



การเข้าสู่ยุคใหม่ของข้อมูล “บิ๊กดาต้า” Transforms a New Era of Data “Big Data”

สุกิจ กุชัยสิทธิ์
sukit.k@bu.ac.th
มหาวิทยาลัยกรุงเทพ

บทคัดย่อ

การอาศัยอยู่ในโลกของข้อมูลข่าวสารไม่ใช่เรื่องง่าย เนื่องจากข้อมูลที่ต้องการจัดเก็บ มีรูปแบบที่มีโครงสร้างที่หลากหลาย และมีความเฉพาะเจาะจง เช่น ข้อมูลที่มีโครงสร้าง ข้อมูลที่ไม่มีโครงสร้าง และข้อมูลกึ่งมีโครงสร้าง นอกจากนี้ ข้อมูลยังมีอัตราการเพิ่มขึ้นอย่างรวดเร็ว จนทำให้มีข้อมูลปริมาณมหาศาลเกินกว่าขีดความสามารถของฐานข้อมูลปกติที่จะจัดเก็บ ประมวลผล และเรียกขึ้นมาใช้งานได้ ซึ่งข้อมูลเหล่านี้ถูกเรียกว่า “บิ๊กดาต้า” บทความนี้จะกล่าวถึง ความหมาย การจัดการ ผลกระทบต่างๆ และการเตรียมความพร้อมรับมือกับบิ๊กดาต้า

Abstract

Living in the world of information is not easy as before. This is due to the fact that several types of stored information have different formats, a variety of structure, and their own specifications such as structure data, unstructured data and semi-structure data. In addition, These data are increasingly fast and the storage of the large amount of data has reached normal database limits of database processing and retrieving. This type of data is named “Big Data”, and this article describes its meaning, ways to handle, various effects of big data, as well as preparation of an organization to cope with big data.

บทนำ

จากอดีต ได้มีการพัฒนาและใช้งานระบบงานธุรกิจต่างๆ ทำให้เกิดข้อมูลต่างๆ ซึ่งจัดเก็บสะสมอยู่ในรูปแบบของระบบฐานข้อมูล จนถึงปัจจุบัน ข้อมูลจึงมีปริมาณมหาศาลมากขึ้น และทำให้ระบบงานธุรกิจต่างๆ เหล่านั้น เข้าถึงข้อมูลได้ช้าและประมวลผลข้อมูลก็ช้ามากขึ้นตามไปด้วย ถึงแม้จะมีการปรับแต่ง (Tuning) ฐานข้อมูลเพื่อเพิ่มการประมวลผลและเข้าถึงข้อมูลแล้วก็ตาม แต่ก็ช่วยเพิ่มประสิทธิภาพได้เพียงระยะเวลาหนึ่งเท่านั้น แต่เมื่อระบบงานธุรกิจดำเนินงานต่อไป ข้อมูลก็จะเพิ่มขึ้นเรื่อยๆ จนทำให้มีปริมาณมหาศาลมากกว่าเดิม ซึ่งจะลดประสิทธิภาพการเข้าถึงและประมวลผลข้อมูลลงอีก ในปัจจุบัน สามารถแบ่งข้อมูลได้ 3 รูปแบบคือ 1) ข้อมูลที่มีโครงสร้าง (Structured Data) 2) ข้อมูลไม่มีโครงสร้าง (Unstructured Data) และ 3) ข้อมูลกึ่งมีโครงสร้าง (Semi-Structure Data)

ตัวอย่างเช่น ข้อมูลการโต้ตอบปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) ทวิตเตอร์ (Twitter) เฟซบุ๊ก (Facebook) ฟิล์มจำพวกมีเดีย (Media) ที่มีขนาดใหญ่ เป็นต้น ส่วนอีกปัจจัยหนึ่ง คือการที่มีข้อมูลใหม่ๆ เกิดขึ้นและเปลี่ยนแปลงรวดเร็ว เช่น ข้อมูลจากพวกมาตรวัดอัจฉริยะ (Smart Meter) ที่วัดผลแบบทันที (Real-Time) ข้อมูลต่างๆ ที่เกิดขึ้นเหล่านั้น ก่อให้เกิดโอกาสอย่างมากมาย โดยการนำข้อมูลมาประมวลผล และวิเคราะห์เพื่อใช้สร้างมูลค่าทางธุรกิจ ข้อได้เปรียบทางการแข่งขัน ช่วยสนับสนุนการตัดสินใจ และช่วยในกำหนดการวางแผนเชิงรุกของการทำงานในอนาคต (ธีรเกียรติ์ เกิดเจริญ, 2555)

ดังนั้นแหล่งข้อมูลต่างๆ ได้ขยายตัวและเพิ่มขึ้นอย่างต่อเนื่อง ทำให้มีปริมาณข้อมูลมีจำนวนมหาศาล และอยู่ในรูปแบบที่หลากหลายและแตกต่างกัน ซึ่งสามารถนำข้อมูลเหล่านั้นมาใช้ให้เกิดประโยชน์และตอบคำถามได้หลากหลายมุมมอง ข้อมูลทั้งหมดเหล่านั้น เรียกว่า บิ๊กดาต้า (Big Data) ในบทความนี้จะกล่าวถึงเรื่องความหมาย ลักษณะ การจัดการข้อมูลที่มีปริมาณมหาศาล และประโยชน์



ของบิกดาต้า เพื่อที่จะได้เตรียมการในการรับมือข้อมูลที่มีปริมาณมหาศาลเหล่านี้

ความหมายของบิกดาต้า (Big Data)

