



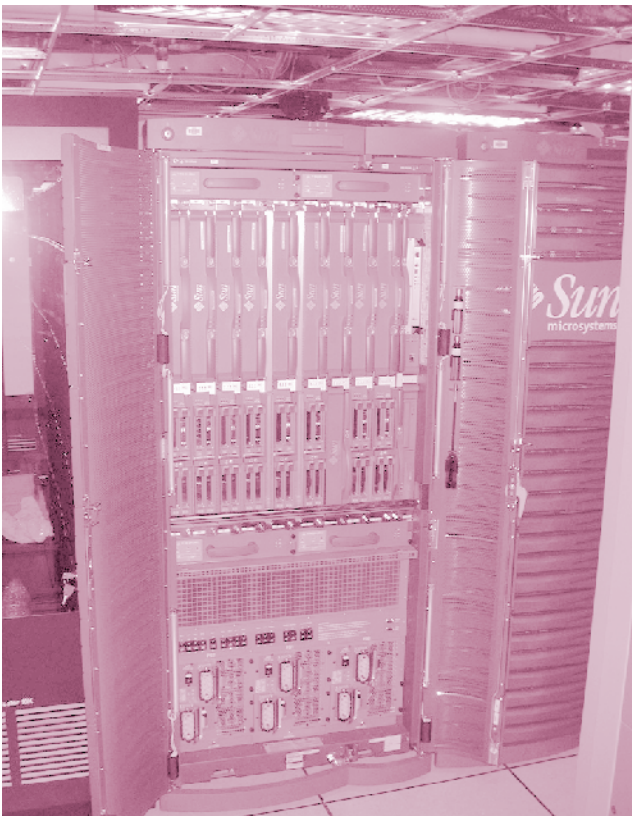
# Big Data

## มวลมหาข้อมูลที่หมุนอยู่รอบกระแสนวัตกรรม

วิษณุฤทธิ์ เมาระพงษ์

ที่ปรึกษาโครงการสารสนเทศของหน่วยงานภาครัฐ  
สังกัดสถาบันวิจัยเทคโนโลยีภาครัฐ  
แห่งมหาวิทยาลัยธรรมศาสตร์

ต่อ จากฉบับที่แล้ว



**ฉบับ**ที่แล้วกล่าวถึงการนำเอา Big Data มาวิเคราะห์ ไม่ใช่เรื่องที่น่าแปลกใจจากตัวเรา และองค์กรของเรา เนื่องด้วยปัจจัยต่างๆ ที่มีผลสำคัญทำให้องค์กรธุรกิจสามารถมองเห็นความเป็นไปได้ที่จะใช้ประโยชน์จากการนำเอา Big Data มาวิเคราะห์ รวมถึงไปถึงปัจจัยและความเป็นไปได้ในการนำเอา Big Data มาวิเคราะห์ และนำไปใช้ประโยชน์ จึงมีองค์กรชั้นนำริเริ่มดำเนินการในเรื่องดังกล่าว และประสบผลสำเร็จ ได้ยกตัวอย่างไปแล้ว 2 องค์กร ฉบับนี้ติดตามกันต่อกับองค์กรที่นำเอา Big Data มาวิเคราะห์ และนำไปใช้ประโยชน์

➡ **Sysomos** ได้ทำการวิเคราะห์พฤติกรรมของผู้ใช้งาน Tweeter ในการ Retweet และตอบกลับบน Tweeter โดยวิเคราะห์

จาก Tweet จำนวน 1.2 พันล้าน Tweet ในปี 2010 พบว่า เมื่อเจ้าของ Tweet ได้ส่งข้อความแล้วไม่มีการตอบกลับคิดเป็น 71% มีการนำไป Retweet ต่อคิดเป็น 6% และมีการตอบกลับ Tweet (@reply) จำนวน 21% เมื่อวิเคราะห์การ Retweet นั้นพบว่า 92% เกิดขึ้นภายในชั่วโมงแรกหลังจากการส่งข้อความ ส่วนการตอบกลับนั้นเกิดขึ้นถึง 96% ในชั่วโมงแรก นั้นหมายความว่า ถ้าเรา Tweet ข้อความไปแล้วแต่ไม่มีการตอบกลับหรือ Retweet ในหนึ่งชั่วโมง โอกาสที่จะมีการตอบกลับหรือ Tweet หลังจากนั้นจะน้อยมาก ซึ่งสามารถนำผลวิเคราะห์ไปประยุกต์ใช้ในเชิงการตลาดได้ในบางกรณีหรือกับ Application อื่นๆ ในทำนองเดียวกัน

