous-traités)



DÉMARCHE SEO



TROUVER 400K NOUVELLES URLS À PROMOUVOIR PARMIS 84M ?

- ▶ Similarités et Classifications
- ▶ Recommandation & intelligence collective
- ▶ OpenData



OUTILS

Collector

Traiter

Analyser



Pig



MESURER : RANKINGS

- ▶ Collecte des réponses Google (30x par recherche = dizaines de millions par mois)
- ▶ Forte croissance de la volumétrie
- ▶ Classifier et Segmenter par produit, par thématiques.



OUTILS

Collector



Traiter



Pig



Stocker



Visualiser

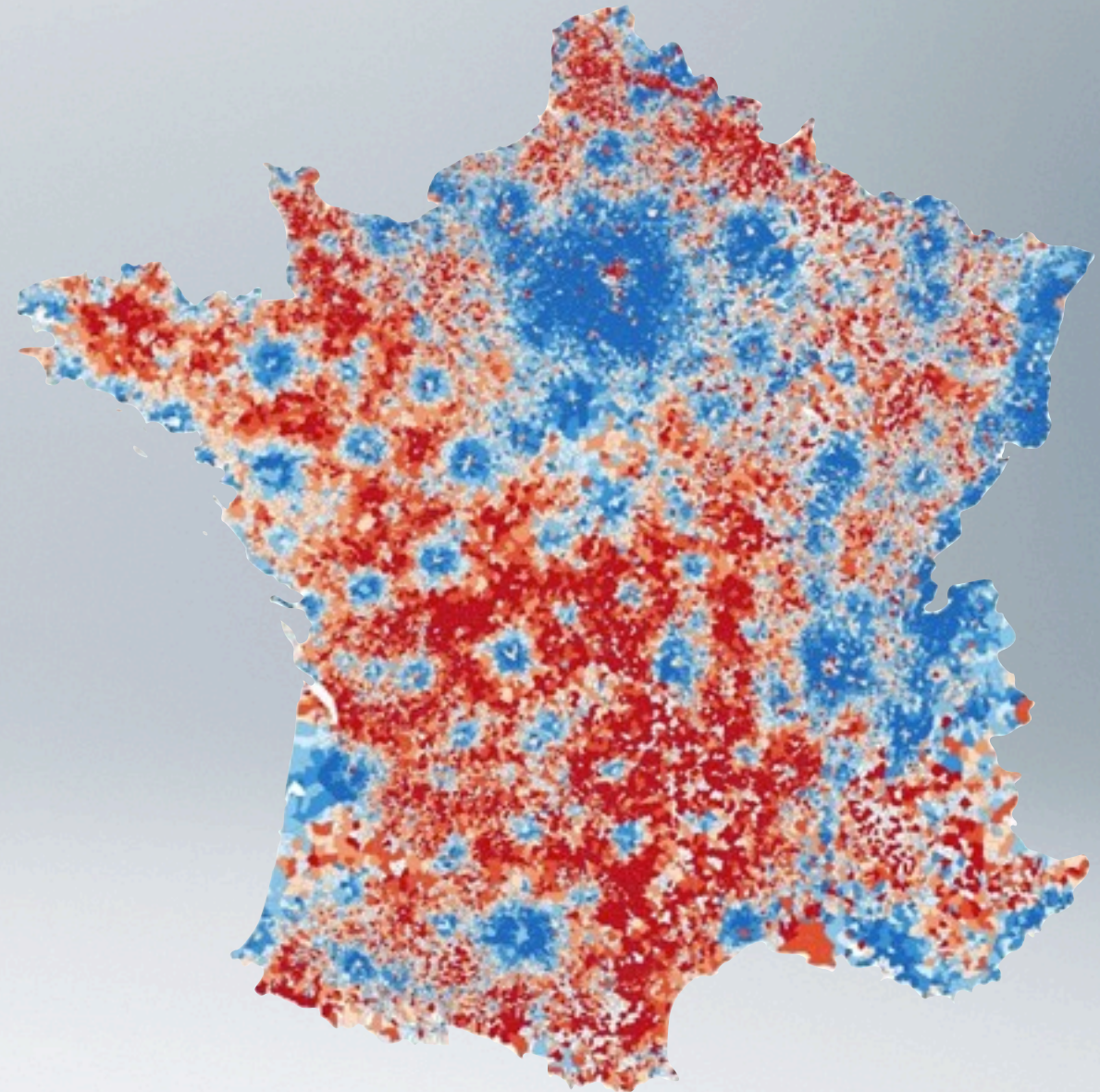
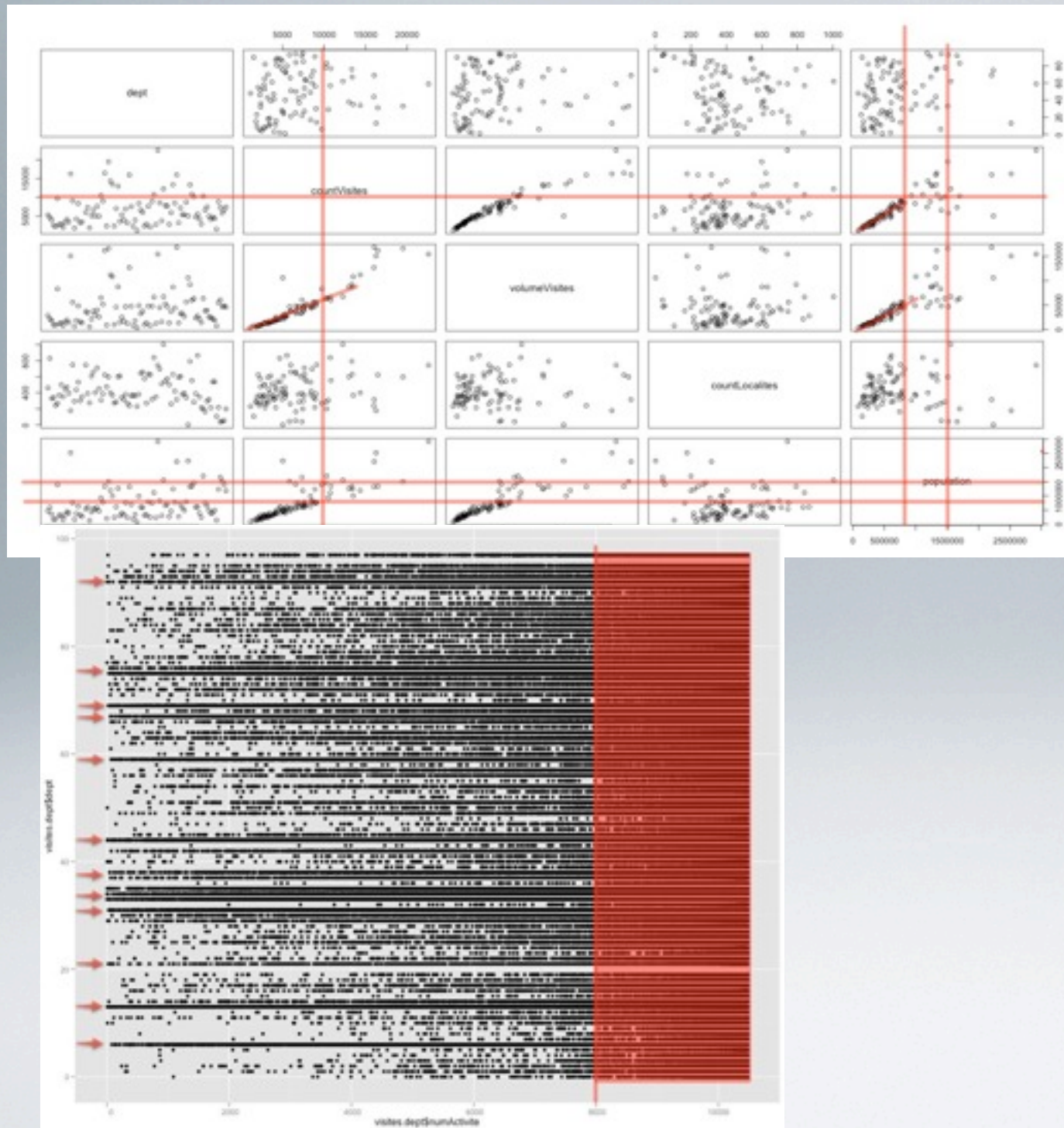


