الأكاديمية العربية الدولية
Arab International Academy

الأكاديمية العربية الدولية

المقررات الجامعية

www.aiacademy.info | care@aiacademy.info

ويكيبيديا

بيانات ضخمة

البيانات الضخمة مصطلح يشير إلى مجموعة بيانات تستعصي لضخامتها أو تعقيدها على التخزين أو المعالجة باحدى الأدوات أو التطبيقات المعتادة لإدارة البيانات. أو ببساطة لتقريب الأفهام، لا يُمكن التعامل معها على حاسوب عادي بمفرده من خلال قاعدة بيانات بسيطة. ومن سمات مجال «**البيانات الضخمة**» استعمال حواسب عديدة لتقاسم الأعمال المطلوبة.

محتويات

سماتها

نبذة

التعريف

أمثلة

العلوم الكبيرة

العلوم والأبحاث

الحكومة

القطاع الخاص

التنمية الدولية

سوق العمل

الهندسة المعمارية

التقنيات

الأنشطة البحثية

النقد

انتقادات نموذج البيانات الضخمة

انتقادات تنفيذ البيانات الضخمة

محاور اساسيه للبيانات الضخمة (3Vs)

انظر أيضًا

وصلات خارجية

المراجع

سماتها

- حجمها،
- وسرعتها (بمعنى معدل تولدها وتغيرها ووجوب معالجتها مباشرة)،
- وتنوعها (انعدام تناسقها وتعدد أنواعها وأنها غير مرتبة)،
- وأهميتها (القيمة المضافة بفهمها)
- وحقيقتها (دقة أدوات القياس ونحو ذلك؛ هل يمكن الاعتماد عليها؟)

وثمة غير السمات الخمسة أعلاه، مثلا عرضتها للسرقة لو كانت حساسة.

نبذة

تشمل التحديات الالتقاط، والمدة، والتخزين[1]، والبحث، والمشاركة، والنقل، والتحليل والتصور[2]. ويرجع الاتجاه إلى مجموعات **البيانات الضخمة** بسبب المعلومات الإضافية المشتقة من تحليل مجموعة واحدة كبيرة من البيانات ذات الصلة، بالمقارنة مع المجموعات المنفصلة الأصغر حجماً مع نفس الحجم الإجمالي للبيانات، مما يسمح بوجود ارتباطات تكشف "الاتجاهات التجارية المحورية، وتحديد جودة البحث، وربط الاستشهادات القانونية، ومكافحة الجريمة وتحديد ظروف حركة تدفق البيانات في الوقت الحقيقي"[3][4][5].

اعتباراً من عام 2012، كانت الحدود المفروضة على حجم مجموعات البيانات الملائمة للمعالجة في مدة معقولة من الوقت خاضعة لوحدة قياس البيانات إكسابايت[6][7]. عادة ما يواجه العلماء عددا من القيود بسبب مجموعات البيانات الضخمة الموجودة في العديد من المجالات، والتي تتضمن الأرصاد الجوية(علم الطقس)، وعلم الجينات(علم الجينوم)[8]، والمحاكاة الفيزيائية[9] المعقدة والبحوث البيولوجية والبيئية[10]. وتؤثر القيود أيضاً على بحث الانترنت(محرك بحث)، وتقنية الأعمال التجارية والتمويل. وتنمو مجموعات البيانات في الحجم بشكل جزئي، ويرجع ذلك لأنها يتم جمعها بشكل متزايد عن طريق أجهزة استشعار المعلومات المتنقلة، والتقنيات الحسية الجوية (الاستشعار عن بعد)[11][12]، وسجلات البرامج، والكاميرات، والميكروفونات، وأجهزة تحديد ذبذبات الإرسال(تحديد الهوية باستخدام موجات الراديو) وشبكات استشعار اللاسلكية. وتضاعفت القدرة التكنولوجية العالمية لتخزين المعلومات للفرد الواحد تقريباً كل 40 شهر من الثمانينات، واعتباراً من عام 2012، ينشيء 2.5 كوينتيليون بايت (2.5×10^{18}) من البيانات يومياً[13]. والتحدي بالنسبة للشركات الكبيرة هو تحديد من يجب أن يمتلك مبادرات البيانات الضخمة التي تنتشر على المنظمة بأكملها.[14]

من الصعب العمل مع البيانات الضخمة باستخدام معظم أنظمة إدارة قواعد البيانات العلائقية وإحصائيات سطح المكتب وحزم المحاكاة، حيث يتطلب الأمر بدلاً من ذلك "برامج متوازية واسعة النطاق تعمل على عشرات أو مئات أو حتى آلاف الخوادم".[15] وما يُعتبر "بيانات ضخمة" يختلف باختلاف قدرات المنظمة التي تقوم بإدارة المجموعة، وعلى قدرات التطبيقات التي تستخدم بشكل تقليدي لمعالجة وتحليل مجموعة البيانات في النطاق الخاص بها. "فبالنسبة لبعض المنظمات، ربما تؤدي مواجهة مئات الغيغا بايت من البيانات لأول مرة إلى إعادة النظر في خيارات إدارة البيانات. وبالنسبة للبعض الآخر، ربما يستغرق الأمر عشرات أو مئات تيرابايت من البيانات قبل أن يصبح حجم البيانات شأناً مهماً".[16]

التعريف

البيانات الضخمة عادة ما تتضمن مجموعات بيانات ذات أحجام تتخطى قدرة البرامج التي يشيع استخدامها لالتقاط وإدارة ومعالجة البيانات في غضون فترة زمنية مقبولة.[17] وبالنسبة لأحجام البيانات الضخمة فهي هدف متحرك باستمرار، فاعتباراً من عام 2012، يتراوح حجمها بين بضع عشرات من تيرابايت إلى العديد من بيتابايت من البيانات في مجموعة واحدة

فقط. ومع هذه الصعوبة، يتم تطوير منصات جديد من أدوات "البيانات الضخمة" للتعامل مع مختلف الجوانب الخاصة بالكميات الكبيرة من البيانات.

