التحليل المرئي للبيانات الضخمة في المكتبات باستخدام Hadoop و Tableau

د. أيمن حبيب حمادي
أستاذ مساعد - كلية الاعلام – الجامعة اللبنانية
دكتوراه في علوم المعلومات والمكتبات
ah021@live.aul.edu.lb

المستخلص:

قدّم هذا البحث مقاربة فكرية حول عناصر تحليل البيانات الضخمة في المكتبات وادوات التعامل معها، لاكتشاف قابلية المكتبات لتطبيق تحليلات البيانات الضخمة في إدارة خدماتها وأنشطتها. وقد اعتمد البحث المنهج التجريبي من خلال العمل التطبيقي القائم على تحليل بيانات موقع التواصل الاجتماعي فيسبوك على خدمة الإعارة خلال جائحة كورونا في مكتبة السبيل العامة في لبنان باستخدام تقنية تحليل البيانات الضخمة "Hadoop"، وقد تم تصوير النتائج بشكل مرئي من خلال برنامج Tableau.

وكان من أهم نتائج تحليل بياناتها الضخمة، وجود ارتباط كبير بين عدد المتفاعلين على فيسبوك ومقاييس معينة مع زيادة استخدام المكتبة بشكل عام وخدمة الإعارة بشكل خاص، مما يدل على أهمية استخدام شبكات التواصل الاجتماعي في المكتبات وخصوصا" في الأوقات التي تكون فيها المكتبة غير قادرة على أن تقوم بدورها بشكل اعتيادي وخاصة في ظل جائحة كورونا.

ثم ذيلت الأطروحة بخاتمة شكلت موجزا" عنها، بالإضافة إلى توصيات عامةٍ تعلقت بالبيانات الضخمة في المكتبات، وقد توقع الباحث أخيرا" أن التحولات والتطورات التي تشهدها المكتبات وخاصة مع ازدياد حجم بياناتها مع مرور السنين، فانهم سيكونون مضطرين البدء في خوض غمار تطبيق البيانات الضخمة والاستفادة من أهميتها وفق الإمكانيات المتاحة لهم ولو بشكل مرحلي بسبب التغييرات التكنولوجية السريعة.

الكلمات المفتاحية:
البيانات الضخمة؛ تطبيقات البيانات الضخمة؛ التحليل المرئي للبيانات؛ هدوب - تابلوه

تاريخ استلام البحث: 2023/10/26 تاريخ إجازة البحث: 2023/11/17

أولا: الاطار المنهجي

تمهيد:

تشكل البيانات الضخمة Big Data اتجاه جديد يعكس تلك البيانات التي يتم إنتاجها من قبل مستخدمي الوسائط الإلكترونية، والتي تتسم بصفة التنوع والتعددية، والتي تصبح معها أدوات المعالجة التقليدية غير قادرة على التعامل معها، فهي تتجاوز قدرة معالجة نظم قواعد البيانات التقليدية.

وتعبّر البيانات الضخمة في مفهومها المبسّط عن كمية هائلة من البيانات المعقدة التي تقاس بوحدات قياس عالية جدا" مثل التيرابايت والبيتابايت الذي يساوي تريليون غيغابايت؛ وهذه البيانات كما يشير (الحكيم،2017) هي مجموعة ضخمة من البيانات المتداخلة بشدة كالتغريدات على توتير، والدردشة على الفيسبوك، والرسائل النصية، والاعجابات بمنتج معين، ومشاركة الحالة أو فيديو مع الأصدقاء، وحجم تداول الأسهم، ومؤشرات وأخبار الطقس وغيرها، مما يصعب معالجتها وإدارتها باستخدام أداة واحدة من أدوات إدارة قواعد البيانات أو بالطرق التقليدية لمعالجة البيانات. وهي تنمو بوتيرة متسارعة للغاية وفي طريقها للمزيد، كما وفي كل يوم تضاف أدوات أكثر وأكثر لإدخال البيانات كالهواتف الذكية، ومستشعرات الفضاء، ومحددات المواقع الجغرافية، وقارئات التأمين الاجتماعي، وقارئات بيانات السير والمرور، ومدخلات معلومات التلاميذ والطلاب والموظفين والعمال في العالم.

وقد غدت البيانات الضخمة قضية محورية تدخل في إطار اهتمام جميع العلوم، على حد سواء، اجتماعية كانت أو تطبيقية، ولم تعد قضية حصرية مصونة على اهتمام العلوم الطبيعية، فضلا عن وجود قدر من الموضوعية والدقة حولها، إلى جانب الاعتقاد بأن البيانات الضخمة تقدم شكلا أعلى من الذكاء والمعرفة (Boyd & Crawford،2012).

مع التطور السريع في تكنولوجيا الاتصالات والمعلومات، ومتابعه من تطوير برامج، وأجهزة تنتج بيانات، بنحو سريع ومستمر، كالأجهزة الذكية، وأجهزة الاستشعار، فضلا عن انتشار تقنيات الجيل الثاني للويب web 2.0، وبخاصة الشبكات الاجتماعية على اختلاف مستوياتها، وبالإضافة الى الاستخدام المتزايد لتطبيقاته، أدى ذلك إلى تضخم ملموس في حجم

البيانات، والتي بات من الصعب التعامل معها، باستخدام أنظمة إدارة قواعد البيانات التقليدية، وإحصاءات سطح المكتب، حيث يتطلب الأمر بدلاً من ذلك "برامج متوازية واسعة النطاق تعمل على عشرات أو مئات أو حتى آلاف الخوادم" تقوم بإدارة المجموعة التي تعرّف بـ "البيانات الضخمة".

ومما زاد الأمر تعقيدا"، التحديث الآني لهذه البيانات بشكل أكبر بكثير من السابق، فضلا" عن أن معظم هذه البيانات باتت غير مهيكلة. وقد أصبحت إدارة هذه البيانات وتحليلها التحدي الأكبر للعديد من المؤسسات، وعلى رأسها المكتبات، ومؤسسات المعلومات، بوصفها المؤسسات الأكبر اهتماما بقضية البيانات والمعلومات.

تواجه مؤسسات المعلومات ولا سيما المكتبات حتمية التعامل مع البيانات الضخمة التي ظهرت نتيجة للطفرة الرقمية المتزايدة حجما ونوعا بسبب استخدام هذه المكتبات لتكنولوجيا المعلومات المختلفة، لذلك هي معنية بالتعامل مع البيانات الضخمة في العديد من الأنشطة؛ فهناك اقتناء للمصادر المختلفة، واشتراكات في قواعد البيانات على الإنترنت، وتحليلات إحصائية للبيانات المستخدمة، وأدوات لإدارة البيانات البحثية.

تجمع المكتبات كمية كبيرة من البيانات، مثل الكتب والمقالات والتقارير البحثية، سواء كانت ورقيّة أو الإلكترونية. وكانت هذه المجموعات المكتبية تتوفر للباحثين أو عامة المستخدمين للعثور على المعلومات الضرورية التي يحتاجونها، حتى أصبحت هذه البيانات كبيرة جدا"، وقد وطرأ عليها تغيير من حيث الشكل مما أثر على فعالية الاستخدام. على الرغم من أن الكثير من بيانات المكتبة قد تم تحويلها إلى شكل رقمي، إلا أن هذه البيانات لم يتم تنقيبها وتحليلها من خلال استخدام تقنيات البيانات الضخمة. على سبيل المثال، المكتبة الجيولوجية في الصين والتي لديها 710.000 سجل من كتب ومجلات إلكترونية وتسجيلات وخرائط ومذكرات، ألا أن معظم هذه السجلات تبقى فردية ومعزولة عن الويب، مما يتطلب دراسة مفصلة حول كيفية عرض البيانات بشكل فعال من خلال استخدام تكنولوجيا البيانات الضخمة. .(Li et al, 2017)

وبالتالي، فقد باتت المكتبات ومؤسسات المعلومات تحتل مكانة مهمة في الواقع الجديد ومتغيراته، كما من المتوقع أن يصبح اختصاصي المكتبات والمعلومات، أخصائيين للبيانات،

وخبراء في إيجاد وتقييم وتحليل البيانات، وتحويلها إلى معلومات متناغمة، تؤدي إلى نتائج تمكن من اتخاذ القرارات الاستراتيجية، وذلك في ضوء فهم متكامل لآليات استخدام البيانات الضخمة لحل المشاكل اليومية.

