مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (41) العدد (1) 2019
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (41) No. (1) 2019

Studying New Classification Standards for Non-Relational Database Systems in Managing Big Data

Dr. Kamal Abed Al Hadi Saloum*
Hani Mohamad Hasan**

(Received 22 / 11 / 2018. Accepted 27 / 12 / 2018)

□ ABSTRACT □

According to the development of internet applications and the huge increase in the produced volume of data daily, the traditional relational database systems showed weakness in the efficiency of managing this data, because of the restrictions imposed by the design model connected to it. Recently, non-relational database appeared to solve this problem without depending on all predesigned structures and reflecting a flexible model that provides successful solution for managing big databases. In short time, the number of these systems has largely developed both qualitatively and quantitatively, and become a global phenomenon in managing database that preceded the appearance of related studies, something that research specialists have recently noticed. As result of the abundance in marketing resources of these systems, carrying out analytical and design studies of its work becomes a necessity. According ,these studies become a user guide for suitable models and types in managing huge data depending on their several classification properties, They also provide an idea about the mechanism of these systems and the techniques used. This issue was discussed in the research.

Keywords: Data, Relational, non-Relational, Structures, Models, classification properties.

***Professor in software engineering and Information systems department ,Faculty of information technology, AlBaath University.**
****Master in software engineering and Information systems, PhD Student In software engineering and Information systems Department, Faculty of information technology, AlBaath University.**

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (41) العدد (1) 2019
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (41) No. (1) 2019

دراسة معايير تصنيفية جديدة لأنظمة قواعد المعطيات غير العلائقية في إدارة البيانات الضخمة

د. كمال عبد الهادي السلوم*

هاني محمد حسن**

(تاريخ الإيداع 22 / 11 / 2018. قُبِل للنشر في 27/ 12 / 2019)

□ ملخّص □

مع تطور تطبيقات الانترنيت والزيادة الكبيرة في حجم البيانات المنتجة يومياً، أظهرت قواعد البيانات التقليدية العلائقية ضعفاً في المقدرة على الإدارة الفعالة لهذه البيانات نتيجة القيود التي يفرضها نموذج التصميم المرتبط بها، فظهرت حديثاً قواعد المعطيات غير العلائقية لتجد حلاً لهذه المشكلة دون الاعتماد على البنى والهيكليات المسبقة التصميم، وعاكسةً لنموذج مرن قدم حلولاً ناجحة لإدارة قواعد البيانات الضخمة، وتطورت هذه الأنظمة كماً ونوعاً خلال فترة زمنية قصيرة، وأصبحت ظاهرة عالمية في إدارة قواعد البيانات سبقت ظهور الدراسات البحثية في هذا المجال، الأمر الذي تنبه له مؤخراً الباحثون الاختصاصيون في هذا المجال، ومع وفرة الموارد التسويقية من هذه الأنظمة بات من الضروري الخوض في دراسات تحليلية وتصميمية شاملة لعملها، لتشكل دليلاً للمستخدمين نحو النماذج والأصناف المناسبة لإدارة بياناتهم الضخمة اعتماداً على خصائص تصنيفية عديدة، وتقديم فكرة عن آلية عمل هذه الأنظمة والتقنيات المستخدمة فيها وهو ما تم التطرق إليه في هذا البحث.

الكلمات المفتاحية: البيانات، العلائقية، غير العلائقية، الهيكليات، النماذج، خصائص تصنيفية.

* أستاذ – قسم هندسة البرمجيات ونظم المعلومات – كلية الهندسة المعلوماتية – جامعة البعث.

** ماجستير هندسة برمجيات ونظم المعلومات– طالب دكتوراه في قسم هندسة البرمجيات ونظم المعلومات – كلية الهندسة المعلوماتية – جامعة البعث.

مقدمة:

بات احتكار الشركات الكبرى المصنعة لأنظمة إدارة قواعد المعطيات العلائقية التقليدية يشكل هاجساً للعاملين والباحثين في هذا الحقل نظراً للتكاليف العالية لهذه الأنظمة ومحدودية أدائها عند نمو حجم البيانات بشكل كبير، لذا صار من الضروري التوجه للبحث عن بدائل حقيقية تتمتع بخاصتين مهمتين: أن تكون أنظمة قواعد المعطيات محمولة وأن تكون مجانية أو مفتوحة المصدر أو غير تجارية، ومن هنا بدأت أنظمة قواعد المعطيات غير العلائقية (NoSQL Database Systems) تكتسب أهميتها ودورها ولاسيما التخزين للبيانات الضخمة في تطبيقات الانترنيت الرائجة، كما أن النمو المتسارع في أعداد ونوعية أنظمة قواعد المعطيات غير العلائقية جعل الحاجة ماسة لدراسات تحليلية وتصنيفية جديدة لاقتراح جملة من المعايير الشاملة تعتمد كدليل بحثي للمستخدمين كافة يساعد في إرشادهم إلى المناسب من أنظمة قواعد البيانات غير العلائقية وفقاً لمتطلبات عملهم سواء كانوا أفراد أم منظمات أم شركات كبرى تعمل في مجال أنظمة قواعد المعطيات.