➡ **Amazon** บริษัท e-Commerce ข้ามชาติรายใหญ่ ใช้ผลการวิเคราะห์ Big Data ในการแนะนำสินค้าอย่างอัตโนมัติ (Recommendation Engine) โดยอาศัยสถิติของการเข้าชม และเลือกซื้อสินค้าของผู้ใช้งาน เพื่อแนะนำสินค้าใหม่ๆ ซึ่งนำไปสู่การเพิ่มยอดขาย ถือเป็นนวัตกรรมแรกๆ ที่อาศัย Big Data เพื่อประโยชน์ทางธุรกิจ ต่อมาเทคนิคในรูปแบบของ Amazon ได้มีการศึกษาอย่างกว้างขวาง และได้แตกแขนงออกเป็นการทำ Analytics (การวิเคราะห์) ขั้นสูง เช่น Click Stream Analysis, Collective Intelligence, Collaborative Filtering เป็นต้น

➡ **Google** มี Big Data ที่สำคัญนั่นคือ สถิติการเข้าชม Web Site ต่างๆ และการใช้ Search Engine ของลูกค้า เพื่อที่สามารถทำ Analytics และเลือกนำเสนอโฆษณาที่ถูกต้อง และเหมาะสมให้กับลูกค้าได้ ในชื่อของ Google AdChoice นอกจากนี้ Google ยังอาจสามารถใช้ข้อมูลอื่นๆ เช่น ข้อความที่ลูกค้าเขียนใน Gmail เพื่อเรียนรู้ถึงความต้องการของลูกค้าได้ นอกจากนี้ Google ยังได้มีทีมงานทางด้าน Big Data และ Analytics ที่ใหญ่ที่สุดของโลก โดยมีผู้เชี่ยวชาญระดับปริญญาเอกเป็นจำนวนมาก ซึ่งจะว่าไปแล้วเครื่องมือแรกเริ่มของ Cloud Computing หรือ Grid Computing นั้น ถือเป็นนวัตกรรมที่พัฒนาขึ้นของ Google อาทิ MapReduce, Google File System แม้กระทั่ง Google Analytics ที่เป็นที่มาของการเริ่มใช้คำว่า Analytics อย่างแพร่หลายในปัจจุบัน

➡ **Netflix Prize** นำ Big Data และ Analytics มาประยุกต์



ใช้ในธุรกิจออนไลน์ โดย Netflix จะมอบรางวัลจำนวนหนึ่งล้านดอลลาร์ ให้กับผู้ที่สามารถประยุกต์ใช้ Big Data ของ Netflix ในการสร้างรูปแบบของ Analytics ที่สามารถเพิ่มยอดขายให้กับ Netflix ได้อย่างน้อย 10% ในการนำเสนอสินค้าประเภทวิดีโอที่ลูกค้าต้องการ

► **คณะวิจัยของ MIT Media Lab** ได้ทำการวิเคราะห์ข้อมูลสถานที่จาก Location-Based Services ใน Mobile Device ของผู้ที่กำลังเดินทางเข้าสู่ที่จอดรถของห้างสรรพสินค้าขนาดใหญ่แห่งหนึ่งในสหรัฐอเมริกา ก่อนที่ห้างสรรพสินค้าจะเปิดทำการในเทศกาลลดราคาที่ใหญ่ที่สุดของปี (Black Friday) จากข้อมูลนี้ ทำให้คณะวิจัยสามารถคำนวณยอดขายที่จะเกิดขึ้นได้อย่างแม่นยำ และสามารถคำนวณถึงกำไรที่จะเกิดขึ้น และโอกาสที่จะเข้าซื้อขายเพื่อทำกำไรในตลาดหุ้น กรณีนี้ถือเป็นตัวอย่งในการเอาชนะตลาดด้วย Big Data และ Analytics ที่ทำให้สามารถล่วงรู้ความเป็นไปของตลาดด้วยการคำนวณ ก่อนที่ความเป็นจริงจะเกิดขึ้น

► **โรงพยาบาลกรุงเทพ** ได้มีการนำเทคโนโลยีการวิเคราะห์ข้อมูลมาใช้ในการดูแลลูกค้า และบริหารจัดการโรงพยาบาล ทำให้ลูกค้า และผู้ป่วยได้รับบริการที่ตรงตามความต้องการเฉพาะบุคคลมากขึ้น สามารถวิเคราะห์แนวโน้มของสุขภาพ และวางแผนการรักษาได้อย่างเหมาะสม ตั้งแต่การป้องกันไปจนถึงการฟื้นฟูหลังการรักษา สำหรับในส่วนของการบริหาร ทำให้สามารถวางแผนการบริหารจัดการในด้านต่างๆ เช่น การลงทุน หรือการเตรียมบุคลากรในด้านต่างๆ และสร้างสรรค์บริการใหม่ๆ ที่เหมาะสมกับลูกค้ามากขึ้น

