Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي (AAJSR)
E-ISSN: 2959-6505
Volume 2, Issue 4, October - December 2024
Page No: 181-194
Website: https://aajsr.com/index.php/aajsr/index
SJIFactor 2024: 5.028 ISI 2024: 0.580 معامل التأثير العربي (AIF) 2024: 0.74

مستقبل قواعد البيانات العلائقية مع تزايد البيانات الضخمة وغير المنسقة (دراسة استعراضية)

هديه سليمان هويدي[1*]، صلاح الهادي جوان [2]، عبدالعزيز امحمد ابوعائشة [3]، مصطفى صلاح الدين حمادي [4]، عبدالقادر ابوريمة احمد[5]
[1]قسم نظم المعلومات، كلية تقنية المعلومات، الجامعة الاسمرية الاسلامية، زليتن، ليبيا
[2]قسم علوم الحاسوب، كلية العلوم، الجامعة الاسمرية الاسلامية، زليتن، ليبيا
[3]قسم علوم الحاسوب، مدرسة العلوم الاساسية، الاكاديمية الليبية للدراسات العليا، طرابلس، ليبيا
[4]قسم علوم الحاسوب، مدرسة العلوم الأساسية، الاكاديمية الليبية للدراسات العليا، طرابلس، ليبيا
[5]قسم علوم الحاسوب، كلية التربية، جامعة بني وليد، بني وليد، ليبيا

The Future of Relational Databases with The Growth of Big and Unstructured Data: (Review Study)

Hadya S Hawedi [1 *], Salah A. Jowan [2], Abd Alaziz Emhemed Abu Aisha [3], Mustafa Salah Eddn Hamadi [4], Abdulghader Abu Reemah A Abdullah [5.]
[1]Department of Information Systems, Faculty of Information Technology, Alasmarya Islamic University, Zliten, Libya
[2]Department of Computer science, Faculty of Science, Alasmarya Islamic University, Zliten, Libya
[3]Department of Computer science, School of Basic Sciences, Libyan Academy for Postgraduate Studies, Tripoli–Libya
[4]Department of Computer science, School of Basic Sciences, Libyan Academy for Postgraduate Studies, Tripoli–Libya
[5]Department of Computer science, Faculty of Education, Bani Walid University, Bani Walid, Libya

*Corresponding author hadia20008@asmarya.edu.ly *المؤلف المراسل
تاريخ الاستلام: 2024-08-19 **تاريخ القبول:** 2024-10-20 **تاريخ النشر:** 2024-10-23

الملخص
أصبحت البيانات في الآونة الأخيرة أكبر وأكبر وأكثر تعقيدا مع النمو الهائل في سوق الإنترنت وزيادة استخدام شبكات التواصل الاجتماعية، وأجهزة الاستشعار لاكتساب البيانات واستخدامات الهواتف الذكية وغيرها. عادة ما تكون مثل هذه البيانات غير منسقة *(unstructured data)* وتنشأ من مصادر متعددة في شكل مختلف والذي يسمى بالبيانات الضخمة *(Big Data)*. تشير البيانات الضخمة إلى مجموعات البيانات التي ليست كبيرة الحجم فقط *(Volume)*، ولكنها أيضًا عالية في التنوع *(Variety)* وسرعة النمو *(Velocity)*، بحيث لا يمكن إدارتها بشكل فعال أو التعامل معها باستخدام نظام قواعد البيانات العلائقية *(RDBMS)*. إن ظهور تقنيات الويب الجديدة والحوسبة السحابية أتى بتحديات وتطبيقات ومفاهيم جديدة مثل قواعد البيانات غير العلائقية *(NoSQL)* التي أصبحت مؤخرًا شائعة جدًا كبديل لقواعد البيانات العلائقية خاصة في التعامل مع البيانات الضخمة. تستعرض هذه الورقة البيانات الضخمة وخصائصها واستخداماتها، كما تهدف أيضا إلى تسليط الضوء على أنواع وخصائص قواعد البيانات غير العلائقية *(NoSQL)* واستخداماتها ومميزاتها وعيوبها ومحاولة الإجابة على السؤال المتعلق بالاستخدام المستقبلي لقواعد البيانات العلائقية. كما تهدف هذه الدراسة إلى توفير فهم أفضل لقواعد البيانات غير العلائقية *(NoSQL)* للطلاب والباحثين والمهتمين بهذا المجال.

الكلمات المفتاحية: البيانات الضخمة، البيانات غير المنظمة، نظام قواعد البيانات العلائقية، قواعد البيانات غير العلائقية.

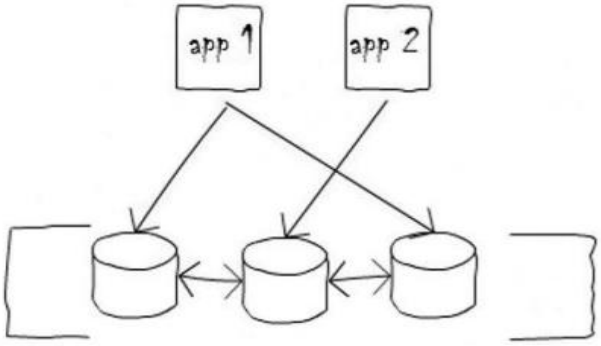
Abstract:
Data recently became larger and larger moreover complicated with the enormous growth of internet and the increase of the social media and the use of mobile phones. Usually, this amount of data is not well structured (unstructured data). This huge amount of data is Originated from various sources and assorted types is called big data. Big data refers to data sets which are not only large in volume but rich in variety and the growth speed. Such as unstructured large amount of data cannot be potently managed or handled using a relational database system (RDBMS). A new challenges, applications, and concepts have been brought by new web technologies and cloud computing such as non- relational database (NoSQL) which became very popular as an alternative to relational databases in order to deal with big data. This paper sheds light on big data, its characteristics and uses also the characteristics of the NoSQL database and its advantages and disadvantages. This paper attempts to give an answer to the question regarding the future use of relational database, also aims to provide a clear and better understanding of NoSQL databases for students, researchers, and those interested in this field.

Keywords: Big Data; Unstructured Data; NoSQL; RDBMS.

المقدمة:

لقد تغيرت التكنولوجيا كثيرًا في هذا العصر الرقمي. لدينا هواتف ذكية وأجهزة كمبيوتر محمولة وأجهزة لوحية تجعل حياتنا أكثر ذكاءً. في وقت سابق كنا نستخدم أجهزة كمبيوتر مكتبية ضخمة لمعالجة كميات هائلة من البيانات، وكنا نستخدم الأقراص المرنة والأقراص الصلبة لتخزينها. الآن يمكننا تخزين البيانات في السحابة وبسبب التحسين المتسارع للتكنولوجيا أصبحنا ننتج الكثير من البيانات، فعلى سبيل المثال يُنشئ كل مستخدم للهواتف الذكية ما يقرب من 40 إكسابايت من البيانات كل شهر في شكل نصوص ورسائل بريد إلكتروني ومكالمات هاتفية ومقاطع فيديو وصور وعمليات البحث والموسيقى وغير ذلك (Aparna, 2021). فإذا تضاعف هذا الرقم في 5 مليارات مستخدم للهواتف الذكية، فهذا يعد كمية ضخمة من البيانات. كذلك تطور التقنيات الأخرى بوتيرة سريعة كأجهزة الاستشعار التي تُستخدم في كل مكان تقريبًا في حياتنا اليومية، والبيانات من آلة إلى آلة والروبوتات والذكاء الاصطناعي وإنترنت الأشياء، ليس فقط التقنيات تتطور ولكن المزيد والمزيد من الناس يستخدمون أنواعًا مختلفة من الأجهزة التي تولد حتى مزيدا من البيانات. كل شيء مما ذكر يولد أنواعًا مختلفة من البيانات - منسقة وشبه منسقة وغير منسقة. إن احدى أكبر التحديات التكنولوجية الرئيسية التي يواجها العالم ليس النمو المتسارع لحجم البيانات فقط، لكن الجزء الصعب هنا هو أن البيانات ليست موجودة بطريقة منسقة حيث يتم إنشاءها وتوليدها بملايين الطرق المختلفة وهي احدى أكبر العوامل لتطور البيانات الضخمة.

