

# Bigdata and the Development of Manpower Potential.

Damras Onchawiang<sup>\*1</sup>

Duangporn Thamma<sup>2</sup>

damras@go.buu.ac.th\*

## Abstract

Developing manpower to meet individual abilities and labor market needs. Strengthening manpower potential will strengthen the country's economy and productivity. However, human resources or manpower in the Eastern Special Development Zone It consists of people of different ages, and many professions. Massive amounts of data are therefore necessary to develop manpower in all dimensions.

will know the situation Know the relationship of variable factors. know the direction or trend know the operating guidelines Which of these will be the most effective actions. It's like "Know him, know us, a hundred battles, a hundred wins." Big Data, refers to the mass of data in digital file formats that are diverse, enormous, complex and fast. Large volumes of data require storage, management, and processing. under the framework of appropriate information governance.

**Keyword:** Bigdata/Manpower Potential/Data Governance

---

<sup>1-2</sup>Lecturer in Educational Technology Department Faculty of Education  
Burapha University

## ข้อมูลมห้ตักับการพัฒนาศักยภาพกำลัองคน

ดำรัส อ่อนเฉวียง\*<sup>1</sup>

ดวงพร ธรรมะ<sup>2</sup>

damras@go.buu.ac.th\*

### บทคัดย่อ

การพัฒนากำลัองคน เพื่อให้สอดคลองกับความสามารถเฉพาะบุคคลและ ความต้องการของตลาดแรงงาน ความเข้มแข็งของศักยภาพกำลัองคนจะสร้าง ความเข้มแข็งเศรษฐกิจและผลิตภาพเพิ่มขึ้นให้กับประเทศ อย่างไรก็ตามทรัพยากรมนุษย์ หรือกำลัองคนในเขตในเขตพัฒนาพิเศษภาคตะวันออก ประกอบด้วยบุคคลหลายช่วงวัย และหลายหลายอาชีพ มวลข้อมูลปริมาณมหาศาลจึงความจำเป็นพัฒนากำลัองคน ทุกมิติ จะทำให้รู้สภาพการณ์ รู้ความล้มพันธ์ของปัจจัยตัวแปร รู้ทิศทางหรือแนวโน้ม รู้แนวทางการปฏิบัติการ เหล่านี้จะเป็นการดำเนินการใดเป็นไปอย่างมีประสิทธิภาพ สูงสุด เข้าทำนอง “รู้เขา รู้เรา รบร้อยครั้งชนะร้อยครั้ง” คำว่าข้อมูลมห้ต (Big Data) บางครั้งเรียกบักดาต้า หมายถึง มวลของข้อมูลรูปแบบไฟล์ดิจิทัลที่มีความหลากหลาย ปริมาณมหาศาล มีความความซับซ้อนและความรวดเร็ว ทั้งนี้ข้อมูลปริมาณมากจำเป็น ต้องมีพื้นที่จัดเก็บ การจัดการ และการประมวลผล ภายใต้กรอบธรรมาภิบาลข้อมูล อย่างเหมาะสม

คำสำคัญ: ข้อมูลมห้ต/ศักยภาพกำลัองคน/ธรรมาภิบาลข้อมูล

<sup>1-2</sup>อาจารย์ประจำสาขาวิชาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

## บทนำ

ปัจจุบันถือได้ว่าเทคโนโลยีเข้ามามีบทบาท และมีอิทธิพลต่อการดำเนินชีวิตประจำวันอย่างหลีกเลี่ยงไม่ได้ เกิดพลเมืองดิจิทัล เกิดสังคมดิจิทัล เกิดการสื่อสารข้อมูลหลากหลายทำให้มีข้อมูลปริมาณมหาศาลที่เพิ่มขึ้นอย่างรวดเร็ว เช่น ข้อมูลดัชนีการค้นคืนของเว็บ ข้อมูลการสื่อสารทางโทรศัพท์เคลื่อนที่ ข้อมูลในเครือข่ายสังคม ข้อมูลการพยากรณ์อากาศ ข้อมูลภาพถ่ายดาวเทียม อีกทั้งข้อมูลมีรูปแบบไฟล์หลากหลาย ไม่ว่าจะเป็น ข้อความ รูปภาพ วิดีโอ มัลติมีเดีย และยังเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาและรวดเร็ว หรือรวมเป็นคุณสมบัติที่เรียกว่าข้อมูลมหัตหรือ Big Data นั่นเอง เมื่อข้อมูลปริมาณมหาศาล มีความหลากหลายของข้อมูล และมีความเร็วทั้งการนำเข้า การประมวลผลและการวิเคราะห์ผล ข้อมูลมหัตจึงมีความสำคัญต่อการดำเนินการทางธุรกิจ หรือการดำเนินการกิจการขององค์กร ตลอดจนการดำเนินการพัฒนาระบบข้อมูลมหัต (Big Data) เพื่อรวบรวมและวิเคราะห์ข้อมูลเกี่ยวกับศักยภาพกำลังคน และนำทางการพัฒนาศักยภาพกำลังคน อันจะเป็นประโยชน์ต่อ กำลังคน ผู้ประกอบการ ผู้ผลิตกำลังคน และเศรษฐกิจของประเทศในภาพรวม

## ความหมายข้อมูลมหัต (Big Data)

สำนักงานราชบัณฑิตยสภา(2563) บัญญัติคำว่า “Big Data” เป็นภาษาไทยว่า “ข้อมูลมหัต” โดยมีความหมายคือ ข้อมูลหลากหลายปริมาณมหาศาลที่เพิ่มขึ้นอย่างรวดเร็ว เช่น ข้อมูลดัชนีการค้นคืนของเว็บ ข้อมูลการสื่อสารทางโทรศัพท์เคลื่อนที่ ข้อมูลในเครือข่ายสังคม ข้อมูลการพยากรณ์อากาศ ข้อมูลภาพถ่ายดาวเทียม

เว็บไซต์วิกิพีเดีย (2564 ) ได้ให้ความหมายว่า Big Data คือ ข้อมูลมหัต หรือ ข้อมูลขนาดใหญ่ หรือนิยมทับศัพท์ บิ๊กเดตา, บิ๊กดาต้า คือชุดข้อมูลที่มีขนาดมหาศาลและความซับซ้อนมาก การดึงข้อมูลอย่างเป็นระบบ หรือจัดการกับชุดข้อมูลที่ใหญ่หรือซับซ้อนเกินกว่าจะจัดการด้วยซอฟต์แวร์แอปพลิเคชันประมวลผลข้อมูลแบบเดิม ข้อมูลที่มีหลายช่อง (คอลัมน์) ให้อำนาจทางสถิติมากกว่า ในขณะที่ข้อมูลที่มีความซับซ้อนสูง (แอตทริบิวต์หรือคอลัมน์มากกว่า) อาจนำไปสู่อัตราการค้นพบเท็จที่สูงขึ้นความท้าทายในการวิเคราะห์ข้อมูลขนาดใหญ่รวมถึง การเก็บข้อมูล การจัดเก็บข้อมูล การวิเคราะห์ข้อมูล การค้นหา การแบ่งปัน การถ่ายโอน การแสดงภาพ การสืบค้น การอัปเดต ความเป็นส่วนตัวของข้อมูล และแหล่งข้อมูล เกี่ยวข้องกับแนวคิดหลักสามประการ ได้แก่ ปริมาณ ความหลากหลาย และความเร็ว

เว็บไซต์ไอ บี เอ็ม ออนไลน์ (IBM, Online, 2017 ) ได้ให้ความหมายว่า Big Data หมายถึง การเก็บข้อมูลที่มีปริมาณมหาศาลเป็นสิ่งที่เป็นไปได้ในปัจจุบัน ข้อมูลรูปแบบใหม่ที่จัดเก็บในปัจจุบันไม่เพียงแต่มีความใหญ่ ทั้งปริมาณ ด้านความซับซ้อนและรูปแบบที่หลากหลายมากขึ้น เช่น เสียง ภาพนิ่ง ภาพเคลื่อนไหว เป็นต้น ซึ่งก็คือการใช้ประโยชน์จากเทคโนโลยีของอุปกรณ์ตรวจจับที่มีขนาดเล็กลง แต่มีประสิทธิภาพมากขึ้น

การ์ทเนอร์ (Gartner, 2019) ได้ให้คำจำกัดความของ Big data ว่าเป็นสินทรัพย์ข้อมูลที่มีปริมาณมาก มีความเร็วสูง และ/หรือมีความหลากหลายสูง ซึ่งต้องการรูปแบบการประมวลผลข้อมูลที่ประหยัดต้นทุนและเป็นนวัตกรรมที่ช่วยทำให้เกิดความเข้าใจอย่างลึกซึ้งซึ่งเกิดการตัดสินใจ และเกิดกระบวนการอัตโนมัติ (Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.)