จากกรณีศึกษาข้างต้น การนำเอา Big Data มาวิเคราะห์ และใช้ประโยชน์ อาจแสดงให้เห็นว่าในอนาคตผลการวิเคราะห์ Big Data สามารถสร้างความได้เปรียบในการแข่งขันทางธุรกิจได้ แต่องค์กรต้องมีความเข้าใจแตกต่างระหว่าง Big Data กับระบบฐานข้อมูลที่ใช้สนับสนุนการทำธุรกิจในยุคปัจจุบัน เพราะ Big Data ไม่สามารถใช้ระบบฐานข้อมูลหรือเครื่องมือทางสารสนเทศโดยทั่วไปมาทำการวิเคราะห์ หรือ Analytics ได้ ตัวอย่างเช่น Big Data ของห้าง Walmart ในสหรัฐอเมริกา มีปริมาณของข้อมูลกว่า 2.5 Petabytes (2.5 x 10<sup>15</sup>

Byte) หรือใช้ Harddisk ขนาด 1 Terabyte จำนวน 2,500 ชิ้นในการจัดเก็บข้อมูลในเวลา 1 ชั่วโมง ข้อมูลปริมาณมหาศาลขนาดนี้ทำให้ความสามารถในการบริหารจัดการ Big Data และแม้กระทั่งการทำ Analytics จำเป็นต้องใช้เทคโนโลยีขั้นสูงในระดับของ Cloud Computing หรือ Grid Computing และ Software ที่มีความสามารถเฉพาะทาง ซึ่งแตกต่างจากการบริหารจัดการระบบฐานข้อมูลทั่วไป และเครื่องมือด้าน Software ที่ใช้ทำงานกับ Big Data ซึ่งเป็นที่นิยมขององค์กรชั้นนำในปัจจุบันซึ่งรู้จักกันในชื่อว่า "Hadoop"

**Hadoop เป็น Open Source Technology** ซึ่งจะทำหน้าที่เป็น Distributed Storage (ส่วนบริหารจัดการเก็บข้อมูลแบบกระจายตัว) ที่สามารถใช้ในการจัดเก็บข้อมูลขนาดใหญ่ที่เป็น Unstructure (ไม่มีรูปแบบที่ตายตัว) และนำมาประมวลผลได้ โดยองค์ประกอบหลักๆ ของ Hadoop จะประกอบด้วย Hadoop Distributed File System (HDFS) ที่ทำหน้าที่เป็น Storage และ MapReduce

MapReduce คือ กลไกการจับคู่ข้อมูลกับ Key ที่กำหนดในแต่ละ record หรือ Transaction และนำมาจัดกลุ่มอีกครั้งเพื่อกรองให้ได้ชุดข้อมูลที่ต้องการตามลำดับความสำคัญ เช่น สมมติว่าเรามีข้อมูลการจัดเก็บค่าอุณหภูมิที่วัดได้ ณ ดอยอย่างขาง จังหวัดเชียงใหม่เป็นระยะเวลา 20 ปี โดยวัดทุกๆ วัน ในทุกๆ 1 นาที โดย Thermometer และจัดเก็บข้อมูลอัตโนมัติ ผลลัพธ์ที่อยากได้คือ อุณหภูมิสูงสุดและต่ำสุดในแต่ละเดือนตลอดเวลา 20 ปี สิ่งที่ MapReduce จะทำก็คือ ขั้นตอนการ Map โดยกำหนด Key หลักเป็นปี และเดือน ส่วนข้อมูลที่จับคู่กับ Key คือ ค่าอุณหภูมิที่วัดได้ในแต่ละช่วงเวลานั้นๆ จับแยกออกมาเป็นชุดๆ จากนั้นก็ทำการ Reduce คือจัดกลุ่มของข้อมูลที่ยกแยกออกมาตาม ปี และเดือน และนำค่าอุณหภูมิมาจัดเรียงหรือหาค่าสูงสุด ต่ำสุดตามที่ต้องการ

**สำหรับโครงสร้างด้านทรัพยากรสารสนเทศที่สนับสนุนการทำงานของ Hadoop จะใช้ Cloud Server จำนวนมากต่อเป็น Cluster กัน (ตัวอย่างที่ Facebook ใช้)**

