

บทบาทของเทคโนโลยีบิ๊กดาต้าต่อธุรกิจสมัยใหม่ และซอฟต์แวร์โอเพนซอร์สที่เกี่ยวข้อง The Role of Big Data Technologies in Modern Business and Related Open Source Software

นันทน์ภัท สัจจิม่า^{1*}

¹ภาควิชาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจ มหาวิทยาลัยพาร์อิสเทอร์น
เลขที่ 120 ถนนมหาราช ตำบลหายยา อำเภอเมือง จังหวัดเชียงใหม่ 50100

บทคัดย่อ

ปริมาณข้อมูลในยุคสังคมดิจิทัลมีปริมาณเพิ่มขึ้นอย่างรวดเร็ว จากหลายแหล่งข้อมูล และหลากหลายลักษณะทั้งแบบมีโครงสร้าง และไม่มีโครงสร้าง บิ๊กดาต้า หมายถึงปริมาณข้อมูลขนาดใหญ่ที่ประกอบด้วย 5 องค์ประกอบ คือ ปริมาณ ความหลากหลาย ความรวดเร็ว ความผันแปร และคุณค่าที่นำเสนอ หากผู้ประกอบการสามารถวิเคราะห์บิ๊กดาต้าเพื่อให้เกิดคุณค่าทางธุรกิจได้ ย่อมส่งผลให้ได้เปรียบทางธุรกิจเหนือคู่แข่ง ปัจจุบันมีเทคโนโลยีบิ๊กดาต้าทั้งในรูปแบบของซอฟต์แวร์โอเพนซอร์ส และซอฟต์แวร์ทางการค้า ขึ้นอยู่กับความเหมาะสมกับประเภทการใช้งานและขนาดของธุรกิจ การใช้เทคโนโลยีดังกล่าวผู้ประกอบการธุรกิจต้องเตรียมความพร้อมทั้งในด้านข้อมูล โครงสร้างพื้นฐานทางด้านเทคโนโลยี และบุคลากร ซึ่งอาจจะเหมาะสมสำหรับธุรกิจขนาดใหญ่ สำหรับธุรกิจขนาดเล็กที่มีต้นทุนไม่มากสามารถใช้บริการบริษัทที่ทำการวิเคราะห์ข้อมูลโดยเฉพาะ หากต้องการทราบเรื่องใดทางบริษัทจะดำเนินการวิเคราะห์ และขายเป็นรายงานให้ในลักษณะการคิดค่าบริการตามการใช้งานจริง (Pay Per Use) ในบทความนี้จะนำเสนอข้อมูลเกี่ยวกับ ฮาดูป (Hadoop) ซึ่งเป็นซอฟต์แวร์โอเพนซอร์ส ที่นิยมใช้ในระบบการจัดเก็บ และประมวลผลข้อมูลบิ๊กดาต้า การใช้งานต้องพัฒนาด้วยโปรแกรมภาษาจาวา ซึ่งไม่สะดวกสำหรับผู้ใช้งานราย จึงทำให้มีการสร้างเครื่องมือเพื่อใช้งานร่วมกับฮาดูป ส่งผลให้การประมวลผลข้อมูลสะดวกและเหมาะสมกับผู้ใช้งานโดยทั่วไป

บทความนี้จะเสนอประโยชน์สำหรับผู้ประกอบการที่ต้องการวิเคราะห์ข้อมูลบิ๊กดาต้าด้วยซอฟต์แวร์โอเพนซอร์ส โดยนำเสนอเกี่ยวกับฮาดูปและเครื่องมือที่ทำงานร่วมกับฮาดูป รวมถึงการวิเคราะห์บิ๊กดาต้าทางด้านการตลาดโดยซอฟต์แวร์ ออราเคิล โซเชียล รีเลชันชิพ เมเนจเม้นท์ (Oracle Social Relationship Management) ในการให้ผลสะท้อนกลับ (Feedback) ของผู้บริโภคจากเว็บไซต์ และเครือข่ายสังคมออนไลน์ เพื่อให้ร้านค้าสามารถวิเคราะห์ข้อมูล และกำหนดทิศทางการทำการตลาดได้อย่างรวดเร็ว

* ผู้เขียนหลัก
อีเมล: nunnapat@feu.edu



คำสำคัญ

บิ๊กดาต้า ฮาดูป เครื่องมือฮาดูป

Abstract

The amount of data in digital society is increasing rapidly from many sources of data both structured and unstructured. Big data is a large volume of data which consists of 5Vs, Volume, Variety, Velocity, Variability and Value proposition. Analyzing Big Data allows entrepreneurs to gain advantages over their competitors. Big Data technology provides in both open source and commercial solutions, depending on entrepreneurs needs. A user has to provide data, infrastructure of technology and people. These components are suitable for enterprise-level applications. Small SMEs can access this service by pay per use from a specific data analysis company to reduce the cost. This article presents Hadoop, which is an open source software framework for storing and processing Big Data. To use hadoop for big data analysis, users must implement solutions via java programming language, which may inconvenient for some. As a result, a number of Hadoop tools have been put forward to ease the problem.

This article is beneficial for entrepreneurs who want to analyze big data with open source software: Hadoop and Hadoop tools. Moreover, Oracle Social Relationship Management also utilize those who analyze big data in marketing to get consumer's feedback from websites and social network in order to analyze and set market direction rapidly by marketer.