เว็บไซต์เอส เอ เอส (www.sas.com, 2563) ได้ให้ความหมายว่า Big Data หมายถึง ข้อมูลที่มีขนาดใหญ่ รวดเร็วหรือซับซ้อนจนยากหรือเป็นไปได้ในการประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมาก

กรมการจัดหางาน (2562) ได้ให้ความหมายว่า Big Data คือ ปริมาณข้อมูลจำนวนมากมหาศาลที่มีอยู่ในองค์กรทุกรูปแบบ ไม่ว่าจะแหล่งที่มาจะมาจกภายในองค์กรหรือภายนอกก็ตาม ทั้งนี้แบ่งออกเป็นข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) และข้อมูลที่มีโครงสร้างไม่ชัดเจน (Unstructured Data)

สถาบันบิกดาต้า (Government Big Data Institute: GBDi) (2563) ได้ให้ความหมายว่า Big Data คือ โลกของข้อมูล แบ่งออกเป็นข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) และข้อมูลที่มีโครงสร้างไม่ชัดเจน (Unstructured Data) เมื่อนำมารวมกันสามารถวิเคราะห์ได้ผลลัพธ์มหาศาล

สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม. (2564). ได้ให้ความหมายว่า Big Data หมายถึง ข้อมูลขนาดใหญ่/ปริมาณมาก หรือข้อมูลจำนวนมากมหาศาล ทุกเรื่อง ทุกแง่มุม ทุกรูปแบบ ซึ่งอาจเป็นข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่เก็บอยู่ในตารางข้อมูลต่างๆ หรืออาจเป็นข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data) เช่น ล็อกไฟล์ (Log files) หรือแม้กระทั่งข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลการโต้ตอบปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) เช่น Facebook, twitter หรือไฟล์จำพวกมีเดีย เป็นต้น ซึ่งอาจจะเป็นข้อมูลภายในองค์กรและภายนอกที่มาจาก การติดต่อระหว่างองค์กร หรือจากทุกช่องทางที่ติดต่อกับลูกค้า แต่ทั้งหมดนี้ ก็ยังคงเป็นเพียงข้อมูลดิบที่รอการนำมาประมวลและวิเคราะห์เพื่อนำผลที่ได้ มาสร้างมูลค่าทางธุรกิจ ข้อมูลเหล่านี้จะไม่ได้อยู่ในรูปแบบที่องค์กรสามารถนำไปใช้ได้ทันที แต่อาจมีข้อมูลที่เป็นประโยชน์ต่อองค์กรบางอย่างแฝงอยู่

จากการศึกษาเกี่ยวกับความหมายข้อมูลมหัต (Big Data) สามารถสรุปได้ว่า ข้อมูลมหัต (Big Data) หมายถึง มวลของข้อมูลรูปแบบไฟล์ดิจิทัลที่มีความหลากหลาย ปริมาณมหาศาล มีความซับซ้อนและความรวดเร็ว

## ความสำคัญของข้อมูลมหัต

ข้อมูลมหัตมีคุณลักษณะที่สำคัญ คือข้อมูลมีปริมาณมากมายมหาศาล มีความหลากหลาย และเกิดขึ้นอย่างรวดเร็ว จึงมีความสำคัญดังต่อไปนี้

- 1) การพรรณนาสภาพการณ์ใด ๆ การวิเคราะห์เชิงพรรณนา (Descriptive analytics) ทำให้ค้นพบนี้ส่งสัญญาณว่ามีบางอย่างผิดปกติหรือถูกต้อง

2) การบริการลูกค้าที่ดีขึ้น เข้าใจลูกค้ามากขึ้น การวิเคราะห์เชิงวินิจฉัย(Diagnostic analytics) ทำให้อธิบายสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อสิ่งที่เกิดขึ้น เช่น ความสัมพันธ์ระหว่างกิจกรรมทางการตลาดกับการซื้อสินค้า

3) ความได้เปรียบด้านการแข่งขันกับคู่แข่ง การวิเคราะห์เชิงพยากรณ์(Predictive analytics) ทำให้ทราบว่าสิ่งใดบ้างที่กำลังจะเกิดขึ้น หรือแนวโน้ม ต่าง ๆ ในอนาคต

4) แนวทางการดำเนินการได้อย่างมีประสิทธิภาพ การวิเคราะห์เชิงแนะนำ(Prescriptive analytics) ทำให้ได้ข้อเสนอแนะว่าต้องดำเนินการใดเพื่อขจัดปัญหาในอนาคตหรือใช้ประโยชน์จากแนวโน้มที่มีแนวโน้มจะเกิดขึ้นอย่างสูงสุด

5) ประสิทธิภาพของปฏิบัติการ ข้อมูลมหัต จะทำให้รู้สภาพการณ์ รู้ความสัมพันธ์ของปัจจัยตัวแปร รู้ทิศทางหรือแนวโน้ม รู้แนวทางการปฏิบัติการเหล่านี้จะเป็นการดำเนินการใดเป็นไปอย่างมีประสิทธิภาพสูงสุด เข้าทำนอง “รู้เขา รู้เรา รบร้อยครั้งชนะร้อยครั้ง”

## ลักษณะข้อมูลมหัต

ด็ก ลานีย์ (Doug Laney, 2001) นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความและกำหนดลักษณะของข้อมูลมหัตประกอบด้วย 3V ดังนี้

1) Volume (ปริมาณ) : องค์กรต่างๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย และอื่นๆ ในอดีต การจัดเก็บข้อมูลถือเป็นปัญหาใหญ่-แต่เมื่อค่าใช้จ่ายในการจัดเก็บบนแพลตฟอร์มต่างๆ เช่น พื้นที่จัดเก็บข้อมูลส่วนกลาง (Data Lake) และ Hadoop ลดลง ภาระนี้จึงบรรเทาลง

2) Velocity (ความเร็ว) : ด้วยการเติบโตของ Internet of Things ข้อมูลจะถูกส่งไปยังธุรกิจต่างๆ ด้วยความเร็วที่ไม่เคยมีมาก่อนและต้องได้รับการจัดการในเวลาที่เหมาะสม แท็ก RFID, เซ็นเซอร์ และสมาร์ตมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับกระแสข้อมูลเหล่านี้ในแบบเรียลไทม์

3) Variety (ความหลากหลาย) : ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูลหุ่น และธุรกรรมทางการเงิน

ไอ บี เอ็ม ออนไลน์ (IBM, Online, 2017) ได้กำหนดองค์ประกอบของข้อมูลมหัตประกอบด้วย 4VS ดังนี้

1) ปริมาณ (Volume) การจัดเก็บข้อมูลที่มีความถี่ในการจัดเก็บก็ยิ่งเพิ่มขึ้น ข้อมูลบางประเภทเคยจัดเก็บรายเดือนอาจถูกเปลี่ยนเป็นจัดเก็บรายวัน หรือในบางกรณีเช่น ข้อมูลตำแหน่งของยานพาหนะในกรณีบริษัทขนส่งอาจจัดเก็บในรูปแบบ Real time เหล่านี้ล้วนทำให้ปริมาณข้อมูลเพิ่มขึ้นเป็นอย่างมาก

2) ความหลากหลาย (Variety) การจัดเก็บข้อมูลที่ประจำวันระหว่างกระบวนการทำงานปกติแล้ว ปัจจุบันองค์กรต่าง ๆ ยังเริ่มจัดเก็บข้อมูลที่มีรูปแบบหลากหลายมากขึ้น เช่น ข้อมูลจาก Sensor ต่าง ๆ ลักษณะ Posts บน Social network, รูปภาพหรือไฟล์วิดีโอ, ระยะเวลาการใช้งานบนเว็บไซต์ เป็นต้น

3) ความเร็ว (Velocity) การจัดเก็บความเร็วในที่นี้หมายถึงความเร็วของข้อมูลที่ถูกผลิต ตัวอย่างที่เห็นได้ชัด ได้แก่ ข้อมูลราคาซื้อขายหุ้นของตลาดหลักทรัพย์ที่ราคาของหลักทรัพย์มีการแสดงผลจากธุรกรรมซื้อขายที่เกิดขึ้นในอัตราที่รวดเร็วเป็นอย่างมาก ทำให้ความต้องการใช้ข้อมูลในลักษณะ Realtime เพื่อใช้วิเคราะห์และตอบสนองต่อเหตุการณ์ที่เกิดขึ้นในตลาดหลักทรัพย์เพิ่มสูงขึ้นอย่างมาก

4) ความไม่แน่นอน (Veracity) เนื่องจากข้อมูลล้วนมีที่มาจากหลากหลายแหล่งไม่ว่าจะมาจากภายในหรือภายนอกองค์กร ข้อมูลภายในองค์กรหมายถึง ข้อมูลที่มาจากการประมวลผลภายในหน่วยงาน ยกตัวอย่าง เช่น สถาบันการเงินจะมีข้อมูลจำนวนบัญชีเงินฝากจากสาขา บัญชีเงินกู้จากฝ่ายสินเชื่อ ประวัติการเดินบัญชีส่วนแหล่งข้อมูลภายนอกองค์กร เช่น สภาพการจราจร สภาพอากาศ ปริมาณนักท่องเที่ยว ด้วยสาเหตุนี้ข้อมูล Big Data จึงมีความไม่แน่นอนจากระดับความซับซ้อนและความหลากหลายของข้อมูล

