สรุปการอบรมเชิงปฏิบัติการ
เรื่อง การใช้งานเครื่องมือ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร "การจัดการข้อมูลและการจัดทำแดชบอร์ด สำหรับนักวิเคราะห์ข้อมูล (Data Analytics) ด้วย Power Query และ Power BI รุ่นที่ ๑"
วันที่ ๒๑ ธันวาคม ๒๕๖๕ เวลา ๐๘.๓๐ – ๑๖.๐๐ น. ณ ห้องประชุม ๑ ชั้น ๓ อาคารนวัตกรรม
วันที่ ๒๒ – ๒๓ ธันวาคม ๒๕๖๕ เวลา ๐๘.๓๐ – ๑๖.๐๐ น. ณ ห้องประชุม ๓ ชั้น ๓ อาคารนวัตกรรม
สำนักงานเศรษฐกิจการเกษตร และผ่านโปรแกรม Zoom Meeting

....

ศูนย์ข้อมูลเกษตรแห่งชาติ (ศกช.) โดยส่วนธรรมาภิบาลข้อมูล (สธข.) ได้จัดอบรมเชิงปฏิบัติการ เรื่อง การใช้งานเครื่องมือ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร "การจัดการข้อมูลและการจัดทำแดชบอร์ด สำหรับนักวิเคราะห์ข้อมูล (Data Analytics) ด้วย Power Query และ Power BI รุ่นที่ ๑" โดยมีวัตถุประสงค์เพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในจัดการข้อมูล วิเคราะห์ข้อมูล และการจัดทำแดชบอร์ด และเพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในการจัดการข้อมูลขนาดใหญ่ (Big Data) ให้สามารถนำข้อมูลไปใช้ประกอบการตัดสินใจบริหารจัดการเชิงนโยบายและกำหนดแนวทาง/มาตรการในการแก้ไขปัญหาต่าง ๆ ได้อย่างเป็นระบบและมีประสิทธิภาพเพื่อนำไปสู่การพัฒนาภาคเกษตร โดยมีผู้เข้าร่วมอบรม ๓๐ คน รายละเอียดของการอบรมและข้อคิดเห็นต่าง ๆ สรุปได้ดังนี้

๑. ภาคบรรยาย เป็นการให้ความรู้ทางวิชาการโดยวิทยากร ผศ.ดร.วรภัทร ไพรีเกรง ผู้อำนวยการหลักสูตรสาขาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรม ด้านเทคโนโลยีและวิศวกรรมศาสตร์ (CITE) มหาวิทยาลัย ธุรกิจบัณฑิตย์ มีหัวข้อบรรยาย ดังนี้

๑.๑ ข้อมูลเบื้องต้นเกี่ยวกับ Power Query และ Power BI

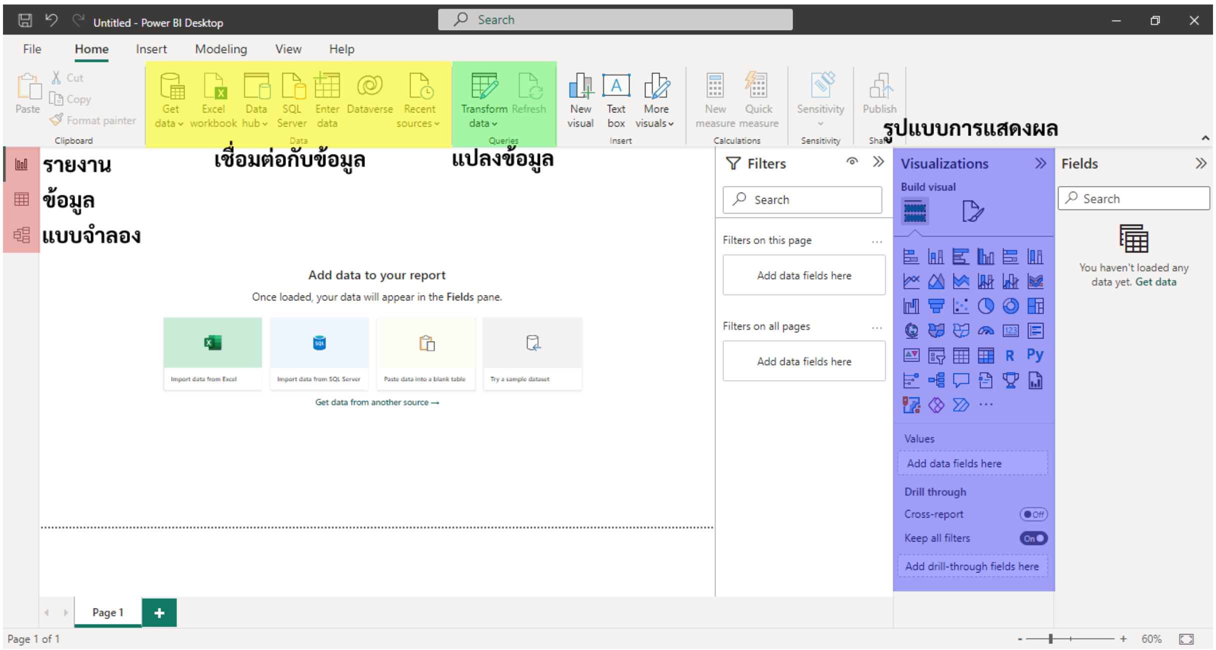
๑) Power Query คือเครื่องมือการเตรียมข้อมูลและการแปลงข้อมูล การจัดตารางหรือคิวรี การจัดคอลัมน์ การจัดการข้อมูลที่บกพร่อง การรวมข้อมูล (Merge) การแยกข้อมูล (Extract) การปรับแต่งข้อมูล (Shaping) การสร้างรูปแบบข้อมูล (Modeling) เพื่อทำให้ข้อมูลพร้อมใช้งาน

๒) ข่าวกรองธุรกิจ (BI: Business Intelligence) คือเทคโนโลยีที่มีความสามารถในการแปลงข้อมูลปกติให้เป็นข้อมูลเชิงลึกที่สามารถนำไปช่วยในการติดสินใจ วิเคราะห์ผล หรือเฝ้าติดตาม เพื่อจัดทำรายงานในรูปแบบต่างๆ

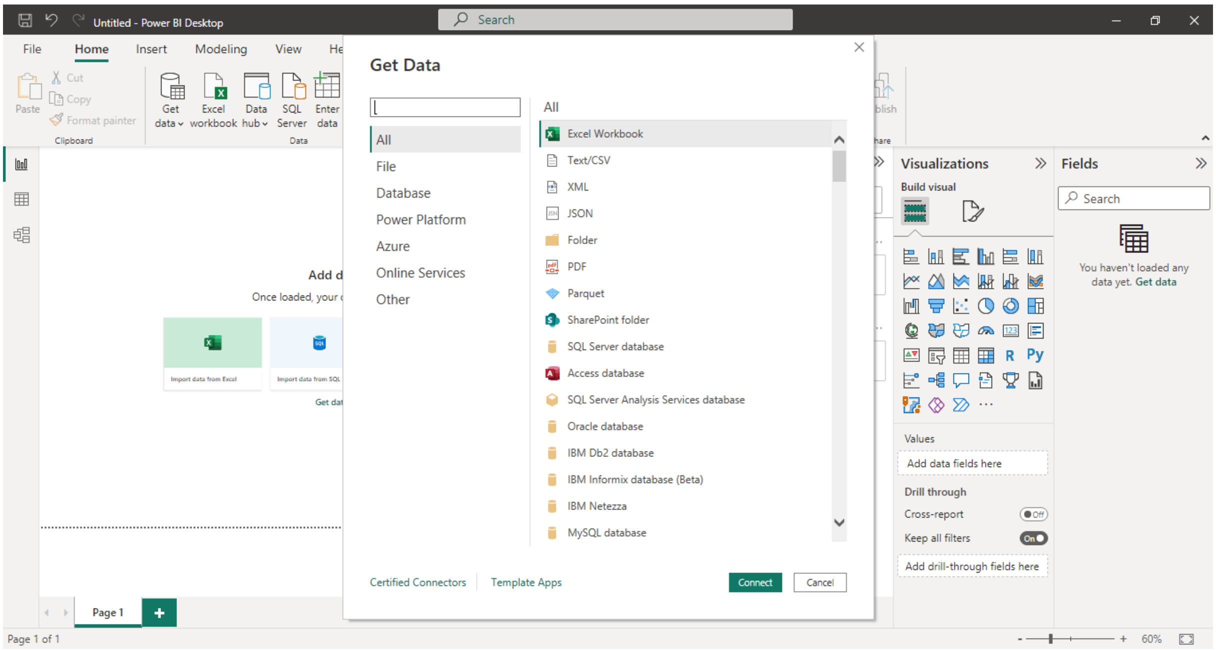
๓) Power BI คือคอลเลกชันของบริการซอฟต์แวร์ แอป และตัวเชื่อมต่อที่ทำงานร่วมกันเพื่อเปลี่ยนแหล่งข้อมูลที่ไม่เกี่ยวข้องของคุณให้เป็นข้อมูลเชิงลึกที่สอดคล้องกัน แสดงข้อมูลได้ และโต้ตอบได้ ข้อมูลของคุณอาจเป็นสเปรดชีต Excel หรือคอลเลกชันของระบบ Cloud และคลังข้อมูลแบบไฮบริดภายในองค์กร Power BI ช่วยให้คุณเชื่อมต่อกับแหล่งข้อมูลของคุณ แสดงภาพและค้นพบเรื่องสำคัญ รวมถึงแชร์สิ่งเหล่านั้นกับบุคคลหรือทุกคนที่คุณต้องการ

