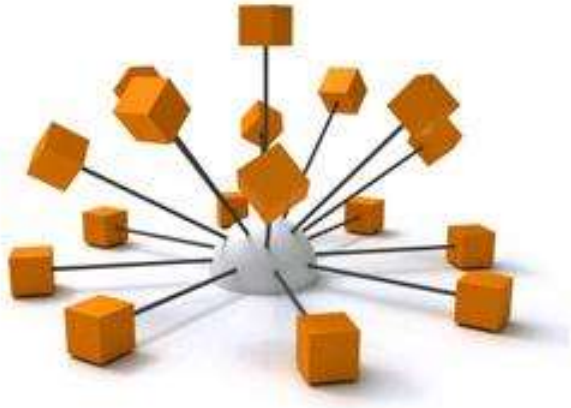
محاضرة : العلوم الاجتماعية في مواجهة البيانات الضخمة

- **للدكتورة: أمينة بن زرارة**

تخصص علوم الإعلام و الاتصال
مخبر الفلسفة و الدراسات الإنسانية والاجتماعية و مشكلات الإعلام و الاتصال
فرقة البحث: تكنولوجيا الاتصال و قضايا المجتمع



BIG DATA

تعريف البيانات الضخمة

تختلف وتتنوع التعريفات والمفاهيم لهذا المجال مابين الخبراء والشركات والمنظمات المتخصصة، حيث يعرف **معهد ماكنزي العالمي** البيانات الضخمة، أنها مجموعه من البيانات التي يفوق حجمها القدرة على معالجتها باستخدام أدوات قواعد البيانات التقليدية، من التقاط ومشاركة ونقل وتخزين وإدارة وتحليل، في غضون فترة زمنية مقبولة [1].

تعرف **شركة جارتنر المتخصصة** في أبحاث واستشارات تقنية المعلومات البيانات الضخمة على أنها "الأصول المعلوماتية كبيرة الحجم وسريعة التدفق وكثيرة التنوع التي تتطلب طرق معالجة مجدية اقتصادياً ومبتكرة من أجل تطوير البصائر والمساعدة على اتخاذ القرارات" [1].

مركز الاحصاء، "مفاهيم عامة حول البيانات الكبيرة"، أدلة المنهجية والجودة، دليل رقم (13)، أبوظبي، صفحة (4).

Laney, D. et al. (2013). "Big Data Means Big Business", Gartner Inc.

تقنية البيانات الضخمة
Big Data Technology

في الدقيقة الواحدة

- يقوم 444,694 مستخدم بالتقديم على وظائف في اللينكدان (LinkedIn).
- يتم إرسال 666,667,41 رسالة على الواتساب.
- 319 يضاف مستخدم جديد على تويتر.
- عقد ما يقارب من 333,208 اجتماع على منصة زوم.
- تحميل 500 ساعة من الفيديوهات على اليوتيوب.
- إرسال 150,000 مشاركة وتحميل 147,000 صورة على الفيسبوك.

مع وجود نحو 4.5 بليون مستخدم على الشبكة العالمية للمعلومات، تشير التقديرات إلى أنه سيتم إنشاء 463 "إكسابايت"(Exabyte) من البيانات يومياً على مستوى العالم - أي ما يعادل 212،765،957 قرص (DVD) .

المصدر : المنتدى الاقتصادي العالمي – 2020

علماً وأنه يمكن متابعة إحصاءات الشبكة العالمية للمعلومات آنياً من خلال الموقع: www.internetlivestats.com

تقنية البيانات الضخمة
Big Data Technology

نشأة البيانات الضخمة

نتيجة للتحديات والصعوبات التي ذكرناها، وجد المختصون والمهتمون بتلك العلوم أنفسهم أمام خيارين، إما إهمال الحجم الهائل من البيانات أو التغلب على تلك التحديات والصعوبات وايجاد حل لها. حيث كان الحل في الأنظمة الموزعة، نتيجة لقدرتها على معالجة البيانات الضخمة على أكثر من جهاز في نفس الوقت. بما ساعد على ظهور تقنية البيانات الضخمة، التي تمثل مصطلح جديد يصف الكم الهائل من البيانات والمعلومات من المصادر المختلفة، التي لا يمكن تخزينها ومعالجتها بالطرق التقليدية وهي واحدة من أهم التقنيات المستقبلية.

