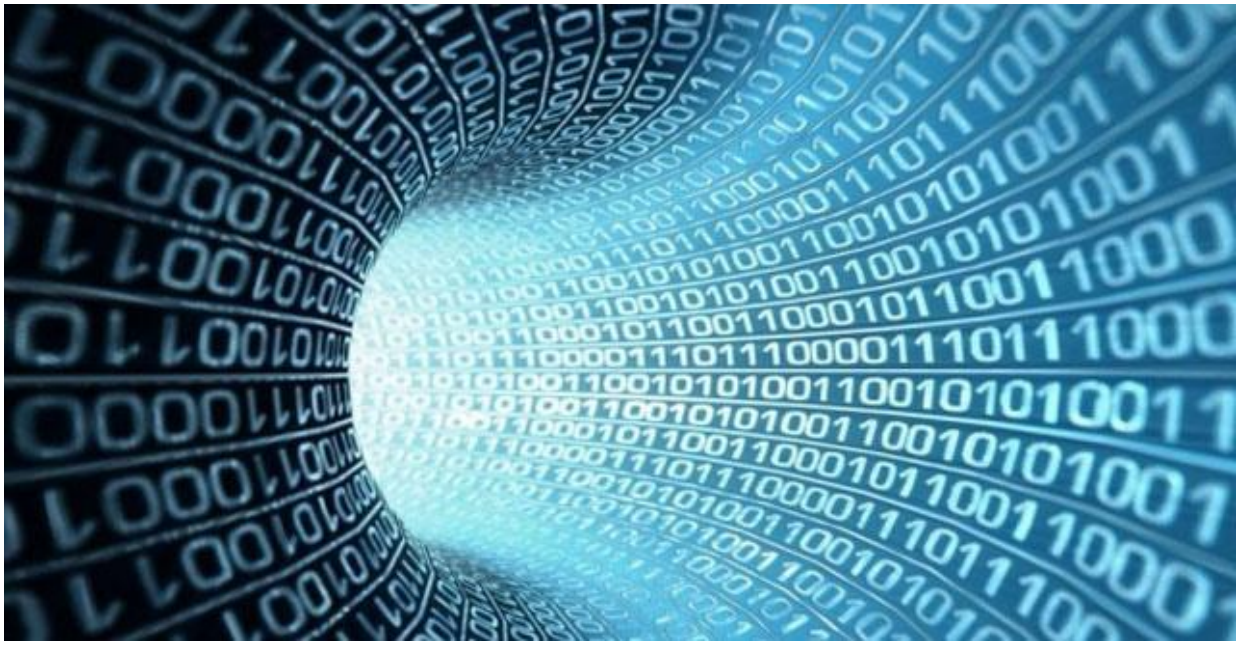
ثورة البيانات الضخمة BIG DATA



اعداد الدكتورة:
بسمة بوكنز

ثورة البيانات

مقدمة :

نسمع كثيرا عن مصطلح البيانات الضخمة Big Dataو سرعة انتشار هذا المجال في سوق العمل.

تعريف:

قبل ان نتطرق لتعريف البيانات الضخمة، يجب علينا معرفة ما هي البيانات؟

البيانات

هي الصورة الخام للمعلومات قبل عمليات الفرز والترتيب والمعالجة ولا يمكن الاستفادة منها بصورتها الأولية قبل المعالجة.

*و يمكن ان **تقسم البيانات الخام الى ثلاثة انواع:***

بيانات مهيكلة: وهى البيانات المنظمة في جداول او قواعد بيانات.

بيانات غير مهيكلة: تشكل النسبة الأكبر من البيانات، وهى البيانات التي يولدها الأشخاص يوميا من كتابات نصية وصور وفيديو ورسائل ونقرات على مواقع الانترنت... الخ.

بيانات شبه مهيكلة: تعتبر نوعا من البيانات المهيكلة الا ان البيانات لا تكون في صورة جداول او قواعد بيانات.