في تقرير بحثي وعدد من المحاضرات المتعلقة به عام 2001،[18] قام "دوغ لاني" محلل مجموعة META Group (المعروفة الآن باسم Gartner) بتعريف تحديات نمو البيانات وفرصها كعنصر ثلاثي الأبعاد، بمعني زيادة الحجم (كمية البيانات)، السرعة (سرعة البيانات الصادرة والواردة) والتنوع (تنوع أنواع البيانات ومصادرها). وتقوم Gartner والكثير من الشركات في هذه الصناعة الآن بالاستمرار في استخدام نموذج "3Vs" لوصف البيانات الضخمة.[19] وفي 2012، قامت Gartner بتحديث تعريفها ليصبح كالتالي: "البيانات الضخمة هي أصول معلومات كبيرة الحجم، عالية السرعة، و/أو عالية التنوع تتطلب أشكال جديدة من المعالجة لتعزيز عملية صنع القرار والفهم العميق و تحسين العملية".

تعريف TBDI للبيانات الضخمة: البيانات الضخمة هو مصطلح ينطبق علي الأجسام الضخمة للبيانات التي تتنوع في طبيعتها سواء أكانت منظمة، غير منظمة أو شبه منظمة، بما في ذلك من المصادر الداخلية أو الخارجية للمنظمة، ويتم توليدها بدرجة عالية من السرعة مع نموذج مضطرب، والتي لا تتفق تماماً مع مخازن البيانات التقليدية والمنظمة وتتطلب نظام إيكولوجي قوي ومعقد مع منصة حوسبة عالية الأداء وقدرات تحليلية للالتقاط ومعالجة وتحويل وكشف واستخلاص القيمة والرؤى العميقة في غضون وقت زمني مقبول".[20]

أمثلة

تتضمن الأمثلة العلم الكبير، سجلات الويب، تحديد الهوية باستخدام موجات الراديو (بالإنجليزية: RFID)، شبكات الاستشعار، الشبكات الاجتماعية، البيانات الاجتماعية [21] (يرجع هذا لثورة البيانات الاجتماعية)، نصوص الانترنت والوثائق، فهرسة بحث الانترنت، تفاصيل سجلات الاتصال، علم الفلك، علوم الغلاف الجوي، علم الجينات، العلوم الكيميائية والبيولوجية وغيرها من البحوث المعقدة وأغلبية المراقبات العسكرية، السجلات الطبية، أرشيفات الصور والتجارة الإلكترونية واسعة النطاق.

العلوم الكبيرة

تُمثل تجارب مصادم الهدرونات الكبير (بالإنجليزية: Large Hadron Collider) حوالي 150 مليون جهاز استشعار تقدم بيانات 40 مليون مرة في الثانية الواحدة. وهناك ما يقرب من 600 مليون تصادم في الثانية الواحدة. وبعد تصفية و تنقيح تسجيلات أكثر من 99.999% من هذه التدفقات، نجد أن هناك 100 تعارض للفائدة في الثانية الواحدة[22][23][24].

- ونتيجة لذلك، بالعمل مع أقل من 0.001% فقط من بيانات تيار الاستشعار، فإن تدفق البيانات من جميع تجارب LHC الأربعة تمثل 25 بيتابايت المعدل السنوي قبل النسخ المتماثل (اعتباراً من 2012). وهذا يصبح تقريباً 200 بيتابايت بعد النسخ المتماثل.
- وإذا تصورنا أن جميع بيانات الاستشعار كان سيتم تسجيلها في LHC، فإن تدفق البيانات كان سيصعب العمل معه للغاية. حيث سيتجاوز تدفق البيانات 150 مليون بيتابايت المعدل السنوي، أو ما يقرب من 500 إكسابايت في اليوم الواحد، قبل النسخ المتماثل. وبالنظر للرقم بشكل نظري، فإنه يصبح مُعادل لـ 500 كوينتيليون بايت (5 1020 x) في اليوم، وهو رقم أعلي 200 مرة تقريباً من جميع المصادر الأخرى المجتمعة في العالم.

العلوم والأبحاث

- عندما بدأ مسح سلوان الرقمي للسماء (بالإنجليزية: SDSS) جمع البيانات الفلكية في عام 2000، فإنه قد جمع بيانات في أسابيعه القليلة الأولى أكثر مما تم جمعه في تاريخ علم الفلك بأكمله. ومع استمراره بمعدل 200 جيجا بايت في الليلة، جمع SDSS أكثر من 140 تيرابايت من المعلومات. وعندما يأتي Large Synoptic Survey Telescope خليفة SDSS إلى أرض الواقع في عام 2016، فمن المتوقع أن يقوم بجمع هذه الكمية من البيانات كل خمسة أيام.[3]
- إن فك رموز الجين البشري تستغرق عادة 10 سنوات حتى تتم العملية، ولكن الآن فإن هذه العملية يمكن إنجازها في أسبوع واحد.[3]
- بالنسبة للعلوم الاجتماعية الحسابية – استخدم "توبياس بريس" وآخرون بيانات اتجاهات جوجل (بالإنجليزية: Google Trends) لإثبات أن مستخدمي الإنترنت من البلدان التي لديها ناتج محلي إجمالي أعلى للفرد (GDP) يتجهون للبحث عن معلومات حول المستقبل أكثر من المعلومات المتعلقة بالماضي. وتشير النتائج إلى أنه قد يكون هناك ارتباط بين السلوك عبر الانترنت والمؤشرات الاقتصادية في العالم الحقيقي.[25][26][27] وقد قام مؤلفو هذه الدراسة بفحص تسجيلات جوجل المصنوعة من قبل مستخدمي الإنترنت في 45 دولة مختلفة عام 2010، وقاموا بحساب نسبة حجم البحث للسنة التالية "2011" مقارنة بحجم البحث في السنة السابقة "2009" والذي أطلق عليه اسم "مؤشر التوجه المستقبلي".[28] ثم قاموا بمقارنة التوجه المستقبلي مع الناتج المحلي الإجمالي للفرد الواحد في كل بلد، حيث وجدوا اتجاه قوي للبلدان التي يقوم فيها مستخدمي جوجل بالبحث عن المستقبل بالحصول على ناتج محلي إجمالي أعلى. وتلمح النتائج لاحتمال وجود علاقة بين النجاح الاقتصادي للبلد وسلوك مواطنيها في البحث عن المعلومات المأسورة في البيانات الضخمة.

