



# Big Data

## มุมมองข้อมูลที่ถูกซ่อนอยู่รอบกระแสนวัตกรรม

วิษณุฤทธิ์ เมาระพงษ์

ที่ปรึกษาโครงการสารสนเทศของหน่วยงานภาครัฐ  
สังกัดสถาบันวิจัยเทคโนโลยีภาครัฐ  
แห่งมหาวิทยาลัยธรรมศาสตร์

ต่อ จากฉบับที่แล้ว



**ฉบับที่แล้ว** Gartner ได้นิยาม Big Data ที่เกิดขึ้นไว้ใน 3 รูปแบบ ได้แก่ Volume (ปริมาณ) Velocity (ความเร็วในการนำข้อมูลเข้าออก) และ Variety (ข้อมูลหลากหลายประเภท) โดยได้เกริ่นรูปแบบแรกไปแล้วนั่นคือ Volume (ปริมาณ) ฉบับนี้มาติดตามกันต่อ

**2. Velocity** คือ การที่ข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็ว เช่น ข้อมูลจาก Social Media ข้อมูลการซื้อขาย ข้อมูล Transaction การเงินหรือการใช้โทรศัพท์ หรือข้อมูลจากการใช้เทคโนโลยี Sensor ในบางกรณีเวลาแค่ 2 นาทีก็สายเกินไปสำหรับกระบวนการที่เวลาที่มีความสำคัญอย่างมาก อาทิ การจับการทุจริต หรือองค์การมีจำเป็นจะต้องใช้ big data ในทันทีที่ไหลผ่านเข้ามาเพื่อที่ใช้ประโยชน์จากข้อมูลเหล่านั้นให้มากที่สุด

ตัวอย่างการใช้ประโยชน์จาก Big Data แบบ Velocity

- การวิเคราะห์ และกั้นกรองเหตุการณ์ 5 ล้านเหตุการณ์ (Transaction) ที่ถูกสร้างขึ้นในแต่ละวันเพื่อตรวจหาการทุจริตที่อาจเกิดขึ้นในสถาบันการเงิน

- การวิเคราะห์ข้อมูลการใช้โทรศัพท์ 500 ล้านครั้งในแต่ละวันแบบ real-time เพื่อคาดการณ์พฤติกรรมของลูกค้าได้รวดเร็ว และแม่นยำยิ่งขึ้น

**3. Variety** คือ การที่ข้อมูลมีความหลากหลายของรูปแบบ ทั้ง Structure และ Unstructure ซึ่งอาจจะอยู่ในรูปทั้ง RDBMS text XML JSON Image หรือ Video Streaming

ตัวอย่างการใช้ประโยชน์จาก Big Data แบบ Variety

- ตรวจสอบภาพวิดีโอสดๆ จากกล้องวงจรปิด 100 กล้อง เพื่อหาสิ่งทีกลุ่มเป้าหมายสนใจ

- ใช้ข้อมูลที่เติบโตขึ้นทั้งที่เป็น Image Video และ Document จำนวน 80% มาปรับปรุงในเรื่องความพึงพอใจของลูกค้า

สำหรับในแง่มุมมองของการนำ Big Data มาวิเคราะห์เพื่อประเมินหรือคาดการณ์ตลาดนั้น จะเพิ่มรูปแบบที่ 4 (ซึ่งน่าจะเรียกว่าคุณลักษณะมากกว่า) เข้าไป นั่นคือ Veracity ซึ่งถ้าจะแปลให้มีความสอดคล้องกับมุมมองของนักการตลาด และผู้บริหารองค์กรที่มองการนำ Big Data มาวิเคราะห์ ก็ควรแปลว่า **“ข้อเท็จจริง” “ความจริงที่น่าเชื่อถือ”** กล่าวคือนอกจาก Big Data จะมีลักษณะของปริมาณข้อมูลที่มีมหาศาล เปลี่ยนแปลงรวดเร็ว และหลากหลายแล้ว ด้วยลักษณะทั้ง 3 ที่กล่าวมา ส่งผลให้เกิดความไม่เชื่อมั่นในคุณภาพของข้อมูลทีวิเคราะห์ได้จาก Big Data เนื่องจากแหล่งกำเนิดข้อมูลทำการสร้างข้อมูลในหลายลักษณะตามพฤติกรรม การเข้าถึง และใช้งาน Application ต่างๆ ในชีวิตประจำวัน หากมีการนำ Big Data มาวิเคราะห์ ประเมิน และแปลค่าข้อมูลในทันที อาจได้ผลลัพธ์ที่คาดเคลื่อนจากข้อเท็จจริง ควรต้องมีการจัดระเบียบข้อมูลเป็นกลุ่มตามความสำคัญ เพื่อสร้างรูปแบบการวิเคราะห์ที่เหมาะสม ซึ่งผ่านการคัดกรองข้อมูลในเบื้องต้นแล้ว ทั้งนี้ก็คงเป็นหน้าที่ของ Software และ Cloud Architecture ที่จะมามีบทบาทสำคัญในการประมวลผล Big Data

