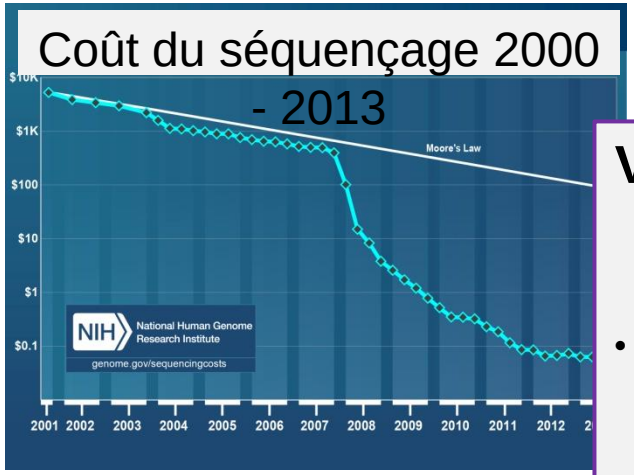
SOGETI
High Tech

Le Big Data comme brique d'assistance aux choix thérapeutiques en oncologie

Teratec

2 Juillet 2014

La volumétrie de la génomique clinique en quelque chiffres…



Volumétrie considérée – Big Data

- Volume de données issues d'un séquençage d'exome **5 Go**
- 3 exomes / patient **15 Go**
- 20% des patients et 50 % des types de cancers (Poumon, Prostate, sein, colon), soit 40,000 patients (France) **630 To**
- Europe Occidentale **3.7 Petaoctets**

Molécules en développement thérapies anti-cancéreuses

- Molécules ave AMM **10**
- Molécules en phase de développement clinique **100**
- Molécules en phase pré-cilinique **1000**

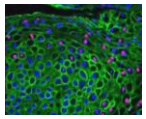
Le Big Data au cœur des enjeux de la médecine personnalisée

Volume / Vitesse

Variétés / Sémantique

Patient

Séquençage



Informations relatives aux molécules

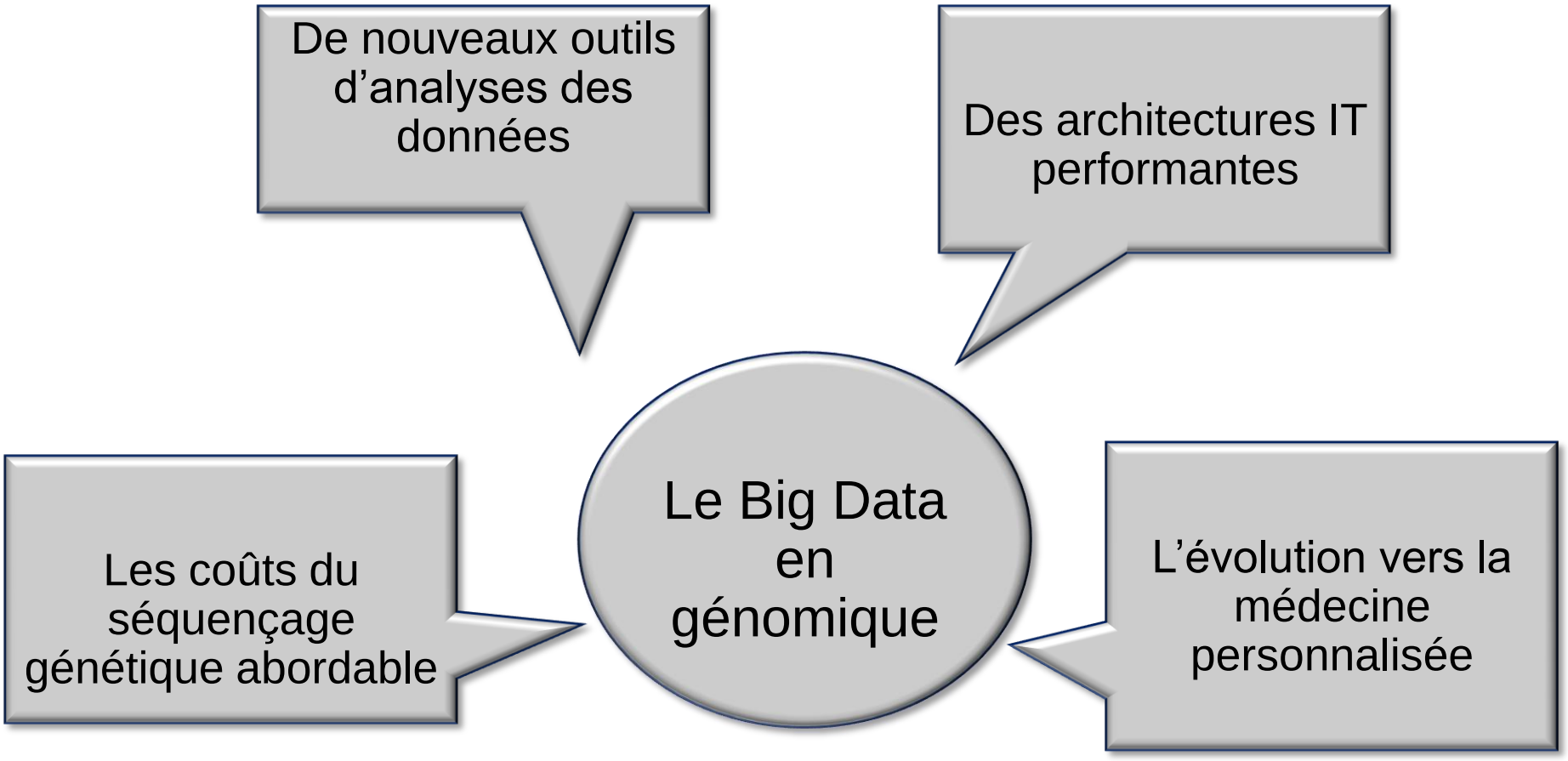
Publications scientifiques

Bio-informatique

Cancérologue

ICE : Outil d'assistance au choix de la thérapie optimale

Pourquoi le Big Data dans l'analyse génomique ?
Une convergence des technologies et de l'évolution des sciences



