



Entrepôts de données multidimensionnelles NoSQL

**Max Chevalier
Mohammed EL Malki
Arlind Akopliku
Olivier Teste
Ronan Tournier**

Sommaire

1 *Contexte - Objectif*

2 *Etat de l'art*

3 *Modélisation*

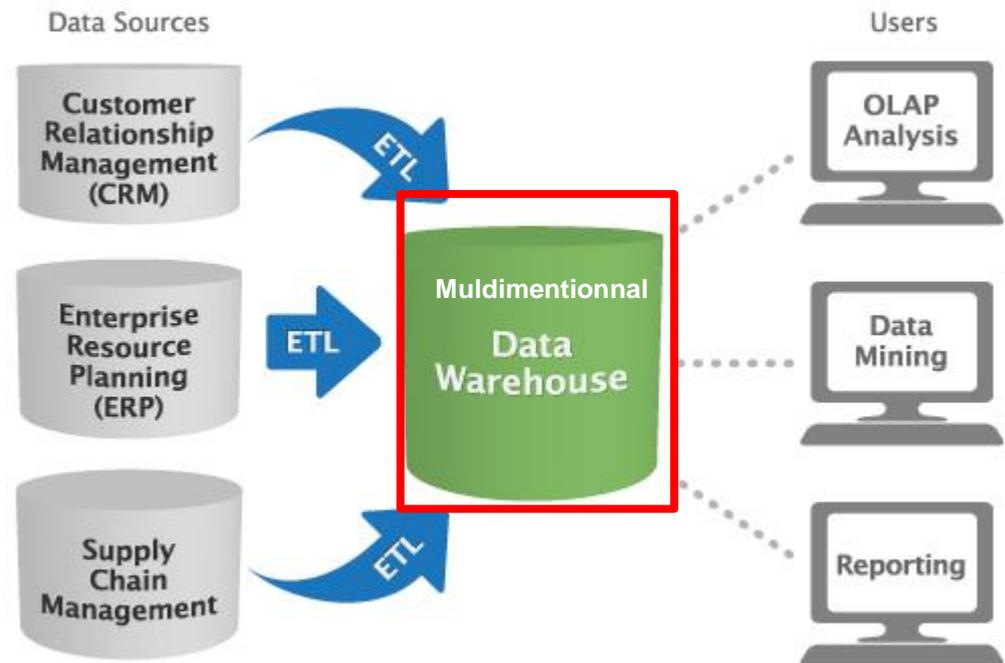
4 *Expérimentations*

Architecture d'un entrepôt de données:

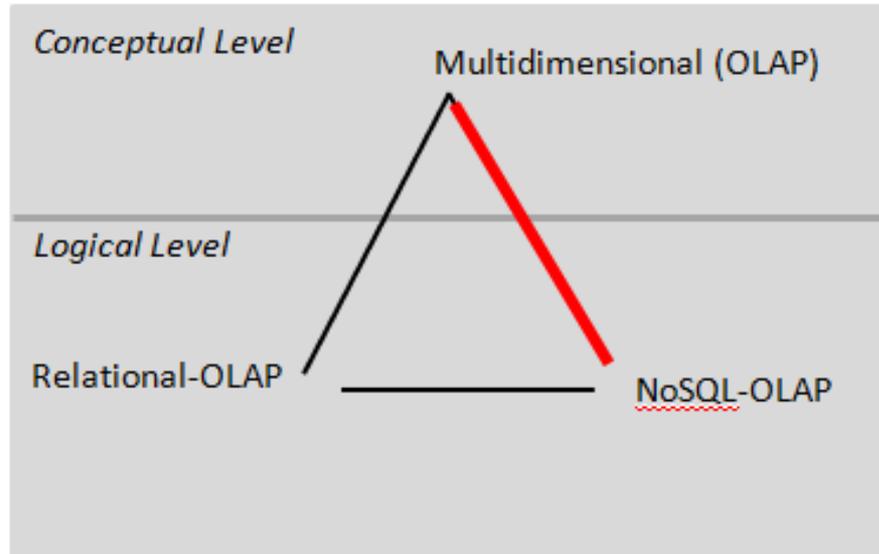
Explosion des données (web...)

