

Entrepôt de données de Santé (EDS)

Développement des fonctionnalités d'analyse des données

Séminaire EDS – 13 septembre 2016

Auteur :

Nicolas PARIS – DSI Département WIND

Sommaire

L'analyse des données de l'EDS

1. Enjeux
2. Défis
3. Etat d'avancement

1

Enjeux

L'EDS : Une vue 360° des patients

■ Intégrer les données cliniques collectées lors des soins à l'AP-HP avec des données d'autres sources

▶ Phénomène

- *Données PMSI (diagnostics, actes)*
- *Résultats de biologie (1 Milliard)*
- *Compte Rendu médicaux (30 Millions)*
- *Imagerie (1 Milliard d'images)*
- *Objets connectés (Millions de valeurs / secondes)*

▶ Génome

- *VCF (1 GO / fichier)*
- *CR de génétique*

▶ Exposome

- *Climat, pollution (chaque instant, partout)*

■ Mieux comprendre l'état de santé des patients : « Phénomène = Génome x Exposome »

L'analyse des megadonnées: une opportunité

■ La recherche, épidémiologie, veille sanitaire

- ▶ Développement d'algorithmes, data-mining

■ Le soin dans le cadre d'une médecine 4P (personnalisée, prédictive, préventive et participative)

- ▶ Visualisation des données patients
- ▶ Aide décisionnelle
- ▶ Aide au codage PMSI

■ Le pilotage opérationnel

- ▶ Tableaux de bords pour les équipes soignantes

■ L'enseignement

- ▶ Recherche de cas similaires, jeux sérieux à partir de données

Les megadonnées posent des défis de taille

2

Défis

Défi technique 1

Comment accumuler ces mégadonnées?

- **Volume** : capacité de stockage, sauvegarde, bande passante
 - **Vélocité** : capacité d'enregistrement et d'interrogation
 - **Variété** : capacité à harmoniser selon de référentiels AP-HP, nationaux et internationaux
- ⇒ **Une plateforme megadonnées basée sur des grappes de machines pour une répartition de la charge**

Défi technique 2

Comment exploiter ces megadonnées ?