البيانات الضخمة ومزاياها ؟

- معالجة سريعة للبيانات
- تعطي نتائج أكثر دقة
- تولد مؤشرات جديدة أو تكميلية
- تنتج بسرعة فائقة
- تنتج بكلفة زهيده
- تحسين في جودة البيانات
- تحسين في التغطية الزمنية والمكانية

استخدامات البيانات الضخمة

- الإتصالات
- تخطيط النقل
- استراتيجيات السياحة
- الرياضة
- الرعاية الصحية
- تحركات المهاجرين

لا يوجد تعريف واحد متفق عليه للبيانات الضخمة. وفق أحد التعريفات:

هي بيانات تولدت من خلال استخدامنا المتزايد للأجهزة الرقمية، والأدوات والمنصات المدعومة من شبكة الإنترنت في حياتنا اليومية. وفي أي لحظة معينة، يستخدم مئات الملايين من الأفراد في أنحاء العالم بعضًا من الهواتف المحمولة في العالم –البالغ عددها من سبعة مليارات إلى ثمانية مليارات هاتف– لإجراء مكالمات، أو لإرسال رسائل نصية أو بريد إلكتروني.

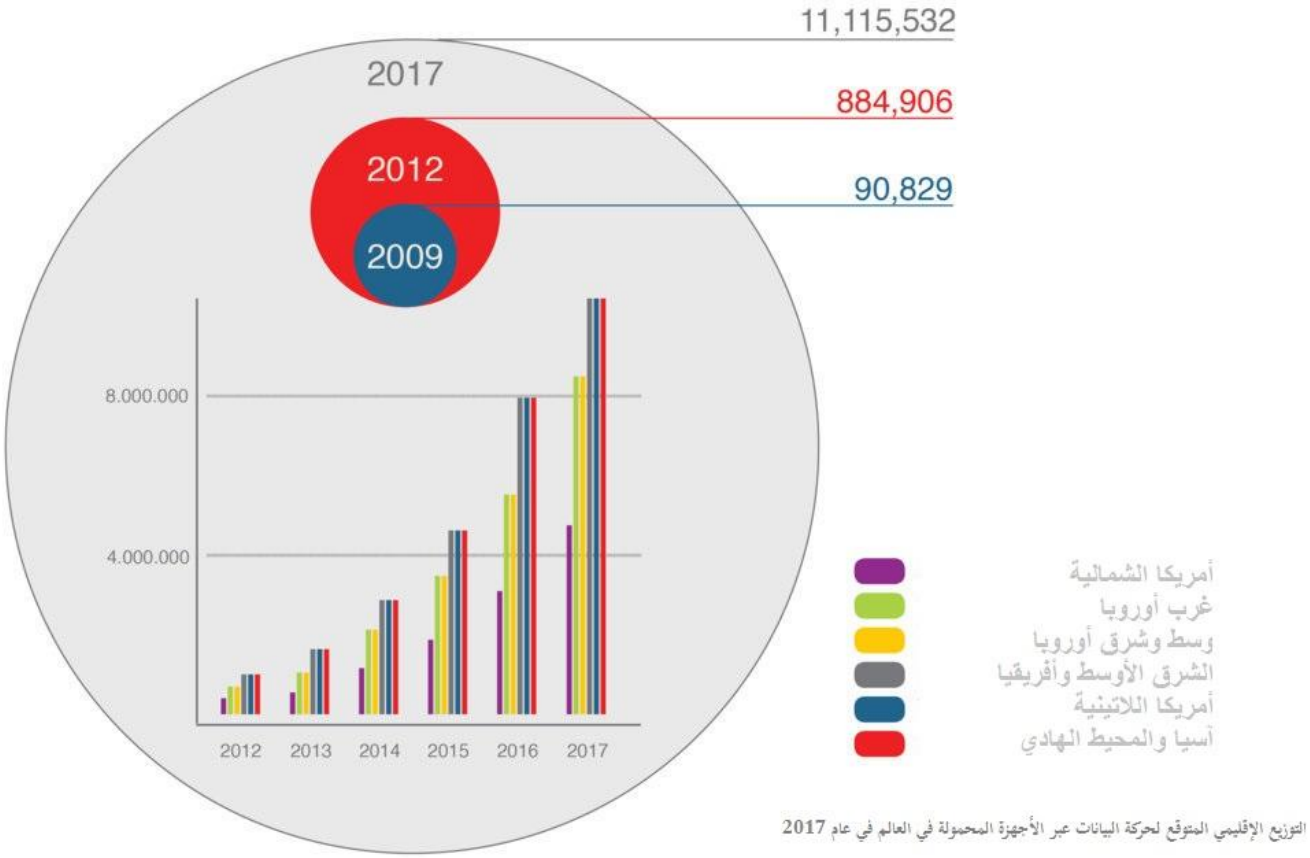
وقد يحولون أموالاً، أو يشترون كتبًا، أو يبحثون في الإنترنت، أو يدفعون ثمن وجبة ببطاقة ائتمان، أو يحدِّثون الحالة على صفحاتهم في 'فيسبوك'، أو يرسلون تغريدات، أو يرفعون ملفات فيديو على موقع 'يوتيوب'، أو ينشرون تدوينات.

