

บรรณนิทัศน์

Big Data Series 1: Introduction to a Big Data Project

ปฐมบทในการทำโปรเจกต์บิ๊กดาต้า

พรรณวดี ประยงค์¹



ชื่อหนังสือ: *Big Data Series 1: Introduction to a Big Data Project*
ปฐมบทในการทำโปรเจกต์บิ๊กดาต้า

ผู้แต่ง: อัสมา กุลวานิชไชยนันท์

ปีที่พิมพ์: 2561

สำนักพิมพ์: พราว เพรส (2002) จำกัด

จำนวนหน้า: 247 หน้า

ราคา: 290 บาท

¹ อาจารย์ประจำกลุ่มวิชาสื่อสารองค์กร คณะวารสารศาสตร์และสื่อสารมวลชน มหาวิทยาลัยธรรมศาสตร์

หลายปีที่ผ่านมา เราได้ยินคำว่า “Big Data” บ่อยครั้ง ดูเหมือนเป็นเทรนด์ใหม่ที่ใครๆ ก็พูดถึง หลายองค์กรเริ่มหันมาสนใจ Big Data พยายามทำความเข้าใจ ต่อยอด และนำไปปรับใช้ โดยเฉพาะในภาคธุรกิจ เพราะเล็งเห็นประโยชน์ของดาต้า (data) ทั้งในแง่ของการบริหารจัดการ รวมถึงการทำความเข้าใจผู้คนหรือลูกค้า แต่ดูเหมือนจะไม่ใช่ว่าเรื่องง่ายสำหรับบางองค์กร เพราะการไปสู่ความสำเร็จในเรื่อง Big Data ไม่ได้เพียงแค่การพัฒนาด้านเทคนิค ระบบปัญญาประดิษฐ์ (artificial intelligence: AI) หรือพัฒนาทักษะด้านดาต้าหรือดิจิทัลให้กับคนในองค์กรเท่านั้น แต่ยังเป็นเรื่องของ การเปลี่ยนแนวคิดหรือ “mindset” ของคนในองค์กรให้เห็นความสำคัญของข้อมูล ปรับเปลี่ยนวิธีการทำงาน วิธีการตัดสินใจด้วยข้อมูล เพื่อจะนำไปสู่ความสำเร็จขององค์กรแบบที่องค์กรระดับโลกอย่างเน็ตฟลิกซ์ (Netflix) อเมซอน (Amazon) เฟซบุ๊ก (Facebook) และกูเกิล (Google) นั้นทำได้มาแล้ว

หนังสือ *Big Data Series 1: Introduction to a Big Data Project* ปฐมบทในการทำโปรเจกต์บิ๊กดาต้า เป็นหนังสือเล่มแรกๆ ที่ผู้เขียนมีโอกาสได้อ่านหลังจากเริ่มสนใจเรื่อง Big Data เหมาะกับผู้ที่ต้องการจะปูพื้นความรู้เพื่อเข้าใจหลักการเบื้องต้นที่อาจเป็นหัวใจสำคัญของการทำ Big Data Project ในองค์กร และเรียนรู้ศัพท์เทคนิคต่างๆ เพื่อนำไปต่อยอดในอนาคต ผู้เขียนเชื่อว่าการเปลี่ยนแปลงเกิดขึ้นทุกวัน คนที่ปรับตัวได้ทันและไม่เกรงกลัวกับเทคโนโลยีที่เข้ามา จะมองเห็นโอกาสที่จะพัฒนาตนเอง พัฒนางานของตนเองในแง่มุมใหม่ๆ ที่ตอบโจทย์มากขึ้น และอย่าลืมว่าการตัดสินใจที่ดีที่สุดอาจไม่ได้มาจากความเชื่อหรือสิ่งที่เคยบอกต่อหรือปฏิบัติกันมา หากแต่มาจากข้อมูลจำนวนมากที่ถูกต้องแม่นยำ เพราะข้อมูลอาจเปรียบได้กับประสบการณ์ที่สั่งสมมานั่นเอง