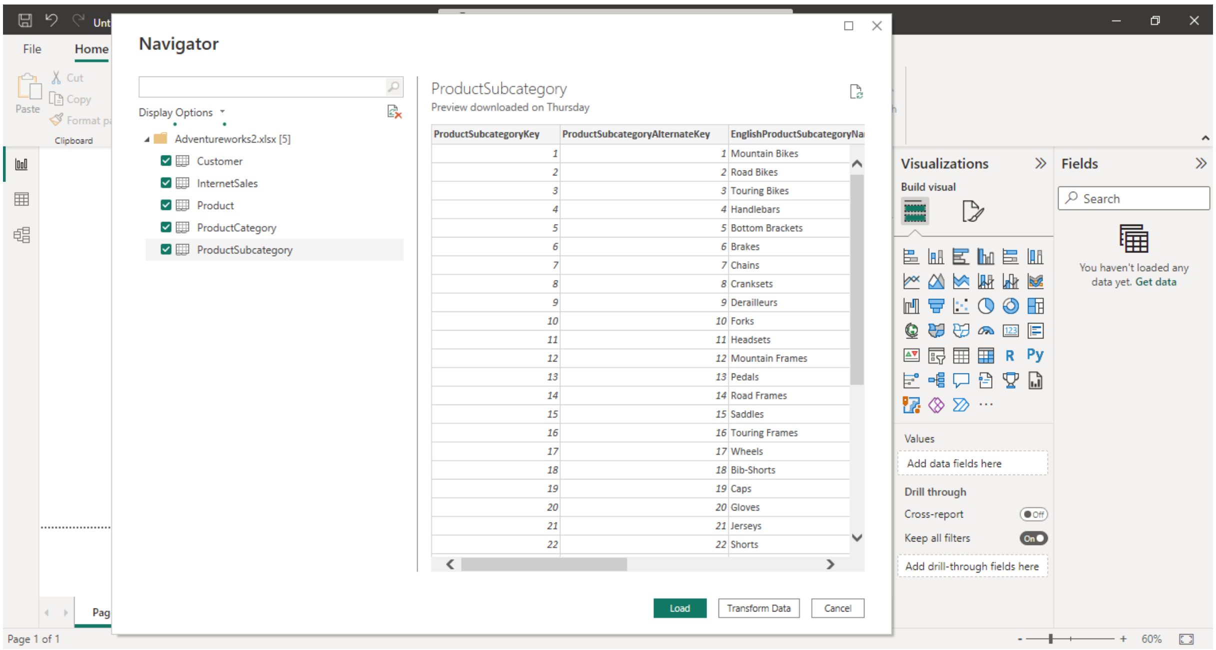
๑.๒ แนะนำโปรแกรม และเริ่มต้นการใช้งาน Power BI