الحكومة

- في عام 2012، أعلنت إدارة أوباما عن مبادرة التنمية وبحوث البيانات الضخمة والتي تناولت كيفية استخدام البيانات الضخمة لمعالجة المشاكل الهامة التي تواجه الحكومة.[29] وقد تألفت المبادرة من 84 برنامج مختلف للبيانات الضخمة موزعة على 6 دوائر.[30]
- والجدير بالذكر، أن تحليل البيانات الضخمة قد لعب دوراً كبيراً في حملة إعادة انتخاب باراك أوباما الناجحة عام 2012.[31]
- وتمتلك الحكومة الاتحادية للولايات المتحدة 6 من أصل 10 أجهزة كمبيوتر تُعد هي الأكثر نفوذا في العالم والتي يُطلق عليها " Supercomputers".[32]
- وتقوم وكالة الأمن الوطني الأمريكي حالياً ببناء مركز بيانات يوتاه " Utah Data Center"، والذي سيكون قادر علي التعامل مع معلومات تقدر مساحاتها بـ يوتابايت والتي جمعتها وكالة الأمن القومي عبر الإنترنت.[33][34]

القطاع الخاص

- يقوم أمازون (موقع) Amazon.com بمعالجة ملايين العمليات الخلفية كل يوم، فضلاً عن استفسارات من أكثر من نصف مليون بائع طرف ثالث. ويعتمد أمازون على تقنية الليينكس بشكل أساسي كي تعمل وسط هذا الكم الهائل من البيانات، واعتباراً من 2005 كانت أمازون تمتلك أكبر 3 قواعد بيانات ليينكس في العالم والتي تصل سعتها إلي 7.8، 18.5 و 24.7 تيرابايت.[35]
- ومن ناحية أخرى، يقوم متجر وول مارت Walmart بمعالجة أكثر من مليون معاملة تجارية كل ساعة، والتي يتم استيرادها إلي قواعد بيانات يُقدر أنها تحتوي على أكثر من 2.5 بيتابايت (2560 تيرابايت) من البيانات – وهو ما يوازي 167 ضعف البيانات الواردة في جميع الكتب الموجودة في مكتبة الكونغرس في الولايات المتحدة.[3]
- أما الفيسبوك فيعالج 50 مليار صورة من قاعدة مستخدميه. ويقوم نظام حماية بطاقات الائتمان من الاحتيال " FICO Falcon Credit Card Fraud Detection System" بحماية 2.1 مليار حساب نشط في جميع أنحاء العالم.

- ووفقاً لأحدث الإحصائيات، فإن حجم البيانات التجارية في جميع أنحاء العالم، عبر جميع الشركات، يتضاعف حجمها كل 1.2 سنة.[36]
- وتقوم شركة Windermere Real Estate باستخدام إشارات GPS مجهولة من ما يقرب من 100 مليون سائق لمساعدة مشتري المنازل الجدد لتحديد أوقات قيادهم من وإلي العمل خلال الأوقات المختلفة لليوم.[37]

التنمية الدولية

بعد عقود من العمل في مجال الاستخدام الفعال لتكنولوجيا المعلومات والاتصالات من أجل التنمية (أو ICT4D)، فقد قيل أنه يمكن للبيانات الضخمة أن تسهم إسهاماً كبيراً في التنمية الدولية.[38][39] من ناحية، فإن ظهور البيانات الضخمة يوفر احتماليات فعالة من حيث التكلفة لتحسين عملية صنع القرار في مجالات التنمية الحيوية مثل الرعاية الصحية، العمالة، الإنتاجية الاقتصاد، الجريمة والأمن، الكوارث الطبيعية وإدارة الموارد. ومن ناحية أخرى، فإن جميع المخاوف المتعلقة بالبيانات الضخمة مثل الخصوصية، تحديات التشغيل البيني، والقوة غير محدودة للحوارزميات المنقوصة تتفاقم في البلدان النامية من خلال تحديات التنمية طويلة الأمد مثل الافتقار إلي البنية التحتية التكنولوجية والاقتصادية بالإضافة لندرة الموارد البشرية. "وهذا أدي إلي إحداث نوع جديد من الفجوات التقنية: فجوة في تقصي البيانات لاتخاذ قرارات مستنيرة".

سوق العمل

لقد تسببت "البيانات الضخمة" في زيادة الطلب علي المتخصصين في إدارة المعلومات لهذا البرنامج الضخم، وقد أنفقت عدد من الشركات العالمية مثل Oracle Corporation، IBM، Microsoft، SAP، EMC و HP أكثر من 15 مليار دولار علي شركات البرمجيات المتخصصة فقط في مجال إدارة البيانات والتحليلات. وفي عام 2010، كانت هذه الصناعة مستقلة بذاتها تساوي أكثر من 100 مليار دولار، كما أنها تنمو تقريباً بمعدل 10% سنوياً، أي حوالي ضعفي قطاع البرمجيات ككل.[3]

تقوم البلدان ذات الاقتصاديات المتقدمة باستخدام التقنيات كثيفة البيانات بشكل متزايد. فهناك 4.6 مليار اشتراك للهواتف النقالة حول العالم، وهناك ما بين 1 مليار إلي 2 مليار شخص يتصل بالإنترنت.[3] وبين عامي 1990 و 2005، أرتقي أكثر من مليار شخص حول العالم بمكانهم إلي الطبقة المتوسطة مما يعني أن هناك الكثير والكثير من الناس الذين يكسبون المال سوف يصبحون أكثر تثقيفاً والذي يؤدي بدوره إلي نمو المعلومات. لقد كانت قدرة العالم الفعالة لتبادل المعلومات من خلال شبكات الاتصال السلكية واللاسلكية هي 281 بيتابايت في 1986، 471 بيتابايت في 1993، 2.2 إكسابايت في عام 2000، 65 إكسابايت في عام 2007 ويتوقع أن تصل كمية البيانات المتدفقة عبر شبكة الانترنت إلي 667 إكسابايت سنوياً بحلول عام 2013.[3]

