

ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากรภาครัฐทุกระดับ

โดย นายองอาจ สุขชนู
นักวิชาการแผนกที่ภาพถ่ายปฏิบัติการ

วัตถุประสงค์

๑. เพื่อให้มีความรู้เกี่ยวกับประเภทข้อมูลต่าง ๆ
๒. เพื่อให้มีความรู้ความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูลและระดับการวิเคราะห์ข้อมูล
๓. เพื่อให้เกิดความเข้าใจและสามารถปรับตัวในการใช้ข้อมูลสำหรับปฏิบัติงาน

สรุปเนื้อหา

Big Data

WHAT IS BIG DATA? Big Data คือข้อมูลที่มีขนาดใหญ่ มีทั้งแบบโครงสร้างปกติและโครงสร้างข้อมูลที่ไม่มีรูปแบบ ซึ่งทั้งหมดเป็นข้อมูลที่ใช้ในเชิงธุรกิจ มักจะถูกนำไปใช้กับงานที่ต้องวิเคราะห์ ข้อมูลที่มีความซับซ้อน และไม่สามารถประเมินขนาดข้อมูลได้



รูปที่ ๑ : WHAT IS BIG DATA ?

ที่มา : <https://www.scnsoft.com/data/what-is-big-data>

รูปแบบของข้อมูล Big Data สามารถเป็นไปได้หลากหลาย ได้แก่

๑. Behavioral Data ได้แก่ ข้อมูลเชิงพฤติกรรมการใช้งานต่าง ๆ เช่น server log พฤติกรรมการคลิกดูข้อมูล หรือข้อมูลการใช้งาน ATM

๒. Image & Sounds เช่น ภาพถ่าย วีดีโอ รูปจาก Google Street View ภาพถ่ายทางการแพทย์ลายมือ ข้อมูลเสียงที่บันทึกไว้

๓. Languages เช่น Text Message ข้อความที่ถูก Tweet เนื้อหาต่าง ๆ ในเว็บไซต์

๔. Records เช่น ข้อมูลทางการแพทย์ ข้อมูลผลการสำรวจที่มีขนาดใหญ่ ข้อมูลภาษี

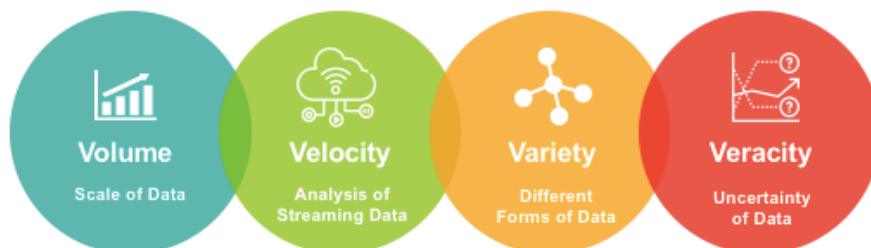
๕. Sensors เช่น ข้อมูลอุณหภูมิ ข้อมูลภูมิศาสตร์

Big Data ประกอบด้วยคุณลักษณะ ๔ ประการ คือ

๑. Volume ข้อมูลที่มีขนาดใหญ่ มีปริมาณข้อมูลมาก ซึ่งสามารถเป็นได้ทั้งแบบ Offline หรือ Online

๒. Variety ข้อมูลที่มีความหลากหลาย สามารถเป็นได้ทั้งที่มีโครงสร้างและข้อมูลที่ไม่สามารถจับ Pattern ได้

๓. Velocity ข้อมูลที่มีการเปลี่ยนแปลงตลอดเวลาอย่างรวดเร็ว มีการส่งผ่านข้อมูลอย่างต่อเนื่อง
 ในลักษณะ Streaming ทำให้การวิเคราะห์ข้อมูลแบบ Manual มีข้อจำกัด
๔. Veracity ข้อมูลมีความไม่ชัดเจน (Untrusted, Uncleaned)



รูปที่ ๒ : ลักษณะของข้อมูล Big Data

ที่มา https://www.dcaasia.co.th/th/solutions/cloud_infrastructure__analytics/Cloud-infrastructure_Big_data_Platform