Difficultés à gérer les données massives

Emergence de nouvelles technologies



Objectif



R-OLAP: Approche Traditionnelle

NoSQL: Mapping au niveau logique

Travaux étudiés indépendamment

4 modèles

- Clé – valeur: Redis
- Document: MongoDB , CouchDB
- Colonne: Hbase, Cassandra
- Graphe: Neo4j

Scalables



**Gestion de Données
massives**

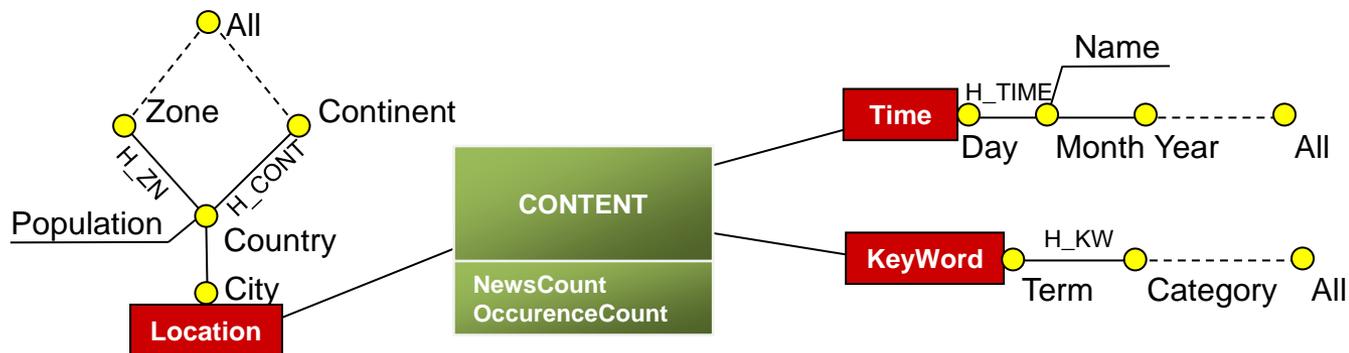
Distribués

Etat de l'art

	Modèle NoSQL	Modélisation			Treillis	Règles de transformation NoSQL
		Fait	Dimension	Hiérarchie		
Implémentation R-OLAP, [Morfonios 2007]		X	X	X	X	
Mapping [Chongxin Li, 2010]	Colonnes					
CNSSB [Dehdouh et al, 2014]	Colonnes	X	X	X		

