تشير البيانات الضخمة إلى مجموعات البيانات الهائلة والمعقدة التي لا تستطيع أنظمة إدارة البيانات التقليدية التعامل معها. وعند جمعها وإدارتها وتحليلها بشكل صحيح، يمكن للبيانات الضخمة أن تساعد المؤسسات على اكتشاف رؤى جديدة واتخاذ قرارات تجارية أفضل.

البيانات - المعلومات -- المعرفة

بداية وقبل التطرق للموضوع الأساسي وهو البيانات الضخمة يجب أن نفهم

أولاً: **ما هي البيانات؟ وما الفرق بينها وبين المعلومات والمعرفة؟**

البيانات، يمكن تعريفها بأنها مجموعة من الحروف، أو الكلمات، أو الأرقام، أو الرموز، أو الصور المتعلقة بموضوع ما. والبيانات في حد ذاتها ليس لها معنى أو قيمة، وهي الصورة الخام للمعلومة [1]، مثال بيانات الموظفين وصورهم.

أما **المعلومات،** فهي البيانات التي تم تحليلها ومعالجتها، بحيث أصبحت تحمل معنى وقيمة، ويمكن الاستفادة منها في اتخاذ القرارات، مثل الحصول على إجمالي عدد الموظفين ومعدل أعمارهم ومعدل سنوات الخبرة، ... الخ.

و**المعرفة،** هي عملية تحليل المعلومات المختلفة وربطُها ببعضها البعض ووجود فَهم واضح لها ومزجها مع الخبرة، وهو ما يوضحه الشكل التالي.

شكل رقم (1)
الفرق بين البيانات والمعلومات والمعرفة

خلاصة القول: البيانات تعني المواد الخام، والمعلومات هي بيانات منظمة، والمعرفة هي معلومات تم إدراكها.

كيف يمكن قياس حجم البيانات ؟

يمكن قياس حجم البيانات باستخدام مجموعة من وحدات القياس المختلفة تبدأ بأقل شيء وهي البت (Bit) والبايت (Byte) والكيلو بايت (Kilobyte)، وصولاً لأعلى شيء متعارف عليه وهو دوميجمقروبايت(Domegemegrottebyte)[2].

البت (Bit): هي أصغر وحدة تخزين ممكنة، تمثل مرور أو عدم مرور تيار كهربائي، وقيمة البت تكون 1 أو 0 حيث يمثل 1 مرور تيار كهربائي، و 0 عدم مروره [3].

وحدة قياس سعة التخزين وهي **البايت** (Byte) ومضاعفاتها مبينة، كما هو موضح في الجدول التالي.

جدول رقم (1)
أنواع وحدات قياس حجم البيانات

وحدة القياس	الاختصار	الحجم
البت Bit	b	0 أو 1
البايت Byte	B	8 بت
كيلوبايت Kilobyte	kB	1024 بايت
ميجابايت Megabyte	MB	1024 كيلو بايت
جيجابايت Gigabyte	GB	1024 ميجابايت
تيرابايت Terabyte	TB	1024 جيجابايت
بيتابايت Petabyte	PB	1024 تيرابايت
إكسابايت Exabyte	EB	1024 بيتابايت
زيتابايت Zettabyte	ZB	1024 اكسابايت
يوتابايت Yottabyte	YB	1024 زيتابايت
زينوتابايت Xenottabyte	XB	1024 يوتابايت
شايلينوبايت Shilentnobyte	SB	1024 زينوتابايت
دوميجمقروبايت Domegemegrottebyte	DB	1024 شايلينوبايت

أساليب تخزين البيانات

تختلف وتتنوع أساليب تخزين البيانات ما بين:

1. **بيانات هيكلية**، يتم تخزينها بطريقة منظمة، مثل البيانات الموجودة في قواعد البيانات وتتميز بسهولة المعالجة والتحليل.
2. **بيانات غير هيكلية**، تمثل النسبة الأكبر من البيانات، وتكون موجودة ومخزنة بطريقة غير منظمة مثل البيانات الموجودة على الشبكة العالمية للمعلومات (الإنترنت) خاصة منها مواقع التواصل الاجتماعي وتطبيقات الهواتف الذكية ولا يمكن معالجتها بسهولة.
3. **بيانات شبة هيكلية**، تخزن بهيئة وترتيب معين مُختلف عن قواعد البيانات [4].