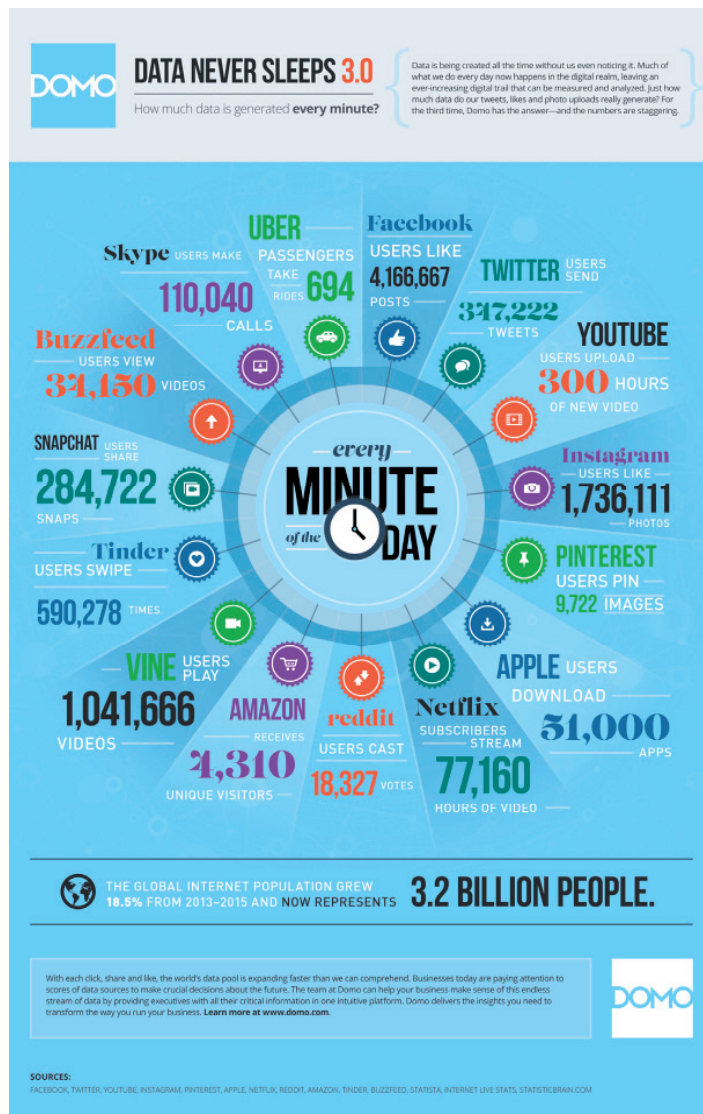
Keywords

Big Data, Hadoop, Hadoop Tools

บทนำ

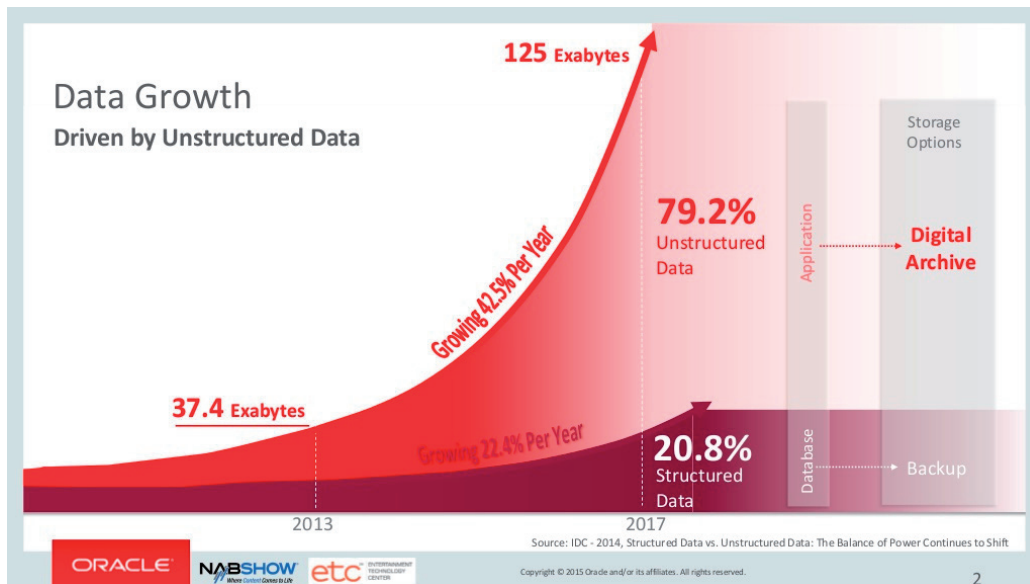
จากการเติบโตอย่างรวดเร็วในการใช้งานผลิตภัณฑ์ และบริการที่เกี่ยวข้องกับยุคดิจิทัล ส่งผลให้ปริมาณข้อมูลที่มีอยู่ในระบบมีปริมาณเพิ่มสูงขึ้นอย่างรวดเร็วและต่อเนื่อง ทั้งจากสื่อสังคมออนไลน์ (Social Media) อินเทอร์เน็ตออฟติงส์ (Internet of Things) การเชื่อมต่อของอุปกรณ์อิเล็กทรอนิกส์ผ่านระบบเครือข่ายอินเทอร์เน็ต และสามารถติดต่อสื่อสารส่งข้อมูลหากันได้ หรือส่งข้อมูลเพื่อเก็บไว้ในฐานข้อมูลในระบบเครือข่ายอินเทอร์เน็ต รวมถึงข้อมูลรายการประจำวันที่เกิดขึ้นจากการดำเนินธุรกิจ (Transaction Logs) ผ่านระบบเครือข่ายดิจิทัล จากภาพที่ 1 จะเห็นว่าข้อมูลในยุคดิจิทัลมีการเพิ่มขึ้นและเปลี่ยนแปลงอย่างรวดเร็วโดยในทุกๆ หนึ่งนาทีก มีการกดไลค์ในเฟซบุ๊กถึง 4,166,667 โพสต์ (Posts) ส่งข้อความในโปรแกรมทวิตเตอร์ 347,222 ทวิต (Tweets) อีเมลหอตวีโอ 300 ชั่วโมงบนยูทูบ และดาวนโหลดแอปพลิเคชัน

ของค่ายแอปเปิล 51,000 แอป (Apps) จากภาพที่ 1 แสดงให้เห็นถึงปริมาณข้อมูลมหาศาลที่ถูกสร้างขึ้นในเวลาเพียง 1 นาที ข้อมูลดังกล่าวเกิดขึ้นจากหลากหลายแหล่งข้อมูล และหลากหลายลักษณะทั้งข้อความ ภาพนิ่ง และภาพเคลื่อนไหว ซึ่งลักษณะข้อมูลในรูปแบบมัลติมีเดียมีโครงสร้างในการจัดเก็บข้อมูลแบบไม่มีโครงสร้าง ลักษณะข้อมูลที่เกิดขึ้นในยุคดิจิทัลมีแนวโน้มเป็นข้อมูลแบบไม่มีโครงสร้างมากยิ่งขึ้น ซึ่งสอดคล้องกับข้อมูลจากภาพที่ 2 แสดงให้เห็นถึงแนวโน้มการเพิ่มขึ้นของข้อมูลแบบไม่มีโครงสร้างที่เพิ่มขึ้น 42.5% ต่อปี



ภาพที่ 1: ปริมาณข้อมูลมหาศาลที่เกิดขึ้นในทุกๆ 1 นาที จากหลายแหล่งข้อมูล และหลายลักษณะทั้งแบบมีโครงสร้างและไม่มีโครงสร้าง

ที่มา: DOMO, 2015



ภาพที่ 2: แผนภาพแสดงแนวโน้มการเพิ่มขึ้นของข้อมูลแบบไม่มีโครงสร้าง
ที่มา: ET Center, 2015

เทคโนโลยีบิ๊กดาต้า คือ เทคโนโลยีในการรวบรวมจัดเก็บฐานข้อมูลปริมาณมหาศาล และมีรูปแบบของข้อมูลที่หลากหลายทั้งแบบมีโครงสร้างและไม่มีโครงสร้าง จากการจัดเก็บข้อมูลปริมาณมหาศาลนั้น ข้อมูลจำเป็นต้องนำมาวิเคราะห์เพื่อทำให้เกิดประโยชน์ เรียกขั้นตอนการวิเคราะห์ข้อมูลนี้ว่า การวิเคราะห์บิ๊กดาต้า (Big Data Analytics) การวิเคราะห์ข้อมูลปริมาณมหาศาลนั้นต้องใช้บุคลากรที่มีความรู้ความสามารถทางด้านสถิติและเครื่องมือเฉพาะทาง การวิเคราะห์ข้อมูลบิ๊กดาต้าสามารถสร้างโอกาสทางธุรกิจให้กับผู้ประกอบการ เช่น กรณีผู้ให้บริการเครือข่ายโทรศัพท์มือถือ มีการนำข้อมูลการใช้งานโทรศัพท์มือถือมาวิเคราะห์พฤติกรรมการใช้งานโทรศัพท์มือถือของลูกค้าเพื่อทำการแบ่งกลุ่มลูกค้า ส่งผลให้ผู้ประกอบการสามารถพัฒนาบริการเพื่อให้ตรงกับความต้องการของลูกค้าเฉพาะกลุ่ม และทำการตลาดเฉพาะกลุ่ม (Individual Marketing) เพื่อเพิ่มยอดขาย

นอกจากนี้การวิเคราะห์ข้อมูลบิ๊กดาต้า ผู้ประกอบการยังสามารถวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ (Social Media Content) เพื่อให้ทราบถึงพฤติกรรมการใช้งานของลูกค้า และความคิดเห็นของลูกค้าที่มีต่อองค์กรธุรกิจ