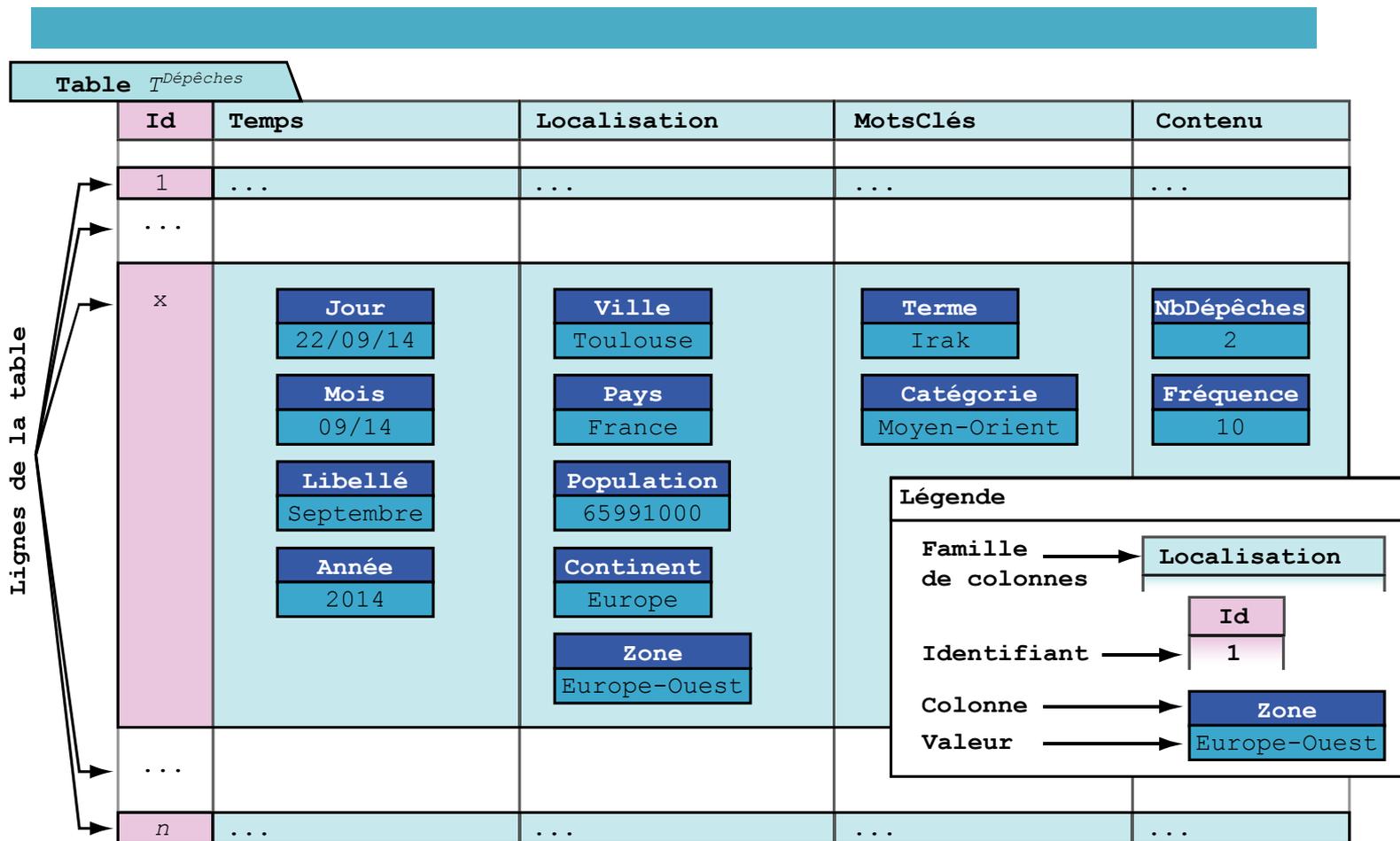
Le schéma en étoile

Le Modèle conceptuel en étoile



Le NoSQL

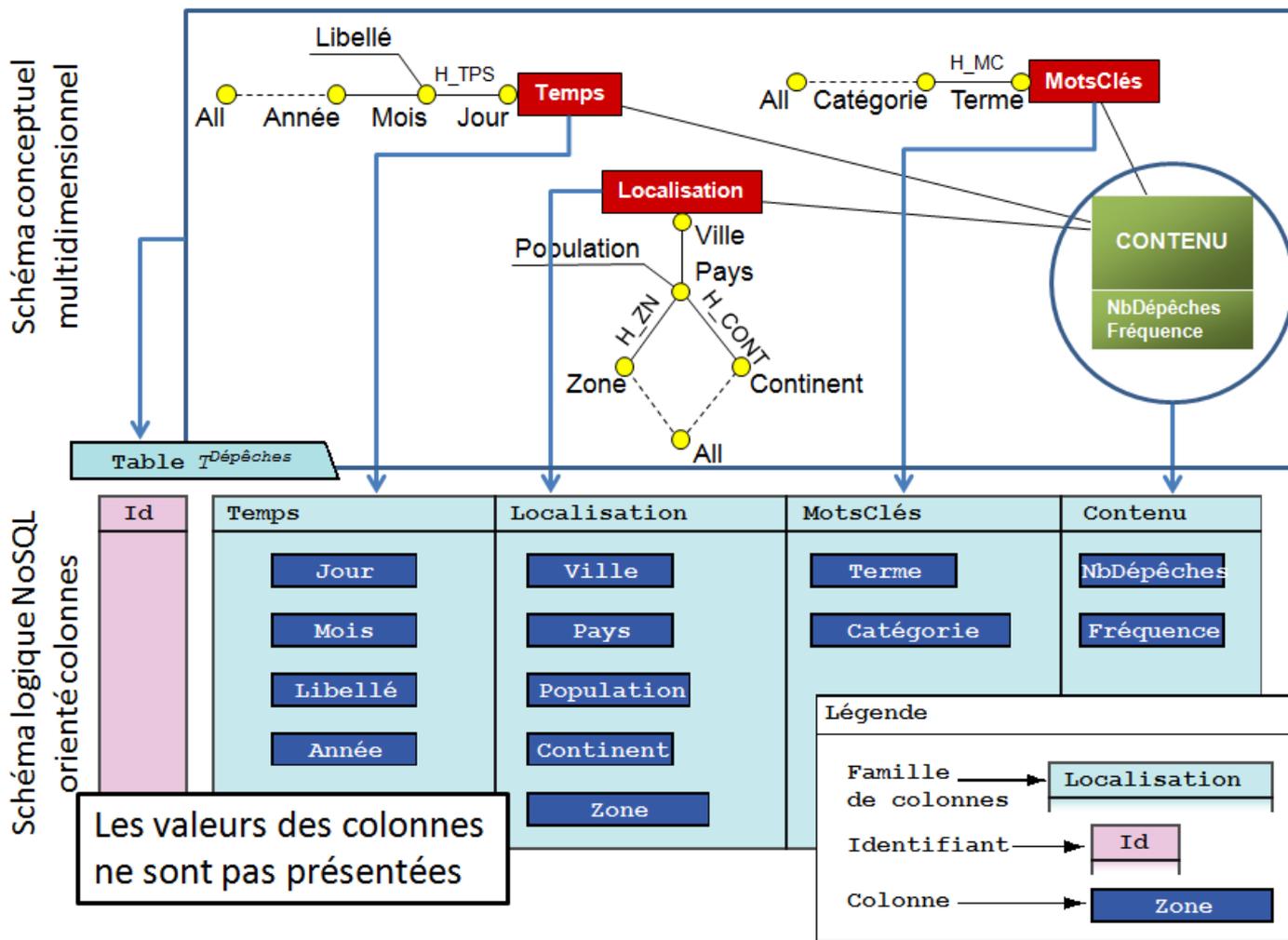
Orienté colonnes: stockage vertical par famille de colonnes



