

Gestion et archivage des big data : étude de cas des données climatiques en Tunisie

Nozha Hizi, Abderrazak Mkadmi

Institut supérieur de documentation; Université de la Manouba

CIDE 21

04-06 Avril 2019

Plan

- ▶ *Contexte et objectifs*
- ▶ *Problématique liée à la gestion et archivage des données climatiques*
- ▶ *Méthodologie de recherche*
- ▶ *Enquête et résultats*
- ▶ *Conclusion et perspectives*

Contexte et objectifs

Entreprises : produisent une masse énorme des données,
gigaoctets ou péta octets;

Apparition du **big data**;

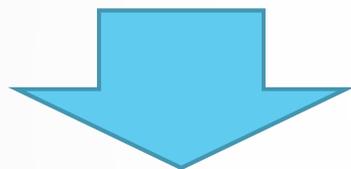
Outils classiques incapables de traiter et stocker l'ensemble des
données;

Besoin d'une nouvelle méthode pour gérer et archiver des mégadonnées ?



Les données climatiques et big data :

- Quantité (données collectées par jour en temps réel) ;
- variété des données (structurées et non structurées) ;
- variété des sources et des outils utilisés pour capter des données ;



Découvrir la méthode de gestion et d'archivage des données climatiques en
 prise à l'ère du big data;

Proposition d'un modèle d'archivage des mégadonnées climatiques;

Problématique

Jusqu'à quel point on pourrait gérer des données climatiques, qualifiées de grosses données, par les moyens de gestion de documents classiques?

Questions spécifiques:

Comment collecter, traiter et analyser les données climatiques à l'ère du big data?

Comment archiver, protéger et pérenniser les données climatiques massives?



Méthodologie de recherche

- ▶ Recherche basée sur une étude qualitative;
 - **Immerger** dans les discours des personnes interrogées et **d'obtenir** une meilleure compréhension sur la manière d'archivage des données climatiques à l'ère du big data;
- ▶ 10 entretiens semi-directifs ;
 - Méthode très utile dans les **études qualitatives** ;
 - collecter des **informations précises** et orienter le discours autour de différents points définis;
 - Des **informations sur les comportements, les attitudes et les représentations des employés**;
 - Comprendre la **manière de gestion et d'archivage** des données climatiques à l'INM ;

Méthodologie d'enquête

- Personnes expertes en météorologie, sismologie, astronomie et en informatique ,des spécialistes en informations;
- Population interrogée : significative ;

Objectifs :

- Informations sur les activités professionnelles, méthodes de gestion des données climatiques: collecte, traitement, archivage, diffusion, exploitation des données, enjeux stratégiques et technologiques, ...
- vérifier l'émergence du big data dans l'INM et proposer un modèle d'archivage des données climatiques.

Big Data ?

Big data ?

- ❑ Données collectées de multiples sources ; de différents types ;
- ❑ Données difficiles à manipuler, transformer, stocker et archiver ;

Mégadonnées

- ✓ Modèle des 3V : volume, variété, rapidité étendu au 6V véracité, valeur et visibilité ;
- ✓ Enjeux stratégiques : trier, traiter, organiser les flux informationnels ;
- ✓ Enjeux technologiques : adopter un nouveau modèle ;



Gestion et archivage des données massives

Enquête et Résultats (1)

Procédures,
techniques, méthodes
de traitement sont
traditionnelles

Big Data \equiv réalité
émergente ;
problématiques
majeures: nouvelles
technologies

Enquête et Résultats (2)

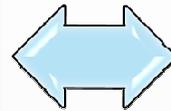
- 
- Repenser des nouvelles solutions :
- Formation des spécialistes ;
 - Méthodes d'archiver le patrimoine climatique ;

Proposition du modèle de gestion des grosses données climatiques :

Modèle de gestion et
d'archivage des données
climatiques



Dématérialiser les
documents et les
processus



Plateforme de gestion
et d'archivage des
données climatiques à
l'ère du big data

Modèle fonctionnel pour l'archivage des données climatiques à l'ère du big data :

□ Concept d'archivage ?