اكتشفت الشركات الكبرى وعلى رأسها Googleأن تقنية قواعد البيانات العلائقية (RDBMS) التقليدية لها قيود خطيرة في بنيتها بحيث لا تستطيع التعامل مع هذا الكم الهائل من البيانات، كما أنها لم تصمم للتعامل مع البيانات الغير منسقة (Moko & Asagba, 2021). ونتيجة لذلك تمكنوا من التوصل إلى تقنيات إدارة بيانات مختلفة أدت إلى خلق اهتمام من الشركات التي لديها مشكلات مماثلة. أدى هذا إلى ولادة قواعد البيانات غير العلائقية (NoSQL) مرنة بما يكفي للتعامل مع مجموعة متنوعة من تنسيقات البيانات، وكذلك تقدم أداءً فائقًا بالإضافة إلى سهولة التوسع للتعامل مع البيانات الضخمة حيث تكون البيانات مخزنة بطريقة موزعة ولكن يتم الوصول إليها وتحليلها وكأنها موجودة على جهاز واحد (شكل 1). هذا النوع الجديد من برامج إدارة البيانات تظهر حلول للتعامل مع قواعد بيانات موزعة على منصات مفتوحة بسرعة نقرة الماوس (Mohanty et al., 2015).



شكل (1): قواعد البيانات الموزعة.

ظهور البيانات الضخمة في الصورة كظاهرة جديدة فتحت الأبواب لظهور أدوات وتقنيات جديدة لتخزين كمية كبيرة من البيانات غير المنسقة وإدارتها Big Data وNoSQL والذي سنقوم باستعراضها بالتفصيل في الفقرات التالية.

اولاً: ما هي البيانات الضخمة (Big Data):

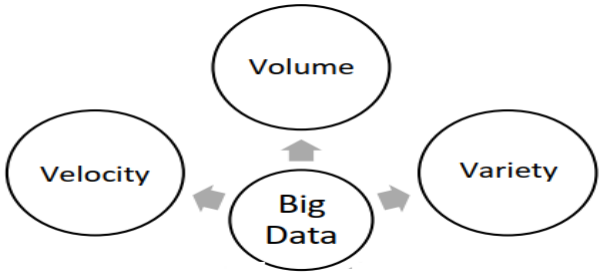
البيانات الضخمة هي المصطلح الذي يصف البيانات، وهي كبيرة ومعقدة لدرجة أنها صعبة أو في بعض الحالات يستحيل معالجتها بالطرق التقليدية (Kovacheva, 2021). جمع مثل هذه الكميات الهائلة من البيانات مفيد لإجراء تحليلات أفضل وأوسع نطاقاً لمشاكل معينة، والتي تعتبر مهمة للشركات. أوضح الباحثان Michael وAndreas (Meier & Kaufmann, 2019)، أن البيانات الضخمة لا يزال ليس لها تعريف ملزم ولكن العديد من الخبراء اتفقوا على ثلاث مقاييس لوصف البيانات الضخمة وأطلق عليها المصطلح ثلاث V's أو (Vs3): الحجم (Volume) والسرعة (Velocity) والتنوع (Variety) انظر (شكل 2).

1- الخصائص الرئيسية للبيانات الضخمة:

الحجم (Volume): الحجم النموذجي للبيانات الضخمة كبير جدا أي بحجم بيتابايت أو إكسابايت ويتزايد بسرعة كل عام. في الماضي كان حجم البيانات الزائد يمثل مشكلة تخزين، ولكن مع انخفاض تكاليف التخزين تظهر مشكلات أخرى، بما في ذلك كيفية تحديد البيانات الملائمة ضمن أحجام البيانات الكبيرة وكيفية استخدام التحليلات لإنشاء قيمة من البيانات ذات الصلة.

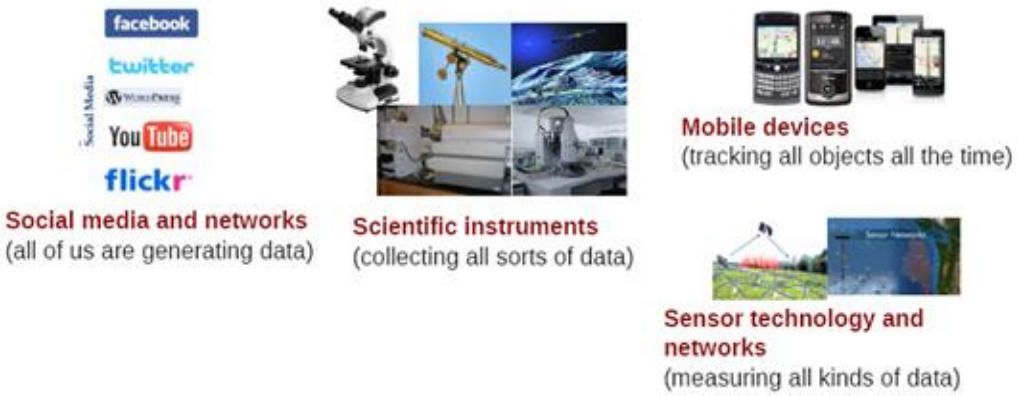
السرعة (Velocity): تتدفق البيانات بسرعة غير مسبوقة ويجب التعامل معها في الوقت المناسب. تعمل علامات RFID وأجهزة الاستشعار والقياس الذكي على زيادة الحاجة إلى التعامل مع السيول من البيانات في الوقت الفعلي تقريبًا. يمثل الرد بسرعة كافية للتعامل مع سرعة البيانات تحديًا لمعظم المؤسسات.

التنوع (Variety): تأتي البيانات اليوم بجميع أنواع التنسيقات. بيانات رقمية منظمة في قواعد البيانات التقليدية، المعلومات التي تم إنشاؤها من تطبيقات خط الأعمال، المستندات النصية الغير منسقة والبريد الإلكتروني والفيديو والصوت وبيانات مؤشر الأسهم والمعاملات المالية. لا تزال العديد من المؤسسات تكافح من أجل إدارة أنواع مختلفة من البيانات ودمجها وإدارتها.



شكل (2): Big Data 3Vs

أصبحت مصادر البيانات أكثر تعقيدًا ويتم جمعها من مصادر مختلفة كما ذكرنا من قبل مثل الذكاء الاصطناعي، والأجهزة المحمولة، والوسائط الاجتماعية، أجهزة الاستشعار، والويب، والوسائط الاجتماعية، إلخ. والتي يتم انشاؤها في الوقت الفعلي بكميات كبيرة جدًا. يمكن رؤية أمثلة توليد البيانات الضخمة في مناطق مختلفة مثل أسواق البورصة، شركات التأمين، محركات الطائرات، الصناعة، الرعاية الصحية، الدعاية والاعلام، الخدمات المالية، التعليم، الطاقة، الزراعة، الخدمات اللوجستية ومواقع التواصل الاجتماعي وغيرها. يوضح (الشكل3) البيانات من مصادر مختلفة. هذا يقودنا إلى النقطة المهمة التالية في هذه الورقة وهي أنواع البيانات المختلفة التي يمكن جمعها (بيانات منسقة - شبه منسقة - غير منسقة).