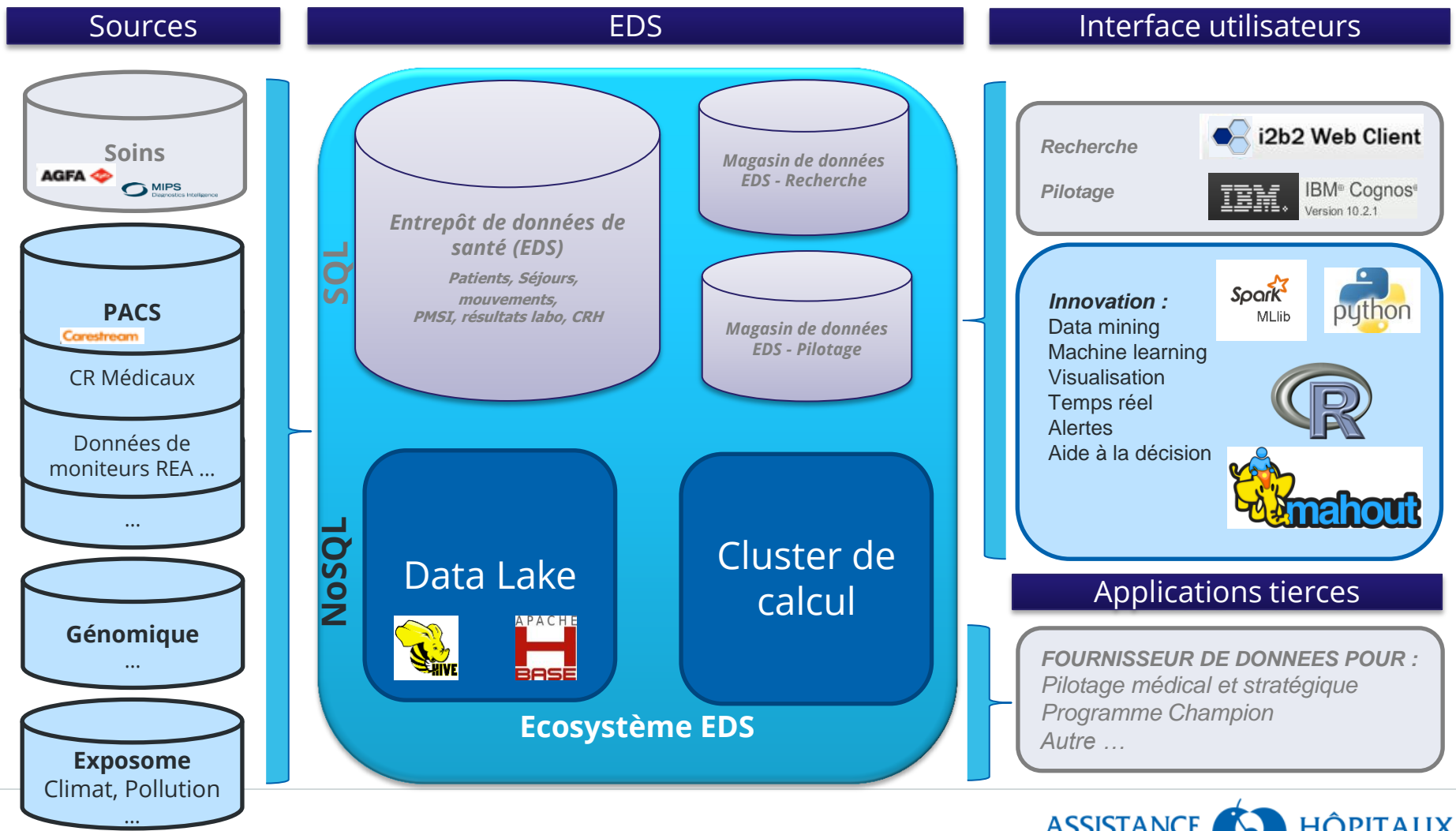
- Télécharger les données pose un problème d'espace
- Analyser les données pose un problème de temps
- Analyser la donnée sur place
 - ▶ Mutualisation de la puissance de calcul
 - ▶ Sans pour autant limiter les outils d'analyse (R, python...)

⇒ **Une plateforme megadonnée distribuée avec une diversité d'outils adaptés aux chercheurs**