บิกดาต้า หมายถึง ปริมาณข้อมูลที่มีขนาดใหญ่มหาศาล เกินกว่าขีดความสามารถในการประมวลผลของระบบฐานข้อมูลธรรมดาที่จะรองรับได้ ปริมาณข้อมูลที่มีขนาดใหญ่มากๆ จะมีอัตราการเพิ่มข้อมูลได้อย่างรวดเร็วมากและจะมีรูปแบบที่ไม่มีโครงสร้างหรือกึ่งโครงสร้าง ซึ่งไม่สามารถอยู่ในระบบฐานข้อมูลที่จะจัดเก็บข้อมูลได้ (Dumbill, 2012)

ประโยชน์ของบิกดาต้ามีหลายประการแต่ที่เห็นได้ชัดเจนมีอยู่ 2 ประการ คือ 1) การวิเคราะห์ (Analytical Use) คือ การวิเคราะห์ข้อมูลที่มีปริมาณมหาศาลเหล่านั้นเพื่อให้เห็นความรู้ที่ซ่อนอยู่ลึก เช่น ข้อมูลสภาพอากาศ จากเครื่องตรวจวัดจำนวนมากบนโลก ทั้งดาวเทียม เรดาร์ บอลลูน ยานตรวจอากาศ และทุ่นลอยในมหาสมุทร ข้อมูลจำนวนมหาศาลที่ได้รับเหล่านี้นำมาสู่การพยากรณ์อากาศที่แม่นยำเป็นรายชั่วโมง ในด้านพันธุศาสตร์ นักวิทยาศาสตร์พยายามถอดรหัสพันธุกรรม เพื่อทำแผนที่ของสิ่งมีชีวิตต่างๆ บนโลก ซึ่งนำไปสู่การค้นพบยาใหม่ๆ เป็นต้น และ 2) การเกิดผลิตภัณฑ์ใหม่ (Enabling New Products) โดยสามารถนำบิกดาต้าที่มีปริมาณข้อมูลจำนวนมหาศาลมาใช้สร้างผลิตภัณฑ์หรือพัฒนาการบริการ เพื่อให้เหมาะสมตามความต้องการของลูกค้าหรือผู้ใช้ เช่น การบริการจัดส่งสินค้าให้กับลูกค้า โดยที่ลูกค้าซื้อของผ่านระบบเครือข่าย การบริการการจองหรือจำหน่ายตั๋วผ่านระบบเครือข่าย การบริการโอนเงินผ่านระบบเครือข่าย เป็นต้น (Dumbill, 2012; Kusnetzky, 2010; Panasas, 2012)

คุณลักษณะของบิกดาต้า

บิกดาต้าเปรียบเสมือนกลุ่มของก้อนเมฆ (Cloud) ขนาดใหญ่ ที่มีข้อมูลจำนวนมากที่รวมกันอยู่ อาทิ ข้อมูลสังคมเครือข่าย ข้อมูลการบริการทางเว็บ (Web Server Logs) ข้อมูลจากอุปกรณ์ตรวจจราจร (Traffic Flow Sensors) ข้อมูลภาพถ่ายดาวเทียม (Satellite Imagery) ข้อมูลด้านการกระจายเสียง (Broadcast Audio Streams) ข้อมูลธุรกรรมทางธนาคาร (Banking Transaction) ข้อมูลด้านการตลาดการเงิน (Financial Market Data) และข้อมูลการสื่อสารจากโทรศัพท์เคลื่อนที่ (Telemetry from Automobiles) เป็นต้น ดังนั้น คุณลักษณะของบิกดาต้าได้มีการจัดแบ่งออกเป็น 3 ลักษณะ ดังภาพที่ 1 (Dumbill, 2012; Wang, 2011)

ภาพที่ 1 แสดงข้อมูลทั้ง 3 ส่วนที่เป็น บิกดาต้า



ที่มา: ดัดแปลงจาก Wang (2011)

1. ปริมาตร (Volume) หมายถึง ข้อมูลที่มีปริมาณมหาศาล ซึ่งโครงสร้างข้อมูลของระบบฐานข้อมูลไม่สามารถจัดเก็บข้อมูลได้ ปริมาณข้อมูลมหาศาลมีประโยชน์เพื่อเป็นข้อมูลที่ใช้ในการตัดสินใจหรือทำนายอนาคต หรือเพื่อเตรียมการวางแผนการทำงานเชิงรุกทางธุรกิจ

การประมวลผลปริมาณของข้อมูลที่มีขนาดใหญ่กว่าการจัดการโดยโครงสร้างพื้นฐานของฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ใช้สถาปัตยกรรมของการประมวลผลของกลุ่มข้อมูลที่มีขนาดมหาศาล โดยจะทำงานไปพร้อมๆ กันระหว่าง คลังข้อมูล (Data Warehouse) หรือฐานข้อมูล (Database) และซอฟต์แวร์ Hadoop

2. ความเร็ว (Velocity) หมายถึง อัตราการเพิ่มขึ้นของข้อมูล ข้อมูลที่เข้าสู่ระบบฐานข้อมูลจะมีอัตราการเพิ่มขึ้นอย่างรวดเร็ว เช่น ข้อมูลที่เกิดขึ้นจากโทรศัพท์เคลื่อนที่ที่ผู้ใช้นำขึ้นเก็บเป็นข้อมูลภาพถ่าย ข้อมูลการพิมพ์การสนทนา ข้อมูลการอัปเดตภาพวิดีโอ หรือข้อมูลการอัปเดตเสียง หรือแม้กระทั่งข้อมูลการสั่งซื้อสินค้า การขนส่ง และการบริการต่างๆ ก็สามารถนำข้อมูลเหล่านั้นเข้าสู่ระบบฐานข้อมูลได้อย่างรวดเร็ว