شكل (3): المصادر المختلفة للبيانات

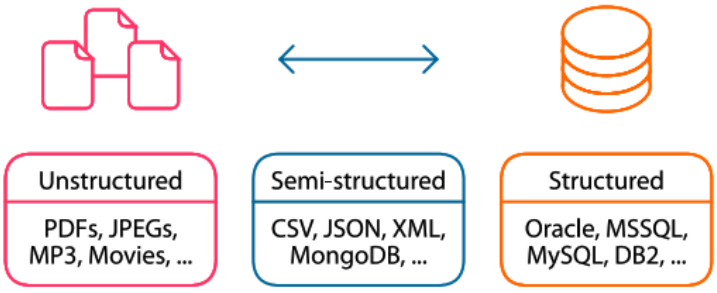
ثانياً: فوائد استخدام البيانات الضخمة:

تُستخدم البيانات الضخمة في الصناعات واسعة النطاق لتوفير قيمة للعملاء بشفافية وسهولة من خلال تحليل سلوك العملاء من البيانات المستمدة من مواقع التواصل الاجتماعي المختلفة، ولقطات CCTV، ومصادر أخرى. يتفق المؤثرون في الصناعة والأكاديميون وأصحاب المصلحة البارزون الآخرون على أن البيانات الضخمة أصبحت عامل تغيير كبير في قواعد العمل في معظم الصناعات. حيث أن الهدف الأساسي لمعظم المؤسسات هو تحسين تجربة العملاء، وخفض التكلفة، والتسويق الموجه بشكل أفضل، وجعل العمليات الحالية أكثر كفاءة أدى الى تغيير كبير في كيفية استثمار وإدارة البيانات وتخزينها للاستخدام المستقبلي. الأمر الذي يجعل المؤسسات المختلفة تبذل أقصى جهودها لتأمين كل جزء محتمل من البيانات والاحتفاظ بها. هناك العديد من الاستخدامات الأخرى للبيانات الضخمة مثل:

- إنشاء عروض للعملاء بناءً على سلوكياتهم عبر الإنترنت.
- تخفيضات التكلفة والوقت
- تطوير منتجات جديدة وعروض محسّنة
- اتخاذ قرارات تجارية أكثر ذكاءً
- كشف الاحتيال
- الإعلان
- الرعاية الصحية
- وسائل الإعلام والترفيه
- التدريب التربوي

ثالثاً: أنواع البيانات:

اعتمادًا على المصادر التي يتم جمع البيانات منها، هناك 3 أنواع من البيانات، منسقة وشبه منسقة وغير منسقة (شكل4). إن تنسيق البيانات مهم ليس فقط لكيفية العمل معها ولكن ما هي الاستنتاجات التي يمكن استخلاصها وما هي الحلول لمشاكل معينة للشركات التي يمكن تقديمها.



شكل (4): أنواع البيانات (Mchergui et al., 2023).

البيانات المنسقة (Structured): هي جميع المعلومات التي يمكن تخزينها ومعالجتها بتنسيق ثابت (Kovacheva, 2021). عادة ما يتم تخزين البيانات في قاعدة بيانات العلائقية. في هذا النوع من قواعد البيانات يتم إنشاء الجداول مسبقًا ويتم تحديد أنواع البيانات التي سيتم تخزينها بالداخل ولا يمكن تغييرها لاحقًا. البيانات المنسقة لها بنية محددة مسبقًا ويمكن تخزينها ليس فقط في قواعد البيانات العلائقية، ولكن في Excel أيضًا. بعض الأمثلة من البيانات المنسقة هي المعاملات المالية والمحاسبية والعناوين ومعلومات العملاء والمواقع وما إلى ذلك.

البيانات شبه المنسقة (Simi-structured): هي البيانات التي لا يمكن تنظيمها في قاعدة بيانات علائقية لأنها لا تحتوي على بنية محددة وثابتة ولكن لا يزال لديها بعض منها (Kovacheva, 2021). البيانات شبه المنسقة لها خصائص من البيانات المنسقة وغير المنسقة. أحد الأمثلة على البيانات شبه المنسقة هو البريد الإلكتروني. تحتوي رسالة البريد الإلكتروني على بعض البنية، حيث تحتوي على البريد الإلكتروني الخاص بـالمرسل، والمستلم، والموضوع، وما إلى ذلك. الرسالة المرسلة من تلقاء نفسها غير منسقة، لكن بقية البريد الإلكتروني بها بعض الحقول المحددة مسبقا.

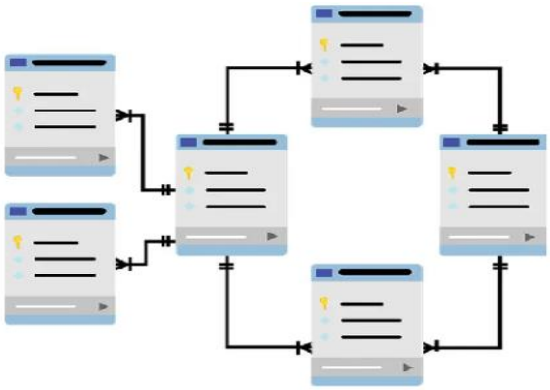
البيانات الغير منسقة (Unstructured): هي نوع من البيانات التي لا تحتوي على أي شكل من أشكال البنية الهيكلية الثابتة ولا يمكن تخزينها في قواعد بيانات علائقية (Kovacheva, 2021). من الصعب جدًا معالجة هذا النوع من البيانات وتحليله لأنه يأتي بشكل مختلف التنسيقات. في الوقت الحاضر البيانات الغير منسقة هي النوع الذي يتم توليده في الغالب، لذلك نحن بحاجة إلى التأكد من امكانية تخزينها وتحليلها بشكل صحيح. من أمثلة البيانات الغير منسقة الصور ومقاطع الفيديو والملفات النصية ومواقع الويب ومحتوى الوسائط الاجتماعية وملفات PDF وغيرها.

رابعاً: قواعد البيانات:

لعبت قاعدة البيانات دورًا مهمًا في المجتمع لعدة قرون، وأي شيء يحمل معلومات باستمرار هو نوع من قواعد البيانات. رسومات الكهوف والألواح الحجرية وورق البردي كلها أمثلة حقيقية لقاعدة البيانات. يرجع تاريخ قاعدة البيانات إلى حد كبير إلى الرغبة في إنشاء هياكل يمكنها تلبية متطلبات تخزين المعلومات بدقة وبشكل دائم على مدى فترات طويلة مع تمكين المستخدمين من البحث بسهولة عن المعلومات داخل البيانات المخزنة. ظهور أجهزة الكمبيوتر فتح إمكانيات جديدة لتكنولوجيا قواعد البيانات، مما سمح بإنشاء المعلومات المخزنة على القرص والوصول إليها بسرعة. ومع ذلك، فإن مجرد وجود النظام المحوسب فقط لا يمكن تفاعلية البيانات بقوة تلقائيًا. إنها تحتاج إلى أنظمة لإدارتها ولغات لعمليات التفاعل مع البيانات من حيث الاستعلام والإضافة والتحدث والعمليات الأخرى التي تجريها على قاعدة البيانات للحصول على المطلوب.

