

Christine Collet: passion pour les bases de données

Genoveva Vargas-Solar
Chargée de Recherches, CNRS, LIG-LAFMIA
genoveva.vargas@imag.fr

<http://vargas-solar.com>





ARCHITECTURES
DES SGBD



DONNÉES



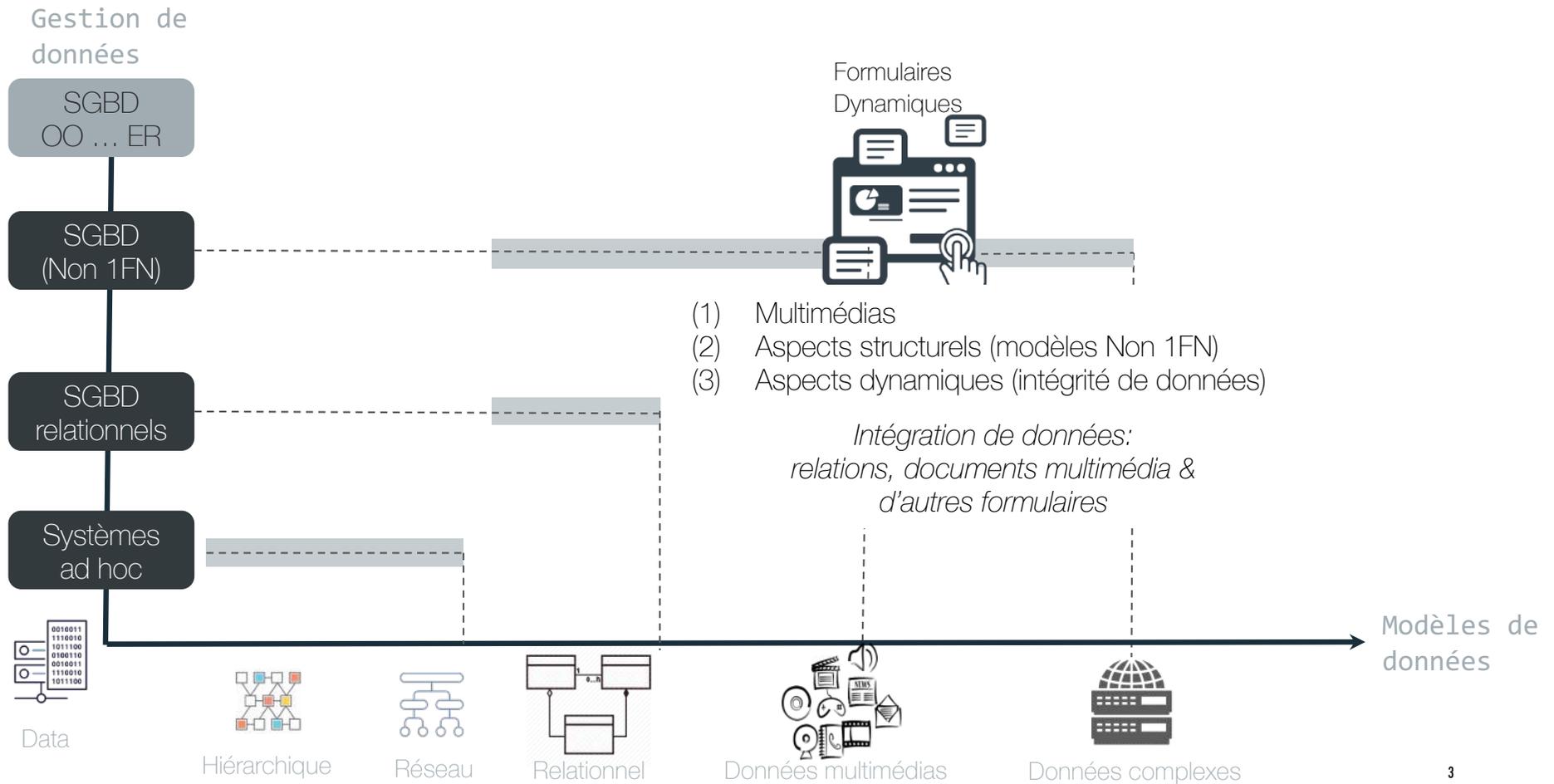
GESTION EFFICACE
DE DONNÉES



001001100100110010011
111001011100101110010
101110010111001011100
010011001001100100110
001001100100110010011
111001011100101110010
101110010111001011100



DONNÉES COMPLEXES ET MULTIMÉDIAS



PRÉSENTATION DE DONNÉES COMPLEXES & MULTIMÉDIAS

create(member, ?a); create a new occurrence ?a of FA "member" : ?a will take the value of the occurrence identifier.
assign(?a.total, count(?a.loans)); for any occurrence ?a, assign to element "total" the result of the function count applied on the set of "loans".

Opérations	Fonctions
Filtrage	
selection	Sélection d'occurrences
prune	ChoiX de certains éléments
rename	Renommer un FA et/ou ses éléments
sort	Ordonnancer des occurrences
Ensemble	
union	Union de deux FA
intersection	Intersection de deux FA
difference	Différence de deux FA
product	Produit cartésien de deux FA
Restructuration	
nest	Création d'un groupe
unnest	Destruction d'un groupe

```

WHEN assign(?a.loans(?i).ref, ?r);
BEFORE FA(CAT,
  FA(CAT, selection (LIBRARYCATALOG,
    =(serial-number, ?r)));
  if( non( exist(CAT)), message('no book corresponding
    to this serial number');
    stop-event; )
  FA(NOTRETURN, selection (member, (selection
    (loans,  $\exists$  selection(loan, and (=(ref,?r), =(return,  $\Phi$  ) ) ) ) ) ) )
  if( exist (NOTRETURN), message('this book is already
    on loan'); stop-event; )
AFTER for ( (CAT, ?c),
  assign(?a.loans(?i).title, ?c.title); );
  