๑) แนะนำโปรแกรม และเริ่มต้นการใช้งาน Power BI

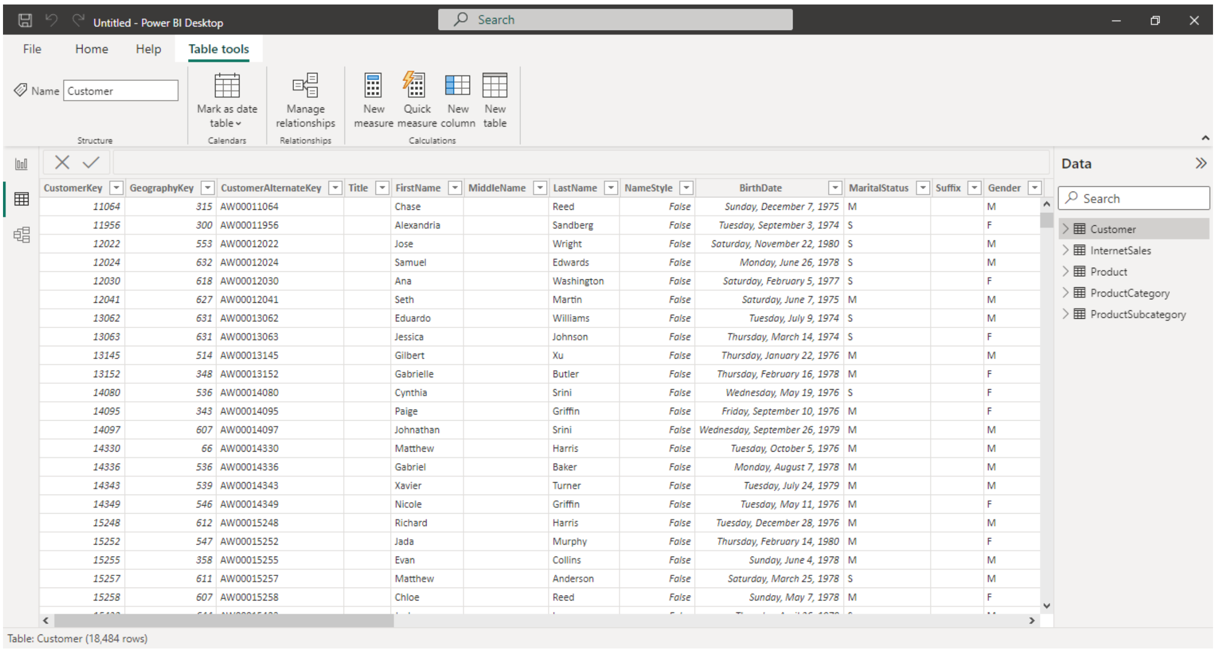


๒) การนำเข้าข้อมูล

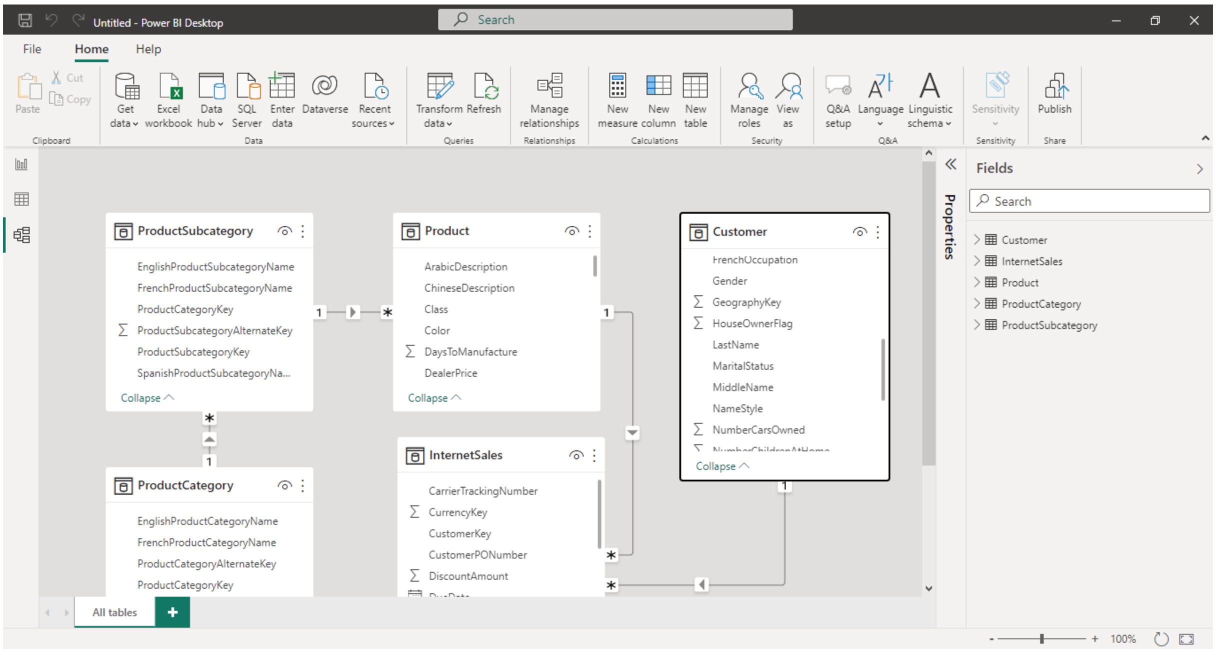




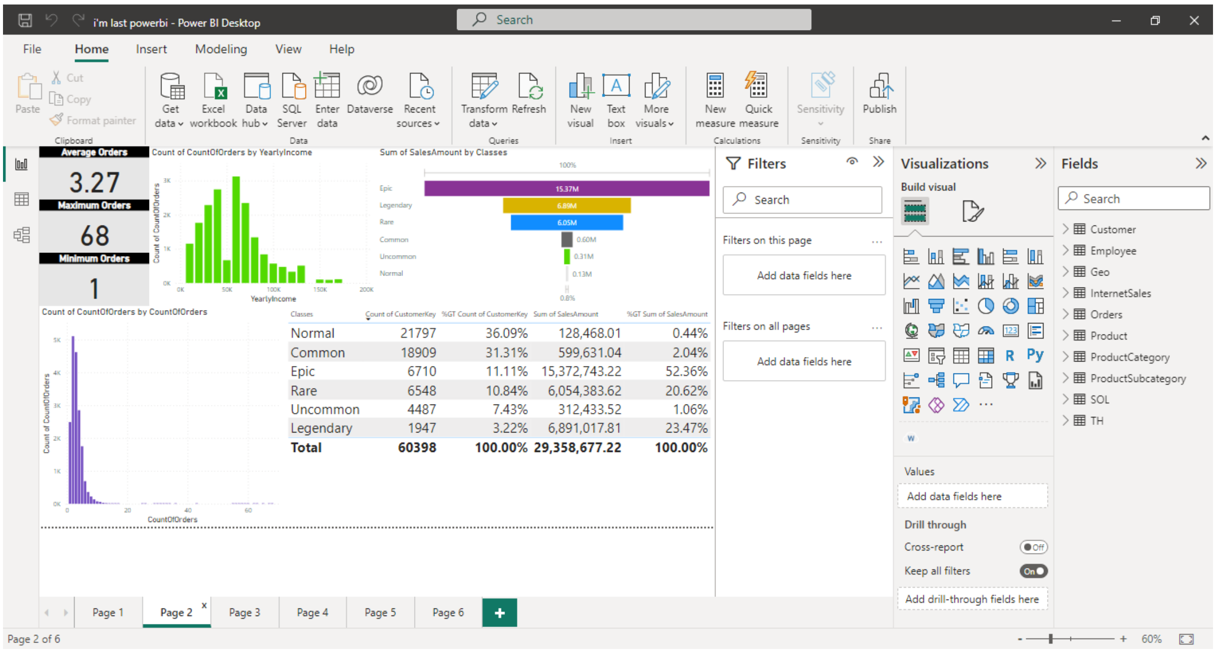
๓) การดูข้อมูล



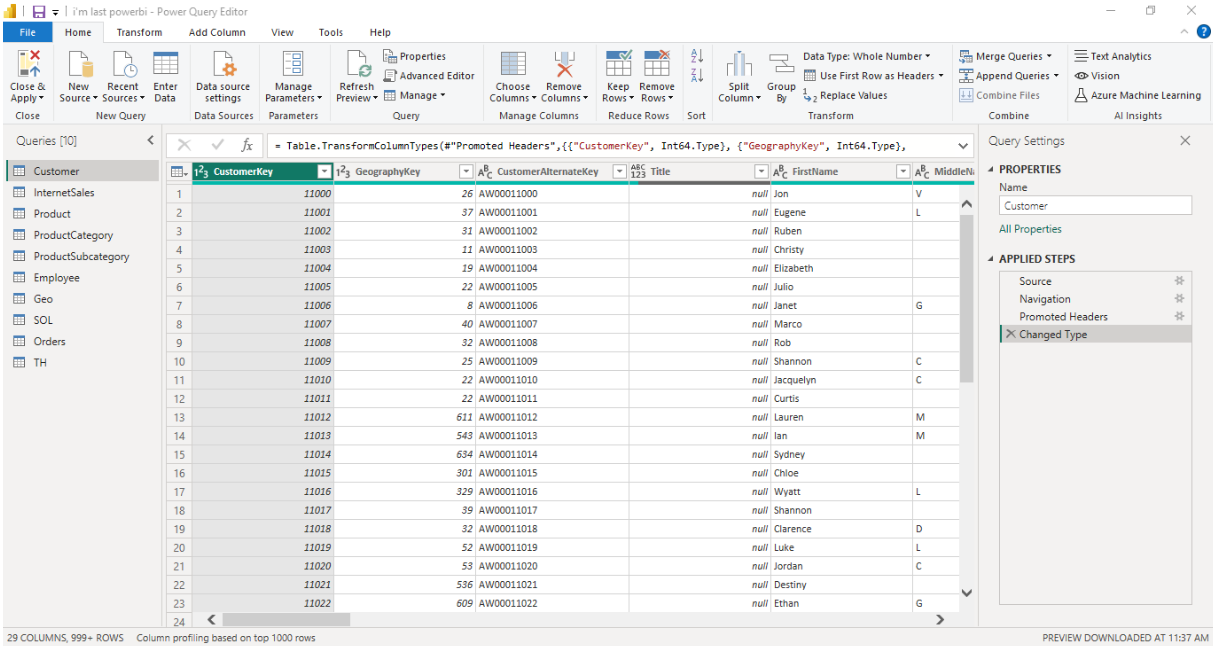
๔) ดูแบบจำลอง



๕) การสร้างรายงาน



๖) การแปลงข้อมูล (Power Query)



๒. การประเมินความพึงพอใจการเข้าร่วมอบรม

เป็นการประเมินความพึงพอใจของผู้เข้าร่วมอบรมที่มีต่อปัจจัยที่เกี่ยวข้องกับการอบรม โดยใช้แบบสอบถาม ซึ่งผลจากการประเมินจะใช้เป็นแนวทางในการพัฒนาคุณภาพการจัดอบรมให้ดียิ่งขึ้นต่อไป โดยมีเกณฑ์ในการประเมิน ดังนี้

คะแนน	การประเมิน	การแปลผล
๕	๔.๐๑ – ๕.๐๐	มากที่สุด
๔	๓.๐๑ – ๔.๐๐	มาก
๓	๒.๐๑ – ๓.๐๐	ปานกลาง
๒	๑.๐๑ – ๒.๐๐	น้อย
๑	๐.๐๐ – ๑.๐๐	น้อยที่สุด

ผู้ตอบแบบสำรวจความพึงพอใจเข้าร่วมอบรมเชิงปฏิบัติการเรื่อง การใช้งานเครื่องมือ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร “การจัดการข้อมูลและการจัดทำแดชบอร์ดสำหรับนักวิเคราะห์ข้อมูล (Data Analytics) รุ่นที่ ๑” ระหว่างวันที่ ๒๑ – ๒๓ ธันวาคม ๒๕๖๕ เวลา ๐๘.๓๐ – ๑๖.๐๐ น. ผ่านระบบออนไลน์ โดย Google Form จำนวน ๒๘ คน พบว่าผู้ใช้บริการส่วนใหญ่เป็นเพศชาย ร้อยละ ๕๐.๐๐ และเพศหญิง ร้อยละ ๕๐.๐๐ โดยมีอายุอยู่ในช่วง ๓๑ – ๔๐ ปี ร้อยละ ๔๖.๔๓ รองลงมาอายุอยู่ในช่วง ๒๐ – ๓๐ ปี ร้อยละ ๔๒.๘๖ และอายุอยู่ในช่วง ๔๑ – ๕๐ ปี ร้อยละ ๑๐.๗๑ สำหรับระดับการศึกษาส่วนใหญ่ จบการศึกษาระดับปริญญาโท ร้อยละ ๔๖.๔๓ รองลงมาระดับปริญญาตรี ร้อยละ ๔๒.๘๓ ระดับปริญญาเอก ร้อยละ ๗.๑๔ และต่ำกว่าปริญญาตรี ร้อยละ ๓.๕๗ ซึ่งมีตำแหน่งส่วนใหญ่เป็นนักวิเคราะห์นโยบายและแผนปฏิบัติการ ร้อยละ ๒๕.๐๐

และรองลงมาเป็นนักวิชาการคอมพิวเตอร์ ร้อยละ ๑๗.๘๖ ตำแหน่งเศรษฐกร ร้อยละ ๑๔.๒๙ ตำแหน่งนักวิชาการสถิติปฏิบัติการ และตำแหน่งนักวิเคราะห์นโยบายและแผน ร้อยละ ๑๐.๗๑ ตำแหน่งเจ้าพนักงานสถิติปฏิบัติงาน นักวิเคราะห์นโยบายและแผนชำนาญการ นักวิชาการตรวจสอบภายในปฏิบัติการ นักวิชาการสถิติชำนาญการ เศรษฐกรชำนาญการ และเศรษฐกรปฏิบัติการ ร้อยละ ๓.๕๗ โดยส่วนใหญ่อยู่ภายใต้สังกัดหน่วยงานสำนักงานเศรษฐกิจการเกษตรที่ ๑ – ๑๒ ร้อยละ ๓๙.๒๙ รองลงมาเป็นศูนย์ข้อมูลเกษตรแห่งชาติ ร้อยละ ๑๔.๒๙ กองนโยบายและแผนพัฒนาการเกษตร ร้อยละ ๑๐.๗๑ ศูนย์ประเมินผล สำนักวิจัยเศรษฐกิจการเกษตร และกองเศรษฐกิจการเกษตรระหว่างประเทศ ร้อยละ ๗.๑๔ สำนักงานเลขานุการกรม กลุ่มตรวจสอบภายใน กลุ่มพัฒนาระบบบริหาร และศูนย์สารสนเทศการเกษตร ร้อยละ ๓.๕๗

รายละเอียด	**ร้อยละ**
เพศ	
- ชาย	๕๐.๐๐
- หญิง	๕๐.๐๐
อายุ	
- ๒๐ – ๓๐ ปี	๔๒.๘๖
- ๓๑ – ๔๐ ปี	๔๖.๔๓
- ๔๑ – ๕๐ ปี	๑๐.๗๑
- ๕๑ – ๖๐ ปี	๐.๐๐
รายละเอียด	**ร้อยละ**
ระดับการศึกษา	
- ต่ำกว่าปริญญาตรี	๓.๕๗
- ปริญญาตรี	๔๒.๘๓
- ปริญญาโท	๔๖.๔๓
- ปริญญาเอก	๗.๑๔
ตำแหน่ง	
- นักวิเคราะห์นโยบายและแผนปฏิบัติการ	๒๕.๐๐
- นักวิชาการคอมพิวเตอร์	๑๗.๘๖
- เศรษฐกร	๑๔.๒๙
- นักวิชาการสถิติปฏิบัติการ	๑๐.๗๑
- นักวิเคราะห์นโยบายและแผน	๑๐.๗๑
- เจ้าพนักงานสถิติปฏิบัติงาน	๓.๕๗
- นักวิเคราะห์นโยบายและแผนชำนาญการ	๓.๕๗
- นักวิชาการตรวจสอบภายในปฏิบัติการ	๓.๕๗
- นักวิชาการสถิติชำนาญการ	๓.๕๗
- เศรษฐกรชำนาญการ	๓.๕๗
- เศรษฐกรปฏิบัติการ	๓.๕๗

หน่วยงาน	
- สำนักงานเศรษฐกิจการเกษตรที่ ๑ – ๑๒	๓๙.๒๙
- ศูนย์ข้อมูลเกษตรแห่งชาติ	๑๔.๒๙
- กองนโยบายและแผนพัฒนาการเกษตร	๑๐.๗๑
- ศูนย์ประเมินผล	๗.๑๔
- สำนักวิจัยเศรษฐกิจการเกษตร	๗.๑๔
- กองเศรษฐกิจการเกษตรระหว่างประเทศ	๗.๑๔
- สำนักงานเลขานุการกรม	๓.๕๗
- กลุ่มตรวจสอบภายใน	๓.๕๗
- กลุ่มพัฒนาระบบบริหาร	๓.๕๗
- ศูนย์สารสนเทศการเกษตร	๓.๕๗

๒.๑ ความพึงพอใจของผู้เข้าร่วมอบรม ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรม การจัดการข้อมูลและการจัดทำแดชบอร์ด สำหรับนักวิเคราะห์ข้อมูล (Data Analytics) ในประเด็นต่าง ๆ อาทิ เช่น ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร เทคนิคหรือวิธีการถ่ายทอดความรู้ ความเหมาะสมของเนื้อหาในการอบรม การสรุปเนื้อหาสาระได้อย่างชัดเจน ความรู้ความเข้าใจเนื้อหา ก่อน-หลัง การเข้าร่วมอบรมเชิงปฏิบัติการ สามารถนำความรู้ ความเข้าใจ และประสบการณ์ที่ได้จากการเข้าร่วมอบรมเชิงปฏิบัติการไปปรับใช้ในการปฏิบัติงานได้ ระยะเวลาการอบรมเชิงปฏิบัติการ การประสานงานและการอำนวยความสะดวก พบว่า ในภาพรวมผู้รับบริการส่วนใหญ่มีความพึงพอใจอยู่ในระดับมากที่สุด คะแนนเฉลี่ย ๔.๕๒ ร้อยละ ๙๐.๔๘ โดยมีรายละเอียดตามประเด็นต่าง ๆ ดังนี้

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๑. ภาพรวมของวิทยากร						**๔.๖๔**
๑.๑ ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร	๖๔.๒๙ (๑๘)	๓๕.๗๑ (๑๐)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๔
๑.๒ เทคนิคหรือวิธีการถ่ายทอดความรู้	๖๔.๒๙ (๑๘)	๓๕.๗๑ (๑๐)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๔
๑.๓ ความเหมาะสมของเนื้อหาในการอบรม	๖๐.๗๑ (๑๗)	๓๙.๒๙ (๑๑)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๑
๑.๔ การสรุปเนื้อหาสาระได้อย่างชัดเจน	๖๗.๘๖ (๑๙)	๓๒.๑๔ (๙)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๘