5) ข้อมูลที่มีรูปแบบและไม่มีรูปแบบ (Structured and Unstructured Data) ลักษณะของข้อมูลในอีกมิติหนึ่งคือ ข้อมูลที่มีลักษณะมีรูปแบบ (Structured) และไม่มีรูปแบบ (Unstructured) ข้อมูลที่มีรูปแบบคือ ข้อมูลที่ระบุได้ มีลักษณะและประเภทที่เหมือนกันในพื้นที่จัดเก็บข้อมูลนั้น ๆ และสามารถค้นหาได้ง่ายเช่น ชื่อ เบอร์โทรศัพท์ ที่อยู่เลขบัตรประชาชน เป็นต้น ส่วนข้อมูลที่ไม่มีรูปแบบคือ ข้อมูลไม่สามารถระบุลักษณะได้แม้จะถูกจัดเก็บในรูปแบบเดียวกัน (Format) เช่น Text file Media รูปถ่ายดาวเทียมหรือกล้องวงจรปิด เป็นต้น

การ์ทเนอร์(Gartner, 2019) กำหนดคุณลักษณะสำคัญของ Big data ที่แตกต่างจากข้อมูลทั่วไปอยู่ 3 ประการคือ

1) ปริมาณ (Volume) ข้อมูลมีปริมาณมากมายมหาศาล ซึ่งเกิดจากการรวบรวมจากแหล่งข้อมูล ที่หลากหลาย อัตราการเติบโตของข้อมูลอยู่ในอัตราก้าวหน้า (Exponential) เมื่อเทียบกับเวลาที่ผ่านไปปริมาณข้อมูลเติบโตจากขนาดเทระไบต์ (Terabyte) เป็นเพตะไบต์ (Petabyte) หรือ เอกซะไบต์ (Exabyte) ซึ่งเทคโนโลยีแบบเดิมที่ใช้งานอยู่ไม่สามารถดำเนินการให้เกิดประโยชน์เพื่อการตัดสินใจได้อย่างมีประสิทธิภาพได้

2) ความหลากหลาย (Variety) ข้อมูลมาจากแหล่งที่หลากหลาย จึงทำให้ข้อมูลมีลักษณะและรูปแบบที่มีความหลากหลายเช่นกัน ทั้งแบบมีโครงสร้าง (Structured data) ซึ่งมีลักษณะเป็นตารางข้อมูลที่มีการจัดเรียงอย่างมีรูปแบบชัดเจนและเป็นระเบียบ สามารถนำมาใช้วิเคราะห์ได้เลยเช่น ข้อมูลที่เก็บไว้ในโปรแกรม Spreadsheet หรือ Microsoft Excel แบบกึ่งโครงสร้าง (Semistructured data) ซึ่งเป็นข้อมูลที่ถูกจัดเก็บอย่างมีรูปแบบในระดับหนึ่งและเป็นข้อมูลที่สามารหค้นหา (Search) หรือแท็ก (Tag) ได้เช่น เว็บไซต์ที่มีการระบุชื่อเพจ คำสำคัญ ในเพจ และวัน เดือนปีที่ปรับปรุงแก้ไขข้อมูล และแบบไม่มีโครงสร้าง (Unstructured data) มักจะหมายถึงข้อความยาว ๆ ข้อความเสียง รูปภาพ ภาพเคลื่อนไหว

3) ความรวดเร็ว (Velocity) ข้อมูลที่เกิดจากการใช้อินเทอร์เน็ตสามารถเกิดขึ้นได้อย่างรวดเร็วตลอดเวลา และต่อเนื่อง รวมทั้งระบบการรวบรวมข้อมูลสามารถกระทำได้ในทันที(Real time) กล่าวคือมีการไหลของข้อมูลเข้ามาสู่ระบบอย่างสม่ำเสมอ ทำให้ข้อมูลมีการเปลี่ยนแปลงตลอดเวลาและรวดเร็ว นอกจากนี้ความรวดเร็วอาจหมายถึงความรวดเร็วในการสร้างและประมวลผลข้อมูล

จากการศึกษาลักษณะของข้อมูลมหัต สามารถสรุปได้ว่า ลักษณะของข้อมูลมหัตคือปริมาณมหาศาลความหลากหลายของข้อมูล และความเร็ว ดังภาพ

ปริมาณ	ความหลากหลาย	ความเร็ว
<ul style="list-style-type: none"> <li>• เทระไบต์</li> <li>• เพตะไบต์</li> <li>• เอกซะไบต์</li> </ul>	<ul style="list-style-type: none"> <li>• ข้อมูลแบบมีโครงสร้าง</li> <li>• ข้อมูลแบบกึ่งโครงสร้าง</li> <li>• ข้อมูลแบบไม่มีโครงสร้าง</li> </ul>	<ul style="list-style-type: none"> <li>• การสร้าง</li> <li>• การนำเข้าข้อมูล</li> <li>• การประมวลผลข้อมูล</li> <li>• Real time</li> </ul>

ภาพประกอบที่ 1 แสดงลักษณะของข้อมูลมหัต

## องค์ประกอบของข้อมูลมหัต

องค์ประกอบของระบบข้อมูลมหัต (Big Data) แบ่งออกเป็น 5 ส่วน (coraline, 2021) ได้แก่

1. แหล่งที่มาของข้อมูล (Data Source) แหล่งที่มาของข้อมูล ซึ่งถือได้ว่าเป็นต้นน้ำ เป็นแหล่งกำเนิดของข้อมูล อาจจะเป็นระบบ โปรแกรม หรือจะเป็นมนุษย์เรา ที่สร้างให้เกิดข้อมูลขึ้นมา ทั้งนี้ เมื่อได้ชื่อว่าเป็น Big Data แล้ว ข้อมูลต่างๆ มักจะมาจากแหล่งข้อมูลที่หลากหลาย นำพามาซึ่งความยากลำบากในการจัดการโครงสร้าง หรือจัดเตรียมให้ข้อมูลที่นำมารวมกันนั้น มีความพร้อมใช้ต่อไป

2. การเชื่อมโยงข้อมูล (Gateway) ช่องทางการเชื่อมโยงข้อมูล การเชื่อมโยงข้อมูล เป็นส่วนที่สำคัญมาก และเป็นปัญหาใหญ่ในการทำ Big Data Project ต้องอาศัยทักษะของ Data Engineer ทั้งการเขียนโปรแกรมเอง และใช้เครื่องมือที่มีอยู่มากมาย ทั้งนี้การจะออกแบบช่องทางการเชื่อมโยงข้อมูลได้อย่างสมบูรณ์แบบ จำเป็นต้องทราบก่อนว่า จะนำข้อมูลใดไปทำอะไรต่อบ้าง มิเช่นนั้น การสร้างช่องทางการเชื่อมที่ไม่มีเป้าหมาย ก็อาจเป็นการเสียเวลาโดยเปล่าประโยชน์

3. การเก็บรักษาข้อมูล(Storage) แหล่งเก็บข้อมูลนี้ ไม่ใช่แค่การเก็บข้อมูลจากแหล่งข้อมูล แต่เป็นการเก็บข้อมูลจากแหล่งข้อมูลหลายๆ แหล่งเอามาไว้เพื่อรอการใช้งาน ซึ่งอาจจะเป็นที่พักข้อมูลให้พร้อมใช้ หรือจะเป็นแหล่งเก็บข้อมูลในอดีตก็เป็นได้

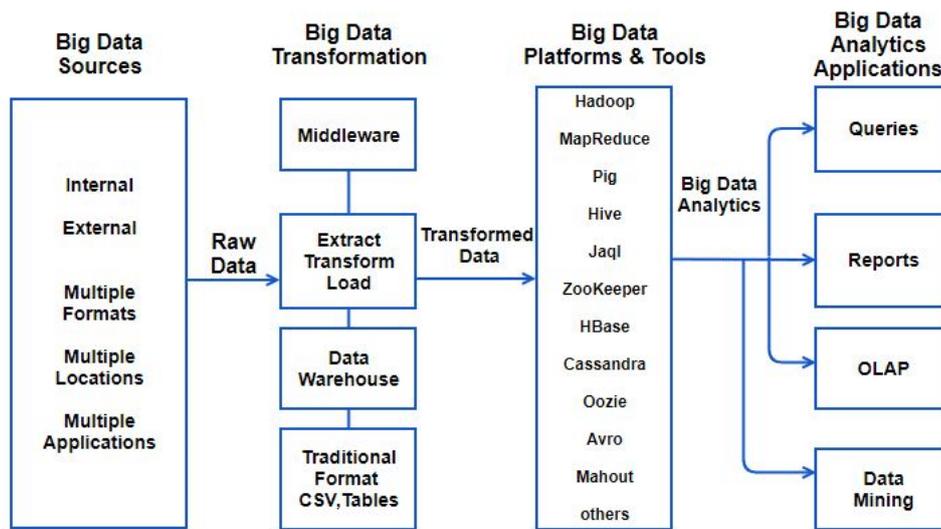
4. การวิเคราะห์ข้อมูล (Analytics) การวิเคราะห์ข้อมูล ส่วนนี้เป็นหน้าที่หลักของ Data Scientist ซึ่งแบ่งงานออกเป็น 2 ลักษณะ คือ การวิเคราะห์เบื้องต้น โดยการใช้วิธีทางสถิติ หรือจะเป็นการวิเคราะห์เชิงลึกโดยการสร้าง Model แบบต่างๆ รวมไปถึงการใช้ Machine Learning เพื่อให้ได้ผลลัพธ์เฉพาะเจาะจงในแต่ละปัญหา และแต่ละชุดข้อมูล