```

CREATE SELECT MODIFY VALIDATE QUIT HELP

Member

Name : _____

Sex : man woman PICTURE

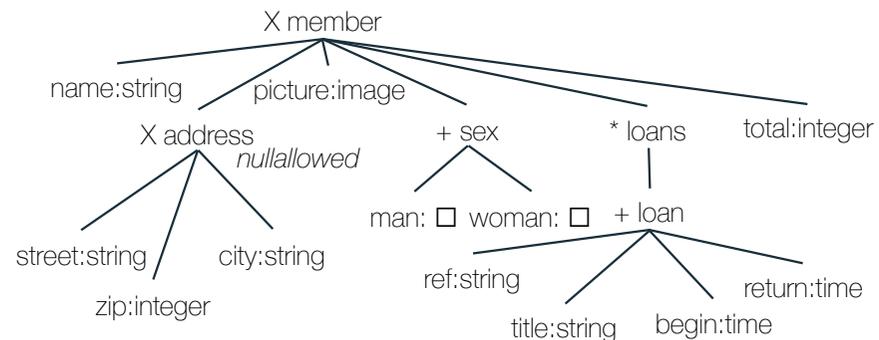
Address Street : _____

Zip : _____ City : _____

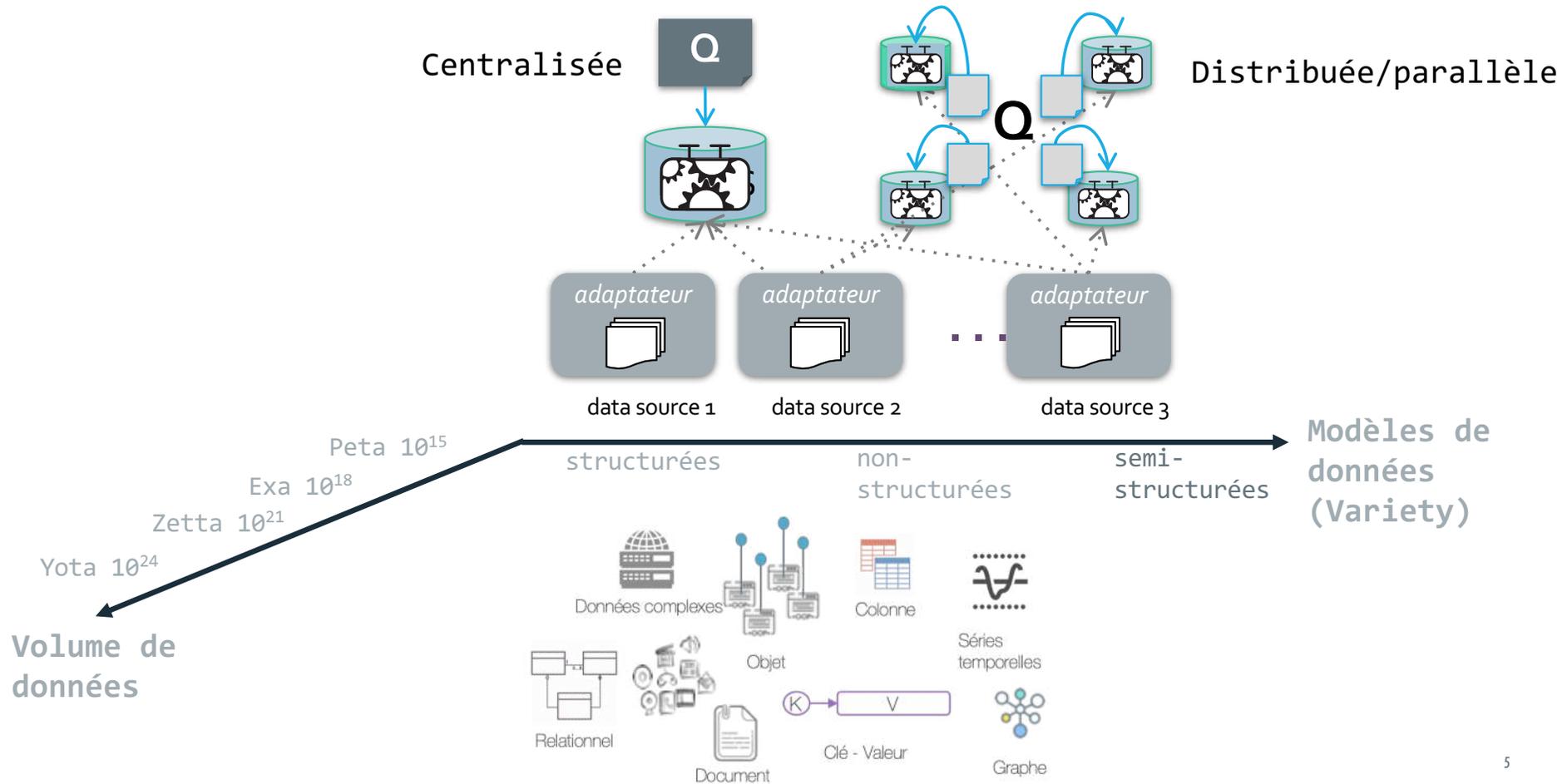
LOANS

Serial number	Title	Loan_date	Return_date
_____	_____	____/____/____	____/____/____
_____	_____	____/____/____	____/____/____
_____	_____	____/____/____	____/____/____
_____	_____	____/____/____	____/____/____
_____	_____	____/____/____	____/____/____