كل هذه الأنشطة يترك أثرًا رقميًّا. وتُشكِّل هذه المعلومات الرقمية في مجموعها الجزء الأكبر من البيانات الضخمة. ومنذ عام 2012، يجري إنتاج أكثر من 1.2 زيتابايت من البيانات سنويًّا؛ 2110 بايت، أو ما يكفي لملء 80 مليار جهاز 'آي فون' سعة 16 جيجابايت (التي قد تُشكِّل على امتدادها أكثر من 100 دورة حول الأرض). (الجدول 1) بل إن حجم هذه البيانات في نمو سريع.[11] لذلك فإن الحجم، والسرعة، والتنوع هي 'الخصائص' الثلاث التي تميز البيانات الضخمة، وغالبًا ما يُضاف إليها القيمة التي يمكن استخراجها منها باعتبارها الخصيصة الرابعة.

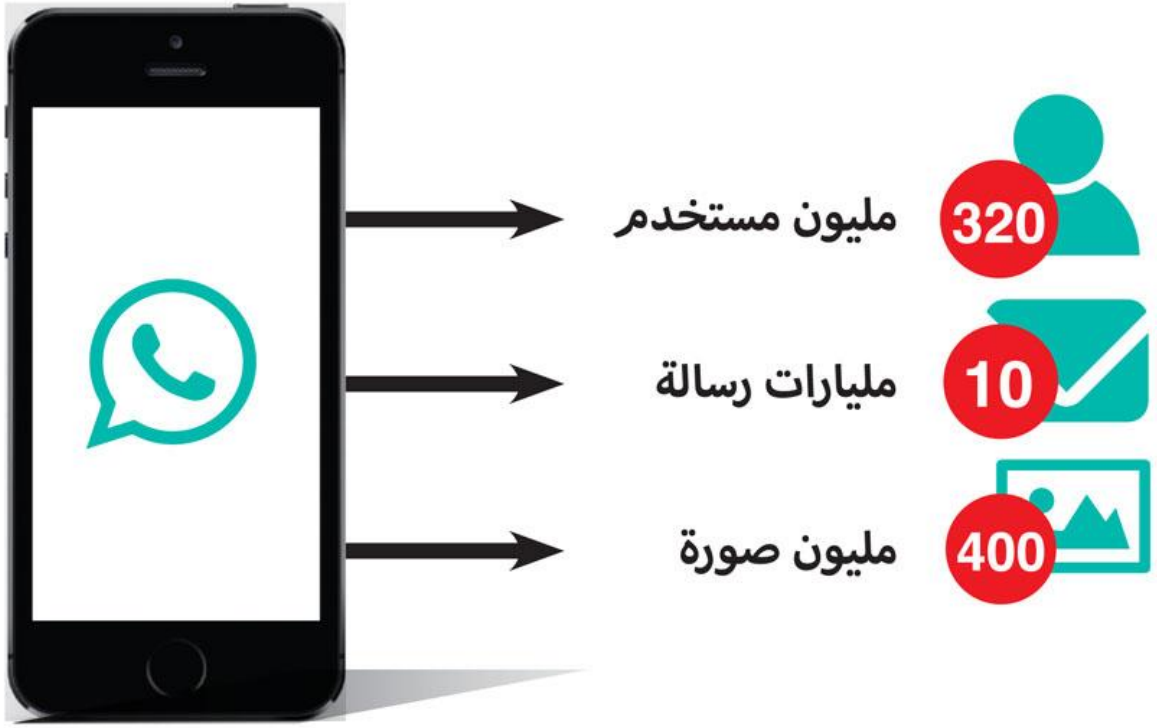
أنواع البيانات

تأتي البيانات الضخمة في أنواع شتى. أحد هذه الأنواع عبارة عن أجزاء صغيرة من بيانات 'صماء' –مثل الأرقام أو الحقائق– التي وصفها أليكس 'ساندي' بنتلاند –وهو أستاذ في معهد ماساتشوستس للتكنولوجيا، الولايات المتحدة الأمريكية– بأنها "فتات الخبز الرقمية".