เริ่มต้นปฐมบทของ Big Data

หนังสือเริ่มจากการเล่าที่มาที่ไปของการใช้ข้อมูลหรือดาต้าที่มีมาตั้งแต่ในยุคสงครามโรมัน โดยใช้เพื่อประโยชน์ทางการทหารและการเมือง และขยับมาสู่การใช้ประโยชน์ในทางธุรกิจ เมื่อคอมพิวเตอร์และอินเทอร์เน็ตมีการพัฒนาอย่าง

ต่อเนื่อง มีราคาที่จับต้องได้ จึงทำให้ระบบการเก็บข้อมูลในองค์กรค่อยๆ ปรับเปลี่ยนจากการบันทึกลงกระดาษมาสู่การบันทึกข้อมูลแบบดิจิทัล เริ่มมีการพูดถึง artificial intelligence (AI) หรือปัญญาประดิษฐ์ โดยในปัจจุบันเป็นที่รู้กันดีว่า AI เป็นส่วนสำคัญที่ช่วยในการจัดการข้อมูลขนาดใหญ่ภายใต้เวลาที่จำกัด ซึ่งอาจไม่สามารถใช้มนุษย์หรือโปรแกรมทั่วไปทำได้ รวมไปถึงการพัฒนา Internet of Things (IoTs) ที่ทำให้อุปกรณ์อิเล็กทรอนิกส์ต่างๆ สามารถส่งผ่านข้อมูลถึงกันได้อัตโนมัติ เหล่านี้ล้วนเป็นเรื่องที่เกี่ยวข้องกับ Big Data ทั้งสิ้น ซึ่งผู้เขียนแนะนำว่า ผู้อ่านอาจหาเวลาเล็กน้อยเพื่อทำความเข้าใจเทรนด์หรือกระแสใหม่ๆ ที่เกิดขึ้นนี้

นอกจากนี้ เมื่อเข้าสู่ยุคของเว็บ 2.0 (Web 2.0) ที่ผู้ใช้งานสามารถสร้างสรรค์เนื้อหาเองได้ อาทิ Facebook.com และการเปิดตัวสมาร์ตโฟน (smart-phone) อย่าง iPhone ในปี 2007 ที่ทำให้คนเราใช้อินเทอร์เน็ตมากขึ้นทุกที่ทุกเวลา ส่งผลให้ข้อมูลบนระบบนั้นเพิ่มขึ้นอย่างรวดเร็ว ไร้ขีดจำกัด การเริ่มเปิดให้บริการ Cloud Computing ของบริษัทยักษ์ใหญ่ด้านเทคโนโลยีต่างๆ ได้เข้ามาเปลี่ยนแปลงกระบวนการเก็บรวบรวมและการประมวลผลข้อมูลของโลกยุคนี้ เมื่อเทคโนโลยีมีราคาถูกลง จับต้องได้ การพัฒนาด้าน Big Data ในองค์กรต่างๆ จึงเป็นไปได้จริงมากขึ้น ผู้เขียนชอบคำกล่าวของ Chris Anderson (อสมากุลวานิชไชยนันท์, 2561: 20) ที่ถูกหยิบยกมาในหนังสือเล่มนี้ที่กล่าวว่า “การมีข้อมูลมากขึ้น ทำให้ model หรือสมการแบบเดิมๆ ตามทฤษฎีเก่าๆ นั้นล้าสมัย เพราะข้อมูลใหม่ๆ สามารถบ่งบอกอะไรใหม่ๆ ได้ชัดเจนกว่า” ดังนั้น เมื่อข้อมูลในระบบเพิ่มขึ้นอย่างรวดเร็วต่อเนื่องในปัจจุบัน ผู้เขียนเชื่อว่า เรากำลังรอไม่ได้อีกต่อไป