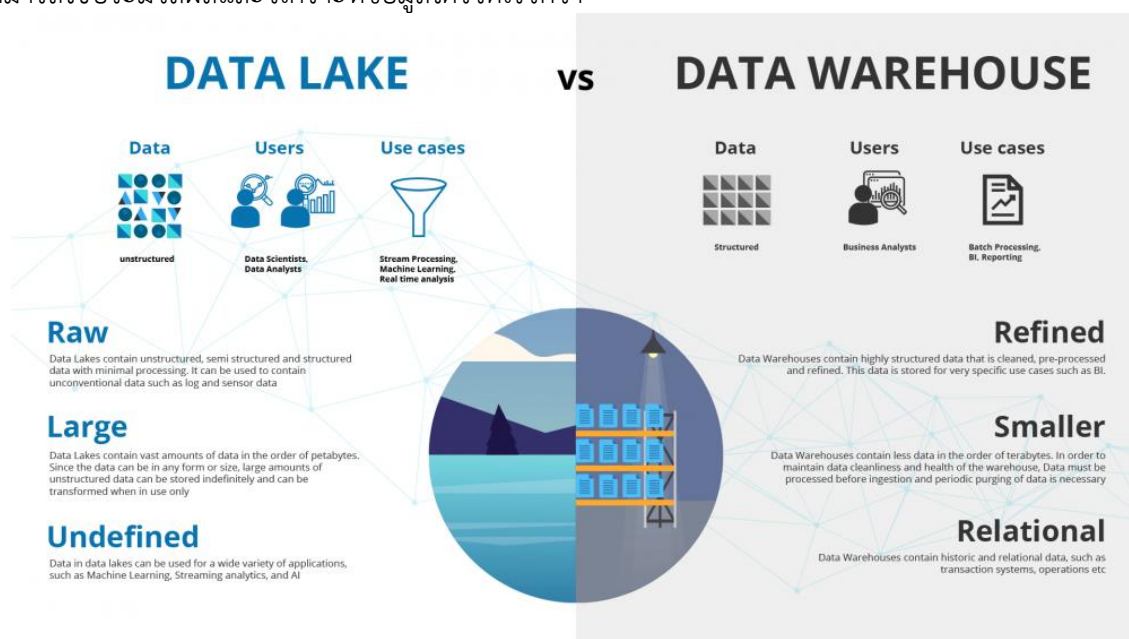
Data Lake

Data Lake เกิดขึ้นเนื่องจากการนำเอาข้อมูลจากแหล่งข้อมูลภายนอกองค์กร ข้อมูลจากเครือข่าย ข้อมูลที่กระจายไปทั่วโลกมาใช้มากขึ้น ปริมาณข้อมูลจากแหล่งภายนอกเพิ่มอย่างต่อเนื่องและมีแนวโน้มที่จะเติบโตแบบก้าวกระโดดมากขึ้น Data Lake แก้ไขข้อจำกัดหลายอย่างของ Data Warehouse ที่ใช้กันมานาน ข้อมูลที่จัดเก็บใน Data Lake คือ

- ข้อมูลดิบจำนวนมากและมีขนาดใหญ่
- ข้อมูลไม่มีรูปแบบแน่นอน
- การเข้าถึงข้อมูลไม่สามารถเข้าถึงได้ง่าย ต้องใช้ความสามารถของเจ้าหน้าที่วิเคราะห์ข้อมูล (Data Scientist)

ความแตกต่างระหว่าง Data Lake กับ Data Warehouse

Data Lake นั้นเก็บข้อมูลทั้งหมด สนับสนุนข้อมูลทุกชนิดไม่ใช่เพียงข้อมูลแบบ Structure ซึ่งผู้ใช้สามารถใช้ประมวลผลและวิเคราะห์ข้อมูลได้รวดเร็วกว่า



รูปที่ ๓ : ความแตกต่างระหว่าง Data Lake กับ Data Warehouse

ที่มา : <https://www.qubole.com/data-lakes-vs-data-warehouses-the-co-existence-argument>

Big Data Analytic

Big Data Analytic คือการวิเคราะห์ข้อมูล Big Data ทำให้มีข้อมูลที่เป็นข้อเท็จจริงที่ผ่านการวิเคราะห์อย่างเป็นระบบเพื่อประกอบการตัดสินใจ โดยระดับการวิเคราะห์เป็นได้หลากหลายแล้วแต่รูปแบบการนำไปใช้งาน

ระดับการวิเคราะห์ข้อมูล

๑. Descriptive Analytics เป็นการวิเคราะห์ในระดับที่บอกว่าเกิดอะไรขึ้น เกิดเหตุการณ์สำคัญตอนไหน จำนวนเท่าไร ถีแค่ไหน ตรงไหนบ้าง

๒. Predictive Analytics เป็นการวิเคราะห์ในระดับที่ซับซ้อนอีกขั้นหนึ่ง เป็นการประเมินว่าจะเกิดเหตุการณ์อะไรขึ้นต่อไป มีการให้ข้อมูลตัวชี้วัดของผลลัพธ์ที่อาจเกิดขึ้นถ้าแนวโน้มยังเป็นเช่นนี้ต่อไป