Total of loans : _____

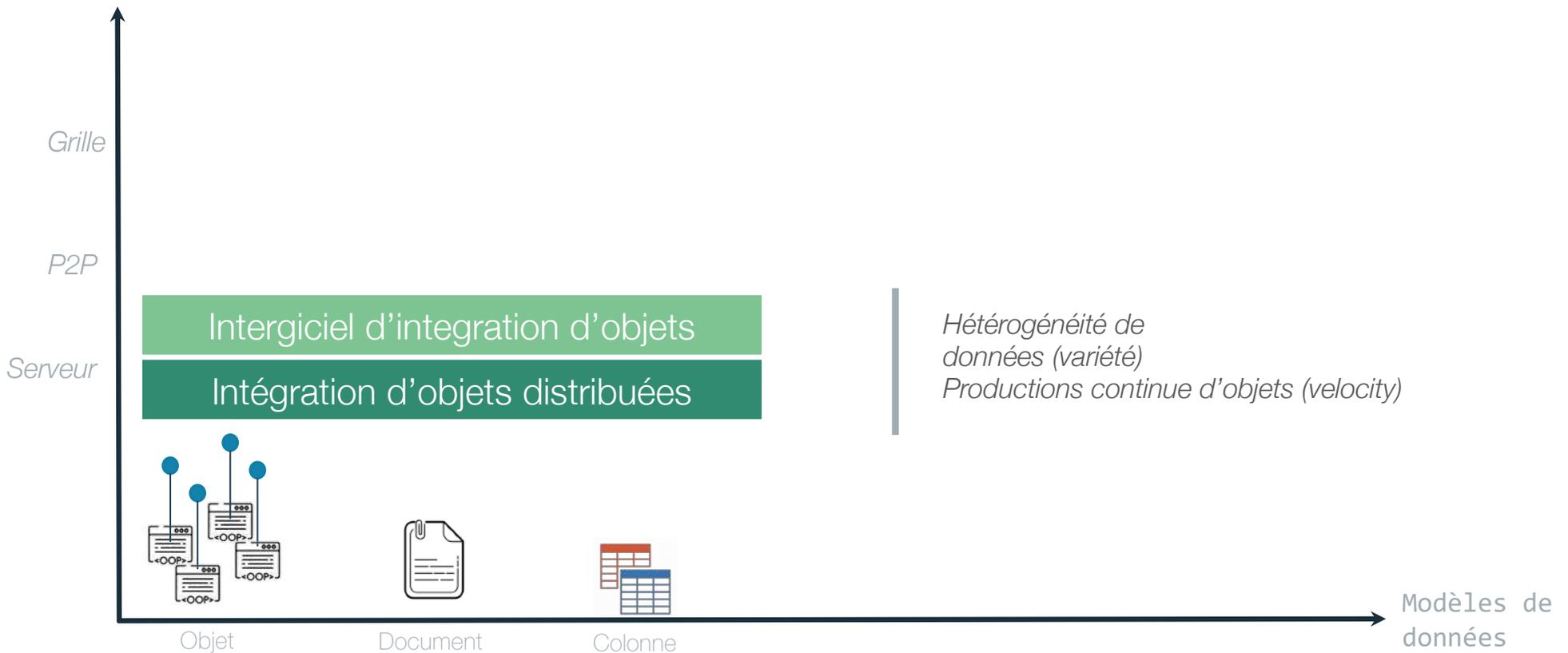


INTÉGRATION DE DONNÉES: CHANGEMENT D'ÉCHELLES



INTÉGRATION DE DONNÉES: CHANGEMENT D'ÉCHELLES

Architectures de
déploiement



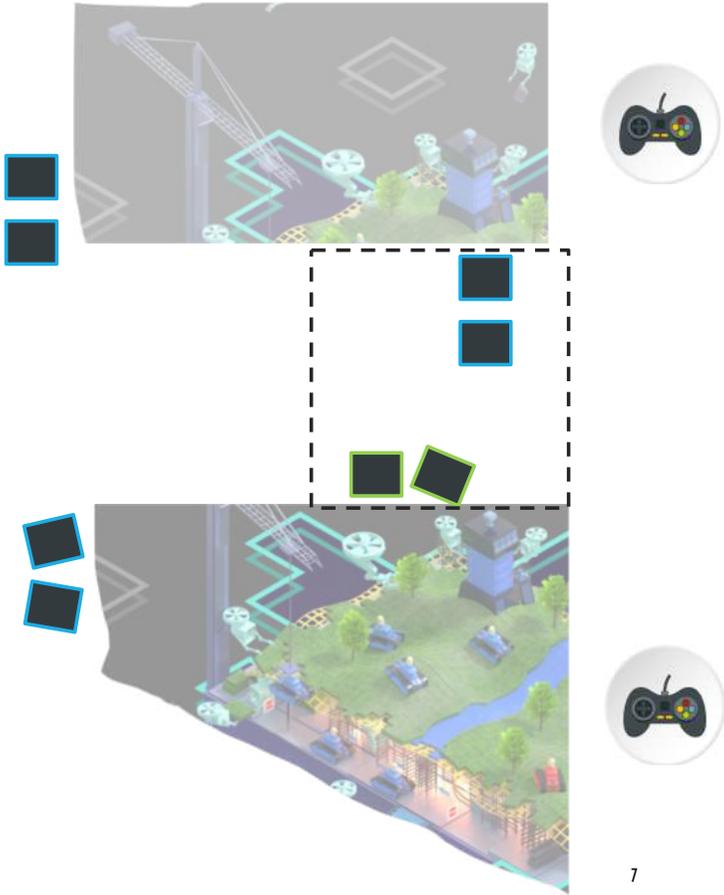
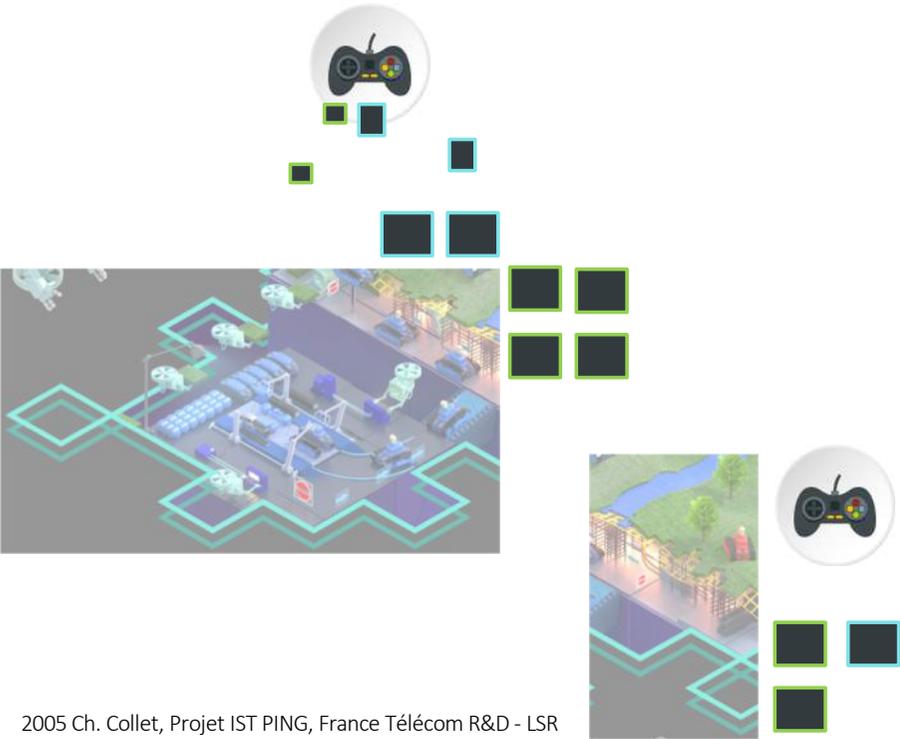
M. Alia, Ch. Collet, A. Lefebvre, Un système d'intégration des données : Une approche à composants, Numéro spécial de la Revue des Sciences et Technologies de l'Information (STI), série L'objet, Vol. 2-3, Hermès, 2004, pp. 117-130.