1- قواعد البيانات العلائقية (RDBMS) أو (SQL):

قاعدة البيانات العلائقية هي بنية بيانات التي تسمح لك بربط المعلومات من "جداول" مختلفة أو أنواع مختلفة من حاويات البيانات. يجب أن تحتوي حاوية البيانات على ما يسمى بالمفتاح أو الفهرس (الذي يسمح للتعرف بشكل فريد على أي جزء ذري من البيانات داخل الحاوية) بحيث حاويات البيانات الأخرى قد تشير إلى هذا المفتاح من أجل إنشاء ارتباط بين ذرات البيانات.

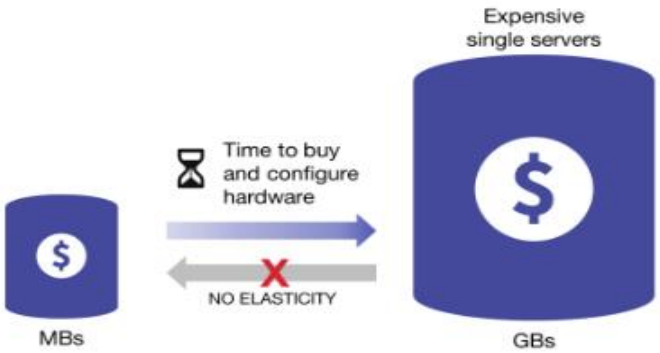


شكل (5): قواعد البيانات العلائقية

أصبحت قاعدة البيانات العلائقية جزءًا لا يتجزأ من معظم تطبيقات البرامج، وفي رأي معظم المبرمجين، لا يمكن الاستغناء عنها نظرًا لسهولة استخدامها وتنظيم البيانات والتخزين المستمر، مما يتيح بيانات ثابتة الاسترداد مع السماح لعدة مستخدمين بالوصول إلى نفس البيانات بشكل متزامن مع القدرة على تعديل تلك البيانات. تتم كتابة معظم برامج قواعد البيانات العلائقية الحالية بواسطة فرق منفصلة من المبرمجين ثم يتم إعادة استخدامها بواسطة تطبيقات أخرى (Jowannet al., 2016). قواعد البيانات SQL هي أنظمة إدارة قواعد البيانات العلائقية (RDBMS) مثل: Microsoft SQL Server و Oracle Database وMySQL وIBM DB2.

1-1 التوسع في قواعد البيانات العلائقية (Scalability):

يعد تحقيق قابلية التوسع والمرونة تحديًا كبيرًا لقواعد البيانات العلائقية حيث تم تصميم قواعد البيانات العلائقية في فترة يمكن فيها الاحتفاظ بالبيانات صغيرة ومرتبة ومنظمة (Allen, 2015). لقد تم تصميم قواعد البيانات العلائقية للتشغيل على خادم واحد من أجل الحفاظ على تكامل تعيينات الجدول وتجنب مشاكل الحوسبة الموزعة وهذا ما يسمى بالتوسع العمودي (Vertical Scalability). مع هذا التصميم إذا احتاج النظام إلى التوسع، فيجب على العملاء شراء أجهزة مملوكة أكبر وأكثر تعقيدًا وأكثر تكلفة مع مزيد من قوة المعالجة والذاكرة والتخزين. تمثل الترقيات أيضًا تحديًا، حيث يجب أن تمر المؤسسة بعملية استحواذ طويلة، ثم غالبًا ما تجعل النظام في وضع عدم الاتصال لإجراء التغيير فعليًا. يحدث كل هذا بينما يستمر عدد المستخدمين في الزيادة، مما يتسبب في المزيد والمزيد من الضغط وزيادة المخاطر على الموارد الغير المتوفرة بشكل كافٍ.



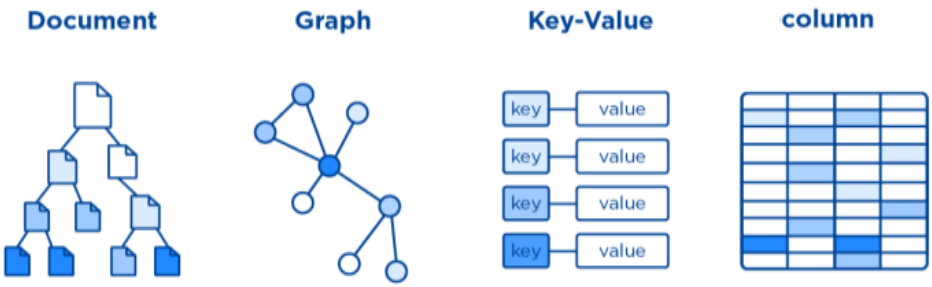
شكل (6): التوسع العمودي لقواعد البيانات العلائقية (Allen, 2015).

2- قواعد البيانات الغير علائقية (Non-RDBMS) أو (NoSQL):

قاعدة البيانات غير العلائقية تقوم فقط بتخزين البيانات بدون آليات واضحة ومنسقة لربط البيانات من مجموعات مختلفة ببعضها البعض. يطلق عليه أيضًا اسم قواعد بيانات NoSQL أو (Not-only

SQL) يستخدم أيضًا كاسم مستعار لأنظمة إدارة قاعدة بيانات الجيل التالي التي تتناول في الغالب بعض النقاط: كونها غير علائقية وموزعة Distributed ومفتوحة المصدر Open-source وقابلة للتوسع أفقيًا Horizontally Scalable. يتم تخزين بيانات NoSQL في واحد من أربعة أنماط لهيكلة البيانات:

- قاعدة بيانات المفتاح-القيمة (Key-Value Database).
- قاعدة بيانات العمود (Column Database).
- قاعدة بيانات المستند (Document Database).
- قاعدة بيانات الرسم (Graph Database).

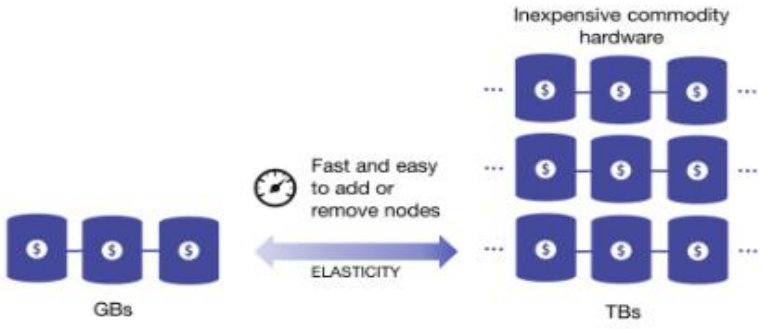


شكل (7): أنواع قواعد البيانات الغير علائقية (Team, n.d.).

تقنيات NoSQL مطلوبة إذا كان التوافر العالي مطلوبًا High Availability التي تتطلبها خدمة الويب. إنها بنية تخزين موزعة على نطاق واسع ولها بنية أساسية مفيدة لنظام إدارة قواعد البيانات. يتم تخزين البيانات الفعلية للقيم الرئيسية في أزواج، أعمدة أو مجموعات من الأعمدة والمستندات والرسوم البيانية. لذلك يتم دعم مفاهيم التكرار Redundancy المختلفة لضمان تجنب الفشل وزيادة توافر Availability أنظمة قاعدة بيانات NoSQL (Kagai, 2019).