مع ظهور مفهوم البيانات الضخمة وتنوع خواصها أصبحت الحاجة لظهور تقنيات جديدة تجعل من عملية تخزين ومعالجة هذه البيانات أمرا" ممكنا"، وأصبحت هناك حاجة ملحة للتأكد من صحة الخدمات المقدمة للمستخدمين. وقد ظهر العديد من الأدوات لتخزين البيانات الضخمة ومعالجتها اعتمدت بمعظمها على نظام Hadoop، فإسم Hadoop أصبح مرادفا للبيانات الضحمة، "فهو برنامج مفتوح المصدر بإطار تخزين موزع لمجموعات كبيرة من البيانات في تجمعات حواسيب (computer clusters)، مما يعطيه خاصية التوسعية (scalability) والاعتمادية (reliability) والعمل دون القلق من فشل الأجهزة. ولديه المقدرة على تخزين بيانات كبيرة من جميع أنواع البيانات وتحليلها، مع القدرة على التعامل مع وظائف متزامنة لا حدود لها تقريبا" (Tome, 2012).

يجدر القول، إن استثمار ذلك الكم الضخم من البيانات وتحليله والاستفادة من نتائجه في اتخاذ القرارات ووضع الخطط التنموية كبناء الاستراتيجيات المستقبلية، يعد ركيزة أساسية في اقتصاديات الدول، فقد أكدت لجنة البيانات الضخمة التابعة لمؤسسة تيك أمريكا (TechAmerica) أن البيانات الضحمة تنطوي على إمكانية تغيير الحكومة بل والمجتمع نفسه (المجلس الاقتصادي والاجتماعي بالأمم المتحدة، 2014).

وفي ضوء ما سبق، سعى هذا البحث الى اختبار منصة للبيانات الضخمة (Cloudera) Hadoop في مكتبة السبيل العامة في لبنان لتحليل البيانات شبه المنظمة وتمثيلها صوريا" عبر برنامج Tableau، على سبيل المثال بيانات وسائل التواصل الاجتماعي وبيانات وسجلات الإعارة وأعداد المستفيدين والتسجيلات الببليوغرافية في الفهارس وغيرها. وحتى يتيسر البحث التطبيقي، فقد تناول هذا البحث تأثير تحليل البيانات الضخمة من شبكة التواصل الاجتماعي فيسبوك على خدمة الإعارة في المكتبة خلال جائحة كورونا.

أهمية البحث:

تأتي أهمية هذا البحث في كون موضـوعها يتناول جانبا مهما" في أنشـطة المكتبة وهو التعامل مع البيانات وتحديدا" في اسـتخدام وتحليل البيانات الضـخمة، والتي تجد المكتبات نفسـها أمام تحد كبير في اسـتخدامها والاسـتفادة منها بكفاءة، وهذا ما يتطلب من المكتبة مواجهته بإيجاد كل متطلبات التعامل معها. وتتجلى أهمية البحث الحالي في معرفة فرص تطبيق هذه التكنولوجيا في المكتبات والتي تساعد في تحسين عملية اتخاذ القرارات عبر اعداد وتجربة منصة متخصصة في التعامل مع البيانات الضخمة. فمن خلال تحديد مصادر توافر البيانات الضخمة، وتحديد نتائج تحليل البيانات الضخمة عبر هذه المنصة سوف يتسنى للمعنيين تأطير أنشطة الاستفادة من البيانات الضخمة في تطوير المكتبات.

أهداف البحث:

انطلق البحث الحالي من هدف رئيسي يرتكز على معرفة آلية تحليل البيانات في مكتبة السبيل العامة باستخدام تقنيات البيانات الضخمة ونظام متخصص لتحليل هذه البيانات، وقد تفرعت أهداف البحث إلى عدة أهداف فرعية، كالتالي؛

- اكتشاف فوائد البيانات الضخمة بالنسبة للمكتبات.
- استكشاف تقنيات وأدوات تحليل البيانات الضخمة التي يمكن تطبيقها في المكتبات.
- التعرف على المصادر التي توفر البيانات الضخمة في المكتبات وتحديد المصادر التي يمكن الاستفادة منها.
- تسليط الضوء على منصة «Hadoop»، وآليات تطبيقها في المكتبات.
- الآثار والحلول العملية للمكتبات عند تطبيق إطار نظام للبيانات الضخمة وتحليلها.

منهج البحث:

بسبب طبيعة هذا البحث الذي يحاول التعريف بمصطلحي البحث: البيانات الضخمة، ومنصة «Hadoop» وتسليط الضوء عليهما، واقتراح بعض النماذج التنبؤية والأدوات الأخرى

المساعدة في تحويل البيانات الضخمة في مكتبة السبيل العامة إلى قيمة مضافة بواسطة نظام «Hadoop» وانطلاقا من أهداف البحث واسئلته الذي يسعى للإجابة عنها، فإن المنهج المستخدم في البحث هو منهج البحث التجريبي حيث يعتبر هو الأقرب لحل المشاكل بالطريقة العلمية وذلك بهدف تجربة المنصة الخاصة بHadoop في المكتبة.

تمثلت "بمكتبة السبيل العامة" كمجتمع دراسة لتحليل بيانات مكتبتها من خلال منصة Hadoop، وذلك لقدرة الباحث على تحصيل بيانات مكتبية تتناسب مع النوعية التحليلية للنظام الذي يستخدم أدوات تخزين ومعالجة البيانات الموزعة كتلك الموجودة في مراكز البيانات.

هذا البحث يحتاج إلى بيانات مكتبية كالبيانات الببليوغرافية وبيانات الاعارة وغيرها من البيانات المتعلقة بمواقع التواصل الاجتماعي وبيانات اقفال المكتبة خلال جائحة كورونا، وقد تم اختيار مكتبة السبيل العامة كإحدى المكتبات العامة الرائدة في لبنان.

حدود البحث:

الحدود الموضوعية: يرتكز البحث على معرفة آليات التعامل مع البيانات الضخمة من خلال تحليلها معالجتها وفرص تطبيق هذه التقنيات في المكتبات اللبنانية.

الحدود المكانية: اقتصر البحث على مكتبة السبيل العامة في بيروت.

الحدود الزمنية: يتناول هذا البحث البيانات الضخمة المتاحة في قاعدة بيانات مكتبة السبيل العامة منذ عام 2018 حتى عام 2021.

مصطلحات البحث:

البيانات الضخمة (Big data): تلك البيانات آنية التحديث، وكثيفة الإنتاج، والتي يتجاوز حجمها قدرة أنظمة قواعد البيانات في المعالجة، والتي تتطلب طرق بديلة لمعالجتها، للحصول على أقصى إفادة منها. (قناوي، 2022)

تطبيقات البيانات الضخمة (Big data applications): مجموعة من البرامج والأدوات التي تتيح معالجة كميات هائلة من البيانات والمعلومات، لاستخراج القيمة المضافة، التي يمكن أن تستفيد منها المؤسسات، ومن أشهر هذه الأدوات Hadoop.(Tome,2012)

هدوب (Hadoop): هو إطار عمل برمجي مفتوح المصدر مكتوبة بلغة Java، يدعم التطبيقات الموزعة للبيانات الكثيفة مثل كلوديرا هادوب (Cloudera Hadoop)، وهو يعمل على تخزين ومعالجة البيانات الضخمة بشكل موزع مثل تخزين عه البيانات على عدة أجهزة ومن ثم توزيع عملية المعالجة على هذه الأجهزة لتسريع نتيجة المعالجة. (Borthakur, 2019)

التحليل المرئي للبيانات (Data visualization): يعد تصور البيانات إحدى خطوات تحليل البيانات وتقديمها للمستخدمين. ويشير تصور البيانات إلى التقنيات المستخدمة لتوصيل البيانات أو المعلومات عن طريق ترميزها ككائنات مرئية (مثل النقاط أو الخطوط أو الأشرطة) الموجودة في الرسومات. (العيسى،2019)

تابلوه (Tableau): هو برنامج لتحليل البيانات بصريًا، يمكّن المستخدمين إنشاء وتوزيع لوحات معلومات تفاعلية وقابلة للمشاركة، تصور الاتجاهات والتغيرات وكثافة البيانات في شكل رسوم بيانية ومخططات. (Patel، 2022)

الدراسات السابقة

بمراجعة الإنتاج الفكري المشير في عدد من الدوريات المتخصصة، وقواعد البيانات العالمية، تبين وجود عدد من الدراسات الأجنبية التي حاولت الربط بين البيانات الضخمة وبيئة المكتبات والمعلومات، كما لوحظ ندرة الدراسات العربية المتخصصة التي تناولت الموضوع.