- méthode de stockage
- ensemble de processus pour assurer la pérennité de l'information

□ Modèle inspiré de l'OAIS (système ouvert d'archivage d'information)?

- Modèle fonctionnel pour l'archivage des big data présenté par VIEIRA Alexandre dans son mémoire de recherche « Archiver les Big Data : un enjeu pour l'archiviste d'aujourd'hui et de demain ? »;
- E - ARK (European Archival Records and Knowledge);
- Projet VITAM (**V**aleurs **i**mmatérielles transmises aux archives pour **m**émoire)

Collecte des données météorologiques (structurées et non structurées) chaque jour et en Giga-octets

Collecte

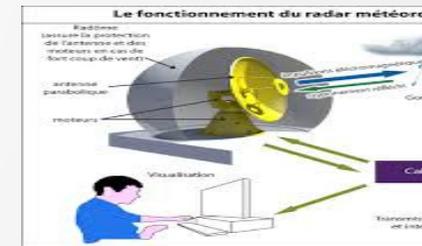


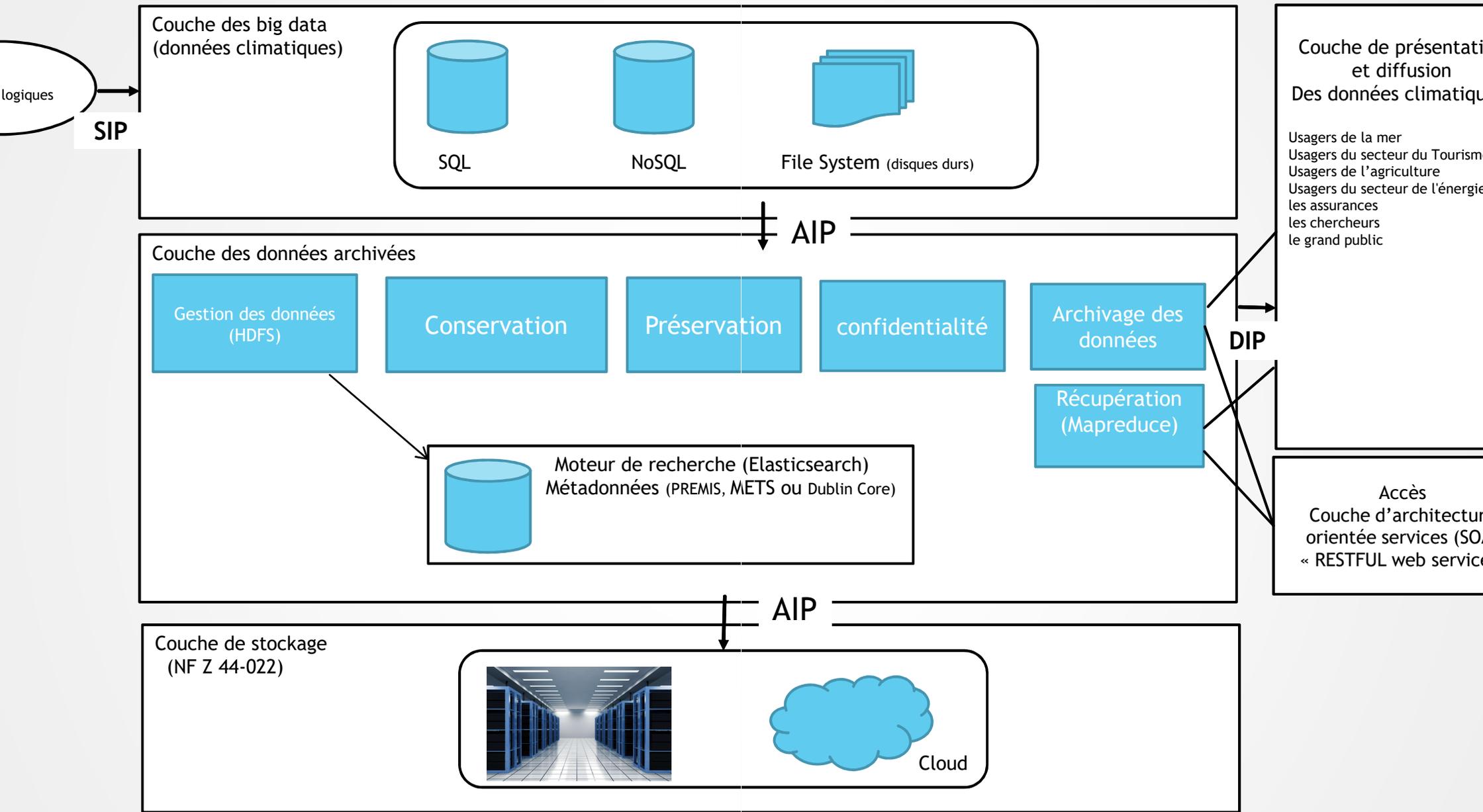
Collecte

Collecte



ORACLE®



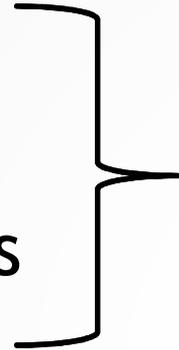


Modèle de gestion et d'archivage des données climatiques à l'ère du big data ¹⁵

Première couche: mégadonnées climatiques

Données structurées

Données non-structurées



Données météorologiques
collectées de différentes sources

C'est l'espace qui contient les données chaudes, stockées dans des espaces fortement sollicités» (A. VIEIRA,2018)

Deuxième couche: données climatiques archivées

Actions d'archivage ; préservation ; communicabilité des données;

Un paquet SIRF:

