النظم والبرمجيات السحابية المستخدمة في معالجة البيانات الضخمة

إعداد

نورهان سامي فتح الله مرسي

أ.د بدوية محمد البسيوني

أستاذ المكتبات والمعلومات بكلية الآداب_ جامعة طنطا

د. مني فاروق شهوان

مدرس المكتبات والمعلومات بكلية الآداب_ جامعة طنطا

❖ **المستخلص**

تهدف هذه الدراسة بشكل عملي إلى حصر وتحليل برمجيات البيانات الضخمة السحابية، وتم التوصل إلى (١٩٠) برنامج، تمَّ تقسيمهم على ستِ فئاتٍ مختلفة لكل فئة خصائصها وإمكاناتها، واعتمدت الباحثة في ذلك على المنهج الميداني كأسلوب من أساليب الوصفي التحليلي، مستخدمة في ذلك قائمة مراجعة لتجميع البيانات التي تخص تلك التطبيقات والنظم والبرمجيات.

وانتهت الدراسة إلى مجموعةٍ من النتائج كان من أهمها: أنَّ برمجيات التمثيل المرئي للبيانات جاءت في المرتبة الأولى بأعلى نسبة، حيث مثلت ٢٥.٣% بواقع (٤٨) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة، في حين جاءت برمجيات معالجة اللغة الطبيعية في المرتبة الأخيرة بنسبة ١١.١% بواقع (٢١) برنامج من إجمالي البرمجيات محل الدراسة ، كما كان من أهم نتائجها تصدر الدول الأجنبية في إنتاج برمجيات البيانات الضخمة وعدم إسهام أي دولة عربية في إنتاج تلك البرمجيات ، وأن أكثر برمجيات البيانات الضخمة مفتوحة المصدر، حيث تمثل تلك البرمجيات نسبة ٧٥.٨% من إجمالي برمجيات البيانات الضخمة محل الدراسة.

❖ **الكلمات الافتتاحية:** البيانات الضخمة – برمجيات البيانات الضخمة – تطبيقات البيانات الضخمة أدوات معالجة وتحليل البيانات الضخمة.

٠/٠التمهيد:

تسعى المؤسسات في استثمار المعرفة، والعمل على تحليل البيانات الضخمة، لتحقق لها قيمة مضافة وتساند متخذي القرار بما تنتجه من معلومات ذات أهمية؛ لأن تخزين ومعالجة وتحليل البيانات الضخمة يشكل تحديًا حقيقياً للكثير من المؤسسات، نتيجة تنوع وتضخم البيانات بشكل معقد جدا، لذلك فإن التعامل مع البيانات الضخمة "Big Data" فى المؤسسات يمثل معضلة حقيقية من حيث حجمها المتزايد بشكل سريع، ومن حيث الحاجة إلى مساحات تخزينٍ هائلة تفوق قدرة تلك المؤسسات المالكة لهذه البيانات، وكذلك قلة العائد من هذه البيانات وعدم استغلالها الاستغلال الأمثل، ولهذا فقد أدركت الشركات المتخصصة في إدارة المعلومات وتقنياتها وشركات برمجيات الحاسب الآلي ونظم الأنطولوجيا وإدارة المعرفة، أهمية التعامل بطرق أكثر قدرة على التحكم في هذه البيانات، والإفادة منها بأفضل الطرق، ليتم معالجتها وتحويلها إلى معلومات يستخرج منها معرفة. (L.Stone ، 2014)

١/٠مشكلة الدراسة:

تتمثل مشكلة الدراسة في حدوث تضخم كبير في حجم البيانات والمعلومات المتداولة والناتجة عن الكثير من المصادر كإنترنت الأشياء والحوسبة السحابية والشبكات الاجتماعية والأجهزة الذكية؛ ونتيجة لضخامة هذه البيانات والمعلومات أصبح من الصعب التعامل معها باستخدام قواعد البيانات التقليدية فيما يعرف بظاهرة البيانات الضخمة "Big Data"، أدى ذلك إلى ظهور نظم وبرمجيات تستطيع معالجة هذا الكم الضخم من البيانات.

٢/٠أهمية الدراسة:

تستمد الدراسة أهميتها من أهمية الموضوع التي تتناوله، والمتمثلة في محاولة تخزين ومعالجة وتحليل البيانات الضخمة، باستخدام النظم والبرمجيات السحابية المستخدمة في معالجة البيانات الضخمة، حيث أصبح الاعتماد على تلك النظم والبرمجيات وسيلة هامة تساعد في التعامل مع البيانات الضخمة والسيطرة عليها، وصولا إلى دعم اتخاذ القرار السليم.

٣/٠ أهداف الدراسة

تهدف الدراسة إلى تحقيق هدف عام وهو " حصر وتحليل النظم والبرمجيات السحابية المستخدمة في معالجة البيانات الضخمة" ويتفرع هذا الهدف إلى مجموعة أهداف فرعية، تتمثل في الآتي:

١- التعرف على مفهوم البيانات الضخمة وأسباب ظهورها وأنواعها وخصائصها.
٢- حصر النظم والبرمجيات المستخدمة في معالجة البيانات الضخمة.
٣- تحليل النظم والبرمجيات المستخدمة في معالجة البيانات الضخمة.
٤- توزيع النظم والبرمجيات محل الدراسة نوعيا وكميا وجغرافيا ولغويا وحسب الإتاحة وقيود الاستخدام.

٤/٠ تساؤلات الدراسة:

تسعى الدراسة الى الاجابة على التساؤلات الاتية:

١- التعرف على مفهوم البيانات الضخمة وأسباب ظهورها وأنواعها وخصائصها.
٢- ما النظم والبرمجيات المستخدمة في معالجة البيانات الضخمة؟
٣- ما السمات النوعية للنظم والبرمجيات المستخدمة في معالجة البيانات الضخمة؟
٤- ما السمات الكمية للنظم والبرمجيات المستخدمة في معالجة البيانات الضخمة؟
٥- ما السمات الجغرافية للنظم والبرمجيات المستخدمة في معالجة البيانات الضخمة؟
٦- ما السمات اللغوية للنظم والبرمجيات المستخدمة في معالجة البيانات الضخمة؟

٥/٠ حدود الدراسة:

يمكن تحديد نطاق الدراسة من خلال الحدود التالية: -

١- **الحدود الموضوعية:** البرمجيات السحابية المستخدمة في معالجة البيانات الضخمة.
٢- **الحدود الزمنية:** منذ نوفمبر ٢٠٢٠ حتى مارس ٢٠٢٢

٦/٠ مجتمع الدراسة:

يشكل مجتمع الدراسة برمجيات البيانات الضخمة المتاحة على شبكة الإنترنت، وذلك من خلال البحث في العديد من محركات البحث وباستخدام المصطلحات باللغة العربية والإنجليزية، ومنها (أدوات البيانات الضخمة – برمجيات البيانات الضخمة – تطبيقات البيانات الضخمة – أدوات معالجة وتحليل البيانات الضخمة - big data applications - big data software - big data tools - Big Data Analytics Tools) وقد تبين من خلال البحث عن وجود (١٩٠) برنامجا، ولم يتم استبعاد أي من هذه البرامج.

٧/٠ منهج الدراسة:

فرضت طبيعة وبنية الدراسة الحالية إتباع المنهج الميداني؛ باعتباره أنسب المناهج العلمية لتحقيق ما ترمى إليه الدراسة والذي تم استخدامه في هذه الدراسة لجمع البيانات المتعلقة ببرمجيات البيانات الضخمة السحابية وتحليلها وتفسيرها.

٨/٠ أدوات جمع البيانات:

اعتمدت الدراسة على الأدوات التالية لجمع المادة العلمية الخاصة بموضوع البحث:

١- **قائمة مراجعة:** تضم الأسئلة التي على أساسها تم تحليل برمجيات البيانات الضخمة السحابية.
٢- **جلسات الاتصال المباشر بالإنترنت.**

٩/٠ مراحل الدراسة:

بناء على طبيعة المشكلة البحثية، وفى إطار موضوع الدراسة ومحاور اهتمامها سعت الباحثة لتنفيذ هذه المراحل المتمثلة في: -

- **المرحلة الأولى: حصر وتجميع برمجيات البيانات الضخمة السحابية المتاحة على الإنترنت.**

تم في هذه المرحلة البحث عن برمجيات البيانات الضخمة السحابية بمختلف أنواعها وذلك حتى ٢٩ نوفمبر ٢٠٢١، وذلك من خلال البحث في العديد من محركات البحث وباستخدام المصطلحات باللغة العربية والإنجليزية، ومنها أدوات البيانات الضخمة – برمجيات البيانات الضخمة – تطبيقات البيانات الضخمة – أدوات معالجة وتحليل البيانات الضخمة – big data - big data tools (Big Data Analytics Tools - big data software - applications).

- **المرحلة الثانية: فحص برمجيات البيانات الضخمة السحابية المتاحة على الإنترنت.**

تم في هذه المرحلة فتح كل برنامج من خلال الموقع الخاص بالبرنامج، وفحصه للتأكد من صلاحية عمله.