POURQUOI HIVE OU PIG

- ▶ Productivité :
Hadoop sans coder du Map/Reduce
- ▶ Traitement de Flux pour PIG
- ▶ HIVE permet de faire du SQL sur Hadoop

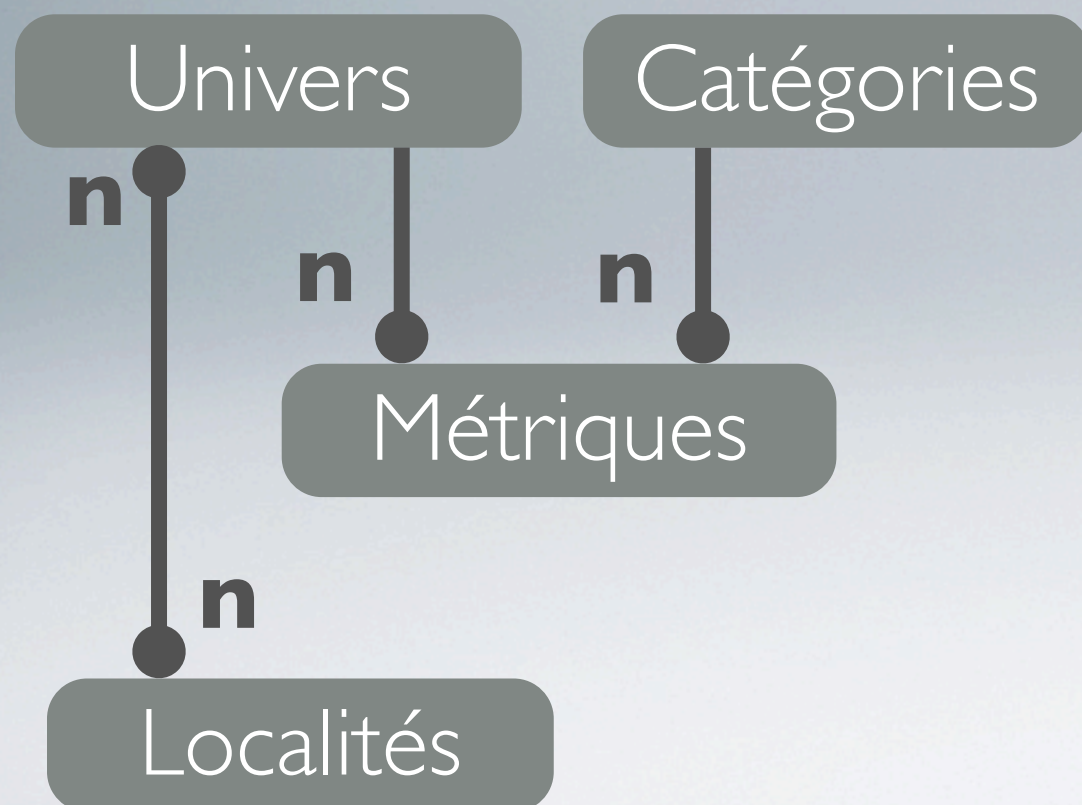


R



MONGODB

- ▶ Base document
- ▶ Stocker l'information avec son contexte
- ▶ Données agrégées et leurs détails



```
{ univers : "restaurants",  
  métrique : 250,  
  regions : ["idf", "paca"],  
  sousMétriques : [  
    { petitesVilles : 80 },  
    { moyennesVilles : 70 },  
    { grandesVilles : 100 },  
  ],  
}
```



EVOLUTION DES MOYENS

- ▶ Scalabilité des outils de traitement de logs
- ▶ Variation des besoins
- ▶ Migration mySQL vers Hadoop/Hive : SQL
- ▶ Plateforme managée : "Simple as a Query"
- ▶ Avoir une boîte à outils de Data-Analyse
- ▶ Dans un cloud mais local (FR)
- ▶ Ouverture en Beta privée à partir du 15 nov

DATAZOOMR



MERCI !

Vincent Heuschling

Gsm : 06 61 88 76 71

Email : vhe@affini-tech.com

Web : <http://www.affini-tech.com>

Twitter : @affinitech & @vhe74