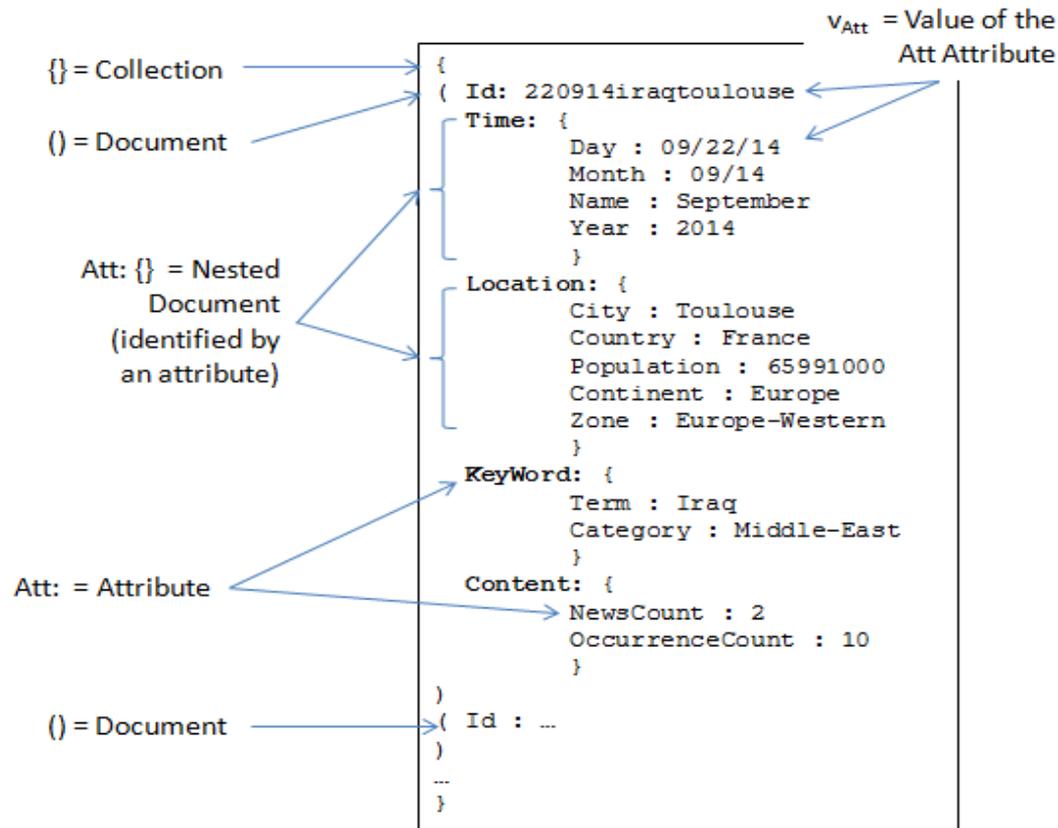
Règles de passage

Modèle OLAP	Fait	Mesure	Dimension	Paramètre	Attribut Faible
Modèles NoSQL					
Orienté Colonnes	Famille de colonnes	Colonne	Famille de colonnes	Colonne	Colonne