الهندسة المعمارية

نظراً لتعقيدات نظم البيانات الضخمة، فقد كان وجود ممارسات متطورة للهندسة المعمارية الخاصة بالبيانات الضخمة أمر لابد منه. إن الإطار المعماري للبيانات الضخمة (BDAF) هو إطار هيكلي لحلول البيانات الضخمة، والذي يهدف إلي المساعدة في إدارة مجموعة من الأعمال الفنية المتميزة وتنفيذ مجموعة من عناصر التصميم المحددة. إن الغرض من (BDAF) هو فرض الالتزام بنهج تصميم متناسق، الحد من تعقيدات النظام، تعظيم إعادة الاستخدام، تخفيض التبعيات وزيادة الإنتاجية.

إن الإطار المعماري للبيانات الضخمة (BDAF) يضم أربعة أجزاء متكاملة: دومين محدد، منصة، الاعتماد علي التفعيل ونموذج محايد تكنولوجياً. وتعتبر مكونات (BDAF) هي نموذج مركزي، تتحكم به الهندسة المعمارية، ويشكل بناء متماسك لمعالجة البيانات الضخمة، بما في ذلك استخراج البيانات، التخزين، المعالجة، التخطيط، التجميع، الإرسال والتواصل، إعداد التقارير، التصور، الرصد، التدقيق والتشغيل الآلي.

في عام 2004، نشرت جوجل بحث عن عملية تُدعي MapReduce والتي استخدمت هندسة معمارية مثل هذه. حيث يوفر الإطار الخاص بـ MapReduce نموذج برمجة متوازي والتطبيق المرتبط به لمعالجة كمية هائلة من البيانات. من خلال MapReduce، يتم تقسيم الأطروحات وتوزيعها عبر العقد المتوازية ومعالجتها بشكل متواز (خطوة the Map). ثم يتم تجميع النتائج بعد ذلك وتسليمها (خطوة the Reduce). لقد كان الإطار ناجح بشكل مذهل، لذا أراد البعض تكرار تلك الخوارزمية. ولذلك، أعتُمد تنفيذ إطار MapReduce من قبل مشروع Apache مفتوح المصدر أطلق عليه اسم Hadoop.[40]

إن MIKE2.0 هو نهج مفتوح لإدارة المعلومات يتناول منهجية التعامل مع البيانات الضخمة من حيث التعديل المفيد لمصادر البيانات، التعقيد في العلاقات المتبادلة والصعوبة في حذف (أو تعديل) السجلات الفردية.[41]

التقنيات

تتطلب البيانات الضخمة تقنيات استثنائية لمعالجة الكميات الكبيرة من البيانات بكفاءة ضمن الوقت المسموح. ويشير تقرير ماكينزي 2011[42] لبعض التقنيات المناسبة التي تتضمن اختبار A/B، تعلم قاعدة المصادقة، التصنيف، التحليل العنقودي، انصهار وتكامل البيانات، الخوارزميات الجينية، التعلم الآلي، معالجة اللغات الطبيعية، الشبكات العصبية، التعرف على الأنماط، الكشف عن الأشياء الشاذة، النمذجة التنبؤية، الانحدار، تحليل وجهات النظر، معالجات الإشارات، التعلم الخاضع والغير خاضع للرقابة، المحاكاة، تحليل السلاسل الزمنية والتصور. إن البيانات الضخمة متعددة الأبعاد يمكن أيضاً أن تمثل مثل tensors،[43] والتي يمكن التعامل معها بكفاءة أكبر من خلال الحسابات التي تعتمد علي الموترة مثل التعلم الفضائي الجزئي متعدد الخطي.[44] والتقنيات الإضافية التي يجري تطبيقها علي البيانات الضخمة تتضمن قواعد بيانات هائلة تتم معالجتها بشكل متوازي (MPP)، التطبيقات المعتمدة على البحث، شبكات البيانات والتعدين، أنظمة الملفات الموزعة، قواعد البيانات الموزعة، البنية التحتية المعتمدة علي التخزين السحابي (التطبيقات، التخزين ومصادر الحوسبة) والإنترنت.[بحاجة لمصدر]

إن بعض وليس كل قواعد البيانات العلائقية MPP لديها القدرة على تخزين وإدارة بيتابايت من البيانات. والمفهوم ضمنياً هو القدرة علي تحميل، مراقبة، النسخ الاحتياطي، و تحقيق الاستخدام الأمثل لجداول البيانات الضخمة في RDBMS.[45]

إن برنامج تحليل بيانات DARPA يستهدف البنية الأساسية لمجموعات البيانات الهائلة، وفي عام 2008 ظهرت هذه التقنية جمهور مع انطلاقة شركة تُدعي Ayasdi.

إن ممارسي عمليات تحليل البيانات الضخمة عادة ما يكونوا معاديين لمساحات التخزين المشتركة الأبطأ،[46] مُفضلين مساحات التخزين المتصلة والمباشرة (DAS) في جميع أشكالها المختلفة بدءاً من محركات الأقراص الصلبة (SSD) وصولاً إلى أقراص الساتا عالية القدرة والموضوعة داخل عقد معالجة متوازية. وإذا نظرنا إلي البنية المعمارية لمساحات التخزين المشتركة SAN و NAS فسوف نجد أنها بطيئة، معقدة وباهظة الثمن. وهذه الصفات لا تتفق مع أنظمة تحليل البيانات الضخمة التي تقوم علي أداء النظام، البنية التحتية والتكلفة المنخفضة.