2- 1 التوسع في قواعد البيانات الغير علائقية (Scalability):

بدلاً من أن تكون مقيدة بحدود معمارية الخادم الفردي، تم تصميم قواعد بيانات NoSQL على نطاق واسع على الأنظمة الموزعة (عادةً مئات من تيرابايت بدلاً من عشرات الجيجابايت). يمكنهم توسيع نطاقها أفقيًا Horizontal Scalability، مما يعني أنها تعمل على خوادم متعددة تعمل معًا، كل جزء يشارك في الحمل كما هو موضح في (شكل8).



شكل (8): التوسع الأفقي لقواعد البيانات الغير علائقية، (Allen, 2015)

باستخدام هذا النهج، يمكن أن تعمل قاعدة بيانات NoSQL عبر مئات الخوادم وبيتابايت من البيانات ومليارات من المستندات ولا تزال قادرة على معالجة عشرات الآلاف من المعاملات في الثانية. ويمكنه القيام بكل هذا على أجهزة حاسوب عادية غير مكلفة، تعمل في أي بيئة (أي الحوسبة السحابية). فائدة

أخرى هي أنه في حالة فشل عقدة واحدة، يمكن للعقدة الأخرى تحمل عبء العمل، وبالتالي التخلص من نقطة فشل واحدة.

خامساَ: البيانات الضخمة وقواعد البيانات NoSQL:

إن الجيل الجديد من التطبيقات مثل ذكاء الاقتصاد والأعمال، الويب 2.0، الشبكات الاجتماعية وغيرها يتطلب معالجة تيرابايت وحتى بيتابايت من البيانات. يتم تحقيق ذلك عن طريق المعالجة الموزعة وهذا هو أحد الأسباب الرئيسية لقوة شركات الويب مثل جوجل، أمازون، الفيسبوك وتويتر وياهو! (Moko & Asagba, 2021). هناك عدة أسباب لتوزيع المعالجة. من ناحية، البرامج يجب أن تكون قابلة للتوسع Scalability ويجب أن تأخذ الاستفادة من أنظمة متعددة وكذلك معماريات وحدة المعالجة المركزية متعددة النواة. على الطرف الآخر، يجب أن تكون خوادم مواقع الويب موزعة عالميًا لوقت استجابة منخفض وتجاوز الفشل. في الوقت الحالي لا تزال أنظمة RDBMS لها مكانها وحالات الاستخدام ومع ذلك فإن الويب قد غير متطلبات التخزين لأنظمة قواعد بيانات للجيل القادم من التطبيقات. هناك العديد من الدراسات البحثية التي تقول إن قواعد البيانات العلائقية (RDBMS) غير مؤهلة للتعامل مع المعالجة الموزعة التي تستخدم عدد كبير جدا من الخوادم للتعامل مع تطبيقات البيانات الضخمة، نذكر بعض منها.

تسلط الورقة (Kalid et al., 2017) الضوء على الأدبيات التي تتناول القضايا وتجد الحلول من خلال استخدام أنظمة متنوعة للبيانات الضخمة والمقارنة بينها في إدارة قواعد بيانات NoSQL: وهي BigTable وDynamoDB وCassandra. يوضح العمل أيضًا أن الشركات الكبيرة التي لا تستخدم أنظمة تخزين صحيحة لا يمكنها تخزين وإدارة البيانات الضخمة. ووجدت الدراسة أن BigTable من جوجل وDynamoDB من أمازون مهمان في التعامل مع البيانات الضخمة ومربحان جدا وCasandra من فيسبوك مزيج من كلا النظامين.

درس الباحثون في الورقة (Razu Ahmed et al., 2018) قواعد بيانات NoSQL لمعالجة البيانات الضخمة، بما في ذلك القضايا المتعلقة بالمعاملات والهيكلية وسلط العمل أيضا الضوء على اتجاهات البحث والتحديات المتعلقة بمعالجة البيانات الضخمة التي اعتقدت الدراسة أن المعلومات كانت مدهشة لمراجعة الأدبيات حول قاعدة بيانات NoSQL الضخمة للبيانات، بما في ذلك البيانات الهيكلية، المشاكل وطرق جمع المعلومات المفيدة التي يتم قياسها في الوقت الفعلي.

أكد الباحثون في دراستهم (Kovacheva, 2021) أن قاعدة بيانات NoSQL تأتي كخيار أفضل لتخزين كميات كبيرة من ملفات بيانات غير منسقة من أنواع مختلفة، لأنه يمكن تغيير نوع البيانات وليست هناك حاجة لتعريفها مقدما. الأشياء الإيجابية الأخرى لقاعدة البيانات غير العلائقية هي التخزين السحابي وإمكانية التوسع السريع Scalability. عند العمل مع البيانات الضخمة، فإن القدرة على التعامل مع أنواع البيانات المختلفة مهم جدًا وكذلك خيار امكانية تغيير النسق الأصلي لهيكل البيانات، والذي يعطي مرونة كبيرة Elasticity لقواعد بيانات الغير علائقية NoSQL. تواجه قواعد البيانات العلائقية صعوبات في هذا الجزء من العمل مع البيانات شبه المنسقة والمنسقة.

تناقش الورقة (Čerešňák & Kvet, 2019) هياكل تخزين قاعدة البيانات والمفاهيم والاختلافات بالإضافة إلى عمليات الإدراج والتحديث والحذف والاختيار. تم اختيار مجموعة من قاعدة البيانات غير العلائقية للتقييم، وتمت مقارنة أدائها مع قاعدة بيانات العلائقية، وخلص إلى أن أداء قاعدة البيانات العلائقية أقوى في البيانات الصغيرة، مقابل البيانات الكبيرة يفضل استخدام قواعد البيانات غير العلائقية.

يقارن البحث (Jose & Abraham, 2020) كفاءة الاستعلامات والأوامر المختلفة في كل من قاعدة البيانات العلائقية وأنظمة NoSQL باستخدام مجموعتين مختلفتين من البيانات ذات الأحجام المختلفة لتسليط الضوء على مزايا قاعدة بيانات NoSQL وعلى قاعدة البيانات العلائقية وخاصة عند التحقيق في البيانات الضخمة. خلصت الدراسة إلى أن الفروق في وقت التنفيذ لوحظت وأظهرت باستخدام الرسوم البيانية. في جميع الحالات، يمكن الاستدلال على أن قواعد بيانات NoSQL تفوق أداء قواعد البيانات العلائقية.

تقارن الورقة (Palanisamy & SuvithaVani, 2020) بين مفاهيم NoSQL وقواعد البيانات العلائقية والقيود الموجودة لديهم وتناول مزايا وأنواع قواعد بيانات NoSQL. من هذه المقارنات، كان الاستنتاج هو أن قاعدة البيانات العلائقية قد تكون خيارًا رائعًا إذا كان لديك بيانات منسقة وتحتاج إلى قاعدة بيانات ارتباطية تقليدية. أما إذا كانت لديك بيانات غير منسقة و/أو منسقة مع إمكانية النمو السريع، فستكون NoSQL هي الخيار الأفضل.