5. การนำข้อมูลไปใช้ประโยชน์ (Result/Action) การใช้ผลการวิเคราะห์ข้อมูลผลลัพธ์ที่ได้จากการวิเคราะห์สามารถนำไปใช้งานได้ 2 รูปแบบ ทั้งในรูปแบบของ Dashboard, Report กล่าวคือ ออกเป็นรายงาน เพื่อให้ Data Analyst นำผลลัพธ์ที่ได้ไปใช้กับงานทางธุรกิจต่อไป หรือจะเป็นการนำไปกระทำเลยโดยที่ไม่ต้องมี “มนุษย์” คอยตรวจสอบ ซึ่งจำเป็นต้องมีการเขียนโปรแกรมเพิ่ม เพื่อให้มีการกระทำออกไป ที่เรียกว่า Artificial Intelligence (AI)

### กรอบงานข้อมูลมหัต (Big Data Framework)

กรอบงานข้อมูลมหัตมีการออกแบบ และนำไปใช้ในหลายองค์กร ทั้งนี้ การออกแบบ และสร้างระบบจัดการข้อมูลมหัต มีลักษณะคล้ายกับการพัฒนาระบบ (System Approach) ในที่นี้จะนำเสนอตัวอย่างดังต่อไปนี้

1) กรอบงานทั่วไปของข้อมูลมหัต (General Framework of Big Data Analytics) แมนดาวา และคณะ (Mandava Geetha Bhargava, Modugula T S Srinivasa Reddy, Shaik Shahbaz, P. Venkateswara Rao & V. Sucharita, 2017) ได้เสนอกรอบงานทั่วไปของข้อมูลมหัต ประกอบด้วย 4 องค์ประกอบหลักดังภาพ



ภาพประกอบที่ แสดงกรอบงานทั่วไปของข้อมูลมหัต

ที่มา : Global Journal of Computer Science and Technology: Software & Data Engineering. Volume 17

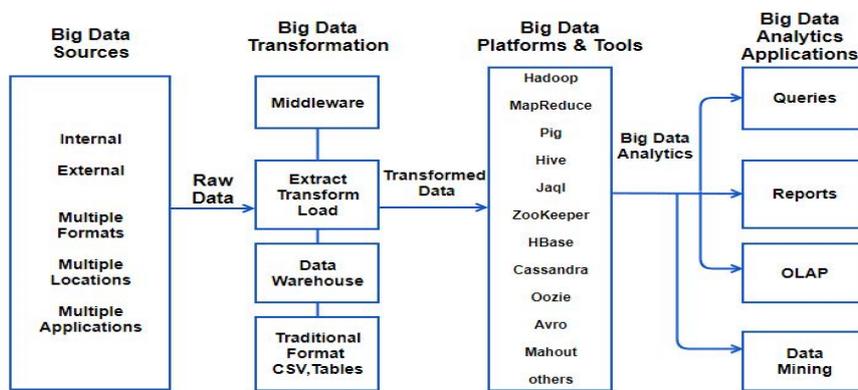
จากภาพ

องค์ประกอบที่ 1 ทรัพยากรข้อมูลมหัต (Big Data Sources) หมายถึงข้อมูล จากภายใน ข้อมูลจากภายนอก ข้อมูลดิจิทัลหลากหลายรูปแบบ ข้อมูลจากหลายแหล่ง และข้อมูลจากหลายแอปพลิเคชัน (Multiple Application)

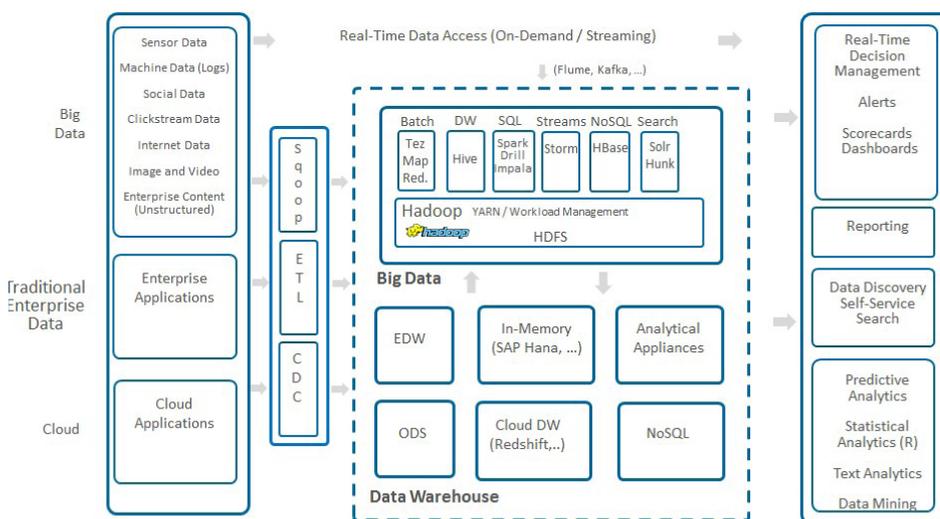
องค์ประกอบที่ 2 การเชื่อมโยงข้อมูลมหัต (Big Data Transformation) หมายถึง คลื่นกลาง (Middle wave), การแยกโหลดการแปลง (Extract Transform load), คลังข้อมูล (Data Warehouse) , รูปแบบดั้งเดิม (Traditional format) และตาราง CSV (CSV table)

องค์ประกอบที่ 3 แพลตฟอร์มบีกดต้าและเครื่องมือ (Big Data Platform & Tool) หมายถึง เครื่องมือปฏิบัติการระบบข้อมูลมหัต ยกตัวอย่างเช่น Hadoop, MapReduce, Pig, Hive, Jaql, Zookeeper, Hbase, Cassandra, Oozie, Avro, Mahout, และอื่น ๆ

องค์ประกอบที่ 4 แอปพลิเคชันการวิเคราะห์ข้อมูลมหัต (Big Data Analytics Application) หมายถึง แบบสอบถาม (Queries), การรายงาน (Report), โอลแลป(Olap) และการทำเหมืองข้อมูล(Data Mining)



2) อัลเบอโต แพน (Alberto Pan, 2016) ได้เสนอ สถาปัตยกรรมเชิงตรรกะสำหรับการวิเคราะห์ข้อมูลมหัต (Logical Architectures for Big Data Analytics) ประกอบด้วย 4 องค์ประกอบหลักดังภาพ



ภาพประกอบที่ แสดงสถาปัตยกรรมเชิงตรรกะสำหรับการวิเคราะห์ข้อมูลมหัต  
ที่มา : <https://www.datavirtualizationblog.com/logical-architectures-big-data-analytics/>

## การวิเคราะห์ข้อมูลมหัต

การวิเคราะห์ข้อมูลมหัตด้วยรูปแบบ และกระบวนการที่เหมาะสมทำให้สามารถนำผลการวิเคราะห์มาใช้ประโยชน์ได้บรรลุเป้าหมายตามที่กำหนดไว้ ดังรายละเอียดต่อไปนี้

5.6.1 รูปแบบของการวิเคราะห์ข้อมูลมหัตสามารถแบ่งการวิเคราะห์ได้ดังนี้ (Bekker, 2019)

1) การวิเคราะห์เชิงพรรณนา (Descriptive analytics) เป็นการวิเคราะห์ที่ง่ายที่สุดเพื่อแสดงผลของเหตุการณ์หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้น ในลักษณะที่ง่ายต่อการทำความเข้าใจ หรือการตัดสินใจ เช่น ความถี่ของการเข้าชมเว็บไซต์ ยอดขายเฉลี่ยต่อสัปดาห์ กำไรในแต่ละเดือน เป็นต้น อย่างไรก็ตาม ข้อค้นพบจากการวิเคราะห์รูปแบบนี้ชี้ให้เห็นว่า การค้นพบนี้ส่งสัญญาณว่ามีบางอย่างผิดปกติหรือถูกต้อง โดยไม่ต้องอธิบายสาเหตุ ด้วยเหตุผลนี้ควรรวมเข้ากับการวิเคราะห์ข้อมูลประเภทอื่นๆ เพื่ออธิบายเหตุผล

2) การวิเคราะห์เชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อสิ่งที่เกิดขึ้น เช่น ความสัมพันธ์ระหว่างอัตราค่าเฉลี่ยไมตรีของผู้สอนกับการแสดงออกทางการเรียนของผู้เรียน ซึ่งช่วยทำให้สามารถตัดสินใจไปในทางที่ถูกต้อง เพื่อนำมาใช้ในการพิจารณาว่าเหตุใดจึงไม่ได้รับกำไรตามเป้าหมายที่กำหนดไว้ การวิเคราะห์เชิงวินิจฉัยให้ข้อมูลเชิงลึกเกี่ยวกับปัญหาเฉพาะ ดังนั้นองค์กรควรรวบรวมข้อมูลเชิงลึกโดยละเอียด เพื่อให้ได้ผลการวิเคราะห์ข้อมูลที่ตรงตามวัตถุประสงค์ที่กำหนดไว้