إن تقسيم المعلومات في الوقت الحقيقي أو شبه الحقيقي هي واحدة من الخصائص المميزة لتحليل البيانات الضخمة. وبالتالي، يتم تجنب الحمول كلما وحيثما كان ذلك ممكناً. إن تكلفة SAN في النطاق اللازم لتطبيقات التحليلات تُعد أعلي بكثير جداً من تقنيات التخزين الأخرى.

هناك مزايا وكذلك يوجد عيوب لمساحات التخزين المشتركة في تحليلات البيانات الضخمة، ولكن ممارسي تحليل البيانات الضخمة لم يحبذوا ذلك بدءاً من عام 2011.[47]

الأنشطة البحثية

في مارس 2012، أعلن البيت الأبيض عن "مبادرة البيانات الضخمة" القومية التي تتألف من 6 إدارات ووكالات فيدرالية تودع أكثر من 200 مليون دولار لمشاريع البيانات الضخمة البحثية.[48]

وقد تضمنت المبادرة National Science Foundation "بعثات في الحوسبة" والتي منحت 10 مليون دولار علي مدي 5 سنوات لمعمل AMPLab إغلاق <ref/> مفقود لوسم <ref> كما تلقي AMPLab أيضاً تمويل من DARPA، وأكثر من اثني عشر راعياً صناعياً ويستخدم البيانات الضخمة لمواجهة مجموعة واسعة من المشاكل بدءاً من الاختناقات المرورية[49] وحتي مكافحة السرطان.[50]

وشملت مبادرة البيت الأبيض أيضاً التزاماً من وزارة الطاقة لتوفير 25 مليون دولار علي مدار 5 سنوات لإنشاء معهد إدارة و تحليل و تصور البيانات (SDAV)[51]، والذي يتم قيادته من قبل معمل لورانس بيركلي الوطني التابع لوزارة الطاقة. ويهدف معهد SDAV تجمع الخبرات من 6 مختبرات وطنية و 7 جامعات لتطوير أدوات جديدة لمساعدة العلماء في إدارة و تصور البيانات علي أجهزة الكمبيوتر العملاقة الخاصة بالإدارة.

هذا وقد أعلنت ولاية ماساشوستس الأمريكية عن مبادرة ماساشوستس للبيانات الضخمة في مايو2012، والتي توفر التمويل من حكومة الولاية وشركات القطاع الخاص لمجموعة متنوعة من المؤسسات البحثية.[52] وقد استضاف معهد ماساشوستس للتكنولوجيا مركز إنتل للعلوم والتكنولوجيا الخاص بالبيانات الضخمة في مختبر MIT لعلوم الكمبيوتر والذكاء الاصطناعي.[53]

وتقوم المفوضية الأوروبية علي مدار عامين بتمويل منتدي القطاعين العام والخاص للبيانات الضخمة من خلال برنامجهم السابع لإشراك الشركات والأكاديميات وغيرهم من أصحاب المصلحة في مناقشة قضايا البيانات الضخمة. ويهدف المشروع إلي تحديد إستراتيجية خاصة بالبحث والابتكار لتوجيه إجراءات الدعم من المفوضية الأوروبية للتنفيذ الناجح لاقتصاد البيانات الضخمة. وسوف تستخدم نتائج هذا المشروع كمدخل لمشروعهم التالي Horizon 2020.[54]

النقد

إن انتقادات نموذج البيانات الضخمة تأتي من ناحيتين، الأولي نابعة من أولئك الذين يشككون في الآثار المترتبة علي النهج نفسه. والثانية تأتي من الذين يشككون في الطريقة التي يتم تنفيذها حالياً.

انتقادات نموذج البيانات الضخمة

"المشكلة الكبيرة هي أننا لا نعرف الكثير عن العمليات التجريبية الأساسية الصغرى التي تؤدي إلي ظهور خصائص الشبكة النموذجية للبيانات الضخمة".[17] في نقدهم للبيانات الضخمة أشار Snijders، Matzat و Reips إلي أنه في كثير من الأحيان يتم طرح افتراضات قوية جداً حول الخصائص الرياضية التي قد لا تعكس علي الإطلاق ما يحدث في الواقع علي مستوي

العمليات الصغرى. وقد وجه مارك غراهام انتقادات واسعة لتأكيد لريس أندرسون بأن البيانات الضخمة سوف توضح نهاية نظرية: مع التركيز بصفة خاصة على فكرة أن البيانات الضخمة سيوف تحتاج دائماً إلى أن يتم وضعها في سياقها الاجتماعي، والاقتصادي والسياسي.[55] حتي إذا كانت هناك شركة تستثمر 8 أو 9 مبالغ مالية لاشتقاق البصيرة من المعلومات المتدفقة من الموردين والعملاء، فإن 40% من الموظفين فقط هم من لديهم مهارات ناضجة بما فيه الكفاية للقيام بذلك. و للتغلب علي هذا العجز، فإن "البيانات الضخمة" مهما كانت شاملة أو تم تحليلها بشكل جيد، فإنه يجب أن تستكمل من قبل "حكم كبير"، وفقاً لمقال نشر في مجلة Harvard Business Review.[56]

وفي نفس السياق، فقد تم الإشارة إلى أن القرارات المستندة علي تحليل البيانات الضخمة تُعد حتميتً "فقد عرفناها من العالم متلما حدثت بالماضي، أو في أحسن الأحوال عرفناها كما هي حالياً.[57] فمن خلال تغذيتها بعدد كبير من البيانات الخاصة بالتجارب السابقة، يمكن للخوارزميات التنبؤ بالتطور ا لمستقبلي إذا كان المستقبل يشبه الماضي. وإذا تغيرت ديناميكيات النظم في المستقبل، فإن الماضي سوف يكون لديه القليل ليقوله عن المستقبل. ولهذا، سيكون من الضروري وجود فهم دقيق لديناميكية النظم، وهو ما يعني ضمنياً وجود نظرية.[58] ورداً علي هذا النقد، فقد أقتُرح ضم مناهج البيانات الضخمة مع المحاكاة الحاسوبية، مثل النماذج القائمة علي وكيل.[57] حيث تقوم هذه النماذج علي نحو متزايد بالتحسن في توقع نتائج التعقيدات الاجتماعية حتي للسيناريوهات المستقبلية الغير معروفة من خلال المحاكاة الحاسوبية التي تعتمد علي مجموعة من الخوارزميات المترابطة مع بعضها البعض. وبالإضافة لذلك، تقوم باستخدام طرق ذات متغيرات متعددة والتي تبحث في البنية الكامنة من البيانات مثل تحليل العامل و تحليل الكتلة، والتي أثبتت فائدتها كمناهج تحليلية تتفوق علي المناهج ثنائية التعدد والي تعمل عادة مع مجموعات البيانات الأصغر حجماً.