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๒. ความรู้ความเข้าใจเกี่ยวกับเนื้อหาวิชาที่เข้าอบรมเชิงปฏิบัติการ						**๔.๓๑**
๒.๑ ก่อน การเข้าร่วมอบรมเชิงปฏิบัติการท่านมีความรู้ความเข้าใจเพียงใด	๑๗.๘๖ (๕)	๗.๑๔ (๒)	๑๔.๒๙ (๔)	๓๙.๒๙ (๑๑)	๒๑.๔๓ (๖)	๒.๖๑
๒.๒ หลัง การเข้าร่วมอบรมเชิงปฏิบัติการท่านมีความรู้ความเข้าใจเพียงใด	๓๕.๗๑ (๑๐)	๖๐.๗๑ (๑๗)	๓.๕๗ (๑)	๐.๐๐	๐.๐๐	๔.๓๒
๒.๓ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมเชิงปฏิบัติการไปปรับใช้ ในการปฏิบัติงานได้เพียงใด	๓๕.๗๑ (๑๐)	๖๐.๗๑ (๑๗)	๓.๕๗ (๑)	๐.๐๐	๐.๐๐	๔.๓๒
๒.๔ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมเชิงปฏิบัติการไปถ่ายทอด ให้เพื่อนร่วมงานได้เพียงใด	๓๙.๒๙ (๑๑)	๕๐.๐๐ (๑๔)	๑๐.๗๑ (๓)	๐.๐๐	๐.๐๐	๔.๒๙
๓. ภาพรวมการจัดการอบรมเชิงปฏิบัติการ						๔.๖๒
๓.๑ ระยะเวลาการอบรมเชิงปฏิบัติการ	๔๖.๔๓ (๑๓)	๕๐.๐๐ (๑๔)	๓.๕๗ (๑)	๐.๐๐	๐.๐๐	๔.๔๓
๓.๒ การประสานงานและการอำนวยความสะดวก	๕๓.๕๗ (๑๕)	๔๖.๔๓ (๑๓)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๕๔
๓.๓ ความพึงพอใจในภาพรวมของการอบรมเชิงปฏิบัติการ	๘๙.๒๙ (๒๕)	๑๐.๗๑ (๓)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๘๙
รวม						**๔.๕๒**

๒.๒ ความพึงพอใจที่มีต่อวิทยากร ผู้เข้าร่วมอบรมมีความพึงพอใจในภาพรวมของวิทยากร

	ร้อยละความพึงพอใจ				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
ด้านความรอบรู้ในเนื้อหาวิชาของวิทยากร	**๗๖.๑**	๒๑.๗	-	๒.๒	-
ด้านเทคนิคหรือวิธีการถ่ายทอดความรู้	**๕๒.๒**	๓๗	๘.๗	๒.๒	-
ด้านความเหมาะสมของเนื้อหาในการอบรม	**๔๗.๘**	๓๙.๑	๑๐.๙	๒.๒	-
ด้านการสรุปเนื้อหาสาระได้อย่างชัดเจน	**๕๒.๒**	๔๑.๓	๖.๕	-	-

๒.๓ ด้านความรู้ ความเข้าใจเกี่ยวกับเนื้อหาในการอบรม

	ร้อยละความพึงพอใจ				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
ความรู้ ความเข้าใจของผู้เข้าร่วมก่อนการอบรม	๒.๒	๘.๗	๒๖.๑	**๓๔.๘**	๒๘.๓
ความรู้ ความเข้าใจของผู้เข้าร่วมหลังการอบรม	๓๒.๖	**๔๓.๕**	๒๑.๗	๒.๒	-
ท่านสามารถนำความรู้ ความเข้าใจ และประสบการณ์ที่ได้จากการเข้าร่วมการอบรมไปปรับใช้ในการปฏิบัติงาน	๑๙.๖	**๖๐.๙**	๑๕.๒	๔.๓	-
ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมการอบรมไปถ่ายทอดให้เพื่อนร่วมงาน	๑๗.๔	**๖๐.๙**	๑๗.๔	๒.๒	๒.๒

๒.๔ ความพึงพอใจที่มีต่อการจัดอบรม

	ร้อยละความพึงพอใจ				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
ระยะเวลาการอบรม	๓๐.๔	**๔๗.๘**	๑๗.๔	๒.๒	๒.๒
การประสานงานและการอำนวยความสะดวก	**๕๒.๒**	๓๗	๑๐.๙	-	-
ความพึงพอใจในภาพรวมของการอบรม	๔๓.๕	**๕๔.๓**	๒.๒	-	-

๒.๕ จุดเด่น ของการเข้าร่วมอบรมที่ประทับใจ

๑) วิทยากรเอาใจใส่ในการให้ความรู้
๒) เนื้อหาเข้าใจง่าย วิทยากรเข้าใจผู้อบรม
๓) ได้ลองปฏิบัติจริงทำให้เห็นภาพรวมการทำงานของ Power BI ชัดขึ้น
๔) ทำให้มีความรู้และความเข้าใจในการใช้เครื่องมือ Power BI และ Power Query สำหรับจัดการชุดข้อมูลทำให้สามารถนำเสนอในรูปแบบที่หลากหลายมากขึ้น
๕) ได้ลงมือปฏิบัติจริงและได้เรียนรู้พร้อมกับการแก้ปัญหา
๖) สามารถนำไปใช้งานได้จริง พัฒนาการทำงานได้ดียิ่งขึ้น
๗) ได้นำไปประยุกต์ต่อยอดใช้ในงานจัดการชุดข้อมูลที่มีขนาดใหญ่
๘) ได้รับความรู้เกี่ยวกับการเชื่องโยงและนำเสนอข้อมูล
๙) ได้เรียนรู้ทักษะและสิ่งใหม่ ๆ
๑๐) เป็นเนื้อที่ดี เป็นประโยชน์ต่อการใช้งานจริง

๒.๖ จุดด้อย ที่ควรปรับปรุง

๑) บางช่วงตามไม่ทันอาจารย์
๒) อยากให้เพิ่มเติมเนื้อหาและจำนวนวันในการอบรมเพิ่มขึ้น
๓) วิทยากรสอนค่อนข้างเร็ว บางเนื้อหายังไม่มีเวลาเจาะลึกหรือฝึกทำจนคล่อง
๔) อาหารว่างควรมีเครื่องดื่มร้อน ชา กาแฟ โกโก้
๕) ควรเพิ่มระยะเวลาอีก ๑ วัน สำหรับการ Workshop นำข้อมูลสำนักงานมาฝึกปฏิบัติจริง
๖) การอำนวยความสะดวกให้กับผู้เข้าอบรม อย่างเช่นการสร้างแอคเค้าไว้รองรับผู้เข้าประชุม
๗) อยากให้มีคอมพิวเตอร์ เพื่ออำนวยความสะดวกในการอบรม
๘) การจัดทำ Workshop ควรทำเป็นกลุ่มย่อย และมีเจ้าหน้าที่คอยให้คำแนะนำประจำกลุ่มย่อย
๙) ควรแจ้งให้ทราบชัดเจนว่าผู้จัดจะรับผิดชอบค่าใช้จ่าย หรือให้เบิกต้นสังกัดมีรายการไหนบ้าง
๑๐) มักมีเสียงแทรกอาจารย์จากผู้เรียนผ่าน Zoom ทำให้รบกวนผู้เรียน

๒.๗ ข้อเสนอแนะเพิ่มเติม

๑) อยากให้มีการจัดอบรมเชิงปฏิบัติการในเรื่องอื่น ๆ ที่เป็นประโยชน์อีก
๒) อยากให้เพิ่มระยะเวลาในการฝึกอบรม
๓) อยากให้เอาข้อมูลจริงของหน่วยงานมาฝึกปฏิบัติ
๔) อยากให้เพิ่มเติมในเรื่องของการจัดการ/เตรียมชุดข้อมูลก่อนนำมาใช้เครื่องมือ Power BI และ Power Query
๕) อบรมค่อนข้างเร็วสำหรับคนที่ไม่มีพื้นฐานการใช้ Power BI ควรมีผู้ช่วยในการให้ความช่วยเหลือผู้เข้ารับการอบรมที่ตามไม่ทัน นอกจากนี้ ควรมีการยกกตัวอย่างที่เป็นด้านเกษตรมากกว่านี้ เพื่อให้เห็นภาพชัดเจนว่าจะสามารถเอาไปประยุกต์ใช้งานที่เกี่ยวข้องด้านใดบ้าง ด้านเอกสารการอบรมควรละเอียดมากกว่านี้
๖) อยากให้อาจารย์เน้นเรื่อง Power Query มากกว่า Power BI เพราะคิดว่า Power BI เมื่อเรียนพื้นฐานแล้วก็สามารถต่อยอดเองได้ง่ายกว่า

สรุปการอบรมเชิงปฏิบัติการ

เรื่อง การใช้งานเครื่องมือ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร “การจัดการข้อมูลและการจัดทำแดชบอร์ด สำหรับนักวิเคราะห์ข้อมูล (Data Analytics) ด้วย Power Query และ Power BI รุ่นที่ ๒”

วันที่ ๑๙ – ๒๑ เมษายน ๒๕๖๖ เวลา ๐๘.๓๐ – ๑๖.๐๐ น. ณ ห้องประชุม ๓ ชั้น ๓ อาคารนวัตกรรม สำนักงานเศรษฐกิจการเกษตร และผ่านโปรแกรม Zoom Meeting

....