تقنية البيانات الضخمة
Big Data Technology

حسب شركة (IBM) تنشأ البيانات الضخمة عن طريق كل شيء من حولنا. في كل الأوقات، كل عملية رقمية وكل تبادل في وسائل التواصل الاجتماعي ينتج البيانات الضخمة، التي تتناقلها الأنظمة وأجهزة الاستشعار والأجهزة النقالة. البيانات الضخمة لها مصادر متعددة تختلف من حيث السرعة والحجم والتنوع. لكي نستفيد من البيانات الضخمة، نحتاج إلى معالجة مثالية وقدرات تحليلية ومهارات [10] .

حسب الإتحاد الدولي للاتصالات (ITU) يشير مصطلح البيانات الضخمة إلى مجموعات البيانات التي تتميز أنها فائقة حجماً وسرعة أو تنوعاً، بالقياس إلى أنواع مجموعات البيانات معهودة الاستخدام [11] .

الخلاصة: تشكل البيانات الضخمة مجالاً هندسياً يوفر مجموعة من الأدوات والتقنيات التي تُستخدم في تخزين ومعالجة البيانات التي تفوق في حجمها قدرات وإمكانيات الطرق والأدوات التقليدية في وقت قياسي.

مصادر البيانات الضخمة

- **المصادر الناشئة عن إدارة البرامج:** سواء أكان برنامج حكومي أو غير حكومي، كالسجلات الطبية الإلكترونية وزيارات المستشفيات وسجلات التأمين والسجلات المصرفية وبنوك الطعام وغيرها.
- **المصادر التجارية أو ذات الصلة بالمعاملات:** تعد مصدرا آخر كالبيانات الناشئة عن معاملات بين كيانين، على سبيل المثال معاملات البطاقات الإئتمانية والمعاملات التي تجرى عن طريق الإنترنت بوسائل منها الأجهزة المحمولة.
- **مصادر معتمدة على شبكات أجهزة الإستشعار وأجهزة التتبع:** فعلى سبيل المثال، التصوير بالأقمار الصناعية، وأجهزة استشعار الطرق، وأجهزة استشعار المناخ وتتبع البيانات المستمدة من الهواتف المحمولة والنظام العالمي لتحديد المواقع وغيرها يمكن أن تكون أحد مصادر البيانات الضخمة.
- **المصادر المتعلقة بسلوك المستخدم:** مثل مرات البحث على الإنترنت عن منتج أو خدمة أوالمعلومات، ومرات مشاهدة إحدى الصفحات على الإنترنت.
- **مصادرالبيانات المتعلقة بالآراء:** مثل التعليقات على وسائط التواصل الإجتماعي مثل فيسبوك وتويتر وغيرها.

تقنيات البيانات الضخمة

وجد العديد من الأدوات والتقنيات التي تستخدم لتحليل البيانات الكبيرة مثل: Hadoop، Map Reduce، HPCC إلا أن Hadoop يُعد من أشهر هذه الأدوات،

وHadoopهو برنامج أو منصة برمجية مفتوحة المصدر مكتوبة بلغة الجافا تستخدم لتخزين ومُعالجة البيانات الضخمة بشكل موزع أي أن تخزين هذه البيانات الضخمة تكون على عدة أجهزة ومن ثم توزع عملية المعالجة على هذه الأجهزة لتسريع نتيجة المعالجة وتعود أو تستدعى كحزمة واحدة. تتكون الأدوات التي تتعامل مع البيانات الضخمة من ثلاثة أجزاء رئيسية وهي: أدوات التنقيب عن البيانات Data Miningوأدوات التحليل Data Analysisوأخيراً أدوات عرض النتائج.Dashboard

ما معنى مفتوح المصدر Open source؟

مفتوح المصدر أو Open sourceهي صفة أو خاصية يتم اطلاقها على برنامج، أو لغة برمجة، أو نظام تشغيل، يوفر مالكها أو مخترعها الأكواد الأساسية والمصدر الأصلي و الخوارزميات الخاصة بها، وذلك من أجل إعادة البرمجة والتعديل، وإعادة التوزيع والمشاركة وفقاً لمتطلبات كل مستخدم.

