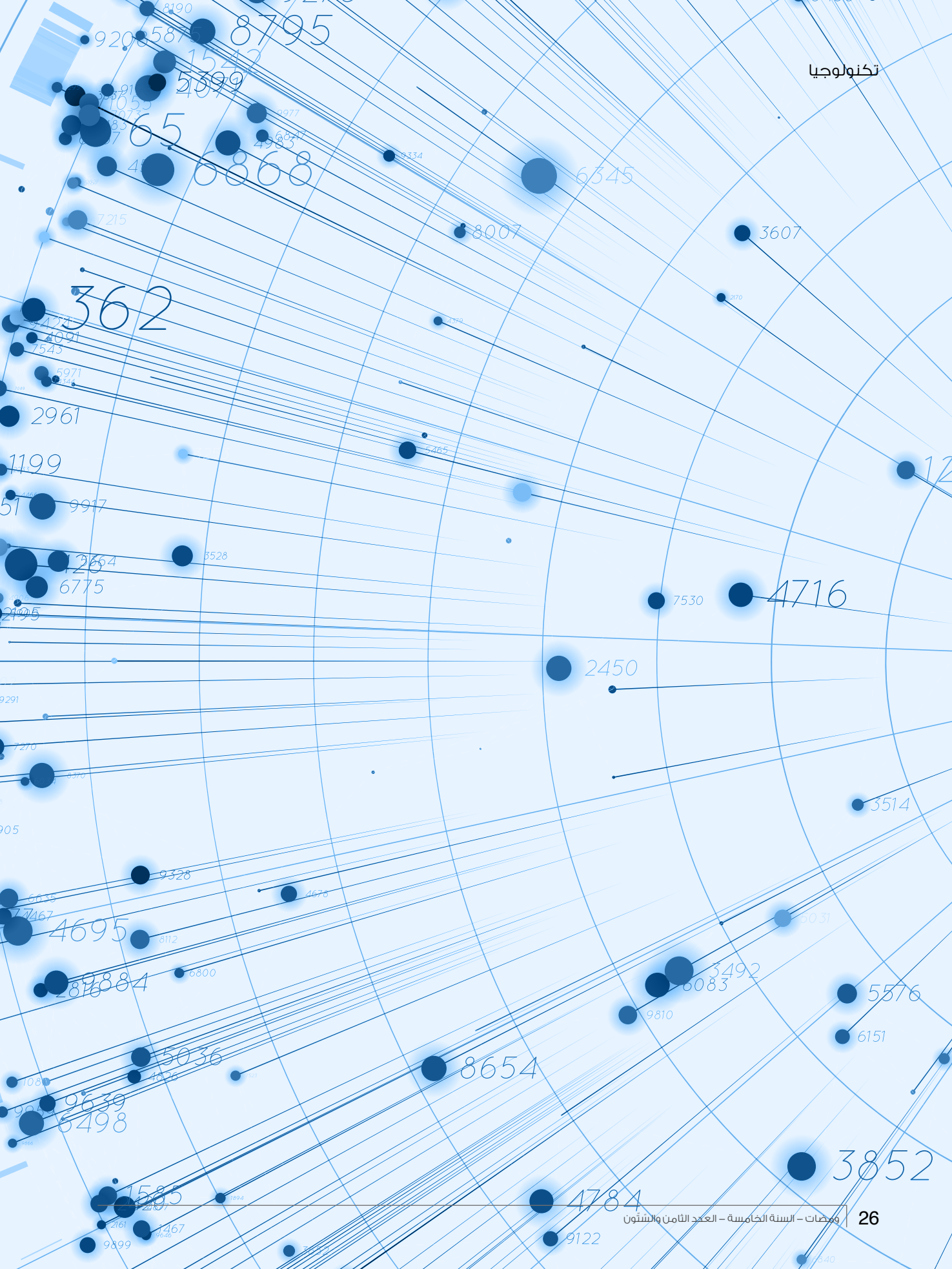




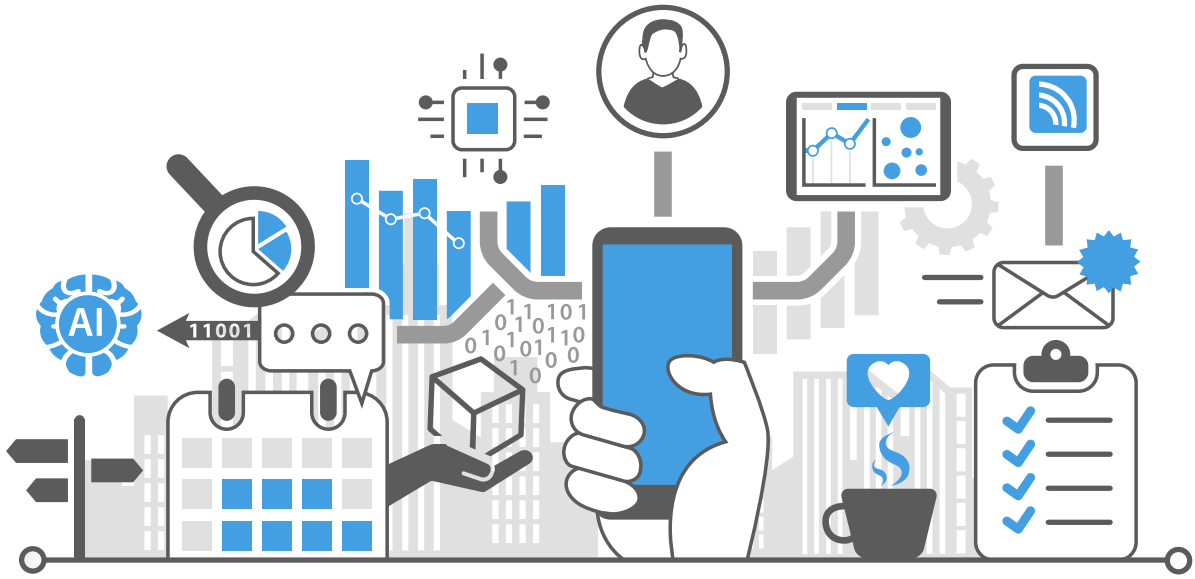
محطات من تاريخ البيانات الضخمة

أ.د. خالد عبد الفتاح محمد*

ازداد الاهتمام في الآونة الأخيرة بالمفاهيم المرتبطة بالبيانات مع النمو الكبير في حجم البيانات المتاحة نتيجة للثورة الرقمية وانتشار تطبيقات الإنترنت في كافة مناحي الحياة. ويُرجع المهتمون بالثورة الرقمية بدايات الانتشار الكبير لعالم البيانات الرقمية إلى ظهور الشبكة العنكبوتية العالمية على يد تيم بيرنر لي في بدايات التسعينيات (1991) من القرن الماضي. وقد أدى انتشار تطبيقات الويب التي تعتمد على لغة تكويد النصوص الفائقة، إلى انتشار مفهوم البيانات المترابطة Linked Data ثم تطوير لغات التكويد، فظهرت لغة التكويد الموسعة XML التي حققت مفهوم البيانات الدلالية Sematic Data.

دراسة هامة بعنوان «ما هو كم المعلومات الموجودة في العالم الآن؟» How Much Information is there in the World. وقد أشار إلى أن التقديرات النظرية تؤكد وجود أكثر من 12 ألف بيتابايت من البيانات على الويب. كما أوضح أن حجم الويب يزداد في مراحلها المبكرة ويتضاعف بمعدل 10 مرات سنوياً. وأن محرك البحث جوجل Google يسعى خلال العشرين عاماً القادمة على الأقل إلى أن يصبح الأداة الرئيسة للبحث عن كافة أنواع البيانات على الإنترنت.