จากการที่เทคโนโลยีได้พัฒนาก้าวไกลในโลกออนไลน์ และถูกนำมาใช้ในธุรกิจและการดำเนินชีวิตประจำวันทำให้ปริมาณข้อมูลที่ผ่านเข้ามาทางเทคโนโลยีต่างๆ เหล่านี้ เข้าสู่ระบบฐานข้อมูลได้ง่าย รวดเร็ว และเพิ่มปริมาณข้อมูลมากขึ้น และทำให้การประมวลผลทำได้ยากลำบาก ดังนั้น จึงต้องหาวิธีประมวลผลข้อมูลให้มีผลดีและถูกต้อง เพื่อหาผลลัพธ์ที่ต้องการ หรือมีการพิจารณาการประมวลผลในขณะที่ข้อมูลนั้นไหลเข้าสู่ระบบฐานข้อมูล

การพิจารณาประมวลผลข้อมูลที่ไหลเข้ามา (Streaming Processing) มี 2 หลักการ คือ หลักการแรก คือ เมื่อข้อมูลเข้ามาสู่ระบบฐานข้อมูลอย่างรวดเร็วจะมีการจัดลำดับข้อมูลเข้าไปในการจัดเก็บข้อมูลตามความต้องการ และปฏิบัติตามลำดับของการ

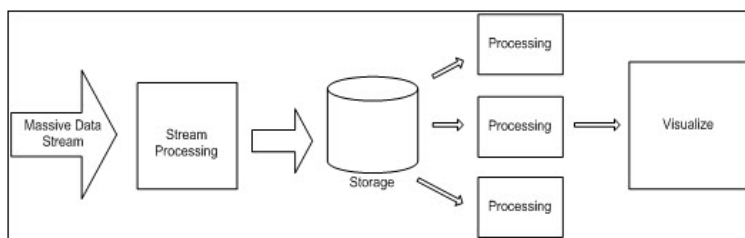
วิเคราะห์ที่เกิดขึ้นจากการเคลื่อนย้ายของข้อมูล ในที่สุดก็จะมีการจัดการในการจัดเก็บข้อมูล โดยจะไม่เก็บข้อมูลที่ไม่ต้องการนั้นลงไป ในฐานข้อมูล ส่วนหลักการที่สอง คือพิจารณาการเคลื่อนย้ายของข้อมูล แอปพลิเคชัน (Application) ที่สามารถใช้สร้างปฏิสัมพันธ์กับ ข้อมูลทันที เช่น Mobile Application และเกมออนไลน์ เป็นต้น ประเภทของผลิตภัณฑ์ที่ใช้เกี่ยวกับการทำงานของข้อมูลที่ไหลเข้ามา (Streaming Data) อาทิ IBM's Info Sphere Stream Twitter's Storm และ Yahoo S4 เป็นต้น

3. รูปแบบที่หลากหลาย (Variety) หมายถึง รูปแบบมีความหลากหลายของรูปแบบข้อมูล ซึ่งอาจจะเป็นรูปแบบที่ไม่มีโครงสร้าง ไม่มีโครงสร้าง และกึ่งมีโครงสร้าง เป็นต้น รูปแบบที่ไม่มีโครงสร้างหรือกึ่งโครงสร้างจะไม่เหมือนข้อมูลที่เรารวบรวมไว้ในระบบฐานข้อมูล เช่น ข้อความ อีเมล รูปภาพ วิดีโอและเสียง เป็นต้น ซึ่งข้อมูลเหล่านี้มีความซับซ้อน และเชื่อมโยงกัน

การจัดการบิกดาต้าโดยรูปแบบการกระจายข้อมูล

เมื่อขนาดของข้อมูลมีปริมาณที่ใหญ่โตมหาศาลที่เรียกว่า บิกดาต้า จึงต้องหาวิธีที่สามารถประมวลผลข้อมูลที่มีปริมาณมหาศาล และข้อมูลนั้นไม่สามารถเคลื่อนย้ายได้ ดังภาพที่ 2

ภาพที่ 2 รูปแบบการแสดงผลการจัดการข้อมูลที่เป็น บิกดาต้า



ที่มา: ดัดแปลงจาก Dumbill (2012)

การจัดการข้อมูลที่เป็นบิกดาต้ามีองค์ประกอบดังต่อไปนี้ (Dumbill, 2012)

1. การจัดเก็บ (Storage)
2. การประมวลผล (Processing)
3. การวิเคราะห์ (Analysis Algorithm)
4. การทำรายงานสรุป (Visualization)

จะมีการเคลื่อนย้ายกลุ่มของข้อมูลเข้าสู่ระบบฐานข้อมูลเพื่อการประมวลผลและนำไปเก็บในส่วนจัดเก็บข้อมูล ซึ่งจะเก็บปริมาณ ข้อมูลมหาศาลที่ไม่สามารถเคลื่อนย้ายได้ ทั้งนี้ ประมวลผลจะต้องเคลื่อนย้ายเข้ามาประมวลผลของข้อมูลให้อยู่ในมุมมองต่างๆ ที่ผู้ใช้ ต้องการ ในปัจจุบันได้มีการพัฒนาซอฟต์แวร์ Hadoop เข้ามาช่วยจัดการปริมาณข้อมูลมหาศาลนี้

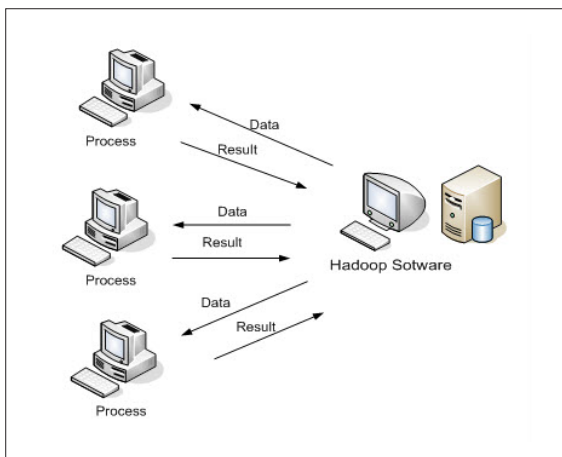
ซอฟต์แวร์ Hadoop

ซอฟต์แวร์ Hadoop เป็นซอฟต์แวร์แบบ Open Source ที่พัฒนาขึ้นมาเพื่อเป็นระบบวิเคราะห์และการจัดการกับข้อมูลปริมาณ มหาศาล โดยเกิดมาจาก Yahoo และถูกนำไปใช้ใน Google