Le séquençage génomique : Algorithmique et Capacité de calcul

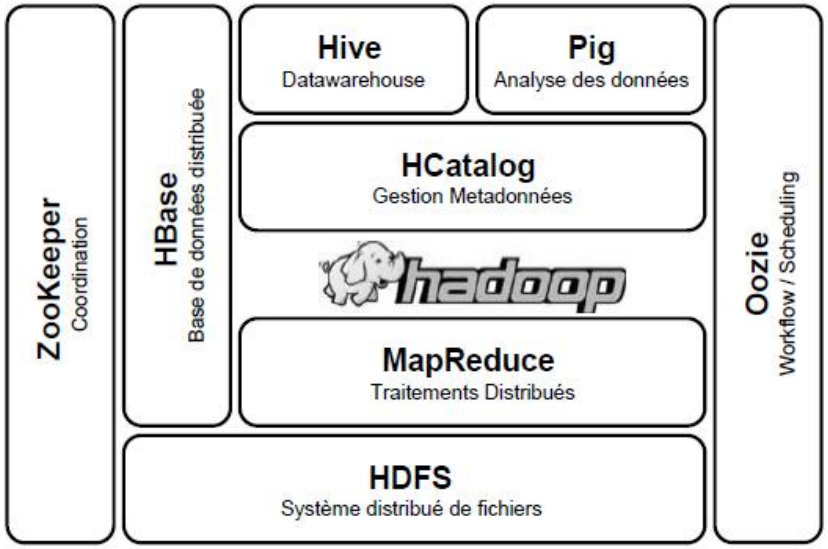
Algorithme en analyse génomique :
- *Algorithme complexe et couteux*
- *Besoin d’analyses à grande échelle*

Les performances des microprocesseurs stagnent :
- *Fréquence des processeurs bloquée*
- *Apparition des multicoeurs*
- *Les performances ne suivent pas l’évolution des masses de données génomiques*

Parallélisme massif
- *La seule solution pour réduire significativement les temps de calcul*

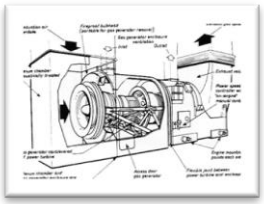
3 jours ➔ 3 heures de traitement

Parallélisation des algorithmes
Infrastructure Hadoop



L'interprétation des données

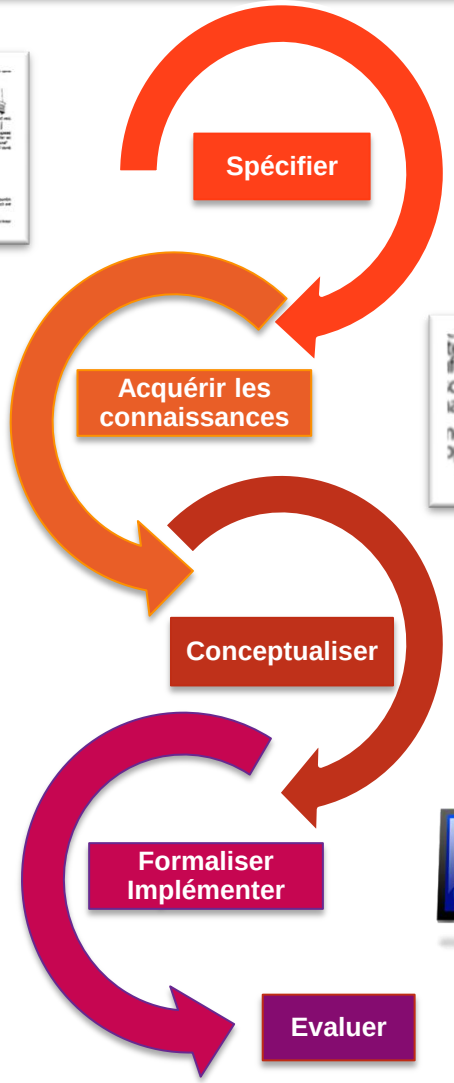
Une approche ontologique pour l'exploitation des données cliniques et scientifiques



Big data mining sur les publications scientifiques : Données hétérogènes, codifications différentes, volumes

Une anomalie génétique a-t-elle déjà été traitée et comment ?

- *Définition d'ontologie*
- *Analyse sémantique*
- *Langage naturel*
- *Auto-apprentissage*



Les bénéfices Big Data dans le projet ICE

- Capacité de stockage des données génomiques ➔ Petaoctets
- Capacité de calcul pour les comparaisons et les annotations ➔ Algorithmes d'analyse
- Capacité de Big Data mining sur des documents scientifiques

➔ Temps de traitement réduit à 3 heures