



# Big Data

## มวลมหาข้อมูลที่หมุนอยู่รอบกระแสนวัตกรรม

วิษณุคุรุ์ เมาระพงษ์

กับริชชาโครงการสารสนเทศของทบวงมหาวิทยาลัย

สภาคณะสถาปัตยกรรมศาสตร์วิศวกรรมศาสตร์

แห่งมหาวิทยาลัยธรรมศาสตร์

ต่อ จากฉบับที่แล้ว

unk ความในตอนที่ผมขอกล่าวถึงแนวคิดและเทคโนโลยีซึ่งเป็น ที่จับตามองมากที่สุดในปี 2014 นั่นคือ “Big Data” แปลเป็น ไทยได้แบบตรงตัวว่า “ข้อมูลขนาดใหญ่” ซึ่งในความเป็นจริงตาม กรอบแนวคิดของเทคโนโลยีนี้แล้ว ก็ไม่ค่อยถูกต้องเท่าไรนัก ต้องให้ นิยามมันว่าเป็น “ข้อมูลปริมาณมหาศาล” ที่อาจไม่สามารถคาดคะเน ได้ว่าจะเติบโต และเพิ่มปริมาณเป็นเท่าไรในภาพรวมทั้งโลก ไม่ว่าจะ อาจะทะลุหน่วยวัดที่เคยกำหนดไว้ตามมาตรฐานสากลภายในไม่ถึง 10 ปีข้างหน้าก็เป็นได้



Big Data ที่ว่านี้ ก็คือข้อมูลที่เกิดขึ้นตั้งแต่ในช่วงที่มีการพัฒนา แนวคิด และเทคโนโลยี Version 2.0 ของ Web หรือที่เรียกว่า “Web 2.0” โดยแต่เดิมข้อมูลต่างๆ จะถูกสร้างขึ้น และบริหารจัดการโดย ระบบสารสนเทศเป็นอิสระจากกัน มีรูปแบบการจัดเก็บที่ต้องกำหนด ชัดเจน และทำงานร่วมกับระบบฐานข้อมูล และอาศัยการนำเข้า-ส่ง ออกข้อมูลเพื่อนำไปใช้งานในระบบสารสนเทศอื่น โดยการสร้างข้อมูล ก็ถูกจำกัดเฉพาะแค่จากงานที่ระบบสารสนเทศทำหรือให้บริการ

เท่านั้น ต่อมามีการพัฒนาช่องทาง การเข้าถึงบริการของผู้ใช้งานใน รูปแบบของ Web Application ที่ใช้ช่องทาง การเชื่อมต่อผ่านเครือข่าย สาธารณะ (Internet) ทำให้ผู้ใช้บริการ มีความสะดวกยิ่งขึ้น ประกอบ กับการริเริ่มแนวคิดของการบูรณาการข้อมูลระหว่างระบบสารสนเทศ เกิดเทคโนโลยี Web Services และแนวคิด SOA (Service-Oriented Architecture) ขึ้น ส่งผลให้เกิดการปรับเปลี่ยนรูปแบบการแสดงผล และใช้งานของระบบสารสนเทศส่วนใหญ่ให้เป็น Web Application ทั้งนี้เพื่อเป็นการอำนวยความสะดวก และลดข้อผิดพลาดในการ ทำงานร่วมกันระหว่างระบบ เพราะจะเป็นการทำงานร่วมกันแบบ อัตโนมัติไม่ต้องมี User ที่เป็นมนุษย์เข้าไปแทรกแซงกระบวนการ ในฐานของผู้ปฏิบัติ แต่จะรับบทบาททำกับการทำงานของระบบ แทน โดยการปรับแต่งค่า Configuration ตามที่ต้องการของกระบวนการ ทำงาน และใช้ข้อมูลร่วมกัน