โครงสร้างของซอฟต์แวร์ Hadoop ประกอบไปด้วย 2 ส่วนหลัก ๆ คือ HDFS (Hadoop Distributed File System) และ MapReduce ซึ่ง HDFS จะเป็นส่วนกระจายข้อมูลและเข้าถึงข้อมูล ส่วน MapReduce จะเป็นส่วนที่ลดโครงสร้างของข้อมูลซึ่งจะทำให้ข้อมูลที่มี ขนาดใหญ่มีขนาดเล็กลง เพื่อส่งไปยัง Server ซึ่งทำหน้าที่ในการประมวลผลพร้อมๆ กัน การประมวลผลจึงเป็นไปอย่างรวดเร็ว และจะ นำผลลัพธ์ที่ได้มารวมที่ส่วนกลาง นอกจากนี้ ยังมีเครื่องมือที่ทำงานคล้ายกับ MapReduce นั่นคือ MPP (Massively Parallel Processing) ซึ่งจะกระจายการประมวลผลของข้อมูลขนาดใหญ่พร้อมกัน และนำผลลัพธ์ที่ได้มารวมกันที่ส่วนกลาง แต่มักจะพบ MPP ในเครื่องมือ

ของระบบคลังข้อมูลระดับสูง (High-End Data Warehouse Appliances) ตัวอย่างเครื่องมือของ MPP ได้แก่ Netezza (IBM) Vertica (HP) Greenplum (EMC) และ Parallel Data Warehouse Edition ใน SQL Server (Microsoft) เป็นต้น ซึ่งเครื่องมือเหล่านี้มีราคาสูง เพราะต้องมีอุปกรณ์พิเศษสำหรับหน่วยประมวลผลกลาง (CPU) และมีระบบเครือข่ายและอุปกรณ์การจัดเก็บที่มีประสิทธิภาพสูง ซึ่งทำให้ซอฟต์แวร์ Hadoop ที่เป็นซอฟต์แวร์ Open Source ได้รับความนิยมมาก (Burst, 2012)

ภาพที่ 3 แสดงการกระจายการทำงานไปยังเครื่องหลายๆ เครื่อง



ที่มา: ดัดแปลงจาก Turner (2011)

จากภาพที่ 3 ซอฟต์แวร์ Hadoop มีการทำงานบนพื้นฐานของการกระจายการประมวลผลไปยัง Server หลายๆ ตัว (Distributed Computing System) ทั้งนี้ Google ใช้ซอฟต์แวร์ Hadoop เป็นเครื่องมือในการลดโครงสร้าง (MapReduce) เพื่อค้นหาและเข้าถึงข้อมูล โดยใช้วิธีการทำดัชนี (Index) ซึ่งมีผลการทำงานที่รวดเร็ว และซอฟต์แวร์ Hadoop จะทำหน้าที่กระจายข้อมูลส่งไปยัง Server หลายๆ ตัว (Hadoop Distributed File System: HDFS) เพื่อประมวลผลพร้อมๆ กัน ซึ่งผลลัพธ์บางส่วนก็จะมารวมกันที่ส่วนกลางทำให้ลดขั้นตอนได้ (Turner, 2011)

ซอฟต์แวร์ Hadoop มีขั้นตอน 3 ขั้นตอน คือ

1. การโหลดข้อมูลโดยใช้ HDFS
2. การลดการทำงานบางส่วน
3. การรับผลลัพธ์ที่ได้จากการทำ HDFS

ตัวอย่างของผู้ใช้งานซอฟต์แวร์ Hadoop มีดังต่อไปนี้

เฟซบุ๊กใช้ซอฟต์แวร์ Hadoop Cluster จำนวน 2 ชุด ชุดแรกประกอบด้วย Server จำนวน 1,100 เครื่อง CPU 8,800 แกน

(Cores) และพื้นที่ 12 เพตะไบต์ (PB (12,000TB)) และชุดที่สองประกอบด้วย Server จำนวน 300 เครื่อง CPU 2,400 แกน และพื้นที่ 3 เพตะไบต์ ในกรณีนี้จะเก็บข้อมูลไว้ที่ระบบฐานข้อมูล MySQL การทำงานของฐานข้อมูลจะเชื่อมต่อกับซอฟต์แวร์ Hadoop และทำการประมวลผล อาทิ การสร้างคำแนะนำบน Homepage ของเพื่อนในเฟซบุ๊กและจะส่งผลลัพธ์กลับไปยังฐานข้อมูล MySQL สำหรับการใช้ในหน้าบริการ (ทูลเวฟ, 2554)

Yahoo ใช้ Server มากกว่า 40,000 เครื่อง CPU มากกว่า 100,000 ชุด สำหรับรองรับระบบ Ads และ Web Search (ทูลเวฟ, 2554)

นอกจากนี้ ยังมีผู้ให้บริการรายใหญ่ๆ มากมาย เช่น Twitter ImageShack Adobe AOL และอื่นๆ ที่ใช้ซอฟต์แวร์ Hadoop ในการจัดเก็บข้อมูลแทนฐานข้อมูล MySQL รวมถึง Microsoft เองก็มีแผนที่จะให้ฐานข้อมูล MS SQL สามารถทำงานเชื่อมต่อกับซอฟต์แวร์ Hadoop ได้เช่นกัน อีกทั้งยังมีผู้ผลิตรายใหญ่ๆ อย่าง IBM และ Supermicro ที่สนับสนุนซอฟต์แวร์ Hadoop ส่วนในวงการการศึกษาและทางภาคธุรกิจเอง ซอฟต์แวร์ Hadoop ก็เป็นทางเลือกที่ดีในงานหลายๆ ประเภท ไม่ว่าจะเป็นงานประมวลผลประสิทธิภาพสูง (High Performance Computing) การประมวลผลภาพ (Image Processing) ระบบสืบค้นสารสนเทศ (Information Retrieval) การเรียนรู้ของเครื่อง (Machine Learning) การวิเคราะห์เครือข่ายทางสังคม (Social Network Analysis) การทำเหมืองข้อมูล (Data Mining) ธุรกิจอัจฉริยะ (Business Intelligence: BI) ความปลอดภัยบนเครือข่าย (Network Security) และการวิเคราะห์จดหมายอิเล็กทรอนิกส์ (Email Analysis) เป็นต้น