ดังนั้น การเข้าสู่เศรษฐกิจดิจิทัลของประเทศไทย หากผู้ประกอบการสามารถนำเอาเทคโนโลยีบิ๊กดาต้ามาใช้ประโยชน์จะสามารถทำได้เปรียบทางธุรกิจเหนือคู่แข่ง

ในหัวข้อถัดไปจะกล่าวถึงคุณลักษณะของบิ๊กดาต้า การจัดการข้อมูลสำหรับบิ๊กดาต้า การวิเคราะห์ข้อมูลด้วยเทคโนโลยีบิ๊กดาต้า การวิเคราะห์บิ๊กดาต้ากับงานด้านการขาย และการตลาด Oracle Social Relationship Management และ ฮาดูป (Hadoop) ซอฟต์แวร์โอเพนซอร์สในการประมวลผลข้อมูลบิ๊กดาต้า

คุณลักษณะของบิกดาต้า

บิกดาต้า หมายถึง ปริมาณข้อมูลขนาดใหญ่ที่ประกอบไปด้วย 5 องค์ประกอบ (5Vs) โดยมีรายละเอียดในแต่ละองค์ประกอบ คือ (ศรีสมรัก อินทจันทร์ยง, 2558, 3-4)

1. ปริมาณ (Volume) องค์การมีข้อมูลในปริมาณมหาศาล รวบรวมจากแหล่งข้อมูลหลากหลาย แหล่งและหลากหลายประเภท ข้อมูลมีอัตราการเติบโตในอัตราก้าวหน้า (Exponential) เมื่อเทียบกับเวลาที่ผ่านไป ปริมาณข้อมูลเติบโตจากขนาดเทระไบต์ (Terabyte) เป็นเพตะไบต์ (Petabyte) หรือเอกซะไบต์ (Exabyte) ซึ่งระบบฐานข้อมูลที่ใช้กันอยู่ไม่สามารถดำเนินการให้เกิดประโยชน์เพื่อการตัดสินใจได้อย่างมีประสิทธิภาพรวมทั้งยังมีต้นทุนในการดำเนินงานที่สูงกว่า

2. ความหลากหลาย (Variety) ความหลากหลายที่มีในบิกดาต้า ไม่ได้มีเพียงความหลากหลายของแหล่งข้อมูล แต่ยังรวมถึงลักษณะรูปแบบของข้อมูลที่แตกต่างกันไปอันเนื่องมาจากแหล่งข้อมูลที่ต่างประเภทกัน แหล่งข้อมูลที่ส่งเข้าสู่บิกดาต้ามีตั้งแต่ระบบสารสนเทศที่ใช้ระบบฐานข้อมูลแบบดั้งเดิม ไปจนถึงฐานข้อมูลเอกสาร (Text Document) ข้อมูลจากการทำงานบนอินเทอร์เน็ต (Web logs, XML) อาร์เอฟไอดี (Radio Frequency Identification, RFID) จีพีเอส (Geographic Positioning System, GPS) การใช้สื่อสังคมออนไลน์ (Social Media) ไปรษณีย์อิเล็กทรอนิกส์ (E-mail) ระบบตัวรับรู้ (Sensor System) มาตรวัดต่างๆ ที่บันทึกข้อมูลแบบทันที (Meter) ข้อมูลจากรายงานมูลค่าหลักทรัพย์ที่เคลื่อนไหวจากห้องค้า (Stock Ticker Data) ภาพเคลื่อนไหว (Video) และเสียง (Audio) เป็นต้น ลักษณะของข้อมูลมีทั้งลักษณะที่มีโครงสร้าง (Structure) ชัดเจน คือ ข้อมูลที่สามารถประมวลผลได้ด้วยเทคโนโลยีฐานข้อมูลแบบดั้งเดิม ไปจนถึงข้อมูลที่มีความเป็นโครงสร้างต่ำหรือไม่มีโครงสร้าง (Semi-Structure and Un-Structure) ซึ่งมีปริมาณประมาณร้อยละ 80 ถึงร้อยละ 85 ของข้อมูลทั้งหมดในองค์การ (Sandra, Delen & Turban, 2015) ส่งผลต่อเทคโนโลยีในการประมวลผล ข้อมูลเหล่านี้เกิดขึ้นในองค์การที่มีรูปแบบการประกอบการ และด้วยปริมาณที่แตกต่างกัน

3. ความรวดเร็ว (Velocity) ข้อมูลที่เกิดจากอินเทอร์เน็ต สื่อสังคมออนไลน์ ตัวรับรู้ มาตรวัด ฯลฯ เหล่านี้สามารถเกิดขึ้นตลอดเวลา รวมทั้งระบบการรวบรวมข้อมูลที่ทำได้อย่างทันที (Real-Time) เสมือนหนึ่งมีการไหลของข้อมูลเข้ามาสู่ระบบอย่างสม่ำเสมอ

4. ความผันแปร (Variability) นอกเหนือจากคุณสมบัติที่มีความหลากหลาย และรวดเร็วในกระแสข้อมูลแล้ว กระแสการไหลของข้อมูลยังมีระดับที่ไม่คงที่ มีความไม่สม่ำเสมอค่อนข้างสูง โดยเฉพาะกระแสข้อมูลที่มาจากสื่อสังคมออนไลน์ที่จะมีผลกระทบต่อบริมาณการไหลของข้อมูลที่สูงมากเมื่อมีเหตุการณ์บางอย่างเกิดขึ้น ทั้งนี้จะมีการเกิดเป็นช่วงๆ ของเวลา

5. คุณค่าที่น่าเสนอ (Value Proposition) ข้อมูลจำนวนมหาศาลจะมีรูปแบบความสัมพันธ์ของข้อมูลที่เห็นความแตกต่างหรือความแปลกใหม่ที่ปรากฏได้ยากจากข้อมูลที่มีจำนวนไม่มากพอ สิ่งเหล่านี้จะเป็นคุณค่าทางธุรกิจอย่างมากถ้าองค์กรค้นพบได้ก่อนคู่แข่ง บิกดาต้าจึงไม่ได้เป็นเพียงแค่การรวบรวมข้อมูล แต่หมายถึงการวิเคราะห์ข้อมูลเพื่อค้นหาสิ่งใหม่ให้กับองค์กร

การจัดการข้อมูลสำหรับบิ๊กดาต้า

การจัดการข้อมูลสำหรับบิ๊กดาต้าประกอบไปด้วย การจัดเก็บ (Storage) การประมวลผล (Processing) การวิเคราะห์ (Analysis Algorithm) และการทำรายงานสรุป (Visualization) โดยผู้ประกอบการที่สามารถวิเคราะห์ข้อมูลจากบิ๊กดาต้าได้ย่อมจะได้เปรียบเหนือคู่แข่ง ทั้งนี้เรื่องสำหรับผู้ประกอบการควรให้ความสำคัญ คือ การประยุกต์ใช้เครื่องมือ (Tools) ทางด้านบิ๊กดาต้าที่เหมาะสม เพื่อตอบโจทย์กลยุทธ์ของธุรกิจ เช่น การวิเคราะห์ทางการตลาด เป็นต้น ในบทความนี้จะมีการนำเสนอเกี่ยวกับเทคโนโลยี Hadoop ซึ่งเป็นหนึ่งในเครื่องมือบิ๊กดาต้าที่ได้รับความนิยมอย่างกว้างขวาง เนื่องจากเป็นซอฟต์แวร์โอเพนซอร์ส ที่สร้างขึ้นมาเพื่อรองรับในการประมวลผลแบบกระจาย (Distributed Processing) สำหรับข้อมูลขนาดใหญ่