- préservation des données brutes et des métadonnées ;
- stockage sur disques ou cloud;

Objet modélisé sous la forme d'un OAIS AIP



Archivage à long terme

Metadonnées

PREMIS : norme de métadonnées :

- enregistrer les informations nécessaires à la préservation des objets numériques
- garantir la convivialité à long terme

METS :

- création et description intégrale des objets numériques;
- composé de sept sections principales (en-tête, métadonnées descriptives, métadonnées administratives, section des fichiers, carte de structure, section des liens, comportement);

Dublin Core

- 15 éléments
- description des données

Troisième couche: stockage des données

cluster de serveurs

groupe de serveurs

reliés entre eux par plusieurs réseaux

fonctionnant comme un seul système



Solutions Cloud

LDPaaS « longterm digital preservation as a service » ;

XaaS « Everything as a service » ;

Format SIRF :

- Stockage en nuage
- Préserver et sécuriser les informations

Solution d'archivage sur le Cloud « OpenStack Swift »

- assure : le stockage ; l'intégrité des données

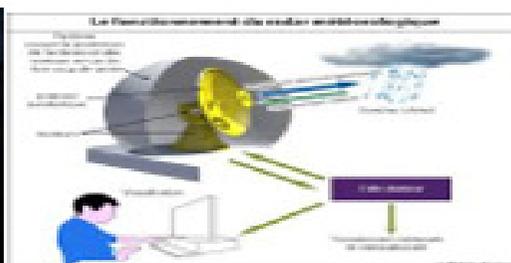
Dernière couche : Accès et sécurité

Accès → SOA (architecture orientée services)
→ API (interface de programmation d'application)
« RESTful » ; Exposition données/documents