إن المدافعين عن خصوصية المستهلك يشعرون بالقلق تجاه تهديدات الخصوصية المتمثلة في زيادة مساحة التخزين وتكامل المعلومات الشخصية، وقد أصدرت لجنة الخبراء توصيات مختلفة لسياسة الخصوصية تتوافق مع مستوي التوقعات.[59]

انتقادات تنفيذ البيانات الضخمة

لقد أثار دانا بويد عدة مخاوف حول استخدام البيانات الضخمة في العلم، ولكنه أغفل عدة مبادئ مثل اختياره لعينة متمثلة في عدد من الأشخاص القلقين جداً من التعامل في الواقع مع كميات ضخمة من البيانات.[60] وقد يؤدي هذا النهج إلي تحيز في النتائج بطريقة أو بأخرى. فالتكامل بين موارد البيانات الغير متجانسة - يمكن أن يعتبره البعض "بيانات ضخمة" وقد لا يعتبره البعض كذلك - وهو ما يمثل تحديات لوجستية و تحليلية هائلة، ولكن العديد من الباحثين يرون أن مثل هذه التكاملات من المحتمل أن تمثل الحدود الجديدة الواعدة في مجال العلوم.[61]

محاور اساسيه للبيانات الضخمة (3Vs)

1. الحجم ان يكون حجم البيانات ضخمة ويقاس حجمها بالبيتا بايت أو الاكسا بايت.
2. السرعة وهو سرعة تدفق البيانات وتداولها.
3. التنوع اشكال مختلفه من البيانات في شكل نصوص.

انظر أيضًا

- البيانات
- بيانات غير مؤكدة
- قاعدة البيانات الموزعة

- تحليلات تنبؤية

وصلات خارجية

- مقالة البصمة الرقمية وتقاسم البيانات (https://www.albayan.ae/knowledge/2014-11-04-1.2235594) للكاتب ناثان إيغل بتاريخ 4\11\2014

المراجع

1. Kusnetzky, Dan. "What is "Big Data?" ". ZDNet. مؤرشف من الأصل في 21 فبراير 2010.
2. Vance, Ashley (22 April 2010). "Start-Up Goes After Big Data With Hadoop Helper". *New York Times Blog*. مؤرشف من الأصل في 18 نوفمبر 2018.
3. اكتب عنوان المرجع بين علامتي الفتح <ref> والإغلاق </ref> للمرجع Economist
4. "E-Discovery Special Report: The Rising Tide of Nonlinear Review". Hudson Global. مؤرشف من الأصل في 11 فبراير 2014. اطلع عليه بتاريخ 01 يوليو 2012. by Cat Casey and Alejandra Perez
5. "What Technology-Assisted Electronic Discovery Teaches Us About The Role Of Humans In Technology — Re-Humanizing Technology-Assisted Review". فوربس (مجلة). مؤرشف من الأصل في 20 فبراير 2019. اطلع عليه بتاريخ 01 يوليو 2012.
6. Francis, Matthew (2012-04-02). "Future telescope array drives development of exabyte processing". مؤرشف من الأصل في 31 مارس 2019. اطلع عليه بتاريخ 24 أكتوبر 2012.
7. Watters, Audrey (2010). "The Age of Exabytes: Tools and Approaches for Managing Big Data". Hewlett-Packard Development Company (Website/Slideshare). مؤرشف من الأصل في 11 يناير 2015. اطلع عليه بتاريخ 24 أكتوبر 2012.
8. Community cleverness required". *Nature*. **455** (7209): 1. 4 September 2008." doi:10.1038/455001a. مؤرشف من الأصل في 9 يوليو 2017.
9. Sandia sees data management challenges spiral". *HPC Projects*. 4 August 2009". مؤرشف من الأصل في 3 أبريل 2016.
10. Reichman, O.J.; Jones, M.B.; Schildhauer, M.P. (2011). "Challenges and Opportunities of Open Data in Ecology". *Science*. **331** (6018): 703–5. doi:10.1126/science.1197962.
11. Hellerstein, Joe (9 November 2008). "Parallel Programming in the Age of Big Data". *Gigaom Blog*. مؤرشف من الأصل في 3 أبريل 2019.
12. Segaran, Toby; Hammerbacher, Jeff (2009). *Beautiful Data: The Stories Behind Elegant Data Solutions*. O'Reilly Media. صفحة 257. ISBN 978-0-596-15711-1. مؤرشف من الأصل في 31 مارس 2019.
13. IBM What is big data? — Bringing big data to the enterprise". 01.ibm.com". مؤرشف من الأصل في 30 مايو 2017. اطلع عليه بتاريخ 05 مارس 2013.
14. Oracle and FSN, "Mastering Big Data: CFO Strategies to Transform Insight into Opportunity" (https://fsn.co.uk/channel_bi_bpm_cpm/mastering_big_data_cfo_strategies_to_transform_insight_into_opportunity#.UO2Ac-TTuys), December 2012 نسخة محفوظة (https://web.archive.org/web/20170505051557/http://www.fsn.co.uk/channel_bi_bpm_cpm/mastering_big_data_cfo_strategies_to_transform_insight_into_opportunity) 05 مايو 2017 على موقع واي باك مشين.
15. Jacobs, A. (6 July 2009). "The Pathologies of Big Data". *ACMQueue*. مؤرشف من الأصل في 20 أبريل 2019.