- **المرحلة الثالثة: تحليل برمجيات البيانات الضخمة السحابية موضوع الدراسة.**

تم في هذه المرحلة الدخول على كل برنامج لتحديد نوعه وإمكاناته وطبيعة عمله، وأسفرت هذه المرحلة عن تحديد ٦ فئات من برمجيات البيانات الضخمة السحابية، وهي:

- برمجيات معالجة وتحليل البيانات الضخمة.
- برمجيات تخزين البيانات الضخمة.
- برمجيات التمثيل المرئي للبيانات.
- برمجيات الذكاء الاصطناعي والتعلم الآلي.
- برمجيات معالجة اللغة الطبيعية NLP.
- برمجيات قواعد البيانات غير العلائقية NoSQL.

تم توزيع هذه البرمجيات نوعيا وكميا وجغرافيا ولغويا وحسب الإتاحة وحسب قيود الاستخدام وتحليلها إحصائيا.

١٠/٠ أسلوب الاستشهادات المرجعية

تم الاعتماد في صياغة الاستشهادات المرجعية في الدراسة على قواعد أسلوب APA الإصدار السابع.

١١/٠ الدراسات السابقة:

وقد قسمت الباحثة الدراسات السابقة إلى قسمين: الأول للدراسات باللغة العربية، الثاني للدراسات باللغة الإنجليزية وقامت بترتيبها من الأحدث للأقدم:

الدراسات باللغة العربية: -

١- دراسة (القصاص، ٢٠١٩)

تهدف هذه الدراسة إلى استكشاف طبيعة البيانات الضخمة Big Data بشكل عام، وفي تخصص المكتبات والمعلومات بشكل خاص، ودراسة التقنيات اللازمة للتعامل مع البيانات الضخمة في كل مراحلها، ومدى تأثير تلك التقنيات على سوق عمل المكتبات، كما اعتمدت هذه الدراسة على المنهج الاستكشافي كمنهج عام لتحديد جوانب الدراسة.

ومن أهم النتائج التي توصلت إليها: أن البيانات الضخمة Big Data في أصلها بيانات عادية، ولكنها تتميز بضخامة الحجم Volume، نتيجة زيادة حجم المستخدمين على مستوى العالم، وتتميز ــ أيضا ــ بالسرعة Velocity من حيث توليد وتحديث هذه البيانات، وبتنوع الشكل والهيئة Format الذي تأتي علية البيانات، وبالتالي تحتاج إلى تقنية أكثر تطوراً في التخزين والمعالجة.

٢- دراسة (البسيونى، ٢٠١٥)

تهدف هذه الدراسة إلى حصر وتحليل مواقع التخزين السحابي المتاحة على الإنترنت، وتوزيعها نوعيًا وجغرافيًا وزمنيًا، وحسب الإتاحة، وبيان إمكاناتها وفئاتها المختلفة، وخصائصها والسمات المميزة بها، وذلك من خلال الاعتماد على المنهج الوصفي التحليلي، وتوصلت الدراسة إلى تحديد ثلاث فئات رئيسة من تلك المواقع، مثلت مواقع رفع الملفات نسبة ٣٩.٦%، بينما بلغت نسبة مواقع استضافة الملفات ٣١.٥%، وبلغت نسبة مواقع النسخ الاحتياطي ٢٨.٩%، وأظهرت الدراسة أن جميع مواقع رفع الملفات موضوع الدراسة تسمح للمستفيد برفع الملفات الموجودة بجهازه الشخصي، أو وسائط التخزين الملحقة به، بينما حرصت ٢٧% منها على إتاحة رفع ملفات منشورة بالفعل على الإنترنت ولها عنوان URL خاص بها، وأن ٢٣.٤% من إجمالي مواقع استضافة الملفات تسمح بإنشاء الملفات وتحريرها، من خلال اعتمادها على محررات خاصة بها.

٣- دراسة (سيد، ٢٠١٥)

تهدف هذه الدراسة إلى التعرف على ماهية التنقيب عن البيانات ومهامه وتطبيقاته، وتحليل أدوات التنقيب عن البيانات مفتوحة المصدر وتقييمها، ومن ثَّم عقد مقارنة بين أدوات التنقيب عن البيانات مفتوحة المصدر. وتوصلت الدراسة للعديد من النتائج، أهمها: هناك مزية يتصف بها بعض الأدوات والتي تتضح من خلال الاستخدام، وهي توفير نموذج السحب والإفلات أثناء عملية التركيب والبناء، للتنقيب عن البيانات، وهي تتوافر بأربع أدوات Orange، Weka RapidMiner.

الدراسات باللغة الإنجليزية: -

١- دراسة (Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2018)

تهدف هذه الدراسة بعمل حصر للتقنيات الحديثة المطورة للبيانات الضخمة، وتوفر عرضًا تفصيلياً لتقنيات البيانات الضخمة الرئيسية، وعقد مقارنات بينها، وفقًا لطبقات النظام المختلفة، مثل: طبقة تخزين البيانات، وطبقة معالجة البيانات، وطبقة استعلام البيانات، وطبقة الوصول إلى البيانات، وطبقة الإدارة. والذي يصنف ويناقش ميزات التقنيات الرئيسية، ومزاياها، وحدودها، واستخداماتها.

٢-دراسة (Khan, Ara Shakil, & Alam, 2017)

توضح هذه الدراسة مفهوم البيانات الضخمة، وتناقش التقنيات الرياضية وتحليلات البيانات التي يمكن استخدامها للبيانات الضخمة، ويعطي تصنيفًا للأدوات والأطر والأنظمة الحالية المتاحة لنماذج حوسبة البيانات الضخمة المختلفة، وإلى جانب ذلك كله يقوم ــ أيضًا ــ بتقييم جدوى حوسبة البيانات الضخمة القائمة على السحابة، ويفحص التحديات والفرص الحالية، ويوفر اتجاهات بحثية مستقبلية في هذا المجال.

٣-دراسة (Hilbert, 2015)

توضح هذه الدراسة الإطار المفاهيمي للبيانات الضخمة، وكذلك التعرف على الفرص والتحديات التي تواجه تحليلات البيانات الضخمة. وكان من أبرز نتائج هذه الدراسة: أن ظهور البيانات الضخمة قدم آفاقًا فعالة من حيث التكلفة، لتحسين عملية اتخاذ القرار في مجالات التنمية الحيوية، مثل: الرعاية الصحية، والإنتاجية، والاقتصادية، والأمن العام، في الوقت نفسه يوجد مجموعة من المخاوف المتعلقة بالخصوصية وندرة الموارد البشرية في البلدان النامية، بسبب النقص الهيكلي طويل الأمد في مجالات البنية التحتية، والموارد الاقتصادية والمؤسسات. والنتيجة هي نوع جديد من الفجوة الرقمية ـــ فجوة في استخدام المعرفة القائمة على البيانات لتوجيه عملية صنع القرار الذكي ـ

٤- دراسة (Hacken, 2015)

تتناول هذه الدراسة البيانات الضخمة وعلاقتها بالحوسبة السحابية، وتطبيقات الموبايل في المكتبات الرقمية، ثم تطرقت إلى تحديات تطبيق البيانات الضخمة في المكتبات الرقمية، وكذلك أوجه الاستفادة منها في المكتبات الرقمية.

"النظم والبرمجيات السحابية المستخدمة في معالجة البيانات الضخمة"

(١/١) الإطار النظري للدراسة

(١/١/١) تعريف البيانات الضخمة

وهناك عدة محاولات كثيرة لوضع تعريف لمصطلح "البيانات الضخمة" من قبل مجموعة من المؤسسات المتخصصة في مجال تكنولوجيا المعلومات وبعض الأشخاص من محللي البيانات وعلماء التقنية، ومن أهم هذه التعريفات:

تعرفها شركة مايكروسوفت بأنها مجموعات ضخمة من البيانات ذات قيمة من الصعب تخزينها وتحليلها وإدارتها باستخدام أدوات إدارة البيانات أو تطبيقات معالجة البيانات التقليدية، ويرجع ذلك بسبب مشكلات الحجم والتعقيد بجانب قيود التكلفة وغياب الآليات المناسبة. (Microsoft, 2016)

(٢/١/١) أسباب ظهور البيانات الضخمة

هناك عدة أسباب أدت لظهور البيانات الضخمة وهي كالآتي:

١. ظهور تقنيات إنترنت الأشياء، والتي تسمح لجميع الأجهزة الاتصال ببعضها عن طريق الإنترنت وإنتاج بيانات جديدة. (Wamba, Akter, Edwards, Chopin, & Gnanzou , 2015)

٢. ظهور مجالات بحثية تنتج بيانات ضخمة جدا وجب تحليلها، مثل: علم الجينوم، والمحاكاة الفيزيائية المعقدة، والبحوث البيولوجية، والبيئية.
٣. ظهور الشبكات الاجتماعية التي ترسل كما ضخما من البيانات على مدار الساعة، ومن مختلف الجهات والهيئات.
٤. القوانين التي تحتم ضرورة بقاء هذه البيانات في قواعد البيانات، لمتابعة المجرمين والمخربين والمتسللين.
٥. انخفاض تكاليف تخزين هذه البيانات كنظم الحوسبة السحابية. (سليم، ٢٠١٩)

(١/١/٣) أنواع البيانات الضخمة

يشير Abhinav Rai إلى أنواع البيانات الضخمة كالآتي:

- **البيانات المهيكلة Structured data**

هي البيانات المنظمة التي يمكن معالجتها وتخزينها واستردادها بشكل ثابت والوصول إليها بسهولة من قاعدة بيانات بواسطة خوارزميات محرك بحث بسيطة

- **البيانات الغير مهيكلة Unstructured data**

هي البيانات الغير منظمة التي تفتقر إلى شكلٍ أو هيكلٍ معينٍ، وهذا ما يجعل الأمر صعباً جداً، ويستغرق وقتاً طويلاً في معالجة تلك البيانات وتحليلها.

