

NewSQL la nouvelles génération des systèmes de gestion de bases de données

Elhadj Oumar Barry

29 décembre

1 Introduction

Mon choix pour les systèmes de gestion de bases de données vu par rapport à l'accroissement d'un point de vue architecture et opérationnel. Il est dû à l'évolution rapide et efficace. Des systèmes informatiques. C'est gestion de données est partie d'une architecture hiérarchique vers une conception relationnelle. Ce pendant nous retiendront que depuis l'évolution du Big Data et en plus bien que le No SQL, répond effectivement aux nouvelles envies des entreprises. Cela a changé les besoins des entreprises qui selon eux ne peuvent pas se passer de la transaction ACID. L'arrivée des systèmes non-relationnels a apporté des solutions par rapport à la distribution. Des données et à la disponibilité des informations. Puisque, sait le système qui manquait de certaines fonctionnalités qui nous proposent. C'est à cet effet, qu'une nouvelle génération des systèmes de gestion appelle New SQL à vue le jour.

2 Présentation Générale

2.1 Historique

La gestion de bases de données date dans les années soixante ou est apparue les premières bases de données « Hiérarchiques » dans lesquelles les informations sont établies par rapport à une arborescence. Et après les arborescences les bases de données réseaux, sont apparues à leur tour qui consiste à établir ou de mettre en place des informations dans leur ensemble pour forme un éventuel lien entre eux [15].

2.2 Les avantages du NewSQL et du NoSQL

Le NewSQL, c'est une nouvelle classe de systèmes de gestion de données qui est en rapport à la fois avec l'architecture et les avantages de No SQL pour complète ou de remplir ses désavantages tout en appliquant l'éventuelle synonymie des données et des propriétés ACID de la méthode SQL.

En d'autres échéances l'architecture NewSQL :

- Il remplit les conditions de gestion de transaction à la difformité des propriétés ACID.
- IL Prend soins les données qui sont en mémoire pour affirmer la réactivité.
- NewSQL évité de faire des perturbations lors des opérations de NoSQL.
- NoSQL, manipuler aussi des grands données de manière qui est sûr et il données des meilleurs résultats optimal par nœud que d'autre solutions classiques comme le SGBDR.

• Et souvent le NoSQL est préfèrent être utilise par des langages SQL pour des requetés nous avons ensuite ici les différents points de NoSQL que nous allons essais d'aborde et de développe qui est aussi des avantages du NoSQL.

- Extension : dans ce cas, ici, NoSQL permet d'une extension horizontale, ce qui signifie qu'il est facile d'ajouter plus de serveurs (nœuds) au système afin d'offrir une grande disponibilité.

- Distribution de la charge : l'absence de quelle que schéma rigide permet au NoSQL de distribué la charge de travail entre les différents nœuds.

- Gestion de gros volume de données : NoSQL, est capable de gérer, de manière performante des volumes qui sont beaucoup plus volumineux et complexes de données quelle que soit leur nature comme : des données qui sont structurées, semi-structurées ou non structurées pour temps, c'est ce qui n'est pas le cas des bases de données SQL.

- Disponibilité continue : dans ce cas, la différente multitude des nœuds joue un rôle très important dans la disponibilité d'information sans faire aucune erreur.
- Coûts bas : ici on parle de la mise en place des réseaux de nœuds, dans le court bas les réseaux des nœuds reste abordable.
- Simplicité : en matière de déploiement, les modèles de données NoSQL restent plus simples que le modèle de données relationnelles.

2.3 Les inconvénients du No SQL

Dans les inconvénients du NoSQL, nous retiendrons que malgré l'élasticité, la rapidité, la disponibilité et de sa simplicité jusqu'à nos jours, il n'a pas réussi d'abord à offrir une bonne cohérence de données en matière de transaction et à celui de la propriété ACID. En effet, c'est ce qui a contribué à l'apparition du NewSQL.