ผลกระทบกับการมาของบิกดาต้า

จากการศึกษาข้อมูลเบื้องต้นและการวิจัยด้านบิกดาต้าของ MEXT และ NSF (ออฟฟิศ ออฟ ไซน์ แอนด์ เทคโนโลยี, 2555) วิเคราะห์ได้ว่า การทำงานของหน่วยงานหรือองค์กรต่างๆ จะได้รับผลกระทบจากเทคโนโลยีบิกดาต้า ซึ่งทำให้เกิดข้อมูลที่มีปริมาณมหาศาล ในปัจจุบัน ได้มีการใช้เทคโนโลยีการสื่อสารทางเครือข่ายอินเทอร์เน็ตอย่างแพร่หลาย การเข้าถึงข้อมูลและอัตราการเพิ่มข้อมูลสามารถทำได้ง่าย สะดวกและรวดเร็ว เนื่องจากได้มีการพัฒนาระบบการสื่อสารจนถึงระดับ 4G ซึ่งเป็นระดับที่มีการส่งข้อมูลข่าวสารได้อย่างรวดเร็ว ทำให้ข้อมูลมีอัตราการเพิ่มขึ้นอย่างรวดเร็วและตลอดเวลา

ผลกระทบจากการเข้ามาของบิกดาต้าจะให้หน่วยงานหรือองค์กรต่างๆ สามารถนำข้อมูลนั้นไปใช้ให้เกิดประโยชน์อย่าง

มากมาย ดังนั้นจึงขอยกตัวอย่างการใช้ประโยชน์ของบิ๊กดาต้าในด้านธุรกิจ และด้านการบริหารจัดการภัยธรรมชาติ เพื่อให้เห็นภาพที่ชัดเจนขึ้น ดังต่อไปนี้

ประโยชน์ทางด้านธุรกิจ เช่น 1) การนำข้อมูลเหล่านั้นไปวิเคราะห์เพื่อหาความต้องการสินค้าใหม่หรือการบริการในรูปแบบใหม่ๆ ให้ตรงกับความต้องการของผู้ใช้ หรือนำมาสร้างมูลค่าทางธุรกิจ 2) การนำข้อมูลเหล่านั้นมาช่วยสนับสนุนการตัดสินใจหรือนำมาสร้างข้อได้เปรียบทางการแข่งขัน 3) การนำข้อมูลเหล่านั้นมาช่วยประเมินระยะเวลาและประเมินงบประมาณค่าใช้จ่ายของการทำโครงการใหม่ๆ ให้ใกล้เคียงกับการปฏิบัติงานจริง หรือช่วยแก้ปัญหาหรือป้องกันปัญหาที่จะเกิดขึ้นในโครงการใหม่ 4) การนำข้อมูลเหล่านั้นมาช่วยในการวิเคราะห์เพื่อนำไปวางแผนการตลาดวางแผนการส่งเสริมการจัดจำหน่าย การวางแผนเชิงรุกของการทำงานในอนาคต หรืออาจจะทำให้เกิดธุรกิจใหม่ๆ ขึ้น ฯลฯ และ 5) การนำข้อมูลเหล่านั้นมาทำให้เกิดสถานการณ์การแข่งขันระหว่างองค์กรธุรกิจที่ทำธุรกิจเดียวกัน ซึ่งจะส่งผลดีต่อผู้บริโภคและทำให้ผู้บริโภคสามารถรู้และเข้าใจข้อมูลสินค้า และการบริการของแต่ละองค์กรธุรกิจ ทำให้ผู้บริโภคมีทางเลือกสิ่งที่ดีที่สุด และทันต่อเหตุการณ์ เป็นต้น

ส่วนประโยชน์ด้านการบริหารจัดการภัยธรรมชาติ เช่น 1) การนำข้อมูลภัยพิบัติจากธรรมชาติ เพื่อการวิเคราะห์ที่ก้าวหน้าการจัดทำโมเดล และการสร้างขีดความสามารถทางคอมพิวเตอร์เพื่อนำไปประยุกต์ใช้ เช่น โมเดลของสถานการณ์ที่คาดว่าจะเกิดขึ้นอันตราย เป็นต้น 2) การนำข้อมูลเหล่านั้นมาปรับปรุงวิธีการฟื้นฟูแก้ไขให้กลับไปสู่สภาพเดิมและพร้อมรับ โดยใช้เทคโนโลยีสารสนเทศที่เอื้อต่อการตรวจวัดด้วยข้อมูลที่เป็นปัจจุบัน การแปลงข้อมูลเป็นภาพ การวิเคราะห์ การทดลอง และการทำนายเพื่อการตัดสินใจในช่วงวิกฤติ 3) การนำข้อมูลมาสร้างนวัตกรรมและความรู้พื้นฐานขั้นสูงเพื่อการสร้างโครงสร้างพื้นฐาน การฟื้นฟูแก้ไขให้กลับไปสู่สภาพเดิมและความยั่งยืน รวมถึงเครือข่ายพื้นฐานในการเผยแพร่ข้อมูลข้างต้น และ 4) การนำข้อมูลเหล่านั้นมารวบรวมสร้างเป็นฐานข้อมูลขนาดใหญ่และการปรับปรุงฐานความรู้ที่กว้างขวาง เพื่อเตรียมรับและตอบสนองของสังคม และความต้องการของโลก รวมถึงมิติต่างๆ ในด้านสิ่งแวดล้อม เศรษฐกิจ สังคม และมนุษยชาติ เป็นต้น