- **البيانات شبه الهيكلية Semi-structured data**

هي تلك البيانات التي تحتوي على كل من البيانات المنظمة والبيانات غير المنظمة، وعلى وجه الدقة فإنه يشير إلى البيانات التي على الرغم لم من أنها لم يتم تصنيفها في مستودع معين (قاعدة بيانات)، إلا أنها تحتوي على معلومات. (Rai, 2019)

(٤/١/١) خصائص البيانات الضخمة:

تذكرMargaret Rouse بأنه يتشكل الأساس التعريفي الأكثر استخدامًا للبيانات الضخمة من ثلاث V هي: الحجم Volume، السرعة Velocity، التنوع Varity:

١- **الحجم (Volume):** كمية كبيرة من البيانات بأحجام التيرابايت إلى زيتابايت.
٢- **السرعة (Velocity):** عندما تكون هناك كميات كبيرة من البيانات، يؤدى ذلك إلى تدفق البيانات بسرعة كبيرة.
٣- **التنوع (Varity):** يمكن أن تأتي البيانات من مصادر بيانات مختلفة، وفى أشكال مختلفة مثل: بيانات المعاملات، والسجلات، والتطبيقات المختلفة، والبيانات المنظمة، كجدول قاعدة بيانات، وبيانات شبه منظمة كبيانات XML، والبيانات غير المنظمة كالنصوص، والصور، والفيديو، والبيانات الصوتية، وغيرها. (Rouse, n.d)

(٢/١) الإطار العملي للدراسة

(١/٢/١) التوزيع النوعي والكمي لبرمجيات البيانات الضخمة محل الدراسة

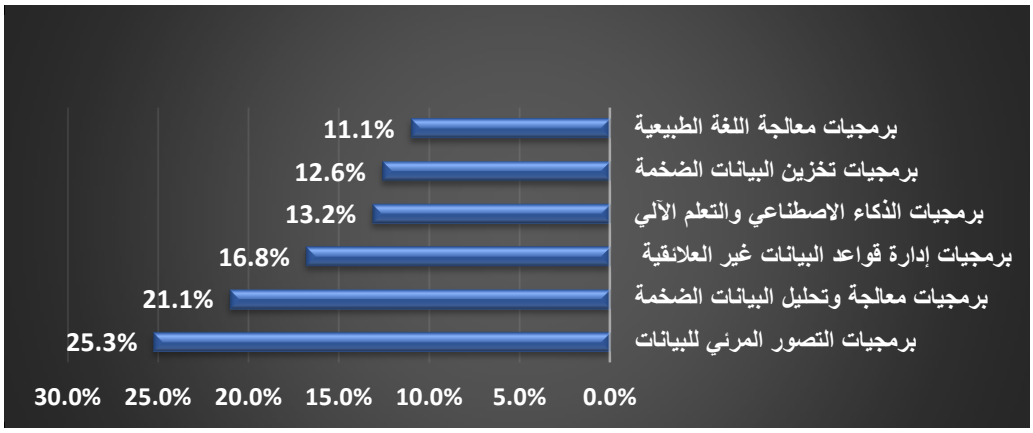
من خلال حصر النظم والبرمجيات الذي قامت به الباحثة، فقد حصرت ١٩٠ برنامجا وجدت أنها تتنوع وفقا لطبيعة عملها، لذا قامت بتقسيمها إلى ست فئات لكل فئة خصائصها وسماتها، موضحة في الجدول التالي:

جدول رقم (١) التوزيع النوعي والكمي لبرمجيات البيانات الضخمة محل الدراسة

م	النوع	عدد البرمجيات	النسبة
1	برمجيات التمثيل المرئي للبيانات	48	%25.3
2	برمجيات معالجة وتحليل البيانات الضخمة	40	%21.1
3	برمجيات قواعد البيانات غير العلائقية NoSQL	32	%16.8
4	برمجيات الذكاء الاصطناعي والتعلم الآلي	25	%13.2
5	برمجيات تخزين البيانات الضخمة	24	%12.6
6	برمجيات معالجة اللغة الطبيعية NLP	21	%11.1
	الإجمالي	190	% 100

بتحليل بيانات الجدول السابق رقم (١) تبين الآتي:

أن برمجيات التمثيل المرئي للبيانات جاءت في المرتبة الأولى بأعلى نسبة حيث مثلت ٢٥.٣% بواقع (٤٨) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة، ويرجع السبب في ذلك لسهولة استخدامها وأن الكثير منها لا يحتاج للخبرة في التعامل معها حيث تناسب المستخدم العادي وليس فقط المطورين والمبرمجين ومحللي ومهندسي البيانات، يليها في المرتبة الثانية برمجيات معالجة وتحليل البيانات الضخمة بنسبة ٢١.١% بواقع (٤٠) برنامج، أما في المرتبة الثالثة فجاءت برمجيات إدارة قواعد البيانات غير العلائقية بنسبة ١٦.٨% بواقع (٣٢) برنامج، وجاءت في المرتبة الرابعة برمجيات الذكاء الاصطناعي والتعلم الآلي بنسبة ١٣.٢% بواقع (٢٥) برنامج، ثم في المرتبة الخامسة جاءت برمجيات تخزين البيانات الضخمة بنسبة ١٢.٦% بواقع (٢٤) برنامج، وأخيرا جاءت برمجيات معالجة اللغة الطبيعية بنسبة ١١.١% بواقع (٢١) برنامج من إجمالي البرمجيات محل الدراسة



شكل رقم (١) التوزيع النوعي والكمي لبرمجيات البيانات الضخمة محل الدراسة

(٢/٢/١) التوزيع الجغرافي لبرمجيات البيانات الضخمة موضوع الدراسة

من خلال الاطلاع على مواقع برمجيات البيانات الضخمة تبين أن السمات المكانية لبرمجيات البيانات الضخمة محل الدراسة موزعة على (٢٣) دولة أجنبية لها إسهام في إنتاج تلك البرمجيات ولا يوجد أي دولة عربية:

جدول رقم (٢) التوزيع الجغرافي لبرمجيات البيانات الضخمة محل الدراسة

م	الدول	عدد البرمجيات	النسبة
1	الولايات المتحدة	143	%73
2	المملكة المتحدة	8	%4.1
3	فرنسا	7	%3.6
4	كندا	6	%3.1
5	ألمانيا	4	%2
6	السويد	3	%1.5
7	الهند	3	%1.5
8	إسرائيل	3	%1.5
9	سويسرا	3	%1.5
10	نيوزلندا	2	%1
11	هولندا	2	%1
12	البرازيل	1	%0.5
13	الصين	1	%0.5
14	النرويج	1	%0.5
15	اليابان	1	%0.5
16	إيطاليا	1	%0.5
17	أستراليا	1	%0.5
18	أوكرانيا	1	%0.5
19	أيرلندا	1	%0.5
20	بيلاروس	1	%0.5
21	تشيكيا	1	%0.5
22	سلوفينيا	1	%0.5
23	لاتفيا	1	%0.5
الإجمالي		196	%100

بتحليل بيانات الجدول السابق رقم (٢) تبين الآتي: أن الولايات المتحدة الأمريكية احتلت الصدارة الأولى انتاجا لبرمجيات البيانات الضخمة محل الدراسة بأعلى نسبة حيث مثلت ٧٣% بواقع (١٤٣) برنامج ، ويرجع ذلك لأن الولايات المتحدة تعد من أهم الدول المتقدمة في مجال التكنولوجيا منذ آواخر

القرن التاسع عشر ، كما أنها تعد موطنا للكثير من العلامات التجارية الرائدة في عالم التكنولوجيا، تليها في المرتبة الثانية المملكة المتحدة بنسبة ٤.١% حيث ظهر لها (٨) برمجيات ، وجاءت في المرتبة الثالثة فرنسا بنسبة ٣.٦% حيث ظهر لها (٧) برمجيات، ثم كندا بنسبة ٣.١% حيث ظهر لها (٦) برمجيات، ثم ألمانيا بنسبة ٢% حيث ظهر لها (٤) برمجيات ، ثم كلاً من السويد والهند وإسرائيل وسويسرا بنسبة ١.٥% بواقع (٣) برمجيات لكل منها ، ثم كل من نيوزلندا وهولندا بنسبة ١% بواقع (٢) برنامج لكل منها ، ثم اثنا عشرة دولة بنسبة ٠.٥% حيث لم يكن في حصيلة إنتاجها سوى برنامج واحد من برمجيات البيانات الضخمة محل الدراسة ، كما اشتركت بعض الدول في انتاج بعض البرمجيات وكانت منها (الهند - الولايات المتحدة) اشتركا في برنامجين Apex Charts و Fusion Charts ، و (الولايات المتحدة - المملكة المتحدة) في برنامج D3.JS ، و (المملكة المتحدة - إيرلندا) في برنامج Instant Atlas ، و (أستراليا - المملكة المتحدة) في برنامج NLTK ، و(نيوزلندا – كندا) في برنامج R