خصائص البيانات الضخمة

هل كل البيانات ذات أحجام كبيرة هي بيانات ضخمة؟

لقد تم وضع مجموعة من الخصائص والمحددات لكي لتصنيف البيانات على أنها بيانات ضخمة ومتعارف عليها باسم (Vs) بدأت بثلاث خصائص وصولاً إلى 10 خصائص نظراً لأنها تبدأ بحرف ال V، وهي:

- **الحجم (Volume) :** هو العامل الرئيس في وصف البيانات أنها ضخمة بحيث يتعدى حجمها واحد تيرا بايت.
- **التنوع (Variety) :** تنوع البيانات ما بين مهيكلة وغير مهيكلة.
- **السرعة (Velocity)** إنتاج معدلات مرتفعة من البيانات في كل لحظة.
- **الدقة / الموثوقية (Veracity)** : أن تكون البيانات موثوقة وصحيحة.
- **القيمة (Value)** : القدرة على تحويل جميع أنواع البيانات إلى بيانات ذات فائدة .

تقنية البيانات الضخمة
Big Data Technology

- **التمثيل البصري (Visualization) :** إمكانية إظهار وعرض البيانات بشكلٍ أفضل بحيث يمكن للشخص الذي يقرأها أن يفهمها بشكل سريع.
- **التباين / التغير (Variability) :** عدد الاختلافات في البيانات نتيجة للتغير سواءً في بنية البيانات أو المعنى أو الشكل.
- **الثغرة الأمنية (Vulnerability) :** الحفاظ على أمن وخصوصية البيانات.
- **الجودة / المصداقية (Validity) :** أن تكون مصادر البيانات دقيقة وأن تكون البيانات موثوقة للاستخدام المقصود.
- **التقلب (Volatility) :** مدة صلاحية البيانات وطول مدة تخزينها.

خامساً : تطبيقات ونماذج من الواقع لاستخدام البيانات الضخمة

هناك العديد من الأمثلة والتطبيقات التي نجحت في الاستفادة من تقنية البيانات الضخمة في عدة مجالات من أهمها:

التعليم

يعتبر تطبيق دايلنجو لتعلم اللغات واحد من التطبيقات والأنظمة التي استطاعت الاستفادة من البيانات الكبيرة عن طريق تخزين البيانات للمستخدمين وتفاعلاتهم وتحليلها باستمرار لمعرفة نجاحات المستخدمين وإخفاقاتهم وأنشطتهم، والعمل على التحسين المستمر لذلك، يعتبر اليوم واحداً من التطبيقات المعروفة في تعليم اللغات.

المبيعات

استطاعت شركة أمازون معالجة ملايين العمليات كل يوم والاستفاده من تحليل السلوك الاستهلاكي لعملائها وتحديد احتياجتهم واقتراح المنتجات والبضائع عليهم مما ساهم في زيادة المبيعات والأرباح.

الترفيه

استطاعت منصة **نتفليكس** من استغلال وتحليل بيانات المستخدمين في تحديد المواد المناسبة واقتراحها عليهم حسب الوقت المناسب لهم.

النقل والمواصلات

استفادت شركة **أوبر** من تحليل بيانات الركّاب في تقديم خدمات وعروض تناسب احتياجتهم وقدراتهم المالية.

الصحة

نجحت شركة **مايكروسوفت** بالتعاون مع أحد المعاهد الهندية في حصولها على كم بيانات ضخم من سجلات، لأكثر من مليون شخص، في مساعدة الأطباء في التعرف بشكل دقيق على أسباب انتشار أمراض العيون، إضافة إلى مساعدة الأطباء على تحديد الوقت المطلوب للمريض حتى تستقر حالته الصحية ويعالج بصره، وكذلك القيام بالعمليات الجراحية للعيون بنجاح.

الخدمات المالية

نجحت شركة أمريكان اكسبريس في استغلال قاعدة بيانات تحتوي على كمية هائلة من البيانات لمالكي البطاقة الإئتمانية من مختلف أنحاء العالم وفهم سلوك المستهلك بشكل أعمق والتنبؤ بالمعاملات المستقبلية.

الصناعة

تعتمد شركة جنرال الكتريك على ابتكار منتجات وطرح حلول رقمية ترتبط باحتياجات ومتطلبات المستهلكين.

الدعاية والتسويق

يعتبر الفيسبوك واحداً من أكبر مصادر البيانات الضخمة وقد نجح باستغلالها والاستفاده منها في تصميم خوارزميات تعرض للمستخدمين إعلانات لبضائع وخدمات تناسب احتياجاتهم واهتماماتهم.