Règles de passage (Colonnes)



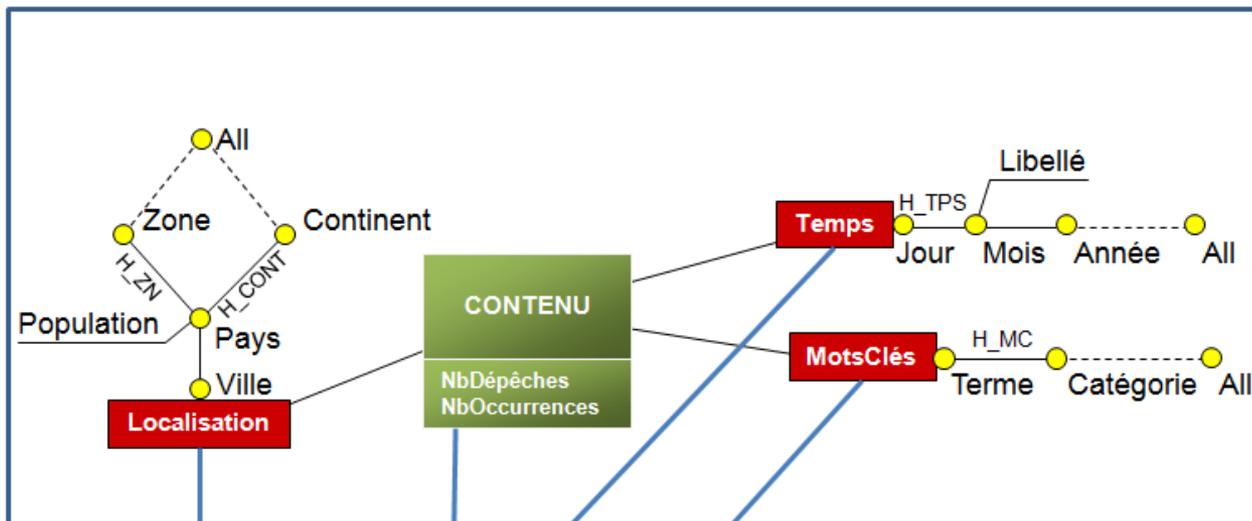
Orienté Documents: stockage horizontal par document



Règles de passage

Modèles NoSQL	Modèle OLAP	Fait	Mesure	Dimension	Paramètre	Attribut Faible
Orienté Documents		Document imbriqué (Attribut composé)	Attribut simple	Document imbriqué (Attribut composé)	Attribut simple	Attribut simple

Règles de passage (Documents)



Légende

- { } Collection
- () Document
- Att: { } Document imbriqué
- Att: Attribut
- v_{Att} Valeur de l'attribut Att

```

{
  (Id : vId
  Temps : {Jour: vJo, Mois: vMo, Libellé: vLi, Année: vAn}
  Localisation: {Ville: vVi, Pays: vPa, Population: vPo,
                 Continent: vCo, Zone: vZn}
  MotsClés : {Terme: vTe, Catégorie: vCa}
  Contenu : {NbDepeches: vNbD, NbOccurrences: vNbO}
  ( Id ... )
  ...
}
                
```