(٣/٢/١) التوزيع اللغوى لبرمجيات البيانات الضخمة موضوع الدراسة

من خلال الاطلاع على مواقع برمجيات البيانات الضخمة تبين أن السمات اللغوية لمواقع برمجيات البيانات الضخمة محل الدراسة تنقسم إلى قسمين هما:

القسم الأول: سمات لغوية خاصة بواجهة موقع البرنامج (اللغة التي يعرض بها الموقع البيانات الخاصة بالبرنامج) حيث تبين أن هناك برمجيات تدعم لغات متعددة وبرمجيات تدعم لغة واحدة وهي الإنجليزية

القسم الثاني: سمات لغوية تدعمها البرمجيات في حد ذاتها، وقد تبين أن جميع برمجيات البيانات الضخمة محل الدراسة تعرض البرنامج باللغة الإنجليزية فقط، فيما عدا البرمجيات التي تعمل on line فإنها تعمل بنفس لغة موقع البرنامج

جدول رقم (٣) اللغات التي تدعمها برمجيات البيانات الضخمة محل الدراسة

م	اللغة	عدد البرمجيات	النسبة
1	اللغة الإنجليزية	186	% 37.7
2	اللغة الألمانية	37	% 7.5
3	اللغة الفرنسية	36	% 7.3
4	اللغة الإسبانية	33	%6.7
5	اللغة البرتغالية	25	%5.1
6	اللغة اليابانية	23	%4.7
7	اللغة الصينية	22	% 4.5
8	اللغة الكورية	18	% 3.6
9	اللغة الإندونيسية	17	% 3.4
10	اللغة الروسية	15	%3.0
11	اللغة العربية	14	% 2.8
12	اللغة التركية	12	% 2.4
13	اللغة الفيتنامية	10	%2.0
14	اللغة التايلاندية	9	% 1.8
15	اللغة الإيطالية	6	% 1.2
16	اللغة البولندية	5	% 1
17	اللغة الهولندية	5	% 1
18	جميع اللغات	5	% 1
19	اللغة النرويجية	4	% 0.8
20	اللغة السويدية	3	% 0.6
21	اللغة المجرية	3	% 0.6
22	اللغة الهندية	2	% 0.4
23	اللغة الأوكرانية	1	% 0.2
24	اللغة التشيكية	1	%0.2
25	اللغة الدنماركية	1	% 0.2
26	اللغة السلوفانية	1	% 0.2
	الإجمالي	494	% 100

بتحليل بيانات الجدول السابق رقم (٣) تبين الآتي:

أن اللغة الإنجليزية تمثل أعلى نسبة في عدد البرمجيات التي تدعمها بنسبة ٣٧.٤% وبواقع (١٨٦) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة، ويرجع ذلك لأنها اللغة الأم والسائدة بالنسبة للدول الأجنبية كما أنها لغة عالمية وأكثر لغات العالم انتشارًا، على عكس باقي اللغات حيث مثلت اللغة الألمانية ٧.٥% بواقع (٣٧) برنامج، يليها في ذلك اللغة الفرنسية بنسبة ٧.٣% بواقع (٣٦)

برنامج، ثم اللغة الإسبانية بنسبة ٦.٧% بواقع (٣٣) برنامج ، في حين مثلت اللغة العربية ٢.٨% بواقع (١٤) برنامج من إجمالي برمجيات البيانات الضخمة ، وأخيرا مثلت اللغة الأوكرانية و الدنماركية و السلوفانية والتشيكية أقل نسبة ٠.٢% بواقع برنامج واحد فقط لكل منها من إجمالي برمجيات البيانات الضخمة محل الدراسة ، وكانت هناك (٥) برمجيات تدعم جميع اللغات بنسبة ١ %

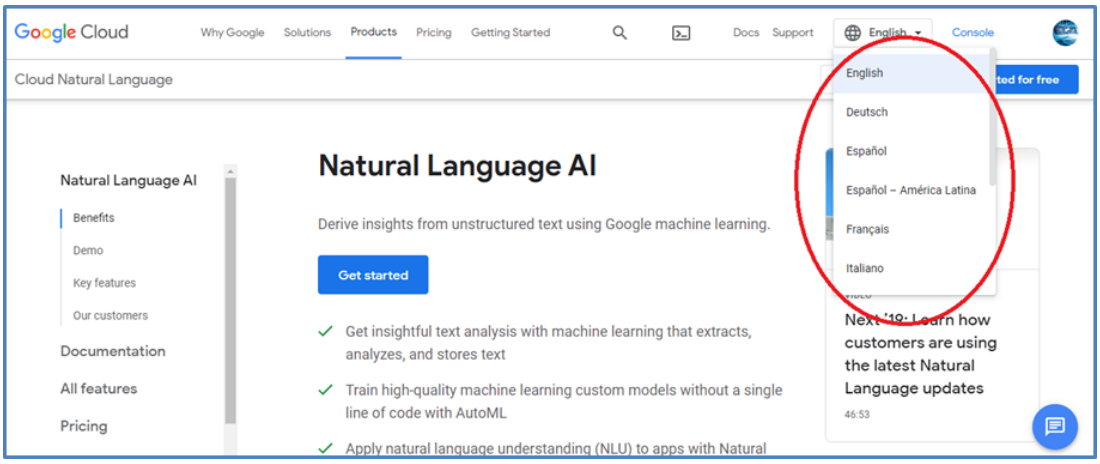
✓ ونظرا لأن مواقع برمجيات البيانات الضخمة تتاح باللغة الإنجليزية ولغات آخري، يتم تقسيم البرمجيات إلى برمجيات أحادية اللغة وبرمجيات متعددة اللغات كالآتي:

جدول رقم (٤) توزيع برمجيات البيانات الضخمة محل الدراسة إلى أحادية ومتعددة اللغات

م	البرمجيات	المجموع	النسبة
1	برمجيات أحادية اللغة	141	% 74.2
2	برمجيات متعددة اللغات	49	%25.8
	الإجمالي	190	% 100

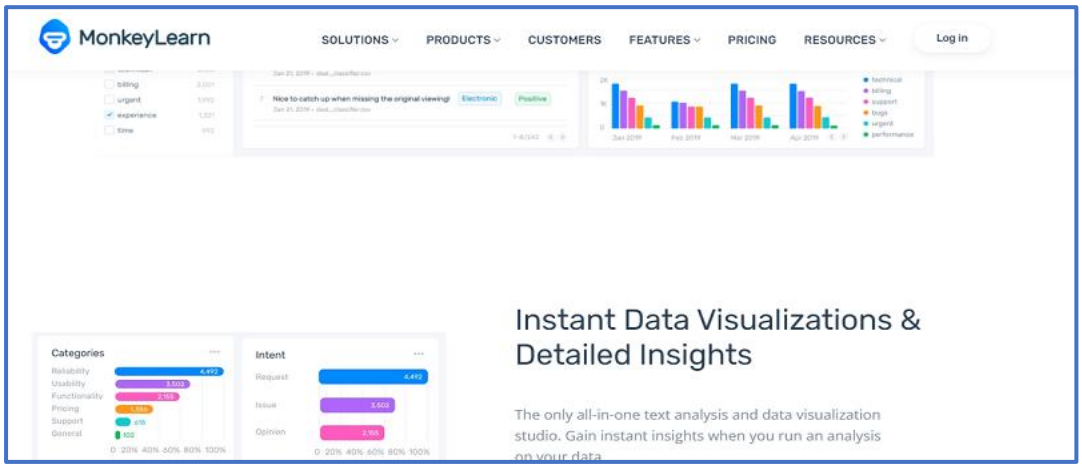
ومن أمثلة البرمجيات متعددة اللغات

برنامج Google Cloud Natural Language API



شكل رقم (٢) عدد اللغات التي يدعمها برنامج Google Cloud Natural Language API

ومن أمثلة البرمجيات أحادية اللغة



شكل رقم (٣) اللغة التي يدعمها برنامج Monkey Learn

(٤/٢/١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب الإتاحة

من خلال الاطلاع على مواقع برمجيات البيانات الضخمة تبين أنها تختلف في طرق الإتاحة للمستفيد منها، كما هو موضح في الجدول التالي: -

جدول رقم (٥) توزيع برمجيات البيانات الضخمة محل الدراسة حسب الإتاحة

م	الاتاحة	عدد البرمجيات	النسبة
1	مفتوح المصدر	144	% 75.8
2	مغلق المصدر	46	%24.2
	الاجمالي	190	% 100

