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 [8].
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" [9].

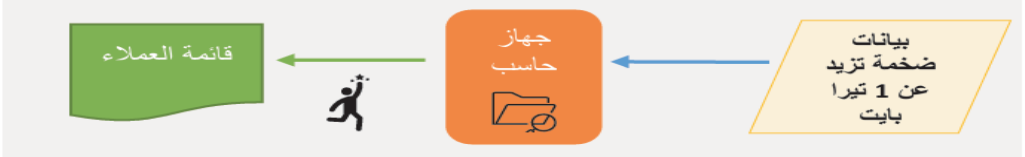
حسب شركة (IBM) تنشأ البيانات الضخمة عن طريق كل شيء من حولنا. في كل الأوقات، كل عملية رقمية وكل تبادل في وسائل التواصل الاجتماعي ينتج البيانات الضخمة، التي تتناقلها الأنظمة وأجهزة الاستشعار والأجهزة النقالة. البيانات الضخمة لها مصادر متعددة تختلف من حيث السرعة والحجم والتنوع. لكي نستفيد من البيانات الضخمة،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

آلية عمل البيانات الضخمة

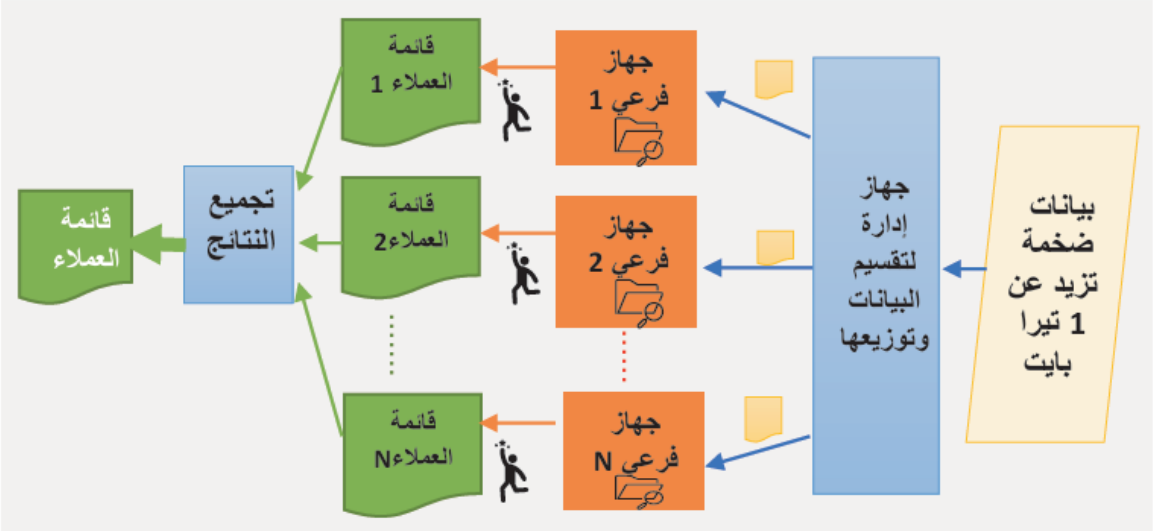
لتبسيط التعريفات والمصطلحات السابقة، دعونا نتعرف على إطار عمل البيانات الضخمة، لنفترض أن شركة لديها بيانات عملاء حجمها يزيد عن واحد تيرا بايت، الذي يمثل الحد الأدنى لحجم البيانات لكي تصنف كبيانات ضخمة. أرادت هذه الشركة أن تحدد قائمة بأسماء العملاء وأرقام هواتفهم ممن يقيمون في منطقة معينة لاستهدافهم في حملة إعلانية. حسب الطريقة التقليدية، هذا الأمر يمكن من خلال الرسم التالي:

شكل رقم (6)
الطريقة التقليدية في معالجة البيانات



معالجة الأمر بالطريقة التقليدية يستغرق وقتاً طويلاً جداً، ربما يصل إلى أيام للحصول على النتائج المطلوبة، أو قد يصعب ذلك. فلنا أن نتخيل الفرق إذا أعطينا مجموعة كتب لشخص واحد للبحث عن الصفحات التي بها كلمة معينة أو أن تستعين بمجموعة أشخاص ونعطي كل منهم كتاب للبحث، فبالطبع الطريقة الثانية هي الأسرع، وهي نفس فكرة عمل البيانات الضخمة. يجسد الرسم التالي آلية عمل تقنية البيانات الضخمة بشكل مبسط. كل مكون ووظيفه بها مجموعة من الوظائف والمكونات والتفاصيل التي تحتاج إلى مختصين لفهمها.

شكل رقم (7)
آلية عمل تقنية البيانات الضخمة في معالجة البيانات



المصدر: من إعداد مؤلف الكتيب.

خصائص البيانات الضخمة

هل كل البيانات ذات أحجام كبيرة هي بيانات ضخمة؟

لقد تم وضع مجموعة من الخصائص والمحددات 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 .

- **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.