تطبيقات من واقع استخدام البيانات الضخمة

Microsoft
الاستفاده من البيانات الضخمة لتقديم حلول لاستفسارات العملاء وتقديم خدمات بشكل أفضل

Duolingo
الاستفاده من بيانات المستخدمين وتفاعلهم في تحسين نمط التعليم

AMERICAN EXPRESS
التنبؤ وتحديد العملاء الواعدين والطريقة التي يمكن بها للشركات الاحتفاظ بالعملاء القدامى

Amazon
الاستفاده من بيانات العملاء في تحديد أنماط السلوك الاستهلاكي ومعرفة احتياجتهم

General Electric
ابتكار منتجات ترتبط باحتياجات ومتطلبات المستهلكين

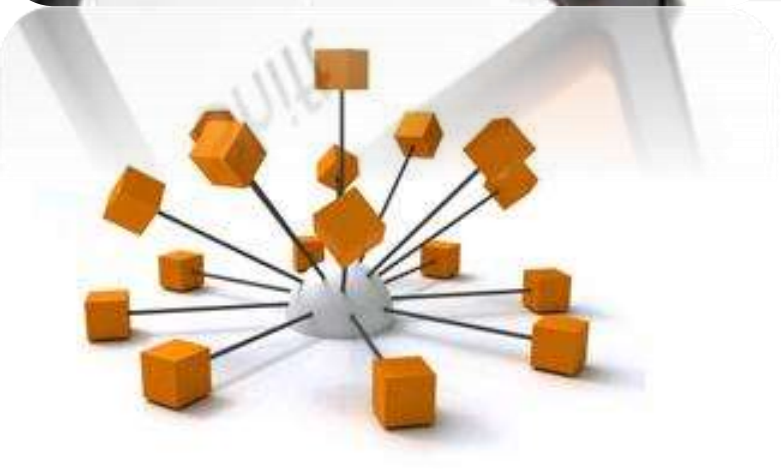
Netflix
تحليل تفضيلات المشاهد والوقت الذي يقضيه في البوابة لتحليل أنماطها واقتراح المواد المناسبه لهم

Facebook
اقتراح اعلانات لخدمات ومنتجات التى تناسبهم

UBER
تحليل بيانات الركاب وتقديم خدمات وعروض تناسب قدراتهم واحتياجتهم

مازال مجال البيانات الضخمة يواجه مجموعة من التحديات والصعوبات التي يجب العمل عليها وإيجاد حلول مناسبة لها للاستفادة منها بشكل بأفضل

ربما من أهم هذه التحديات، المخاطر المرتبطة بالمستخدم البسيط خصوصاً فيما يتعلق بحريته الشخصية، ما يدعو بضرورة الحفاظ على البيانات والمعلومات الشخصية قدر الإمكان، بحيث لا تكون سلعة تستغلها الشركات والتي أصبحت تعرف عن الأشخاص أكثر مما يعرفون عن أنفسهم.



تخزين البيانات الضخمة

1- الكم الهائل من المعلومات لا يتم تخزينه في مكان واحد .فمثلا معلومات طلبة واساتذة جامعة قالمة مخزنة في خادم جامعة قالمة، معلومات زبائن اتصالات الجزائر مخزنة في خوادم الشركة، الصور والفيديوهات المنشورة على منصة فيسبوك مثلا مخزنة في خوادم الشركة في أمريكا...الخ

2- بالنسبة للشركات الكبرى مثل فيسبوك ويوتيوب التي تشارك في الثانية الواحدة كم هائل جدا من البيانات، يتم حفظ البيانات في ما يطلق عليه " مزارع الخوادم" او " مراكز البيانات " data centers هذه الاخيرة هي مدن واسعة جدا من الاجهزة المرتبطة ببعضها البعض، والتي يتم فيها تخزين البيانات

3- سياسة الحماية التي تتبعها شركات حفظ البيانات تعتمد على توزيع عدة نسخ للبيانات عبر مراكز متعددة، والتي في حالات كثيرة يتم بنائها في اماكن سرية تحت الأرض حتى تتمتع بحماية أكبر خاصة من العمليات الهجومية في حالة حدوث الحروب. ولكن هذه الطريقة لها مخاطر أيضا خاصة عند تعرض المدن للفياضانات او الزلازل.

MERCI POUR VOTRE ATTENTION
ENCORE DES
QUESTIONS?
KeepCalmAndPosters.com