OBJETS PERSISTANTS DANS LES JEUX VIRTUELS RÉPARTIS

● *disponibilité*

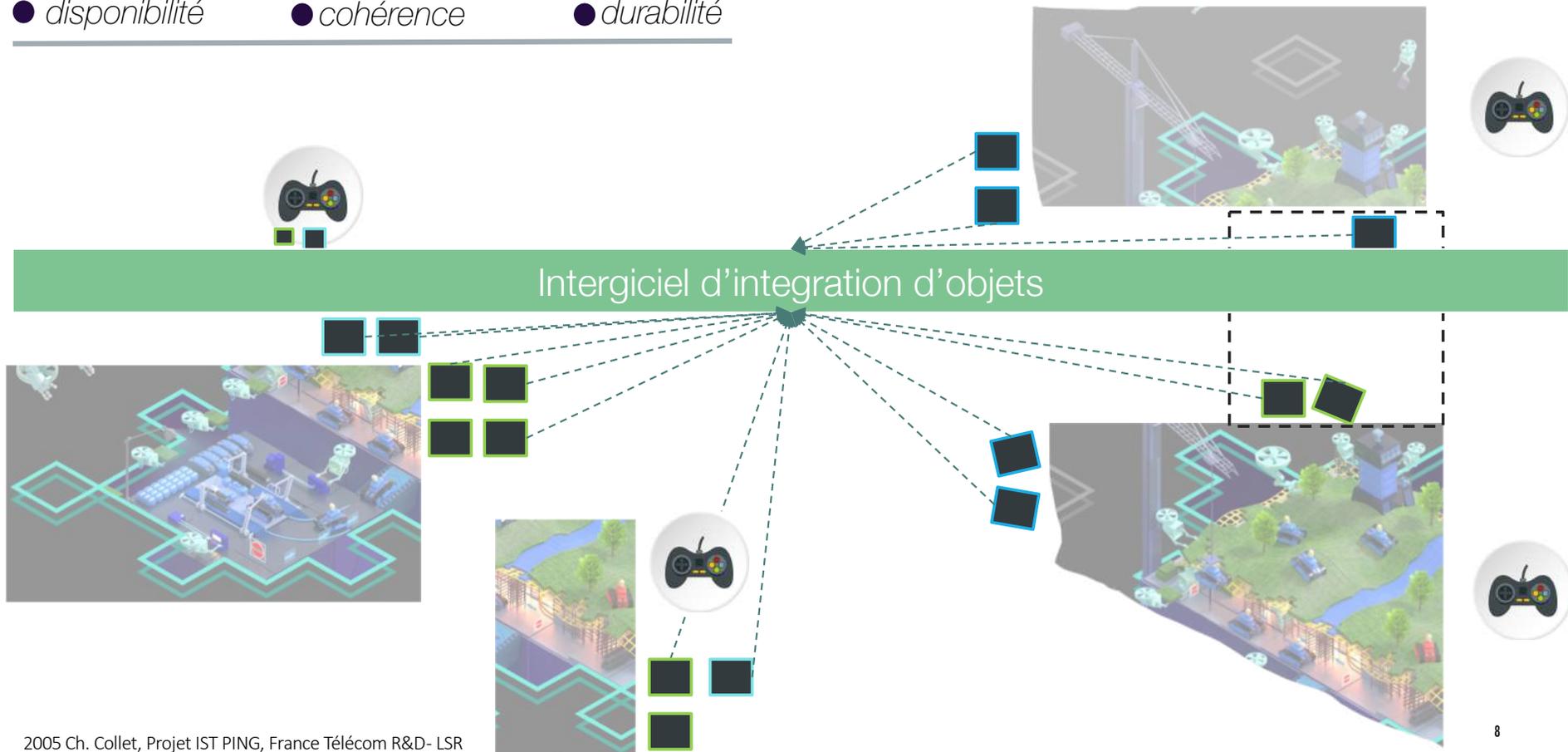
● *cohérence*

● *durabilité*



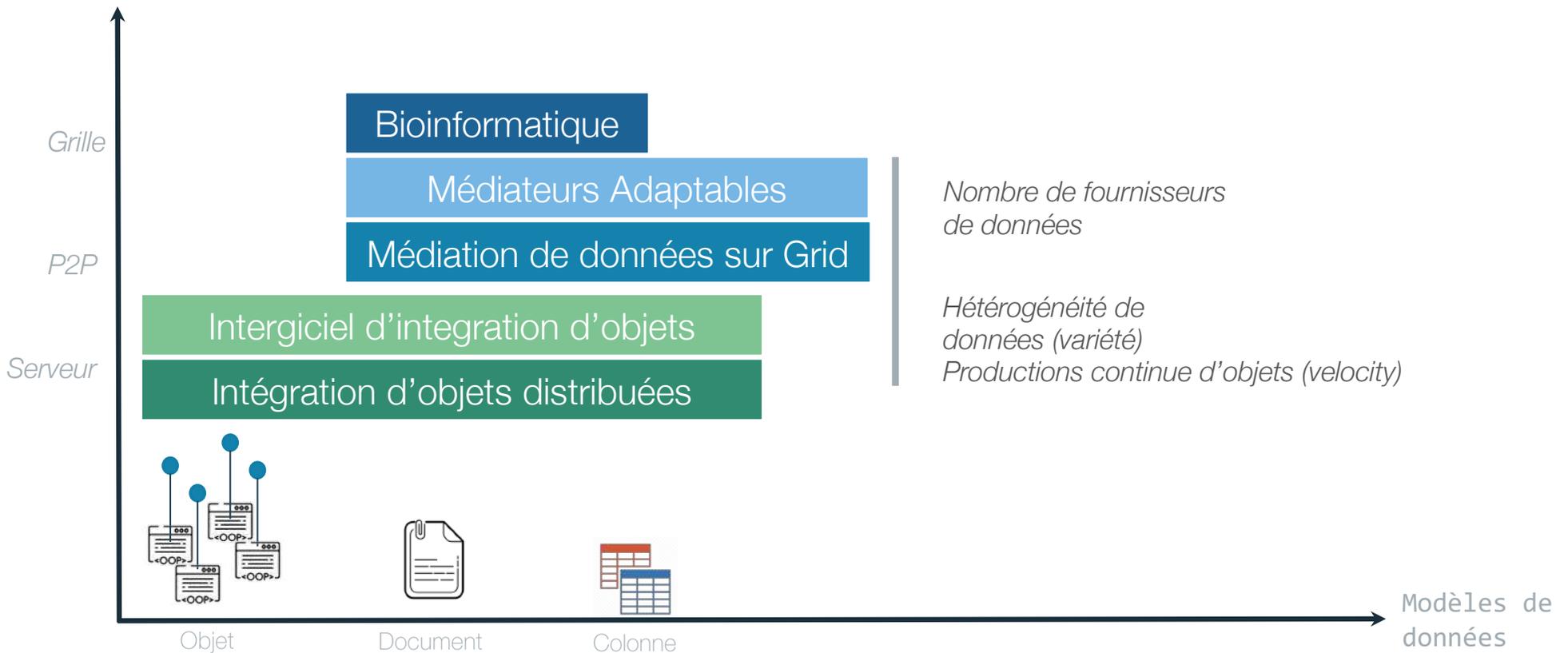
OBJETS PERSISTANTS DANS LES JEUX VIRTUELS RÉPARTIS