أهمية البحث وأهدافه:

يأتي هذا البحث مع تنامي استخدام الشبكات الحاسوبية وزيادة كمية البيانات المنتجة، حيث تتعدى حجومها يومياً الملايين من الجيجابايت وباتت الحاجة ملحة لأنظمة قواعد المعطيات NoSQL غير العلائقية، لإدارة فعالة لهذه الكمية الضخمة من البيانات، حيث ازداد عدد هذه الأنظمة بشكل مضطرد وسريع و أضحت كل شركة أو منظمة تبني نظامها الخاص لإدارة قواعد بياناتها العملاقة دون معايير واضحة تؤطر عملها حيث بلغ عدد هذه المنظومات في عام 2013 ما يربو على 150 نظاماً، و تجاوزت حالياً عتبة 700 برنامج على المستوى العالمي، مما أسس لفوضى عارمة في نمو هذه الأنظمة بعيداً عن النمذجة والتصنيفات القائمة حالياً في الدراسات والأبحاث المعتمدة مما جعل الحاجة ماسة لدراسات تحليلية وتصنيفية جديدة لهذه الأنظمة و وضع معايير يتم الالتزام بها بالحد الأدنى لدى بناء نظام معطيات غير علائقي في محاولة ضمن هذا البحث لإيجاد اعتراف مبدئي للمعايير بهذه الأنظمة عند التزامها بحد أدنى من متطلبات وشروط بناء قواعد البيانات، فالتصنيفات المطروحة لا تلبي المطلوب ولا تؤطر كامل أنظمة قواعد المعطيات غير العلائقية المتوفرة حالياً، حيث تصنف البعض كلياً، والبعض الآخر جزئياً أما الباقي فخارج التصنيفات المقترحة، كما أن وفرة الموارد التسويقية من هذه الأنظمة يجعل من الصعب أن نعلم فيما إذا كان حل معين مناسب لحالة الاستخدام الخاصة بنا، لذا تم العمل من خلال هذا البحث على ذكر التصنيفات الموجودة حالياً لهذه الأنظمة غير العلائقية، وتبيان محدوديتها وعدم شموليتها من خلال دراسة نقدية لمعيار التصنيف المُعتمد فيها والمُستند على النموذج التصميمي (The Designed Model) المستخدم في هيكلة قاعدة البيانات وتخزينها واقتراح أصناف جديدة إضافةً إلى تلك التصنيفات المعروفة حالياً في الدراسات البحثية، وذلك بالاعتماد على النموذج التصميمي لقاعدة البيانات، بالإضافة إلى عرض معايير جديدة للتصنيف تعتمد على لغات الاستعلام و الاستخدام والانتماء لنظريات قواعد البيانات المعروفة (CAP,BASE,ACID) ومعايير تتعلق بمواصفات الاتساقية والموثوقية والثبات وقابلية التوسعة وغيرها، وتصنيف تلك الأنظمة اعتمادا على معيار الاستخدام على المستوى العالمي وذلك باقتراح عدة مستويات لتصنيف الاستخدام تبعاً لمؤشرات دولية معتمدة بهذا الخصوص، وبذلك نكون قد عرضنا مجموعة واسعة من المعايير الجديدة في تصنيف أنظمة قواعد البيانات غير العلائقية (NoSQL Database Systems)، تشكل دليلاً شاملاً لدى المستخدمين الراغبين باختيار واعتماد المناسب من هذه الأنظمة لإدارة بياناتهم الضخمة أو غير المهيكلة.

طرائق البحث ومواده:

قدمت أنظمة قواعد المعطيات غير العلائقية حلاً مثالياً لإدارة البيانات الكبيرة (Big Data)، وتصدت لمشكلة الإدارة غير الفعالة لأنظمة قواعد المعطيات العلائقية التقليدية للبيانات المخزنة في مثل هذه الحالات، وتعتمد أنظمة قواعد البيانات غير العلائقية على قاعدة مهمة: وهي أن حصول مستخدم قاعدة البيانات الضخمة على إجابة سريعة ودقيقة هو أكثر فائدة بالنسبة له من الحصول على إجابة دقيقة وبطيئة، بمعنى أنه في عالم إدارة قواعد البيانات الضخمة يمكن لنا أن نستغني لحظياً عن اتساقية (Consistency) البيانات مقابل أن نحصل على توافرية (Availability) عالية للنظام وبياناته، أي أن هذه الأنظمة تحرص على الاستجابة العالية لطلبات المستخدمين دون استخدام تقنيات قفل البيانات(Lock Data) كما هو الحال في أنظمة قواعد المعطيات العلائقية التي تفرض قيوداً صارمةً لا تسمح بموجبها سوى لعملية نفاذ وحيدة بالتعديل على محتوى قاعدة البيانات في وقت معين، بينما تنتظر بقية الطلبات على قائمة الانتظار حتى إتمام التعديل المناسب من قبل المستخدم الحالي للبيانات، مما يسبب تأخيراً زمنياً في الاستجابة لطلبات المستخدمين، أما في أنظمة قواعد المعطيات غير العلائقية فإن إدارة عمليات النفاذ للبيانات تتم من نظام قاعدة المعطيات بطريقة تحقق توافرية عالية للبيانات دون الاستغناء عن اتساقيتها و صحتها ولو بدا ذلك لفترة وجيزة زمنياً، لأن الموافقة الفعلية لتعديل البيانات لا ترسل من قبل النظام إلى المستخدمين إلا بعد تطبيقه للاتساقية الزمنية التي تكفل أنه بعد تنفيذ سلسلة من المناقلات المتزامنة فإن قاعدة البيانات تنتقل إلى حالة تؤكد صحة البيانات وتلتزم معايير الاتساقية، وهو ما يمنع حدوث أخطاء أثناء تعديل المستخدمين للبيانات وبنفس الوقت يُظهر النظام سرعة في الاستجابة لطلبات المستخدمين، وهذه المرونة هي غير متوفرة في أنظمة إدارة قواعد البيانات العلائقية بسبب القيود التي يفرضها نموذج التخزين المستخدم والمتعلق باستخدام الجداول ومفهوم المناقلات وقفل البيانات (Lock Data) جراء عملية اتصال أو تحديث معينة.