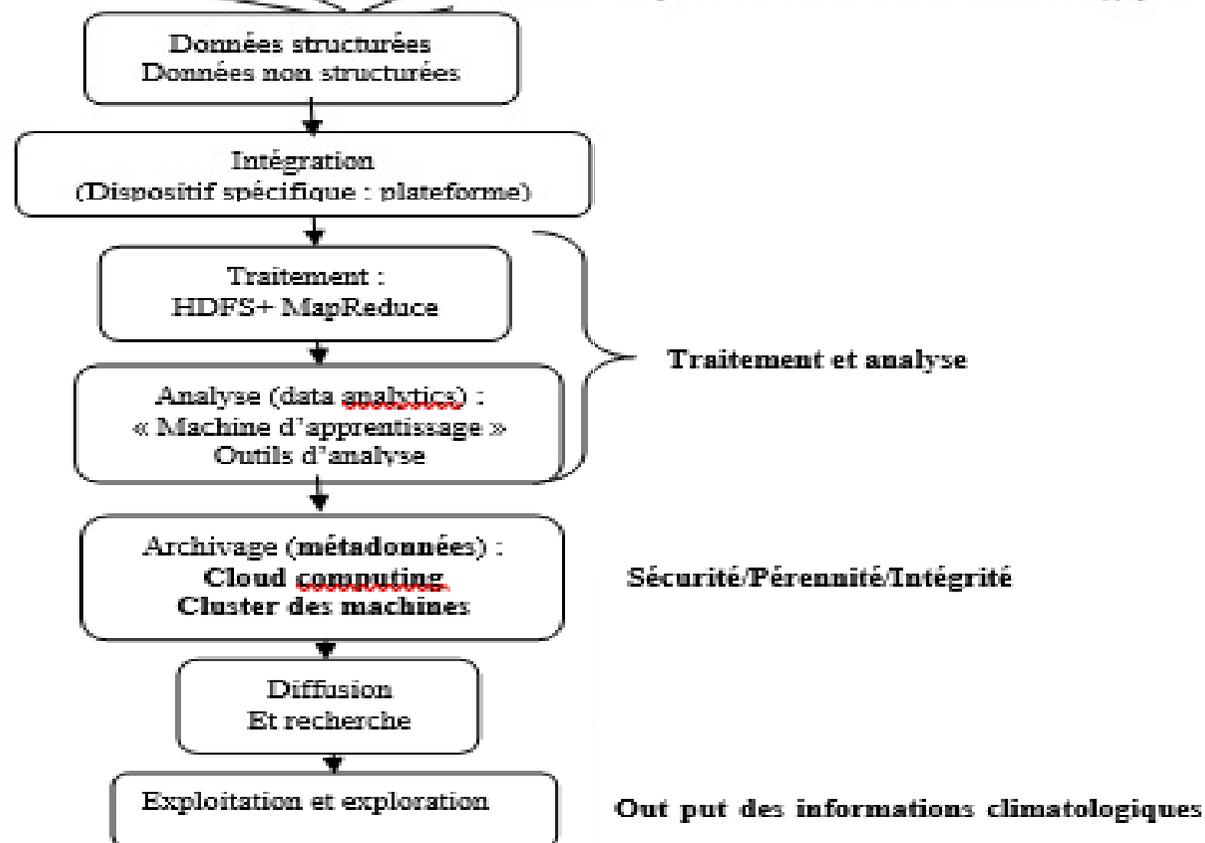
SOA: forme d'architecture de médiation ; mettre en œuvre des services (format d'échange XML) ;

API RESTful : permet à l'utilisateur d'accéder à des services Cloud ; le client ne connaît pas la structure et le contenu des informations stockées ;

Plan de gestion et d'archivage des données climatiques à l'ère du big data



Collecte / In put : des observations météorologiques



En résumé, la plateforme de gestion et d'archivage des données climatiques à l'ère du big data nous permet de faire :

- la collecte des données en temps réel ;
- la gestion et l'analyse intelligente d'une quantité volumineuse des données;
- la préservation et la confidentialité de ces mégadonnées à travers le temps;
- l'archivage et le stockage des données climatiques à long terme;
- la recherche et la récupération des données.

Conclusion

Donnée climatique :

- Importante;
- non modifiable (intégrité);
- conservée à long terme (pérennité, accessibilité, sécurité).

Ce qui impose :

- ▶ Un modèle fonctionnel adapté aux données climatiques à l'ère du big data;
- ▶ Des nouvelles méthodes de gestion et d'archivage des données climatiques à l'INM;

Les contraintes:

Sujet : nouveau, complexe, difficile;

Concept « big data » ↔ nouveau, flou ↔ secteurs publics ;

Absence des idées :

- Applications du big data ;
- Technologie de gestion et d'archivage.

Perspectives

Big data : large, compliqué

Du big Data au smart data ?

analyse des données (data analytique) ↔ intelligence artificielle

. Analyse des données stockées pour assurer des prévisions
météorologiques;

data vs capta ?

Modélisation d'une plateforme de partage d'information à l'ère du big data entre plusieurs services en Tunisie (météo, agriculture, transports, etc..)

***Merci pour
votre attention***