ในปัจจุบัน การนำเอา Big Data มาวิเคราะห์นั้น ไม่ใช่เรื่องที่น่าหวาดกลัวจากตัวเรา และองค์กรของเรา เนื่องด้วยปัจจัยต่างๆ ที่มีส่วนสำคัญทำให้องค์การธุรกิจสามารถมองเห็นความเป็นไปได้ที่จะใช้ประโยชน์จากการนำเอา Big Data มาวิเคราะห์ ซึ่งได้แก่

- **ราคาของอุปกรณ์จัดเก็บข้อมูล** อาทิ Harddisk USB Flash Drive Memory Card นั้น พัฒนาไปอย่างรวดเร็วจนราคาต่อ-

หน่วยพื้นที่จัดเก็บถูกลดอย่างมาก ทำให้องค์การขนาดกลางหรือแม้แต่มูลนิธิทั่วไปสามารถซื้อเข้ามาใช้เก็บข้อมูลขนาดใหญ่ได้

๖ **เทคโนโลยีการประมวลผลรูปแบบใหม่** อาทิ การประมวลผลแบบ Multi-core ซึ่งปัจจุบันสามารถพบเห็นได้ใน Computer ระดับใช้งานทั่วไป หรือแม้แต่ใน Mobile Device การประมวลผลแบบ Many-core อุปกรณ์ประมวลผลภาพ และ 3D Grapic ซึ่งสามารถให้ประสิทธิภาพการประมวลผลที่สูงมากในราคาที่ไม่แพงมากเกินไปนัก ทำให้องค์การโดยทั่วไปก็สามารถจัดหาเข้ามาใช้เพื่อการประมวลผลข้อมูลมหาศาลนี้ได้

๗ **เทคโนโลยี Cloud Computing** ที่เข้ามามีบทบาทในการเปลี่ยนแปลงรูปแบบการลงทุนด้านโครงสร้างพื้นฐานของเทคโนโลยีสารสนเทศ และการสื่อสาร ให้เป็นรูปแบบของการเช่าใช้งาน ทำให้บุคคลทั่วไปสามารถเข้าถึงทรัพยากรด้านเทคโนโลยีสารสนเทศ ทั้งในส่วนของการประมวลผลและหน่วยจัดเก็บข้อมูลขนาดใหญ่ได้โดยเสียค่าใช้จ่ายน้อยกว่าการลงทุนเพื่อจัดซื้อระบบเองเป็นอย่างมาก ซึ่งปัจจัยข้อนี้ถือว่าเป็นจุดเปลี่ยนสำคัญที่ทำให้ Big Data เป็นสิ่งที่เข้าถึงได้อย่างเป็นรูปธรรม

๘ **ระบบ Software ที่ใช้บริหารจัดการ และประมวลผล Big Data** ระบบสารสนเทศที่ใช้ในการบริหารจัดการ และประมวลผลข้อมูลปริมาณมหาศาลนี้ ได้ถูกพัฒนาอย่างต่อเนื่องจนถึงจุดที่สามารถนำออกมาผลิตเป็นผลิตภัณฑ์ให้ผู้ใช้งานทั่วไปได้ซื้อเข้ามาใช้งาน ซึ่งทำให้ผู้ใช้ไม่จำเป็นต้องพัฒนาเครื่องมือด้าน Software เอง โดย Software ที่นำมาใช้กับ Big Data อาทิ ฐานข้อมูลแบบ NoSQL เช่น MongoDB หรือเครื่องมือที่ใช้ในการกระจายการประมวลผลไปยัง Computer อื่นๆ เช่น MapReduce Hadoop เป็นต้น