ถัดมา หนังสือได้เน้นย้ำความหมายและคุณสมบัติของ Big Data ว่าต้องประกอบได้ด้วย 4V ได้แก่ (1) Volume ซึ่งแน่นอนว่า ข้อมูลต้องมีขนาดใหญ่ และขยายตัวต่อไปได้ ไม่หยุดอยู่กับที่ (2) Velocity คือ ความเร็วของการสร้างและการประมวลผลข้อมูล เพื่อนำไปใช้ได้ทันที (3) Variety หมายถึง ความหลากหลายของข้อมูล หรือประเภทของข้อมูลที่หลากหลาย และ (4) Veracity

คือ ความแม่นยำของข้อมูล ซึ่งแน่นอนว่า ข้อมูลขนาดใหญ่ที่ไหลเข้ามาเป็นจำนวนมากและหลากหลายนั้น ต้องอาศัยการทำความสะอาดข้อมูล หรือที่เรียกว่า Cleansing Data ก่อนนำไปใช้งาน ทั้งนี้ ในปัจจุบันคุณสมบัติของ Big Data ได้ถูกเพิ่มคุณสมบัติอื่นๆ เข้ามามากขึ้น อาทิ Value คือ ข้อมูลนั้นๆ ต้องมีคุณค่า หรือนำไปสู่คุณค่าบางประการ และ Complexity อันหมายถึงความซับซ้อนของข้อมูล ซึ่งหากผู้อ่านสนใจสามารถไปศึกษาเพิ่มเติมได้ในหนังสือที่เกี่ยวข้องกับ Big Data ต่อไป

การเริ่มต้นสู่ Big Data Project

แม้หนังสือเล่มนี้จะเน้นไปที่ภาคปฏิบัติ แต่ก็ชี้ให้เห็นว่า ทฤษฎีก็เป็นสิ่งสำคัญเพื่อปูพื้นฐานสำหรับคนที่ต้องการทำ Big Data Project ในอนาคต โดยหนังสือเล่มนี้มีภาคต่อแล้วถึง 3 เล่ม และเล่มแรกนี้ได้พาเราไปสู่จุดเริ่มต้นว่า “หากจะเริ่มต้นทำ Big Data Project ต้องทำอะไร” โดยผู้เขียนหนังสือเสนอว่า ให้สำรวจดูก่อนว่า ในการตัดสินใจเรื่องต่างๆ ในองค์กร มีส่วนไหนในองค์กรที่ยังถูกตัดสินใจด้วย “คน” อยู่บ้าง และถัดมาคือ สำรวจดูว่าเรามีข้อมูลอะไรอยู่บ้างในองค์กร ทั้งที่เก็บไว้แล้ว ซึ่งอาจนำมาต่อยอดใช้งานได้ และข้อมูลใหม่ๆ ที่น่าจะเป็นประโยชน์ เพื่อออกแบบแนวทางการเก็บข้อมูลเหล่านั้นต่อไป

ในบทแรกของหนังสือเล่มนี้ ผู้เขียนหนังสือคือ อสมา กุลวานิชไชยนันท์ ยังพยายามเน้นย้ำว่า การมีข้อมูลขนาดใหญ่หรือ Big Data ในองค์กรนั้น ไม่ได้มีประโยชน์อะไร หากเราไม่ได้นำมาใช้งานต่อยอด เพราะไม่ได้นำไปสู่ผลลัพธ์ใดๆ เลย หรือบางครั้งแม้การมีข้อมูลจำนวนไม่มาก แต่ก็สามารถสร้างคุณประโยชน์มหาศาลได้ หากข้อมูลนั้นมีคุณภาพก็จะนำไปสู่ผลลัพธ์ที่ถูกต้อง ดีกว่าข้อมูลที่ถูกรวบรวมไปเรื่อยๆ อย่างไรก็ตาม ไร้ทิศทาง และท้ายที่สุด สิ่งที่คุณเขียนเห็นด้วยอย่างยิ่งกับหนังสือเล่มนี้คือ การปรับเปลี่ยน “mindset” หรือแนวคิดในการทำงานให้ไม่ยึดติดอยู่กับสิ่งเดิมๆ แต่พร้อมเผชิญการเปลี่ยนแปลงที่แม้จะล้มลุกคลุกคลาน หรือผิดพลาด ก็จะไปสู่การพัฒนาแบบก้าวกระโดดได้ ซึ่งเป็นหลักการขององค์กรยุคใหม่ที่ต้องปรับตัวให้ทันกับเทคโนโลยีที่เปลี่ยนแปลง