การวิเคราะห์ข้อมูลด้วยเทคโนโลยีบิ๊กดาต้า

การวิเคราะห์ข้อมูลด้วยเทคโนโลยีบิ๊กดาต้าประกอบด้วย 3 ส่วน คือ

1. ข้อมูล (Data) การวิเคราะห์ข้อมูลบิ๊กดาต้าเหมาะกับการประมวลผลข้อมูลที่มีความหลากหลาย โดยลักษณะของข้อมูลที่จัดเก็บเป็นทั้งข้อมูลที่มีโครงสร้างและไม่มีโครงสร้าง จัดเก็บอยู่ร่วมกัน มาจากแหล่งข้อมูลที่หลากหลาย มีการเปลี่ยนแปลงตลอดเวลาและมีแนวโน้มในการเติบโตของข้อมูลที่สูงขึ้น ดังนั้นผู้ประกอบการควรพิจารณาข้อมูลว่าเหมาะสมต่อการวิเคราะห์ด้วยบิ๊กดาต้าหรือสามารถใช้การจัดการข้อมูลแบบเดิมก็เพียงพอ (ศรีสมรัก อินทจันทร์ยง, 2558, 6)

2. โครงสร้างพื้นฐาน (Infrastructure) ควรพิจารณาความพร้อมทางด้านเทคโนโลยีขององค์กร ที่เกี่ยวข้องกับสื่อที่ใช้ในการจัดเก็บข้อมูล และความสามารถในการประมวลผลข้อมูล ดังนี้

- 2.1 ด้านการจัดเก็บข้อมูลปริมาณมหาศาลสามารถใช้บริการแหล่งเก็บข้อมูลบนก้อนเมฆ (Cloud Storage) ซึ่งมีผู้ให้บริการตั้งแต่ระดับนานาชาติ เช่น Amazon หรือ Google ไปจนกระทั่งผู้ให้บริการอินเทอร์เน็ต (Internet Service Provider: ISP) และ ผู้ขาย (Vendor) ในแต่ละประเทศ

- 2.2 สำหรับฐานข้อมูลบิ๊กดาต้า เหมาะสำหรับเทคโนโลยี NoSQL ที่มุ่งเน้นการจัดการข้อมูลที่มีพลวัต (Dynamic) เช่น Apache Cassandra ซึ่งสามารถอ่านเขียนข้อมูลได้อย่างรวดเร็ว ทำให้สามารถทำงานกับข้อมูลปริมาณมหาศาลของบิ๊กดาต้าได้ และยังเป็นซอฟต์แวร์ในรูปแบบโอเพนซอร์ส

- 2.3 ด้านการประมวลผลข้อมูลบิ๊กดาต้าในปัจจุบันเทคโนโลยีที่นิยมใช้ได้แก่ Hadoop ซึ่งเป็นเทคโนโลยีการประมวลผลแบบขนาน (Parallel Processing) มีความสามารถในการบริหารจัดการงานประมวลผลโดยกระจายงานออกไปยังการประมวลผลแบบกลุ่มเมฆ (Cloud Processing) เพื่อประมวลผลและรวบรวมผลลัพธ์กลับมาให้ผู้ใช้ ซึ่งสามารถรองรับการทำงานด้านบิ๊กดาต้าได้อย่างมีประสิทธิภาพ อีกทั้งยังเป็นซอฟต์แวร์ในรูปแบบโอเพนซอร์ส (กมลภัทร บุญคำ, 2558, 5)

3. บุคลากร เป็นหัวใจสำคัญของบิ๊กดาต้า คือ สามารถวิเคราะห์ข้อมูลปริมาณมหาศาลเพื่อค้นหาสิ่งใหม่ หรือหาความสัมพันธ์ระหว่างข้อมูลที่ไม่เคยทราบมาก่อน ดังนั้นองค์กรจึงจำเป็นต้องพัฒนาบุคลากรให้มีความเข้าใจในข้อมูล ลักษณะของธุรกิจ และเทคโนโลยีที่ใช้ เช่น นักวิทยาศาสตร์ข้อมูล (Data Scientist)

หัวหน้าทีมข้อมูล (Chief Data Officer) สถาปัตยกรรมข้อมูล (Data Architecture) และการวิเคราะห์ธุรกิจอัจฉริยะ (BI Analysis) เป็นต้น (ธนชาติ นุ่มนนท์, 2558, 2)

การวิเคราะห์ที่บิดาทำงานด้านการขายและการตลาด

การวิเคราะห์ที่บิดาทำงานด้านการขายและการตลาด มีความสำคัญอย่างยิ่งสำหรับผู้ประกอบการ เนื่องจากปัจจุบันมีการแข่งขันทางธุรกิจมากขึ้น ผู้ประกอบการที่สามารถนำข้อมูลขนาดใหญ่มาวิเคราะห์ได้ ย่อมส่งผลให้ผู้ประกอบการรายนั้นได้เปรียบคู่แข่ง

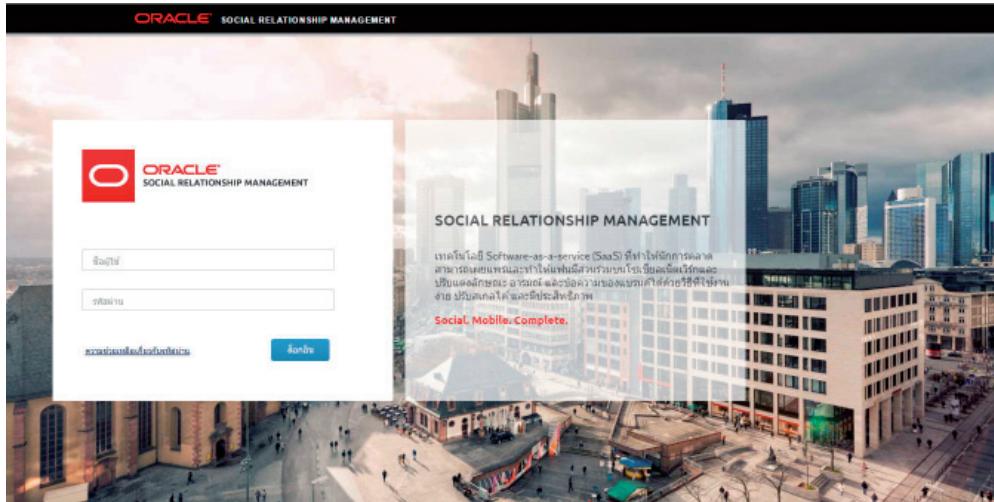
วิธีการแก้ปัญหา (Solution) ที่เกี่ยวกับการตลาด (Marketing-Based Applications) เป็นวิธีการแก้ปัญหาที่รวบรวมข้อมูลแนวคิดของผู้บริโภค หรือผู้ใช้สินค้าและบริการขององค์กรมาทำการวิเคราะห์เพื่อที่จะได้ตอบสนองต่อผู้ใช้สินค้าและบริการได้ตรงใจ ข้อมูลแนวคิดที่องค์กรต้องการ คือ ข้อมูลในเชิงพฤติกรรม (Behavioral Data) ตามมาด้วยข้อมูลความคิดเห็น (Attitudinal Data) ข้อมูลแสดงความชอบ (Clickstream Data) รวมด้วยข้อมูลในเชิงข้อเท็จจริงของผู้ใช้สินค้า (Description Data) โดยจุดประสงค์ของวิธีการแก้ปัญหานี้ คือ ความสามารถในการเพิ่มยอดขาย (Upselling) และการขยายช่องทางการขาย (Cross-Selling) สำหรับธุรกิจ นอกจากนี้ข้อมูลจากสื่อสังคมออนไลน์ (Social Media Content) มีความจำเป็นต่อธุรกิจขององค์กร โดยวิธีการของบิดาจะทำการดึงข้อมูลจากสื่อสังคมออนไลน์ ไม่ว่าจะผ่านบริการของเจ้าของสื่อแต่ละที่ หรือจะผ่านบริษัทกลางที่ทำหน้าที่ให้บริการมาทำการประมวลผลโดยใช้วิธีการแก้ปัญหาที่เรียกว่า การวิเคราะห์อารมณ์และความรู้สึกจากข้อความ (Sentiment Analysis) ซึ่งในปัจจุบันมีซอฟต์แวร์ที่สามารถทำได้ และสามารถติดต่อกับสื่อสังคมออนไลน์ได้ครบในตัวเองกัน เพียงแต่ต้องแยกซื้อเป็นทางเลือกไป โดยจุดประสงค์ของ Solution นี้คือ ต้องการทราบพฤติกรรมการใช้งานของผู้บริโภค รวมถึงความคิดเห็นของผู้บริโภคที่มีต่อองค์กร (ณรงคฤทธิ์ มโนมัยพิบูลย์, 2557, 19)