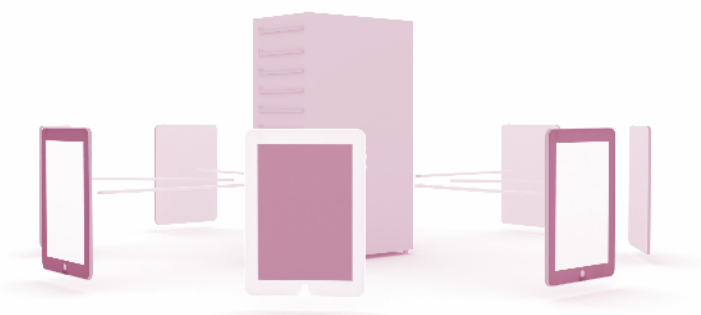
อย่างไรก็ตามการบูรณาการเชื่อมโยงข้อมูลระหว่างระบบ สาธารณเทคโนโลยีที่ใช้อัตโนมัติของการทำงานร่วมกัน และการให้บริการ ผ่าน Web Application นั้น ก็ต้องมีกำหนดรูปแบบมาตรฐานเพื่อ สร้างความเข้าใจในระบบสารสนเทศที่มีความแตกต่าง และหลากหลาย ในแง่ของเทคโนโลยีที่ใช้ในการพัฒนา สามารถทำการสร้าง จัดเก็บ และ นำข้อมูลที่ถูกสร้างขึ้นไปใช้งานได้อย่างเหมาะสม เข้าใจในเนื้อหาของ ข้อมูลได้ถูกต้อง จึงเกิดเป็นมาตรฐานของเทคโนโลยีที่เกี่ยวข้องขึ้นมา อย่างมากมาย ซึ่งรวมถึงมาตรฐานในการพัฒนารูปแบบของการ แสดงผลข้อมูลสารสนเทศแบบ Web Application ซึ่งถูกเรียกใช้งาน ผ่าน Web Browser อาทิ XML (Extensible Markup Language) JSON (JavaScript Object Notation) Ajax (Asynchronous JavaScript and XML) HTML5 Java Framework รุ่นใหม่ๆ JQuery เป็นต้น

จากการกำหนดมาตรฐาน และพัฒนาเทคโนโลยีที่เกี่ยวข้อง ในยุค Web 2.0 เกิดการนำไปประยุกต์ใช้งานอย่างต่อเนื่อง และแพร่ หลาย ส่งผลให้หัวใจสำคัญที่เป็นผลลัพธ์จากเทคโนโลยี Web 2.0 เกิด ขึ้น นั่นคือ Social Network ซึ่งข้อมูลไม่ได้ถูกสร้างขึ้นจากการใช้งาน ระบบสารสนเทศเพื่อการทำงาน หรือให้บริการเฉพาะเรื่องเพียงเท่านั้น ต้นกำเนิดของข้อมูลจะมาจากผู้ใช้งานที่เข้าใช้งาน Web Application บน Social Network โดยปริมาณข้อมูลจะเติบโตขึ้นอย่างรวดเร็ว และ

เป็นข้อมูลเฉพาะส่วนบุคคล ที่แปรผันตรงกับการใช้งาน Social Network ของบุคคลนั้นๆ สำหรับ Mobile Technology ที่มีการพัฒนาแบบก้าวกระโดดก็กลายมาเป็นตัวคูณของจำนวนข้อมูลที่เกิดขึ้นจาก Social Network ซึ่งผู้ใช้ ใช้งานผ่านเครื่อง PC Laptop โดย Social Network Application ถูก Port ไปสู่การใช้งานบน Mobile Device ที่เข้าถึงได้ง่ายใช้งานสะดวก ผมขอเรียกว่าเกิดเป็น “วัฒนธรรมนิ้วสัมผัส” (Finger Touch Culture) ซึ่งก็ไม่ได้หยุดอยู่เพียงแค่ Social Network Application เท่านั้น เนื่องจากการเติบโตของตลาด Mobile Device และ Mobile Application Software อื่นๆ ก็ทยอย Port สู่ Mobile แทบทั้งหมด ทำให้ความถี่ในการสร้างข้อมูลเพิ่มขึ้น ปริมาณข้อมูลก็เพิ่มขึ้นเป็นเงาตามตัว

ด้วยเหตุดังกล่าว ข้อมูลปริมาณมหาศาล ที่เติบโตอย่างต่อเนื่อง รวดเร็ว และยังเป็นข้อมูลที่กระจายไปแล้วสะท้อนพฤติกรรมของผู้ใช้งาน หรือผู้บริโภค ในเกือบทุกแง่มุม อาจเป็นประโยชน์อย่างยิ่งในมุมมองธุรกิจ และการตลาด หากมีการนำมาวิเคราะห์อย่างเหมาะสม ซึ่งนำไปสู่คำถามว่า แล้วจะวิเคราะห์ได้อย่างไร ก็ข้อมูลมีปริมาณมหาศาลอย่างนี้ แตกรูปแบบการจับเก็บยังหลากหลายไม่ได้เป็นไปตามเงื่อนไขการจัดเก็บของจำพวกระบบฐานข้อมูลด้วยซ้ำ ทำให้การนำมาประมวลผล ปรับรูปแบบ จัดเก็บ และ Query ข้อมูลมีความยุ่งยาก และซับซ้อน ด้วยธรรมชาติของการออกแบบข้อมูลสารสนเทศที่ใช้ในระบบ Web Application จะเป็นในลักษณะของ File System ขนาดเล็กที่อาศัยเทคโนโลยีพวก XML Standard ซึ่งใช้พื้นที่น้อยไม่เปลืองหน่วยประมวลผล (CPU) และง่ายต่อการเชื่อมโยงระหว่างระบบหรือในรูปแบบฐานข้อมูลที่มีความยืดหยุ่นสูงลดการสร้างเชื่อมโยง (Relational) ภายใต้วงเนื้อข้อมูลในแต่ละตาราง หรือไม่มีเลย ซึ่งปัจจุบันถูกเรียกว่าระบบข้อมูลแบบ NoSQL สำหรับคำตอบของทรัพยากรที่ใช้ในการประมวลผล Big Data หลายท่านคงเดาออก นั่นคือ เทคโนโลยี Cloud Computing ซึ่งเป็นเทคโนโลยีที่สนับสนุนให้การนำเอา Big Data มาใช้ประโยชน์เกิดขึ้นได้อย่างเป็นรูปธรรม