อย่างรวดเร็วตามหลักของ “Agile” ที่ตอบรับการเปลี่ยนแปลงหรือ “welcome change” ได้ทุกช่วงของการพัฒนา

Big Data มาจากไหน และเราจะต้องพึ่งพาใครในองค์กรบ้าง

ในบทความนี้ ผู้เขียนหนังสือเล่มนี้ได้เล่าถึงที่มาของ Big Data อันมาจากข้อมูลบนระบบเพิ่มขึ้นอย่างก้าวกระโดดตั้งแต่ปี 2010 เป็นต้นมา โดยพบว่า มีข้อมูลบนโลกเพิ่มขึ้นถึง 10 เท่าจากปี 2010 ถึง 2020 สาเหตุจากการเกิดขึ้นของเว็บ 2.0 (Web 2.0) โซเชียลมีเดีย (social media) และสมาร์ทโฟน (smart-phone) ที่ทำให้เราสามารถสร้างเนื้อหาและแบ่งปันกันได้ไม่รู้จบ ลองนึกดูว่า ตั้งแต่มีโซเชียลมีเดียและโทรศัพท์สมาร์ทโฟน เราโพสต์หรือแชร์เนื้อหาเพิ่มขึ้นมากแค่ไหน ด้วยเหตุนี้ การพัฒนาโครงการที่เกี่ยวข้องกับ Big Data จึงเป็นไปได้จริงมากขึ้น อย่างไรก็ตาม แม้จะมีข้อมูลไหลเข้ามาในระบบเป็นจำนวนมาก แต่หลายองค์กรก็ยังไม่สามารถนำข้อมูลเหล่านั้นมาใช้ได้ในทันที เพราะข้อมูลบนระบบส่วนใหญ่เป็น Unstructured Data หรือข้อมูลแบบไม่มีโครงสร้าง ซึ่งจากชื่อคงคาดเดาได้ว่า ไม่สามารถใช้งานได้ทันที จะต้องอาศัยผู้เชี่ยวชาญอย่าง Data Scientist มาจัดการข้อมูล และปรับเปลี่ยนให้เป็น Structured Data หรือข้อมูลแบบมีโครงสร้างเสียก่อน จึงจะนำมาใช้งานได้ โดยข้อมูลส่วนใหญ่บนโซเชียลมีเดียมักจะเป็นข้อมูลประเภท Unstructured Data ทำให้มีข้อมูลประเภทนี้บนระบบถึงร้อยละ 80 ของข้อมูลทั้งหมด โดยจะต้องอาศัยการ Clean Data หรือการทำความสะอาดข้อมูลก่อนนำมาใช้เพื่อไม่ให้เกิดข้อผิดพลาด หลายองค์กรจึงต้องปวดหัวเล็กน้อยในการหาผู้เชี่ยวชาญเข้ามาแนะนำหรือดำเนินการ ซึ่งในหนังสือเล่มนี้ผู้เขียนได้อธิบายขั้นตอนของการจัดการข้อมูลไว้ให้ผู้อ่านศึกษาอย่างละเอียด รวมถึงยกตัวอย่างที่จะทำให้เข้าใจกระบวนการขั้นตอนต่างๆ มากขึ้น อย่างไรก็ตาม หากเราไม่ได้มีข้อมูลที่มากนัก เราอาจเริ่มวางแผนเก็บข้อมูลอย่างมีทิศทาง และจัดแบ่งประเภทของข้อมูลอย่างเป็นระบบ ก็จะช่วยให้นำเอาข้อมูลจำนวนไม่มากเหล่านั้นมาใช้ประโยชน์ได้ก่อน