กรณีศึกษาเว็บไซต์อาลีบาบา (Alibaba) ซึ่งมีการใช้เทคโนโลยีที่บิดาเพื่อวิเคราะห์และตัดสินใจในการให้บริการสินเชื่อแก่ร้านค้า ตลอดทั้งกระบวนการมีการวิเคราะห์ข้อมูลร้านค้าอัตโนมัติโดยไม่มีการแทรกแซงจากบุคคลอื่น จากผลการดำเนินงานมีรายงานหนี้สูญ (Bad Loans) เพียง 0.3 เปอร์เซ็นต์ซึ่งต่ำกว่าธนาคารพาณิชย์ทั่วไป (Chen, Mao & Lin, 2014, 198)

Oracle Social Relationship Management

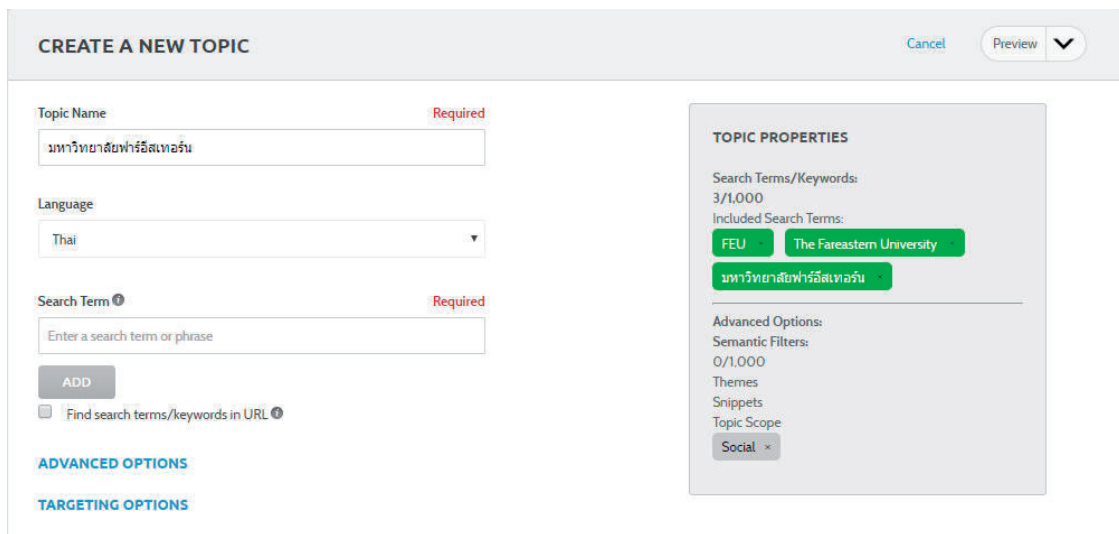
Oracle Social Relationship Management คือ ซอฟต์แวร์ที่ทำงานบน Cloud มีฟังก์ชันในการจับ Feedback ของผู้บริโภคจากเว็บไซต์ และเครือข่ายสังคมออนไลน์ อาทิ Facebook, Twitter, LinkedIn, Google +, YouTube และสื่อสังคมออนไลน์อื่นๆ โดยนำข้อมูลที่ได้จากโพสต์ และความคิดเห็นต่างๆ มาประมวลผล และยังมีเครื่องมือที่สามารถติดต่อกับกลุ่มผู้บริโภคได้ในระยะเวลารวดเร็ว ซึ่งจะช่วยตอบโต้ทางธุรกิจให้กับนักการตลาดที่ต้องการความรวดเร็วของข้อมูลได้ มีเครื่องมือในการติดตามและวิเคราะห์ข้อมูล เพื่อกำหนดทิศทางในการทำการตลาดได้อย่างทันท่วงที ส่งผลให้สามารถแข่งขันและก้าวเป็นผู้นำทางการตลาดได้โดยง่าย (ASTV ผู้จัดการออนไลน์, 2557)

หน้าจอการใช้งานซอฟต์แวร์ Oracle Social Relationship Management ผ่านโปรแกรมเว็บเบราว์เซอร์
ดังภาพที่ 3



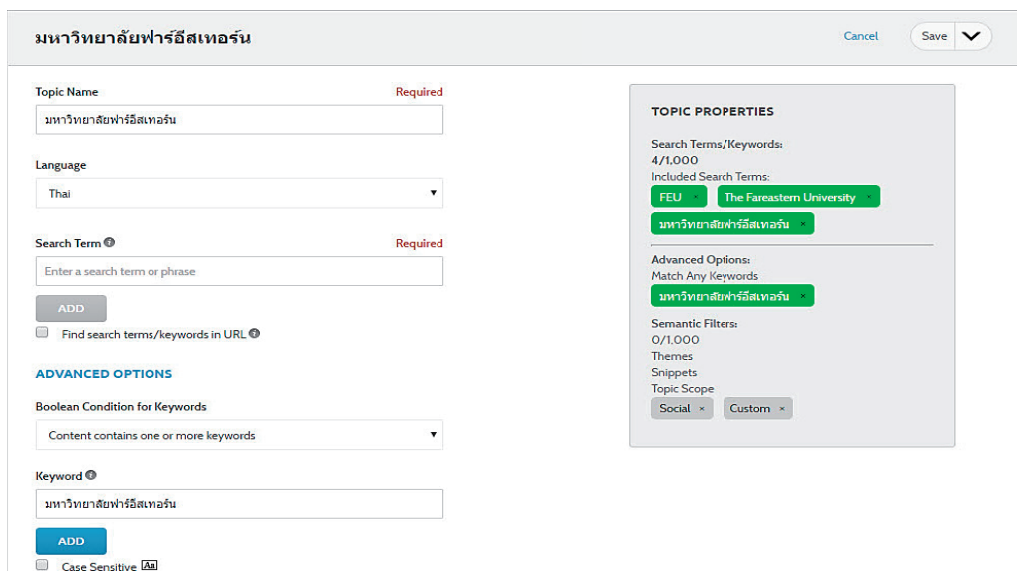
ภาพที่ 3: ซอฟต์แวร์ Oracle Social Relationship Management
ที่มา: ORACLE, 2016

การใช้งานซอฟต์แวร์ Oracle Social Relationship Management กรณีที่ต้องการทราบความคิดเห็นหรือข้อมูลบนเครือข่ายสังคมออนไลน์ที่เกี่ยวข้องกับองค์กร สามารถระบุคำที่ต้องการค้นหา โดยสามารถระบุคำที่ต้องการค้นหาได้มากกว่าหนึ่งคำ ดังภาพที่ 4