๓. Prescriptive Analytics เป็นรูปแบบการวิเคราะห์ที่มีความซับซ้อนและยากที่สุด เพราะไม่เพียงพยากรณ์หรือทำนายว่าจะอะไรจะเกิดขึ้น แต่ยังให้คำแนะนำในทางเลือกต่าง ๆ และผลของทางเลือกต่าง ๆ ว่าจะมีผลดีและผลเสียอย่างไร โมเดลของ Prescriptive Analytics สามารถปรับเปลี่ยนได้ตามข้อมูลที่เพิ่มเติมเข้ามามากขึ้นและ Prescriptive Analytics นี้ยังเป็นการใช้ข้อมูลที่มากที่สุดและเกี่ยวกับเรื่อง Big Data อย่างมาก

รูปแบบการวิเคราะห์ แบ่งออกเป็น ๓ ประเภท

๑. Data Mining คือการวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก (Big Data) เพื่อหาความสัมพันธ์ของข้อมูล โดยการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และความน่าจะเป็นที่เกิดขึ้น เพื่อให้ได้องค์ความรู้ใหม่ ที่สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่าง ๆ เช่น ตลาดหลักทรัพย์ ธุรกิจ

๒. Text Mining เป็นเทคนิคเพื่อค้นหารูปแบบ (Pattern) จากข้อความมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อความคือกระบวนการที่กระทำกับข้อความ (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง หลักคณิตศาสตร์ หลักการประมวลเอกสาร (Document Processing) หลักการประมวลผลข้อความ (Text Processing) และหลักการประมวลผลภาษาธรรมชาติ (Natural Language Processing)

๓. Machine Learning เป็นศาสตร์ของโครงสร้างโมเดลคณิตศาสตร์ มุ่งเน้นที่จะสร้างองค์ความรู้จากข้อมูลโดยเริ่มจากการสร้างโมเดลทางคณิตศาสตร์ที่มีความยืดหยุ่น และสามารถปรับตัวเองเข้ากับข้อมูลที่ได้รับ เพื่อที่จะสามารถทำนายอนาคตได้

Descriptive Analytics เป็นการวิเคราะห์ในระดับที่บอกว่าเกิดอะไรขึ้น เพื่อบรรยายข้อมูลลักษณะข้อมูลนั้น ตัวอย่างเช่น

การเลือกตั้ง

- จัดเก็บข้อมูลของผู้มีสิทธิ์เลือกตั้ง การเปลี่ยนแปลงที่อยู่ของผู้มีสิทธิ์เลือกตั้ง ผู้สมัครรับเลือกตั้ง
- ข้อมูลเกี่ยวกับรายงานการทุจริตเลือกตั้งที่ส่งเข้ามา
- ข้อมูลความคิดเห็นของผู้มีสิทธิ์เลือกตั้งเกี่ยวกับวิธีการจัดการเลือกตั้ง สถานที่และความสะดวก
- ให้บริการตรวจสอบข้อมูลการเลือกตั้ง สถานที่จัดเลือกตั้ง ทางอินเทอร์เน็ต

e-commerce

- Cross-channel Analytics : ใช้ตรวจวัดความพึงพอใจของลูกค้าจากประสบการณ์ของลูกค้าที่มีต่อผู้ให้บริการ
- Event- Analytics : ช่วยในการตรวจวัดว่าผู้ใช้งานมีปฏิสัมพันธ์กับเนื้อหาในเว็บไซต์ของท่านอย่างไร เช่น ต้องการทราบจำนวนครั้งของปุ่มที่ถูกกดหรือมีขึ้นของข้อความหรือหัวข้อใดที่ถูกใช้งานมาก

Predictive Analytics ช่วยให้สามารถพิสูจน์ทราบเหตุการณ์ก่อนที่จะเกิดขึ้นจริง เช่น การพิสูจน์ทราบแบบแผนการขาย จากกรอบของเวลาที่ผ่านมาเพื่อการพยากรณ์ปริมาณของสินค้าที่จะต้องจัดเก็บในสต็อกสำหรับขายรอบต่อไป ตัวอย่างเช่น

งานราชการ

- การตรวจสอบการทุจริต การตรวจสอบภัยคุกคาม
- การรักษาความปลอดภัยในระบบไอที การปฏิบัติตามกฎระเบียบและการวิเคราะห์
- การปฏิบัติตามกฎหมายวิชาชีพ-การทุจริตและการละเมิด
- การพยากรณ์แนวโน้มอาชญากรรมและการป้องกัน