แน่นอนว่า การทำงานกับ Big Data นั้นไม่ใช่เรื่องง่าย และจำเป็นต้องอาศัยผู้เชี่ยวชาญเข้ามาช่วยในองค์กร มีความเกี่ยวข้องทั้งฝั่งที่เป็นผู้วางกลยุทธ์ขององค์กร ฝั่งธุรกิจ และฝั่งเทคนิค หนังสือ *Big Data Series 1: Introduction to a Big Data Project* ประมวลบทในการทำโปรเจกต์บิกดาต้า ได้กล่าวถึงสามตำแหน่งงานหลักที่เกี่ยวข้องกับระบบนิเวศของการทำ Big Data Project ได้แก่ (1) Data Engineer หรือวิศวกรข้อมูล ผู้ทำหน้าที่เป็นด่านแรกที่เข้าไปจัดการกับข้อมูล ออกแบบระบบการเก็บข้อมูล ก่อนที่ (2) Data Scientist หรือนักวิทยาศาสตร์ข้อมูล จะนำข้อมูลดังกล่าวไปวิเคราะห์พัฒนาโมเดล และส่งต่อให้กับ Data Engineer นำไปใช้อีกครั้งหนึ่ง เพื่อสร้างเป็นระบบหรือแอปพลิเคชันสำหรับ (3) Data Analyst หรือนักวิเคราะห์ข้อมูล ที่จะเข้ามาใช้งานหรือวิเคราะห์ข้อมูลเหล่านี้เพื่อนำไปใช้ประโยชน์ในองค์กรต่อไป อาจเป็นการใช้เพื่อแก้ปัญหาในองค์กร หรือตอบโจทย์ทางธุรกิจต่างๆ ที่ฝั่งกลยุทธ์หรือฝั่งธุรกิจวางไว้ ซึ่งตำแหน่ง Data Analyst นั้นจำเป็นจะต้องมีทักษะทางการสื่อสารที่ดีด้วย เพราะต้องนำเสนอข้อมูลที่วิเคราะห์ให้ฝ่ายอื่นๆ เข้าใจ ทั้งนี้ หนังสือเล่มนี้ได้เล่าถึงความซับซ้อนของงานทั้งสามตำแหน่ง และยกตัวอย่างเครื่องมือและเทคโนโลยีสมัยใหม่สำหรับคนที่ทำงานในสายงานเหล่านี้ โดยผู้เขียนยังแสดงความกังวลต่อการขาดแคลนทรัพยากรคนที่มีความเชี่ยวชาญทั้งสามด้านข้างต้นในประเทศไทย ซึ่งพบว่า มหาวิทยาลัยต่างๆ ไม่สามารถผลิตบัณฑิตได้เพียงพอกับความต้องการ ทำให้เด็กรุ่นใหม่อาจไปเลือกลงเรียนคอร์สออนไลน์หรือคอร์สระยะสั้นที่อาจทำให้เกิดปัญหาตามมา อาทิ ความไม่ต่อเนื่อง และโอกาสในการได้ลงมือปฏิบัติจริง โดยเป็นที่น่ากังวลว่า ไม่เฉพาะแต่ในระดับองค์กรเท่านั้น แม้แต่ในระดับประเทศเอง ก็อาจทำให้เกิดช่องว่างของการพัฒนาทางเทคโนโลยีของประเทศได้ หากขาดแคลนบุคลากรที่มีทักษะเหล่านี้

เลือกไม่ได้...และคงต้องไปต่อกับ Big Data

ตลอดเวลาที่อ่านหนังสือเล่มนี้ ทำให้ผู้เขียนคิดเสมอว่า “เราคงไม่มีทางเลือกมากนักว่า จะใช้ Data หรือไม่ เพราะข้อมูลมีอยู่ทุกที่ และไหลบ่าเข้ามา