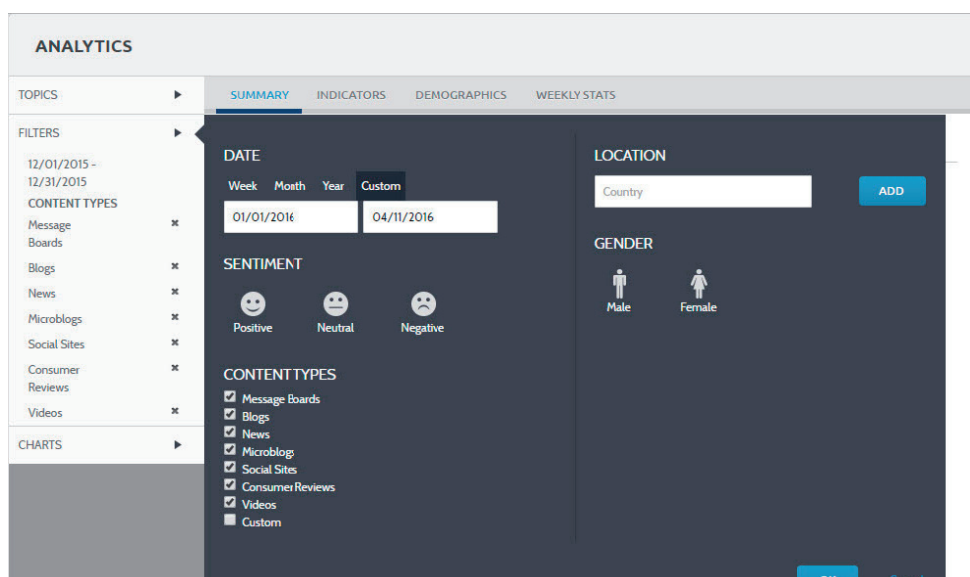
ภาพที่ 4: หน้าจอการเลือกใช้เมนู Search Term เพื่อกำหนดคำที่ต้องการค้นหา
ที่มา: ORACLE, 2016

นอกจากนั้นภาพที่ 5 แสดงหน้าจอการกำหนดเงื่อนไขเพิ่มเติมสำหรับคำที่ต้องการค้นหา เช่น เนื้อหาที่ทำการวิเคราะห์นั้นจะประกอบไปด้วยคำที่ต้องการค้นหาอย่างน้อยหนึ่งคำ



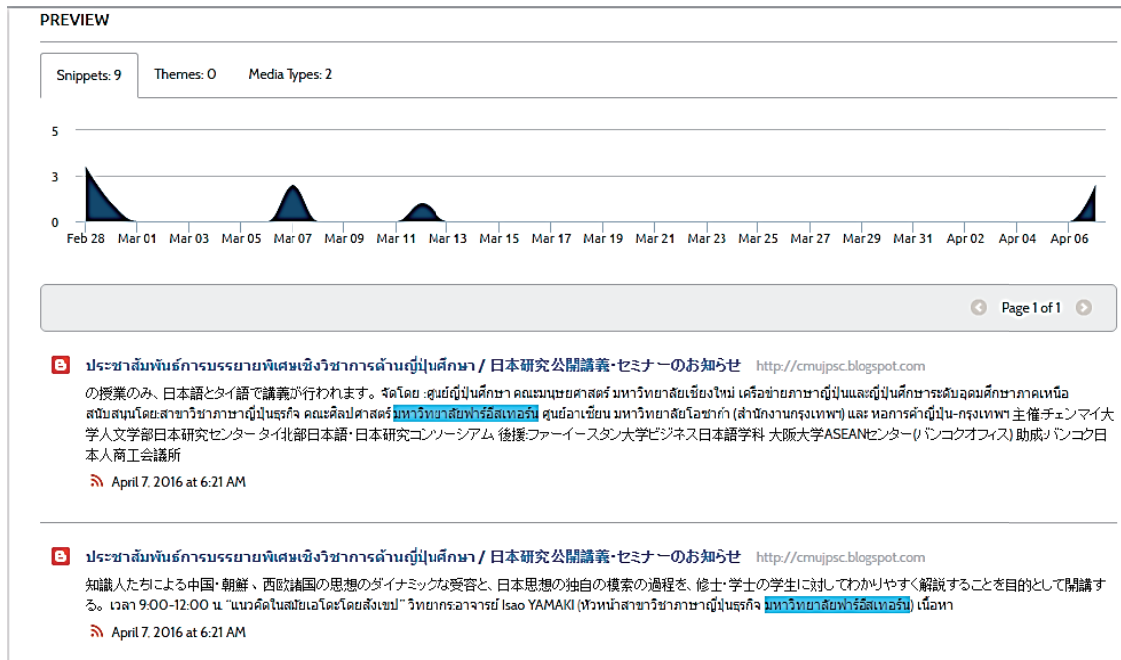
ภาพที่ 5: หน้าจอการเลือกใช้เมนู ADVANCED OPTION
ที่มา: ORACLE, 2016

หน้าจอสำหรับการกรองไฟล์ (Filter) ข้อมูลในการวิเคราะห์ข้อมูล ดังภาพที่ 6



ภาพที่ 6: หน้าจอการเลือกใช้เมนู Filter
ที่มา: ORACLE, 2016

ผลการตรวจสอบผลสะท้อนกลับ (Feedback) ของผู้บริโภคจากเว็บไซต์และเครือข่ายสังคมออนไลน์ ที่กล่าวถึงคำหลัก คือ มหาวิทยาลัยพะเยา ในช่วงเวลาที่กำหนด ดังภาพที่ 7



ภาพที่ 7: หน้าจอการเลือกใช้เมนูตรวจสอบ (Preview)

ที่มา: ORACLE, 2016

Hadoop : ซอฟต์แวร์โอเพนซอร์สในการประมวลข้อมูลบิ๊กดาต้า

1. Hadoop

ซอฟต์แวร์ที่สำคัญที่นิยมใช้ในระบบบิ๊กดาต้า คือ Hadoop ซึ่งเป็นซอฟต์แวร์โอเพนซอร์ส ทำหน้าที่ในการแบ่งการประมวลผลและจัดเก็บข้อมูลขนาดใหญ่ได้ สามารถเพิ่มขนาดได้อย่างไม่จำกัด โดยจะประกอบไปด้วยส่วนหลัก คือ Hadoop Distributed File System (HDFS) ทำหน้าที่เป็นที่เก็บข้อมูลแบบขนาน และกระจาย และใช้ MapReduce ที่เป็นเทคนิคการประมวลข้อมูลแบบขนาน โดยนักพัฒนาสามารถพัฒนาโปรแกรมโดยใช้ภาษาจาวาวิเคราะห์ข้อมูลในรูปแบบของฟังก์ชันการ Map และ Reduce โดยระบบจะกระจายข้อมูลไปประมวลแบบขนานบนเครื่องหลายๆ เครื่อง การประมวลผลจะเป็นแบบกลุ่ม (Batch Offline) ใช้เวลาพอสมควรในการประมวลผล ทั้งนี้ขึ้นอยู่กับขนาดข้อมูล (ธนชาติ นุ่มนนท์, 2557, 1)

2. Hadoop Ecosystem

ระบบ Hadoop จะมีส่วนประกอบหลักอยู่สองส่วนคือ HDFS และ MapReduce ซึ่งค่อนข้างจะไม่สะดวกกับผู้ใช้งานที่มีความต้องการอื่นๆ เช่น การประมวลผลโดยใช้ภาษา SQL การเขียนหรืออ่านข้อมูลแบบ Random Access หรือการถ่ายโอนข้อมูลจากแหล่งอื่นๆ จึงมีการพัฒนาเครื่องมือในรูปแบบอื่นๆ ร่วมกับ Hadoop เพื่อให้ได้ประสิทธิภาพดียิ่งขึ้น ดังตัวอย่างของเครื่องมือต่อไปนี้