إن معظم الأبحاث التي عملت في مجال تصنيف أنظمة قواعد المعطيات غير العلائقية أجمعت على اعتماد نموذج البيانات المستخدم في التخزين كقاعدة رئيسية في عمليات التصنيف، فاستخدمت أربعة مجموعات رئيسية لتصنيف هذه الأنظمة [1] وهي:

1- قواعد البيانات مفتاح/ قيمة (Key-Value Database): وفيها يتم تخزين البيانات باستخدام المفهوم الحاسوبي للقاموس، وبالتالي قاعدة البيانات هي عبارة سلسلة مكونة من مفاتيح وقيم مرتبطة بها، حيث أن كل قيمة مخزنة يميزها رقم وحيد و غير مكرر على مستوى قاعدة البيانات، والقيم تمثل البيانات المخزنة من مختلف الأنماط، حيث يمكن أن تكون مستندات أو صور أو نص أو سلسلة محارف أو جدول وغيرها..وكل مفتاح يرتبط عبر مؤشر بالقيمة المخزنة التي يعبر عنها و التي يتم حفظها ضمن مصفوفات، و بالتالي البحث عن البيانات يتم حصراً عبر مفاتيحها، مما يعطي نموذج التخزين المستخدم بساطة كبيرة في التنجيز ويكسب قاعدة البيانات فعالية كبيرة في الأداء، ومن أشهر قواعد البيانات التي تعتمد هذا النموذج هي Redis.

2- قواعد البيانات المرتبطة بالوثائق (Document-Oriented Database): تمثل البيانات المخزنة بمجموعة كبيرة جداً من الوثائق المختلفة الحجوم والمحتوى والهيكلية والنمط حيث تتنوع بين وثائق JSON أو BSON وغيرها، ويتم تمييز كل مستند برقم وحيد مرتبط به وغير مكرر على مستوى قاعدة البيانات، مما يعطي مرونة في الاستعلامات بحيث لا نسترد كامل الوثيقة المخزنة و إنما الأجزاء التي تهمنا منها، ومن الأمثلة الشائعة على الأنظمة التي تستخدم نموذج الوثائق في التخزين هي MongoDB,CouchDB.

3- قواعد البيانات الموجهة بالأعمدة (Column-Oriented Database): وهي تسمى أيضاً BigTable وهي قريبة لمفهوم أنظمة قواعد المعطيات العلائقية من خلال استخدامها لمفهوم الجدول حيث تحفظ البيانات على شكل أعمدة تختلف في حجومها ومستوى الامتداد الشاقولي لتخزين البيانات فيها، وهذا ما يجعله مثالياً للتطبيقات التي تتعامل مع كميات ضخمة من البيانات والمخزنة على تنوع كبير من العناقيد (Clusters) لأن نموذج التخزين المستخدم يمكن أن يجزأ بشكل فعال على المخدمات، كما أنها أوجدت حلاً لحجز مساحة تخزين مسبقة لقيم ربما تكون فارغة أو خالية (Null Value) كما هو الحال في الجداول المستخدمة في قواعد البيانات العلائقية، ووفرت الكثير من عمليات المسح للصفوف أثناء البحث حيث حصرتها ضمن عمود معين لإيجاد القيمة المطلوبة، ومن أكثر الأنظمة شيوعاً في استخدام هذه الآلية في التخزين هي Cassandra ,HBase.

4- قواعد البيانات المعتمدة على الرسم البياني (Graph-Oriented Database): هو نموذج يستخدم التمثيل البياني لحفظ المعطيات مستخدماً مفاهيم العقد والحواف(الروابط) والخصائص التي تحدد وتميز علاقات الربط بين العقد لتشكيل مخططات ضخمة تحدد الارتباطات المعقدة بين البيانات المخزنة وهي مناسبة لتخزين بيانات تتعلق بالشبكات الاجتماعية وخرائط الطرق والخطوط الجوية لرحلات الطيران، ومن الأمثلة الشهيرة عن الأنظمة التي تستخدم هذه التقنية هي Neo4j,HyperGraphDB.

لا بد أن نشير هنا إلى أنه لا توجد لغة استعلام محددة مرتبطة بأنظمة قواعد البيانات غير العلائقية، فلكل نظام لغة الاستعلام المناسبة وفقاً لنموذج التخزين المُعتمد فيه، ولكن أغلب الاستعلامات تدعم استرجاع البيانات بالاستناد إلى قيمة معينة للمفتاح الذي يكون قد تم توليده أوتوماتيكياً أو مُخزن بشكل مسبق في البيانات، حيث يوجد هناك جهد لإنشاء لغة استعلام خاصة بهذه الأنظمة وهي قريبة من حيث الصياغة من لغة SQL ولكنها تعتمد على تنفيذ استعلامات على بيانات غير مهيكلة [2].

إن التعامل مع الأغراض المخزنة في قواعد البيانات يتطلب توفر آلية محددة تكفل صحة البيانات واتساقها لدى إجراء عدة عمليات نفاذ متزامنة من قبل المستخدمين على مخازن البيانات، حيث تنشط في هذا الجانب مجموعة من نظريات التخزين، و تقدم كل منها فلسفة معينة في التصميم لقاعدة المعطيات بما يكفل الحفاظ على سلامة البيانات ومن هذه النظريات:

1- **نظرية CAP:** حيث تضمن هذه الفلسفة في التصميم اثنان من الخصائص التالية فقط في نفس الوقت:

أ- الاتساقية: الحفاظ على صحة البيانات و سلامتها طوال تخزينها في قاعدة البيانات.

ب- التوافرية: كل الاستعلامات المنفذة من قبل المستخدمين تحصل على رد يوضح نجاح تنفيذ العملية على البيانات أو عدم نجاحها.

ج- تجزئة البيانات: لا يتوقف النظام عن العمل إلا إذا تعرض لعطل كامل، وفي الحالة الطبيعية وبالرغم وجود أعطال جزئية تبقى البيانات الموزعة فرعياً تعمل بشكل اعتيادي.

إن أنظمة قواعد البيانات التي تعتمد نموذج CAP تحقق خاصيتين على التوازي من الخصائص الثلاث المذكورة أعلاه.

2- **نظرية BASE:** تعتمد على الخصائص التالية:

أ- التوافرية : النظام متاح دائماً.

ب- أمان النظام: هذه الخاصية تدل أنه يمكن لحالة النظام أن تتغير مع مرور الوقت وبالتالي ننتقل به من وضع آمن إلى وضع آمن آخر من ناحية استقرار البيانات واتساقها.