๙ **ราคาของอุปกรณ์ Sensor ประเภทต่างๆ ถูกลง และมีรูปแบบที่หลากหลายขึ้น** Sensor ที่กล่าวถึง อาทิ Accelerometer หรือ Sensor ที่มีไว้สำหรับตรวจจับลักษณะการเคลื่อนไหวของ Mobile Device จำพวก Smart Phone โดยเป็นการตรวจจับแบบ 3 แกน (3-Axes) ประโยชน์ในการใช้งานที่เห็นกันอยู่เป็นประจำก็คือ การปรับทิศทางการแสดงผลหรือการใช้งานที่ต้องอาศัยการเอียงเครื่องไปในทิศทางต่างๆ เช่น ไม่ว่าเราเอียงเครื่องไปทางไหน หน้าจอก็จะปรับให้แสดงผลในทิศทางเดียวกันโดยอัตโนมัติ Sensor ตรวจจับอุณหภูมิ ความดัน แสง เสียง GPS และอื่นๆ Sensor เหล่านี้เป็นกลไกสำคัญที่นำไปสู่การผลิตข้อมูลในรูปแบบ Big Data



๑๐ **การเติบโตของ Social Network** คือปัจจัยที่ส่งผลให้ผู้ใช้งานทั่วไปมีการผลิตข้อมูลปริมาณมากขึ้น ซึ่งเริ่มต้นตั้งแต่มีการพัฒนา Web 2.0 โดยผู้ใช้ทั่วไปสามารถสร้าง Content และ Upload ขึ้นไปบน Web Application รวมถึงระบบ Social Networks ต่างๆ ที่ผู้ใช้ติดต่อ สื่อสาร ปฏิสัมพันธ์ กับผู้ใช้คนอื่นๆ ในหลากหลายรูปแบบ อาทิ การ Post ข้อความ การกด Like Status การ Retweet เป็นต้น ข้อมูลที่เกิดจากการสื่อสาร และได้ตอบกันระหว่างผู้ใช้นั้นมีปริมาณมหาศาล ซึ่งสามารถทำการจัดเก็บไว้เพื่อนำมาประมวลผลได้

*จากปัจจัย และความเป็นไปได้ในการนำเอา Big Data มาวิเคราะห์ และนำไปใช้ประโยชน์ จึงมีองค์การชั้นนำริเริ่มดำเนินการในเรื่องดังกล่าว ซึ่งประสบผลสำเร็จ อาทิ*

๑ **Nate Silver** ใช้ Big Data ในการคาดการณ์ผลการเลือกตั้งประธานาธิบดีสหรัฐอเมริกาเมื่อปี 2012 โดยรวบรวมข้อมูลจากผลสำรวจ (Poll) ก่อนการเลือกตั้งจากทุกรัฐในอเมริกา และนำมาประมวลผล เพื่อหาโอกาส และเปอร์เซ็นต์ความคลาดเคลื่อน ซึ่งการคาดการณ์ของ Nate Silver นั้น แม่นยำในทุกุณนั้นคือ คาดการณ์ได้ถูกต้องว่า บารัค โอบามาจะแพ้หรือชนะในรัฐไหนบ้าง และ Nate Silver ยังสรุปผลวิเคราะห์ได้อย่างถูกต้องว่าถ้าหาก บารัค โอบามาชนะการเลือกตั้งในรัฐโอไฮโอ ก็จะมีชนะการเลือกตั้งทั่วประเทศได้ขึ้นดำรงตำแหน่งประธานาธิบดีสหรัฐอเมริกา

๒ **Last.fm เป็นบริษัทที่ให้บริการเล่นเพลง On-line ผ่าน Internet** ได้ทำการวิเคราะห์สัดส่วนจำนวนครั้งในการเล่นเพลงคริสต์มาสในแต่ละปีตามวัน ซึ่งแน่นอนว่าสัดส่วนจำนวนครั้งในการเล่นเพลงคริสต์มาสจะเพิ่มขึ้นเรื่อยๆ เมื่อเข้าไปใกล้วันคริสต์มาสจนมีค่าสูงที่สุดในวันคริสต์มาสอีฟ และสัดส่วนจำนวนครั้งที่เล่นจะลดลงอย่างรวดเร็วหลังวันคริสต์มาส แต่ที่น่าแปลกใจก็คือว่า วันที่สัดส่วนขึ้นถึง 50 เปอร์เซ็นต์ ของจำนวนครั้งที่เล่นในวันคริสต์มาสอีฟนั้นจะช้าลงทุกปี นั่นคือในปี 2005 วันที่มีการเล่นเพลงคริสต์มาสคิดเป็น 50 เปอร์เซ็นต์ ของวันคริสต์มาสอีฟเดียวกัน นั่นคือวันที่ 30 พฤศจิกายน ในขณะที่ของปี 2011 คือวันที่ 6 ธันวาคม ซึ่งจากข้อมูลก็สรุปผลวิเคราะห์ได้ว่า ผู้คนเริ่มเข้าสู่บรรยากาศวันคริสต์มาสช้าลงเรื่อยๆ