จะเห็นได้ว่า ประโยชน์ของการนำบิ๊กดาต้ามาใช้ยังมีอีกมากมายมหาศาล ดังนั้น จึงต้องเตรียมพร้อมนำข้อมูลบิ๊กดาต้ามาใช้ก่อนให้เกิดประโยชน์สูงสุดและให้เกิดปัญหาน้อยที่สุด

การเตรียมความพร้อมรับเทคโนโลยีบิ๊กดาต้า

ในโลกธุรกิจปัจจุบัน องค์กรต่างๆ กำลังตื่นตัวและเตรียมความพร้อมที่จะรองรับเทคโนโลยีบิ๊กดาต้า เพื่อนำข้อมูลมหาศาลเหล่านั้นไปใช้ให้เกิดประโยชน์สูงสุด พร้อมทำนายนสิ่งต่างๆ ที่จะเกิดขึ้นในอนาคต การผสมผสานเทคโนโลยีระดับสูงเพื่อสนับสนุนนวัตกรรมและเทคโนโลยีการบริหารจัดการข้อมูลขนาดใหญ่หรือบิ๊กดาต้า และเพื่อตอบสนองเทคโนโลยีที่มีการพัฒนาขึ้นอย่างรวดเร็วและด้วยขนาดข้อมูลที่อัตราการเพิ่มขึ้นอย่างรวดเร็วนี้ จึงมีการพัฒนานวัตกรรมแบบใหม่เพื่อการบริหารจัดการข้อมูลแบบต่างๆ และเพิ่มประสิทธิภาพการทำงานร่วมกับโปรแกรมประยุกต์หลากหลายรูปแบบ โดยสามารถปรับเปลี่ยนเทคโนโลยีการใช้งานได้อย่างยืดหยุ่น รวดเร็ว และมีประสิทธิภาพสูงสุด

บริษัทชั้นนำต่างๆ ทางเทคโนโลยี อาทิ EMC² Dell Cloudera IBM Oracle และ Microsoft เป็นต้น ได้เตรียมความพร้อมที่จะรับเทคโนโลยีนี้ที่จะเข้ามาในเมืองไทยทั้งด้านซอฟต์แวร์และด้านฮาร์ดแวร์ โดยการนำเทคโนโลยีบริหารจัดการจัดเก็บข้อมูลและสำรองข้อมูล เพื่อเพิ่มประสิทธิภาพการทำงานร่วมกับโปรแกรมประยุกต์ในปัจจุบันและอนาคต ด้วยปริมาณข้อมูลที่มีขนาดเพิ่มมากขึ้นและเพิ่มขยายตลอดเวลา กรรมวิธีการจัดเก็บและสำรองข้อมูลจึงมีความสำคัญมากที่จะช่วยรองรับการทำงานของโปรแกรมประยุกต์แบบต่างๆ ให้สามารถตอบสนองกับผู้ใช้งานได้อย่างรวดเร็วขึ้น รูปแบบเทคโนโลยีดังกล่าวยังเพิ่มความสามารถในการบริหารจัดการและเรียกคืนข้อมูลได้อย่างถูกต้องและรวดเร็ว ทำให้การทำงานของโปรแกรมหรือ Application สามารถตอบรับกับผู้ใช้งานได้อย่างลงตัวและรองรับขีดความสามารถในการขยายงานได้อีกในอนาคต โดยขอยกตัวอย่างบริษัทที่ได้เตรียมความพร้อมในการรับเทคโนโลยีบิ๊กดาต้า ดังนี้

บริษัท Dell Corporation (Thailand) จำกัด มีนวัตกรรมที่สามารถตอบสนองการจัดการที่จัดเก็บข้อมูล (Storage) ด้วยสถาปัตยกรรม Fluid Data ที่เน้นประโยชน์ 3 ประการ คือ

1. ประสิทธิภาพ ความคล่องตัวในการเข้าถึงข้อมูลได้ในเวลาอันรวดเร็ว
2. ความพร้อมใช้งานของระบบการทำงานของโปรแกรมหรือ Application
3. ในอนาคตสามารถเก็บข้อมูลสารสนเทศด้วยคุณลักษณะ ดังนี้

3.1 เทคโนโลยีการจัดข้อมูลซ้ำซ้อน (De-Duplication) และการบีบอัดข้อมูล (Compression) ไว้ในทุกๆ ส่วนของที่จัดเก็บข้อมูล ทำให้ช่วยประหยัดเนื้อที่ในการเก็บข้อมูล

3.2 สามารถทำยูนิไฟต์ในระดับไฟล์ บล็อก และออบเจกต์ โดยการรวมระบบโครงสร้างพื้นฐานทั้งหมดไว้เป็นผืนเดียว ซึ่งจะช่วยให้การค้นหาข้อมูลเร็วขึ้น

3.3 รวมระบบ Server ทั้งหมดเป็นผืนเดียว (Seamless Server Integration) และส่งข้อมูลถึงกันระหว่างเซิร์ฟเวอร์ได้โดยตรง

3.4 การรวมระบบทั้งหมดไว้บนคลาวด์ (Seamless Cloud Integration) และมองเห็นโครงสร้างพื้นฐานทั้งหมดเป็นผืนเดียว ซึ่งเป็นนวัตกรรมที่ตอบสนองต่อแนวโน้มของบิ๊กดาต้า โดยจัดเป็น กลุ่มการจัดเก็บข้อมูล (Dell Big Data Retention Solution) กลุ่มการวิเคราะห์ข้อมูลเชิงธุรกิจ (Dell Big Data Analytics Solution) และรูปแบบการนำเสนอ (Business Ready Solutions) เพื่อให้ผู้ใช้สามารถเข้าระบบได้ทันที (บล็อกนั้น, 2555)