ศูนย์ข้อมูลเกษตรแห่งชาติ (ศกช.) โดยส่วนธรรมาภิบาลข้อมูล (สธข.) ได้จัดอบรมเชิงปฏิบัติการ เรื่อง การใช้งานเครื่องมือ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร “การจัดการข้อมูลและการจัดทำแดชบอร์ด สำหรับนักวิเคราะห์ข้อมูล (Data Analytics) ด้วย Power Query และ Power BI รุ่นที่ ๒” โดยมีวัตถุประสงค์เพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในจัดการข้อมูล วิเคราะห์ข้อมูล และการจัดทำแดชบอร์ด และเพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในการจัดการข้อมูลขนาดใหญ่ (Big Data) ให้สามารถนำข้อมูลไปใช้ประกอบการตัดสินใจบริหารจัดการเชิงนโยบายและกำหนดแนวทาง/มาตรการในการแก้ไขปัญหาต่าง ๆ ได้อย่างเป็นระบบและมีประสิทธิภาพเพื่อนำไปสู่การพัฒนาภาคเกษตร โดยมีผู้เข้าร่วมอบรม ๓๐ คน รายละเอียดของการอบรมและข้อคิดเห็นต่าง ๆ สรุปได้ดังนี้

๑. ภาคบรรยาย เป็นการให้ความรู้ทางวิชาการโดยวิทยากร ผศ.ดร.วรภัทร ไพรีเกรง ผู้อำนวยการหลักสูตรสาขาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรม ด้านเทคโนโลยีและวิศวกรรมศาสตร์ (CITE) มหาวิทยาลัย ธุรกิจบัณฑิตย์ มีหัวข้อบรรยาย ดังนี้

๑.๑ ข้อมูลเบื้องต้นเกี่ยวกับ Power Query และ Power BI

๑) Power Query คือ เครื่องมือการเตรียมข้อมูลและการแปลงข้อมูล การจัดตาราง การจัดคอลัมน์ การจัดการข้อมูลที่บกพร่อง การรวมข้อมูล (Merge) การแยกข้อมูล (Extract) การปรับแต่งข้อมูล (Shaping) การสร้างรูปแบบข้อมูล (Modeling) เพื่อทำให้ข้อมูลพร้อมใช้งาน

๒) ข่าวกรองธุรกิจ (BI: Business Intelligence) คือ เทคโนโลยีที่มีความสามารถในการแปลงข้อมูลปกติให้เป็นข้อมูลเชิงลึกที่สามารถนำไปช่วยในการติดสินใจ วิเคราะห์ผล หรือเฝ้าติดตามเพื่อจัดทำรายงานในรูปแบบต่างๆ

๓) Power BI คือ เครื่องมือในการวิเคราะห์ข้อมูลธุรกิจ (Business Analytics Tool) และสร้างรายงาน สร้าง Dashboard เพื่อให้ผู้ใช้งานนำไปประกอบการตัดสินใจ แบบรวมศูนย์ สามารถอัพเดตได้อย่างทันที Power BI สามารถเชื่อมต่อแหล่งข้อมูล (Data Source) ที่เป็นที่นิยมมากมาย เช่น Excel, Database, Website ด้วยหลักการของ Design Once View Anywhere

๑.๒ แนะนำโปรแกรม และเริ่มต้นการใช้งาน Power BI

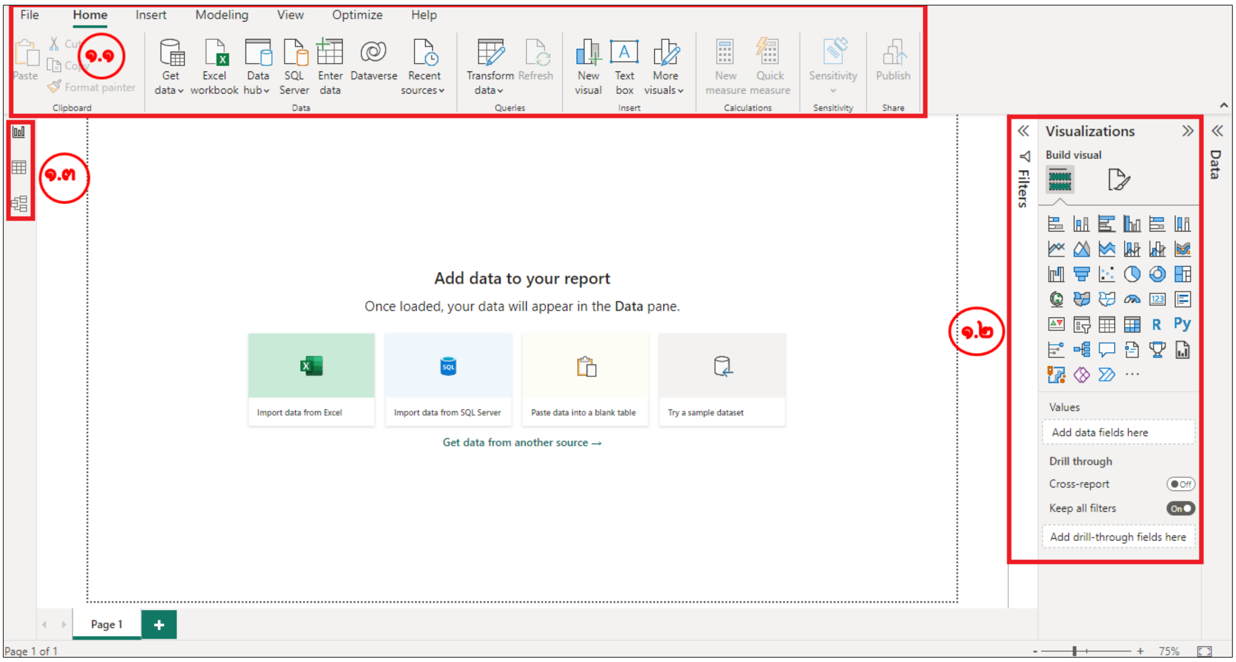
๑) แนะนำโปรแกรม และเริ่มต้นการใช้งาน Power BI โดยโปรแกรม จะมีส่วนประกอบหลักของโปรแกรม ตามภาพที่ ๑ ดังนี้

๑.๑) แถบเมนู เครื่องมือ คำสั่งต่างๆ ที่ใช้ในการจัดการข้อมูล

๑.๒) แถบเมนู ชุดเครื่องมือในการสร้างรายงาน Data Visualization รูปแบบต่างๆ จากข้อมูลที่นำเข้า เช่น Table, Chart, Bar Chart, Pie Chart, Line, Map เป็นต้น

๑.๓) แถบเมนู ตัวเลือกมุมมองของข้อมูล ประกอบด้วย ๓ มุมมอง คือ

- **Report view :** เป็นมุมมองที่จะใช้ในการแสดงผลรายงาน ในรูปแบบต่างๆ เช่น ตัวเลข แผนภูมิ กราฟ วงกลม กรวย มาตรวัด และแผนที่ เป็นต้น
- **Datasheet View** : เป็นมุมมองที่จะใช้ในการแสดงข้อมูลที่ได้ Get Data เข้ามา โดยจะแสดงผลใน รูปแบบตาราง (Table) ซึ่งจะเป็นรูปแบบที่อ่านได้อย่างเดียว
- **Relationships View** : เป็นมุมมองที่จะใช้ในการแสดงความสัมพันธ์ของข้อมูล จากแหล่งข้อมูลต่างๆ เมื่อมีการ Get Data มาเรียบร้อยแล้ว สามารถที่จะสร้างความเชื่อมโยงความสัมพันธ์ข้อมูลระหว่างตารางได้



ภาพที่ ๑

๒) การนำเข้าข้อมูล

การนำเข้าข้อมูล สามารถนำเข้าจากไฟล์ข้อมูลประเภทต่างๆได้ เช่น Microsoft Excel, CSV , XML , Text หรือฐานข้อมูลต่างๆ ซึ่งรวมไปถึงข้อมูลที่อยู่บน Cloud และ web site ได้เช่นเดียวกัน โดยวิธีการนำเข้าข้อมูล ตามภาพที่ ๒ - ๓ มีขั้นตอนดังนี้

๑) คลิก เมนู “Home”

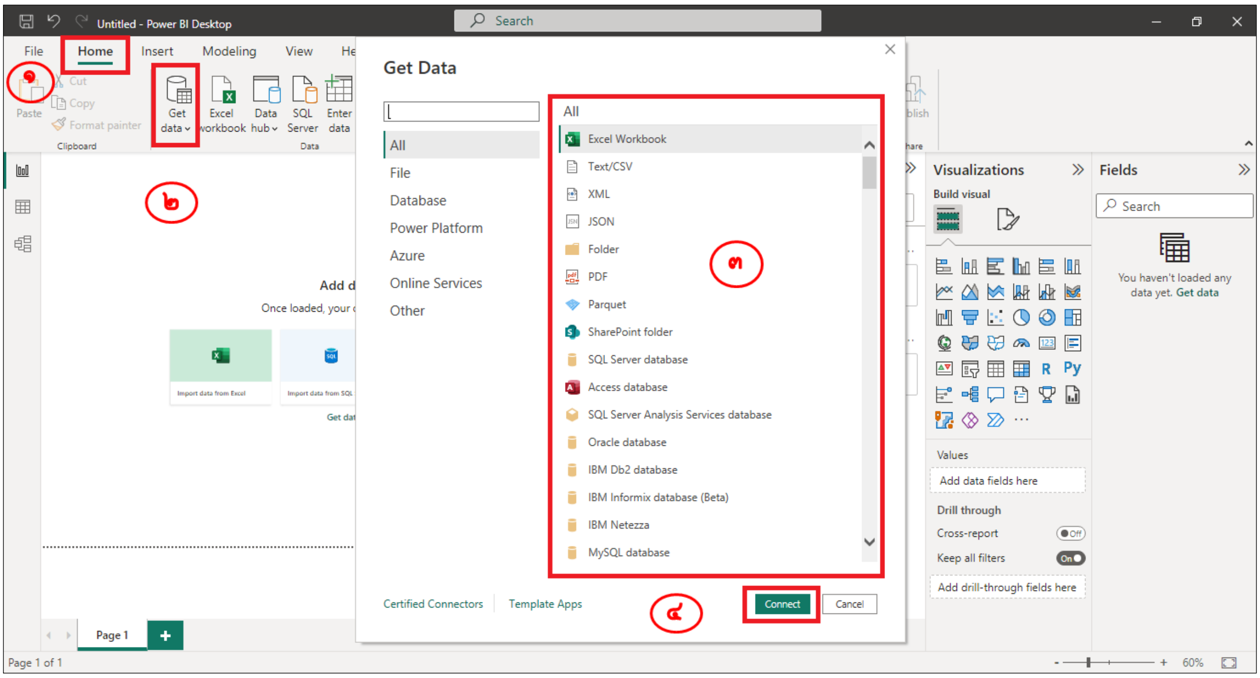
๒) คลิก เมนู “Get data”