شكل رقم (2)
أساليب تخزين البيانات

بيانات هيكلية STRUCTURED DATA	البيانات غير هيكلية UNSTRUCTURED DATA	بيانات شبه هيكلية SEMI-STRUCTURED DATA
بيانات تم تنظيمها في جدول أو قاعدة بيانات. مثل البيانات الخاصة بالموظفين أو عملاء البنك أو معاملات المؤسسة المختلفة.	تمثل النسبة الأكبر من البيانات، مثل البيانات التي تنتجها المواقع الإلكترونية يوميا من نصوص وصور وفيديوهات وخاصة مواقع التواصل الاجتماعي ، وكذلك تطبيقات الأجهزة الذكية.	تعد على أنها نوع من البيانات المهيكلة ولكنها ليست مبينة في جداول أو قواعد بيانات، ولكنها تخزن في ملفات بشكل منظم مثل : JSON,HTML and XML

البيانات التقليدية والبيانات الضخمة ما الفرق؟

يختلف مفهوم البيانات التقليدية عن مفهوم البيانات الضخمة بشكل أساسي في أنواع البيانات المستخدمة، وكمية البيانات التي يتم التعامل معها، والأدوات اللازمة لتحليلها. تتكون البيانات التقليدية في المقام الأول من بيانات منظمة مخزنة في قواعد بيانات علائقية. تُنظم هذه القواعد البيانات في جداول محددة بوضوح، مما يُسهل الاستعلام عنها باستخدام أدوات قياسية مثل لغة SQL. عادةً ما تتضمن تحليلات البيانات التقليدية أساليب إحصائية، وهي مناسبة لمجموعات البيانات ذات التنسيقات المتوقعة والأحجام الصغيرة نسبيًا....

مصادر البيانات

بالنسبة لمصادر البيانات، فتنقسم إلى ثلاثة مصادر رئيسة [5]:

1-المصدر الأول: البيانات التي تنتجها المؤسسات

البيانات التي يتم إنتاجها من قبل الشركات والمؤسسات خلال عَملها بشكل يومي، ويتم تخزينها في قواعد بيانات أو ملفات تكون خاصة بالشركة أو المؤسسة.

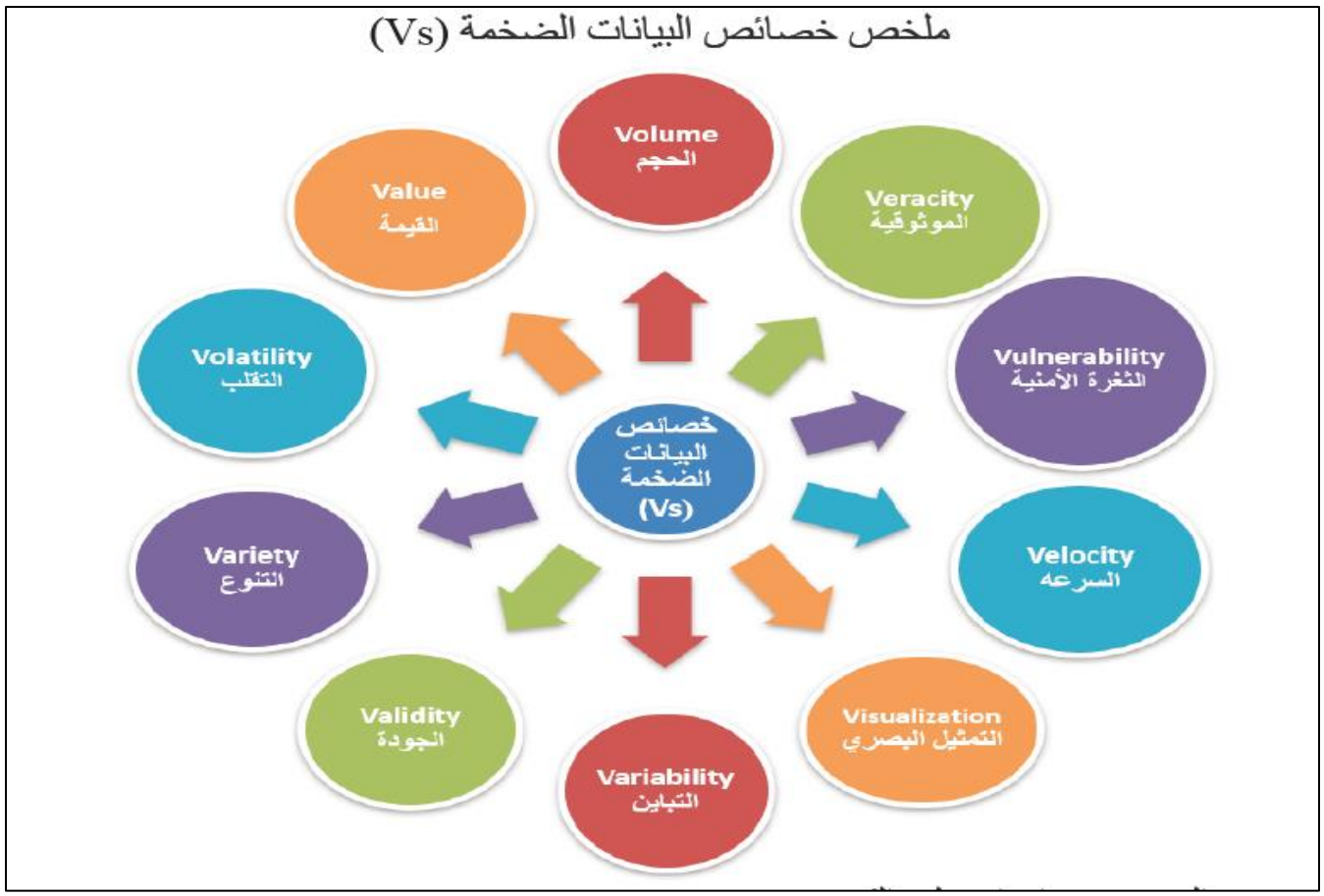
2-المصدر الثاني: البيانات التي ينتجها الإنسان