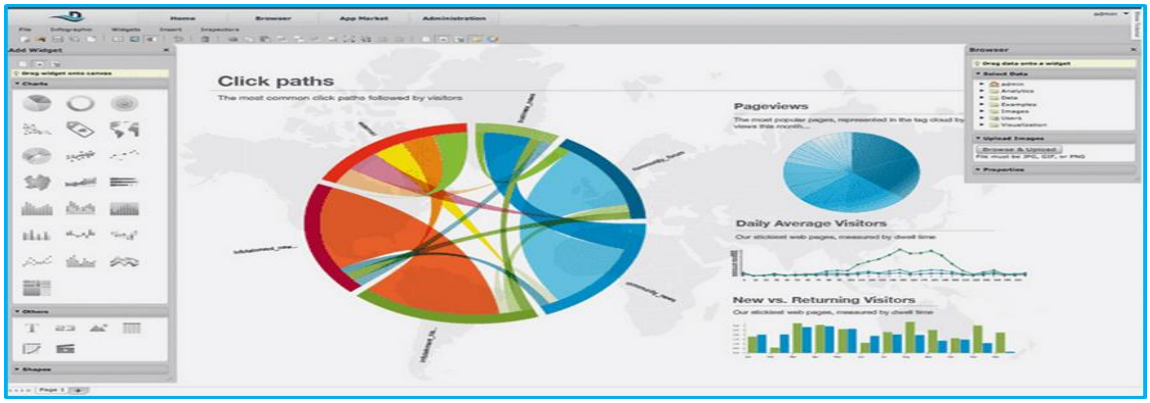
بتحليل بيانات الجدول السابق رقم (٥) تبين الآتي: أن أكثر برمجيات البيانات الضخمة مفتوحة المصدر، حيث تمثل تلك البرمجيات نسبة ٧٥.٨% % بواقع (١٤٤) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة، بينما مثلت برمجيات البيانات الضخمة مغلقة المصدر نسبة ٢٤.٢% بواقع (٤٦) برنامجا، ويرجع ذلك لعدة أسباب منها تفضيل العديد من الأشخاص البرامج مفتوحة المصدر لأن لديهم سيطرة أكبر على هذا النوع من البرامج، مما يمكنهم من فحص الكود للتأكد من أنه لا يفعل أي شيء لا يريدون القيام به، كما أنها أكثر أمانًا واستقرارًا من البرامج مغلقة المصدر الاحتكارية. نظرًا لأنه يمكن لأي شخص عرض وتعديل البرامج مفتوحة المصدر، فقد يكتشف شخص ما الأخطاء أو الإغفالات التي ربما فاتها المؤلفون الأصليون للبرنامج ويصححها دون طلب إذن من المؤلفين الأصليين.

ومن أمثلة البرمجيات مفتوحة المصدر

شكل رقم (٤) برنامج Apache pig كأحد برمجيات البيانات الضخمة مفتوحة المصدر

ومن أمثلة البرمجيات مغلقة المصدر

برنامج Datameer



شكل رقم (٥) برنامج Datameer أحد برمجيات البيانات الضخمة مغلقة المصدر

(٥/٢/١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب قيود الاستخدام

من خلال الاطلاع على مواقع برمجيات البيانات الضخمة تبين أنها تختلف موزعة حسب قيوم الاستخدام، كما هو موضح في الجدول التالي: -

جدول رقم (٦) توزيع برمجيات البيانات الضخمة محل الدراسة حسب قيود الاستخدام

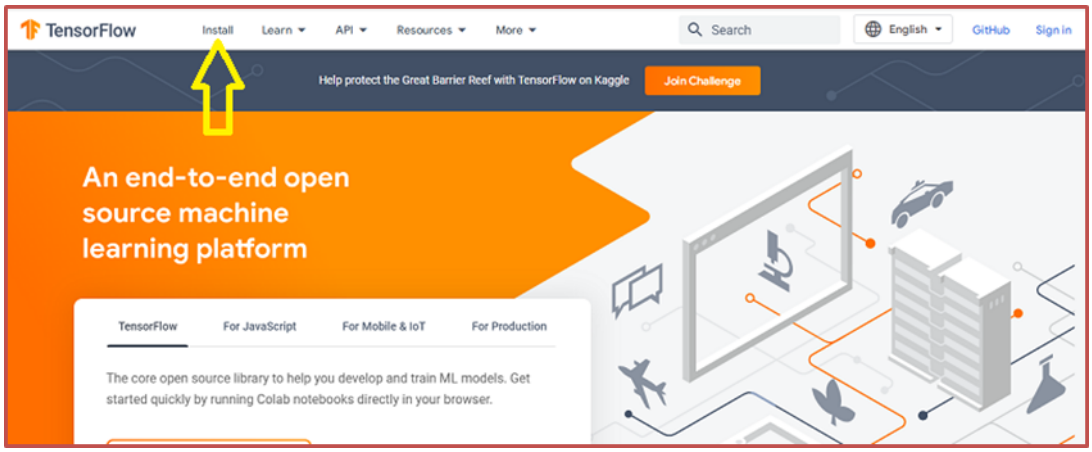
م	قيود الاستخدام	العدد	النسبة
1	مجاني	145	% 76.3
2	مدفوع	45	%23.7
	الاجمالي	190	% 100

بتحليل بيانات الجدول السابق رقم (٦) تبين الآتي:

أن حوالي ٧٦.٣% من إجمالي برمجيات البيانات الضخمة متاحة بشكل مجاني بواقع (١٤٥) برنامج، بينما وصلت البرمجيات التجارية المدفوعة ٢٣.٧% بواقع (٤٥) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة، ويوضح الشكل التالي رقم (١٣) توزيع برمجيات البيانات الضخمة محل الدراسة حسب قيود الاستخدام.

ومن أمثلة برمجيات البيانات الضخمة المجانية

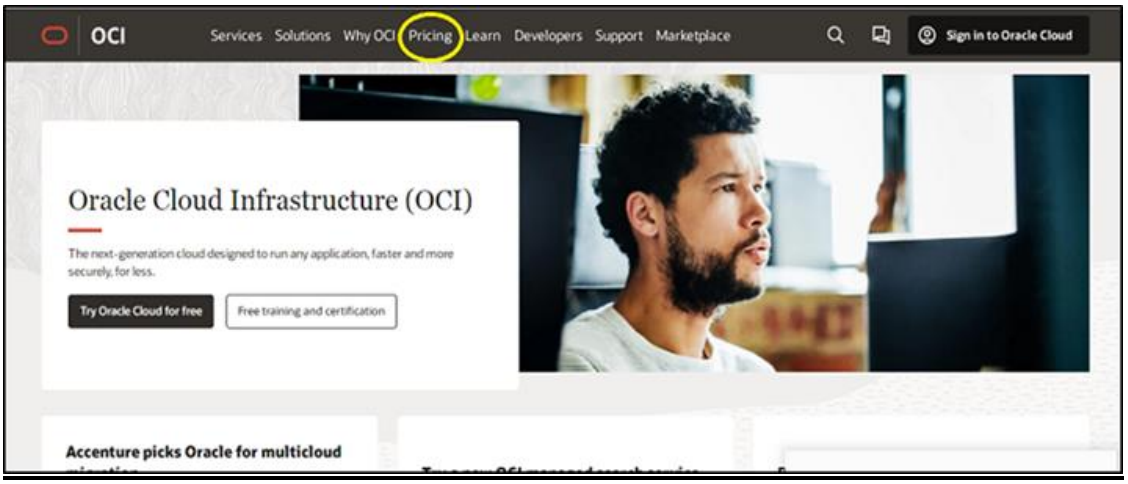
برنامج TensorFlow



شكل رقم (٦) برنامج Tensor Flow كأحد برمجيات البيانات الضخمة المجانية

ومن أمثلة برمجيات البيانات الضخمة المدفوعة

Oracle Cloud



شكل رقم (٧) Oracle Cloud كأحد برمجيات البيانات الضخمة المدفوعة

(٦/٢/١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب التراخيص

من خلال الاطلاع على برمجيات البيانات الضخمة تبين أن هناك اختلافاً في ترخيص برمجيات البيانات الضخمة، منها ما يعتمد على رخص برمجيات مفتوحة المصدر مثل رخصة جنو العمومية GNU، ورخصة MIT، ورخصة أباتشي ٢.٠ ومنها ما يعتمد على التراخيص الاحتكارية التي تقع تحت ملكية الشركات، ويوضح الجدول التالي رقم (٧) التوزيع حسب تراخيص برمجيات البيانات الضخمة محل الدراسة

جدول رقم (٧) توزيع برمجيات البيانات الضخمة محل الدراسة حسب التراخيص

م	الترخيص	عدد البرمجيات	النسبة
1	Apache License 2.0	78	%41.1
2	Proprietaryامتلاكي / احتكاري	42	% 22.1
3	رخصة جنو العمومية GNU	24	% 12.6
4	MIT License	21	% 11.1
5	BSD License	15	% 7.9
6	Elastic License 2.0	2	%1.1
7	غير محدد	2	% 1.1
8	Eclipse Public Licenseرخصة إكليبس العمومية	1	% 0.5
9	Microsoft Public Licenseرخصة مايكروسوفت العامة	1	% 0.5
10	ISC License	1	% 0.5
11	EULAترخيص المستخدم النهائي	1	% 0.5
12	PSFLرخصة مؤسسة برمجيات بايثون	1	% 0.5
13	SSPLرخصة جانب الخادم العامة	1	%0.5
	الإجمالي	190	% 100

