

# Stage M2R CEDRIC

## Approche de dénormalisation de schémas pour implantation NoSql

[Faten Atigui \(ISID\)](#), [Nicolas Travers \(Vertigo\)](#)

**Mots-clés : Dénormalisation de schéma, fusion de schémas, modèle de coût,  
systèmes NoSql**

### **Description :**

Les systèmes d'information doivent faire face à une quantité toujours plus grande de données et cherche à prendre en compte toutes les dimensions de leur éco-système afin de répondre aux exigences du métier. Ce volume toujours plus grand, complexe et dynamique (connu comme les 3V) a mis à mal les techniques traditionnelles de bases de données relationnelles et les entrepôts de données. Ainsi, pour des problèmes de passage à l'échelle, les bases de données NoSQL (HBase<sup>1</sup>, Cassandra<sup>2</sup>, MongoDB<sup>3</sup>, Néo4J<sup>4</sup>, etc.) ont vu le jour depuis une dizaine d'années et tentent de répondre à ces besoins. De nouvelles solutions sont proposées chaque année en vue de cibler une optimisation particulière, toutefois, ces fonctionnalités restent ad-hoc.

En conséquence, le choix de la bonne solution NoSQL en fonction des besoins métiers est fondamentale pour le système d'information. Il peut avoir d'énormes impacts sur le passage à l'échelle et la pérennité de la solution. Ce choix implique une connaissance précise du besoin, en termes de volumes et dynamique des données, d'interrogations et de contraintes sur le système. De plus, être capable de faire la corrélation entre les besoins et les solutions demande une réelle expertise sur le marché de la *Data*, ce qui très souvent s'avère donner une orientation commerciale plutôt que qualitative.

Le but de ce stage est donc de proposer un outil d'aide au choix d'orientation technologique et de conception d'un système d'information, en reposant sur une méthodologie de modélisation des données, simulation de distribution des données et un modèle de coût adaptatif (pour intégrer de nouvelles fonctionnalités NoSQL).

Ainsi, ce sujet s'inscrit à la fois à un niveau d'abstraction pour la modélisation de SI, mais également à un niveau physique pour favoriser l'optimisation du système. C'est deux approches traditionnellement séparées (architecture ANSI-SPARC) se doivent d'interagir finement pour s'adapter à un contexte d'optimisation extrêmement contraint et complexe.

### **Sujet du stage :**

---

<sup>1</sup> <https://hbase.apache.org/>

<sup>2</sup> <http://cassandra.apache.org/>

<sup>3</sup> <https://mongodb.com>

<sup>4</sup> <https://neo4j.com/>

Le stage débutera par l'étude des méthodes de dénormalisation de schéma conceptuels (entités/associations ou diagrammes de classes UML) pour le NoSQL de la littérature [1,2,3,4], ainsi que les différentes solutions NoSQL existantes [5,6,7] pour comprendre les fonctionnalités spécifiques de chacune. Ensuite, l'objectif de ce stage est de :

1. Proposer une approche de dénormalisation de schéma conceptuel en reposant sur les besoins métiers (application et requêtes). La spécificité de l'approche sera de fournir plusieurs solutions dénormalisées candidates ;
2. Proposer un modèle de coût générique pour chaque fonctionnalité offerte par les solutions NoSQL, celui-ci permettra d'estimer le coût d'une approche ;
3. En reposant sur les statistiques et la dynamique des données et des requêtes, notre approche évaluera les différentes solutions envisagées sur différentes implantations systèmes (HBase, Cassandra, MongoDB, Néo4J) grâce au modèle de coût.

La finalité est de proposer des orientations d'implémentation pondérées, tout en donnant les avantages et les inconvénients de chaque solution envisagée.

## **Bibliographie**

[1] Abdelhedi F, Ait Brahim A, Atigui, Gilles Zurfluh. MDA-Based Approach for NoSQL Databases Modelling. DaWaK 2017: 88-102.

[2] A. Chebotko, A. Kashlev, S. Lu. A Big Data Modeling Methodology for Apache Cassandra. IEEE International Congress on Big Data (BigData Congress). 2015

[3] Robert T. Mason. NoSQL Databases and Data Modeling Techniques for a Document-oriented NoSQL Database. Proceedings of Informing Science & IT Education Conference (InSITE) 2015, 259-268.

[4] DE FREITAS, Myller Claudino, SOUZA, Damires Yluska, et SALGADO, Ana Carolina. Conceptual Mappings to Convert Relational into NoSQL Databases. In : ICEIS (1). 2016. p. 174-181.

[5] RAUT, A. B. NOSQL Database and Its Comparison with RDBMS. International Journal of Computational Intelligence Research, 2017, vol. 13, no 7, p. 1645-1651.

[6] R. Fournier S'niehotta, P. Rigaux, N. Travers. Bases de données documentaires et distribuées. CNAM, 2017 <http://b3d.bdpedia.fr/>

[7] N. Travers. Maîtrisez les bases de données NoSQL. OpenClassrooms, 2017