● *disponibilité* ● *cohérence* ● *durabilité*



INTÉGRATION DE DONNÉES: CHANGEMENT D'ÉCHELLES

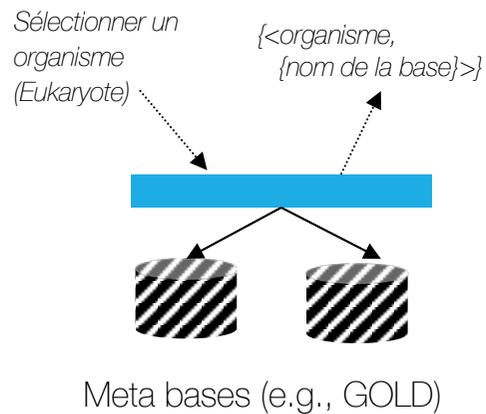
Architectures de déploiement



Collet, Christine, et al. "Towards a mediation system framework for transparent access to largely distributed sources." International Conference on Semantics for the Networked World. Springer, Berlin, Heidelberg, 2004.

MÉDIATION DE BASES DE DONNÉES BIOLOGIQUES

Chercher les bases de données



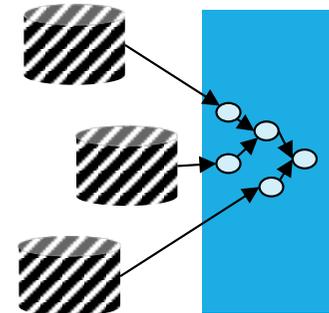
Choisir les bases de données

Information cartographique (SGD, NCBI)

Expression des gènes (SMD, GeO)

Configurer le système de médiation

Intégration de données



MÉDIATION DE BASES DE DONNÉES BIOLOGIQUES

Système de médiation
ad hoc

Interrogation

Corréler les niveaux d'expression de chaque
gène avec sa localisation génomique &
Observer son évolution

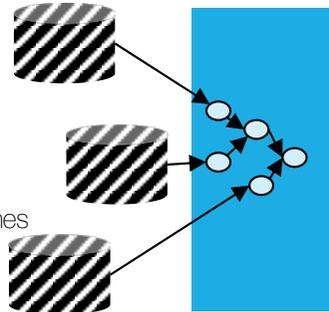
Intégration de données

Evaluation interactive &
partielle

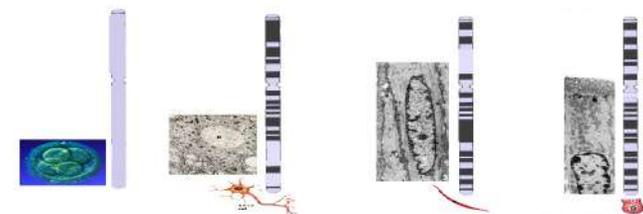
Zones ouvertes / fermées

Information
cartographique
(SGD, NCBI)

Expression des gènes
(SMD, GeO)



Chercher des corrélations
d'expression de gènes

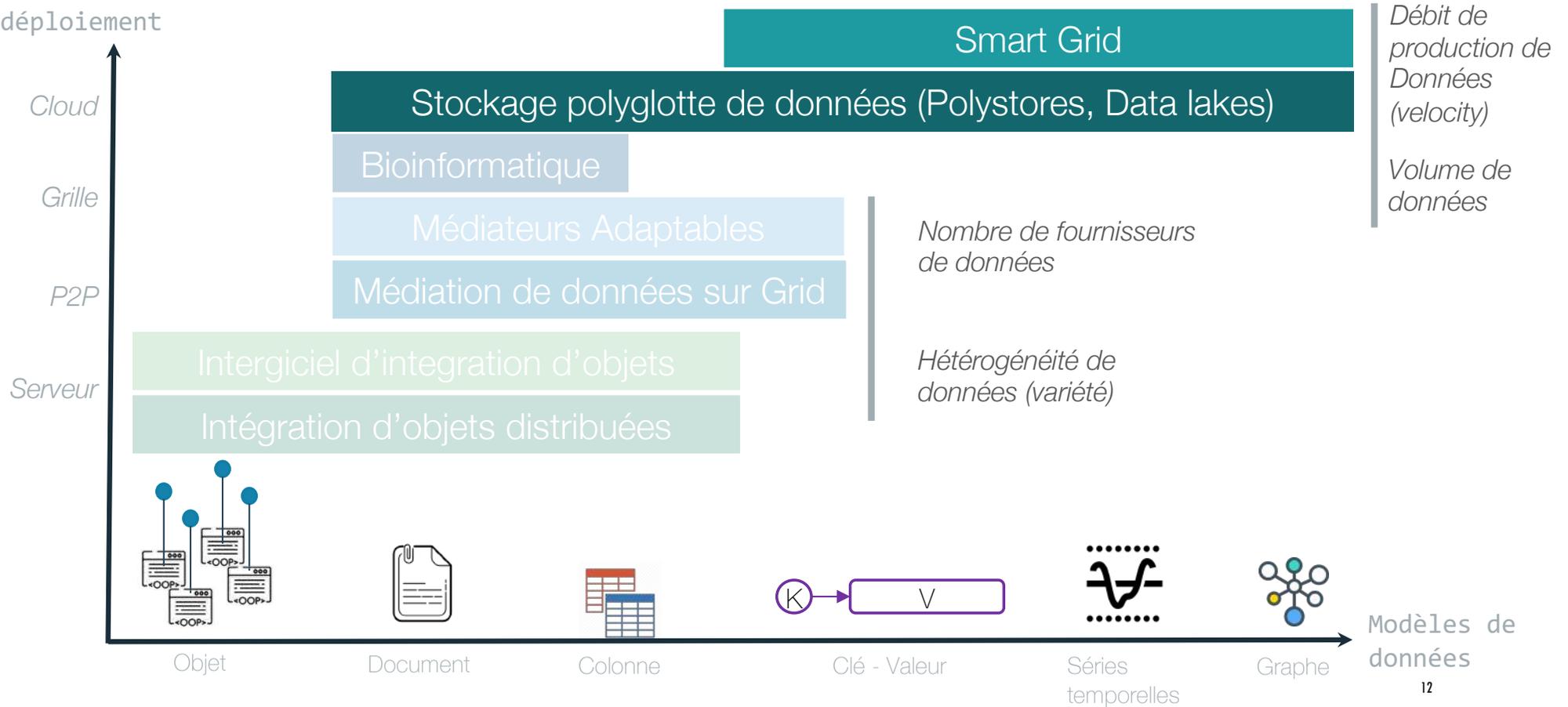


Noyaux des cellules différenciées : Zones brillantes (ouvertes)
Zones obscures (fermées)