บริษัท Oracle Corporation (Thailand) จำกัด ได้มีนวัตกรรมที่เตรียมความพร้อมของการมาของบิ๊กดาต้า โดยผ่านผลิตภัณฑ์ที่ชื่อว่า Oracle Big Data Appliance ซึ่งแบ่งการจัดการข้อมูลประเภทบิ๊กดาต้า ออกเป็น 3 ขั้นตอน ได้แก่ 1) ขั้นตอนการรวบรวมและจัดเก็บข้อมูล (Data Acquisition) 2) ขั้นตอนการจัดเตรียมข้อมูล (Data Organization) และ 3) ขั้นตอนการวิเคราะห์ข้อมูล (Data Analysis) ซึ่ง บริษัท Oracle มีสถาปัตยกรรมที่สามารถรองรับปริมาณการใช้งานที่สูงได้ โดยได้เลือกใช้ผลิตภัณฑ์ NoSQL Database ในการรองรับบริการ ซึ่งมีความยืดหยุ่นมากในการสนับสนุนลักษณะข้อมูลแบบต่างๆ และสามารถขยายตัวเพื่อรองรับปริมาณงานได้สูงโดยการขยายในแนวราบ (Scale Out) ในการจัดเตรียมข้อมูล ทางบริษัท Oracle เลือก Hadoop ของ Apache มาเป็นโครงสร้างพื้นฐานในการจัดเก็บข้อมูล และการประมวลผลโดยทำงานอยู่บนระบบปฏิบัติการ Oracle Linux ส่วนการวิเคราะห์ บริษัท Oracle ใช้โปรแกรม Oracle R Enterprise จากซอฟต์แวร์ฟรีชื่อ "R" ที่ใช้กันแพร่หลายในการคำนวณทางด้านสถิติและกราฟิก นอกจากนี้ บริษัท Oracle ยังได้ต่อเชื่อมบิ๊กดาต้ากับผลิตภัณฑ์ของ บริษัท Oracle ที่มีอยู่เดิม เช่น Oracle Database 11g Oracle Exadata และ Oracle Exalytics โดยผ่านเครื่องมือที่ได้พัฒนาขึ้นมาใหม่ในกลุ่มที่เรียกว่า Oracle Big Data Connectors (ณรงค์ฤทธิ มโนมัยพิบูลย์, 2555)

ในด้านผู้ใช้บริการเทคโนโลยี องค์กรต่างๆ ได้เตรียมความพร้อมของการมาของบิ๊กดาต้า โดยส่งบุคลากรในองค์กรไปอบรมความรู้เรื่องเทคโนโลยีของบิ๊กดาต้า การจัดการกับบิ๊กดาต้า กับองค์กรชั้นนำต่างๆ ทางด้านเทคโนโลยี เพื่อสร้างความรู้และความเข้าใจที่จะนำข้อมูลบนบิ๊กดาต้าไปใช้ให้เกิดประโยชน์สูงสุด และวางแผนการทำงาน การวางแผนกลยุทธ์เชิงธุรกิจ การจัด

งบประมาณค่าใช้จ่ายต่างๆ ที่จะเกิดขึ้น อาทิ การอบรมให้ความรู้เรื่องบิ๊กดาต้าแก่บุคลากร การอบรมเครื่องมือที่จะนำเอาเทคโนโลยีบิ๊กดาต้ามาใช้ การซื้อเครื่องมือที่จะใช้การจัดการข้อมูลบนบิ๊กดาต้า และการบำรุงรักษาเครื่องมือ เป็นต้น

จากตัวอย่างที่กล่าวมาเบื้องต้นนั้นจะเห็นได้ว่า องค์กรชั้นนำต่างๆ ทางด้านเทคโนโลยีได้เตรียมความพร้อมรับมือกับเทคโนโลยีบิ๊กดาต้าที่กำลังจะเข้ามาในประเทศไทย โดยจัดการอบรมสัมมนา และให้ความรู้ด้านต่างๆ พร้อมทั้งเทคโนโลยีด้านซอฟต์แวร์และฮาร์ดแวร์แก่บุคลากร ส่วนด้านผู้ใช้บริการเทคโนโลยีขององค์กรต่างๆ ก็ได้ส่งบุคลากรมาอบรม สัมมนา ในด้านความรู้ ความเข้าใจและการจัดการ เพื่อเป็นการต้อนรับการมาของเทคโนโลยีนี้

บทสรุป

กระแสบิ๊กดาต้าถือเป็นแนวโน้มที่จะมีความสำคัญมากขึ้นอย่างต่อเนื่อง เนื่องจากข้อมูลในปี 2011 มีอัตราการเติบโตเพิ่มขึ้น 47 เท่า และมีข้อมูลเกิดขึ้น 1.8 หมื่นล้านกิกะไบต์ (GB) พร้อมกับคาดการณ์ว่า ในปี 2016 จะต้องใช้พื้นที่เก็บข้อมูลเพิ่มขึ้นร้อยละ 850 ("เดลินิวส์แนวโน้ม 'Big Data' อีก 4 ปีใช้พื้นที่มหาศาลเก็บข้อมูล", 2555) ซึ่งทำให้ปริมาณข้อมูลมหาศาลจะอยู่ในระดับเทราไบต์ (TB) และอาจเพิ่มสูงขึ้นไปอีก การเข้าถึงข้อมูลจะต้องใช้เวลาและข้อมูลบางส่วนไม่ได้มีการจัดระเบียบ จึงจำเป็นต้องจัดระเบียบการเก็บข้อมูล ในบางองค์กร มีการพัฒนาต่อยอดเพื่อวิเคราะห์เชิงธุรกิจ เพื่อสร้างโอกาสในการทำธุรกิจ ขณะเดียวกันจะต้องมีการจัดการกับบิ๊กดาต้า จึงทำให้ฝ่ายเทคโนโลยีสารสนเทศต้องเผชิญกับความท้าทายในเรื่องต่างๆ อาทิ การเก็บข้อมูลจำนวนมากมหาศาล วิธีลดค่าใช้จ่ายในการดูแลรักษาข้อมูล การรักษาระดับการบริการตามที่ตกลงไว้ การทำตามกฎเกณฑ์และข้อกำหนดของรัฐบาลในการเก็บข้อมูลอย่างเคร่งครัด การเข้าถึงข้อมูลที่ซับซ้อนได้ อีกทั้งยังจำเป็นต้องปรับปรุงประสิทธิภาพของฐานข้อมูล การจัดเก็บข้อมูลและระบบการวิเคราะห์ข้อมูล