اهتم الكثير من الباحثين وعلى رأسهم مجموعة الباحثين في معهد مكنزي الدولي (McKinsey Global Institute, 2011) ومجموعة من الباحثين في شركة IBM بالبحث في البيانات الضخمة بشكل عام وبتطبيقاتها خلال السنوات القليلة الماضية، وقد تم الاعتراف بالأهمية الكبيرة للبحث في البيانات الضخمة بشكل جيد: فتكنولوجيا البيانات الضخمة تمكن من الحصول على رؤى أعمق وأكثر قيمة للبيانات، ومن ثم اتخاذ قرارات أكثر في الوقت المناسب.

كما اشـارت مجموعة من الدراسـات التي تتعلق بتحليل البيانات الضـخمة في المكتبات والاسـتفادة منها ، كدراسـة (Hota, 2022) حول إطار تطبيق البيانات الضخمة وتحليل جدوى اسـتخدامها في المكتبة، وتهدف هذه الدراسـة إلى وتقديم نظرة شـاملة للقراء لفهم إطار تطبيق البيانات الضـخمة في المكتبات، واقتراح إطار تطبيق البيانات الضـخمة في المكتبات، وتوصـلت الدراسـة للعديد من النتائج التي تشـتمل على التحديات الأسـاسـية في تطبيق البيانات الضخمة في المكتبات ودقة البيانات، وتقليل حجم البيانات وضـغطها، وسـرية البيانات وأمنها، ونظام معاجلة البيانات الضـخمة والتكنولوجيا المتعلقة بها وتتمثل قيمة هذه الدراسـة في توافر العديد من التطبيقات والحلول العملية للمكتبات لتطبيق إطار البيانات الضخمة.

بينما دراسـة (Mutula، 2016) بعنوان" Big Data Industry: Implication for the Library & Information Sciences" التي هـدفـت إلى التعرف، على تـأثير صـنـاعـة البيـانـات الضـخمة على علوم المكتبات والمعلومات، بما في ذلك مناقشـة فوائد اسـتخدام تحليلات البيانات الضخمة، في جميع المجالات الأكاديمية، وتزايد النمو، والاهتمام بالبيانات الضخمة، بسبب انخفاض تكلفة تخزين البيانات، وقد اتخذت الدراسة المنهج التحليلي منهجا رئيسا" لها، من خلاك تحليل برامج العديد من مدارس علوم المكتبات والمعلومات، في الولايات المتحدة الأمريكية التي أضـافت البيانات الضـخمة لها، بما في ذلك جامعة Pittsburgh، وقد عكسـت الدراسـة أن البيانات الضـخمة لها تأثير هائل في جميع المجالات الأكاديمية، والتي سـاعدت في خلق مهارات جديدة في مختلف القطاعات الأكاديمية، بما في ذلك علوم المعلومات، في مجالات إدارة البيانات، والتنظيم، والأرشفة، والبحث والاسترجاع، للبحوث متعددة التخصصات.

وجاءت دراسـة (البراشـدية، 2018) بعنوان" اسـتخدام البيانات الضـخمة في المكتبات الأكاديمية "والتي هدفت الدراسـة إلى اسـتكشـاف التقنيات وأدوات تحليل البيانات الضـخمة لكي يمكن تطبيقها في المكتبات الأكاديمية، فضـلا عن اكتشـاف فوائد البيانات الضـخمة بالنسـبة للمكتبات الأكاديمية. وقد أجرت بحث منهجي وشـامل عن الأبحاث المنشـورة على مدار 7 سـنوات حول البيانات الضـخمة للمكتبات الأكاديمية. وأظهرت النتائج وجود (37) دراسـة سـابقة فقط عن البيانات الضـخمة في المكتبات الاكاديمية. وقد قامت الدراسـة بمراجعتها وتحليلها.

وأظهرت دراسة (Mishra، 2015) إدارة البيانات الضخمة في المكتبات من حيث جمعها ومعالجتها، وما يترتب من تحليلها في إنتاج معلومات جديدة لإفادة المستفيد النهائي منها، ومن أبرز ما أشارت إليها الدراسة عن مجالات استخدام البيانات الضخمة في المكتبات: قواعد البيانات على الإنترنت، وفهرس الإتاحة على الخط المباشر، والتحليل الإحصائي لبيانات الاستخدام، وإدارة بيانات البحوث الرقمية، ومواقع التواصل الاجتماعي.

بينما دراسة (Cervone، 2017) بعنوان" Evaluating social media presence: A practical application of big data & analytics in information organizations" التي قامت على تقييم فعالية وسائل التواصل الاجتماعي بالمكتبات بوصفها أحد مصادر البيانات الضخمة، حيث قامت بتحليل عدد كبير من تغريدات ومشاركات مستخدمي المكتبات، ومن ثم استكشاف علاقة ذلك بفاعلية الأداء بالمكتبات، ومؤسسات المعلومات، وانعكاسات ذلك على خدمات المعلومات بها، وقد اعتمدت الدراسة المنهج الوصفي التحليلي منهجا رئيسا" لها، وقد كشفت الدراسة عن مجموعة من النتائج، لعل وجود العديد من التقنيات، مثل تحليل الشبكات الاجتماعية، والتحليل الإحصائي، والترميز الجغرافي، والتي يمكن استخدامها كجزء من مشروع لتقييم فعالية وسائل التواصل الاجتماعي بوصفها أحد أهم مصادر البيانات الضخمة، كما تبين إمكانية إجراء تحليلات قوية لهذه التغريدات والمشاركات بالاعتماد على هذه التقنيات مجتمعة.

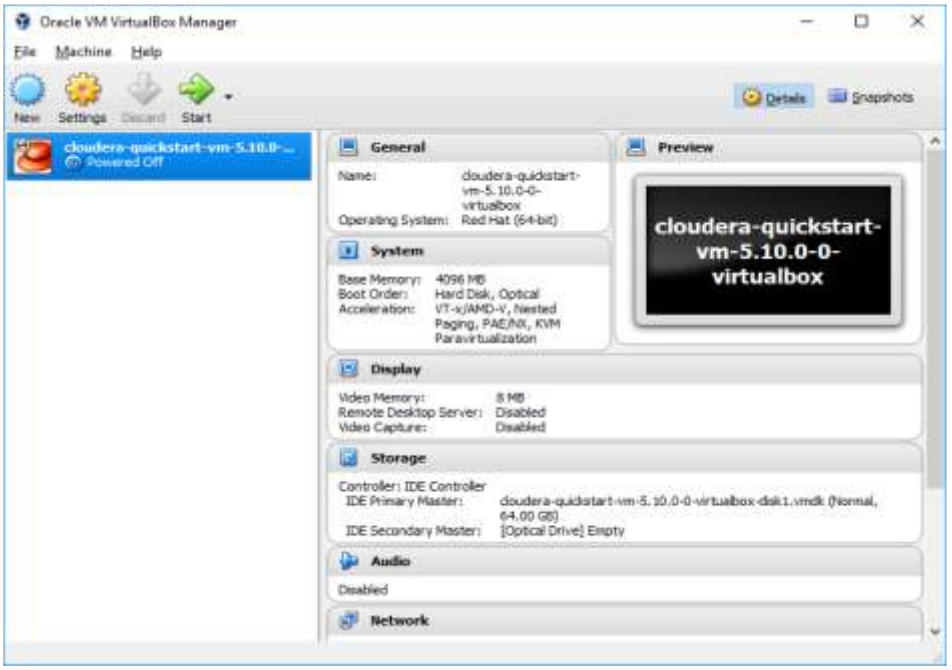
وناقشت (Fister, 2015) أيضا" بحث لها بعنوان" Big Data or Big Brother? Data, Ethics, & Academic Libraries, Library Issues " حول البيانات الضخمة ومسائل تتعلق بتأثير البيانات الضخمة على مسألة الخصوصية التي تتعلق بحياة الناس، ودور مواقع التواصل الاجتماعي في تعريض خصوصية الأفراد للخطر، كما ناقشت دور المكتبات الأكاديمية في مواجهة مسألة الخصوصية حيث تواجه هذه المكتبات العديد من التحديات في عصر البيانات الضخمة، وسوف يترتب عليها دعم استخدام البيانات والمحافظة عليها باعتبارها ذات قيمة متزايدة في النظام البيئي للمعرفة، وتطرقت الباحثة إلى الاستفادة من تحليل البيانات الضخمة في تنمية مجموعات المكتبات وخدماتها.