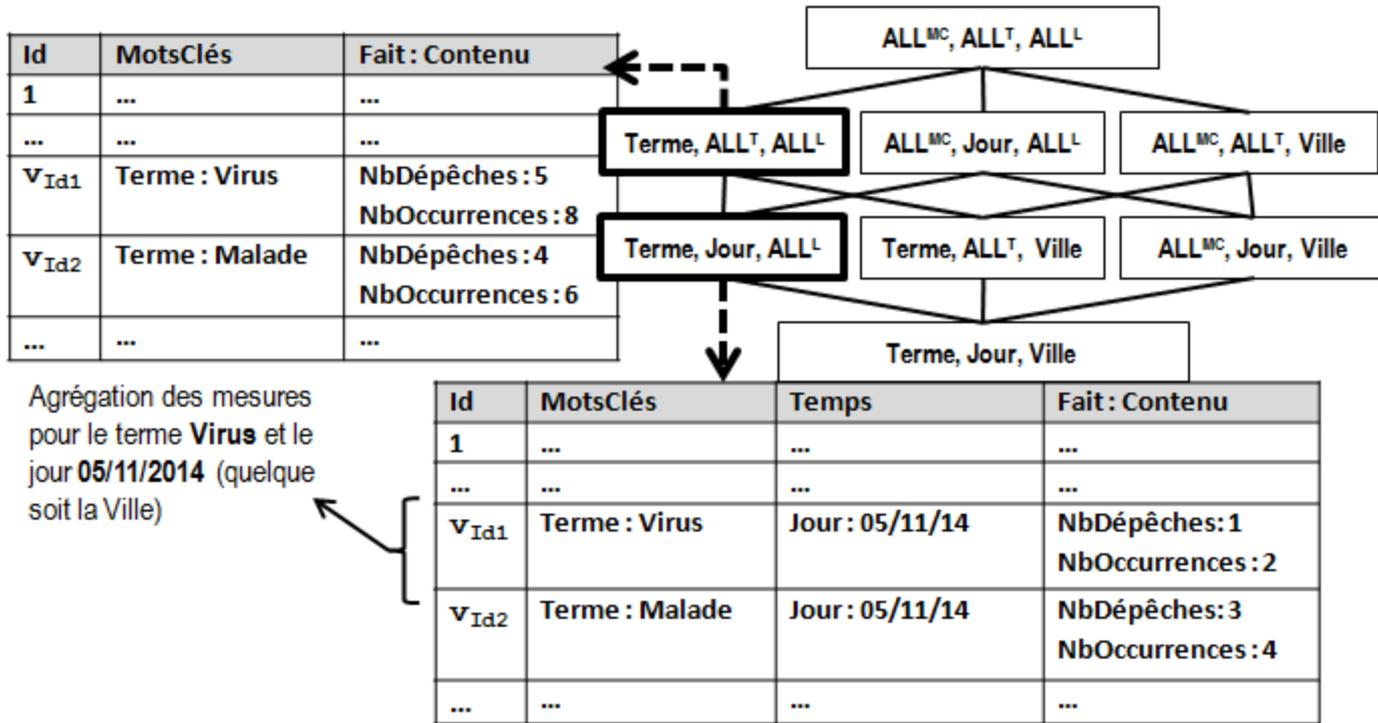
Les valeurs des attributs sont présentées pour des raisons de clarté

Treillis d'agrégats

Relationnel: Techniques de pré-calcul représentées sous forme de treillis

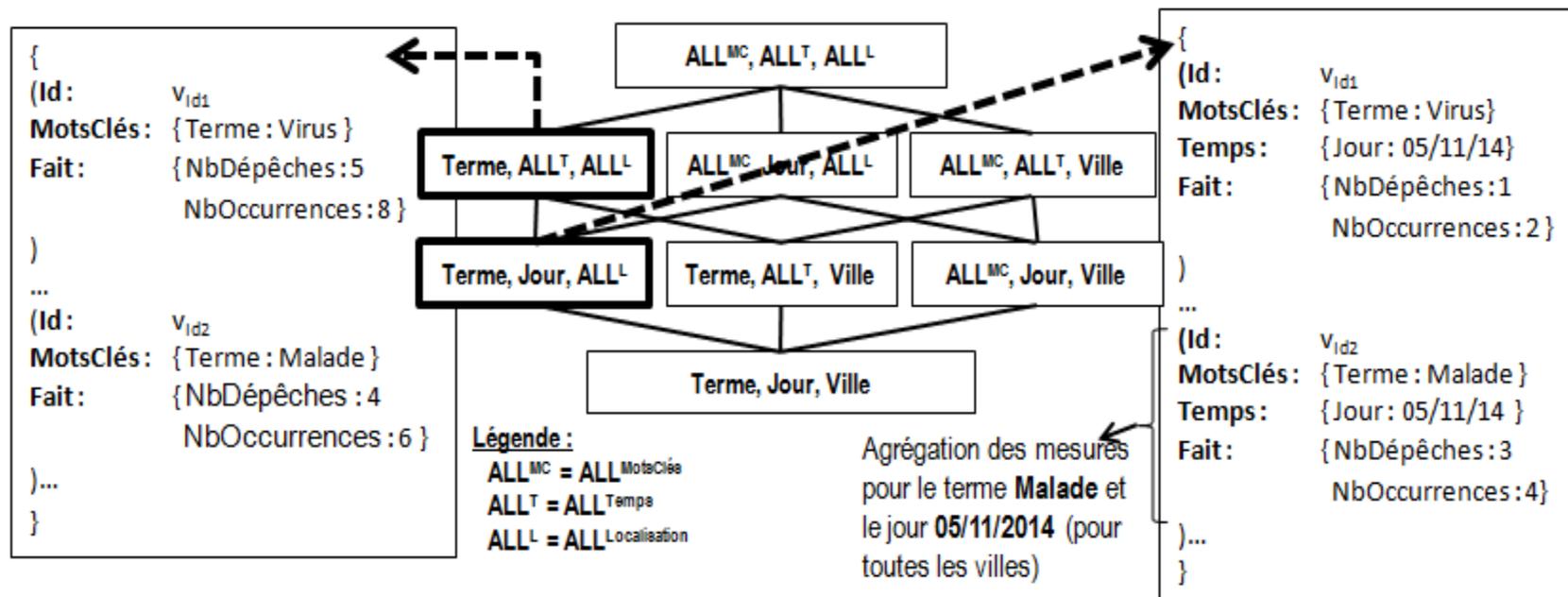
Transposer ce mécanisme dans les deux modèles NoSQL