2.4 Quatre types de bases de données NoSQL

Il existe quatre types de bases de données qui sont en parallèle à l'approche NoSQL. Nous avons :

1. Bases de données clé-valeur :

Dans cette base de données Clé-valeur, chaque enregistrement est composé d'une clé, qui est constituée en fonction d'une opération appelée hachage et d'une valeur qu'on peut trouver grâce à l'opération de hachage, grâce à cette approche qui garantit la notion de colonnes et de tables inexistante, mais néanmoins qui garantit les quatre types d'opérations que peut connaître une base de données qui est : création, lecture, modification, et suppression [17, 12].

2. Bases de données orientées documents :

Ici, il s'agit des données qui sont bien organisées dans les collections de telles sortes qu'on n'a pas besoin d'avoir de schéma qui nous indique les colonnes. Et sans oublier aussi que cette approche utilise des clés et des valeurs, mais qui est mis dans un enregistrement appelé document. Et les formats les plus utilisés sont celui de JSON et de XML. Dans ses formats chaque enregistrement peut contenir plus ou moins de paires de clés-valeur que celui des documents et d'autres [17].

3. Bases de données orientées colonnes :

C'est une base de données où l'absence des schémas, permet un accroissement facile et horizontal de la base de données qui sont enregistrées par famille de colonnes et les nombres de colonnes qui sont disponibles est différents d'un enregistrement de l'une de l'autre [17, 10].

4. Bases de données orientées Graphes :

Cette base de données est basée sur des théories des graphes, et ce sont des bases de données qui sont similaires aux bases de données relationnelles. Et ils sont regroupés dans des ensembles de nœuds et cette approche est souvent utilisée pour la gestion de données spatiales, sociales ou financières [17, 12, 10].

2.5 L'incommodités du NoSQL

Bien que nous connaissons la capacité de la rapidité, élasticité, la simplicité, la disponibilité de l'information du No SQL, à ce jour, il n'a pas réussi de donner une bonne cohésion en matière de gestion des transactions et aux propriétés ACID. C'est grâce à ce problème de cohérence qui a contribué à l'apparition du No SQL. On doit se poser la question de savoir, qu'est-ce que le New SQL? Et de la fonctionnalité clé des systèmes New SQL? Dans les paragraphes suivants, nous essayerons de répondre à ses questions.

3 Qu'est-ce que le New SQL?

3.1 Le New SQL

Depuis, arrive de ce groupe de recherche nommé « 451 Research » ils ont enfin proposé le concept NewSQL en 2011 comme une succession à ces prédécesseurs. Qui a pour objectif de donner une idée de ce que l'approche NewSQL ainsi que les différences catégories de cette base de données qui la compose

Qu'est-ce que le New SQL?

C'est une suite logique qui combine à la fois les avantages qui est apporté par la technologie NoSQL tout en ne pas changeant la propriété ACID du système SQL que le NoSQL n'arrive pas à offrir.

En d'autre sur constance, la technologie NewSQL :

Dans la technologie, le NoSQL prend en charges la gestion de transition par les propriétés ACID. Le NoSQL, gères les données qui se trouve dans la mémoire pour garantir la réactivité. Il permet encore de jouer avec les grands volumes de données de manière solide et efficace et avec une capacité de distribution ainsi qu'une adaptation métrise de nouveaux nœuds de données [13].

Ce tableaux consiste a utilisé pour le langage SQL pour les requêtes

Fonctionnalité	SQL	NoSQL	NewSQL
----------------	-----	-------	--------

Relationnel	Oui	Non	Oui
-------------	-----	-----	-----

ACID	Oui	Non	Oui
------	-----	-----	-----

Language SQL	Oui	Non	Oui
--------------	-----	-----	-----

Extensible	Non	Oui	Oui
------------	-----	-----	-----

Complexité des reuetes	Bah	Haut	Très haut
------------------------	-----	------	-----------