ج- الاتساقية الزمنية: وتعني أن النظام سوف ينتقل بعد تنفيذ سلسلة من المناقلات إلى حالة تؤكد صحة البيانات وتلتزم معايير الاتساقية، بمعنى أنه يمكن لنظام قاعدة البيانات الاستغناء عن الاتساقية لحظياً مقابل تحقيق استجابة عالية للطلبات ليعود النظام سريعاً بعدها لفرض حالة من الاستقرار والصحة للبيانات عند أقرب فرصة متاحة زمنياً.

3 – **نظرية ACID**: تُستخدم في قواعد البيانات العلائقية ويرتكز عملها على الخصائص الأربع التالية:

أ- الكتلية: وتعني أن المناقلة إما أن تنفذ في قاعدة البيانات أو لا تنفذ أي لا يوجد تنفيذ جزئي لبعض تعليمات المناقلة في إما تنفذ أو ترفض ككتلة واحدة.

ب- الاتساقية: وتعني الانتقال من حالة صحيحة ومتناسقة للبيانات إلى حالة أخرى صحيحة ومتناسقة مع الحفاظ على تكاملية البيانات واستمرار كافة العمليات المطلوبة على مخازن المعطيات.

ج- العزل: تضمن عملية نجاح حصول عدة مناقلات متزامنة في آن معاً على قواعد البيانات، وذلك عن طريق تنفيذ كل مناقلة بمعزل عن الأخرى.

د- الاستمرارية: بمجرد الانتهاء من تنفيذ مناقلة معينة بنجاح، سوف يتم تطبيق كافة تعديلاتها الموجبة على قاعدة البيانات حتى لو انهار النظام لاحقاً.

بناءً على البنية والتصميم البرمجي المعتمد في عدد كبير من برامج أنظمة قواعد معطيات NoSQL المتوفرة حالياً ما يزال قسم منها ضمن التصنيف المعروض أعلاه، و بعضها مصنف جزئياً والبعض الآخر خارج التصنيفات المذكورة في الدراسات البحثية ذات الصلة، مما حتم علينا العمل على عرض مجموعات تصنيفية جديدة ملائمة للواقع و أكثر تمثيلياً للبرامج المتاحة بالاعتماد على نموذج التخزين، وعليه نقترح إضافة النماذج التالية إلى التصنيف المعتمد في الدراسات السابقة:

1- قاعدة البيانات متعددة النماذج(هجينة) (Hybrid Models Database): وتمثل حاويةً لبيانات مخزنة معاً وفقاً لنماذج متعددة في التخزين كنموذج القيمة/ مفتاح أو نموذج الوثائق أو نموذج الحفظ بالأعمدة وغيرها، وهنا لابد من الإشارة أنه يمكن لنا أن نصنف نظام Cassandra ضمن هذا الصنف حيث أنها تعتمد نموذج القيمة / مفتاح بالإضافة إلى نموذج الأعمدة، وبالتالي نخرجها من تصنيفها المدرج بالأبحاث السابقة ضمن تصنيف البيانات الموجهة بالأعمدة لتدخل تحت هذا الصنف الهجين الذي ينتمي إليه عدد من أنظمة قواعد المعطيات NoSQL (سوف نعرض بعضاً منها في الجدول رقم (1)).

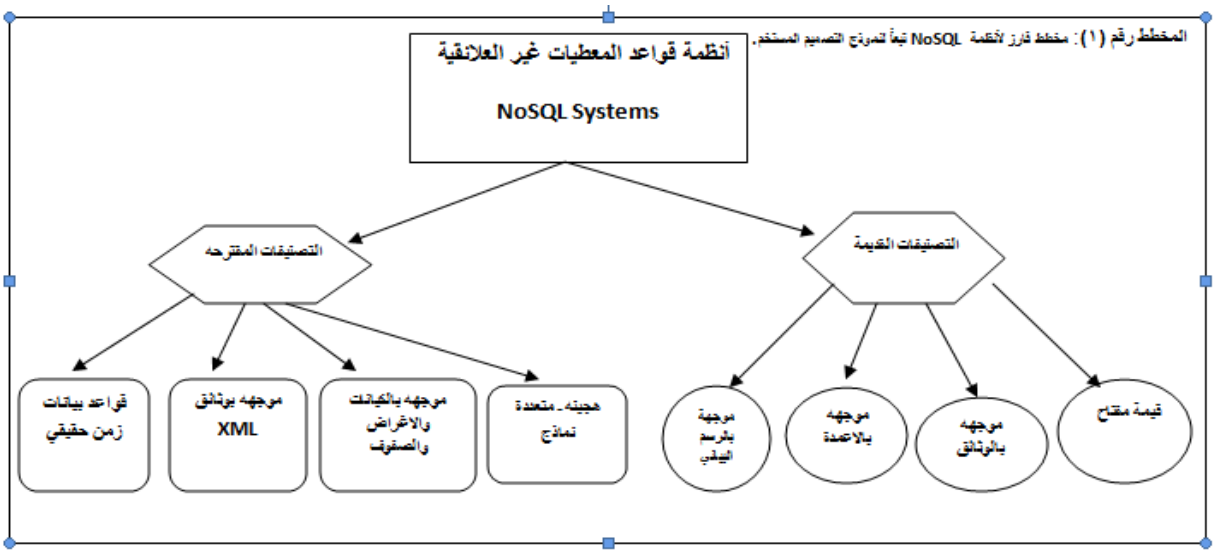
2- قاعدة البيانات الموجهة بالأغراض والكيانات(Entity–Object Oriented Database): تستند على حفظ البيانات ضمن أغراض (Objects) أو صفوف (Classes) وتعتمد في تصميمها على لغات البرمجة الغرضية التوجه مثل لغة Java ولغة C++ ولغة C# ولغة Python وغيرها من اللغات الداعمة لمفهوم الغرض في حفظ البيانات حيث تتمثل المعطيات بسلسلة من العقد الموزعة المخزنة على المخدمات، وسوف نبين أمثلة متعددة عن هذه الأنظمة من خلال الجدول اللاحق رقم (1).