16. Magoulas, Roger; Lorica, Ben (فبراير 2009). "Introduction to Big Data". *Release 2.0*. (11 (Sebastopol CA: O'Reilly Media. مؤرشف من الأصل في 4 يونيو 2010.

17. Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet science. *International Journal of Internet Science, 7*, 1-5. http://www.ijis.net/ijis7_1/ijis7_1_editorial.html https://web.archive.org/web/20191123051001/http://www.ijis.net:80/ijis7_1/ijis7_1_editorial.html) نسخة محفوظة 2019-11-23 على موقع واي باك مشين.

18. Douglas, Laney. "3D Data Management: Controlling Data Volume, Velocity and Variety" (PDF). Gartner. مؤرشف من الأصل (PDF) في 5 يونيو 2019. اطلع عليه بتاريخ 06 فبراير 2001.

19. Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Gartner. مؤرشف من الأصل في 10 يوليو 2011. اطلع عليه بتاريخ 13 يوليو 2011.

20. Douglas, Laney. "The Importance of 'Big Data': A Definition". Gartner. مؤرشف من الأصل في 7 أبريل 2020. اطلع عليه بتاريخ 21 يونيو 2012.

21. Big social data analysis. E. Cambria, D. Rajagopal, D. Olsher, and D. Das. In: R. :2013 Akerkar (ed.) Big Data Computing, ch. 13, Taylor & Francis

22. "LHC Brochure, English version. A presentation of the largest and the most powerful particle accelerator in the world, the Large Hadron Collider (LHC), which started up in 2008. Its role, characteristics, technologies, etc. are explained for the general public". *CERN-Brochure-2010-006-Eng. LHC Brochure, English version*. CERN. مؤرشف من الأصل في 19 مارس 2019. اطلع عليه بتاريخ 20 يناير 2013.

23. "LHC Guide, English version. A collection of facts and figures about the Large Hadron Collider (LHC) in the form of questions and answers". *CERN-Brochure-2008-001-Eng. LHC Guide, English version*. CERN. مؤرشف من الأصل في 7 أبريل 2020. اطلع عليه بتاريخ 20 يناير 2013.

24. Brumfiel, Geoff (19 January 2011). "High-energy physics: Down the petabyte highway". *Nature*. **469**: 282–83. doi:10.1038/469282a. صفحات. مؤرشف من الأصل في 30 يوليو 2017.

25. Preis, Tobias; Moat,, Helen Susannah; Stanley, H. Eugene; Bishop, Steven R. (2012). "Quantifying the Advantage of Looking Forward". *Scientific Reports*. **2**: 350. doi:10.1038/srep00350. PMC 3320057. PMID 22482034.

26. Marks, Paul (April 5, 2012). "Online searches for future linked to economic success". *New Scientist*. مؤرشف من الأصل في 24 أبريل 2015. اطلع عليه بتاريخ April 9, 2012.

27. Johnston, Casey (April 6, 2012). "Google Trends reveals clues about the mentality of richer nations". *Ars Technica*. مؤرشف من الأصل في 8 مايو 2012. اطلع عليه بتاريخ April 9, 2012.

28. Tobias Preis (2012-05-24). "Supplementary Information: The Future Orientation Index is available for download" (PDF). مؤرشف من الأصل (PDF) في 6 يوليو 2017. اطلع عليه بتاريخ 24 مايو 2012.

29. Kalil, Tom. "Big Data is a Big Deal". White House. مؤرشف من الأصل في 10 يناير 2017. اطلع عليه بتاريخ 26 سبتمبر 2012.

30. Executive Office of the President (2012). "Big Data Across the Federal Government" (PDF). White House. مؤرشف من الأصل (PDF) في 11 ديسمبر 2016. اطلع عليه بتاريخ 26 سبتمبر 2012.

31. "How big data analysis helped President Obama defeat Romney in 2012 Elections". Bosmol Social Media News. 8 February 2013. مؤرشف من الأصل في 5 مارس 2016. اطلع عليه بتاريخ 09 مارس 2013.

32. Hoover, J. Nicholas. "Government's 10 Most Powerful Supercomputers". *Information Week*. UBM. مؤرشف من الأصل في 16 أكتوبر 2013. اطلع عليه بتاريخ 26 سبتمبر 2012.

33. Bamford, James. "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)". *Wired Magazine*. مؤرشف من الأصل في 17 مارس 2012. اطلع عليه بتاريخ 18 مارس 2013.

34. Groundbreaking Ceremony Held for $1.2 Billion Utah Data Center". National Security Agency Central Security Service. مؤرشف من الأصل في 16 أبريل 2016. اطلع عليه بتاريخ 18 مارس 2013.

35. Layton, Julia. "Amazon Technology". Money.howstuffworks.com. مؤرشف من الأصل في 18 مايو 2019. اطلع عليه بتاريخ 05 مارس 2013.

36. eBay Study: How to Build Trust and Improve the Shopping Experience"." Knowwpcarey.com. 2012-05-08. مؤرشف من الأصل في 12 أغسطس 2014. اطلع عليه بتاريخ 05 مارس 2013.

37. Predicting Commutes More Accurately for Would-Be Home Buyers - The New York Times (http://bits.blogs.nytimes.com/2013/03/12/predicting-commutes-more-accurately-for-would-be-home-buyers/) نسخة محفوظة (https://web.archive.org/web/20170706150116/https://bits.blogs.nytimes.com/2013/03/12/predicting-commutes-more-accurately-for-would-be-home-buyers/) 06 يوليو 2017 على موقع واي باك مشين.

38. UN GLobal Pulse (2012). Big Data for Development: Opportunities and Challenges (White paper by Letouzé, E.). New York: United Nations. Retrieved from http://www.unglobalpulse.org/projects/BigDataforDevelopment نسخة محفوظة (https://web.archive.org/web/20200601215207/https://www.unglobalpulse.org/document/big-data-for-development-opportunities-and-challenges-white-paper/) 2020-06-01 على موقع واي باك مشين.