ตลอดเวลาในโลกยุคดิจิทัล ดังนั้นเราคงต้องเรียนรู้และเดินไปข้างหน้า” หลายคนคงเคยมีโอกาสดูเห็นสถิติ “Internet Minute” ที่เผยแพร่ออกมาอยู่ทุกๆ ปี เพื่อเล่าว่าในหนึ่งนาที คนเราทั้งโลกทำอะไรบ้างบนอินเทอร์เน็ต ไม่ว่าจะเป็นการเสิร์ชหาข้อมูลบนกูเกิล (Google) ก็ครั้งในหนึ่งนาที การทวีตบนทวิตเตอร์ (Twitter) ก็ข้อความในหนึ่งนาที ไปจนถึงจำนวนชั่วโมงของการดูหนังหรือซีรีส์บนเน็ตฟลิกซ์ (Netflix) ของคนทั่วโลกในหนึ่งนาที สิ่งเหล่านี้คือสิ่งที่ผู้เขียนหนังสือเล่มนี้พยายามสะท้อนให้เห็นว่า ในหนึ่งนาทีนั้นมีข้อมูลไหลเข้ามาบนระบบมากมาย ข้อมูลเกิดขึ้นแบบเรียลไทม์ และข้อมูลเหล่านี้ล้วนมีความซับซ้อนและมีคุณค่าองค์กรไหนที่สามารถเข้าถึงและเข้าใจข้อมูลได้ก่อนย่อมมีโอกาสได้เปรียบกว่าคู่แข่งหรือองค์กรอื่นๆ

ในส่วนสุดท้ายของบรรณนิทัศน์นี้ ผู้เขียนขอกล่าวถึงประโยชน์บางประการของการทำ Big Data Project ที่หนังสือ *Big Data Series 1: Introduction to a Big Data Project* ปรุบทในการทำโปรเจกต์บิกดาต้า ได้สรุปไว้ เพื่อว่าจะจุดประกายให้หลายๆ คนอยากลองพัฒนาโมเดล Big Data และนำไปใช้งานในองค์กรของตนเองบ้าง ประโยชน์ประการแรก คือ ทำให้การตัดสินใจต่างๆ ในการทำงานนั้นรวดเร็วและแม่นยำมากขึ้น โดยผู้เขียนหนังสือได้เปรียบเทียบว่าในการตัดสินใจต่างๆ มักใช้ประสบการณ์ในการตัดสินใจ คนที่มีประสบการณ์มากก็มักจะตัดสินใจได้รวดเร็วและแม่นยำมากขึ้น ซึ่งประสบการณ์ในความเป็นจริงแล้วก็เปรียบได้กับ “ข้อมูล” แบบ Unstructured Data ที่หากเราจดบันทึกไว้อย่างเป็นระบบก็จะกลายเป็นข้อมูลที่มีค่าได้นั่นเอง อย่างไรก็ตาม ในโลกยุคนี้ที่มีข้อมูลมหาศาลและซับซ้อน และด้วยสภาพแวดล้อมเปลี่ยนแปลงไปอย่างรวดเร็ว ประสบการณ์เดิมๆ จึงไม่อาจตอบโจทย์ได้ทั้งหมด ด้วยเหตุนี้จึงจำเป็นต้องใช้ AI หรือปัญญาประดิษฐ์ที่สามารถคิด คำนวณ และตัดสินใจได้อย่างแม่นยำ รวดเร็ว และชัดเจนกว่ามนุษย์โดยอาศัยข้อมูลจาก Big Data มาประมวลผล หรือที่เราเรียกว่า Machine Learning หนังสือเล่มนี้ได้ยกตัวอย่างกรณีศึกษาของแบรนด์ Adidas ที่ใช้ข้อมูลทั้งในระบบของ Adidas เอง และข้อมูลจากสื่อออนไลน์ที่มีอยู่ทั่วโลก มาประมวลผลด้วย AI เพื่อใช้ในการตัดสินใจออกแบบรองเท้ารุ่นต่างๆ