Distribué	Non	Oui	Oui
-----------	-----	-----	-----

3.2 Fonctionnalités clés des systèmes NewSQL

Dans cette partie, nous allons découvrir quelques fonctionnalités importantes dans l'approche NewSQL ainsi que quelque retouche ajoutée qui a vraiment jouer un rôle très important dans l'accroissement la performance, la disponibilité et le rapport des bases de données. Nous avons : la Stockage en mémoire, Partitionnement/Sharding, Contrôle de la concurrence, Index secondaires, Réplication, Restauration de données

3.3 Stockage en mémoire

De nos jours, le stockage en mémoire est devenu abordable pour gérer des quantités importantes de données, la majorité des systèmes qui existe gèrent des données par lecture et part l'écriture depuis vers des disques de stockages, mais au paravent celle-ci était coûteuse, et maintenant qu'il est abordable pour pouvoir gerber la qualité des données à l'exception des grands volumes. Ceci uniquement de garantir une grande disponibilité de l'information vers la rentrée et la sortie [20]. Et sans oublier qu'il y a d'autres systèmes qui utilisent cette approche le plus souvent pour sauvegardes la base de données pour uniquement d'éviter tous qui est incident qui peut provoquer une perte de données en cas d'une coupure d'électricité ou d'une panne. Et nous avons aussi d'autres entreprises qui mettent en avant cette fonctionnalité du système NewSQL comme le « Volt DB [8] et SAP HANA [9] » pour le but de gagner une popularité

3.4 Partitionnement/Sharding

Ce partitionnement consiste à faire beaucoup de choses comme par exemple, il sert à optimiser la lecture et l'écriture de données, et entre temps un système intermédiaire a été mis en place pour assurer l'indexation des schards et aux données qui les contiens, ce qui permet à l'accélération des étapes ou processus de la lecture et de l'écriture. Et ce partitionnement sert à découper la base de données en plusieurs couches de données. Et cet ensemble forme un cluster, chaque couche utiliser les mêmes systèmes de gestion de données, mais ce système ne s'occupe d'une celle partie des enregistrements appelée schards.

3.5 Contrôle de la concurrence

Ici, la qualité de système de l'information consiste à une grande partir dans sa capacité de gérer des requêtes d'un grand volume concurrentiel pour assurer une grande partie d'utilisation d'une application. Bien qu'il a une forte utilisation d'une grande partir de la public, chaque utilisateur se sent le seul à utiliser ce système [20]. Il existe un principe appelle verrouillage en deux phases, ce processus consiste à verrouiller l'enregistrement concernant une transaction qui s'effectuait en cours et aucune modification soit écrite n'est interrompus par celui-ci tant que la transaction n'est pas (to end). Et les systèmes NewSQL évités l'utilisation des deux principes de verrouillage en deux phases appelé (2 PL ou 2 Phase Locking). Et il existe une approche qui se basse sur le principe du versioning appelle le contrôle concurrentiel multi version, il est le seul qui exécute les opérations qui concernent les données tout en gardant un suivi de modifications sous le format de versions. Et cette opération permet d'éviter les latences en lecture et en même temps, il permet d'augmenter la performance du système [20].

3.6 Index secondaires

Comment les index sont utilisés? Ici, en se basant dans un système de gestion de bases de données, les index sont souvent utilisés pour accroître la performance dans le but de s'appliquer sur les clés primaires surtout dans un environnement distribué. Pourtant, l'utilisation d'indexation classique semble inefficace dans le sens où les nœuds sont partitionnés dans les différentes bases d'une clé primaire. Le plus souvent pour offrir plus de caractère quand il s'agit de l'utilisation du reverse lookup, les index secondaires s'appliquent sur les champs appelés non-primaires. Cependant, dans les systèmes distribués, il est recommandé de veiller que cette indexation soit partagée sur chaque cluster. En plus de ça de la maintenance de ses index, il peut vite s'avérer être complète en cas où il s'agit d'un grand volume de données. Pour résoudre ce problème de veiller à cette indexation soit pour partager sur chaque cluster, il nous a fallu l'apparition d'un nouveau système de NewSQL qui est adapté à une architecture décentralisée qui vise à fractionner, c'est stocker des index de chaque bloc dans des nœuds qui lui est concerné [20].