[13] ويقال إن هذه البيانات 'بنيوية' لأنها تشكل مجموعات من المتغيرات التي يمكن توصيفها، وتصنيفها، وتنظيمها بسهولة (في أعمدة وصفوف على سبيل المثال) بغرض التحليل المنهجي. ومن الأمثلة على ذلك سجلات تفاصيل المكالمات الهاتفية التي تجمعها شركات المحمول (الجدول 2). وسجلات تفاصيل المكالمات الهاتفية تُعدّ واصفات للبيانات (بيانات عن البيانات) تسجل استخدام المشتركين لهواتفهم المحمولة –بما في ذلك رمز تحقيق الهوية، وموقع برج الهاتف الذي حدد مسار المكالمة لكل من المتصل والمتلقي، كحد أدنى– ووقت المكالمة ومدتها. وتحفظ شركات الاتصالات الكبرى سجلات تفاصيل المكالمات الهاتفية يوميًّا.

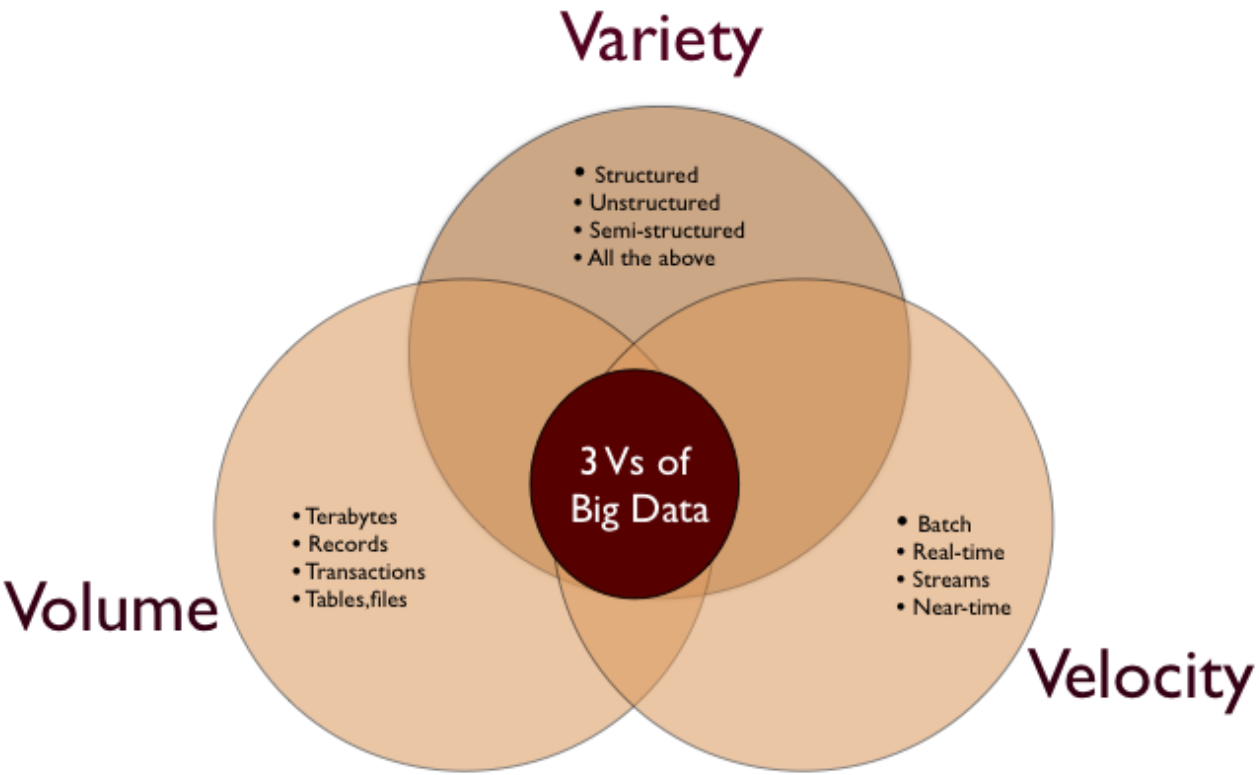
بيانات الأجهزة المحمولة في العالم – نمو الحركة وتوقعاتها (تيرابايت في الشهر)