2.1 Apache Hive เป็นเครื่องมือสำหรับผู้ต้องการสืบค้น (Query) ข้อมูลที่เก็บใน HDFS ด้วยภาษาลักษณะ SQL แทนที่จะต้องเขียนโปรแกรม MapReduce ด้วยภาษาจาวา โดย Hive จะทำหน้าที่ในการแปล SQL Like ให้มาเป็น MapReduce แล้วทำการประมวลผลแบบ Batch ดังตัวอย่างในภาพที่ 8

```
SELECT age FROM people
WHERE province = 'bangkok'
GROUP BY city
ORDER BY age DESC
```

ภาพที่ 8 ตัวอย่าง HiveQL

ที่มา: Siam HTML, 2558

2.2 Apache Pig เป็นเครื่องมือคล้ายกับ Hive ที่ช่วยให้ประมวลผลโดยไม่ต้องเขียนโปรแกรม MapReduce ซึ่ง Pig จะใช้โปรแกรมภาษา Script ง่ายๆ ที่เรียกว่า Pig Latin ทำให้ผู้เขียนโปรแกรมสามารถสืบค้นข้อมูลด้วยคำสั่งที่ไม่ซับซ้อน ดังตัวอย่างในภาพที่ 9

```
A = LOAD 'myfile.txt' USING PigStorage() AS (t, u, v);
B = GROUP A BY t;
C = FOREACH B GENERATE group, COUNT(A.t) as mycount;
D = ORDER C BY mycount;
STORE D INTO 'mysortedcount' USING PigStorage();
```

ภาพที่ 9: ตัวอย่างโค้ด Pig Latin

ที่มา: Siam HTML, 2558

2.3 Sqoop เป็นเครื่องมือในการถ่ายโอนข้อมูลระหว่างฐานข้อมูลที่อยู่ในรูปแบบของตาราง (Table) บนระบบการจัดการฐานข้อมูลเชิงสัมพันธ์ (Relation Database Management System: RDBMS) เช่น SQL Server, Oracle, Oracle หรือ MYSQL กับข้อมูลบน HDFS ของ Hadoop

2.4 Flume เป็นเครื่องมือในการดึงข้อมูลจากระบบอื่นๆ แบบ Real Time เข้าสู่ HDFS เช่น การดึง Log จาก Web Server การดึงข้อมูลเหล่านี้จะต้องมีการติดตั้ง Agent ที่เครื่อง Server



2.5 HBase เป็นเครื่องมือที่จะทำให้ Hadoop สามารถอ่านและเขียนข้อมูลแบบ Real-Time Random Access ได้โดยจะทำให้เป็น BigTable ที่เก็บข้อมูลได้ไม่จำกัดแถว (Row) หรือคอลัมน์ (Column) ซึ่ง HBase จะเป็นเสมือนการทำให้ Hadoop เป็น NoSQL Database

2.6 Oozie เป็นเครื่องมือในการทำ Workflow จะช่วยให้สามารถนำคำสั่งประมวลผลต่างๆ ของระบบ Hadoop เช่น MapReduce, Hive หรือ Pig มาเชื่อมต่อในรูปของ Workflow ได้

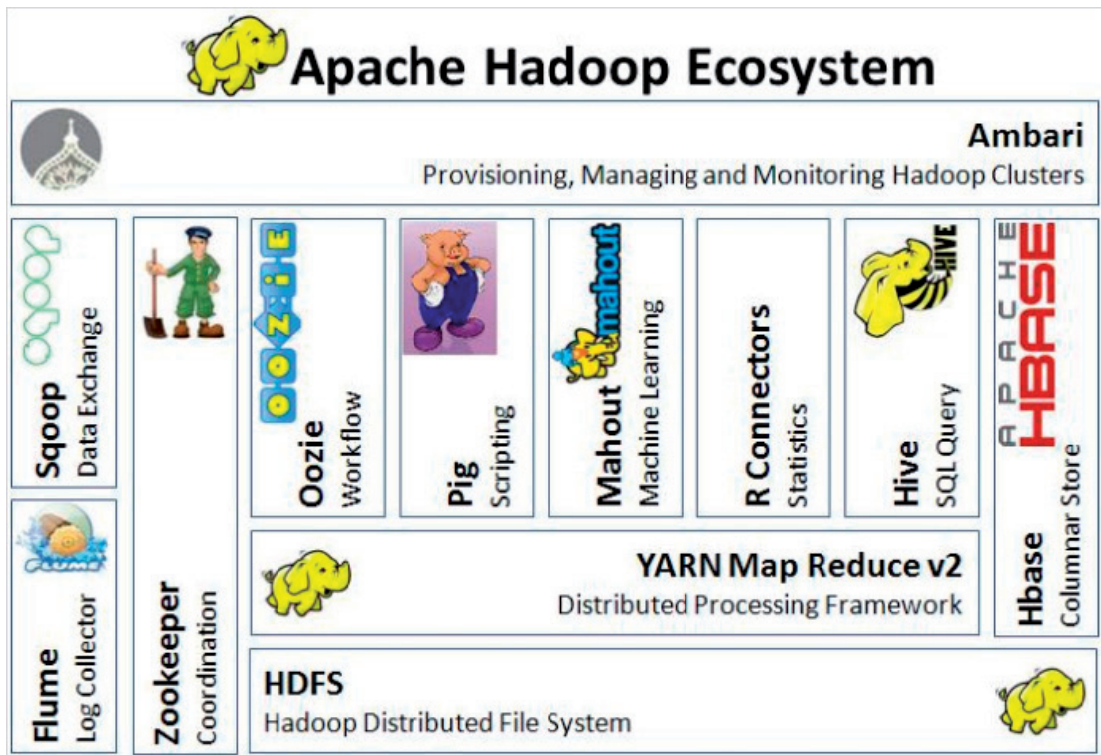
2.7 Hue เป็นเครื่องมือที่ช่วยทำ User Interface ของ Hadoop ให้ใช้งานได้ง่ายขึ้นกว่าการใช้ Command Line

2.8 Mahout เป็นเครื่องมือของ Data Scientist ที่ต้องการทำ Predictive Analytics ของข้อมูลบน Hadoop โดยใช้ภาษาจาวา ทั้งนี้ Mahout สามารถใช้ Algorithm ที่เป็น Recommender, Classification และ Clustering ได้เช่นกัน (ธนชาติ นุ่มนนท์, 2557, 3)

2.9 Yarn เป็นเทคโนโลยีในการบริหารคลัสเตอร์ใน Hadoop รุ่นที่ 2 คือ ระบบปฏิบัติการแบบกระจายตัวขนาดใหญ่ (Large-scale Distributed Operating System) โดยแยกเอาความสามารถในการเรื่องการทำ Scheduling และการบริหารทรัพยากรออกมาจากองค์ประกอบด้านการประมวลผลข้อมูล ทำให้ Hadoop รองรับการผลิตที่หลากหลาย และรองรับแอปพลิเคชันได้กว้างขวางมากขึ้น เช่น ทำให้ Hadoop Clusters สามารถทำภารกิจวีเบอินเทอร์แอคทีฟ และใช้งานแอปพลิเคชันสตรีมมิงดาต้าไปพร้อมกับ Batch ของ MapReduce ได้ด้วย