3.7 Réplication

Afin, de garantir un rapport irrécusable des informations qui sont transmises par les différents nœuds qui les constituent il s'agira tout simplement de mettre en rapport le résultat d'une transaction d'écriture à travers différents nœuds avant de la considérer comme valide, et nous verrons que de cette manière la lecture de données assure un très bon rapport entre les informations qui sont communiquées vers l'application [20].

3.8 Restauration de données

Si les points centralisés se basent sur la restauration créée et sur l'application des changements enregistrés dans les fichiers logs. De là, le système qui désigne un nœud appelé « Master ». Assure la synchronisation des données vers les autres nœuds à travers la réplication « slaves ».

NB : si les nœuds principaux arrêtent de fonctionner il y aura directement un mécanisme de restauration se déclenche et qui consiste à désigner un autre nœud secondaire.

NB : pour maintenir de la garantie de la disponibilité de l'information. Il faut qu'une fois les nœuds finalisent sa restauration, il devient directement un nœud secondaire et prend la formule d'un nouveau master [20].

3.9 Classification des systèmes New SQL

Dans cette classification, nous distinguons deux à trois grandes familles de système de gestion de bases de données qui sont : les routeurs de données, le service (DBaaS).

3.10 Routeurs de données

Les routeurs de données aussi appelle (sharding Middlewares), est un peu similaires au système qui a été utilisé dans les années 2000 par quelque grande entreprise beaucoup plus reconnu comme Google, Facebook, eBay, etc. Il existe aussi d'autre entreprise qui à travers le sharding de couper les bases de données en petit partir appelés « shards » pour après les stocker dans l'architecture [20].

3.11 Bases de données en tant que service (DBaaS)

C'est un système qui proposés en tant que service à travers la plateforme de cloud computing. Souvent, c'est systèmes sont déployés sur des machines virtuelles et qui est mise en possession a des entreprises pour maintenir leur maintenance.

4 Conclusion

Nous avons vu que les différents personnalités apportées par la technologie New SQL ont permis à d'autre secteur d'activité qui veiller fonctionnalise NoSQL tout en maintenant les propriétés ACID qui est le cas du secteur financier, Et étant donné que certain des familles des systèmes New SQL gagnent en popularité chaque jour et cela peut permettre de d'éclanche une série de défis que ces systèmes vont d'avoir résoudre tout en suivent les processus d'avancement de la technologie et ses aussi la manière de voir la capacité à répondre aux besoin de changement qui peut permettre la suivie de ce système .

Enfin, c'est en réalisant ce travail qui m'a permis d'enrichir mes connaissances en système de gestion de bases de données.

5 Acronymes

SGBD : Système de Gestion de Bases de Données

SQL : Structured Query Language

NoSQL : Not Only SQL

ACID : Atomicity Consistency Isolation Durability

JSON : JavaScript Object Notation

XML : Extensible Markup Language

2PL : 2 Phase Locking

DBaaS : Database as a Service

OLTP : Online Transaction Processing [article]

6 Références

- [8] VoltDB. URL : <https://www.voltdb.com/>.
- [9] What is SAP HANA database? | Architecture Benets.
URL : <https://www.sap.com/products/hana/what-is-sap-hana.html#overview>.
- [10] Qu'est-ce qu'une base NoSQL? Les cas Datastax et MongoDB, February 2016. URL : <https://www.digora.com/fr/blog/definition-base-nosql-datastax-mongodb>.
- [12] Ayoub Bilem and Walid Lahcen Sakhi. Etude-comparative-des-bases-dedonnees-SQL.pdf.
URL : <http://dspace.univ-tlemcen.dz/bitstream/112/11962/1/Etude-comparative-des-bases-de-donnees-SQL.pdf>.
- [15] Claude Chrisment. Bases de données relationnelles. page 16.
- [20] Andrew Pavlo and Matthew Aslett. What's Really New with NewSQL? ACM SIGMOD Record, 45(2) :4555, September 2016. doi :10.1145/3003665.3003674.