ثالثا": التطبيق العملي للبحث

تثبيت وتشغيل منصة Cloudera Hadoop و Tableau

قبل أن نبدأ اعداد Hadoop لابد من اختيار نظام لتشـــغيل الاجهزة الافتراضـــية مثل VirtualBox من اجل تشغيل جهاز افتراضي معد من قبل شركة Cloudera والذي يحتوي على Hadoop.

- من بعدها سيتم استيراد ملف (import Cloudera) بعد تحميله كما يوضح الشكل رقم (1).

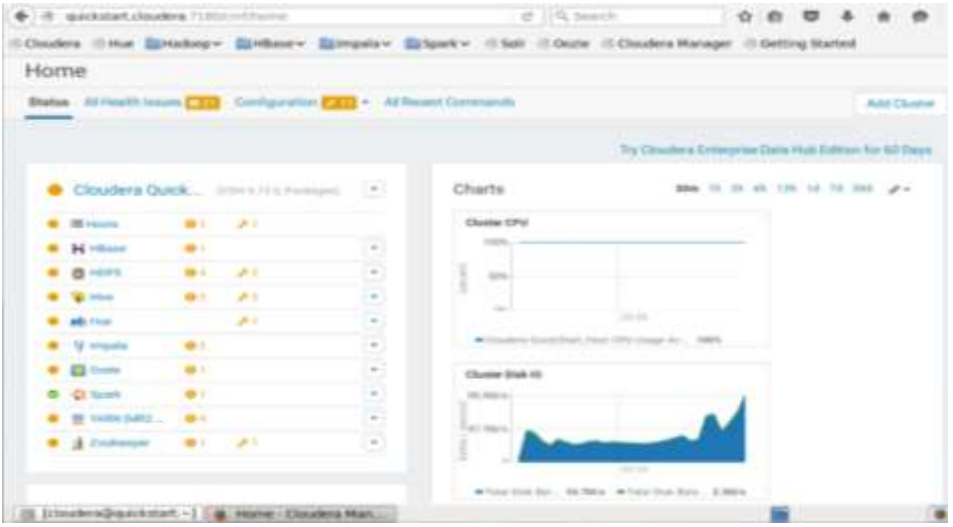


الشكل رقم (1): Cloudera داخل VirtualBox

بعد ان تكتمل العملية، سنجد انه قد تكون لدينا جهاز معد من قبل Cloudera مثبت فيه Hadoop ومعد وجاهز للاستخدام.

تصميم النظام:

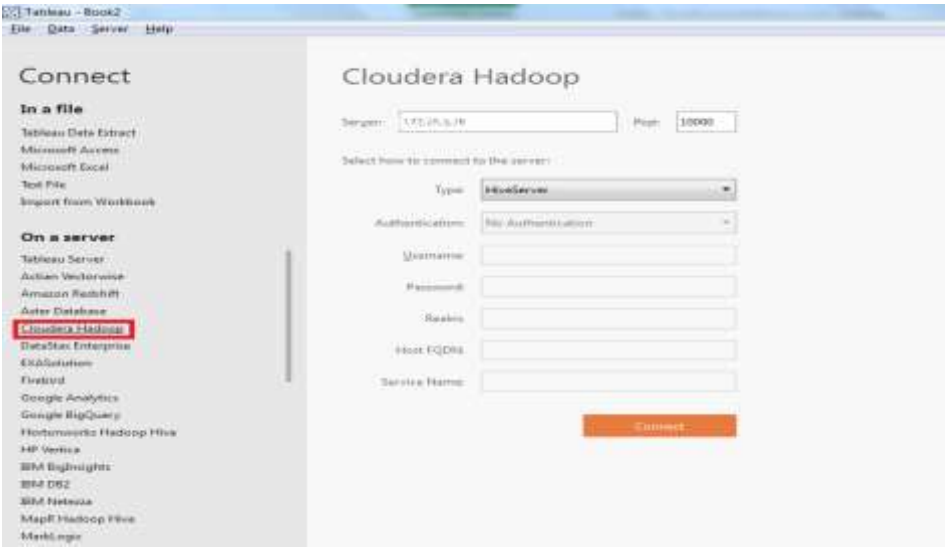
الرسم التالي يوضح معمارية نظام التحليل والمعالجة والادوات المستخدمة فيه، والذي سيتم توضيح الوظائف المستخدمة في هذا الباب.



الشكل رقم (2): الصفحة الرئيسية لمنصة Cloudera مع ادواتها

تثبيت برنامج Tableau

تم تنزيل Tableau Desktop لمدة 15 يومًا من الفترة التجريبية. ولاتصاله بقاعدة بيانات MySQL، تم تنزيل برامج تشغيل "اتصال قاعدة البيانات المفتوحة" ODBC (Programming Interface (API) for accessing a database) كما يوضح الشكل رقم (3) و(4).



الشكل رقم (3): ربط Tableau مع Cloudera

الشكل رقم (4): واجهة استخدام Tableau

آلية تطبيق منصة Cloudera Hadoop و Tablaeu

لقد تناول هذا القسم آليات تطبيق Cloudera Hadoop وتكامله مع Tableau بالاضافة الى عرض ومناقشة وتحليل مختلف النتائج التي كشف عنها القسم العملي من هذا البحث.

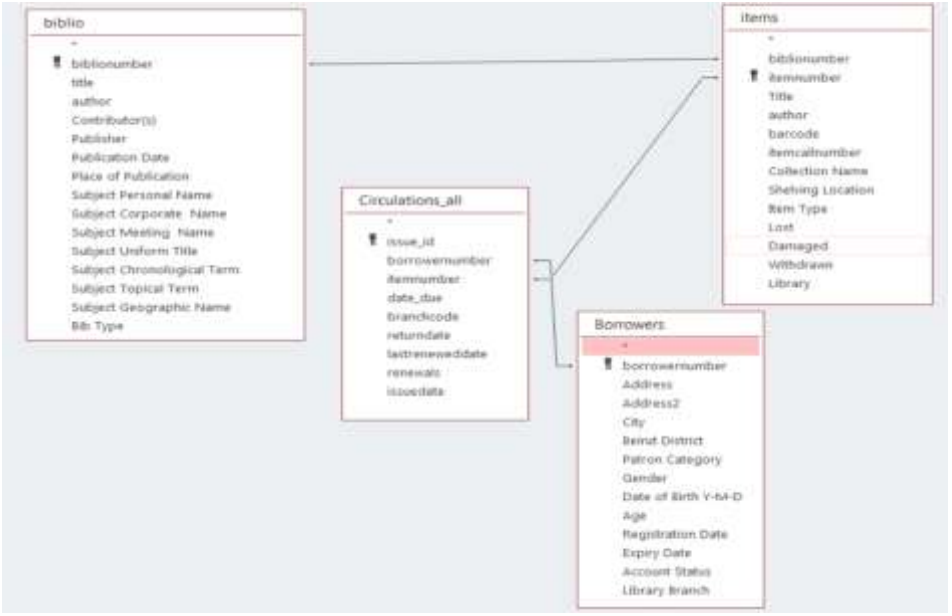
وحتى يتيسر البحث التطبيقي، فقد تم توجيهه للعمل على معالجة تأثير تحليل البيانات الضخمة من شبكة التواصل الاجتماعي فيسبوك على خدمة الإعارة في المكتبة خلال جائحة كورونا.