التوزيع الإقليمي المتوقع لحركة البيانات عبر الأجهزة المحمولة في العالم في عام 2017


320
مليون مستخدم
10
مليارات رسالة
400
مليون صورة

خصائص البيانات الضخمة



الاستخدامات الفعلية والمحتملة للبيانات الضخمة لأغراض التنمية

التطبيقات	التفسير	الأمثلة	التطبيقات والمحاذير
تصنيف مبادرة النبض العالمي الأمم المتحدة			
1. وصفي	يمكن للبيانات الضخمة توثيق ما يحدث ونقله	يشبه هذا التطبيق إلى حد كبير تطبيق "التوعية الآنية"، على الرغم من أن أهدافه أقل طموحًا. وأي معلومات مصورة (مخطط معلومات بياني)، بما في ذلك الخرائط التي تعالج كميات هائلة من البيانات، لتصيّرها واضحةً للقارئ ومقروءةً بسهولة، تُعدُّ مثالاً للتطبيق الوصفي	ينطوي وصف البيانات دائمًا على تبنّي خيارات وافتراضات –عن طبيعة البيانات وكيفية عرضها– ينبغي توضيحها وفهمها؛ فمن المعروف أن الرسوم البيانية والخرائط أيضًا قد تكون مضللة

2. تنبؤي	يمكن للبيانات الضخمة أن تفسر ما يحتمل أن يحدث، بغض النظر عن السبب	يشير أحد أنواع "التنبؤ" إلى ما قد يحدث لاحقًا– ويُعدُّ حفظ الأمن على أساس تنبؤي مثالاً على ذلك. ويشير نوع آخر إلى توقُّع الظروف السائدة من خلال البيانات الضخمة –كما في حالات التنبؤ بالمستويات الاجتماعية الاقتصادية باستخدام سجلات تفاصيل المكالمات الهاتفية في أمريكا اللاتينية وكوت ديفوار	تنطبق عليها تعليقات تطبيقات "الإنذار المبكر" و"التوعية الآنية" المماثلة

3. إرشادي، أو تشخيصي	قد تسلط البيانات الضخمة الضوء على أسباب حدوث الأشياء، وماذا يمكن عمله حيالها	حتى الآن، ليس هناك أي أمثلة واضحة تقريبًا على هذا التطبيق في سياقات التنمية. ويُعَدُّ مثال بيانات سجلات تفاصيل المكالمات الهاتفية المستخدمة لإظهار إمكانية "تحسين" خطوط الحافلات في أبيدجان هو الأقرب إلى حالة يمكن للتحليل فيها تحديد روابط سببية، وتشكيل سياسات	تنطبق عليها معظم تعليقات تطبيق "التعقيبات الآنية". وعلى وجه التحديد، يتطلب أي مثال على تنفيذ تطبيق تشخيصي وجود المقدرة على تحديد العلاقة السببية. ويعمل التطبيق الإرشادي بشكل أفضل من الناحية النظرية عندما تدعمه نظم وحلقات التعقيبات على تأثير أعمال السياسات