๓) จากนั้นจะปรากฏหน้าต่าง “Get Data” เลือกประเภทข้อมูล Excel

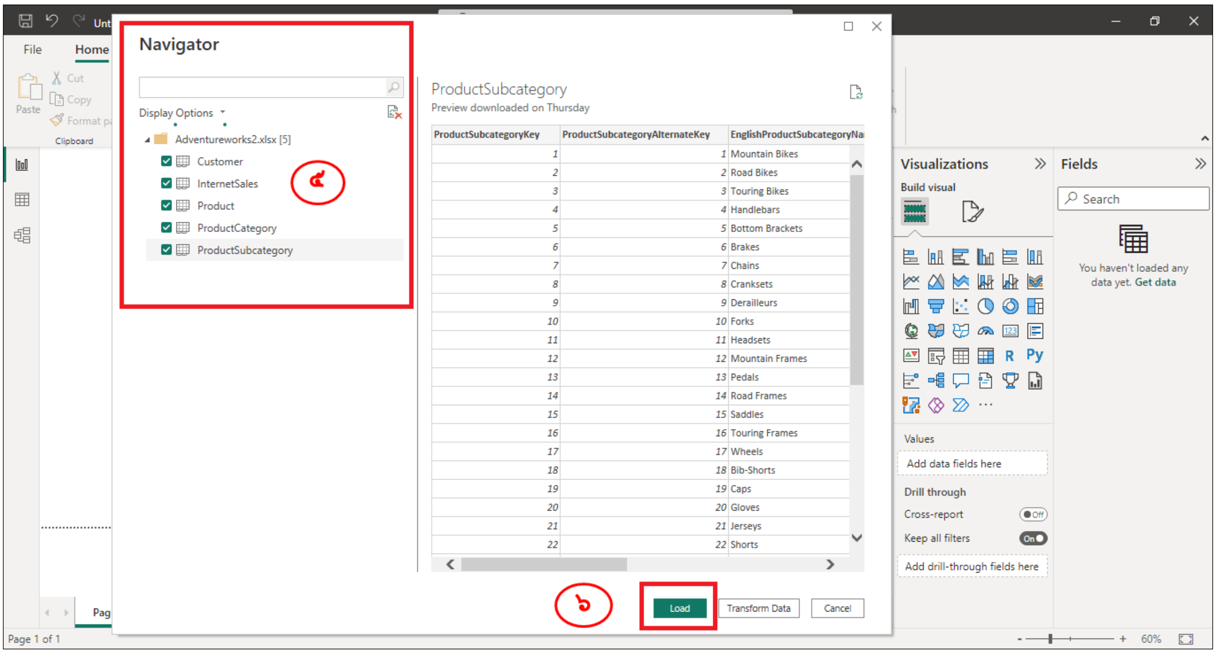
๔) คลิก Connect

๕) เลือกชุดข้อมูลภายในไฟล์ excel

๖) คลิก Load



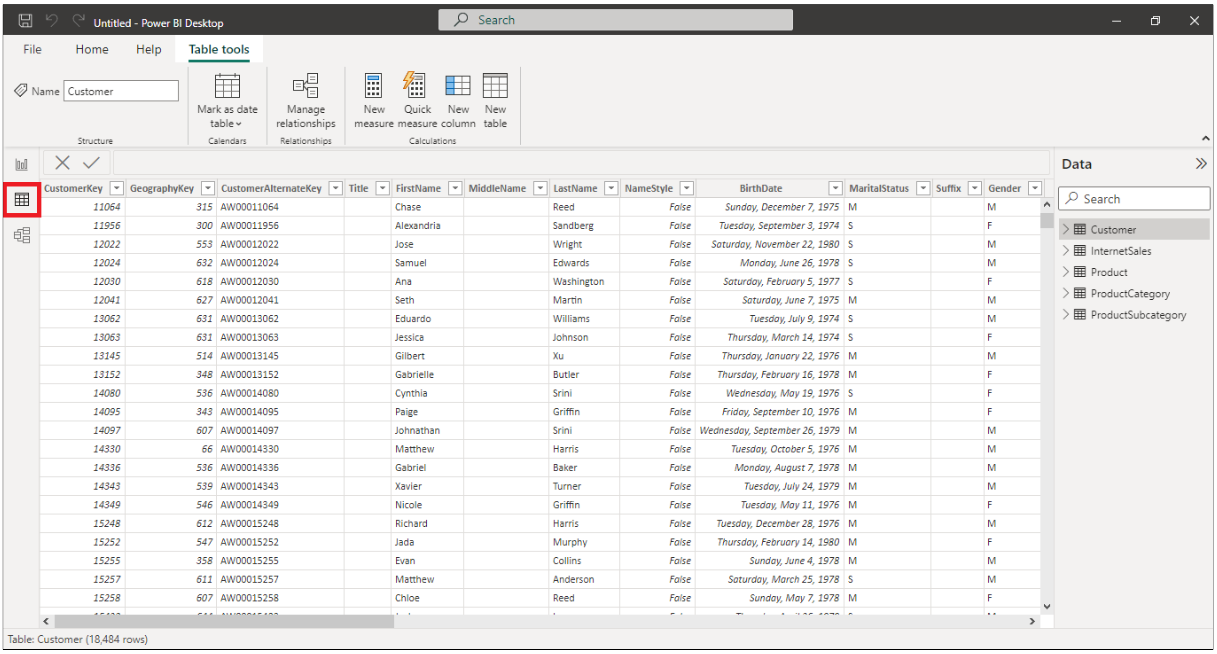
ภาพที่ ๒



ภาพที่ ๓

๓) การดูข้อมูล (Datasheet View)

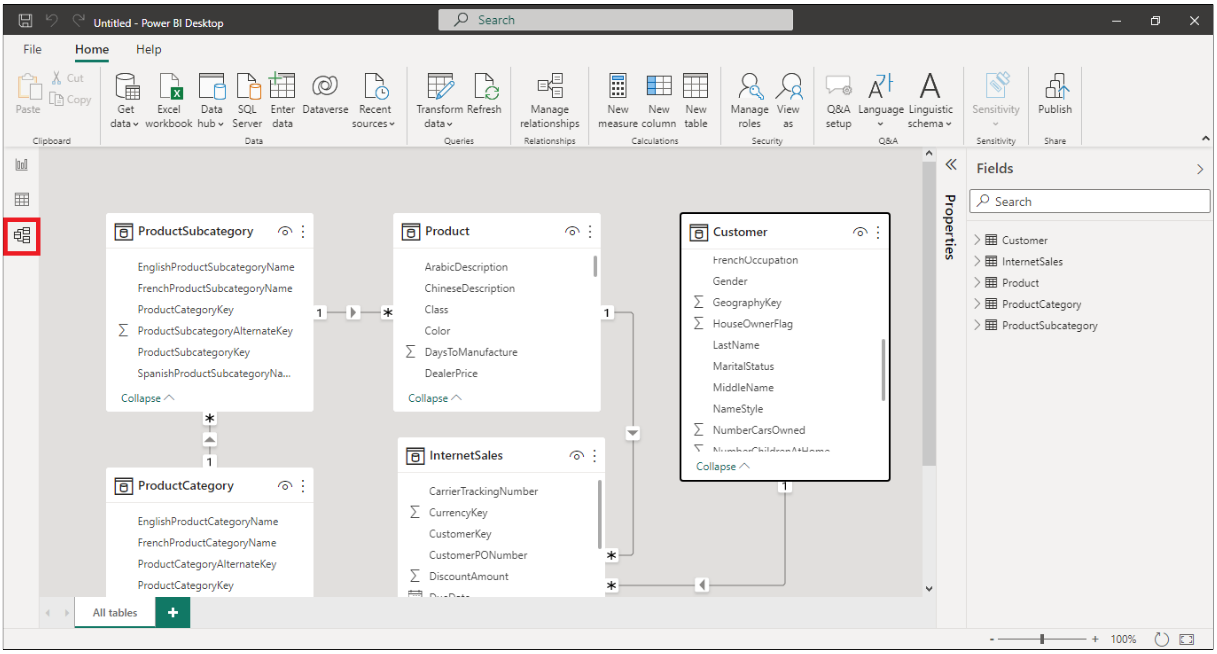
มุมมองที่ใช้ในการแสดงข้อมูลที่ได้ Get Data เข้ามา โดยจะแสดงผลในรูปแบบตาราง (Table) ซึ่งจะเป็นรูปแบบที่อ่านได้อย่างเดียว ตามภาพที่ ๔ ดังนี้



ภาพที่ ๔

๔) การแสดงความสัมพันธ์ของข้อมูล (Relationships View)

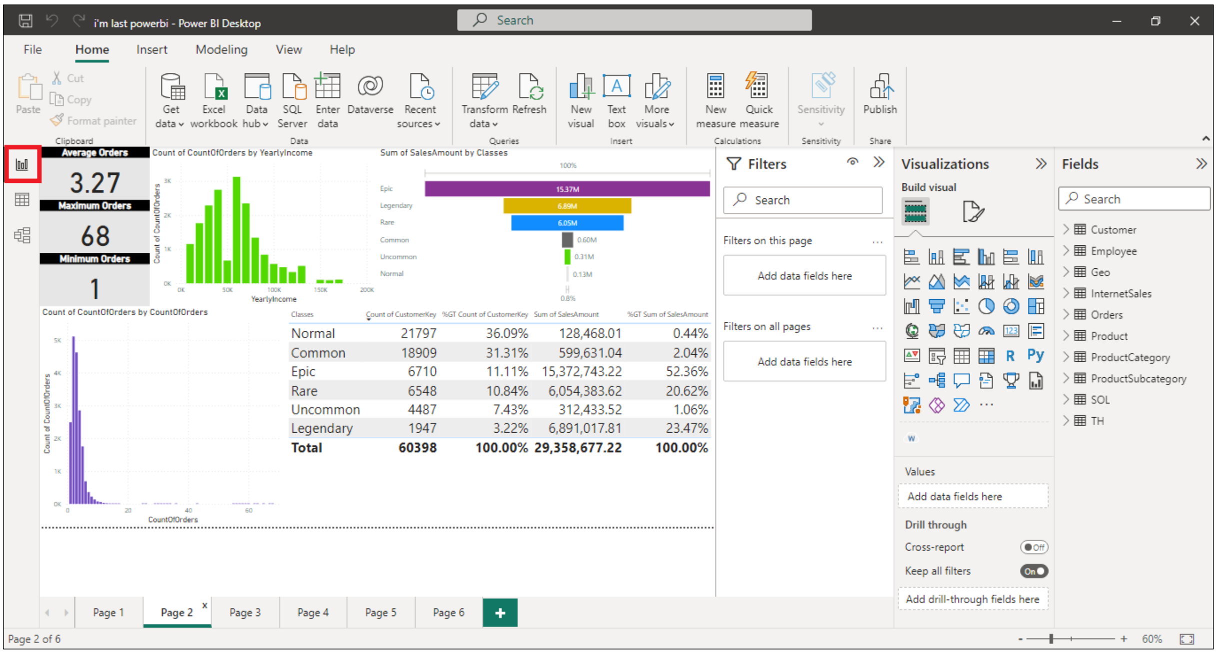
มุมมองที่ใช้ในการแสดงความสัมพันธ์ของข้อมูลจากแหล่งข้อมูลต่างๆ เมื่อ Get Data มาเรียบร้อยแล้ว สามารถที่จะสร้างกำหนดความเชื่อมโยงความสัมพันธ์ของข้อมูลระหว่างตารางได้ ดังภาพที่ ๕