تمثلت قاعدة بيانات مكتبة السبيل العامة من ستة جداول مفصلة تحتوي على:

1- جدول بيانات المستفيدين (borrowers): كافة بيانات المستفيد الشخصية.
2- جدول البيانات الببليوغرافية (bibliographic): كافة البيانات الببليوغرافية المتعلقة بمصادر المعلومات.
3. جدول بيانات مصادر المعلومات (Items): كافة البيانات المتعلقة بالمصدر.
4. جدول بيانات الإعارة (Circulations_all): كافة البيانات المتعلقة بعمليات الاعارة.
5. جدول أيام الاقفال في المكتبة (daysclosed): عدد أيام الاقفال في المكتبة خلال الشهر منذ سنة 2018 حتى 2021.

6. جدول بيانات المتفاعلين على الفيسبوك (facebook_ins): بيانات المتفاعلين على فيسبوك المكتبة.

والرسم التالي يوضح هيكلية البيانات داخل قاعدة البيانات.



الشكل رقم (5): ربط جداول قاعدة البيانات بين خاناتها

تنقيب البيانات:

البيانات المدرجة ليست بحاجة لعمليات معالجة أولية نسبة لكونها تقريبا" بيانات شبه مهيكلة، حيث لا توجد بها بيانات ناقصة مفقودة. مع العلم، ان معظم البيانات الضخمة تخضع للتنقيب كونها تحوي قيما" شاذة وأخرى مكررة، يعالجها النظام كنظام Spark عبر عدة خوارزميات لمعالجة هذه القيم الشاذة مثل StandardScaler و Normalizer و MinMaxScaler.

رابعا": عرض وتحليل النتائج

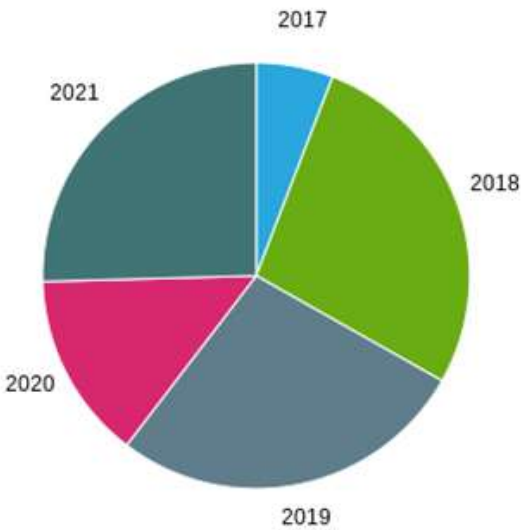
فيما يلي نستعرض نتائج تحليل البيانات من سنة 2018 حتى سنة 2021 وذلك لمعرفة أهمية تطبيق أدوات تحليل البيانات الضخمة على مكتبة السبيل العامة.

❖ <u>الخطوة الأولى</u>: بداية في هذه الخطوة، اخترت الخصائص التي سيتم اجراء عمليات التجميع Clustering عليها، وجمعت جميع بيانات الاعارة من خلال الاستعانة بالبيانات من سنة 2017 حتى 2021.

جدول رقم (1): مجموع الإعارة في كل سنة

```
SELECT SUBQUERY.year, SUM(SUBQUERY.CountOfCollection) as total from
(
SELECT items.`collection name` , Year(issuedate) AS year, Count(items.`collection name`) AS CountOfCollection FROM Circulations_all INNER JOIN i
GROUP BY items.`collection name` , Year(issuedate)
ORDER BY  year, CountOfCollection DESC
) AS SUBQUERY
GROUP BY Year;
```

	subquery.year	total
1	2017	3512
2	2018	14496
3	2019	14376
4	2020	9550
5	2021	15201



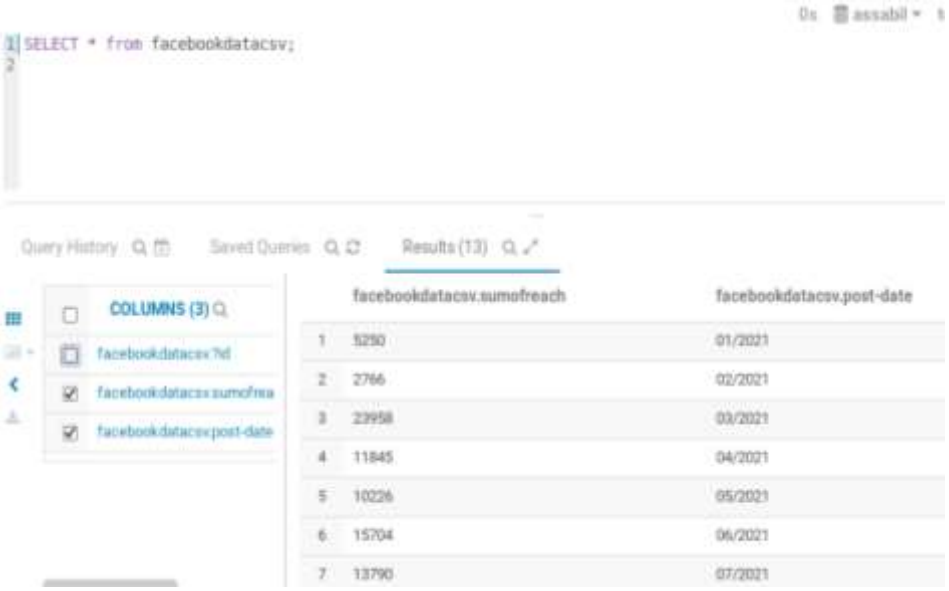
شكل رقم (6): مجموع الإعارة في كل سنة

❖ <u>الخطوة الثانية:</u> الحصول على بيانات المتفاعلين على فيسبوك المكتبة من خلال خاصيةFacebook Insights التي توفرها شركة فيسبوك.

جدول رقم (2): المتفاعلين على فيسبوك

Post ID	Permalink	Post Message	Type	Posted	Reach
4447768	sabilingo/posts/4989376444447768	جمعية السبيل تتمنى لكم عاماً سعيداً مليئاً بالثقافة والقراءة. ك	Photo	12/30/2021 5:10	1867
1174619	sabilingo/posts/4988774601174619	ons. #assabil #cultureforall #documentaires	Photo	12/30/2021 0:26	564
2176223	sabilingo/posts/4978758562176223	بتعرفوا انو صار فيكن تزوروونا كل يوم، من الإثنين للسبت ا	Photo	12/27/2021 0:30	25575
0638794	sabilingo/posts/4960799520638794	نسعى من خلال عملنا إلى إرضاء جميع الأذواق، ونعزز مح	Photo	12/23/2021 0:44	985
1105169	sabilingo/posts/4956135771105169	sabil #cultureforall #e_resources #youtube	Photo	12/22/2021 0:16	591
8679356	sabilingo/posts/4947060568679356	مشاركة الانسان ما يشعر به مع الآخرين واحدة من أهم الأمو	Photo	12/20/2021 2:26	661
0004237	sabilingo/posts/4933811760004237	احتفلوا معنا باليوم العالمي للغة العربية مع فرقة الكنجاتي و	Link	12/17/2021 5:31	334
6737245	sabilingo/posts/4933148346737245	#assabil #cultureforall #colleen_mccullough	Photo	12/17/2021 2:33	412
0107795	sabilingo/posts/4932776180107795	all #english_book_club #english_litterature	Photo	12/17/2021 0:35	508
3975821	sabilingo/posts/4927429253975821	assabil #cultureforall #disney #kids #books	Photo	12/16/2021 0:23	624
7751053	sabilingo/posts/4923010267751053	مشاركة الانسان ما يشعر به مع الآخرين واحدة من أهم الأمو	Photo	12/15/2021 3:18	1046
1147601	sabilingo/posts/4922378121147601	Al kamandjati - الكمنجاتي إنَّ الّذي مَلَأَ اللّغاتِ مَحاسِنً	Photo	12/15/2021 0:06	1311
5059761	sabilingo/posts/4916589855059761	السبيل وناديّ لكلّ الناس يعرضان فيلم "الى أين؟" للمخرج ه	Photo	12/13/2021 4:54	474
9782884	sabilingo/posts/4902691959782884	prêt. #assabil #cultureforall #freud #books	Photo	12/9/2021 3:07	634

جدول رقم (3): مجموع المتفاعلين على فيسبوك حسب الشهر والسنة فقط



❖ <u>الخطوة الثالثة</u>: الحصول على مجموع ايام اقفال المكتبة من سنة 2018 حتى 2021.