Zone fermée → gènes non exprimés

INTÉGRATION DE DONNÉES: CHANGEMENT D'ÉCHELLES

Architectures de déploiement





LES DONNÉES DANS LE « SMART GRID »

LES DONNÉES DANS LE « SMART GRID »



Echantillonnage de données

Utilisation des techniques mathématiques et éthiques pour construire des échantillons robustes à partir des systèmes de collecte continue des données

Fusion de collections de données

Intégration de collections de données préexistantes & des qualités différentes pour construire une vue intégrée des données clé sur l'énergie

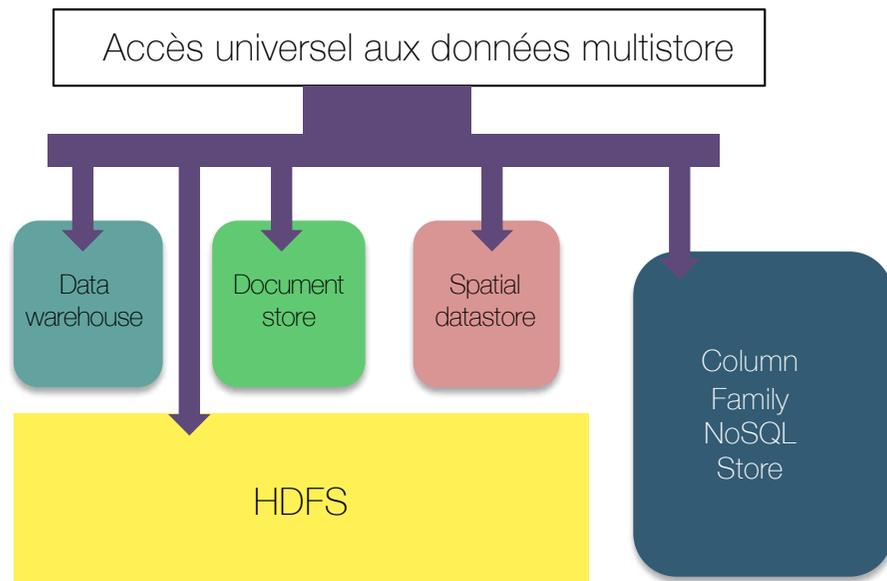
Innovation guidée par les données

Utilisation de innovante de la science et la recherche pour gérer de manière proactive la privacité des données et construction des vues analytiques sur l'énergie pour ajouter de la valeur à cette connaissance

Accès interactif aux données

Interrogation de données personnalisée et amicale pour explorer des collections de données et comprendre la distribution & la consommation d'énergie

MULTI-STORE HYBRIDE



Teradata, MongoDB, Elasticsearch,
Cassandra, HBase, HDFS ...

différents modèles, langages de requêtes et architectures

Evaluation efficace de requêtes universelles

- Structures d'indexation et jointures hybrides
- Moteur d'exécution de requêtes multi-store et optimisation
- Intégration de données

→ Recommandation de présentation de données + contraintes de données Smart Grid

Gestion de métadonnées

- 1- Extraction de métadonnées
- 2- Traçabilité de données
- 3- Gestion de catalogue de métadonnées
4. Découverte de métadonnées