بتحليل بيانات الجدول السابق رقم (٧) تبين الآتي: أن أكثر البرمجيات محل الدراسة تعتمد على Apache License 2.0 في ترخيصها (٧٧) برنامج بنسبة ٤١.١ % ، ويرجع ذلك لأن معظم برمجيات البيانات الضخمة هي من إنتاج مؤسسة برمجيات أباتشي التي تعتمد على ترخيصها ، يليها الترخيص الاحتكاري (الامتلاكى) Proprietary بواقع (٤٢) برنامج بنسبة ٢٢.١% ، يليها رخصة جنو العمومية GNU بواقع (٢٥) برنامج بنسبة ١٣.٢% ، يليها ترخيص MIT بواقع (٢١) بنسبة ١١.١% ، يليها ترخيص BSD بواقع (١٥) برنامج بنسبة ٧.٩% ، يليها Elastic License 2.0 بواقع (٢) برنامج بنسبة ١.١% ، يليها كل من رخصة إكليبس العمومية و رخصة مايكروسوفت العامة و ترخيص ISC و رخصة مؤسسة برمجيات بايثون PSFL و رخصة جانب الخادم العامة SSPL بواقع برنامجا واحدا بنسبة ٠.٥% ، ويوجد (٢) برنامج غير محدد ترخيصهما

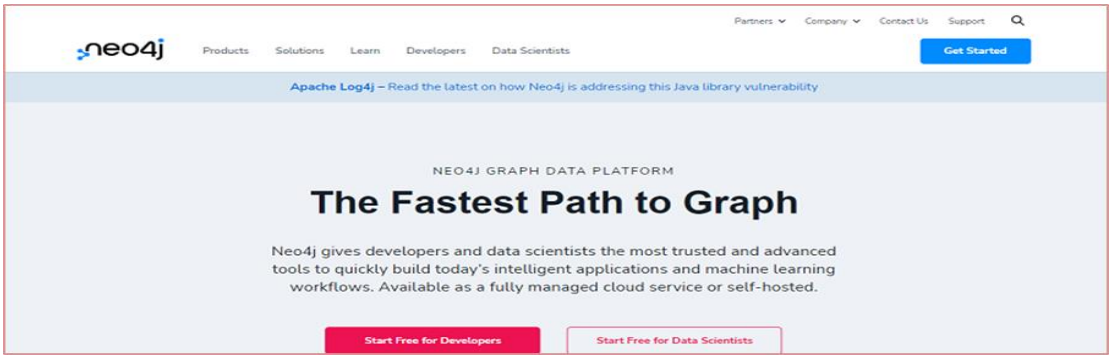
ومن أمثلة برمجيات البيانات الضخمة التي تعتمد على ترخيص Apache License 2.0

برنامج Apache Spark

شكل رقم (٨) أحد برمجيات البيانات الضخمة التي تعتمد على ترخيصApache License 2.0

<u>ومن أمثلة برمجيات البيانات الضخمة التي تعتمد على ترخيص GNU</u>

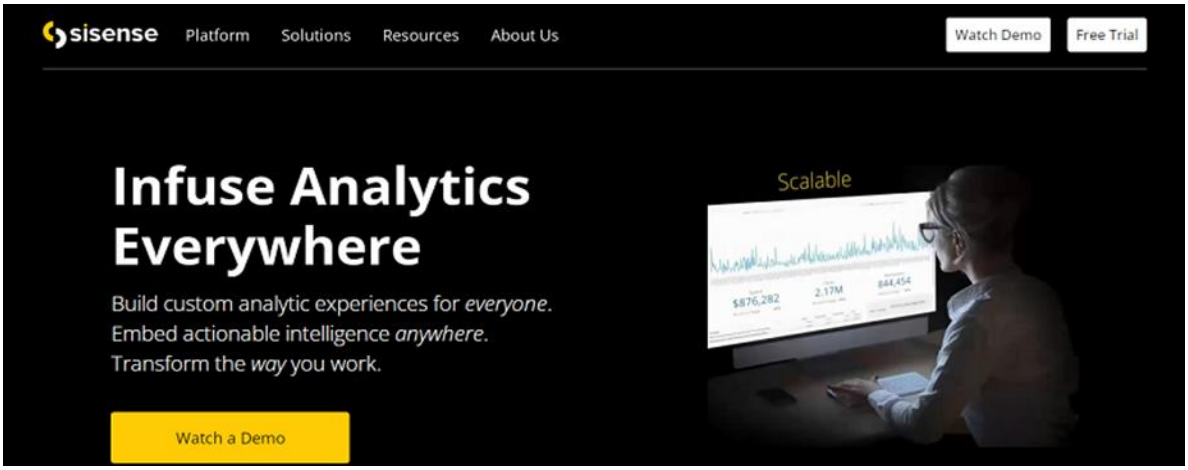
برنامج neo4j



شكل رقم (٩) أحد برمجيات البيانات الضخمة التي تعتمد على ترخيص GNU

<u>ومن أمثلة برمجيات البيانات الضخمة التي تعتمد على ترخيص MIT</u>

برنامج Sisense



شكل رقم (١٠) أحد برمجيات البيانات الضخمة التي تعتمد على ترخيص MIT

(٧/٢/١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب مطوري البرمجيات

من خلال الاطلاع على مواقع برمجيات البيانات الضخمة تبين أن تطوير البرمجيات يخضع لفئتين، كما هو موضح في الجدول التالي: -

جدول رقم (٨) توزيع برمجيات البيانات الضخمة محل الدراسة حسب مطوري البرمجيات

م	مطوري البرمجيات	عدد البرمجيات	النسبة
1	شركات ومؤسسات	164	% 86.3
2	أفراد	26	% 13.7
	الاجمالي	190	% 100

بتحليل بيانات الجدول السابق رقم (٨) تبين الآتي: أن معظم مطوري برمجيات البيانات الضخمة محل الدراسة هم من الشركات التكنولوجية والمؤسسات، حيث مثل إنتاج الشركات والمؤسسات نسبة ٨٦.٣% بواقع (١٦٤) برنامج، في حين مثل إنتاج الأفراد من المطورين والمبرمجين نسبة ١٣.٧% بواقع (٢٦) برنامجا من إجمالي برمجيات البيانات الضخمة محل الدراسة، ويرجع ذلك لأن إنتاج البرمجيات ضخم ومكلف يحتاج إلى شركات ومؤسسات راعية تتبناه وتقدمه إلى سوق البرمجيات

ومن أهم الشركات والمؤسسات التي كان لها إسهام في إنتاج برمجيات البيانات الضخمة محل الدراسة:

جدول رقم (٩) الشركات والمؤسسات المنتجة لبرمجيات البيانات الضخمة محل الدراسة

م	الشركة / المؤسسة	عدد البرمجيات	النسبة
1	مؤسسة برمجيات أباتشي	35	%40.2
2	Google	11	%12.6
3	أمازون	8	%9
4	Microsoft	3	%3.4
5	IBM	4	%4.6
6	شركة Oracle	3	%3.4
7	Facebook	2	%2.3
8	Andrew W. Mellon Foundation	2	%2.3
9	معهدSAS	2	%2.3
10	شركة Elastic	2	%2.3
11	Apple	1	%1.1
12	Twitter	1	%1.1
13	المعهد الفيدرالي للتكنولوجيا	1	%1.1
14	Social Media Research Foundation	1	%1.1
15	جامعة ليوبليانا	1	%1.1
16	جامعة كونستانز	1	%1.1
17	جامعة لومير ليون ٢	1	%1.1
18	جامعة ستانفورد	1	%1.1
19	جامعة وايكاتو	1	%1.1
20	معهد تكنولوجيا الأعمال الإلكترونية بجامعة هونغ كونغ	1	%1.1
21	مختبر أبحاث UC Berkeley AMPLab	1	%1.1
22	مختبر أبحاث T&AT	1	%1.1
23	المعمل البحثي ببوليتكنك ميلانو	1	%1.1
24	CNCF	1	%1.1
25	شركات آخري	1	%1.1
	الإجمالي	87	%100

بتحليل بيانات الجدول السابق رقم (٩) تبين الآتي:

أن مؤسسة برمجيات أباتشي احتلت الصدارة الأولى في أكثر الشركات إنتاجا لبرمجيات البيانات الضخمة محل الدراسة بأعلى نسبة حيث مثلت ٤٠.٢% بواقع (٣٥) برنامجا، تليها في المرتبة الثانية شركة جوجل بنسبة ١٢.٦% حيث أنتجت (١١) برنامج، وجاءت في المرتبة الثالثة شركة أمازون بنسبة ٩% حيث أنتجت (٨) برمجيات، ثم شركة IBM بنسبة ٤.٦% حيث أنتجت (٤) برمجيات، ثم شركة Microsoft و Oracle بنسبة ٣.٤% بواقع (٣) برمجيات لكل منها ، ويرجع ذلك لأنهم من أكبر الشركات التقنية في العالم المهيمنة على سوق التكنولوجيا ، ثم كل من شركة Facebook ومؤسسة أندرو دبليو ميلون Andrew W. Mellon و معهد SAS وشركة Elastic بنسبة ٢.٣% بواقع (٢) برنامج لكل منها، وأخيرا شركة Apple و شركة Twitter و المعهد الفيدرالي للتكنولوجيا و مؤسسة أبحاث وسائل التواصل الاجتماعي Social Media Research Foundation و جامعة ليوبليانا و جامعة كونستانز وجامعة لومير ليون ٢ و جامعة ستانفورد و جامعة وايكاتو ومعهد تكنولوجيا الأعمال الإلكترونية (ETI) بجامعة هونغ كونغ و مختبر أبحاث UC Berkeley AMPLab و مختبر أبحاث AT&T و المعمل البحثي ببوليتكنك ميلانو و Cloud Native Computing Foundation (CNCF) و٧٩ شركة من الشركات الأخرى بنسبة ١.١% لكل منها برنامجا واحدا