องค์กรธุรกิจต่างๆ จะต้องเตรียมความพร้อมที่จะรับเทคโนโลยีบิ๊กดาต้าที่จะเข้ามา เพื่อการบริหารข้อมูลปริมาณมหาศาลที่จะมีอัตราการเพิ่มอย่างรวดเร็วและนำข้อมูลเหล่านั้นไปใช้ให้เกิดประโยชน์ต่างๆ ทั้งในเรื่องการแข่งขัน การวางแผนเชิงรุกของการทำงานในอนาคต การตัดสินใจ การเกิดสินค้าและการบริการใหม่ๆ เพื่อตอบสนองความต้องการของลูกค้า เป็นต้น โดยอุตสาหกรรมของประเทศไทยที่น่าจะได้ใช้ก่อนคือ ธนาคาร และการเงิน โทรคมนาคมและสุขภาพ

ดังนั้น ในอนาคตบี๊กดาต้าและซอฟต์แวร์ที่จัดการบี๊กดาต้า (Hadoop Software) จะมีบทบาทในระดับองค์กรขนาดใหญ่ (Enterprise) มากขึ้น และการจัดเก็บข้อมูลต่างๆ ก็อยู่ใน Server หลายๆ Server และรูปแบบการดึงข้อมูลมาใช้งานจะเป็นลักษณะ

กลุ่มก้อนเมฆ ซึ่งจะลดต้นทุนในเรื่องของฮาร์ดแวร์ ในขณะที่ผู้ดูแลระบบเองก็ควรจะต้องเริ่มศึกษาเทคโนโลยีใหม่ๆ เพิ่มเติม นอกเหนือจากระบบจัดการฐานข้อมูล แบบเดิมๆ เช่นกัน

บรรณานุกรม

- ณรงค์ฤทธิ์ มโนมัยพิบูลย์. (2555). Oracle กับ Big Data. สืบค้นเมื่อ 6 พฤศจิกายน 2555, จาก <http://www.firstlogic.co.th/portal/page/portal/firstlogic/thai/products/oracle/database/BigData/BigDATA.pdf>
- เดลินิวส์แนวนอน 'Big Data' อีก 4 ปีใช้พื้นที่มหาศาลเก็บข้อมูล. (2555, 10 มิถุนายน). ไทยรัฐออนไลน์. สืบค้นเมื่อ 31 กรกฎาคม 2555, จาก <http://www.thairath.co.th/content/tech/267228>
- ทรวูฟ. (2554, 10 พฤศจิกายน). จัดการข้อมูลขนาดใหญ่ด้วย Apache Hadoop สำหรับองค์กร. สืบค้นเมื่อ 31 กรกฎาคม 2555, จาก <http://www.throughwave.co.th/2011/11/10/enterprise-apache-hadoop/>
- ธีรเกียรติ์ เกิดเจริญ. (2555, 6 มิถุนายน). Big Data ยุคข้อมูลใหญ่หิมามา มาถึงแล้ว. สืบค้นเมื่อ 30 สิงหาคม 2555, จาก <http://nano-in-thailand.blogspot.com/2012/06/big-data.html>
- บล็อกนั้น. (2555, 8 มิถุนายน). ยุทธศาสตร์ Big Data และ Enterprise Storage ของ Dell. สืบค้นเมื่อ 27 สิงหาคม 2555, จาก <http://www.blognone.com/node/33147>
- ออฟฟิศ ออฟ ไซน์ แอนด์ เทคโนโลยี. (2555, 7 มิถุนายน). MEXT และ NSF ร่วมวิจัยด้าน Big Data. สืบค้นเมื่อ 7 พฤศจิกายน 2555, จาก http://www.ostc.thaiembdc.org/test2012/stnews_July12_1
- Brust, A. (2012, March 2). MapReduce and MPP: Two sides of the Big Data coin? Retrieved February 12, 2013, from <http://www.zdnet.com/blog/big-data/mapreduce-and-mpp-two-sides-of-the-big-data-coin/121>
- Dumbill, E. (2012, January 11). What is big data? An introduction to the big data landscape. Retrieved July 24, 2012, from <http://radar.oreilly.com/2012/01/What-is-big-data.html>
- Kusnetzky, D. (2010, February 16). What is "Big Data?". Retrieved July 30, 2012, from <http://www.zdnet.com/blog/virtualization/what-is-big-data/1708>
- Panasas. (2012, 13 March). Big Data. Retrieved July 27, 2012, from <http://www.panasas.com/solutions/big-data>
- Turner, J. (2011, January 12). Hadoop: What it is, how it works, and what it can do. Retrieved August 1, 2012, from <http://radar.oreilly.com/2011/01/What-is-hadoop.html>
- Wang, R. (2011, February 27). Monday's musings: Beyond The three V's of big data - Viscosity and virality. Retrieved September 6, 2012, from <http://www.forbes.com/sites/raywang/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/>