ทั้งสี่ที่น่าจะได้รับความนิยมในซีซั่นหน้า สไต์ล และวัสดุที่น่าจะเป็นที่ต้องการของลูกค้า ทำให้ทีมออกแบบไม่ต้องเสียเวลาไปประชุมถกเถียงหรือคาดเดากันเอาเอง และมีความแม่นยำมากกว่า จนกลายมาเป็นรองท้ารุ่นที่ได้รับความนิยมไปทั่วโลก ประโยชน์อีกประการหนึ่งของการพัฒนา Big Data Project ที่หนังสือเล่มนี้กล่าวถึงคือ ทำให้เกิดสินค้าและบริการใหม่ๆ ที่แน่นอนว่าสามารถนำมาซึ่งรายได้ให้กับองค์กรได้มากขึ้น โดยผู้เขียนหนังสือได้ยกกรณีศึกษาของธนาคารแห่งหนึ่งในสหรัฐอเมริกาที่ยอมลงทุนกับการวางรากฐาน ทั้งการเก็บและการประมวลผลข้อมูลก่อนคู่แข่งรายอื่นๆ ในตลาด เพื่อใช้ทำความเข้าใจลูกค้าแต่ละราย และออกแบบผลิตภัณฑ์ใหม่ที่ตอบโจทย์กับลูกค้าแต่ละกลุ่ม อาทิ การออกบัตรเครดิตที่ตรงกับการใช้งานของลูกค้าแต่ละคนให้มากที่สุด ผลลัพธ์ที่ได้คือทำให้บริษัทมีส่วนแบ่งทางการตลาดเพิ่มสูงขึ้นอย่างเห็นได้ชัด

การเข้าสู่ยุคดิจิทัลและการเติบโตของเทคโนโลยีสมัยใหม่ ไม่ว่าจะเป็น AI, IoT และ Big Data ทำให้หลายองค์กร ไม่เว้นแม้แต่ภาครัฐ ต่างต้องปรับตัวและลงทุนด้าน Data มากขึ้น ทั้งการวางโครงสร้างพื้นฐานด้านข้อมูล (Data Infrastructure) การเก็บ การวิเคราะห์ และการใช้งานข้อมูลอย่างเป็นระบบ และรวมถึงการวางระบบด้านความปลอดภัยและการคุ้มครองข้อมูลส่วนบุคคลดังที่มีการบังคับใช้กฎหมายในหลายๆ ประเทศทั่วโลก ซึ่งประเทศไทยเองก็เป็นหนึ่งในประเทศที่เริ่มมีการบังคับใช้ พ.ร.บ.คุ้มครองข้อมูลส่วนบุคคล หรือที่เราเรียกกันว่า PDPA ไปเมื่อเร็วๆ นี้ ดังนั้น การทำความเข้าใจเรื่อง Big Data ในวันนี้จึงมีความจำเป็นอย่างยิ่งสำหรับทุกๆ องค์กร ซึ่งอาจยังไม่ต้องไปไกลถึงการมีตำแหน่ง Data Scientist และ Data Engineer ในองค์กร แต่อาจต้องเริ่มจากการมองหา Data และใช้ประโยชน์จาก Data รวมถึงพัฒนาคนที่จะเป็น Data Analyst ในองค์กรขึ้นมาให้ได้ ซึ่งคงจะไม่ยากเกินไปนักสำหรับหลายๆ องค์กรที่จะเริ่มต้น

บรรณานุกรม

อสมมา กุลวานิชไชยนันท์ (2561), *Big Data Series 1: Introduction to a Big Data Project*
ปฐมบทในการทำโปรเจกต์บิ๊กดาต้า, กรุงเทพฯ: พราว เพรส (2002).