(٨/٢/١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب لغات البرمجة

من خلال الاطلاع على برمجيات البيانات الضخمة تبين أن تطوير تلك البرمجيات يتم باستخدام لغات البرمجة حيث تم استخدام ٢٧ لغة برمجة، كما هو موضح في الجدول التالي: -

جدول رقم (١٠) توزيع برمجيات البيانات الضخمة محل الدراسة حسب لغات البرمجة

م	لغات البرمجة	عدد البرمجيات	النسبة
1	Java	84	%23
2	Python	67	%18.3
3	JavaScript	51	%13.9
4	C++	37	%10.1
5	C	17	%4.6
6	Go	16	%4.4
7	Scala	13	%3.6
8	SQL	10	%2.7
9	Ruby	10	%2.7
10	R	8	%2.2
11	C#	7	%1.9
12	HTML	6	%1.6
13	.NET	5	%1.4
14	PHP	4	%1.1
15	TypeScript	4	%1.1
16	CSS	4	%1.1
17	Erlang	4	%1.1
18	Cython	3	%0.8
19	Fortran	3	%0.8
20	Objective-C	2	%0.5
21	Wolfram Language	2	%0.5
22	Perl	2	%0.5
23	Shell	2	%0.5
24	Node.js	2	%0.5
25	Looker	1	%0.3
26	DAX query	1	%0.3
27	XML	1	%0.3
	الإجمالي	366	%100

بتحليل بيانات الجدول السابق رقم (١٠) تبين الآتي:

أن عدد البرمجيات التي اعتمدت على لغة الجافا Java (٨٤) برنامجا بنسبة ٢٣% من إجمالي برمجيات البيانات الضخمة محل الدراسة، يليها لغة البايثون حيث مثلت نسبة ١٨.٣% بواقع (٦٧)

برنامج، يليها لغة الجافا سكريبتJavaScript بنسبة ١٣.٩%، ويرجع ذلك لأنهم يتم اعتبارهم أهم اللغات المخصصة لمعالجة البيانات، يليهم لغة C++ بنسبة ١٠.١% بواقع (٣٧) برنامج، ثم لغة C بنسبة ٤.٦% بواقع (١٧) برنامج، يليها لغة Go بنسبة ٤.٤% بواقع (١٦) برنامج، وأخيرا مثلت لغة Looker وDAX query وXML أقل نسبة ٠.٣% بواقع برنامجا واحدا ويرجع ذلك لأنها لغات برمجة فرعية مكملة للغات الأساسية التي تعتمد عليها معالجة البيانات، ويتضح أيضا من تحليل بيانات الجدول أن هناك برمجيات اعتمدت على أكثر من لغة في برمجتها وتطويرها مثل برنامج Apache MXNet حيث اعتمد على C++ و Python و R و Java و JavaScript و Scala و Go و Perl ، وبرنامج Heroku اعتمد على Java و Node.js و Scala و Python و PHP وGo

(٩/٢/١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب أنظمة التشغيل

من خلال الاطلاع على برمجيات البيانات الضخمة تبين أن هناك تنوع في أنظمة التشغيل التي تعتمد عليها برمجيات البيانات الضخمة محل الدراسة، فمنها ما يعتمد على أنظمة سطح المكتب ومنها ما يعتمد على أنظمة الهواتف الذكية، ومنها ما يتم تشغيله على الويب " on line"، ويوضح الجدول التالي رقم (١١) التوزيع حسب أنظمة تشغيل البرمجيات محل الدراسة.

جدول رقم (١١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب أنظمة التشغيل

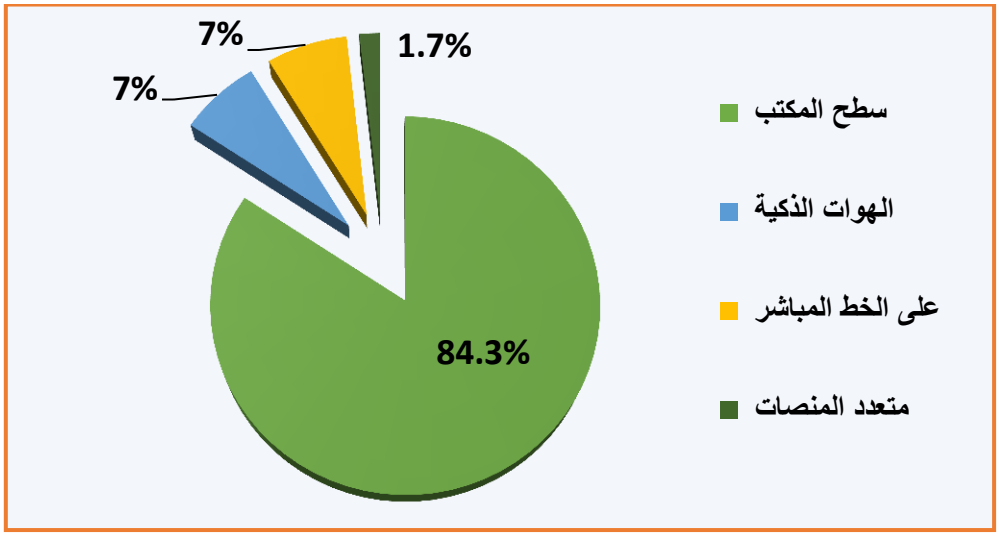
م	نظم التشغيل		عدد البرمجيات	النسبة
1	سطح المكتب	Windows	185	% 32.3
		Mac OS X	136	% 23.7
		Linux	135	% 23.6
		Unix	24	% 4. 2
		Debian	2	% 0.3
		Raspbian	1	% 0.2
	المجموع		483	% 84.3
2	الهواتف الذكية	Android	21	%3.7
		IOS	19	% 3.3
	المجموع		40	% 7
3	مباشر (On line)	Web App	40	7%
4	متعدد المنصات		10	% 1.7
	الإجمالي		573	% 100

بتحليل بيانات الجدول السابق رقم (١١) تبين الآتي:

١-أن ٨٤.٣% من إجمالي برمجيات البيانات الضخمة محل الدراسة تعمل على أنظمة سطح المكتب، فنظام Windows يستخدم بنسبة ٣٢.٣% بواقع (١٨٥) برنامج، بينما مثلت نسبة نظام Mac OS X ٢٣.٧% بواقع (١٣٦) برنامج، يليها نظام Linux يستخدم بنسبة ٢٣.٦% بواقع (١٣٥) برنامج، يليها نظام Unix يستخدم بنسبة ٤. ٢%بواقع (٢٤) برنامج، ثم نظام Debian يستخدم بنسبة ٠.٣% بواقع (٢) برنامج، ومثل نظام Raspbian أقل نسبة في أنظمة سطح المكتب حيث بلغ ٠.٢% بواقع (برنامجا واحدا).

٢-وأن ٧% من إجمالي برمجيات البيانات الضخمة محل الدراسة تعمل على أنظمة الهواتف الذكية (نظام Android بنسبة ٣.٧% بواقع (٢١) برنامج – نظام IOS بنسبة ٣.٣% بواقع (١٩) برنامج) في حين بلغت نسبة البرمجيات التي تعمل على الخط المباشر أيضا ٧% بواقع (٤٠) برنامج.

٣- وأن ١.٧% من البرمجيات متعددة المنصات تعمل على كافة أنظمة التشغيل (سطح مكتب – هواتف ذكية – on line) بواقع (١٠) برمجيات.



شكل رقم (١١) توزيع برمجيات البيانات الضخمة محل الدراسة حسب أنظمة التشغيل

❖ نتائج الدراسة

وقد سعت الدراسة إلى تحقيق أهدافها من خلال الإجابة على التساؤلات التي حددتها ضمن أهداف الدراسة وفيما يلي أهم النتائج التي توصلت إليها الدراسة:

١- افتقار كثير من مجتمع المعلومات العربي، والدراسات العربية، بإدراك مفهوم وأهمية البيانات الضخمة.

٢- البيانات الضخمة في أساسها بيانات عادية، ولكنها تتصف بضخامة الحجم نتيجة زيادة البيانات المولدة من المصادر المتعددة، وتتميز بالسرعة نتيجة زيادة معدل تدفق البيانات القادمة بسرعة كبيرة، كما تتصف بالتنوع في الشكل والهيئة، وبالتالي تتطلب تقنيات أكثر تطورا للقيام بعمليات

التخزين والمعالجة، ليتم تحويل تلك البيانات من صورتها الخام إلى قيمة معرفية، يتم الاعتماد عليها فيما بعد، وهذا هو جوهر الاستفادة.

٣- كشفت الدراسة أن الحوسبة السحابية حلاً مناسبًا لمتطلبات تخزين ومعالجة البيانات الضخمة، حيث توفر حلول تخزين قابلة للتطوير وفعالة.