สุขภาพและสาธารณสุข

- ข้อมูลประกอบการรักษา เช่น ข้อมูลประวัติผู้ป่วย ประวัติรักษาพยาบาล ประวัติการแพ้ยา หรือข้อมูลประวัติครอบครัวที่ใช้ประกอบการวิเคราะห์
- ผู้ผลิตยาและเวชภัณฑ์ใช้ข้อมูลมากและหลากหลาย เพื่อหาสาเหตุการเจ็บป่วยที่แท้จริง การวิเคราะห์เพื่อจำเพาะเจาะจงกลุ่มผู้ป่วยที่จะทดลองและติดตามผลการรักษา จากยาและเวชภัณฑ์ที่พัฒนาขึ้น และเพื่อการพัฒนากลยุทธ์ด้านการตลาดของยานั้น
- การวิเคราะห์หลักและรูปแบบการแพร่เชื้อ เพื่อใช้ในงานวิจัยทางแพทย์

Digital Media

- วิเคราะห์ปริมาณ Click Stream ขนาดใหญ่
- การกำหนดเป้าหมายการโฆษณา การวิเคราะห์ พยากรณ์และแนวโน้มในเรื่องต่างๆ
- การป้องกันการใช้งานผิดวิธี และการฉ้อโกงทางเว็บไซต์
- การวิเคราะห์กราฟการใช้งานบนสังคมออนไลน์และการแบ่งส่วนแบ่งประเภทลูกค้า

การเงิน

- การปฏิบัติตามระเบียบและการรายงาน
- การวิเคราะห์และบริหารจัดการความเสี่ยง การวิเคราะห์การรักษาความปลอดภัย
- การตรวจสอบทุจริตและบริหารจัดการความเสี่ยง การวิเคราะห์แบบแผนทางการค้าขายที่ผิดปกติ

Big Data Analytics กับการบริหารภาครัฐ

องค์กรภาครัฐมีความจำเป็นที่จะต้องสร้างมูลค่าจากการวิเคราะห์ Big Data โดยมีแนวทางดังนี้

๑. รับฟังความเห็น รวบรวมข้อมูล และปรึกษากับผู้มีส่วนได้ส่วนเสีย ทุกภาคส่วนเกี่ยวกับศักยภาพ การเพิ่มประสิทธิภาพและการกำหนดนิยามทางเทคนิคที่เกี่ยวข้องรวมถึงเป้าหมายที่จะทำให้เกิดประโยชน์ ในการนำเอาข้อมูลมารวมใช้อย่างมีประสิทธิภาพในรูปแบบ Open Source ให้มากขึ้น

๒. วางแผนการลงทุนในการจัดโครงสร้าง รวบรวมข้อมูลและการบริหารจัดการข้อมูลโดยอยู่บนพื้นฐานของความเชื่อมั่นต่อผู้มีส่วนได้ส่วนเสีย ซึ่งการจัดการข้อมูลต่างๆ จะต้องเป็นไปตามกฎหมาย และหลักจริยธรรมที่เกี่ยวข้อง รวมทั้งต้องสื่อสารให้ทุกภาคส่วนรวมทั้งประชาชนให้มีความเข้าใจ และมั่นใจในเรื่องความเป็นส่วนตัวของข้อมูล (Privacy)

๓. มีความเข้าใจและมีทักษะทางธุรกิจและทักษะทางเทคนิค ความท้าทายที่เกิดขึ้นจากพลังของผลการวิเคราะห์ Big Data ในระดับชาตินั้น ผู้บริหารและเจ้าหน้าที่ในหน่วยงานภาครัฐจะต้องมีความเข้าใจ และมีทักษะทางธุรกิจและทักษะทางเทคนิคอย่างมืออาชีพโดยจะต้องทำให้ Big Data สามารถเชื่อมโยง ให้เกิดประโยชน์กับผู้ที่มีส่วนได้ส่วนเสียทุกกลุ่มโดยเฉพาะอย่างยิ่งประชาชน

๔. เตรียมความพร้อมภายใต้ความเปลี่ยนแปลงของเทคโนโลยี เพื่อตอบสนองความต้องการของผู้มีส่วนได้ส่วนเสียที่มีกฎระเบียบที่หลากหลาย จึงทำให้การออกแบบระบบ Big Data Analytics ต้องใช้สถาปัตยกรรมด้านไอทีแบบผสมผสานที่สามารถปรับปรุงขนาดได้ (Scalability) และมีความยืดหยุ่น (Flexibility) โดยองค์กรภาครัฐที่ทำหน้าที่ดูแลระบบจะต้องเตรียมความพร้อมภายใต้การเปลี่ยนแปลงของเทคโนโลยีอย่างรวดเร็ว