وقد أشار تروسكويسكي وماريوس Truskowski & (Morris, 2003) في كتابهما عن تطور أنظمة التخزين إلى أن عام 1996 هو نقطة التحول الرئيسة في هذا المجال، حيث أصبح التخزين الرقمي للبيانات في هذا العام أكثر فاعلية من جانب التكلفة، حيث أصبحت تكاليف التخزين الرقمي «منذ عام 1996» أقل تكلفة من التخزين الورقي، وقد صاحب ذلك بدايات انتشار المكتبات والأرشفات الرقمية. في عام 1997 نشر مايكل ليسك



بدايات البيانات الضخمة

Peter Lyman and Hal Varian وهما من كبار المحللين الاقتصاديين في شركة جوجل تحديد كم المعلومات الرقمية المتاحة في العالم الآن في دراسة بعنوان «How Much Information» ومعدل نموها السنوي. وخلصا إلى أن إجمالي الإنتاج السنوي من المطبوعات والأفلام والمحتوى المسموع والبصري يتطلب حوالي 1,5 مليار جيجابايت من التخزين أي ما يعادل 250 ميجابايت للشخص الواحد على وجه الأرض.

وفى عام 2001 نشر دوج لاني Doug Laney المحلل بوكالة جارتر Gartner مقالة بعنوان «الأبعاد الثلاثة لإدارة البيانات: الحجم والسرعة والتنوع» 3D Data Management: Controlling Data Volume, Velocity and Variety. وقد حدد دوج في المقالة الخصائص الثلاث التي تميز البيانات الضخمة والمتمثلة في الحجم وسرعة معالجتها ومدى التنوع في هذه البيانات وخاصة البيانات غير المهيكلة.