تصنيف بديل			
1. الإنذار المبكر	إن الكشف المبكر عن الحالات الشاذة في كيفية استخدام السكان للأجهزة والخدمات الرقمية يمكن أن يتيح استجابة أسرع في أوقات الأزمات	حفظ الأمن على أساس تنبؤي، استنادًا إلى فكرة أن تحليل البيانات التاريخية يمكن أن يكشف عن تركيبات معينة من العوامل المرتبطة بزيادة معدلات الأعمال الإجرامية في منطقة معينة، ويمكن استخدامه لتخصيص موارد الشرطة. ونظام اتجاهات الإنفلونزا من جوجل هو مثال آخر، حيث يتم تحليل عمليات البحث عن مصطلحات معينة (مثل 'سيلان الأنف'، أو 'حكة العين')؛ للكشف عن بداية موسم الإنفلونزا – على الرغم من أن دقته محل خلاف	يفترض هذا التطبيق إمكانية ملاحظة بعض الانتظام في السلوك البشري ونمذجتها. وتشمل التحديات الرئيسة للتنبؤ بالسياسات نزوع معظم أنظمة الكشف عن الخلل ونماذج التنبؤ إلى الإفراط في التنبؤ – أي أن يغلب عليها 'نتائج إيجابية زائفة'

2. النوعية الآنية	يمكن للبيانات الضخمة رسم تمثيل مفصل وحالي للواقع، من شأنه أن يمد التصميم بالمعلومات، ويقيد استهداف البرامج والسياسات	وجد باحثون – باستخدام البيانات الصادرة عن شركة Orange– درجة عالية من الارتباط بين الشبكات الاجتماعية وتوزيع اللغة في كوت ديفوار، مما يدل على أن هذه البيانات قد توفر معلومات حول الجماعات اللغوية في بلدان لا تتوافر عنها هذه البيانات	يكمن رونق هذا التطبيق وحجته الأساسية في فكرة أن البيانات الضخمة قد تكون بديلة لبيانات سيئة أو نادرة. ولكن النماذج التي تُظهر ارتباطًا كبيرًا بين المؤشرات 'المستندة إلى البيانات الضخمة' وتلك 'التقليدية' غالبًا ما تتطلب توافر المؤشرات التقليدية. ويعني تعبير 'الآنية' هنا استخدام بيانات رقمية عالية التواتر للحصول على صورة للواقع في أي وقت من الأوقات

3. التعقيبات الآنية	تُمكِّن القدرة على رصد السكان آنيا من إدراك مواضع الخلل في السياسات والبرامج، وإجراء التعديلات اللازمة	تستخدم الشركات الخاصة بالفعل تحليلات البيانات الضخمة. وعند استخدامها لأغراض التنمية، قد يشمل هذا التطبيق تحليل تأثير إجراءات سياسات معينة مثل سن قوانين مرور جديدة– آنيا	على الرغم من رونقه، هناك عدد قليل (إن وُجِدَ) من الأمثلة الواقعية لهذا التطبيق. ما يشكل تحديًا عند التأكد من أن أي تغيير ملحوظ يمكن أن يُعزى إلى التدخل أو 'العلاج'. ومع ذلك، قد تحتوي البيانات عالية التواتر أيضًا على 'تجارب طبيعية' –مثل انخفاض مفاجئ في أسعار سلعة معينة على الإنترنت– والتي يمكن الاستعانة بها لاستنتاج العلاقة السببية

مصادر البيانات الضخمة:

قامت اللجنة الاقتصادية لأوروبا، بتقديم تقرير بعنوان "ماذا تعنيه البيانات الضخمة للإحصاءات الرسمية" آذار مارس و قد أوردت فيه تصنيفا لمصادر البيانات الضخمة على النحو التالي:

المصادر الناشئة عن إدارة أحد البرامج، سواء كان برنامجا حكوميا أو غير حكومي، كالسجلات الطبية الإلكترونية وزيارات المستشفيات وسجلات التأمين والسجلات المصرفية وبنوك الطعام.

المصادر التجارية أو ذات الصلة بالمعاملات، الناشئة عن معاملات بين كيانين، على سبيل المثال معاملات البطاقات الائتمانية والمعاملات التي تجرى عن طريق الإنترنت بوسائل منها الأجهزة المحمولة.