3- قاعدة البيانات الموجهة بوثائق XML (Pure XML Document– Oriented Database): نقصد هنا النموذج الصافي منها، أي أن كل الوثائق المخزنة هي وثائق معتمدة على لغة التأشير الموسعة xml بشكل كامل ويختلف حجمها ومحتواها تبعاً لما تمثله من بيانات، وتُخزن ضمن هذه الأنظمة وبأحجام كبيرة وتدار بفعالية عبر مجموعة من أنظمة قواعد المعطيات NoSQL المفتوحة المصدر و المتخصصة في إدارة مخازن المعطيات XML، والبعض يعتبر هذه الأنظمة ضمن تصنيف قواعد البيانات الموجهة بالوثائق بشكل عام، ولكن رأينا أنه من الأفضل فصلها في

تصنيف خاص بها نظراً لخصوصية وثائق XML وقدرتها علي لعب دور هام كلغة تخاطب بين التطبيقات التي يتم تشغيلها على منصات عمل مختلفة، كون لغة XML لغة معيارية قابلة للفهم من مختلف البرمجيات وأنظمة التشغيل المتاحة، وسوف نعرض أيضاً من خلال الجدول اللاحق رقم (1) مجموعة من هذه البرمجيات التابعة لهذا الصنف.

4- قواعد البيانات الزمنية (Real Time-Oriented databases): وهي تمثل مجموعة من قواعد البيانات الموزعة المصممة لمعالجة وتحليل معدل تواتر عالي لسلسلة من المعطيات في الزمن الحقيقي، وتتضمن كافة الأدوات والإمكانات القادرة على التعامل مع البيانات الضخمة المولدة ضمن قاعدة المعطيات، وسنسرد بعضاً من هذه الأنظمة في الجدول التالي رقم (1).

وسنعرض من خلال المخطط رقم (1) رسماً بيانياً يوضح الفرز الحالي لأنظمة قواعد المعطيات غير العلائقية وفقاً للتصنيفات المذكورة أعلاه:



المناقشة والنتائج :

نلاحظ وجود الكثير من أنظمة قواعد المعطيات غير العلائقية المنتشرة حديثاً، تنتمي فعلياً إلى التصنيفات الجديدة وهي غير مشمولة ضمن التصنيفات السابقة الموجودة بالدراسات البحثية القديمة، ومن خلال تتبعنا للنموذج التصميمي للأنظمة الجديدة وآلية عملها وحفظها للبيانات والهيكليات المستخدمة بها، اخترنا عينة من خمسة وعشرين نظاماً عالمياً لإدارة البيانات الضخمة، وبالتدقيق في نموذج التصميم المستخدم وبإسقاطه على التصنيفات الجديدة، تم فرز هذه الأنظمة وفق الجدول التالي:

الجدول (1): أصناف جديدة لأنظمة القواعد غير العلائقية مستندة للنموذج

اسم الصنف	قاعدة البيانات متعددة النماذج	قاعدة البيانات الموجهة بالأغراض	قاعدة البيانات الموجهة بوثائق XML	قواعد البيانات الزمنية
أمثلة	WonderDB	Perst	Berkeley DB XML	eXtremeBD
	AlchemyDB	Objectivity	Qizx	InfluxDB
	RockallDB	HSS Database	BaseX	PipeLineDB

Raik Ts	Sedna	NEO	GunDB	
Axibase	eXist	JADE	FoundationDB	
GridDB	JEntigrator	FarmerD	CortexDB	

إن الجزم بانتماء كامل أنظمة برامج NoSQL للنظرية BASE هو أمر قد يكون مقبولاً من الناحية النظرية للوهلة الأولى، ولكن بمتابعة عمل هذه الأنظمة واستخداماتها ونماذج التصميم المختلفة، يتبين لنا أن لهذه الفرضية شواذ كثيرة، تجعلها تقترب إلى كل من نظريتي CAP وACID كنظريات معتمدة في حفظ واسترجاع البيانات من قواعد المعطيات.

فالتوافرية تركز على بقاء النظام يعمل بطريقة صحيحة مستجيباً لطلبات المستخدمين مدة طويلة من الزمن وتقنيات NoSQL توفر هذه الخاصية بطريقة أبسط من أنظمة قواعد المعطيات العلائقية[3].

يمكن لنا التأكد أنه كلما زادت نسبة معدل التوافرية في أنظمة NoSQL تقل نسبة الاتساقية للبيانات والعكس صحيح، أي أن تحقيق الاتساقية يقلل من احتمالات التوافرية نظراً للقيود الجديدة التي تفرضها الاتساقية على النظام، وعليه إن قسماً كبيراً من أنظمة NoSQL تنتمي إلى التصنيف الجزئي ضمن نظرية CAP عبر تحقيق خاصيتين معاً على التوازي التوافرية و العنقدة الجزئية (AP) أو الاتساقية و العنقدة الجزئية (CP) وقلة من هذه الأنظمة تجمع على التوازي خاصيتي التوافرية و الاتساقية (CA)، كما أن أنظمة قواعد المعطيات غير العلائقية كلما ارتفع بها معدل الاتساقية المحققة كما هو الحال في نظام Cassandra، كلما اقتربت أكثر من تحقيق نظرية ACID التي تضبط عمل أنظمة قواعد المعطيات العلائقية، أي أن ابتعاد أنظمة المعطيات NoSQL عن تحقيق التوافرية وتطبيقها للاتساقية بشكل صارم، يقربنا من نموذج الجداول وقواعد المعطيات العلائقية.

كلما زادت نسبة إنجاز التوافرية ضمن أنظمة NoSQL كلما ابتعدت بنية النظام الفعلية عن النموذج العلائقي الخاص بقاعدة البيانات.