جدول رقم (4): مجموع ايام اقفال المكتبة خلال كل سنة

```
SELECT `year-lk` , Sum(daysclosed.nbclosed) AS SumOfclosed
FROM daysclosed
GROUP BY `year-lk`;
```

Query History | Saved Queries | Results (5)

COLUMNS (2)

	year-lk	STRING_TYPE
	sumofclosed	DOUBLE_TYPE

	year-lk	sumofclosed
1	2017	64
2	2018	67
3	2019	74
4	2020	247
5	2021	138

■ وهذا الجدول يلخص العلاقة بين مجموع ايام اقفال المكتبة ومجموع الاعارة خلال فصـول السنة.

جدول رقم (5): ربط العلاقة بين مجموع الاعارة وعدد أيام الاقفال حسب فصول السنة

السنة	الفصل	عدد ايام الإقفال	مجموع الإعارة
2018	Q1	17	3882
2018	Q2	15	3712
2018	Q3	18	3730
2018	Q4	17	4864
2019	Q1	18	4460
2019	Q2	20	4534
2019	Q3	11	4331
2019	Q4	17	3003
2020	Q1	44	3596
2020	Q2	73	997
2020	Q3	69	1616
2020	Q4	61	2341

السنة	الفصل	عدد ايام الإقفال	مجموع الإعارة
2021	Q1	70	1682
2021	Q2	27	3896
2021	Q3	24	5235
2021	Q4	11	4513

Q1: الفصل الأول من السنة

Q2: الفصل الثاني من السنة

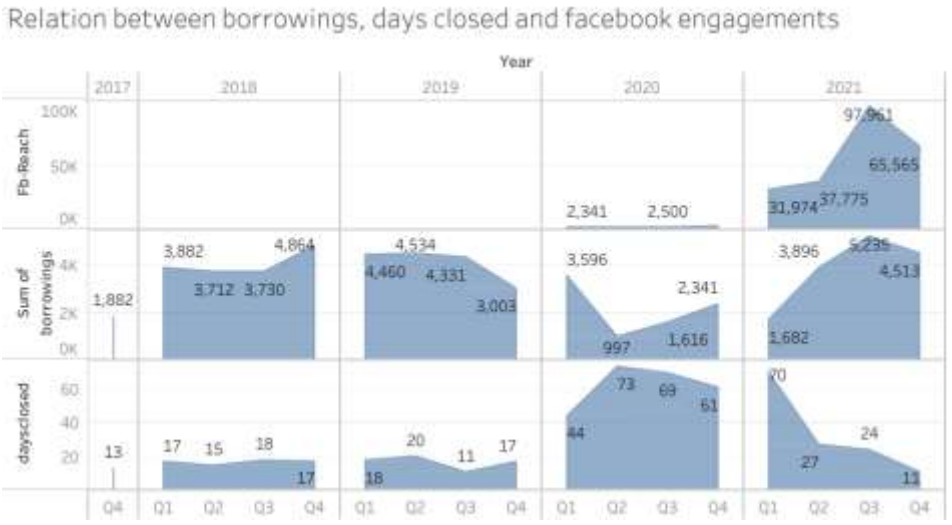
Q3: الفصل الثالث من السنة

Q4: الفصل الرابع من السنة

لا بد من تحديد المتوسط الحسابي لمجموعة البيانات (مجموع الإعارة)؛ فالمتوسط الحسابي يحدد مركز هذه البيانات أو منتصفها.

المتوسط الحسابي = مجموع الإعارة كل سنة/عددها= 3525

❖ <u>الخطوة الرابعة</u>: تنفيذ تطبيق Tableau لتحليل البيانات صوريا" وذلك من خلال إنشاء علاقة منطقية بالبيانات بصريًا.(العلاقة بين مجموع الاستعارة، أيام اقفال المكتبة، ومجموع المتفاعلين على فيسبوك المكتبة).



شكل رقم (7): ربط العلاقة بين مجموع الاستعارة، أيام اقفال المكتبة، ومجموع المتفاعلين على فيسبوك المكتبة

أظهر هذا التحليل عدد المتفاعلين التي تم جمعها من صفحة الفيسبوك الخاصة بمكتبة السبيل خلال سنة 2020 وسنة 2021 فقط، لأن المكتبة لم تستخدم شبكة التواصل الاجتماعي Facebook قبل هذا التاريخ. كما أثار أيضا" تأثير جائحة كورونا على خدمة الإعارة في مكتبة السبيل العامة والتي واجهت خيارات صعبة بشأن الخدمات التي تقدمها والكيفية التي تفعل بها ذلك، وتتفاوت هذه الخيارات من فرض قيود محدودة إلى الإغلاق الكامل. وفقد شمل التحليل سنة 2018 حتى سنة 2021 ضمنا" وقد جاءت النتيجة على النحو التالي:

1. في عام 2018، جاءت النتيجة إيجابية وكافية لأن أيام الاغلاق اقتصرت فقط على الأعياد الرسمية حيث بلغ عدد أيام الإقفال خلال الفصل الأول ب 17 يوما" ومجموع الإعارة 3882 وهي أعلى من قيمة المتوسط الحسابي العام. وقد بلغ عدد أيام الإقفال خلال الفصل الثاني من السنة ب 15 يوما" ومجموع الإعارة 3712 وهي أعلى أيضا" من قيمة المتوسط الحسابي العام، في حين كانت النتيجة متقاربة مع الفصل الثالث أي اشهر (تموز – آب – أيلول) بمجموع 3730 اعارة وعدد أيام الاغلاق ب 18 يوما". أما الفصل الأخير من السنة فقد شهدت المكتبة ارتفاعا" ملحوظا" بأعداد الإعارة، فقد ناهز ال 4864 وهو أعلى أيضا" من قيمة المتوسط الحسابي العام، وعدد إيام الاقفال بلغت 17 يوما".
2. عام 2019، كانت أرقامه متقاربة من عام 2018 لأنه أيضا" كان عاما" طبيعيا" واقتصرت الأعطال على الأعياد والمناسبات الرسمية، حيث كان الارتفاع الملحوظ في بداية العام أي في الفصول السنة الأولى، الثانية والثالثة لتنخفض تدريجيا" في الفصل الأخير من السنة. وقد بلغ مجموع الاعارة 3730 وعدد أيام الاغلاق ب 18 يوما" في الفصل الأول، في حين بلغ عدد أيام الإقفال خلال الفصل الثاني من السنة ب 20 يوما" ومجموع الإعارة 4534 وهي أعلى أيضا" من قيمة المتوسط الحسابي العام، أما الفصل الثالث من السنة بمجموع 4331 للإعارة وعدد أيام الاغلاق ب 11 يوما". ثم أخيرا انخفض مجموع الإعارة في الفصل الأخير من السنة ليصبح 3003 وهو أقل قيمة من المتوسط الحسابي العام، وعدد أيام الاغلاق 17 يوما".
3. ففي عام 2020 أظهر التحليل انخفاضا" كبيرا" في عدد الناشطين على المنشورات المتعلقة بالمكتبة حيث سجل عدد المتفاعلين بما يقارب 2500 متفاعل على جميع المنشورات كمتوسط حسابي خلال فصل من فصول هذه السنة. في حين سجل مجموع الإعارة ادنى

مستوياته حوالي 2137 كمتوسط حسابي عام وهو قيمة متدنية جدا" نظرا لتأثير جائحة كورونا الذي وصل إلى لبنان، حيث واجهت المكتبات حول العالم خيارات صعبة بشأن الخدمات التي ستقدمها والكيفية التي ستفعل بها ذلك، واضطرت جميع أنظمة المكتبات العامة الى الإغلاق بالكامل ومن بينها مكتبة السبيل العامة. فقد أدت تداعيات الوباء الى اقفال معظم شهور هذه السنة والذي ألقى بظلاله السلبية على خدمات المكتبة ومنها خدمة الإعارة. ففي الفصل الأول من السنة فقد بلغ عدد الاقفال 44 يوما"، في حين كان مجموع الإعارة ب 3596 وهو أعلى قيمة من المتوسط الحسابي العام. أما في الفصل الثاني فقد وصل معدل الاغلاق الى ذروته بحيث بلغ 73 يوما"، مما أثر بطبيعة الحال على مجموع الإعارة، فقد وصل الى أدنى مستوياته حيث بلغ 997 اعارة فقط وهو أقل بكثير من المتوسط الحسابي العام. في القصل الثالث من السنة ارتفع نسبيا" مجموع الإعارة ليبلغ 1616 وهو معدل ما زال متدنيا نسبتا" الى المتوسط الحسابي العام، في حين بلغ عدد إيام الإغلاق 69 يوما". أما الفصل الأخير من السنة فقد شهد ارتفاعا" بسيطا" في عدد الإعارة ليسجل 2341 وهو عدد اقل قيمة من المتوسط الحسابي العام، في حين بلغ عدد الاقفال 61 يوما".