ภาพที่ ๕

๕) การสร้างรายงาน (Report view)

มุมมองที่ใช้ในการแสดงผลรายงาน ในรูปแบบต่างๆ Table, Bar Chart, Pie Chart, Slicer, Donut Chart, Matrix, Map, Card เป็นต้น โดยสามารถเลือกใช้รูปแบบในการแสดงผลรายงานตามความเหมาะสมของข้อมูลได้อย่างหลากหลาย ดังภาพที่ ๖



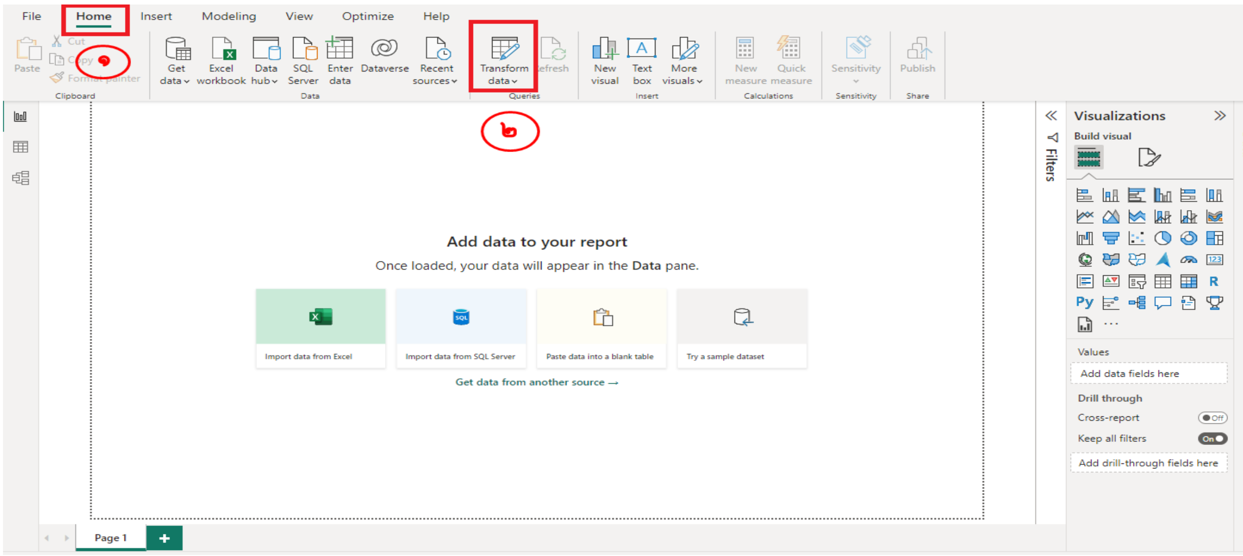
ภาพที่ ๖

๖) การแปลงข้อมูล (Power Query)

Power Query เป็นเครื่องมือสำหรับการเตรียมข้อมูลและการแปลงข้อมูล การจัดตาราง การจัดคอลัมน์ การจัดการข้อมูลที่บกพร่อง การรวมข้อมูล (Merge) การแยกข้อมูล (Extract) การปรับแต่งข้อมูล (Shaping) การสร้างรูปแบบข้อมูล (Modeling) เพื่อทำให้ข้อมูลพร้อมใช้งาน การเข้าใช้งานเมนู Power Query มีขั้นตอนตามภาพที่ ๗ ดังนี้

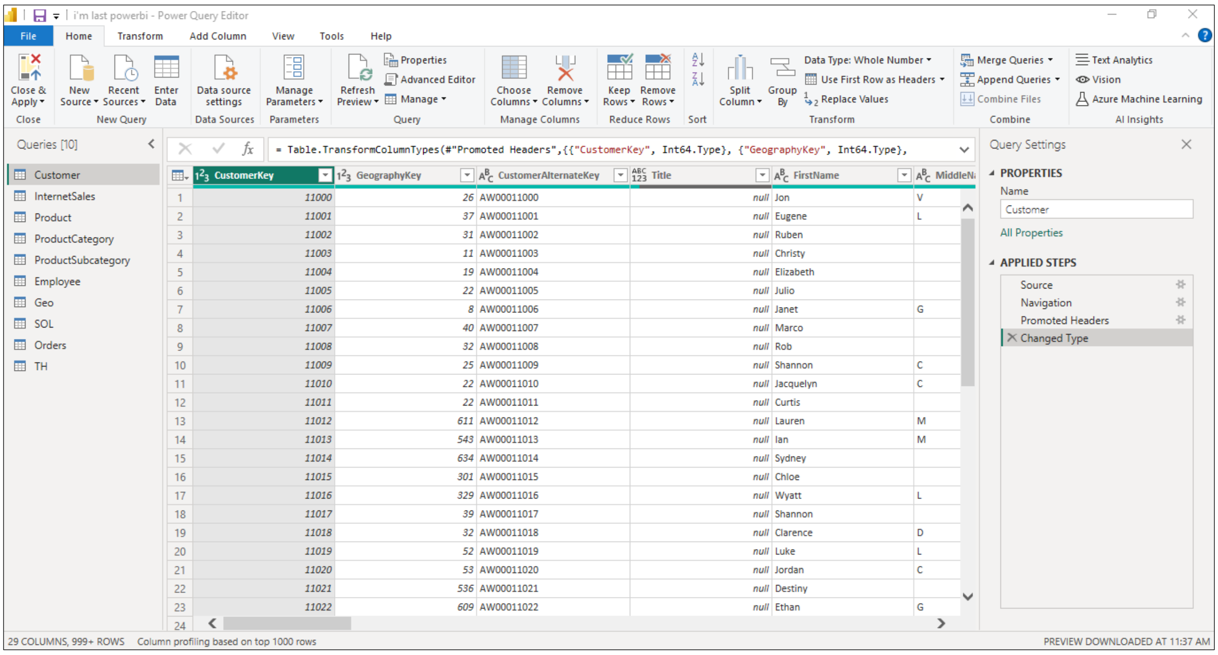
๑) คลิก เมนู "Home"

๒) คลิก เมนู "Transform data"



ภาพที่ ๗

หลังจากคลิก เมนู "Transform data" แล้ว จะแสดงหน้ารายละเอียดของข้อมูลแบบตาราง ดังภาพที่ ๘



ภาพที่ ๘

๗) ตัวอย่าง การนำเข้าข้อมูลผ่าน Web โดยวิธีการนำเข้าข้อมูลผ่าน Web มีขั้นตอนตามภาพที่ ๙ - ๑๓ ดังนี้

๑) คลิก เมนู “Home”

๒) คลิก เมนู “Get data”