تتمتع أنظمة NoSQL بمصداقية أو اعتمادية (Reliability) عالية كلما زادت نسبة الاتساقية فيها والعكس صحيح، حيث أن خاصية الاعتمادية تمثل قدرة النظام على العمل دون حدوث أخطاء لمدة زمنية محددة، وكلما زادت التوافرية قلت الموثوقية ومعها الاتساقية، وكلما زادت الاتساقية قلت التوافرية وازدادت الاعتمادية، حيث تعد أنظمة MongoDB و Neo4j من الأنظمة التي تتمتع باعتمادية عالية كونها تعتمد اتساقية كبيرة في التنجيز.

خاصية الثبات (Robustness) لدى أنظمة NoSQL تُعبر عن قدرة النظام على التأقلم مع الأخطاء واستيعابها خلال مرحلة التنفيذ، حيث تشتهر أنظمة قواعد المعطيات التقليدية بأنها ثابتة (Robustness)، ولكن هناك العديد من الاستفسارات التي تطرح عند مناقشة الموضوع في سياق NoSQL، حيث يعتبر معظم الباحثين أن أنظمة NoSQL الرائجة هي ثابتة وخاصة نظام Cassandra لاعتماده من قبل شركة الفيس بوك في الشبكات الاجتماعية ولكن ثباته لا يمكن مقارنته بأنظمة قواعد المعطيات العلائقية.

سنعرض فيما يلي الجدول رقم (2) لنبين خصائص الجودة لمجموعة من أنظمة قواعد المعطيات غير العلائقية NoSQL .

الجدول(2) تصنيف أنظمةNoSQL تبعاً لخصائص الجودة

الخاصية	الموجهة بالأعمدة	الموجهة بالوثائق	قيمة/مفتاح	الموجهة بالمخططات
التوافرية	جيدة	ممتازة	ممتازة	متوسطة
الاتساقية	متوسطة	متوسطة	جيدة	متوسطة
الاعتمادية	جيدة	جيدة	جيدة	متوسطة
الثبات	جيدة	جيدة	جيدة	لا بأس
CAP	AP/CP	CP	AP	AP

جميع المقاييس المتوفرة في الجدول تم اعتمادها كمعدل وسطي للخصائص العائدة لكل برنامج ينتمي إلى هذه الصنوف من خلال سلسلة من الدراسات البحثية في هذا المجال[4][5][6]، وبالتالي من خلال تتبع بسيط للجدول نلاحظ وبشكل عام أنه في أنظمة NoSQL كلما زادت التوافرية قلت الاتساقية وبالتالي الاعتمادية، واقتربنا نحو نظرية AP والنموذج BASE كفلسفة تصميم، وكلما زادت الاتساقية زادت الموثوقية (الاعتمادية) واقتربنا نحو نظرية CP وبالتالي نحو نموذج ACID والتي تضبط عمل قواعد المعطيات العلائقية.

إن عدداً كبيراً من مستخدمي قواعد البيانات يضعون في سلم أولوياتهم الاعتماد على الأكثر رواجاً من أنظمة قواعد المعطيات غير العلائقية، والتي تتميز بنسب استخدام عالية عالمياً، وهم يرون أن البرامج الأكثر شهرة هي البرامج الأكثر نجاعة في إدارة قواعد بياناتهم وأكثر موثوقية كونها مجربة ولديها كم هائل من الزبائن والمستخدمين.

يعد الانتشار الذي حققته هذه الأنظمة أكبر دليل على أهميتها وجدارتها في إدارة قواعد المعطيات، أي أن هذا الصنف من المستخدمين يغض الطرف عن النماذج التصميمية لأنظمة قواعد المعطيات، ويتجاهل كل الخصائص التي توفرها هذه الأنظمة والمتعلقة بالنظريات القواعدية الناظمة لحفظ المعطيات، بالإضافة إلى ما يترتب عليها من مواصفات جودة تتعلق بالاتساقية والثبات و التوافرية و الاستمرارية وغيرها من الصفات أو فيما إذا كان النظام مناسباً لطبيعة و هيكلية البيانات المراد تخزينها وإدارتها.

أي أن المعيار الثابت لديهم في استخدام أنظمة قواعد المعطيات غير العلائقية هو مؤشر سوق العمل المرتبط بشهرة هذه الأنظمة وانتشارها في الاستخدام العالمي وتوسعها أفقياً و شاقولياً، وعليه وجدنا أنه من المناسب في هذا البحث أن نعتمد تصنيفاً يستند إلى هذا المعيار كدليل بحثي مرجعي في هذا المجال، لذا افترضنا وجود ثلاثة مجموعات رئيسية في التصنيف الجديد، وهنا اعتمدنا في استنتاج مؤشرات التصنيف على زيارة الموقع الالكتروني للأنظمة المقترحة والاطلاع على عدد زوارها و النسخ المحملة منها، وعلى القاعدة المعرفية العالمية المتعلقة بأنظمة قواعد المعطيات العلائقية وغير العلائقية والمزودة من قبل فريق عمل SOLID IT والمختص في تقديم الإرشاد للزبائن نحو التطبيقات المناسبة لهم لإدارة قواعد بياناتهم وذلك في ظل النمو الهائل في عدد وحجم أنظمة قواعد المعطيات والتي تقدم كل منها بديلاً مقترحاً لإدارة المعطيات.

التصنيف المقترح حسب معيار الاستخدام العالمي [7]:

1- الأنظمة الأكثر استخداماً عالمياً: وتشمل مجموعة من الأنظمة التي تقدم تسهيلات هامة في إدارة قواعد البيانات، تبعاً للميزات التي تقدمها والمرتبطة بنموذج التصميم والتخزين بالإضافة إلى الخدمات الأخرى المتعلقة بقواعد

المعطيات والتي توفرها هذه الأنظمة بسهولة مما جعلها أكثر رواجاً عالمياً وتتدرج في تسلسل أكثر مئة نظام معطيات مستخدم عالمياً، وسنسرد في الجدول رقم (3) أسماء عدد من هذه الأنظمة.