مصادر شبكات أجهزة الاستشعار، على سبيل المثال، التصوير بالأقمار الصناعية، وأجهزة استشعار الطرق، وأجهزة استشعار المناخ.

مصادر أجهزة التتبع، على سبيل المثال تتبع البيانات المستمدة من الهواتف المحمولة والنظام العالمي لتحديد المواقع.

مصادر البيانات السلوكية، على سبيل المثال، مرات البحث على الإنترنت عن منتج أو خدمة ما أو أي نوع آخر من المعلومات، ومرات مشاهدة إحدى الصفحات على الإنترنت.

مصادر البيانات المتعلقة بالآراء، على سبيل المثال، التعليقات على وسائط التواصل الاجتماعي.

أمثلة من الواقع على البيانات الضخمة:

لدى واتس آب أكثر من مليار مستخدم، و يتم تداول اكثر من 42 مليار رسالة و حوالي 1.6 مليار صورة بشكل يومي.

فيسبوك تتعامل مع أكثر من 50 مليار صورة من مستخدميها.

تتعامل مع حوالي 100 مليار عملية بحث في الشهر. Googleجووجل

أهمية البيانات الضخمة:

للبيانات الضخمة اهمية عالية فهي تقدم ميزة تنافسية عالية للشركات اذا استطاعت الاستفادة منها ومعالجتها لأنها تقدم فهما أعمق لعملائها ومتطلباتهم ويساعد ذلك على اتخاذ القرارات المناسبة و الملائمة داخل الشركة بطريقة أكثر فعالية و ذلك بناء على المعلومات المستخرجة من قواعد بيانات العملاء وبالتالي زيادة الكفاءة والربح وتقليل الخسائر.

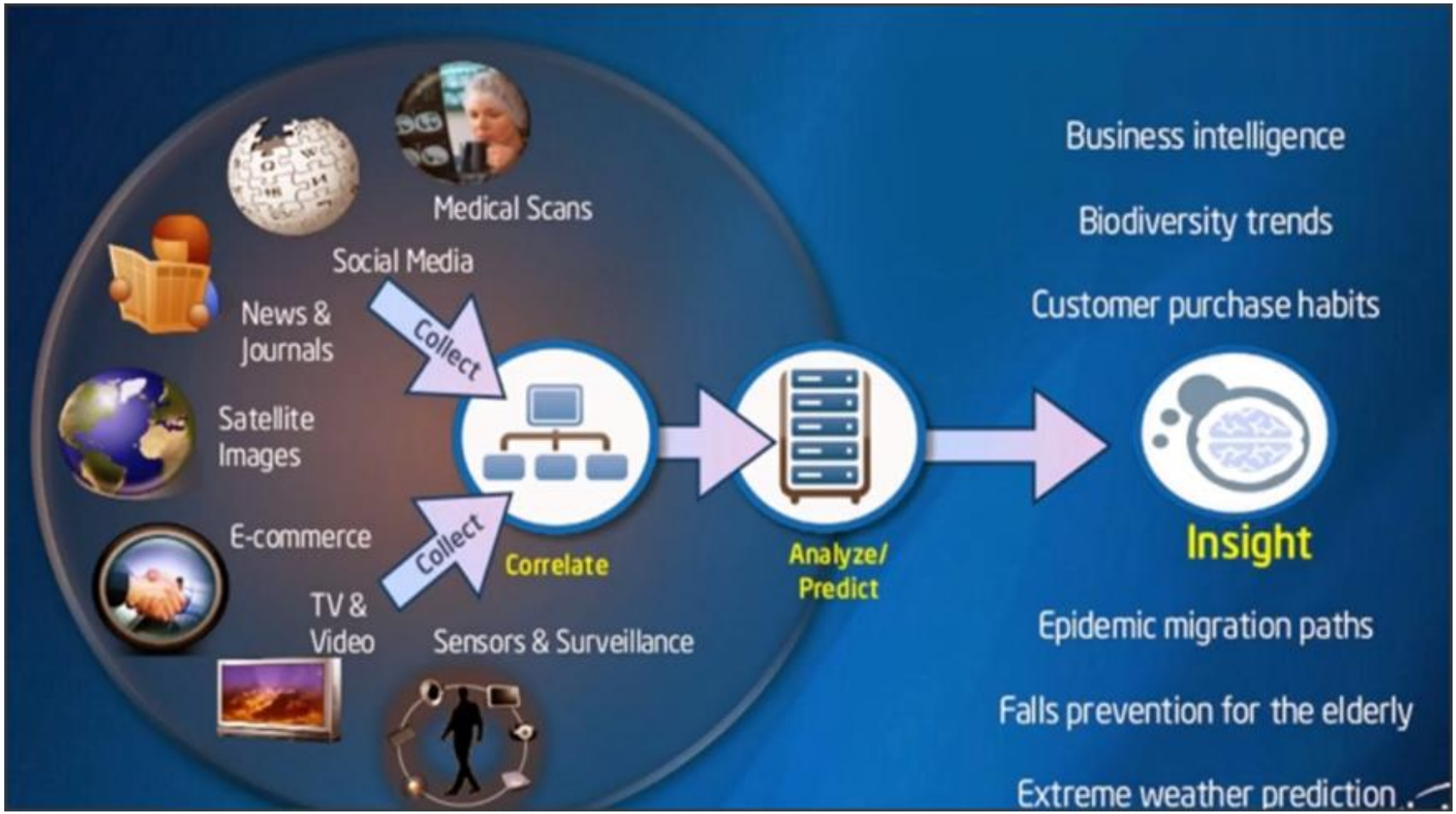
فباستخدام تقنيات و أدوات تحليل البيانات الضخمة استطاعت وول مارت تحسين نتائج البحث عن منتجاتها عبر الأنترنت بنسبة 10-15% بينما في تقرير لماكينزى-وهى شركة رائدة في مجال استشارات الاعمال- ان القطاع الصحي بالولايات المتحدة لو كان يستخدم تقنيات تحليل البيانات الضخمة بفاعليه وكفاءه لكان قد أنتج أكثر من 300 مليون دولار أمريكى كفائض سنوى من ميزانيه الصحة ثلثيها بسبب خفض تكاليف الانفاق بنسبة 8%