ก่อนอื่นเรามาทำความรู้จัก Big Data ให้มากขึ้นตามที่นิยามไว้ในระดับสากล เริ่มที่ Gartner ได้นิยาม Big Data ที่เกิดขึ้นไว้ใน 3



รูปแบบ ได้แก่ Volume (ปริมาณ) Velocity (ความเร็วในการนำข้อมูลเข้าออก) และ Variety (ข้อมูลหลากหลายประเภท) โดย

1) Volume คือ การที่มีจำนวนข้อมูลมากเกินกว่าที่ระบบฐานข้อมูลในรูปแบบเดิมๆ จะสามารถจัดการได้

ตัวอย่างของ Big Data ในรูปแบบ Volume นั้นได้แก่ Project ทางดาราศาสตร์ชื่อ GALEX ซึ่งถ่ายภาพของท้องฟ้าทั้งหมด 360 องศา โดยใช้กล้องโทรทรรศน์ดูดาว ซึ่งนักวิจัยจะนำภาพถ่ายเหล่านั้นมาวิเคราะห์จำแนกประเภทของดวงดาวที่ถ่ายภาพมาได้ ซึ่งภาพถ่ายท้องฟ้าที่มีความละเอียด 2,400 Pixel ต่อหนึ่งองศา คำนวณคร่าวๆ ได้ว่าภาพถ่ายทรงกลมของท้องฟ้าจะมีจำนวน Pixel ประมาณ 238 GigaPixel นั่นคือต่อหนึ่งภาพเท่านั้น ขนาดข้อมูลของภาพถ่ายทั้งหมดของ Project GALEX รวมกันประมาณ 20 TeraByte ซึ่งคือข้อมูลที่เริ่มจัดเก็บตั้งแต่ปี 2003 ปัจจุบัน 10 ปีผ่านไป ปัจจุบัน Project GALEX ได้เปลี่ยนไปใช้กล้องโทรทรรศน์ดูดาวรุ่นใหม่ซึ่งจะผลิตข้อมูลขนาดมหาศาลถึงหลัก PentaByte โดย 1 PentaByte ขนาดเท่ากับ 1 ล้าน GigaByte

Big Data ในรูปแบบ Volume อีกตัวอย่างหนึ่ง คือ จำนวนจุดข้อมูลที่มหาศาล ตัวอย่างเช่น จำนวนการ Tweet ที่ผู้ใช้งาน Tweeter ส่งถึงกัน ใน 1 สัปดาห์ อยู่ที่ประมาณ 1,000 ล้าน Tweet โดยถ้าเราตั้งวิเคราะห์จำนวนครั้งเฉลี่ยที่การ Tweet 1 ครั้ง ถูก Retweet ต่อๆ กันไป ก็ต้องวิเคราะห์จาก 1,000 ล้าน Tweet นี้โดยปกติถ้าจะวิเคราะห์ Big Data คงต้องทำในศูนย์วิจัยด้วย Super Computer ขนาดใหญ่เท่านั้น แต่ในปัจจุบันความก้าวหน้าของเทคโนโลยีสารสนเทศและการสื่อสาร ได้สร้างความเป็นไปได้ในการนำเอา Big Data มาวิเคราะห์ให้กับองค์กรต่างๆ รวมถึงผู้ใช้งานทั่วไป

ตัวอย่างการใช้ประโยชน์จาก Big Data แบบ Volume

- ➡ เปลี่ยนข้อมูล 12 Terabyte จากการ Tweet ในแต่ละวัน ให้เป็นการวิเคราะห์ความเชื่อมั่นของผลิตภัณฑ์ที่ดีขึ้น
- ➡ แปลงข้อมูลจากการอ่าน Meter ไฟฟ้าประจำทุกปี ประมาณ 350 พันล้านครั้ง ให้เป็นข้อมูลการคาดการณ์การใช้พลังงานไฟฟ้าที่ดีกว่าเดิม