قامت الورقة (Sirish Shetty & Akshay, 2019) بتحليل مقارن لاكتشاف أن كانت قواعد البيانات غير العلائقية تعمل بشكل أفضل من قاعدة البيانات العلائقية. بالنظر إلى مجموعة البيانات المستخدمة للتجربة، فإن الفرضية هي أن MongoDB (نظام إدارة قواعد بيانات غير علائقية) يعمل بشكل أفضل فقط للإدراج والحذف والاختيار، بينما لا تزال Oracle (نظام إدارة قواعد بيانات علائقية) تعمل بشكل أفضل مع عملية التحديث في البيئة المتاحة.

تطوير تطبيقات البيانات الضخمة وخاصة تطبيقات الويب أصبح ذات أهمية متزايدة في السنوات الأخيرة. في الواقع يعتمد عدد متزايد من المؤسسات عبر مجموعة متنوعة من القطاعات على المعلومات المستسقاة من كميات هائلة من البيانات. ومع ذلك عند التعامل مع كميات كبيرة من البيانات، تكون أنظمة ومنصات قواعد البيانات العلائقية التقليدية غير فعالة، مع أوقات استجابة بطيئة ونقص في قابلية التوسع والأداء والدقة لمعالجة تعقيد مشاكل البيانات الضخمة (Oussous et al., 2018). أيضا صعوبة تعامل قاعدة البيانات العلائقية مع البيانات غير المنسقة. بينما تستفيد قواعد البيانات العلائقية من SQL، التي تم تحسينها للعمل مع البيانات المنسقة، يمكن أن تصبح SQL معقدة للغاية عند التعامل مع البيانات غير المنسقة (Palanisamy & SuvithaVani, 2020).

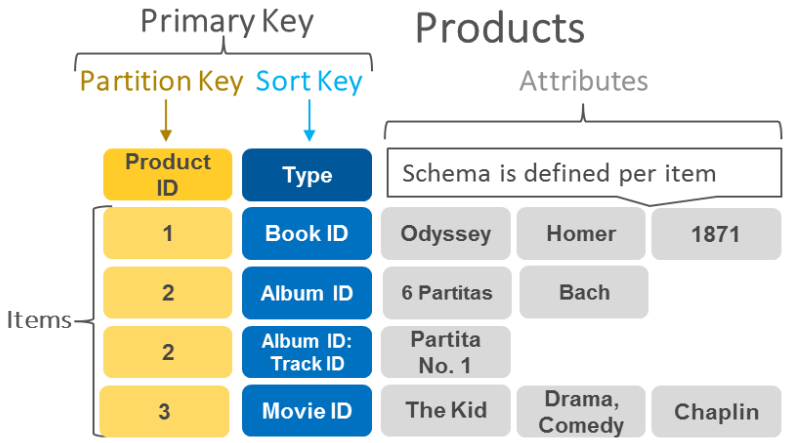
تعديل مخطط البيانات وتطويره: مع إنشاء التطبيق وتطويره، لا يمكن تعديل مخطط قاعدة البيانات العلائقية بمرور الوقت، ولا يمكن أن تستوعب التعديلات أنواعًا مختلفة من البيانات (Sahatqija et al., 2018). أصبح من الواضح أن قواعد البيانات العلائقية التقليدية كانت غير متوافقة مع متطلبات مجموعة متنوعة من التطبيقات وبياناتها. أدى ذلك إلى إدخال تقنية جديدة تُعرف باسم NoSQL بين المنظمات التجارية والشركات.

سادساً: أنواع قواعد البيانات NoSQL:

تقوم أنظمة NoSQL بتخزين واسترداد البيانات بتنسيقات مختلفة. هناك طرق مختلفة لتصنيف قواعد بيانات NoSQL ينتج عنها فئات وفئات فرعية مختلفة. اقترح بن سكوفيلد التصنيف الأكثر عمومية وشائعة الاستخدام في عرضه التقديمي في مؤتمر 2010 CodeMash والذي تضمن مقارنة موجزة لفئات قواعد بيانات NoSQL المختلفة وقواعد البيانات العلائقية (Scofield, 2010). يمكن وصفها بإيجاز على النحو التالي.

1-6 قواعد بيانات Key-Value:

هذه طريقة قوية وفعالة لأنظمة قواعد بيانات NoSQL. يمكن عادةً تخزين البيانات كأزواج (مفتاح – قيمة) بحيث تكون البيانات مخزنة في أزواج مصفوفة أو جداول تجزئة على هيئة أبجدية رقمية بمفاتيح وقيم فريدة مثل JSON أو BLOB أو متغير سلسلة (Moniruzzaman & Hossain, 2013). يجب أن تحتوي الجداول المستقلة عمودين، مع عمود واحد يتم الاحتفاظ به كمفتاح أساسي (PK) بينما يعمل الآخر كحامل القيم المنطقية. يُعرف هذا الإجراء باسم مخزن الصف ويسمى أيضًا Tuple Store لأن البيانات التي تنتمي إلى سجل واحد يتم تخزينها معًا. تتضمن بعض الأمثلة على نظام إدارة قواعد البيانات ذات (القيمة – المفتاح) Redis Riak, HyperDex, Amazon DynamoDB Dynomite, Aerospike, Voldemorte, CouchBase, FoundationdB, MemcacheDB, OrientDB, Azure Table Storage (ATS), Berkeley DB etc.



شكل (9): Key-Value Stores Database (Amazon Web Services, 2019).

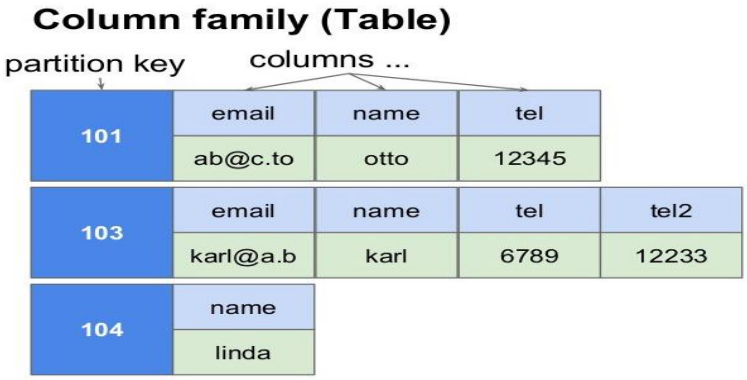
فيما يلي بعض حالات استخدام قواعد بيانات Key-Value:

- تخزين واسترداد معلومات جلسة المستخدم لتطبيق الويب
- تخزين ملفات تعريف المستخدم وتفضيلاته
- تخزين بيانات عربة التسوق الخاصة بالمستخدم للمتاجر عبر الإنترنت.

2-6 قواعد بيانات Column-Oriented:

هذه النوع من قواعد البيانات ينظم البيانات في جداول، تشبه إلى حد كبير نظام RDBMS. على الرغم من أنهم يخزنون المحتوى حسب الأعمدة بدلاً من الصفوف. تخزن البيانات في بنية موزعة ويرتبط كل مفتاح بسمة واحدة أو أكثر. تُعرف قواعد البيانات هذه أيضًا باسم BigTable أو قواعد البيانات العمودية. تخزن قاعدة بيانات العمود بياناتها بطريقة يمكن تجميعها بسرعة مع نشاط إدخال / إخراج أقل. إنه يوفر أداءً عاليًا للغاية وبنية قابلة للتطوير بدرجة كبيرة. يمكن أن تكون مفيدة لتخزين البيانات واستخراج البيانات وتطبيق التحليلات (Mukherjee, 2019). أمثلة على هذا النظام:

Cassandra, Apache Hadoop Hbase, Hypertable, Accumulo, Druid, Vertica, etc.