3) การวิเคราะห์เชิงพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อคาดการณ์สิ่งที่กำลังจะเกิดขึ้น หรือแนวโน้ม ต่าง ๆ ในอนาคต ใช้การวิเคราะห์เชิงพรรณนาและการวินิจฉัยเพื่อ และบริบท ประกอบกัน และเพื่อคาดการณ์แนวโน้มในอนาคต เครื่องมือที่มีค่าสำหรับการคาดการณ์ โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วกับแบบจำลองทางสถิติหรือเทคโนโลยีปัญญาประดิษฐ์ (Artificial intelligence) เช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ การพยากรณ์ความเสี่ยงในการลงทุนเพื่อขยายกิจการ เป็นต้น การวิเคราะห์แบบพยากรณ์เป็นการวิเคราะห์ขั้นสูงและนำเสนอข้อดีมากมาย เช่น การวิเคราะห์ที่ซับซ้อน หรือการเรียนรู้เชิงลึก และแนวทางเชิงรุกที่สามารถคาดการณ์ได้ อย่างไรก็ตาม การพยากรณ์เป็นเพียงการประมาณการ ซึ่งความถูกต้องแม่นยำนั้นขึ้นอยู่กับคุณภาพของข้อมูลและความเสถียรของสถานการณ์เป็นอย่างมาก ดังนั้นจึงต้องมีการดูแลอย่างระมัดระวัง และการเพิ่มประสิทธิภาพอย่างต่อเนื่อง

4) การวิเคราะห์เชิงแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด จุดประสงค์ของการวิเคราะห์เชิงแนะนำคือเพื่อให้ข้อเสนอแนะว่าต้องดำเนินการใดเพื่อขจัดปัญหาในอนาคตหรือใช้ประโยชน์จากแนวโน้มที่มีแนวโน้มจะเกิดขึ้นอย่างสูงสุด ที่คาดว่าจะเกิดการวิเคราะห์รูปแบบนี้เป็นทั้งการพยากรณ์สิ่งที่เกิดขึ้นในอนาคต การวิเคราะห์เชิงแนะนำใช้เครื่องมือและเทคโนโลยีขั้นสูง นอกจากนี้ การวิเคราะห์ข้อมูลประเภทที่ล้ำสมัยนี้ไม่เพียงต้องการข้อมูลภายในในอดีตเท่านั้น แต่ยังต้องใช้ข้อมูลภายนอก การประเมินข้อดีข้อเสียสาเหตุระยะเวลาระยะของสิ่งที่เกิดขึ้น และการให้คำแนะนำทางเลือกต่าง ๆ ที่มีอยู่รวม

ถึงผลที่จะเกิดขึ้นของแต่ละทางเลือกเช่น การระบุโอกาสที่ผู้เรียนจะเข้าเรียนตรงเวลา การระบุโอกาสในการซื้อสินค้าหรือบริการซ้ำ เหล่านี้จะพิจารณาจากประวัติ การซื้อสินค้าของลูกค้าแต่ละคน โดยอาศัยข้อมูลเชิงประวัติและข้อมูลจากภายนอก มาใช้ในการวิเคราะห์

5.6.2 กระบวนการวิเคราะห์ข้อมูลมหัต การวิเคราะห์ข้อมูลจาก Big data หรือ Data analytics เริ่มต้นจากการนำข้อมูลมาทำให้อยู่ในรูปแบบไฟล์ดิจิทัลที่พร้อม จะประมวลผลได้ด้วยเทคโนโลยีหรือชุดคำสั่ง และแบบจำลองที่สร้างขึ้นเพื่อนำข้อมูล ที่ได้ผ่านการวิเคราะห์แล้วมาใช้หรือแปลความหมาย รายละเอียดดังต่อไปนี้ (coraline, 2021)

1) ความเข้าใจเกี่ยวกับธุรกิจ (Business Understanding) นั่นคือ การทำความเข้าใจ และวางเป้าหมายเชิงธุรกิจ จะเห็นได้ว่า ผู้มีหน้าที่ในการวาง Scope ของการทำโครงการ คือ Business Analyst แต่การจะวาง Scope ของการทำ โครงการได้ จำเป็นต้องทราบถึงแนวทางในการแก้ปัญหาขั้นต้น รวมไปถึง สามารถ มองเห็นในสิ่งที่ต้องการได้

2) ความเข้าใจเกี่ยวกับ (Data Understanding) คือ ศึกษาแนวทางการได้มาซึ่งข้อมูล รวมไปถึง การเลือกปัจจัยที่คิดว่ามีส่วนสำคัญในการประมวลผล ในส่วนนี้ ผู้มีหน้าที่หลัก คือ Data Engineer เพราะเป็นขั้นตอนในการวางโครงสร้าง ตั้งแต่ต้นน้ำ จนถึงการจัดเก็บให้อยู่ในรูปแบบที่พร้อมใช้

3) การเตรียมข้อมูล (Data Preparation) คือ การจัดการข้อมูล ให้อยู่ในรูปแบบที่พร้อมนำไปวิเคราะห์ต่อ รวมไปถึงการทำความสะอาดข้อมูล หน้าที่นี้ Data Engineer และ Data Scientist จะต้องทำด้วยกัน ที่สำคัญที่สุด ขั้นตอนนี้จะ เป็นการกระทำย้อนไปย้อนมา ร่วมกับการสร้าง Model หมายถึงว่า จะทำต้อง วนไปวนมา จนกว่าจะแน่ใจว่า Model ที่ได้ เป็น Model ที่ตอบโจทย์ที่สุด จึงเป็น ที่มาว่า การจะทำความสะอาดข้อมูล จะต้องเกิดจากการสร้าง Model เพราะทราบได้ ที่ยังไม่เคยนำข้อมูลไปใช้จริง ก็คงไม่สามารถทราบได้ว่า จะต้องทำความสะอาด หรือ จัดเตรียมให้อยู่ในรูปแบบใด

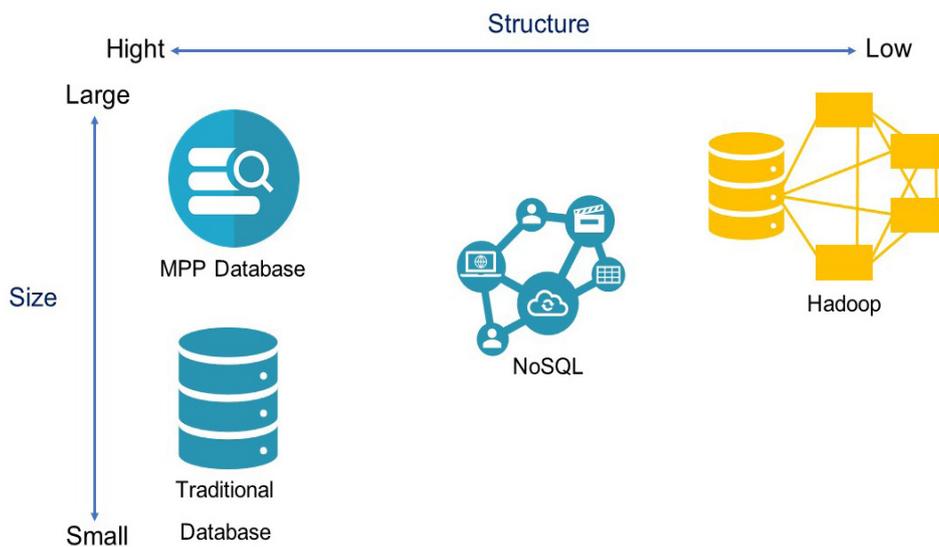
4) การสร้างแบบจำลอง(Modeling) คือ การสร้างสมการที่ซับซ้อน เหตุผลที่ต้องสร้าง แบบจำลอง (Model) เพราะธุรกิจต้องการความเป็นเลิศ ต้องการ ผลลัพธ์ที่เป็นเอกเทศ เพื่อให้เกิดการแข่งขันทางการตลาด ในการทำ Data Science ไม่มี Model ไหนเป็น Model ที่ดีที่สุด จึงจำเป็นต้องมีการสร้าง Model หลายๆ แบบ เพื่อนำไปสู่กระบวนการประเมินผลลัพธ์ที่ได้จาก Model ในลำดับต่อไป

5) การประเมินผล (Evaluation) คือ การประเมินผลลัพธ์ที่ได้จาก Model แต่ละโจทย์จะมีความต้องการที่ต่างกัน เช่น ต้องการ Model ที่ทำนายได้แม่นยำ ที่สุด ต้องการ Model ที่ประมวลผลได้เร็วที่สุด หรือต้องการ Model ที่นำไปใช้ได้ ง่ายที่สุด และเช่นกัน ขั้นตอนนี้ ผู้มีหน้าที่สำคัญ คือ Business Analyst (ในขณะที่ Data Scientist เป็นผู้นำเสนอ Model แต่ละประเภทให้ Business Analyst เป็นผู้เลือก)