4. في عام 2021، أظهرت نتائج التحليل تنشيط وتفعيل صفحة المكتبة على الFacebook، بحيث أصبحت تنشر كل ما هو جديد وتهتم وتستغل أكبر قدر ممكن من إمكانات هذه الشبكة، من خلال زيادة عدد منشوراتها والترويج لها من نشر الأخبار اليومية عن المكتبة ومجموعاتها وإقامة المعارض وتقديم الخدمات والإعلان عن الورش. كما طورت المكتبة صفحتها على الFacebook للوصول إلى مستخدميها في كل زمان ومكان وبالمثل حرص المختصون بإدارة هذه الشبكات على تطوير الخطط والسياسات المختلفة لتشجيع المستفيدين على الوصول إلى المكتبة واستعارة مجموعاتها، واستخدامها بسهولة وتلبية احتياجاتهم في أي وقت ومن أي مكان. كما تمكنت المكتبة من خلق برامج لتوعية المستفيدين بأهمية المكتبات لذلك سجل الحد الأقصى من عدد المتفاعلين 97961 متفاعلا"، مما انعكس إيجابا" وبشكل سريع على عملية الإعارة حيث بلغ 5235 وهو المعدل الأعلى خلال جميع السنوات على الرغم من استمرار جائحة الذي خلق تحديات جديدة للمكتبات العامة خلال الكارثة، ظلت مكتبة السبيل وفيّة لرسالتها كمؤسسات مجتمعية،

واستمرت في العمل والاستجابة للوضع الجديد من خلال إيجاد طرق مبتكرة لخدمة روادها على الرغم من التأثير الكبير لهذه الجائحة. استمرت مكتبة السبيل في حالة الاغلاق في الفصل الأول من السنة ليبلغ 70 يوما"، بينما سجل مجموع الإعارة 1682 وهو رقما" أقل قيمة من المتوسط الحسابي لعام، ليرتفع بعدها معدل الإعارة بشكل كبير ليسجل 3896 ويكون بذلك قد تخطى المتوسط الحسابي العام مع عدد أيام اقفال بلغ 27 يوما". أما الفصل الثالث من السنة فقد ارتفع مجموع الإعارة بحد كبير ليبلغ ذروته خلال جميع السنين ليبلغ 5235 اعارة وهو الرقم اعلى قيمة من المتوسط الحسابي العام مع عدد أيام اقفال ناهز 24 يوما". في الفصل الأخير من السنة كانت ارقامه متقاربة من الفصل الثالث حيث سجل مجموع الإعارة 4513 في حين بلغت أيام الاقفال 11 يوما".

من هنا يمكننا القول أن على الرغم من تداعيات الوباء وما رافقها من إجراءات مثل إغلاق مكتبة السبيل وإلغاء برامجها، وما ألقته من تأثيرات سلبية على خدماتها بشكل عام وخدمة الإعارة بشكل خاص كالنتائج التي ظهرت عام 2020، فرض عليها مواجهة هذه التحديات والقيام بتطوير واستخدام أدوات من أجل تقديم أفضل الخدمات لروادها خلال حقبة انتشار الجائحة وتعويض الخلل الحاصل، كزيادة استخدام مواقع التواصل الاجتماعي الخاص بها مثل، Facebook للتواصل مع المستفيدين وتسويق خدمات المكتبة والمجموعات المكتبية المتواجدة فيها، بالإضافة الى الضوابط الاحترازية بما يخص الارشادات السلامة العامة في المكتبة في ظل هذه الجائحة.

أخيرا"، يمكننا القول أن مكتبة السبيل واجهت تحديات كبرى بسبب وباء كورونا في السنة الأولى من الجائحة مما حتّم عليها اتخاذ إجراءات لحماية العاملين بها وحماية المستفيدين من خدمات المعلومات، واستطاعت أن تستدرك الأمر حيث استطاعت أن تواجه هذه التحديات بشكل ناجحٍ من خلال ما اظهرته النتائج في عام 2021 بزيادة كبيرة في مجموع الإعارة من خلال الترويج لخدماتها ومجموعاتها عبر مواقع التواصل الاجتماعي، والتي برهنت أنها قادرة على تحسين العلميات داخل المكتبة وزيادة فعاليتها من ناحية الدقة والسرعة، بالإضافة الى تحسين الخدمات المقدمة إلى المستفيدين في بيئة المكتبات يجعل ذلك نظرة المستخدمين إلى المكتبة تتغير يوما بعد يوم.

التوصيات:

في ضوء النتائج التي تم التوصل إليها يوصي البحث بما يأتي:

- ✓ أن تهتم المكتبات اللبنانية بموضوع البيانات الضخمة، وتبدأ نقاشاً مع الجهات الأخرى المهتمة بالبيانات في نفس المؤسسة من أجل وضع خطة لتوعية العاملين بأهمية البيانات الضخمة ومجالات الاستفادة منها في المكتبة والمؤسسات ذات العلاقة.
- ✓ أن تولي المكتبات اللبنانية اهتماماً خاصاً بتحليل البيانات الضخمة، وأدوات التحليل غير التقليدية، وغيرها من أدوات وبرمجيات التنقيب عن البيانات وخاصة مفتوحة المصدر.
- ✓ أن تولي أقسام المكتبات والمعلومات في الجامعات اللبنانية الاهتمام الكافي بما يسمى في دراسات المعلومات بعلم البيانات (Data science) وعلم التنقيب عن البيانات (Data mining) لأن هذين الموضوعين يمثلان التوجهات الحديثة في أقسام المعلومات.
- ✓ عمل دورات تدريبية لأمناء المكتبات تهدف إلى إكساب مهارات التعامل مع البيانات الضخمة وجمعها وتحليلها والاستفادة منها في تطوير المكتبات.
- ✓ الاستفادة من شبكات التواصل الاجتماعي كأداة لجمع البيانات الضخمة من أجل تحليلها والاستفادة منها في تنمية وتحسين واقع المكتبات.
- ✓ ضرورة جمع وتحليل البيانات الجديرة بالاهتمام لاسيما البيانات الناتجة من عمليات الإعارة والإرجاع والشبكات الاجتماعية.
- ✓ تكثيف توجيه الاستثمار للبيانات الضخمة الناتجة من المكتبات في إدارة المقتنيات وتقييم الأداء وجودة الخدمات.
- ✓ العمل على اهتمام المدراء بما يعزز أدوارهم في استشراف مستقبل هذه المكتبات في استخدام البيانات الضخمة.

الخاتمة:

تضافرت جهود متواصلة في هذا البحث لتقديم رؤية أكاديمية شاملة، وذلك من خلال التعرف على ماهية البيانات الضخمة وتحليلاتها، والتعرف على أهمية تحليلات البيانات الضخمة وعناصر تحليلات البيانات الضخمة وادوات التعامل مع البيانات الضخمة، وعرض وتحليل البيانات الخاصة بمكتبة السبيل العامة.