وقد شهدت نشأة الجيل الثاني للويب Web 2.0 كثافة كبرى في عدد موردي الخدمات، لأن الاستفادة يسهم فيها بالغالبية العظمى من البيانات والمعلومات والمحتوى المتاح على الويب

ظهر مصطلح البيانات الضخمة لأول مرة في عام 1999 في إطار دراسة عن الوقت الحقيقي لاستكشاف عرض مرئي لمجموعات بيانات حجمها جيجا بايت Visually Exploring Gigabyte Datasets in Real Time والتي نشرتها جمعية الحواسيب الآلية. وأشارت الدراسة إلى أن الأهمية الحقيقية تكمن في تحليل البيانات وليس تخزينها فقط، حيث إن الغرض من الحوسبة هو التعامل مع البيانات ببصيرة وتحليل عميق وليس على أنها مجرد أرقام. وقد صاحب ظهور مصطلح البيانات الضخمة أول استخدام لمصطلح إنترنت الأشياء «Internet of Things-IoT»، لوصف العدد المتزايد من الأجهزة عبر الإنترنت وإمكانية اتصال بعضها ببعض، دون وجود «وسيط بشري». وتم استخدام المصطلح كعنوان لعرض تقديمي مقدم لشركة Procter and Gamble من قبل رائد تكنولوجيا تحديد الهوية باستخدام ترددات الراديو RFID كيفين أشتون Kevin Ashton الذي وضع معايير استخدام ترددات الراديو بمعهد ماساشوستس للتكنولوجيا Massachusetts Institute of Technology MIT.

في عام 2000 حاول بيتر ليمن وهال فارين



من خلال الخدمات التفاعلية. فعلى سبيل المثال يستخدم ملايين الأشخاص مواقع التواصل الاجتماعي مثل فيسبوك وتويتر في تحميل ومشاركة البيانات وتفاعلاتهم على تلك المنصات.

وقد شهد نفس العام إنشاء هادوب Hadoop وهي عبارة عن منصة مفتوحة المصدر تم تطويرها خصيصاً لتخزين وتحليل مجموعات البيانات الضخمة. وقد ساعدت مرونة تلك المنصة على تيسير التعامل مع البيانات غير المهيكلة.

المحمولة لأول مرة في نقل وتبادل البيانات الرقمية بكثافة تفوق استخدام أجهزة الكمبيوتر المحمولة والمنزلية. وأشار استطلاع لرأي المديرين التنفيذيين للشركات الكبرى أعدته شركة EG Accenture إلى أن 88 % منهم يفضلون التحليلات التي تعتمد على البيانات الضخمة وأن تلك البيانات تمثل أهمية قصوى لأعمالهم.

وقد تنبأت مؤسسة ماكينزي أنه بحلول عام 2018 ستواجه الولايات المتحدة الأمريكية عجزاً كبيراً يتراوح بين 140 ألف إلى 190 ألف في علماء البيانات المحترفين. كما أشار التقرير إلى تفاقم مشكلات خصوصية وأمن البيانات والملكية الفكرية التي ستطلب حلول لتحقيق الاستفادة القصوى من البيانات الضخمة.

في عام 2009 أصدرت مؤسسة ماكينزي العالمية تقريراً بعنوان The Next Frontier for Innovation, Competition and Productivity by McKinsey Global Institute، أشارت فيه إلى أن الشركة الأمريكية متوسطة الحجم التي يبلغ عدد موظفيها 1000 تقوم بتخزين أكثر من 200 تيرابايت من البيانات. كما أشار أريك شميدت Eric Schmidt المدير التنفيذي لشركة جوجل في أحد المؤتمرات عام 2010 إلى أن حجم البيانات التي يتم إنتاجها كل يومين يعادل حجم ما أنتجته الحضارة البشرية منذ بدايتها حتى عام 2003. وأشارت التنبؤات أنه بحلول عام 2020 سينخفض ذلك من يومين ليوم واحد.

من ثم يمكن القول إن البيانات الضخمة ليست ظاهرة جديدة أو معزولة، ولكنها جزء من تطور طويل لتجميع واستخدام البيانات. مثله مثل التطورات الرئيسية الأخرى في تخزين ومعالجة البيانات والإنترنت. تعد البيانات الضخمة مجرد خطوة أخرى ستحدث التغيير في الطريقة التي تدار بها الأعمال وآليات التواصل، كما ستضع الأسس التي سيتم بناء عليها العديد من التطورات في المستقبل.

في عام 2010 تم نشر تقرير بعنوان «How Much Information? 2010 report» أشار إلى وجود 9.57 زيتابايت (9.57 Zettabytes) أي ما يعادل (9.57 trillion gigabytes) بمعدل 12 جيجابايت لكل فرد كل يوم. وقد تنبأ التقرير بإنتاج حوالي 14.7 Exabytes من المعلومات الجديدة هذا العام.

في عام 2014 بدأ استخدام تطبيقات الأجهزة

الأهمية الحقيقية تكمن في تحليل البيانات عن طريق تحليلها ودراسة أبعادها بعمق وبصيرة.

* مدير مركز المعرفة الرقمية بمؤسسة محمد بن راشد آل مكتوم للمعرفة