

สรุปความรู้ที่ได้จากการเข้าอบรม “ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากร  
ภาครัฐทุกระดับ(Big Data)”

ผ่านการพัฒนาทางไกลด้วยระบบการฝึกอบรมผ่านสื่ออิเล็กทรอนิกส์  
(TDGA e-Learning - สำนักงานพัฒนารัฐบาลดิจิทัล)

กันยายน ๒๕๖๖ - เมษายน ๒๕๖๖

โดย นายธีรพันธ์ ศิลาน้อย ตำแหน่ง นายช่างสำรวจชำนาญงาน  
สังกัด กลุ่มสำรวจเพื่อทำแผนที่ สำนักงานพัฒนาที่ดินเขต ๖

หลักสูตรความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากรภาครัฐทุกระดับ(Big Data) เพื่อเรียนรู้องค์ประกอบของ Big Data ซึ่งเป็นการเรียนรู้เกี่ยวกับความหมายของ Big Data และการนำไปใช้ให้เกิดประโยชน์ กระบวนการในการจัดเก็บ Big Data รวมถึงเรียนรู้เกี่ยวกับเครื่องมือที่ใช้ในการวิเคราะห์ Big Data และแนวทางในการวิเคราะห์ Big Data เพื่อการทำงานภาครัฐ และการเรียนรู้พื้นฐานเกี่ยวกับเครื่องมือวิเคราะห์ข้อมูล (Hadoop) เพื่อการทำงานเกี่ยวกับข้อมูลขนาดใหญ่

จากการศึกษาในหลักสูตรมีทั้งหมด จำนวน ๙ หัวข้อ ประกอบด้วย

- หัวข้อที่ ๑ ทำความรู้จัก Big data
- หัวข้อที่ ๒ จะนำ Big data ไปใช้ประโยชน์ได้อย่างไร
- หัวข้อที่ ๓ Big data มีกระบวนการจัดเก็บอย่างไร
- หัวข้อที่ ๔ Hadoop จะหาได้จากที่ใด
- หัวข้อที่ ๕ Hadoop มีกระบวนการทำงานอย่างไร
- หัวข้อที่ ๖ Hadoop มีกระบวนการทำงานอย่างไร
- หัวข้อที่ ๗ Hadoop กับการประยุกต์ใช้ในโลกธุรกิจ
- หัวข้อที่ ๘ Big data กับการสร้างอาชีพและรายได้
- หัวข้อที่ ๙ Big Data Analytics กับการบริหารภาครัฐ

โดยสรุปได้ดังนี้

ความหมายของ Big Data คือ ข้อมูลขนาดใหญ่ มีทั้งแบบโครงสร้างปกติและโครงสร้างข้อมูลที่ไม่มีรูปแบบ ซึ่งทั้งหมดเป็นข้อมูลที่ใช้ในเชิงธุรกิจ มักจะถูกใช้กับงานพวกที่ต้องวิเคราะห์ข้อมูลที่มีความซับซ้อน และไม่สามารถประเมินขนาดข้อมูลได้

รูปแบบของข้อมูล Big Data สามารถเป็นไปได้หลากหลาย ได้แก่

๑. Behavioral Data ได้แก่ ข้อมูลเชิงพฤติกรรมการใช้งานต่างๆ เช่น Server Log พฤติกรรมการคลิกดูข้อมูล หรือข้อมูลการใช้ ATM

๒. Image & Sounds ตัวอย่างเช่น ภาพถ่าย วิดีโอ รูปจาก Google Street View ภาพถ่ายทางการแพทย์ ลายมือ ข้อมูลเสียงที่ถูกบันทึกไว้

๓. Languages ตัวอย่างเช่น Text Message ข้อความที่ถูก Tweet เนื้อหาต่างๆ ในเว็บไซต์

๔. Records ตัวอย่างเช่น ข้อมูลทางการแพทย์ ข้อมูลผลสำรวจที่มีขนาดใหญ่ ข้อมูลภาษี

๕. Sensors ตัวอย่างเช่น ข้อมูลอุณหภูมิ Accelerometer ข้อมูลทางภูมิศาสตร์

Big Data ประกอบด้วยคุณลักษณะ ๔ ประการ คือ

๑. Volume ข้อมูลมีขนาดใหญ่ มีปริมาณข้อมูลมากซึ่งสามารถเป็นไปได้ทั้งข้อมูลแบบ Offline หรือ Online

๒. Variety ข้อมูลที่มีความหลากหลาย สามารถเป็นไปได้ ทั้งที่มีโครงสร้างและข้อมูลที่ไม่สามารถจับ Pattern ได้

๓. Velocity ข้อมูลมีการเปลี่ยนแปลงตลอดเวลาอย่างรวดเร็ว มีการส่งผ่านข้อมูลอย่างต่อเนื่องในลักษณะ Streaming ทำให้การวิเคราะห์ข้อมูลแบบ Manual มีข้อจำกัด