شكل (10): Column-Oriented Database (freeCodeCamp.org, 2021).

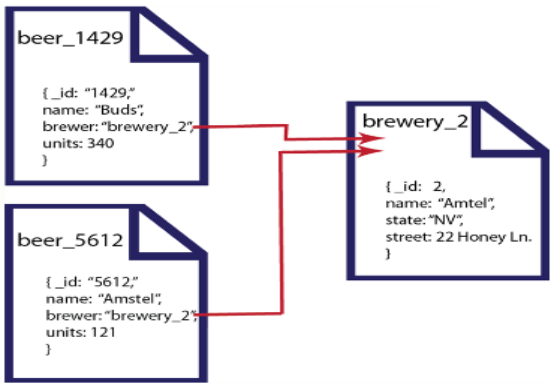
فيما يلي بعض حالات استخدام قواعد بيانات Column-Oriented:

- منصات التدوين.
- نظم إدارة المحتوى.
- الأنظمة التي تحافظ على العدادات.
- الخدمات التي انتهت صلاحية استخدامها.
- الأنظمة التي تتطلب طلبات كتابة مكثفة.

3-6 قواعد بيانات Document-Oriented:

يتم تخزين البيانات القائمة في هذا النوع كمستندات ويتم تنظيمها كمجموعة من المستندات. هذه الوثائق مرنة جدا بحيث يمكن أن يحتوي كل مستند على أي عدد من الحقول. هذه مصممة لتخزين واسترجاع وإدارة المعلومات الموجهة للمستندات. توفر مخازن المستندات أداءً رائعًا وقابلية التوسع الأفقي. تكون هذه المستندات ذات تنسيقات قياسية مثل XML وPDF وJSON وما إلى ذلك. في قاعدة البيانات العلائقية، سيكون السجل داخل نفس قاعدة البيانات نفس حقول البيانات ويتم الاحتفاظ بحقول البيانات غير المستخدمة فارغة، ولكن في حالة تخزين المستندات، قد يحتوي كل مستند بيانات متشابهة وغير متشابهة. تتم معالجة المستندات الموجودة في قاعدة البيانات باستخدام مفتاح فريد يمثل ذلك المستند. قد تكون هذه المفاتيح عبارة عن سلسلة بسيطة أو سلسلة تشير إلى URI أو المسار (Mukherjee, 2019). أمثلة على ذلك:

MongoDB, Apache CouchDB, Elasticsearch, MarkLogic, CouchBase, RethinkDB, Clusterpoint, DocumentDB, OrientDB, etc.



شكل (11): Document-Oriented Database (Moot, 2019)

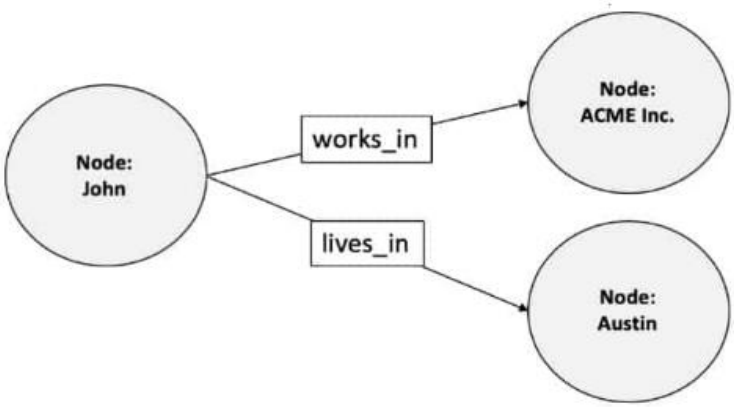
فيما يلي بعض حالات استخدام قواعد بيانات Document -Oriented:

- منصات التجارة الإلكترونية.
- أنظمة إدارة المحتوى.
- منصات التحليلات.
- منصات التدوين.

4-6 قواعد بيانات Graph:

تطبق قواعد البيانات هذه نظرية الرسم البياني لعلوم الكمبيوتر لتخزين واسترجاع البيانات. تركز على الترابط بين أجزاء مختلفة من البيانات. يتم تصور وحدات البيانات كعقد ويتم تحديد العلاقات فيما بينها من خلال الحواف التي تربط العقد. يمكن لقواعد بيانات الرسم البياني التنقل في ملايين السجلات باستخدام تقنية التقارب الخالي من الفهرس، حيث تتكون كل عقدة من مؤشر مباشر يشير إلى العقدة المجاورة. يمكن استخدام قواعد بيانات الرسم البياني لمجموعة متنوعة من التطبيقات مثل تطبيقات الشبكات الاجتماعية والمعلوماتية الحيوية وإدارة المحتوى والأمن والتحكم في الوصول وإدارة الشبكة والسحابة وما إلى ذلك (Mukherjee, 2019). مثال على هذا النوع:

Neo4J, OrientDB, Giraph, AllegroGraph, MarkLogic, InfiniteGraph, Virtuso, Stardog, etc.



شكل (12): Graph-based Database

فيما يلي بعض حالات استخدام قواعد بيانات Graph-based:

- كشف الاحتيال.
- البحث القائم على الرسم البياني.
- عمليات الشبكة وتكنولوجيا المعلومات.
- الشبكة الاجتماعية.

سابعاَ: مستقبل قواعد البيانات مع NoSQL:

في الوقت الحاضر كل مؤسسة تتعامل مع كمية ضخمة من السجلات من مجموعة متنوعة من المصادر بسرعات كبيرة. قواعد البيانات العلائقية في بعض الأحيان غير فعالة بالنسبة للشركات لمعالجة والتحقيق في الكم الهائل من البيانات متعددة الأوجه وغير منسقة. نظرًا لأن NoSQL عبارة عن قواعد بيانات ذات نموذج أقل-مخطط أو بدون-مخطط، فهي فعالة للغاية للتعامل مع كمية كبيرة من البيانات. في الغالب تغرق جميع المؤسسات بكميات كبيرة من البيانات كل ثانية من مجموعة متنوعة من المصادر التي يتم استردادها من الإنترنت. باستخدام هذه البيانات، يمكن إجراء التحقق من صحة البيانات من أجل الاستخدام الأفضل أو الأكثر فاعلية لها وللقيام بالتنبؤات المستقبلية. استخدام تحليلات NoSQL في الوقت الحقيقي هو استجابة أسرع بكثير، يمكن أن تكون الأوقات أقل من دقيقة لجميع الاستفسارات المعقدة. في قواعد بيانات SQL، ترتبط جداول قواعد بيانات SQL بالمفاتيح الأساسية الخارجية بينما في NoSQL يمكن استخدام نموذج بيانات مختلف للتعامل مع كمية هائلة من البيانات. عندما يريد المستخدم استخدام أزواج المفتاح والقيمة، يمكن استخدام قواعد بيانات المفتاح-القيمة، ويمكن استخدام قواعد بيانات الرسم لمؤشرات البيانات، ويمكن إضافة المزيد من العقد إلى المجموعة وهو قابل للتطوير بسهولة بدلاً من استخدام الآلات الكبيرة. يمكن استخدام NoSQL في العديد من التطبيقات المتقدمة. NoSQL تجعل التعلم الآلي أسرع، الآلاف من المعاملات كل ثانية يمكن مراقبتها بسهولة باستخدام NoSQL، لذا فهو مفيد عند التعامل مع عمليات الكشف عن الاحتيال في المعاملات المصرفية. أيضًا، كميات كبيرة من البايتات من البيانات التي يتم إنشاؤها من الشبكات الاجتماعية، وبيانات المناخ، والصور المبتكرة، ومحادثات البيانات، وهذه هي البداية فقط. بالنسبة لهذه السيناريوهات، توجد أنواع عديدة من العناصر، على سبيل المثال الصور والفيديو والصوت يمكن دمجها وتخزينها في قاعدة البيانات. قواعد بيانات NoSQL المختلفة قابلة للتوسع لاستيعاب المعلومات. مع نمو البيانات الضخمة، فإن عملية تطوير NoSQL تنمو بشكل أسرع بين جميع مؤسسات الويب والمطورين. تخدم قواعد بيانات NoSQL مؤسسات الويب بشكل جيد للتأثير على أهدافها التحليلية في عصر تضخم البيانات المتسارع. نشاطات NoSQL سريعة جدا بحيث توفر معالجة خالية من الأخطاء مقارنة بقواعد البيانات العلائقية التقليدية. تعالج NoSQL مشاكل المؤسسات التي تتطلع إلى تنظيم متطلبات العملاء المتنوعة والأسواق المتنامية المتنوعة. تتفوق هذه التقنية على التوقعات في معالجة غالبية البيانات غير المنسقة حيث نرى أن معظم المؤسسات تنضم إلى أشهر المنتجات مفتوحة