البيانات الناتجة عن الإنسان سواءً منها المستندات الموجودة على أجهزة الحاسب الآلي أو تلك المخزنة في الأجهزة المحمول المختلفة كالرسائل النصية وتطبيقات الأجهزة الذكية، إضافة إلى تلك البيانات التي يدرجها الإنسان في الشبكة العالمية للمعلومات مثل البريد الإلكتروني ومواقع التواصل الاجتماعي، من مشاركة للصور والمحادثات والتعليقات وتكون أغلبها بيانات غير منظمة.

3-المصدر الثالث: البيانات التي تنتجها الآلات

هي جميع البيانات التي يتم إنتاجها تلقائياً من دون تدخل الإنسان، مثل بيانات كاميرات المراقبة وأجهزة الاستشعار والأقمار الاصطناعية.

خصائص البيانات الضخمة: تُعرف "خصائص البيانات الضخمة" بأنها: الحجم، والسرعة، والتنوع، والمصداقية، والقيمة، وهي خمس سمات تميز البيانات الضخمة عن أنواع البيانات الأخرى. توضح هذه السمات كيف تختلف البيانات الضخمة عن مجموعات البيانات التقليدية، وما يلزم لإدارتها بفعالية. الحجم: تُوصف البيانات الضخمة بأنها "ضخمة" نظرًا لكثرتها. إن الكم الهائل من البيانات المُولّدة اليوم - من تطبيقات الويب، وأجهزة إنترنت الأشياء، وسجلات المعاملات، وغيرها - قد يُشكّل تحديًا لأي



ما هى اهم تلك الخصائص؟

1- **الحجم** : البيانات الضخمة تسمى "ضخمة " لانها اصبحت اكثر وفرة فالكم الهائل من البيانات المولدة من تطبيقات الويب واجهزة انترنت الاشياء وسجلات المعاملات وغيرها تشكل صعوبة لاي مؤسسة في ادارتها وغالبا ما تعاني انظمة تخزين البيانات ومعالجاتها التقليدية من صعوبة التعامل معها على نطاق واسع.

2- **القيمة**: تشير القيمة الى الفوائد الملموسة التي يمكن للمؤسسات جنيها من البيانات الضخمة وتشمل هذه الفوائد :

- تحسين العمليات التجارية.
- تحديد فرص تسويقية جديدة.

حيث يعد تحليل البيانات الضخمة عنصرا اساسيا في هذا الامر حيث يعتمد غالبا على التحليلات المتقدمة والتعلم الالي والذكاء الاصطناعي لتحويل المعلومات الخام الى رؤى قابلة للتنفيذ.

3- المصداقية :

تشير المصداقية إلى دقة البيانات وموثوقيتها. ونظرًا لأن البيانات الضخمة تأتي بكميات هائلة ومن مصادر متنوعة، فقد تحتوي على تشويش أو أخطاء، مما قد يؤدي إلى اتخاذ قرارات خاطئة.

4- السرعة: هي مدى سرعة تدفق البيانات الى النظام وتتحرك البيانات الضخمة بسرعة فائقة اليوم،تصل البيانات اسرع من اي وقت مضى بدءا من تحديثات وسائل التواصل الاجتماعي في الوقت الفعلي وصولا الى سجلات تداول الاسهم عالية التردد يوفر التفق السريع للبيانات الية تدعم اتخاذ القرارات.

5- التنوع :يشير المصطلح هذا الى السيغ المختلفة التي يمكن ان تتخذها البيانات الضخمة فالى جانب البيانات المهيكلة التقليدية تشمل البيانات الضخمة بيانات غير مهيكلة مثل : النصوص- الصور مقاطع الفيديو كما تشمل بيانات شبه مهيكلة مثل ملفات XML التي تحتوي على بعض الخصائص التنظيمية

يتطلب ادارة هذا النوع من البيانات قواعد بيانات مثل NoSQL وبحيرات البيانات .

بقية الخصائص يمكن اختصارها بالشكل التالي :

- **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.