6) การปรับใช้ (Deployment) คือ การนำ Model ไปใช้งานจริง อาจจะเป็นการสร้างระบบขึ้นมาเพื่อให้เกิดการทำงานอย่างอัตโนมัติ หรือจะเป็น การใช้ Model เฉพาะเมื่อจำเป็นต้องใช้ แตกต่างกันตรงที่ หากเป็นระบบอัตโนมัติ ก็จำเป็นต้องนำไปขึ้นเป็นระบบ (คล้ายๆ การทำ Software หรือ อาจจะเป็นการทำ Software หรือ Application รูปแบบหนึ่งก็เป็นได้) ซึ่งผู้ที่ทำหน้าที่ให้ Requirement หรือ ออกแบบช่องทางการใช้งาน คือ Business Analyst เพราะถือว่า เป็นการส่งมอบ งาน และนำ Model นั้นไปใช้จริงในธุรกิจ

## การกำหนดเทคโนโลยีด้านข้อมูลมหัต (Big Data) ขององค์กร

การทำระบบข้อมูลมหัต (Big Data) จำเป็นต้องกำหนดวัตถุประสงค์ และแนวทางการดำเนินการให้ชัดเจนก่อนการกำหนดเทคโนโลยีที่จะต้องนำมาใช้งานให้สอดคล้องตามวัตถุประสงค์ ซึ่งเทคโนโลยีแต่ละแบบจะมีความเหมาะสมกับข้อมูลที่แตกต่างกัน (ชนชาติ นุ่มนนท์, 2557) ดังภาพ



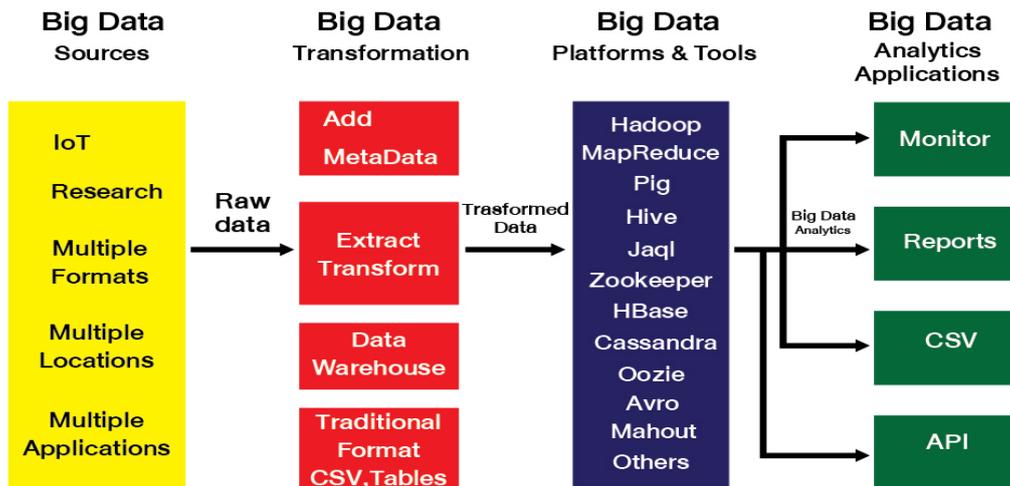
ภาพประกอบที่ 3 เปรียบเทียบเทคโนโลยีการเก็บข้อมูลแบบต่าง ๆ

จากภาพ

- 1) Traditional Database คือเทคโนโลยีฐานข้อมูล SQL แบบเดิม สำหรับข้อมูลที่เป็น โครงสร้าง (structure) ในระดับ (Gigabyte) ถึงเทระไบต์ (Terabyte) และมีความเร็วในการประมวลผลไม่มากนัก
- 2) MPP Database คือเทคโนโลยีสำหรับข้อมูลขนาดใหญ่หลายเทระไบต์ (Terabyte) ที่เป็นโครงสร้าง (structure) โดยมีความสามารถในการประมวลผลข้อมูลขนาดใหญ่ได้อย่างรวดเร็ว ตัวอย่างของ MPP มี อาทิเช่น Oracle Exadata, SAP HANA, Amazon Redshift หรือ Datawarehouse อย่าง Teradata หรือ Greenplum
- 3) NoSQL คือเทคโนโลยีในการเก็บข้อมูลแบบกึ่งโครงสร้าง (semi-structure) ขนาดใหญ่ โดยไม่ได้ใช้คำสั่งในการประมวลผลที่เป็น SQL ตัวอย่าง เช่น mongo DB, Cassandra หรือ Dynamo DB
- 4) Hadoop คือเทคโนโลยีในการเก็บข้อมูลที่เป็นแบบไม่มีโครงสร้าง (unstructured) ซึ่งสามารถจะเก็บข้อมูลขนาดใหญ่ได้เป็นเพตะไบต์ (Petabyte)

## ข้อมูลมห้ดกับการพัฒนาศักยภาพกำลังคน

ข้อมูลมห้ด (Bigdata) หมายถึงข้อมูลหลายหลายมีมวลมหาศาลและเกิดขึ้นรวดเร็ว ส่วนระบบข้อมูลมห้ด หมายถึงโครงสร้างสำหรับการจัดการข้อมูลมห้ด หรือบางครั้งเรียกรวมงานข้อมูลมห้ด (Big Data Framework) อยางไรก็ตาม ข้อมูลมห้ด สำหรับการพัฒนาศักยภาพกำลังคน หมายถึงฐานข้อมูลหลากหลายที่มีความเกี่ยวข้งและสัมพันธ์กับการพัฒนาศักยภาพกำลังคน รายละเอียดดั่งภาพ



จากภาพ ประกอบด้วยองค์ประกอบหลัก 4 ส่วนสำคัญ ดังนี้

องค์ประกอบที่ 1 หรือพยากรข้อมูลมห้ด (Big Data Sources) หมายถึงข้อมูลดิบจากแหล่งข้อมูลดั่งต่อไปนี้

1.1) ข้อมูลจากอินเทอร์เน็ต (Internet Of Thing : IoT) หมายถึง ข้อมูลจากอุปกรณ์ สิ่งของเครื่องใช้ที่แตกต่างและหลากหลายที่เชื่อมต่อกับเครือข่าย และสามารถเก็บบันทึกและแลกเปลี่ยนข้อมูลกันได้

1.2) ข้อมูลจากการวิจัย(Research Data) หมายถึง ข้อมูลที่นักวิจัยใช้ในการทำรายงานวิจัยตามที่สอดคล้องกับการพัฒนากำลังคนรองรับ EEC นับตั้งแต่เริ่มเกิดคำถามหรือปัญหาการวิจัย จนถึงได้ผลงานวิจัยออกมาเป็นฉบับสมบูรณ์

1.3) ข้อมูลดิจิทัลหลากหลายรูปแบบ (Multiple Formats) หมายถึง ข้อมูลจำนวนมากมหาศาล ทุกเรื่อง ทุกแง่มุม ทุกรูปแบบ ซึ่งอาจเป็นข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่เก็บอยู่ในตารางข้อมูลต่างๆ หรืออาจเป็นข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data) เช่น ล็อกไฟล์ (Log files) หรือแม้กระทั่งข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลการโต้ตอบปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) เช่น Facebook, twitter หรือ โพสต์จำพวกมีเดีย

1.4) ข้อมูลจากหลายแหล่ง (Multiple Locations) หมายถึง ข้อมูลจากแหล่งข้อมูลหลากหลาย เช่น กำลังคน ผู้ประกอบการ องค์กร สถาบันพัฒนากำลังคน และภาคีเครือข่าย (Partners) เป็นต้น

1.5) ข้อมูลจากหลายแอปพลิเคชัน (Multiple Application) หมายถึง ข้อมูลจากแอปพลิเคชันต่างๆ เช่น Google Form, Google Chome, MOOCs, Youtube, Facebook, twitter และวารสารวิชาการออนไลน์เข้าถึงได้จาก [https://so04.tci-thaijo.org/index.php/MPDIJ\\_EEC/index](https://so04.tci-thaijo.org/index.php/MPDIJ_EEC/index) เป็นต้น

องค์ประกอบที่ 2 การเชื่อมโยงข้อมูลมหัต (Big Data Transformation) หมายถึง ขั้นตอนในการสกัดข้อมูลดิบ เพื่อให้เป็นข้อมูลที่เป็นประโยชน์