المصدر مثل Cassandra وMongoDB وRedis، إلخ. توفر NoSQL أيضًا بديلاً أقل تكلفة لتحميل البيانات واسترجاعها.

الخاتمة:

قمنا في هذه الورقة البحثية بمراجعة ودراسة البيانات الضخمة وطرق التعامل مع الزيادة السريعة بشكل استثنائي في حجم البيانات وأوضحنا أن قواعد البيانات العلائقية تواجه صعوبات عند التعامل معها بسبب حجمها والبيانات الغير منسقة. لذلك نحتاج إلى طرق وهيكليات جديدة للتعامل مع البيانات الضخمة وغير المنسقة. قواعد بيانات NoSQL هي قواعد بيانات غير علائقية تقوم بتخزين البيانات بطريقة مختلفة عن العلاقات الجدولية المستخدمة في قواعد بيانات SQL. في حين أن قواعد بيانات SQL هي الأفضل استخدامًا للبيانات المنسقة، فإن قواعد بيانات NoSQL مناسبة للبيانات المنسقة وشبه المنسقة وغير المنسقة. هذا العمل استعرض أيضًا أنواع هيكليات قواعد بيانات NoSQL المتنوعة للبيانات الضخمة المرتبطة بها وأهميتها واستخداماتها.

قائمة المراجع:

[1] Aparna, M. (2021). BIGDATA AND ITS APPLICATIONS. International Journal for Innovative Engineering and Management Research, 288–297.

[2] Moko, A., & Asagba, P. (2021). Big Data and NoSQL Databases Architecture: A review.ResearchGate.https://www.researchgate.net/publication/356716269_Big_Data_and_NoSQL_Databases_Architecture_A_Review.

[3] Mohanty, S., Nathrout, K., Barik, S., & Prof, A. (2015). A study on evolution of data in traditional RDBMS to big data analytics. ResearchGate. https://doi.org/10.17148/IJARCCE.2015.41049.

[4] Kovacheva, M. (2021). Storing Big Data in NoSQL databases compared to SQL - advantages and problems, 10th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education, Unwe, Sofia, Bulgaria, 2021.

[5] Meier, A., & Kaufmann, M. (2019). SQL & NoSQL databases: Models, Languages, Consistency Options and Architectures for Big Data Management. Springer Vieweg.

[6] Jowan, S. A., Swese, R. F., Aldabrzi, A. Y., & Shertil, M. S. (2016). Traditional RDBMS to NoSQL database: new era of databases for big data. *Journal of Humanities and Applied Science (JHAS),* 29(29), 83-102.

[7] Kalid, S., Syed, A., Mohammad, A., & Halgamuge, M. N. (2017). Big-data NoSQL databases:A comparison and analysis of “Big-Table”, “DynamoDB”, and “Cassandra”. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA) (89–93).

[8] Razu Ahmed, Md., Arifa Khatun, Mst., Asraf Ali, Md., & Sundaraj, K. (2018). A literature review on NoSQL database for big data processing. International Journal of Engineering & Technology, 7(2), 902. https://doi.org/10.14419/ijet.v7i2.12113.

[9] Čerešňák, R., & Kvet, M. (2019). Comparison of query performance in relational a non-relation database. Transportation Research Procedia, 40, 170–177. https://doi.org/10.1016/j.trpro.2019.07.027.

[10] Jose, B., & Abraham, S. (2020). Performance Analysis of NoSQL and Relational Databases with MongoDB and MySQL. Materials Today: Proceedings, 24(3), 2036–2043. https://doi.org/10.1016/j.matpr.2020.03.634.

[11] Palanisamy, S., & SuvithaVani, P. (2020). A survey on RDBMS and NoSQL Databases MySQL vs MongoDB. 2020 International Conference on Computer Communication and Informatics (ICCCI).

[12] Sirish Shetty, B., & Akshay, K. (2019). Performance Analysis of Queries in RDBMS vs NoSQL. 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). https://doi.org/10.1109/icicict46008.2019.8993394.
[13] Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. Journal of King Saud University - Computer and Information Sciences, 30(4), 431–448.
[14] Francis, K. K. (2019). NoSQL Databases for Big Data Management: Review and Application in Mobile Commerce. https://doi.org/10.13140/RG.2.2.10239.87204.
[15] Sahatqija, K., Ajdari, J., Zenuni, X., Raufi, B., & Ismaili, F. (2018). Comparison between relational and NOSQL databases. 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). https://doi.org/10.23919/mipro.2018.8400041.
[16] Scofield, B. (2009). "NoSQL –Death to Relational Databases(?)". Presentation given at RubyConf. [Slide show]. PPT. https://www.slideshare.net/bscofield/nosql-death-to-relational-databases.
[17] Moniruzzaman, A. B. M., & Hossain, S. A. (2013). "Nosql database: New era of databases for big data analytics-classification, characteristics and comparison". arXiv preprint arXiv:1307.0191.
[18] Mukherjee, S. (2019). The Battle between NoSQL Databases and RDBMS. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3393986.
[19] Allen, M. (2015). Relational databases are not designed for scale. MarkLogic. https://www.marklogic.com/blog/relational-databases-scale/.
[20] What Is a Key-Value Database? (2019). Amazon Web Services, Inc. https://aws.amazon.com/nosql/key-value/.
[21] Moot, S. (2019). Document-oriented databases versus Relation databases. Medium. https://medium.com/@samual.r.moot/document-oriented-databases-versus-relation-databases-929350c28d7c.
[22] Mchergui, A., Hajlaoui, R., Moulahi, T., Alabdulatif, A., & Lorenz, P. (2023). Steam computing paradigm: Cross-layer solutions over cloud, fog, and edge computing. IET Wireless Sensor Systems. https://doi.org/10.1049/wss2.12051.
[23] Team, M. (n.d.). Exploring the NoSQL family. https://morioh.com/a/b7adc58a5a4e/exploring-the-nosql-family.
[24] freeCodeCamp.org. (2021, July 9). The basics of NoSQL databases—and why we need them. https://www.freecodecamp.org/news/nosql-databases-5f6639ed9574/.