

---

## Data schema does matter, even in NoSQL Systems!

### *De l'importance d'un schéma de données dans les systèmes NoSQL*

**Paola Gómez<sup>1</sup>, Rubby Casallas<sup>2</sup>, Claudia Roncancio<sup>1</sup>**

1. *LIG, Université Grenoble Alpes*  
{paola.gomez-barreto,claudia.roncancio}@imag.fr

2. *TICSw, Universidad de los Andes, Bogotá, Colombia*  
rcasalla@uniandes.edu.co

Article accepté et présenté à la conférence internationale RCIS, co-localisée avec INFORSID 2016 à Grenoble. La version longue de l'article, en anglais, est disponible dans les actes de RCIS.

---

*RESUME. La plupart de systèmes « NoSQL » n'utilisent pas de schéma de bases de données. Ces systèmes n'offrent donc pas de fonctions de gestion de schéma et la définition de la gestion des structures de données utilisées reviennent à l'application. Les données peuvent être semi-structurées et l'utilisateur a de nombreuses possibilités de structuration.*

*L'absence de schéma prédéfinie présente certainement des avantages de flexibilité. Cependant, ceci a aussi un coût en termes de volume des données stockées, de performances des requêtes, de clarté du code applicatif et, a fortiori, la mise au point et la maintenance des applications. Ceci est démontré dans cet article où nous présentons une étude de l'impact de la structuration des données dans la base de documents MongoDB.*

*Nous avons expérimenté avec une variété d'alternatives de structuration et une série de requêtes avec complexité croissante. Cet article présente notre analyse et conclusions suite à cette expérimentation.*

*ABSTRACT. A Schema-less NoSQL system refers to solutions where users do not declare a database schema and, in fact, its management is moved to the application code. This paper presents a study that allows us to evaluate, to some extent, the data structuring impact.*

*The decision of how to structure data in semi-structured databases has an enormous impact on data size, query performance and readability of the code, which influences software debugging and maintainability. This paper presents an experiment performed using MongoDB along with several alternatives of data structuring and a set of queries having*

*increasing complexity. This paper introduces an analysis regarding the findings of such an experiment.*

*MOTS-CLES : NoSQL, systèmes orientés document, MongoDB.*

*KEYWORDS: NoSQL, schema-less systems, document-oriented systems, MongoDB.*

---

## **Extended abstract (non numéroté) : résumé anglais long**

☛ Les revues *European Journal of Electrical Engineering (EJEE)*, *Revue des Composites et des matériaux avancés (RCMA)*, *Traitement du signal (TS)* demandent un extended abstract d'environ une page avant l'introduction.

### **1. Introduction AR\_Inter1**

L'ensemble du texte est composé en police times new roman (TNR) 10, interligné simple, sans césure dans le texte, justifié. Chaque paragraphe commence par un alinéa de 0,5 cm et est suivi d'un espace de 6 pts : AR\_texte.

Les différentes sections sont numérotées de l'introduction jusqu'à la conclusion. Les remerciements et la bibliographie (ainsi que l'extended abstract) ne sont pas numérotés. Les intertitres sont alignés à gauche sans alinéa, comme suit. Les espaces au-dessus s'annulent quand ils sont précédés par un autre inter.

### **3. Intertitre de niveau 1 ; 18 pts avant, 12 après AR\_Inter1**

#### **3.1. Intertitre de niveau 2 : 18 pts avant, 12 après AR\_Inter2**

##### **3.1.1. Intertitre de niveau 3 ; 12 pts avant, 6 après AR\_Inter3**

##### **3.1.1.1. Intertitre de niveau 4 ; 6 pts avant 6 après AR\_Inter4**

### **2. Particularités AR\_Inter1**

#### **2.1. Énumérations AR\_Inter2**

Les énumérations sont introduites par des tirets, de différentes tailles s'il existe plusieurs niveaux ; chaque terme de l'énumération commence par une minuscule et se termine par un point-virgule, ou une virgule quand les termes de l'énumération sont courts, exemple :

- les auteurs utilisent... AR\_enumeration1 ;
- les auteurs utiliseront uniquement des tirets AR\_enumeration2.

L'espace entre les paragraphes de l'énumération est de 3 pts.

#### **2.2. Les notes de bas de page**

Voir modèle ci-dessous <sup>1</sup>

---

1. Exemple de note de bas de page. Times new roman, corps 9, interligne simple, sans alinéa, non exposant, justifié. Filet retrait droit 9,5 cm, 2 pt avant et après le filet. AR\_notesbp

### 2.3. Figures, tableaux, algorithmes AR\_Inter2

Les figures, illustrations et tableaux (préférer les niveaux de gris et les filets supérieurs à 0,5) sont incorporés dans le texte, centrés dans la largeur de page et numérotés de 1 à n dans l'article. Des légendes explicites les accompagnent, composées en TNR italique 10 et centrées.