التطبيقات الرئيسية للبيانات الضخمة في الفينتيك

تُستخدم البيانات الضخمة في مجالات متعددة، مما يحول القطاع المالي من نموذج تقليدي إلى نموذج ذكي. إليكم أبرز التطبيقات:

1. **إدارة المخاطر :(Risk Management)** تساعد في تحديد المخاطر المحتملة عبر تحليل أنماط البيانات، مثل التنبؤ بمخاطر الإقراض باستخدام خوارزميات تكتشف البيانات غير الطبيعية. على سبيل المثال، في البنوك، تقلل تكاليف التقييم الائتماني بنسبة تصل إلى 47%، وتعزز الشمول المالي للفئات ذات التاريخ الائتماني الضعيف .
2. **كشف الاحتيال :(Fraud Detection)** تعتمد على التحليل الفوري للمعاملات للكشف عن الأنشطة المشبوهة، مثل مراقبة التحويلات عالية القيمة أو من مواقع غير مألوفة. تقنيات مثل تحليل الرسوم البيانية (Graph Analytics)تكشف عن شبكات الاحتيال المنظم، مما يقلل الخسائر المالية ويعزز الامتثال لقوانين مكافحة غسيل الأموال.(AML)
3. **تخصيص الخدمات :(Personalization)** تُمكن من تقديم خدمات مخصصة بناءً على سلوكيات العملاء، مثل اقتراح استثمارات أو قروض مخصصة. دراسات تشير إلى زيادة الإيرادات بنسبة 5-15% عبر التخصيص.
4. **الشمول المالي :(Financial Inclusion)** تساهم في توسيع الوصول إلى الخدمات المالية للأفراد غير المصرفيين، من خلال نماذج الائتمان البديلة التي تعتمد على بيانات مثل استخدام الهواتف المحمولة أو وسائل التواصل. مثال بارز هو منصة M-Pesa في كينيا، التي استخدمت البيانات لتقديم خدمات الدفع الرقمي للسكان الريفيين.
5. **التداول والاستثمار :(Trading and Investing)** تُدعم التداول الآلي عبر تحليل السوق الفوري، كما في استخدام جولدمان ساكس لمعالجة اللغة الطبيعية (NLP) لتقييم الاستثمارات بناءً على السياق الإعلامي.

لتوضيح التطبيقات، إليكم جدول يلخص الاستخدامات الرئيسية مع أمثلة:

الفائدة الرئيسية	أمثلة عملية	الوصف الرئيسي	التطبيق

تقليل الخسائر وتعزيز الشمول.	نماذج الائتمان البديل في Klarna.	تحليل أنماط البيانات للتنبؤ بالمخاطر في الإقراض والتأمين.	إدارة المخاطر
تقليل الخسائر المالية بنسبة عالية.	استخدام ML في الدفعات عبر SoFi.	مراقبة المعاملات الفورية للكشف عن الأنشطة المشبوهة.	كشف الاحتيال
زيادة الرضا والإيرادات.	روبو-أدفايزر في الاستثمار.	تقسيم العملاء بناءً على السلوكيات لعروض مخصصة.	تخصيص الخدمات
وصول لـ2 مليار شخص غير مصرفي.	M-Pesaللدفعات الرقمية في أفريقيا.	تقييم الائتمان باستخدام بيانات غير تقليدية.	الشمول المالي
تحسين الدقة في الأسواق المتقلبة.	NLPفي جولدمان ساكس.	تحليل السوق الفوري للقرارات الآلية.	التداول

الفوائد الرئيسية

توفر البيانات الضخمة فوائد متعددة، منها تحسين الكفاءة التشغيلية من خلال تقليل التكاليف وأتمتة العمليات، وزيادة رضا العملاء عبر التخصيص – حيث يفضل 80% من العملاء الخدمات المخصصة. كما تعزز الابتكار، مثل في التأمين الرقمي (InsurTech) حيث تُعدل الأسعار ديناميكياً بناءً على بيانات الوقت الفعلي من الأجهزة الذكية. دراسة على بنك تجاري دولي في مصر أظهرت ارتفاع معامل الكفاءة إلى 1.0 بعد تبني البيانات الضخمة.

التحديات والمخاطر

رغم الفوائد، تواجه البيانات الضخمة تحديات مثل خصوصية البيانات، حيث تتطلب الامتثال لقوانين مثلGDPR ، والتحيزات الخوارزمية التي قد تعزز التمييز .كما هناك نقص في المهارات المتخصصة، ومشكلات في جودة البيانات، مما يتطلب استثمارات في التدريب والتنظيم. في السياق العربي، أظهرت دراسات تحديات ثقافية في تبني التكنولوجيا في البنوك.

أمثلة عملية ودراسات حالة

- **في الإقراض** :تستخدم شركات مثل Klarna و Affirmالبيانات الضخمة للإقراض عند نقاط البيع، مما يقلل التكاليف ويوسع الوصول في الدول النامية.
- **في التأمين** :تستفيد Greater Than من بيانات الطرق لتخصيص التأمين السياراتي.
- **في الشمول** :في الصين، ساهمت البيانات الضخمة في توسيع الوصول المالي للأسر المنخفضة الدخل.