3

Etat d'avancement

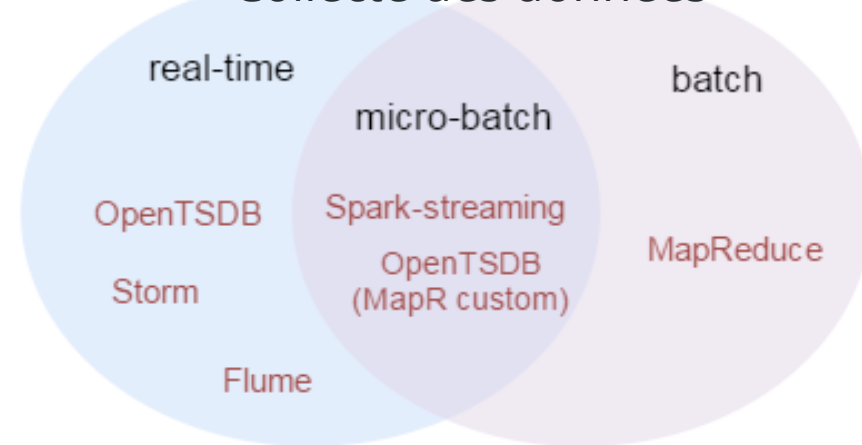
Architecture globale



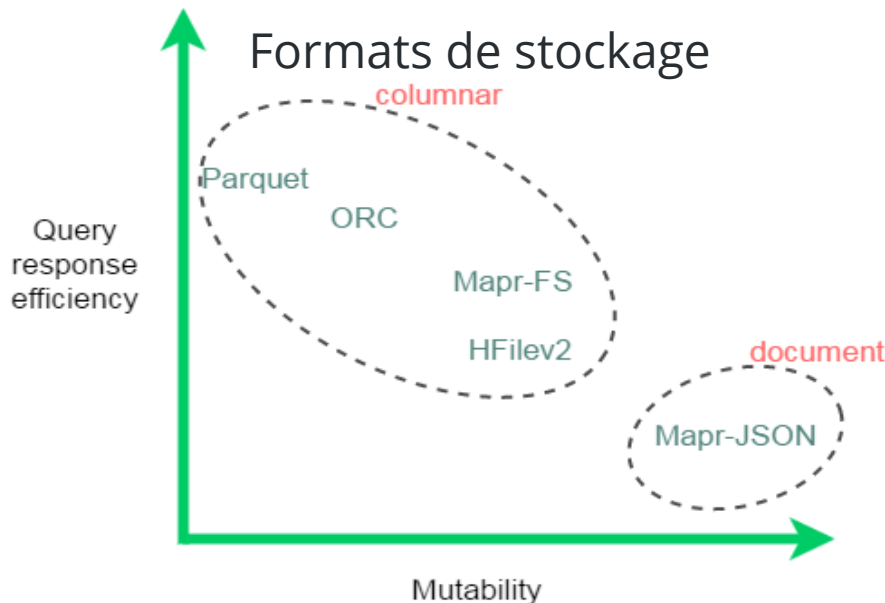
Phase 1 (2016) – Prototypage

- ▶ Evaluation comparative des plateformes
- ▶ Cas d'usage en réanimation (données physiologiques)

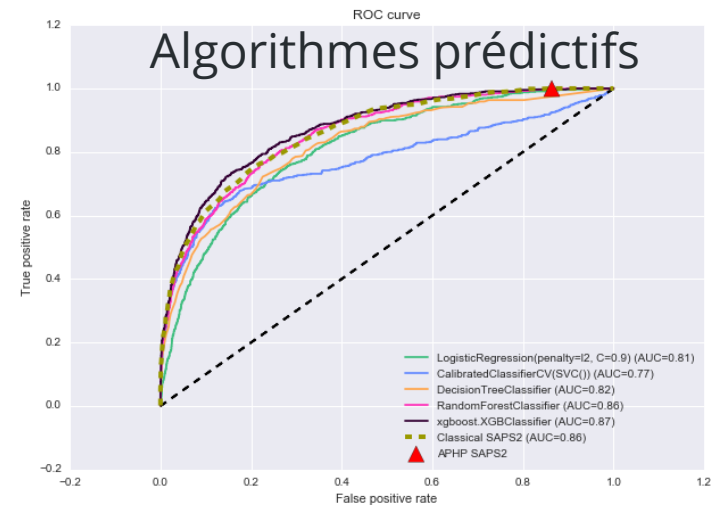
Collecte des données



Formats de stockage



Algorithmes prédictifs



Phase 2 (2017) – Industrialisation

■ Nouvelles données

- ▶ Images du PACS
- ▶ Génomique en lien avec la plateforme de bioinformatique
- ▶ Intégration de données climatiques/pollution

■ Nouveaux cas d'usage (projets soumis au CSE)

- ▶ Valeur prédictive de fracture d'un score de densitométrie osseuse calculé sur des scanners abdominaux
- ▶ Liens entre pollution et constantes respiratoires des patients hospitalisés en réanimation

■ Organiser l'accès à la plateforme megadonnées

- ▶ Sécurité, cloisonnement
- ▶ Anonymisation

Merci de votre attention

Contacts

nicolas.paris@aphp.fr

[SAP] Pacte Wind Siège AP-HP <pacte.wind@aphp.fr>