٤- تبين من الدراسة أن الحوسبة السحابية والبيانات الضخمة مكملة لبعضها البعض، فنجد أن الحوسبة السحابية تقنية فعالة توفر حلولا لمشكلة النمو السريع في البيانات الضخمة، في حين أن التخزين التقليدي لا يمكنه تلبية متطلبات التعامل مع البيانات الضخمة، بالإضافة إلى الحاجة إلى تبادل البيانات بين مواقع التخزين الموزعة المختلفة، وهذا ما توفره الحوسبة السحابية.

٥- كشفت الدراسة أن هناك مواضع تربط بين علم البيانات وعلم المعلومات، وهي أن البيانات يمكن أن تصبح موضوعا للبحث في علم المعلومات، وفي علم البيانات على حدٍ سواء.

٦- بينت الدراسة أن علم البيانات يصب كل اهتمامه على **كم البيانات**، في الوقت الذي تشكل **جودة البيانات** أحد الاهتمامات الرئيسية في علم المعلومات.

٧- توصلت الدراسة إلى وجود **١٩٠ برنامجا** من برمجيات البيانات الضخمة، وأن تلك البرمجيات تتنوع وفقا لطبيعة عملها.

٨- أظهرت الدراسة وجود **٦ فئات** لبرمجيات البيانات الضخمة السحابية، لكل فئة خصائصها وسماتها، وهي: (برمجيات التصور المرئي للبيانات - برمجيات معالجة وتحليل البيانات الضخمة - برمجيات قواعد البيانات غير العلائقية NoSQL - برمجيات الذكاء الاصطناعي والتعلم الآلي - برمجيات تخزين البيانات الضخمة - برمجيات معالجة اللغة الطبيعية NLP).

٩- أسفرت الدراسة أن **برمجيات التصور المرئي للبيانات** جاءت في المرتبة الأولى بأعلى نسبة، حيث مثلت **٢٥.٣%** بواقع (٤٨) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة، في حين جاءت برمجيات معالجة اللغة الطبيعية في المرتبة الأخيرة بنسبة **١١.١%** بواقع (٢١) برنامج من إجمالي البرمجيات محل الدراسة.

١٠- أوضحت الدراسة تصدر **الدول الأجنبية** في إنتاج برمجيات البيانات الضخمة، وعدم إسهام أي **دولة عربية** في إنتاج تلك البرمجيات.

١١- كشفت التحليلات أن **الولايات المتحدة الأمريكية** احتلت الصدارة الأولى لأكثر الدول انتاجا لبرمجيات البيانات الضخمة محل الدراسة بأعلى نسبة، حيث مثلت **٧٣%** بواقع (١٤٣) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة.

١٢- تبين من الدراسة أن **اللغة الإنجليزية** تحتل المرتبة الأولى في عدد البرمجيات التي تدعمها بنسبة **٣٧.٤%** ، وبواقع (١٨٦) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة.

١٣- أظهرت الدراسة أن البرمجيات **أحادية اللغة** أعلى نسبة من البرمجيات **متعددة اللغات**، حيث تمثل نسبة **٧٤.٢%** من إجمالي برمجيات البيانات الضخمة محل الدراسة.

١٤- تبين من التحليلات أن أكثر برمجيات البيانات الضخمة **مفتوحة المصدر**، حيث تمثل تلك البرمجيات نسبة **٧٥.٨%** من إجمالي برمجيات البيانات الضخمة محل الدراسة.

١٥- تبين من التحليلات أن أكثر برمجيات البيانات الضخمة **مجانية**، حيث تمثل تلك البرمجيات نسبة **٧٦.٣%** من إجمالي برمجيات البيانات الضخمة محل الدراسة.

١٦- أظهرت الدراسة أن معظم البرمجيات تعتمد على ترخيص **Apache License 2.0** ، بنسبة **٤١.١%**، بواقع (٧٧) برنامج من إجمالي برمجيات البيانات الضخمة محل الدراسة.

١٧- كشفت الدراسة أن معظم مطوري برمجيات البيانات الضخمة محل الدراسة هم من **الشركات التكنولوجية والمؤسسات**، حيث مثل إنتاج الشركات والمؤسسات نسبة **٨٦.٣%**، في حين مثل إنتاج **الأفراد** من المطورين والمبرمجين نسبة **١٣.٧%**.

١٨- أوضحت الدراسة أن أكثر الشركات إنتاجا لبرمجيات البيانات الضخمة محل الدراسة هي **مؤسسة برمجيات أباتشي**، حيث احتلت الصدارة الأولى بأعلى نسبة، حيث مثلت **٤٠.٢%** من إجمالي البرمجيات محل الدراسة.

١٩- كشفت الدراسة أن أكثر لغات البرمجة استخداما في تطوير برمجيات البيانات الضخمة محل الدراسة هي **لغة الجافا**، حيث مثلت أعلى نسبة **٢٣%**، بواقع (٨٤) برنامج من إجمالي البرمجيات محل الدراسة.

٢٠- بينت الدراسة أن معظم برمجيات البيانات الضخمة تعمل على أنظمة سطح المكتب، حيث مثلت أعلى نسبة **٨٤.٣%** من إجمالي البرمجيات محل الدراسة.

❖ المصادر والمراجع

قائمة المصادر العربية

١- سليم، عمر. (٢٠ يناير، ٢٠١٩). البيانات الضخمة. تاريخ الاسترداد ٤ يوليو، ٢٠١٩، من bimarabia: shorturl.at/knxJV/

٢- سيد، أحمد فايز أحمد. (٢٠١٥). أدوات التنقيب عن البيانات مفتوحة المصدر : دراسة تحليلية تقيمية. *مجلة جامعة طيبة للآداب والعلوم الإنسانية*، ٥(١٠). تم الاسترداد من shorturl.at/kpCLS

٣- البسيونى، بدوية محمد. (يوليو، ٢٠١٥). برمجيات الحوسبة السحابية ودورها فى تطوير خدمات المعلومات: دراسة تطبيقية على مواقع التخزين السحابي بالإنترنت. *المجلة الدولية لعلوم المكتبات والمعلومات*، ٢(٣)

٤- القصاص ، علا نبيل. (٢٠١٩). مشروعات البيانات الرقمية الضخمة في المكتبات: دراسة استكشافية. المنوفية، مصر: قسم المكتبات والمعلومات . كلية الآداب . جامعة المنوفية . رسالة ماجستير غير منشورة.

قائمة بالمصادر الأجنبية

1. Hacken, R. (2015). Big Data: Challenges and Opportunities for Digital Libraries. *International Conference on Computing in Engineering and the Sciences.*
2. Hilbert, M. (2015, December 13). Big Data for Development: A Review of Promises and Challenges. *Wiley Online Library, 34*(1). Retrieved from https://doi.org/10.1111/dpr.12142
3. Khan, S., Ara Shakil, K., & Alam, M. (2017). *Big Data Computing Using Cloud-Based Technologies: Challenges and Future Perspectives.* New

Delhi: Department of Computer Science, Jamia Millia Islamia. Retrieved from shorturl.at/epryL

4. Microsoft. (2016, July 4). What is big data? Retrieved December 20, 2018, from Microsoft: https://docs.microsoft.com/en-us/previous-versions/msp-n-p/dn749868(v=pandp.10)

5. Nafi, W. (2018). The Impact of Big Data on Business Intelligence: A Field Study on Jordanian Telecommunication Companies. Master Thesis.Faculty of Business. Middle East University.
6. Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018, October). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences, 30*(4), 431-448. Retrieved from https://www.researchgate.net/publication/317557529_Big_Data_Technologies_A_Survey
7. Rai, A. (2019, January 6). What is Big Data – Characteristics, Types, Benefits & Examples. Retrieved June 10, 2019, from upgrad: https://www.upgrad.com/blog/what-is-big-data-types-characteristics-benefits-and-examples/
8. Rouse, M. (n.d). What's big data? Definition. Retrieved June 10, 2019, from techtarget: https://searchdatamanagement.techtarget.com/definition/big-data
9. Wamba, S. F., Akter, S., Edwards, A. J., Chopin, G., & Gnanzou , D. (2015, July). How ‘Big Data’ Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study. International Journal of Production Economics, 165, 234-246. Retrieved from https://www.sciencedirect.com/science/article/pii/S0925527314004253

Cloud Systems and Software Used to Process Big Data

By

Nourhan Samy Fathallah Morsy

Prof. Dr. Badwia Mohamed Elbasuoni

Professor of Libraries and Information, Faculty of Arts,

Tanta University

Dr. Mona Farouk Shahwan

Lecturer of libraries and information, Faculty of Arts,

Tanta University

Abstract:

This study aims to identify and analyze cloud big data software; 190 programs were reached and divided into six different categories. The researcher relied on the descriptive analytical approach using a checklist to collect data related to those applications and software.
The study has reached a set of results; the most prominent of them are: the data visualization software ranked first with the highest percentage, representing 25.3% (48) out of the total big data software under study, while natural language processing software came in the last rank with 11.1% or (21) programs out of the total the software under study; foreign countries lead the production of big data software, while the Arab countries have not contributed to the production of such software; and most of the big data software is open source and represents 75.8% of the total big data software under study.

Keywords: Big data - Big data software - Big data applications - Big data processing and analysis tools**.**