2.10 Spark เป็นเครื่องมือสำหรับทำ Data Processing ที่สร้างบน Hadoop โดยมีความเร็วในการประมวลผลเร็วกว่า MapReduce ของ Hadoop ถึง 10-100 เท่า และยังช่วยให้ผู้เขียนโปรแกรมสามารถเขียน Query ได้ง่ายเหมือนกับการใช้ HiveQL (กมลภัทร บุญคำ, 2558, 5)

จากที่กล่าวมาแสดงให้เห็นถึงพันธมิตรของ Hadoop (Hadoop Ecosystem) ที่ช่วยเสริมในเรื่องการจัดการข้อมูล การเข้าถึงข้อมูล และการติดต่อแลกเปลี่ยนข้อมูลกับระบบต่างๆ ให้มีความสะดวกมากยิ่งขึ้น โดยมี Hadoop เป็นแกนกลางในการทำงาน ดังแสดงรายละเอียดในภาพที่ 10



ภาพที่ 10: Hadoop Ecosystem

ที่มา: Bappalige, 2557

สำหรับผู้ที่ต้องการใช้ Hadoop ในการแก้ปัญหาการจัดการบิ๊กดาต้าสามารถเลือกใช้เครื่องมือที่กล่าวมาแล้วในข้างต้นตามความเหมาะสมของผู้ประกอบการแต่ละราย ยกตัวอย่างเช่น องค์กรที่มีข้อมูลขนาดใหญ่ ต้องการวิเคราะห์ข้อมูลแบบอินเตอร์แอคทีฟ (Interactive) มีการประมวลผลแบบ Real-Time สามารถเลือกใช้การประมวลผลโดยใช้ Spark เนื่องจากมีความเร็วในการประมวลผลสูง แต่กรณีองค์กรที่มีข้อมูลขนาดใหญ่แต่ไม่มีความจำเป็นในการประมวลผลในระดับสูงที่ซับซ้อนอย่าง Spark สามารถเลือกใช้เครื่องมือตัวอื่นที่มีความเหมาะสมในด้านของต้นทุนแทนเครื่องมือดังกล่าว

ปัจจุบันมีซอฟต์แวร์ทางการค้า (Commercial Software) ที่นำเทคโนโลยีบิ๊กดาต้ามาพัฒนาเครื่องมือเพิ่มเติมเพื่อให้สามารถใช้งาน หรือจัดการได้ง่ายขึ้น นอกจากนี้ยังมีบริการหลังการขายในรูปแบบต่างๆ เพื่อให้องค์กรสามารถได้ข้อมูลที่มีประโยชน์ต่อธุรกิจ ได้แก่ Oracle BDA, Cloudera, IBM BigInsights, Amazon EMR และ MapR เป็นต้น



สรุป

จากที่กล่าวมาการใช้งานบิ๊กดาต้าจะอยู่ในกลุ่มของผู้ประกอบการขนาดใหญ่ที่มีความพร้อมทั้งทางด้านข้อมูล โครงสร้างพื้นฐาน และบุคลากร โดยการได้มาซึ่งข้อมูลที่เป็นประโยชน์สำหรับงานด้านการตลาดของผู้ประกอบการ สามารถจัดหาได้โดยเทคโนโลยีบิ๊กดาต้าทั้งที่เป็นการพัฒนาจากซอฟต์แวร์ Hadoop และ Ecosystem Hadoop ในรูปแบบซอฟต์แวร์โอเพนซอร์ส หรือจัดหาโปรแกรมในรูปแบบของซอฟต์แวร์ทางการค้าที่มีให้บริการอยู่หลายบริษัท ทั้งนี้ขึ้นอยู่กับความเหมาะสมของแต่ละองค์กรสำหรับธุรกิจขนาดกลางและขนาดย่อม เนื่องจากไม่มีระบบโครงสร้างพื้นฐานของตนเอง และไม่สามารถซื้อข้อมูลจากหลายๆ แหล่งข้อมูลเนื่องจากมีราคาสูง แนวทางในการจัดการข้อมูลสำหรับบิ๊กดาต้าของผู้ประกอบการขนาดกลางและขนาดย่อม คือ การใช้บริการบริษัทที่เกี่ยวกับการวิเคราะห์บิ๊กดาต้าโดยเฉพาะ เช่น Insightsquared, Google, IBM, SAP และ Oracle เป็นต้น ซึ่งเป็นระบบที่มีโครงสร้างพื้นฐานและบุคลากรที่มีความพร้อมในด้านการวิเคราะห์บิ๊กดาต้า โดยผู้ประกอบการต้องการทราบเรื่องใดบริษัทจะดำเนินการวิเคราะห์และขายเป็นรายงานให้กับผู้ประกอบการในลักษณะการคิดค่าบริการตามการใช้งานจริง

เอกสารอ้างอิง

- กมลภัทร บุญคำ. (2558). Turning Big Data into Data Analytics. สืบค้นเมื่อ 12 เมษายน 2559, จาก <http://www.theeleader.com/turning-big-data-data-analytics/>.
- ณรงค์ฤทธิ์ มโนมัยพิบูลย์. (2557). ประยุกต์ Big Data ในงานธุรกิจ. Gmagz. (37), 19-20.
- ธนาชาติ นุ่มนนท์. (2557). Hadoop Ecosystem สำหรับการพัฒนา Big Data. สืบค้นเมื่อ 30 มีนาคม 2559, จาก <https://thanachart.org/2014/10/18/ปี-2014-hadoop-ecosystem-สำหรับการพัฒนา-big-data/>.
- ธนาชาติ นุ่มนนท์. (2558). ปี 2015 จะเป็นปีเริ่มต้นของ Big Data Analytics. สืบค้นเมื่อ 30 มีนาคม 2559, จาก <https://thanachart.org/2015/01/02/ปี-2015-จะเป็นปีเริ่มต้นของ-big/>.
- ศรีสมร อันทุจันทรียง. (2558). Big Data จำเป็นสำหรับทุกองค์กรหรือไม่. วารสารนักบริหาร. 38 (146), 1-8.
- สุรนาถ เนียมคำ. (2558.). Big Data คืออะไร? + วิธีใช้ Hadoop/Spark บน Cloud Dataproc. สืบค้นเมื่อ 5 เมษายน 2559, จาก <http://www.siamhtml.com/getting-started-with-big-and-hadoop-spark-on-cloud-dataproc/>.
- ASTV ผู้จัดการออนไลน์. (2557). ISSP ผนึกพลังพันธมิตร Oracle ให้บริการ Oracle SRM. สืบค้นเมื่อ 13 เมษายน 2559, จาก <http://www.manager.co.th/Cyberbiz/ViewNews.aspx?NewsID=9570000115734>.
- Bappalige, S. P. (2557). An introduction to Apache Hadoop for big data. Retrieved 12 April 2016, from <https://opensource.com/life/14/8/intro-apache-hadoop-big-data>.



- Chen, M.; Mao, S. & Liu, Y. (2014). Big Data: A Survey. **Mobile New Application**. 19, 171-209.
- DOMO. (2015). **Data Never Sleep 3.0**. Retrieved 13 April 2016, from <https://www.domo.com/learn/data-never-sleeps-3-0>.
- ET Center. (2015). **Effective Big Data Solution for Content Owners**. Retrieved 13 April 2016, from <http://www.slideshare.net/ETCenter/5-onprem-and-the-cloud-effective-big-data-solutions-and-services-for-content-owners>.
- ORACLE. (2016). **Oracle Social Relationship Management**. Retrieved 13 April 2016, from <https://cas.vitruue.com/logout>.
- Sandra, R.; Delen, D. & Turban, E. (2015). **Business Intelligence and Analytics: Systems for Decision Support**. (10th ed.). N.J.: Pearson Education.