2- الأنظمة متوسطة الاستخدام عالمياً: يعتمدها عدد مهم من الشركات والمنظمات والأشخاص وتقدم حلولاً برمجية لإدارة قواعد المعطيات ولكنها أقل جودة في إدارة قواعد البيانات من مثيلاتها السابقة.

3- الأنظمة قليلة الاستخدام عالمياً: معظمها أنظمة تجريبية جاءت لمعالجة إشكالات معينة في إدارة قواعد البيانات، وقد يكون من أسباب عدم انتشارها في الاستخدام العالمي هو ضعف عملية تسويقها و إظهار ميزاتها للعاملين والمهتمين بهذا المجال، حتى أن البعض منها حديث العهد في عالم قواعد المعطيات وكثير من مستخدمي تلك الأنظمة لم يعلموا بها أو يتنبهوا لخصائصها وميزاتها بعد.

وسنعرض من خلال الجدول(3) عدداً من البرامج كممثل عن كل مجموعة تصنيفية:

الجدول(3) تصنيف حسب معيار الاستخدام العالمي

الصنف	أمثلة
مجموعة 1	MongoDB,Redis,Cassandra,Cloudkit,HBase,AmazonDynamoDB, CouchDB, Hazecast, LevelDB
مجموعة 2	RocksDB,GridGain, JBase,Geode, BaseX, Giraph, Bigchain DB,MapDB
مجموعة 3	Lockijs,Raik Ts, Scalaris, Sensei DB, Denso DB, Badger, WhiteDB, LedisDB, SawyDB, JasDB

لا توجد لغة استعلام خاصة بنموذج معين من أنظمة قواعد المعطيات غير العلائقية، لكن هناك محاولات لإنشاء لغات استعلام تستفيد من لغة الاستعلام SQL المستخدمة في قواعد البيانات العلائقية من حيث البنية مع بعض التحسينات في تركيبة التعليمات بما يتوافق ونموذج التصميم المستخدم في التخزين لدى نظام المعطيات مع إضافات مقترحة في أساليب الفهرسة وتحليل الاستعلامات لنحصل على خرج متطابق مع صنف ومتطلبات الاستعلام. وهنا سنعرض مجموعة مقترحة من الأصناف لأنظمة NoSQL تبعاً للغة الاستعلام المستخدمة في قواعد المعطيات:

1- الصنف الأول: تشمل مجموعة من أنظمة NoSQL التي تعتمد لغة الاستعلام CQL المستخدمة في نظام Cassandra، وهي لغة ليست ببعيدة عن سياق التعليمات في لغة الاستعلام المهيكلة SQL مع تحسينات تتعلق بالتوافق مع نموذج المعطيات الموجه بالأعمدة.

2- الصنف الثاني: يمثل عدداً من أنظمة NoSQL الموجهة بالوثائق من نمط وثائق نصية أو وثائق BSON أو JSON و تعتمد لغة استعلام تسمى SQL++ ومثلها لغة الاستعلام N1QL.

3- الصنف الثالث: تضم عدداً من أنظمة NoSQL التي تعتمد وثائق XML في التخزين، حيث ينشط استخدام لغات استعلام من نوع XQuery التي تراعي خصوصية وهيكلية ملفات XML ومثلها نموذج XPath الذي يمثل قالباً هاماً في تمثيل ملفات XML.

هناك عدد من مستخدمي قواعد البيانات غير العلائقية يبحثون عن مزايا خاصة بالنظام تتعلق بعمليات الحفظ والاسترجاع من ناحية الاستجابة للحصول على نتائج سريعة للاستعلامات، أو من ناحية الزمن المستغرق في عمليات حفظ البيانات بعد إدخالات لبيانات بحجوم كبيرة، وهنا يدخل عامل الأداء ليؤسس لمعيار مهم في تقييمات المستخدمين لاعتماد نظام ما، حيث يأخذ بعين الاعتبار الزمن المستغرق من قبل النظام لتنفيذ عملية ما، أكانت إدخالاً أو إظهاراً أو تحديثاً لمحتوى معين من البيانات [8]، وهنا يلعب نموذج التصميم لنظام إدارة قاعدة البيانات والطريقة التي يحفظ بها للمعطيات عاملاً مهماً في التحكم بعملية أداء النظام [9]، حيث يمكن لنا أن نميز هنا بين صنفين هامين من قواعد البيانات غير العلائقية وهما:

1- قواعد البيانات المحسنة للقراءة: وهي قواعد بيانات مناسبة للعمليات المتواترة في الاستعلام عن البيانات حيث توفر للمستخدمين سرعة وجودة عالية في عمليات استرجاع البيانات وعرضها ضمن زمن تنفيذ مثالي، وتعد بنية قاعدة البيانات محركاً أساسياً للأداء في مثل هذه الحالات، حيث تعتبر قواعد البيانات غير العلائقية من نمط قيمة/مفتاح وقواعد البيانات الموجهة بالوثائق مثالية في مثل هذه الحالات، فالزمن المستغرق في الوصول إلى البيانات المطلوبة وفقاً للاستعلام هو زمن قصير نسبياً إذا ما قورن ببقية أنظمة قواعد البيانات غير العلائقية ومن الأمثلة المشهورة هنا في هذا الصنف هي CouchDB و MongoDB و Redis.

2- قواعد البيانات المحسنة بالكتابة: وهي مثالية لمتطلبات أرشفة البيانات وتخزينها لعمليات استرجاع لاحقة وتقدم نموذجاً عالي الأداء خلال تنفيذ عمليات الإدخال للبيانات وحفظها، حيث إن إنجاز قاعدة البيانات باستخدام ذواكر إضافية للكتابة، وبضبط التحكم بحيث يتم التحديث والإدخال على LogFile لتنقل بشكل مباشر إلى الحفظ المؤقت في الذاكرة الرئيسية، ومن بعدها للتخزين الدائم على القرص (الذاكرة الثانوية)، أي أن تعدد استخدام الذواكر في أنظمة قواعد البيانات يجعلها مثالية لعمليات إدخال البيانات وحفظها وتعطي أداء سريع بوجود حجم كبير من البيانات في عمليات الكتابة، وتعتبر النماذج الهجينة التي تعتمد الدمج بين نموذجي مفتاح/قيمة والنموذج الموجه بالأعمدة (BigTable) كما هو الحال في نظام Cassandra، أنظمة مثالية لعمليات تخزين وكتابة البيانات وأرشفتها [10] [11].