Treillis d'agrégats dans le NoSQL



Treillis simplifié en modèle orienté colonnes

Treillis d'agrégats dans le NoSQL



Treillis simplifié en modèle orienté documents

Les Agrégats sont fait sur le modèle

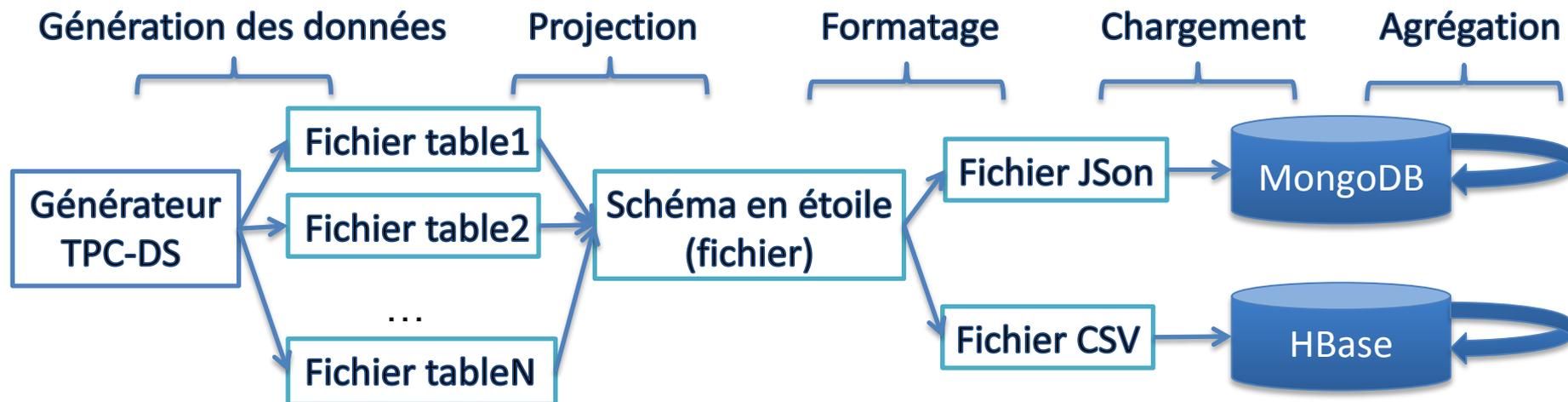
Motivation: valider faisabilité

Environnement: Cluster : 3 Nœuds (3 PC de 8Go RAM)