ARCHITECTURES
DES SGBD



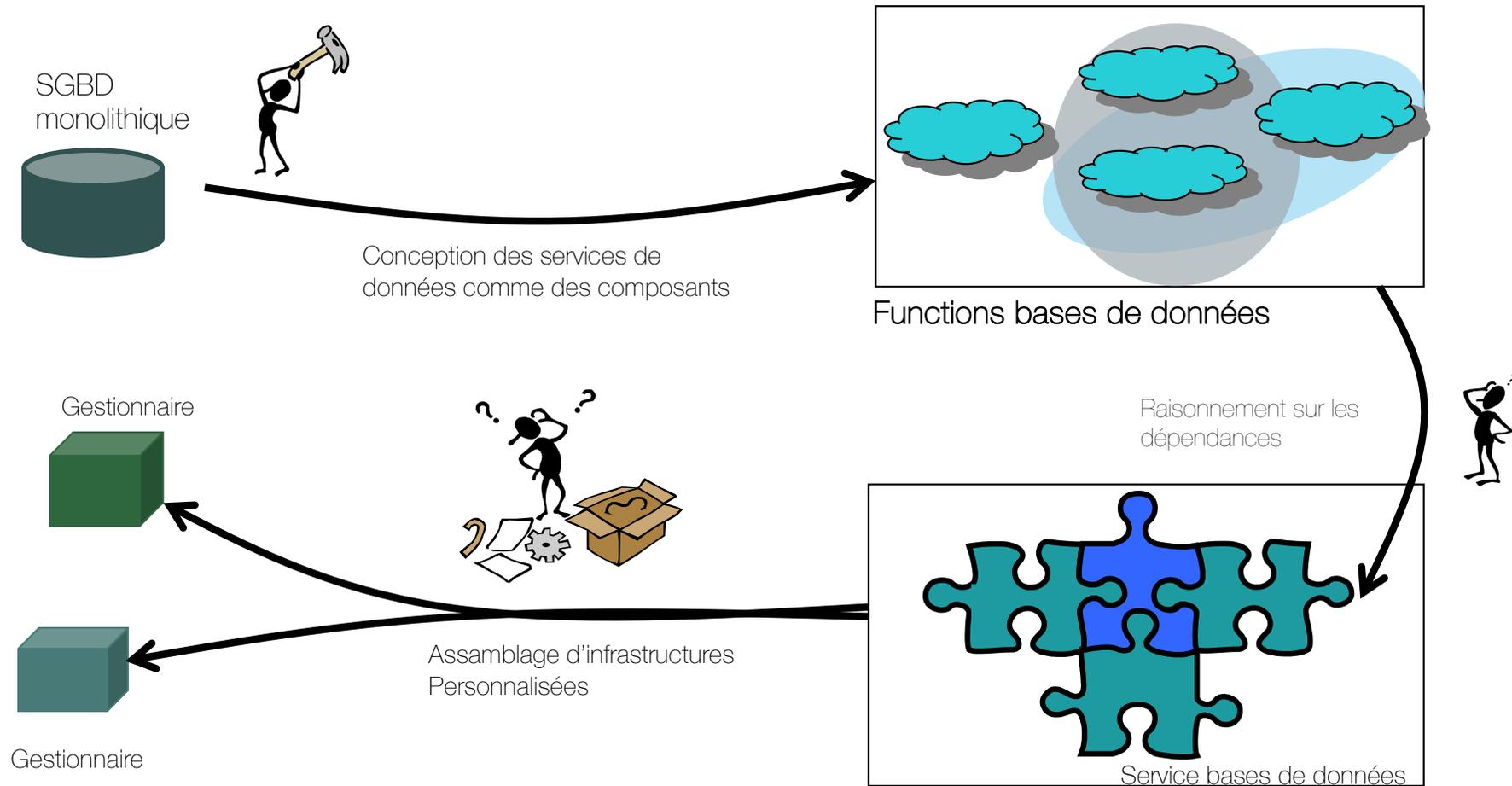
DONNÉES

GESTION
EFFICACE DE
DONNÉES

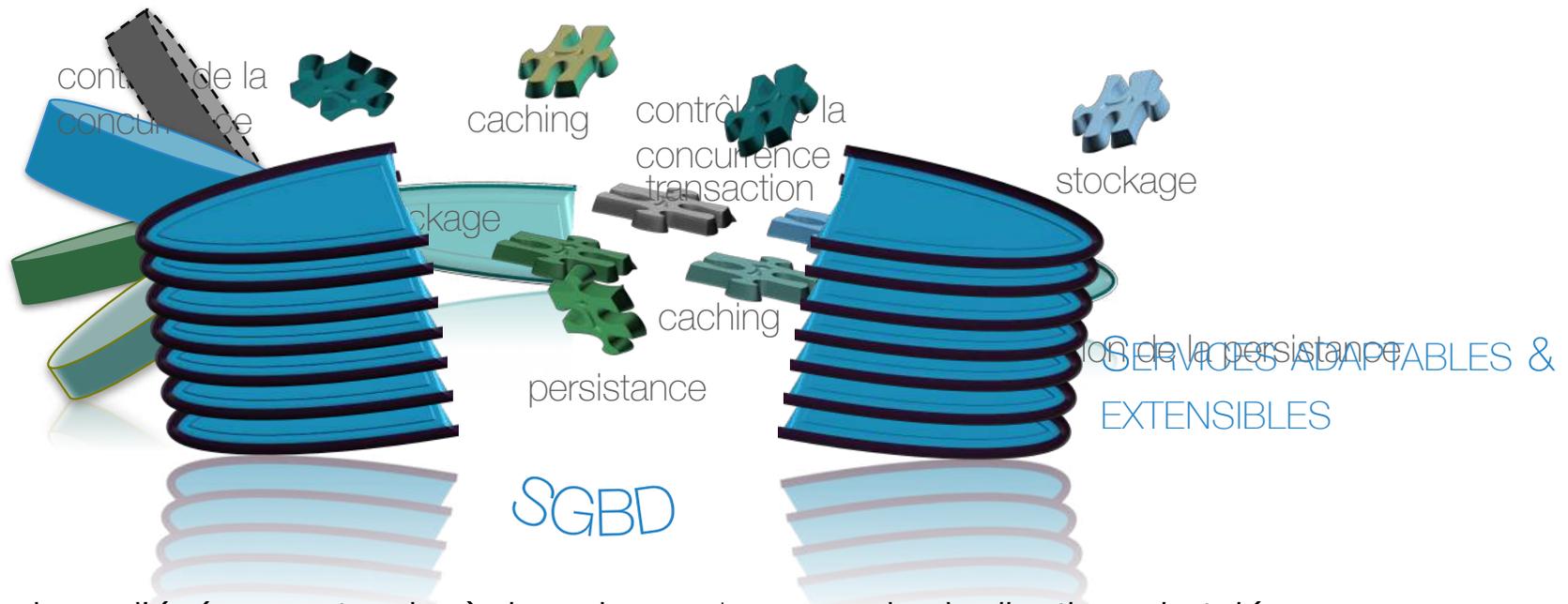


001001100100110010011
111001011100101110010
101110010111001011100
010011001001100100110
001001100100110010011
111001011100101110010
101110010111001011100

SERVICES BASES DE DONNÉES DU RÉSEAU: NODS



GESTIONNAIRES DE DONNÉES AD-HOC: SYSTÈMES NODS



Services d'événements, de règles, de persistance, de duplication, de tolérances aux fautes, de cache, de transactions



```
001001100100110010011  
111001011100101110010  
101110010111001011100  
010011001001100100110  
001001100100110010011  
111001011100101110010  
101110010111001011100
```



ARCHITECTURES
DES SGBD

DONNÉES

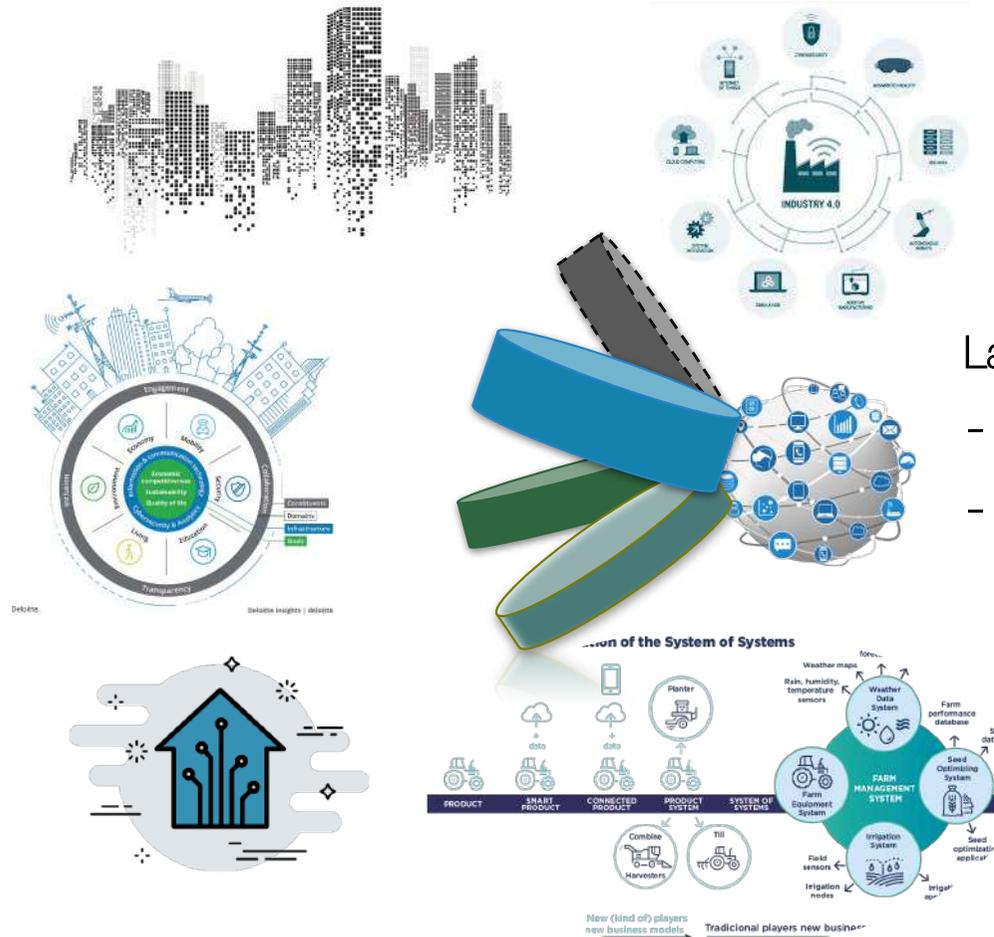
GESTION EFFICACE
DE DONNÉES



```
001001100100110010011  
111001011100101110010  
101110010111001011100  
010011001001100100110  
001001100100110010011  
111001011100101110010  
101110010111001011100
```