الاستنتاجات والتوصيات:

بالتدقيق في بنية أنظمة قواعد المعطيات غير العلائقية من ناحية النموذج التصميمي والآليات التي تستخدمها هذه الأنظمة لإدارة قواعد البيانات وحفظها، نجد أنه يوجد لدينا خيارات كثيرة في عملية تصنيف هذه الأنظمة تبعاً لعوامل عديدة تم ذكرها في البحث، ولكن النمو المتسارع في عدد هذه الأنظمة وبنيتها والتغيير في أساليب حفظها و معالجتها للبيانات، يجعل من استمرار التحديث للدراسات البحثية المتعلقة بها أمراً ضرورياً بهدف مواكبة التطورات المتسارعة لهذه الأنظمة، ولكن الثابت الوحيد هنا هو أن نوعية البيانات المراد حفظها في قاعدة البيانات يفرض على النظام المتعلق بإداراتها استخدام بنى و هيكليات محددة والتقيد بطرائق معالجة محددة لهذه المعطيات حيث أن طبيعة البيانات المخزنة من ناحية كونها مهيكلة أو شبه مهيكلة تتطلب آليات محددة لتخزين البيانات والسماح بالاستعلام اللاحق عنها من قبل المستخدمين.

كما تبين لنا أن عدداً من أنظمة قواعد المعطيات غير العلائقية كفيل بإدارة ناجحة لبيانات المستخدمين ولكن ضعف عمليات تسويق هذا النظام أو غيره من هذه الأنظمة، يبقيه مجهولاً لدى الكثير من المستخدمين حيث يسارع معظمهم إلى استخدام الشائع من هذه الأنظمة والأكثر رواجاً عالمياً، مع العلم أن البعض من الأنظمة قليلة الشهرة قد يكون أكثر كفاءة في تحقيق إدارة فعالة وناجحة لبيانات المستخدمين.

لا بد لنا في الدراسات المستقبلية من التركيز على نماذج التصميم لهذه الأنظمة و وضع معايير تتعلق بآلية حفظ البيانات واسترجاعها والشروط الواجب تحققها حتى يتم السماح للمستخدمين بالتعديل على محتوى قاعدة المعطيات دون المساس بصحة المعلومات المحفوظة، مما يعطي نظام قاعدة البيانات وثوقية أكبر، كما ينبغي توجيه الشركات والمؤسسات العاملة في إدارة قواعد البيانات إلى الاعتماد على التصنيفات الجديدة في عملية تطوير أي نظام قاعد معطيات غير علائقي، بحيث نخرج تدريجياً من حالة الفوضى واللامعيارية في بناء هذه الأنظمة و نعزز خصائصها الإيجابية ونحقق حد أدنى من المواصفات التي ينبغي تحقيقها في أي نظام قاعدة بيانات غير علائقي، لنحقق شروط جودة معينة ينبغي توافرها في أي نظام يتعامل مع إدارة المعطيات تحقيقاً للفعالية والأمان اللازم للبيانات المخزنة.

المراجع :

1. GESSERT,F;WINGREATH,W;FRIEDRICH,S;RITTER,N. *NoSQL database system: survey and decision guidance*. Springer-Verlag , Berlin, 2016, 13.

2. MAKRIS,A;TSERPES,K; ANDRONIKOU,V; ANAGNOSTOPOULOS,D. *A Classification of NoSQL Data Stores Based On Key Design Characteristics.* Procedia Computer Science, university of Athens, Greece , 2016, 13.

3. MONIRUZZAMAN,A ; HOSSAIN, S. *NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison*. Daffodil International University, International Journal of Database Theory and Application, 2013, 15.

4. SHARMA,S. *An Extended Classification and Comparison of NoSQL Big Data Models. Center for Survey Statistics and Methodology.* Iowa State University, Ames, Iowa, USA,2016,20.

5. OUSSOUS,A; BENJELLOUN,F; LAHCEN,A ; BELFKIH,F. *Comparison and Classification of NoSQL Database for Big Data.* Mohammed V University. Rabat, Morocco, 2015,10.

6. LOURENCO,J; CABRAL,B; CARREIRO,P; VIEIRA,M; BERNARDINO,J. *Choosing the right NoSQL database for the job: a quality attribute evaluation.* journal of big data, Portugal, 2015, 26.

7. SOLID IT ,10 Sept. 2018, < *http://www.db-engines.com*>.

8. MOHAMED,M; ALTRAFI,O; ISMAIL,M. *Relational vs. NoSQL Databases Survey.* International Journal of Computer and Information Technology, India, vol. 03, no. 32014, pp.598–601.

9. ZAKI,A,K. *NoSQL DATABASES: NEW MILLENNIUM DATABASE FOR BIG DATA, BIG USERS, CLOUD COMPUTING AND ITS SECURITY CHALLENGES*. International journal of research in engineering and technology, India, 2014, eISSN:2319-1163 | pISSN: 2321-7308.

10. PORE,S; PAWAR,B .*Comparative Study of SQL & NoSQL Databases*. International Journal of Advanced Research in Computer Engineering & Technology, India ,Volume 4 Issue 5, 2015,20.

11. FARAJ,A; RASHID,B; SHAREEF,T. *Comparative Study of relational and non relations database performances using Oracle and MongoDB System*. International Journal of Computer Engineering and Technology, Iraq, Volume 5, Issue 11,2014,1