تحليل البيانات الكبيرة

أدى هذا إلى ضرورة الحاجة إلى استحداث طرق جديدة وفعالة لتحليل هذا الكم الهائل من المعلومات والاستفادة منها في تحسين العملية التعليمية وتطوير الأبحاث في مجال التعليم، وهو ما يعرف بعلم المعلومات و يعرف هذا العلم بأنه استخراج المعرفة من كميات كبيرة من البيانات كانت منظمة أو غير منظمة، و يعتبر تمديدا لمجال التحليلات التنبؤية وتنقيب البيانات .

لماذا تستخدم الشركات الكبيرة نظام تحليل البيانات الضخمة:

- تحسين العمليات الداخلية، مثل إدارة المخاطر، إدارة علاقات العملاء، والخدمات اللوجستية.
- تحسين المنتجات والخدمات القائمة.
- تطوير الخدمات والمنتجات الجديدة.
- الاستفادة من المعلومات واستهداف العروض المناسبة لعملائها في الوقت المناسب.


Medical Scans
Social Media
News & Journals
Satellite Images
E-commerce
TV & Video
Sensors & Surveillance
Collect
Collect
Correlate
Analyze/ Predict
Insight
Business intelligence
Biodiversity trends
Customer purchase habits
Epidemic migration paths
Falls prevention for the elderly
Extreme weather prediction

الخاتمة:

في زمننا هذا نشهد انفجارا ضخما في البيانات فتحليل و معالجة هذه البيانات يزيد بشكل رئيسي في فهم و استيعاب متطلبات العملاء و بالتالي زيادة الكفاءة و الانتاجية و تقليل الخسائر بالنسبة للشركات. إلا أن هنالك العديد من التحديات والمعوقات التي تعيق استخدام البيانات الضخمة أو التوسع في استخدامها والتي ممكن تناولها في مقال آخر بإذن الله تعالى . ومع مرور الوقت والتقدم التكنولوجي، فمن المتوقع إحراز تقدم كبير فيما يتعلق بمعالجة تحديات ومعوقات استخدام البيانات الضخمة بشكل أوسع.

شكراً