2.1) บอกรายละเอียดของข้อมูล (Add Metadata) หมายถึงข้อมูลที่ใช้บ่งชี้ บอกลักษณะ หรือให้ข้อมูลเกี่ยวกับข้อมูล ซึ่งมีอยู่มากมายหลายลักษณะ สำหรับการพัฒนาศักยภาพกำลังคน ได้นำเมตาดาต้า(Metadata) มาช่วยในการลงรายการเกี่ยวกับกำลังคน ข้อมูลศักยภาพกำลังคน ความต้องการใช้กำลังคนที่มีศักยภาพ และการพัฒนาศักยภาพกำลังคนต่างๆ ที่เพิ่มจำนวนขึ้นอย่างรวดเร็ว ทั้งทรัพยากรสารสนเทศที่เป็นดิจิทัล และไม่เป็นดิจิทัล เนื่องจากเมตาดาต้า (Metadata) ช่วยเพิ่มความสามารถในการค้นคืน และเข้าถึงข้อมูลเกี่ยวกับกำลังคน ข้อมูลศักยภาพกำลังคน ความต้องการใช้กำลังคนที่มีศักยภาพ และการพัฒนาศักยภาพกำลังคนต่างๆ ช่วยสงวนรักษาบริบทอื่นๆ ที่สำคัญของทรัพยากรสารสนเทศเกี่ยวกับการพัฒนาศักยภาพกำลังคน ขยายขอบเขตเกี่ยวกับการพัฒนาศักยภาพกำลังคน และบ่งชี้คุณสมบัติเฉพาะของเกี่ยวกับการพัฒนาศักยภาพกำลังคน มาตรฐานคุณวุฒิและระดับความสามารถต่างๆ ซึ่งจะทำให้เกิดความเข้าใจที่ตรงกัน ทำให้สะดวก และไม่เกิดความสับสนในการใช้ทรัพยากรสารสนเทศเกี่ยวกับการพัฒนาศักยภาพกำลังคน

2.2) การแยกโหลดการแปลง (Extract Transform load) หมายถึง การนำข้อมูลจากหลากหลายแหล่งข้อมูล เช่น จากอินเทอร์เน็ต จากการจัดเก็บจากผู้ใช้งาน จากแอปพลิเคชัน มาแปลงรูปแบบให้เหมาะสม (Transform) แล้วเก็บ (Load) เข้าไปในดาต้าแวร์เฮาส์ (Data Warehouse) หรือ ดาต้าเลค (Data Lake)

2.3) คลังข้อมูล (Data Warehouse) หมายถึง การสร้างที่เก็บข้อมูลส่วนกลางสำหรับนำไปวิเคราะห์ เป็นคลังเก็บข้อมูลที่ผ่านมากระบวนการสารสนเทศแล้ว

2.4) รูปแบบไฟล์ดั้งเดิม (Traditional format) และตาราง CSV (CSV table) หมายถึง เทคโนโลยีจัดการข้อมูลแบบเดิมสำหรับข้อมูลที่เป็นโครงสร้าง (Structure) ในระดับกิกะไบต์ (Gigabyte) ถึง เทระไบต์(Terabyte) และมีความเร็วในการประมวลผลไม่มากนัก อย่างไรก็ตามรูปแบบไฟล์ (File Formats) ที่ใช้กับงานข้อมูลมหัต (Big Data) ที่นิยมใช้ 4 สกุล (ยุทธการ อังมณีกุล, 2563) ดังต่อไปนี้

- CSV ย่อมาจาก Comma-Separated Values เป็น text file ที่เก็บข้อมูลแบบตาราง โดยเป็นแบบ row-based คือ แต่ละบรรทัดเท่ากับ 1 row ในตาราง และแต่ละ column คั่นด้วยเครื่องหมาย comma (,) ซึ่งจะถูกเรียกว่า เดลิมิเตอร์ (Delimiter) ไฟล์สกุลCSV ไฟล์เดี่ยวไม่สามารถแสดงความสัมพันธ์ของข้อมูลได้ ดังนั้นจำเป็นต้องใช้ ไฟล์สกุลCSV หลายไฟล์ๆ มาประกอบกันเพื่อเชื่อมความสัมพันธ์

- JSON ย่อมาจาก JavaScript Object Notation เก็บข้อมูลเป็นลำดับชั้น (Hierarchy) ทำให้สามารถเก็บความสัมพันธ์ของข้อมูลไว้ในไฟล์เดียวกันได้เลย ไม่ต้องเก็บหลายไฟล์เหมือนไฟล์สกุล CSV

- Parquet เป็น binary file แบบ column-based เก็บข้อมูลแบบ column-based จะเก็บข้อมูลทั้งคอลัมน์เรียงต่อกันไปที่ละ disk block เป็น binary file เราก็จะเปิดอ่านและแก้ไขข้อมูลตรงๆ เหมือน CSV และ JSON ไม่ได้ ซึ่งอาจจะดูไม่สะดวกนัก แต่ในงาน Big Data เรามักจะใช้ Apache Spark หรือ Python ในการจัดการข้อมูล

- Avro เก็บข้อมูลเป็น Binary file แบบ row-based และเก็บ meta-data หรือ schema ในรูปแบบ JSON ที่ส่วนต้นของไฟล์

องค์ประกอบที่ 3 แพลตฟอร์มบีเก็ดต้าและเครื่องมือ (Big Data Platform & Tool) หมายถึง เครื่องมือปฏิบัติการระบบข้อมูลมหัด ยกตัวอย่างเช่น Hadoop, MapReduce, Pig, Hive, Jaql, Zookeeper, Hbase, Cassandra, Oozie, Avro, Mahout, และอื่น ๆ

องค์ประกอบที่ 4 แอปพลิเคชันการวิเคราะห์ข้อมูลมหัด (Big Data Analytics Application) หมายถึง การกระบวนการวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อค้นหารูปแบบความสัมพันธ์ของข้อมูลเหล่านั้นที่ซ่อนอยู่ข้างใน ประกอบด้วย

1) การแสดงผล (Monitor) หมายถึง การนำเสนอข้อมูลของระบบทางหน้าจอ

2) การรายงาน (Report) หมายถึง การแสดงผลลัพธ์ตามวัตถุประสงค์ของการวิเคราะห์ข้อมูล ประกอบไปด้วย รูปแบบการวิเคราะห์ข้อมูล 4 รูปแบบดังนี้

2.1) การวิเคราะห์เชิงพรรณนา (Descriptive analytics) เป็นการอธิบายสภาพการณ์ของเหตุการณ์หรือกิจกรรมต่าง ๆ เกี่ยวกับกำลังคน ความต้องการกำลังคน การพัฒนาศักยภาพกำลังคน ในลักษณะที่ง่ายต่อการทำความเข้าใจ

2.2) การวิเคราะห์เชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกำลังคน ความต้องการกำลังคน การพัฒนาศักยภาพกำลังคน

2.3) การวิเคราะห์เชิงพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อคาดการณ์สิ่งที่กำลังจะเกิดขึ้น หรือแนวโน้มต่าง ๆ เกี่ยวกับกำลังคน ความต้องการกำลังคน การพัฒนาศักยภาพกำลังคน

2.4) การวิเคราะห์เชิงแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์เพื่อให้ข้อเสนอแนะว่าต้องดำเนินการใดเกี่ยวกับกำลังคน ความต้องการกำลังคน การพัฒนาศักยภาพกำลังคน เพื่อขจัดปัญหาได้ในอนาคต

3) การสร้างคุณค่าร่วมกัน (Creating shared value : CSV) หมายถึง แนวทางการดำเนินธุรกิจที่ให้ความสำคัญกับสร้างคุณค่าของข้อมูลมหัดให้กับองค์กรและสังคมไปพร้อมกันเพื่อความสำเร็จในระยะยาว บทบาทของระบบข้อมูลมหัด ต่อสังคมบนฐานคิดแบบระยะยาว (Long-term Thinking) เน้นการสร้างประโยชน์แก่องค์กรและสังคมให้เกิดเป็นผลประโยชน์ร่วมกัน

4) ช่องทางการเชื่อมต่อ (Application Programming Interface : API) หมายถึง การเชื่อมต่อจากระบบหนึ่งไปสู่อีกระบบหนึ่ง เพื่อให้ซอฟต์แวร์ภายนอกเข้าถึงและอัปเดตข้อมูลนั้นๆได้ แต่ยังคงอยู่ในขอบเขตที่ถูกกำหนดไว้

## ธรรมาภิบาลข้อมูลกับการพัฒนาศักยภาพกำลังคน

ข้อมูลจัดเป็นทรัพย์สินที่สำคัญในการดำเนินงานขององค์กร โดยเฉพาะข้อมูลมหัต (Big Data) ซึ่งถูกนำมาใช้ประโยชน์ทั้งในเชิงเศรษฐกิจและสังคม ดังนั้นการกำหนดกฎกติกาและขอบเขตของอำนาจในการตัดสินใจที่เกี่ยวข้องกับข้อมูล หรือที่เรียกว่าธรรมาภิบาลข้อมูล (Data Governance) จึงเป็นเรื่องสำคัญ การจัดทำธรรมาภิบาลข้อมูล คือการกำกับดูแลข้อมูลรวมถึง กระบวนการ บุคลากร และเทคโนโลยีที่จำเป็นในการจัดการและปกป้องสินทรัพย์ข้อมูลขององค์กร เพื่อรับประกันว่าข้อมูลนั้นมีความถูกต้อง ความสมบูรณ์ ความปลอดภัย สามารถเชื่อถือและค้นหาได้ง่าย กฎหมายที่เกี่ยวข้องกับข้อมูล ข่าวสาร หรือสิทธิส่วนบุคคล ในประเทศไทย มีประเด็นที่อาจส่งผลกระทบต่อหลักแนวคิดธรรมาภิบาลข้อมูล ภาครัฐ ซึ่งหากหน่วยงานในประเทศไทยต้องการสร้างหรือปรับปรุงระบบภายใน ให้มีธรรมาภิบาลข้อมูลภาครัฐ จะต้องพิจารณาประเด็นหลัก ๆ ดังนี้ (ราชกิจจานุเบกษา, 2563)