Figure 1. Légende figure sous la figure AR\_legendefig (12 pts avant, 24 après)

Tableau 1. Légende au-dessus du tableau AR\_legendetab (12 pts avant, 12 après)

Marges et format (en cm) AR_tableau (TNR 9)			
Haut	Bas	Gauche	Droit
5,35	5,35	4,5	4,5
Disposition des en-têtes (en cm)			
En-tête : 4,4		Pied de page : 4,4	
Times new roman 9, centré		Times new roman 9, centré	

Algorithme 1. Légende algorithme idem légende tableau

---

```

1: generateurRepresentationEtatsActions (ObjectifsScenario, ObjetsFixes, Carte,
DescriptionZone,
2:           DescriptionGroupeUnites, ComportementsHautNiveau, Joueurs)
3: {
4:   PointsTactiques ← ∅
5:   ZonesStrategiques ← ∅
6:   pour tout  $o_i \in ( \text{ObjectifsScenario} \cup \text{ObjetsFixes} \cup \text{Carte} )$  faire
7:     si estCaseImportante(  $o_i$  ) alors
8:       PointsTactiques ← PointsTactiques  $\cup \{o_i\}$ 
9:       ZonesStrategiques ← ZonesStrategiques  $\cup \{\{o_i\}\}$ 
10:    fin condition
11:  fin boucle

```

---

### 2.4. Citations AR\_Inter2

Pour mettre en exergue du texte une citation d'auteur. En Times new roman 10 italique indenté à gauche de 1 cm. AR\_citation.

### 3. Équations AR\_Inter1

Les équations sont numérotées de 1 à n, en times new roman 10, interligne simple, 6 pts avant, 12 après. Avec une tabulation centrée pour commencer et une tabulation à droite pour le numéro entre parenthèses. AR\_équation. ex :

$$P_r(\sigma_r) = 1 - \exp \left[ - \left( \frac{\sigma_r}{\sigma_0} \right)^m \right] \quad (1)$$

$$P_r(x) = \frac{i}{n+1} \quad (2)$$

Remerciements AR\_Inter3 : inter de niveau 3 non numéroté

Le texte des remerciements se place après la conclusion, retrait gauche de 0,3 cm. Espace après 6 pts. AR\_merci

### Bibliographie non numérotée et références

Les références bibliographiques s'écrivent par ordre alphabétique, en times new roman 9, interligne simple, espacées de 6 points. Justifiées, avec un alinéa négatif de première ligne de 5 mm. AR\_bibli. Les références sont introduites dans le texte comme suit :

- (Dupont, 2000) pour 1 auteur ;
- (Dupont et Durand, 2001) pour 2 auteurs ;
- (Dupont *et al.*, 2002) pour 3 auteurs et plus ;
- 2 citations des mêmes auteurs : (Dupont *et al.*, 2001 ; 2011) ;
- 2 citations du même auteur, même date : (Durant, 1990a ; 1990b) ;
- 2 citations d'auteurs différents : (Dupont, 2008 ; Durand, 1991) ;
- Citation avec postfixe, par ex. la page : (Dupont *et al.*, 1989, 20) ;
- Citation avec préfixe, par ex. : (voir Dupont *et al.*, 1991).

Selon les phrases, on aura :

- On peut citer la méthode de Durant et Dupont (1991) qui permet de baser ...
- Durant et Dupont (1991) ont proposé une nouvelle méthode basée sur ...
- La recherche porte sur une approche multicritère (cf Durant et Dupont, 1991).

*Pour les ouvrages, thèses, rapports internes et articles en ligne :*  
titre en italique, le reste en romain

[Modèle ouvrage]

Vial S. (2010). *Court traité du design*, PUF, Paris.

[Modèle thèse, mémoire]

Amato É.-A. (2008). *Le jeu vidéo comme dispositif d'instanciation. Du phénomène ludique aux avatars en réseau*. Thèse en Sciences de l'Information et de la Communication, Université Paris 8.

[Modèle rapport interne]

Dupont J., Durand G., (2012). *Titre du Rapport interne*. Rapport de recherche Lavoisier n° 55, mars 2012.

[Modèle article en ligne]

Soegaard M. (2008). *Interaction Styles*, [http://www.interaction-design.org/encyclopedia/interaction\\_styles.html](http://www.interaction-design.org/encyclopedia/interaction_styles.html)

*Pour les articles de revues, ou issus d'actes de conférences, ou chapitres de livres : Titre de l'article en romain, Revue, Conférence, Livre en italique.*

[Modèle revue]

Pignier N. (2012). Le plaisir de l'interaction entre l'utilisateur et les objets TIC numériques. *Revue Interfaces numériques*, vol. 1, n° 1, p. 123-152.

[Modèle colloque, conférence]

Ouellet P. (2001). Sémiotique de l'empathie. L'expérience esthétique de l'autre, *Actes du colloque Sémio 2001*, Pulim, Limoges.

[Modèle chapitre de livre]

Dupont R. (2009). Titre du chapitre. *Titre de l'ouvrage*. Paris, Lavoisier, p. 53-68.

Article reçu le : [AR\\_1religne\\_soumission](#)

Article accepté le : [AR\\_soumission](#)

**Annexe(s) éventuellement [AR\\_Inter1](#)**

***Annexe 1. S'il y en a plusieurs [AR\\_Inter2](#)***

***Biographie ([AR\\_Inter2](#))***

☛ Les revues *Traitement du signal (TS) et Techniques et sciences informatiques (TSI)* demandent à leurs auteurs une biographie courte : 5 lignes maximum, avec leurs domaines d'étude et de recherche

Robert Dupont est chercheur en ... Il est actuellement en charge de au sein de l'équipe du laboratoire. Etc. [AR\\_biblio](#).