๔. Veracity ข้อมูลมีความไม่ชัดเจน (Untrusted, Uncleaned)

Data lake คือการนำเอาข้อมูลจากแหล่งข้อมูลภายนอกองค์กร ข้อมูลจากเครือข่ายข้อมูลที่กระจายไปทั่วโลกมาใช้มากขึ้น

ข้อมูลที่จัดเก็บ คือ

- ข้อมูลดิบจำนวนมากและมีขนาดใหญ่
- ข้อมูลไม่มีรูปแบบที่แน่นอน
- การเข้าถึงข้อมูลไม่สามารถเข้าถึงได้ง่ายต้องใช้ความสามารถของเจ้าหน้าที่วิเคราะห์ข้อมูล (Data Scientist)

Big Data Analytics กับการบริหารภาครัฐ ได้แก่

๑. รับฟังความคิดเห็น รวบรวมข้อมูล และปรึกษากับผู้ที่มีส่วนได้ส่วนเสีย
๒. วางแผนการลงทุนในการจัดโครงสร้าง
๓. มีความเข้าใจและมีทักษะทางธุรกิจและทักษะทางเทคนิค
๔. เตรียมความพร้อมภายใต้การเปลี่ยนแปลงของเทคโนโลยี
๕. เจ้าหน้าที่ภาครัฐจะต้องปรับ Mindset ในการเข้าร่วมกับทุกภาคส่วน
๖. ปรับปรุงวิธีคิดและกระบวนการเพื่อทำให้เกิดการแลกเปลี่ยนข้อมูลและ การใช้ข้อมูลร่วมกันระหว่างหน่วยงานภาครัฐ
๗. กำหนดแนวทางและการบริการให้คำปรึกษาในด้าน Big Data Analytics ให้แก่ทุกภาคส่วน

Hadoop หรือชื่อเรียกอย่างเป็นทางการคือ Apache Hadoop เป็น software framework สำหรับจัดการกับชุดข้อมูลขนาดใหญ่ (large data set) และการประมวลผลแบบกระจาย (distributed processing) บนเครื่องคอมพิวเตอร์ที่เชื่อมต่อกันเป็น กลุ่ม (cluster)

มีหน้าที่หลักๆ อยู่ ๒ หน้าที่ คือ

๑. บริหารจัดการหน่วยเก็บข้อมูล (storage management)
  ๒. บริหารจัดการทรัพยากรสำหรับการประมวลผล (computing resource management)
- สรุป

Hadoop เป็น software framework ที่ถูกออกแบบมาเพื่อจัดการกับ Big Data โดยเฉพาะ ดังนั้นมันจึงตอบโจทย์ของ Big data ได้เกือบทุกข้อ ระบบ Ecosystem ก็ยังมีเครื่องมือดีๆให้นักพัฒนาสามารถนำมาใช้งานหรือพัฒนาต่อยอดได้มากมาย แถมยังประหยัดค่าใช้จ่าย แต่ดูเหมือนว่าจะไม่เหมาะกับผู้ใช้งานทั่วไป การติดตั้งและใช้งาน Hadoop จำเป็นต้องมีความรู้ทางด้าน IT ค่อนข้างลึกในด้าน System Admin และ Software Development จึงจะสามารถนำ Hadoop มาใช้งานได้อย่างมีประสิทธิภาพ

โดยสรุป หน่วยงานภาครัฐทุกภาคส่วนควรเร่งศึกษาเรื่อง Big Data ทั้งเรื่องแหล่งที่มาของข้อมูล การรวบรวมข้อมูล และการวิเคราะห์ข้อมูลที่ได้รวบรวมมา เพื่อนำข้อมูลมาปรับใช้ให้เกิดประโยชน์ต่อหน่วยงานของตนเอง ให้เกิดประสิทธิภาพและประโยชน์ตรงตามวัตถุประสงค์ขององค์กรให้มากที่สุด เช่น หน่วยงานของกรมพัฒนาที่ดิน ควรมีการรวบรวมข้อมูลจากทุกหน่วยงานใน Big Data เช่น พฤติกรรมของเกษตรกร พฤติกรรมของผู้บริโภค ความต้องการของตลาดทั้งในและนอกประเทศ การใช้ประโยชน์ที่ดิน สภาพพื้นที่ ชนิดของดิน การเพาะปลูก ชนิดพืชที่ปลูก สภาพภูมิอากาศของแต่ละพื้นที่ ปริมาณน้ำฝนแต่ละปี โรคของพืช ผลผลิต ต้นทุนการผลิต รายได้จากการเพาะปลูก ภัยธรรมชาติที่เกิดในแต่ละปี ความต้องการพืชของตลาดในช่วงแต่ละเดือนในปีนั้นๆ เพื่อนำข้อมูลที่ได้จาก Big Data นำมาวิเคราะห์ และส่งต่อข้อมูลที่ได้ให้เกษตรกรนำไปพัฒนาที่ดินของตนต่อไป