- 1) การกำหนดสิทธิ หน้าที่ และความรับผิดชอบของผู้ซึ่งมีหน้าที่เกี่ยวข้องกับการบริหารจัดการข้อมูลของหน่วยงาน
- 2) การวางแผนการดำเนินงาน การปฏิบัติตามแผนการดำเนินงาน การตรวจสอบและการรายงานผลการดำเนินงาน และการปรับปรุงแผนการดำเนินงานอย่างต่อเนื่อง เพื่อให้ระบบบริหารและกระบวนการจัดการข้อมูลมีประสิทธิภาพ สามารถเชื่อมโยง แลกเปลี่ยน และบูรณาการข้อมูลระหว่างกันทั้งภายในและภายนอกหน่วยงาน และคุ้มครองข้อมูลให้มีประสิทธิภาพ
- 3) การกำหนดมาตรการควบคุมและพัฒนาคุณภาพข้อมูล เพื่อให้ข้อมูลมีความถูกต้องครบถ้วน เป็นปัจจุบัน มั่นคงปลอดภัย และไม่ถูกละเมิดความเป็นส่วนตัว รวมทั้งสามารถเชื่อมโยงแลกเปลี่ยน บูรณาการ และใช้ประโยชน์ได้อย่างมีประสิทธิภาพ
- 4) การวัดผลการบริหารจัดการข้อมูล โดยอย่างน้อยประกอบด้วย การประเมินความพร้อมของธรรมาภิบาลข้อมูลภาครัฐในระดับหน่วยงาน การประเมินคุณภาพข้อมูล และการประเมินความมั่นคงปลอดภัยของข้อมูล
- 5) การกำหนดหมวดหมู่ของข้อมูล เพื่อกำหนดนโยบายข้อมูลหรือกฎเกณฑ์เกี่ยวกับผู้มีสิทธิเข้าถึงและใช้ประโยชน์จากข้อมูลต่างๆ ภายในหน่วยงาน สำหรับให้ผู้ซึ่งมีหน้าที่เกี่ยวข้องปฏิบัติตามนโยบายหรือกฎเกณฑ์ได้อย่างถูกต้อง และสอดคล้องตามกฎหมายที่เกี่ยวข้อง อันจะนำไปสู่การบริหารจัดการข้อมูลภาครัฐอย่างเป็นระบบ
- 6) การจัดทำคำอธิบายชุดข้อมูลดิจิทัลของภาครัฐและบัญชีข้อมูลให้มีความถูกต้อง ครบถ้วนและเป็นปัจจุบัน

## สรุปข้อมูลมหัศจรรย์กับการพัฒนาศักยภาพกำลังคน

ข้อมูลมหัศจรรย์ (Bigdata) ตามรูปศัพท์อธิบายถึงปริมาณข้อมูลที่มีมหาศาลอย่างไรก็ตามโดยทั่วไปข้อมูลมหัศจรรย์เป็นโครงสร้างพื้นฐานที่รองรับการนำเข้าข้อมูล การเก็บข้อมูล และการวิเคราะห์ข้อมูลเรียกอีกอย่างว่า ศาสตร์ข้อมูล สิ่งสำคัญสำหรับการพัฒนาศักยภาพกำลังคนไม่ใช่ปริมาณของข้อมูลเท่านั้น หากแต่สิ่งสำคัญก็คือการที่องค์กรจัดการกับข้อมูลเกี่ยวกับสภาพปัจจุบันของกำลังคน ความต้องการกำลังคน การผลิตกำลังคน ข้อมูลที่นักวิจัยใช้ในการทำรายงานวิจัยตามที่สอดคล้องกับการพัฒนากำลังคน จากแหล่งข้อมูลหลากหลาย เช่น กำลังคน ผู้ประกอบการ องค์กร สถาบันพัฒนากำลังคน และภาคีเครือข่าย (Partners) เป็นต้น ข้อมูลมหาศาลเหล่านี้นำไปสู่ข้อมูลเชิงลึกเพื่อการตัดสินใจที่ดีกว่า การใช้ข้อมูลขนาดใหญ่และการวิเคราะห์เชิงพยากรณ์เป็นกุญแจสำคัญในการทำความเข้าใจวิธีการสร้างสรรค์ผลิตภัณฑ์และผลลัพธ์ที่ได้ ซึ่งจะช่วยลดต้นทุน ลดเวลา พัฒนาศักยภาพกำลังคนเต็ม เพิ่ม ใหม่ โดยกำหนดข้อเสนอที่ดีที่สุด และ ตัดสินใจอย่างชาญฉลาด เพื่อการพัฒนาศักยภาพกำลังคนให้บรรลุผลได้อย่างมีประสิทธิภาพ

### เอกสารอ้างอิง

- กรมการจัดหางาน (2562) *การจัดการความรู้ : ความรู้เบื้องต้นเกี่ยวกับ Big Data และ Machine Learning*. กรมการจัดหางาน: กรุงเทพฯ
- คอราไลน์. (2021). *Data Science Process Cycle*. (ออนไลน์) ได้จาก <https://www.coraline.co.th/single-post/2018/11/26/data-science-process-cycle> (สืบค้นเมื่อ 1 มีนาคม 2564)
- คอราไลน์. (2021). *5 ลำดับขั้นในการดำเนินโครงการ Big Data*. (ออนไลน์) : ได้จาก <https://www.coraline.co.th/single-post/5-steps-to-big-data-project> (สืบค้นเมื่อ 1 มีนาคม 2564)
- ธนชาติ นุ่มนนท์. (2557). *การวางกลยุทธ์ด้าน Big Data ขององค์กรและ Technology ด้าน Data ต่างๆ*. (ออนไลน์) ได้จาก <https://thanachart.org/2014/12/06/3645/> (สืบค้นเมื่อ 1 มีนาคม 2564)
- ยุทธการ อังมณีกุล. (2563). *File Formats ที่ใช้กับงาน Big Data มีอะไรบ้าง?*. (ออนไลน์) ได้จาก <https://yuthakarn.medium.com/big-data-file-format-csv-json-parquet-avro-14ed7e024a19> (สืบค้นเมื่อ 1 มีนาคม 2564)
- ราชกิจจานุเบกษา. (2563). *ประกาศคณะกรรมการพัฒนารัฐบาลดิจิทัลเรื่อง ธรรมนูญข้อมูลภาครัฐ*. 31 มีนาคม 2563. หน้า 47-48
- ราชบัณฑิตยสภา. (2563). *การบัญญัติศัพท์ของคณะกรรมการจัดทำพจนานุกรมศัพท์สถิติศาสตร์ คำว่า Big data ว่า ข้อมูลมหัศจรรย์*. (ออนไลน์) ได้จาก <https://royalsociety.go.th/tag/big-data/> (สืบค้นเมื่อ 1 มีนาคม 2564)
- วัลย์ลักษณ์ อมรสิริพงศ์. (2562). *เทคโนโลยี Big data กับงานห้องสมุดในยุคดิจิทัล*. *วารสารห้องสมุด* 63(2) 1-22

- วิกิพีเดีย. (2564). *Big data*. [ออนไลน์] ได้จาก [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data) [สืบค้นเมื่อ 1 มีนาคม 2564]
- สถาบันบิ๊กดาต้า (Government Big Data Institute: GBDi). (2563). [ออนไลน์] ได้จาก <https://bigdata.go.th/big-data-101/what-is-big-data/> [สืบค้นเมื่อ 1 มีนาคม 2564]
- สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม. (2564) *Big Data?* [ออนไลน์] :<https://www.ops.go.th/main/index.php/knowledge-base/article-pr/657-big-data> [สืบค้นเมื่อ 1 มีนาคม 2564]
- Bekker, A. (2019). *4 types of data analytics to improve decision-making*. Retrieved 2019 July 21. [online]. Available from <https://www.scnsoft.com/blog/4-types-of-data-analytics> [Access 3/1/2021]
- Big Data : What it is and why it matters*.(2021) [online]. ได้จาก [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html) [Access 3/1/2021]
- Doug Laney. (2001). "3D Data Management: Controlling Data Volume, Velocity, and Variety", *Gartner*, file No.949. 6 February 2001, [online]. Available from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity- and-Variety.pdf> [Access 3/1/2021]
- Gartner. (2019). *Big data*. [online]. Available from <https://www.gartner.com/it-glossary/big-data/> [Access 3/1/2021]
- IBM, Online. (2017). *What is big data? More than volume, velocity and variety...* . [online]. Available from <https://developer.ibm.com/blogs/what-is-big-data-more-than-volume-velocity- and-variety/> [Access 3/1/2021]
- Mandava Geetha Bhargava, Modugula T S Srinivasa Reddy, Shaik Shahbaz, P. Venkateswara Rao & V. Sucharita. (2017). *Potential of Big Data Analytics in Bio-Medical and Health Care Arena: An Exploratory Study*. *Global Journal of Computer Science and Technology: Software & Data Engineering*. 17 (2)