Logiciels: HBase, MongoDB

Jeu de données: TPC-DS

Expérimentations



Résultats

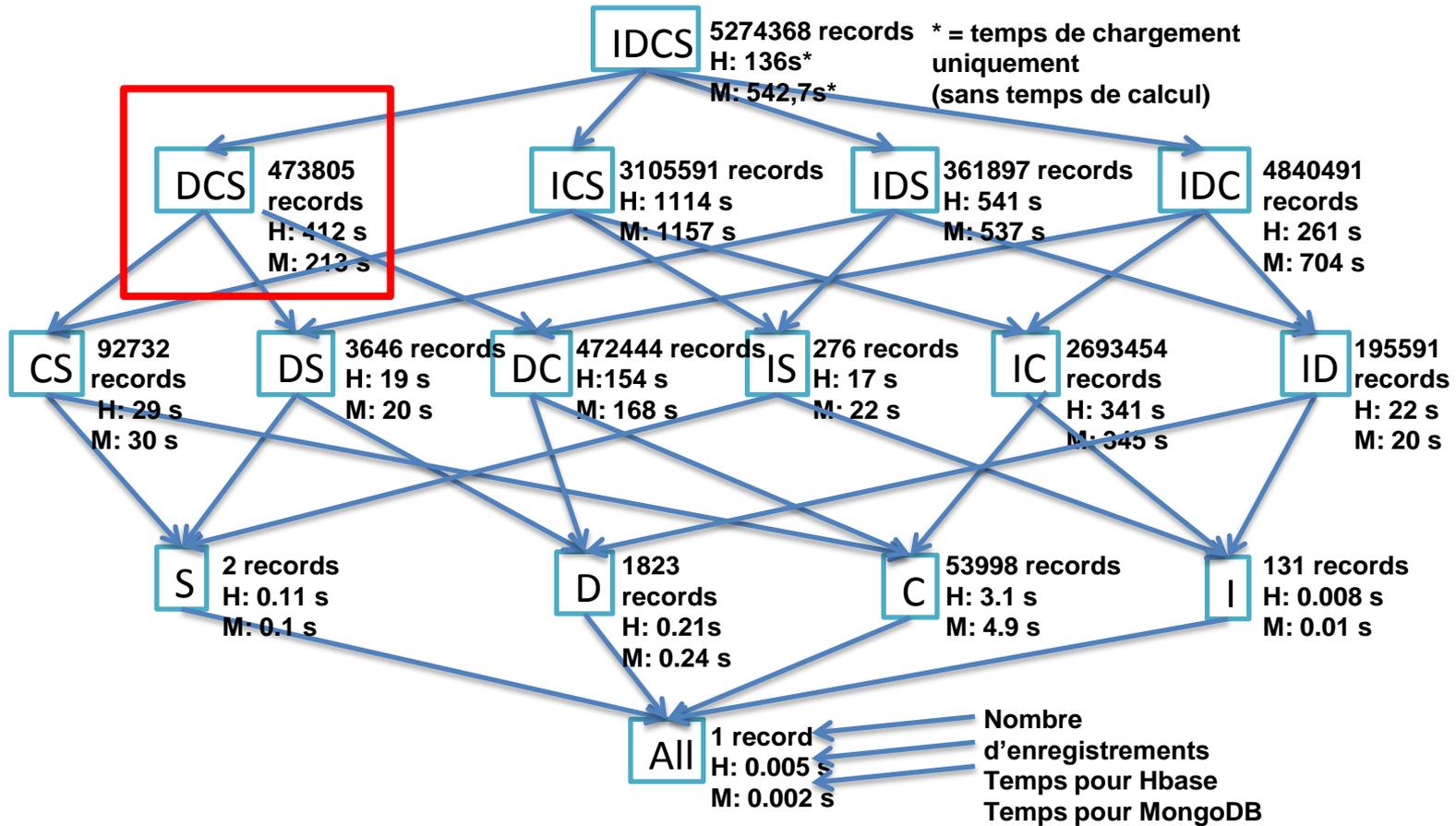
Volume des données	1GB	10GB	100GB
MongoDB	9.045m	109.58m	132m
HBase	2.26m	2.078m	10,3m

MongoDB: Création d'indexes

HBASE: Données déjà distribuées

Format de fichier différent(Json, Csv)

Treillis d'agrégats dans le NoSQL



Conclusion

- **Bilan**

Développer des processus de transformations dans le NoSQL colonnes et documents

Considération du treillis d'agrégats dans les transformations

- **Perspectives:**

Transformation alternative intra-modèle NoSQL

Migration inter-modèles NoSQL

Benchmark adapté au NoSQL (RCIS 2015)

MERCI