ENVIRONNEMENTS UBIQUITAIRES

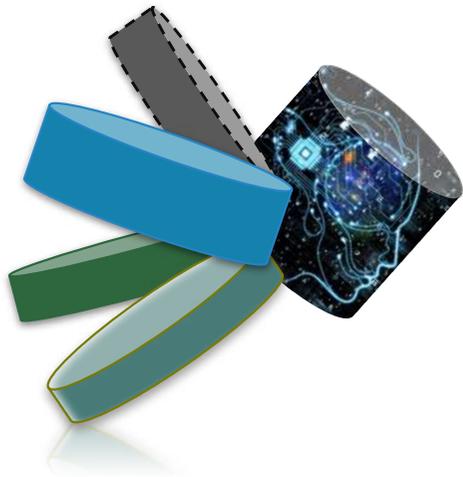


La configuration des SGBD est une tâche difficile

- Complexité des fonctions de gestion de données
- Besoins spécifiques des systèmes/applications/utilisatrices

VERS UN GESTION DE DONNÉES AUTONOME

Confluence entre la gestion de données et l'IA



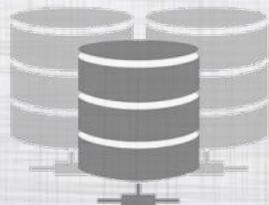
→ Proposer des techniques d'apprentissage automatique pour améliorer le comportement des SGBD

Les SGBD peuvent prendre des mauvaises décisions lors de l'optimisation de requêtes

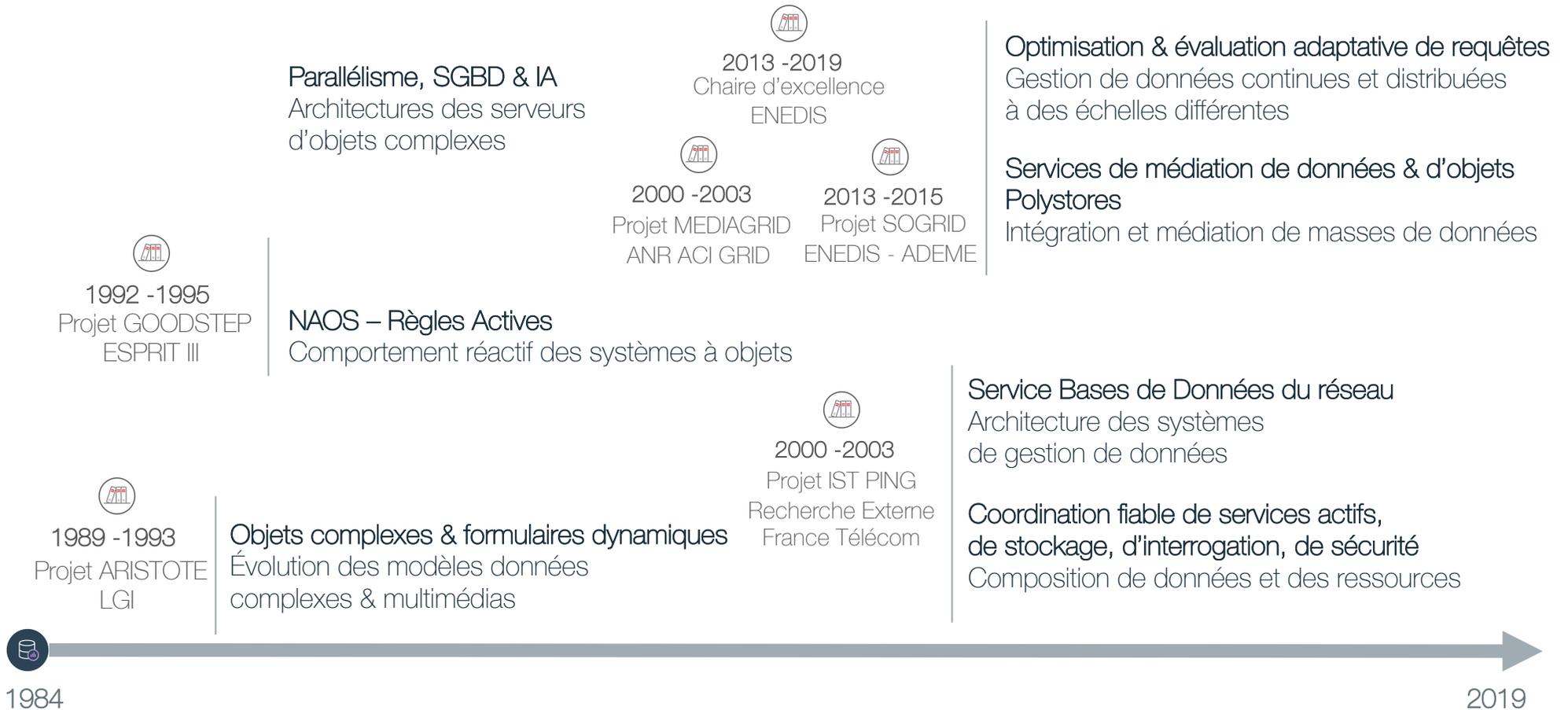
- Absence de méta-données suffisantes / pertinentes
- Mauvaises hypothèses sur les inter-dépendences des données

→ Utiliser des techniques d'apprentissage pour corriger des erreurs sur l'estimation de coût des requêtes

PASSION POUR LES BASES DE DONNÉES



CONTRIBUTIONS



"Ne pas partager est une perte de temps". –

Michèle Bernier