๕. เจ้าหน้าที่ภาครัฐจะต้องปรับ Mindset ในการเข้าร่วมกับทุกภาคส่วน ต้องปรับตัวให้องค์กรภาครัฐเป็นส่วนหนึ่งของระบบนิเวศอย่างกลมกลืน และจะต้องไม่เป็นผู้สร้างอุปสรรคให้เกิดขึ้นในการสร้างระบบ Big

Data Analytics เจ้าหน้าที่ภาครัฐจะต้องปรับ Mindset ในการเข้าร่วมกับทุกภาคส่วนในลักษณะพันธมิตร (Partnership) ให้ได้

๖. ปรับปรุงวิธีคิดและกระบวนการเพื่อทำให้เกิดการแลกเปลี่ยนข้อมูลและการใช้ข้อมูลร่วมกันระหว่างหน่วยงานภาครัฐ ภาครัฐจะต้องปรับปรุงวิธีคิดและกระบวนการเพื่อทำให้เกิดการแลกเปลี่ยนข้อมูลและการใช้ข้อมูลร่วมกันอย่างมีประสิทธิภาพ โดยไม่หวังข้อมูลว่าเป็นของหน่วยงานตัวเอง ซึ่งการแลกเปลี่ยนข้อมูลและใช้ข้อมูลร่วมกันจะเป็นกุญแจสำคัญที่นำไปสู่การมองเห็นความรู้ใหม่ๆที่เกิดจากข้อมูลในมิติใหม่และมีผลต่อการสร้างแหล่งข้อมูล Big Data ได้อย่างรวดเร็ว โดยรัฐบาลจะต้องสร้างสภาพแวดล้อมด้านเทคโนโลยีที่ปลอดภัยและสร้างกรอบการกำกับดูแลที่เหมาะสมในด้านการแลกเปลี่ยนข้อมูลและใช้ข้อมูลร่วมกัน ซึ่งอาจวางนโยบาย Open Government เพื่อเปิดเผยข้อมูลภาครัฐให้มากขึ้นเพื่อนำไปสู่การทำ Big Data ให้ประสบความสำเร็จอย่างแท้จริง

๗. การกำหนดแนวทางและการบริการให้คำปรึกษาในด้าน Big Data Analytics ให้แก่ทุกภาคส่วน เป็นอีกหนึ่งปัจจัยแห่งความสำเร็จ ซึ่งองค์กรภาครัฐจะต้องมีส่วนร่วมให้มากขึ้นกว่าที่เป็นมาในอดีต ในรูปแบบการให้บริการที่ปรึกษา โดยจะต้องสร้างแนวทางและมาตรการเพื่อให้เกิดความร่วมมือในการให้ข้อมูลเชิงลึกก็ยิ่งจะทำให้เกิดประโยชน์จาก Big Data Analytics มีความทรงพลังมากขึ้น

การนำองค์ความรู้ไปปรับใช้ในการปฏิบัติงาน

มีความเข้าใจเกี่ยวกับลักษณะข้อมูลของ Big Data มากขึ้นทำให้สามารถทำ Data Cleansing เพื่อตรวจสอบ คัดกรอง จัดระเบียบหมวดหมู่ข้อมูลต่าง ๆ ให้มีความถูกต้องเหมาะสมและถูกต้องตามลักษณะของข้อมูลนั้น ๆ เพื่อที่จะใช้วิเคราะห์ผลได้อย่างถูกต้องและรวดเร็วอย่างมีประสิทธิภาพ เช่น การจัดทำฐานข้อมูลของผู้รับบริการการตรวจสอบที่เขา ภูเขา ความลาดชัน ๓๕ % เป็นต้น

ประโยชน์ที่ได้รับ

๑. ทำให้ผู้เรียนมีความรู้เกี่ยวกับประเภทข้อมูล Big Data
๒. มีความรู้ความเข้าใจระดับและรูปแบบการวิเคราะห์ข้อมูล Big Data Analytic ทำให้สามารถปฏิบัติงานภายในองค์กร และระหว่างหน่วยงานได้อย่างรวดเร็วและมีประสิทธิภาพ
๓. สามารถประยุกต์ใช้ความรู้ที่ได้รับและปรับตัวกับการปฏิบัติงานในภาครัฐและการบริการประชาชน

แหล่งที่มา

หลักสูตร : ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากรภาครัฐทุกระดับ

บรรยายโดย : สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

หน่วยงาน : สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

รูปแบบหลักสูตร : e-Learning

ช่วงเวลาฝึกอบรม : สิงหาคม ๒๕๖๗