39. WEF (World Economic Forum), & Vital Wave Consulting. (2012). Big Data, Big Impact: New Possibilities for International Development. World Economic Forum. Retrieved August 24, 2012, from http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development نسخة محفوظة (https://web.archive.org/web/20200601215225/https://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development/) 2020-06-01 على موقع واي باك مشين.

40. Webster, John. "MapReduce: Simplified Data Processing on Large Clusters" (http://research.google.com/archive/mapreduce-osdi04.pdf), "Search Storage", 2004. Retrieved on 25 March 2013. نسخة محفوظة (https://web.archive.org/web/20091214180106/http://research.google.com/archive/mapreduce-osdi04.pdf) 23 ديسمبر 2009 على موقع واي باك مشين.

41. Big Data Definition". MIKE2.0". مؤرشف من الأصل في 25 سبتمبر 2018. اطلع عليه بتاريخ 09 مارس 2013.

42. Manyika, James; Chui, Michael; Bughin, Jaques; Brown, Brad; Dobbs, Richard; Roxburgh, Charles; Byers, Angela Hung (May 2011). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute. مؤرشف من الأصل في 5 مارس 2013.

43. Future Directions in Tensor-Based Computation and Modeling" (PDF". مايو 2009). مؤرشف من الأصل (PDF) في 17 أبريل 2018.

44. Lu, Haiping; Plataniotis, K.N.; Venetsanopoulos, A.N. (2011). "A Survey of Multilinear Subspace Learning for Tensor Data" (PDF). *Pattern Recognition*. **44** (7): 1540–1551. doi:10.1016/j.patcog.2011.01.004. مؤرشف من الأصل (PDF) في 10 يوليو 2019.

45. Monash, Curt (30 April 2009). "eBay's two enormous data warehouses". مؤرشف من الأصل في 31 مارس 2019.

Monash, Curt (6 October 2010). "eBay followup — Greenplum out, Teradata> 10 petabytes, Hadoop has some value, and more". مؤرشف من الأصل في 31 مارس 2019.
46. "Storage area networks need not apply". CNET News (April 1, 2011). مؤرشف من الأصل في 18 أكتوبر 2013.
47. "How New Analytic Systems will Impact Storage". سبتمبر 2011. مؤرشف من الأصل في 4 أبريل 2016.
48. Obama Administration Unveils "Big Data" Initiative:Announces $200 Million In New R&D Investments" (PDF). The White House. مؤرشف من الأصل (PDF) في 29 نوفمبر 2016.
49. Timothy Hunter; Teodor Moldovan; Matei Zaharia; Justin Ma; Michael Franklin; Pieter Abbeel; Alexandre Bayen (أكتوبر 2011). *Scaling the Mobile Millennium System in the Cloud*. مؤرشف من الأصل في 31 مارس 2019.
50. David Patterson (5 December 2011). "Computer Scientists May Have What It Takes to Help Cure Cancer". The New York Times. مؤرشف من الأصل في 13 يوليو 2017.
51. "Secretary Chu Announces New Institute to Help Scientists Improve Massive Data Set Research on DOE Supercomputers". "energy.gov". مؤرشف من الأصل في 3 أبريل 2019.
52. "Governor Patrick announces new initiative to strengthen Massachusetts' position as a World leader in Big Data". Commonwealth of Massachusetts. مؤرشف من الأصل في 7 يناير 2015.
53. "Big Data @ CSAIL". Bigdata.csail.mit.edu. 2013-02-22. مؤرشف من الأصل في 3 مايو 2019. اطلع عليه بتاريخ 05 مارس 2013.
54. "Big Data Public Private Forum". Cordis.europa.eu. 2012-09-01. مؤرشف من الأصل في 3 أبريل 2019. اطلع عليه بتاريخ 05 مارس 2013.
55. Graham M. (2012). "Big data and the end of theory?". *The Guardian*. مؤرشف من الأصل في 24 يوليو 2013.
56. "Shah, Shvetank; Horne, Andrew; Capellá, Jaime; Good Data Won't Guarantee Good Decisions. Harvard Business Review". HBR.org. مؤرشف من الأصل في 1 أبريل 2016. اطلع عليه بتاريخ 08 سبتمبر 2012.
57. "Big Data for Development: From Information- to Knowledge Societies" (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145), Martin Hilbert (2013), SSRN Scholarly Paper No. ID 2205145). Rochester, NY: Social Science Research Network; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145 (https://web.archive.org/web/20200407070054/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145) نسخة محفوظة 7 أبريل 2020 على موقع واي باك مشين.
58. Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired Magazine, (Science: Discoveries). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (https://web.archive.org/web/20140327145018/http://www.wired.com/science/discoveries/magazine/16-07/pb_theory/) نسخة محفوظة 2014-03-27 على موقع واي باك مشين.
59. Ohm, Paul. "Don't Build a Database of Ruin". Harvard Business Review. مؤرشف من الأصل في 2 سبتمبر 2013.
60. Danah Boyd (2010-04-29). "Privacy and Publicity in the Context of Big Data". *WWW 2010 conference*. مؤرشف من الأصل في 22 أكتوبر 2018. اطلع عليه بتاريخ 18 أبريل 2011.
61. Jones, MB; Schildhauer, MP; Reichman, OJ; Bowers, S (2006). "The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere" (PDF). *Annual Review of Ecology, Evolution, and Systematics*. **37** (1): 519–544. doi:10.1146/annurev.ecolsys.37.091305.110031. مؤرشف من الأصل (PDF) في 8 يوليو 2019.

مجلوبة من "https://ar.wikipedia.org/w/index.php?title=بيانات_ضخمة&oldid=49934888"

آخر تعديل لهذه الصفحة كان يوم 20 أغسطس 2020، الساعة 20:46.

النصوص منشورة برخصة المشاع الإبداعي. طالع شروط الاستخدام للتفاصيل.