ان انتشار تطبيقات البيانات الضخمة بشكل واسع في كل المجالات، مما أدى إلى زيادة الاهتمام ببنائها بطريقة صحيحة خالية من الأخطاء والمشاكل وبشكل يلبي حاجات المستخدمين، حيث أنّ زيادة استخدام هذه التطبيقات مع مرور الوقت جعل الحاجة ملحة لتقديم طرائق وأدوات لدعم عملية اختبارها لاختبار وتحليل بيانات مكتبة السبيل العامة تطلب وجود منصة **Hadoop** بأجزائه المختلفة، ومن خلال هذه الدراسة قدمنا هذا التحليل ضمن بيئة تحاكي بيئة **Hadoop**.

يمكن اعتبار النتائج المقدمة في هذه الدراسة كأساس لدراسات مستقبلية، حيث أنّ مكتبة السبيل قدمت لاختبار تطبيق البيانات الضخمة، لذا يمكن الاستفادة من هذه الدراسة لتكون أساساً لتطوير أي مكتبة تنوي تحليل بياناتها ضمن بيئة Hadoop الفعلية، بالإضافة الى آلية ربطه ببرنامج Tableau لتظهر النتائج بالشكل المرئي المناسب. لذلك تحليل البيانات الضخمة في المكتبات باستخدام هادوب وتابلوه (Hadoop and Tableau) يعد توليفة قوية للاستفادة من البيانات الضخمة وتحويلها إلى رؤى قيمة ومعلومات مفيدة، حيث يمكن استخدام هادوب لتحليل ومعالجة البيانات الضخمة، ثم تصور النتائج باستخدام تابلوه بشكل بديهي وبصري وان استخدام هادوب وتابلوه معًا يمكن أن يوفر قدرات تحليل قوية ورؤى عميقة من البيانات الضخمة. ويمكن تحديد النماذج والأنماط والعلاقات في البيانات وتصورها بشكل بصري ممتع وفعال.

بشكل عام، استخدام هادوب وتابلوه في تحليل البيانات الضخمة في المكتبات يمكن أن يساهم في تحويل البيانات إلى معلومات قيمة واتخاذ قرارات أفضل. تلك التكنولوجيات توفر إمكانية الوصول إلى البيانات بشكل أسرع وتحويلها إلى تقارير وتحليلات سهلة الفهم، مما يعزز فهم البيانات بشكل كبير.

كما إن اســتخراج القيمة من هذا الكم الهائل من البيانات وتجميعها واســترجاعها في الوقت والمكان المناســب يتطلب من أخصــائي المعلومات وعي كبير بمفهوم البيانات الضــخمة ومن جهة أخرى لابد له من اكتســاب مجموعة من المهارات والكفاءات من أجل إدارة هذا الكم الهائل من البيانات والتعامل معها بمرونة، وهذا التمكن سيؤدي لا محالة بأخصائي المعلومات إلى امتلاك مجموعة من المهارات والكفاءات الجديدة تتلاءم وذلك الأمر الذي يتطلب تدريب أكاديمي مناسب مع ضرورة مواكبته، ومن جهة أخرى تكوين ذاتي لأخصائي المعلومات للتأقلم مع المهارات والأدوار جديدة المطلوبة منه في ظل البيانات الضخمة.

على الرغم أن المكتبات اللبنانية بعيدة كثيرا" عن مواكبة مفهوم البيانات الضخمة وغير قادرة كفاية على الخوض في غمار الاستفادة من خصائصه نظرا" لمحدودية وعي ومعرفة مهينيها بهذا المفهوم الجديد وبمزاياها وسبل الاستفادة منه، ولكن التحولات والتطورات التي تشــهدها المكتبات وخاصــة مع ازدياد حجم بياناتها مع مرور الســنين، لابد أنهم ســيكونون مضــطرين البدء في خوض غمار تطبيق البيانات الضــخمة والاســتفادة من أهميتها وفق الإمكانيات المتاحة لهم ولو بشــكل مرحلي لأنه لا مجال للتأجيل أو التردّد أو التخوّف المســتمر من أي تغيير خاصة التغييرات التكنولوجية.

المراجع

المراجع باللغة العربية

البراشـــدية، حفيظة.(2018). اســـتخدام البيانات الضـخمة في المكتبات الأكاديمية. المؤتمر الدولي الرابع والعشـرين للمكتبات المتخصـصـة- فرع الخليج العربي. جامعة السـلطان قابوس. مسقط

الحكيم، مازن ســـمير.(2017).البيانات الضـخمة-Big data. مجلة شـــهرية تصـــدر عن كلية العلوم. جامعة بغداد: العراق.

العيســى، أيوب.(2019). تصـــوير البيـانـات: مقـدمـة. تم اســـترجـاعهـا بتـاريخ 9/4/2021 مـــــن:https://units.imamu.edu.sa/rcentres/en/asa/Pages/Introduction-to-Data-visualization.aspx

قناوي، يارة. (2022). آليات تطبيق نظم الذكاء الاصـطناعي في بيئة البيانات الضـخمة دراسـة وصفية تحليلية. المجلة العلمية للمكتبات والوثائق والمعلومات، 4(11.2)، 49-88.

المجلس الاقتصـــادي والاجتماعي بالأمم المتحدة.(2014).البيانات الضـخمة وتحديث النظم الإحصائية.

المراجع الأجنبية

Borthakur, D. (2019). HDFS Architecture Guide. Retrieved 08 23, 2021, from https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication and society, 15(5), 662-679.

Cervone, F.(2017). Evaluating social media presence: A practical application of big data & analytics in information organizations. Digital Library Perspectives, 33(1), 2-7.

Fister, B. (2015) Big Data or Big Brother? Data, Ethics, & Academic Libraries, Library Issues: Briefings for Faculty & Administrators 35(4).

Hota, P. K., & Hota, L. (2022). Big Data Ameliorates Library Services. The Changing Landscape for the Libraries and Librarians: Issues & Challenges, 110.

Li, Jun & Lu, Ming & Dou, Guowei & Wang, Shanyong. (2017). Big data application framework & its feasibility analysis in library. Information Discovery & Delivery.

McKinsey Global Institute. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved 07 29, 2022, from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

McKinsey.(2022). Author Talks: In defense of big data. Retrieved May, 17, 2022 from https://www.mckinsey.com/featured-insights/mckinsey-on-books/author-talks-in-defense-of-big-data

Mishra, R.(2015), BIG DATA: Opportunities and Challenges, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.6, June- 2015, pg. 27-35

Mutula, S. (2016). Big Data Industry: Implication for the Library & Information Sciences. Air. J. Lib. Arch. & Inf. Sc. 26. 2 : 93-96.

Patel, A. (2022). Data Visualization Using Tableau. Retrieved 5th January 2022 https://www.theseus.fi/bitstream/handle/10024/652129/Patel_Ashwin.pdf?sequence=4

Tome, W.(2012). Hadoop: The Definitive Guide - O'reilly media, USA ,third edition,07.05.2012) p 22.

Arab International Journal of Information Technology & Data
Vol. 4, No. 1 January - March 2024

Data Visualization for Big Data in Libraries using Hadoop and Tableau

Dr. Ayman Hamade
Lebanese University, Faculty of Information, Library information Science, Beirut, Lebanon
ah021@live.aul.edu.lb

Abstract:

This research presents an intellectual approach concerning the use of big data analysis in libraries and the tools for dealing with it, aiming to explore the potential of libraries to apply big data analytics in managing their services and activities. The research adopted an experimental methodology through practical work involving the analysis of data from the Facebook, specifically focusing on the lending service during the COVID-19 pandemic at the Al-Sabil Public Library in Lebanon, using the "Hadoop" big data analysis technology. The results were visually represented using Tableau software.

One of the significant findings of the analysis of the library's big data was a strong correlation between the number of interactions on Facebook and specific metrics, along with increased library usage in general and the lending service in particular. This indicates the importance of utilizing social networks in libraries, especially during times when the library is unable to perform its regular role, particularly in the context of the COVID-19 pandemic.

The thesis concluded with a concise summary, along with general recommendations related to big data in libraries. The researcher also anticipated that as libraries undergo transformations and developments, especially with the growth of their data over the years, they will inevitably need to delve into the realm of big data application and harness its significance to the best of their available resources, even if gradually, due to rapid technological changes.

Keywords:

Big data; Big data applications; Data Visualization; Hadoop; Tableau