๓) จากนั้นจะปรากฏหน้าต่าง “Get Data” เลือกประเภทข้อมูล Web

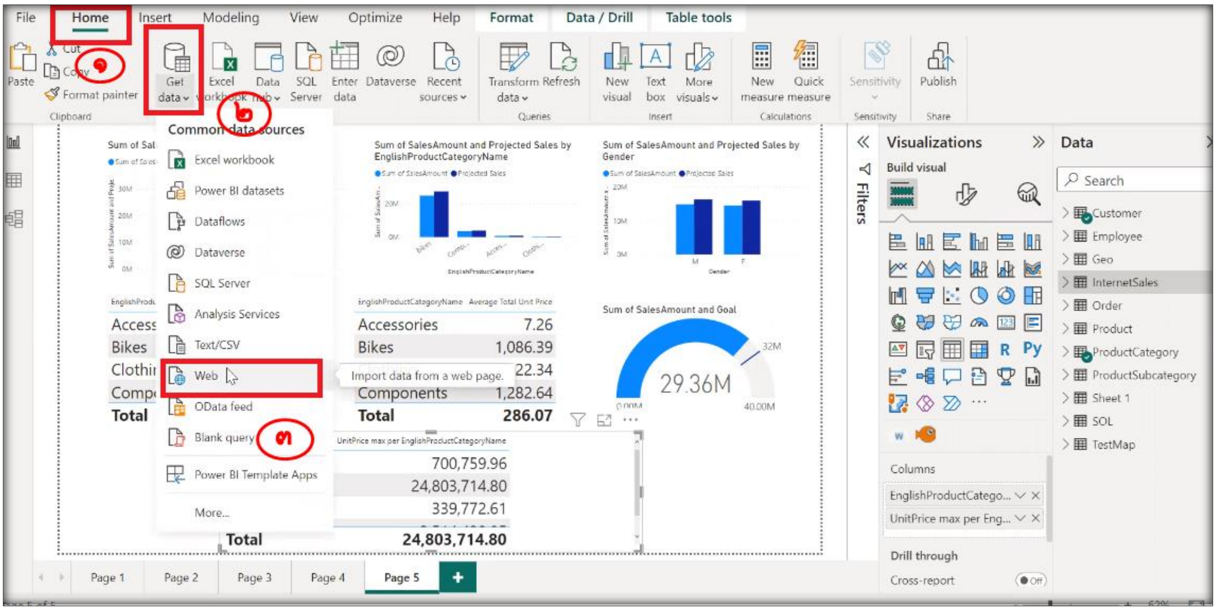
๔) ระบุ URL Web ที่ต้องการข้อมูล

๕) คลิก OK

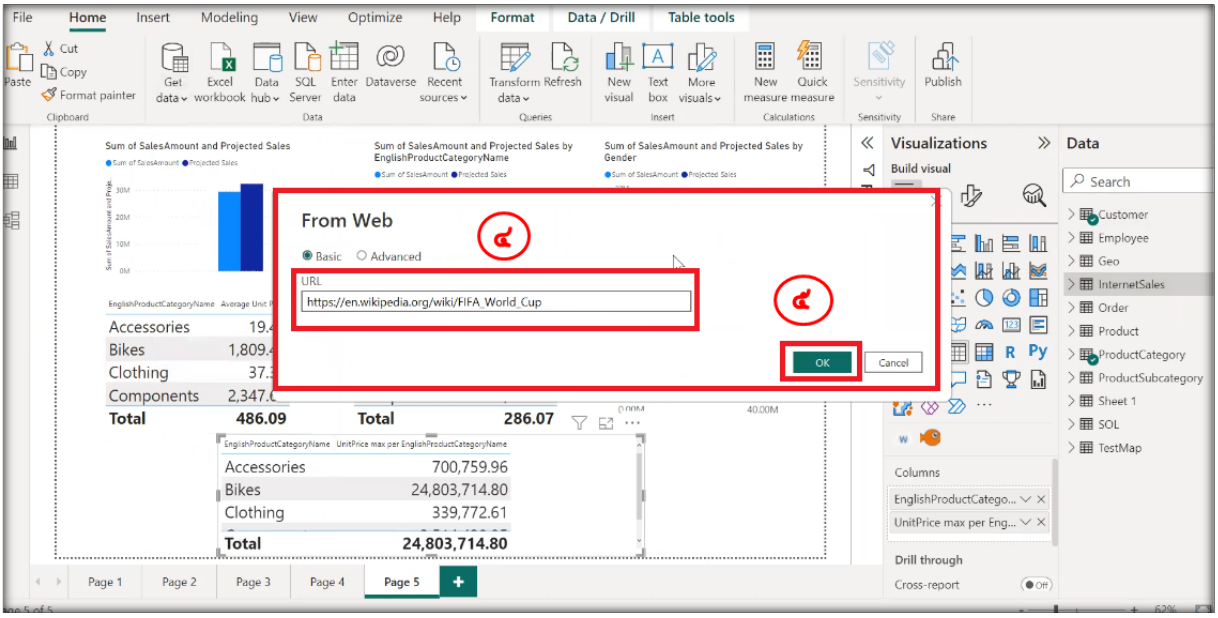
๖) เลือกตารางชุดข้อมูลที่ต้องการ

๗) ตรวจสอบ จัดการข้อมูล เพื่อความถูกต้อง ก่อนนำไปสร้างเป็นหน้ารายงาน Dashboard

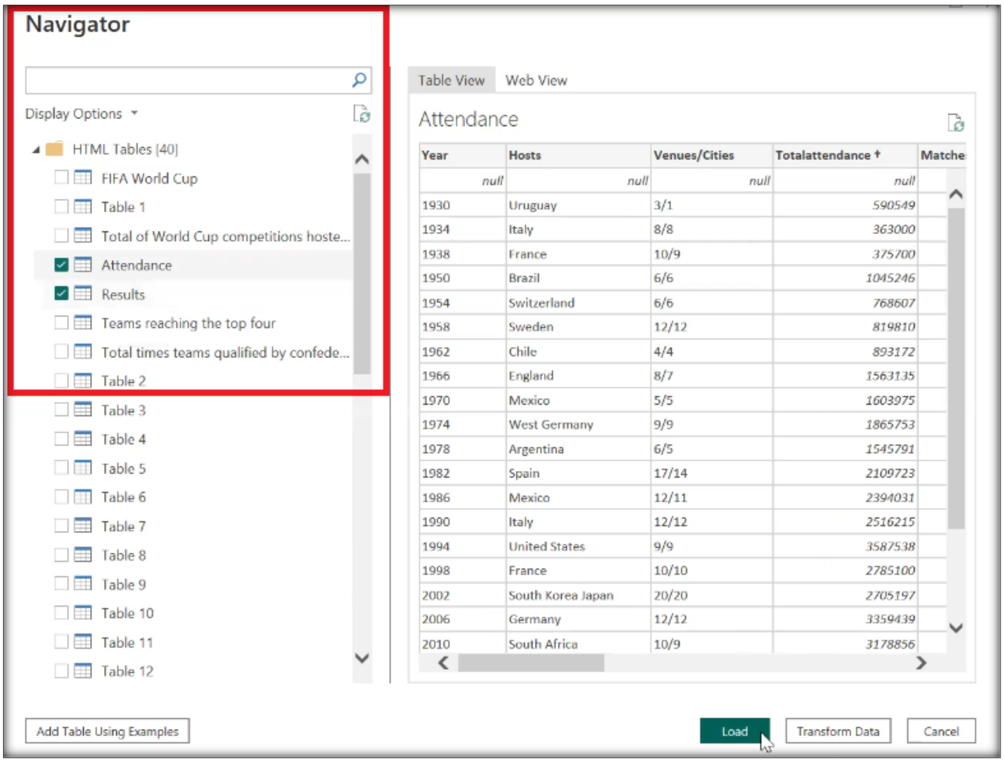
๘) การสร้างรายงาน (Report view) โดยหลังจากที่ได้ตรวจสอบ และจัดการข้อมูลเรียบร้อยแล้ว ดำเนินการสร้างหน้ารายงาน Dashboard เพื่อแสดงผลสรุปข้อมูล



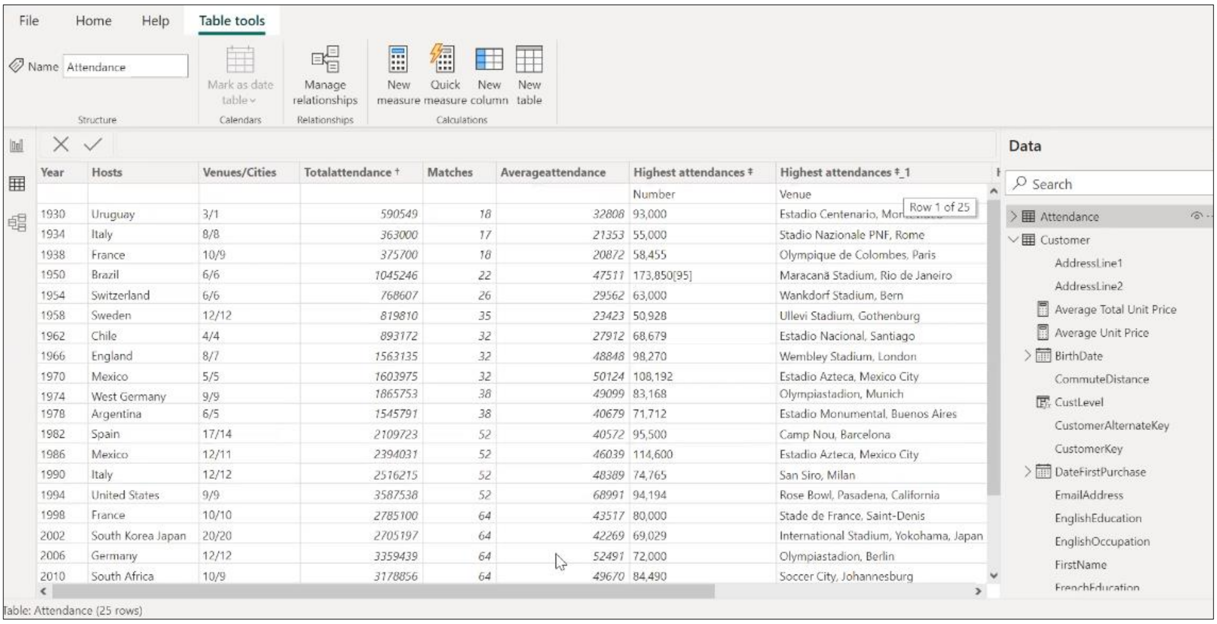
ภาพที่ ๙



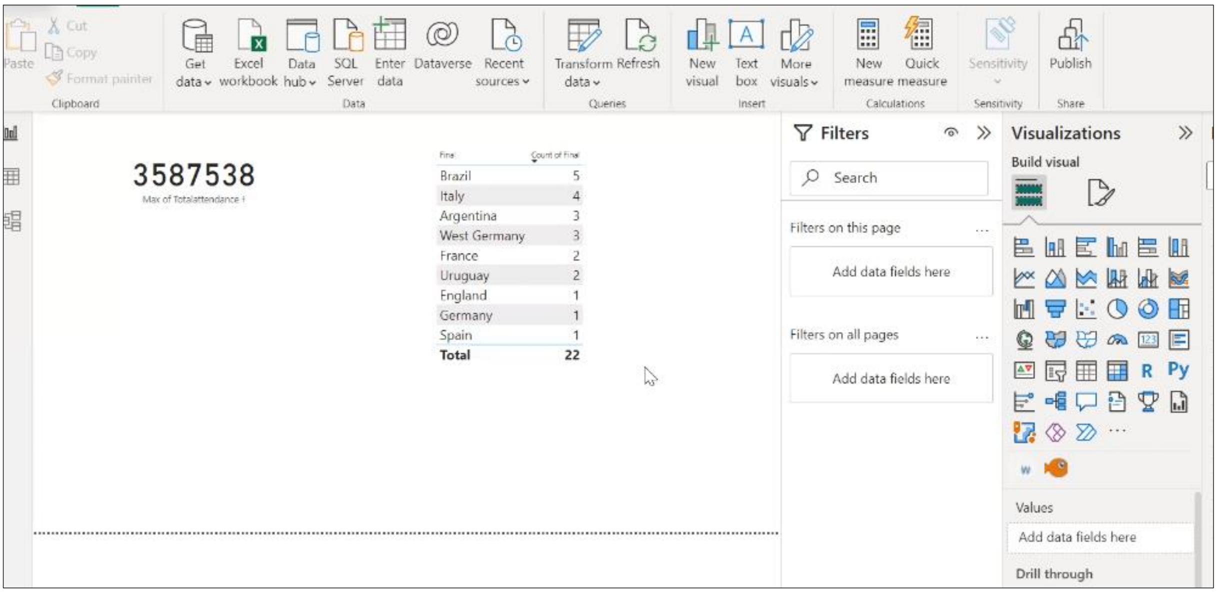
ภาพที่ ๑๐



ภาพที่ ๑๑



ภาพที่ ๑๒



ภาพที่ ๑๓

๑.๓ การ Workshop โดยใช้ข้อมูลจากสำนักงานเศรษฐกิจการเกษตร

๑. การสร้างข้อมูลแผนที่แบบเลือกเฉพาะรายจังหวัดหรือรายอำเภอในแผนที่ประเทศไทย เพื่อนำไฟล์ไปใช้ในโปรแกรม power bi โดยมีขั้นตอนตามภาพที่ ๑๔ - ๒๐ดังนี้

๑) เข้า website https://mapshaper.org/ และกด select เพื่อนำเข้าไฟล์แผนที่

๒) เลือกไฟล์แผนที่ชื่อ th_adm1.json จากโฟลเดอร์ \data_workshop (สศก.) \data\map

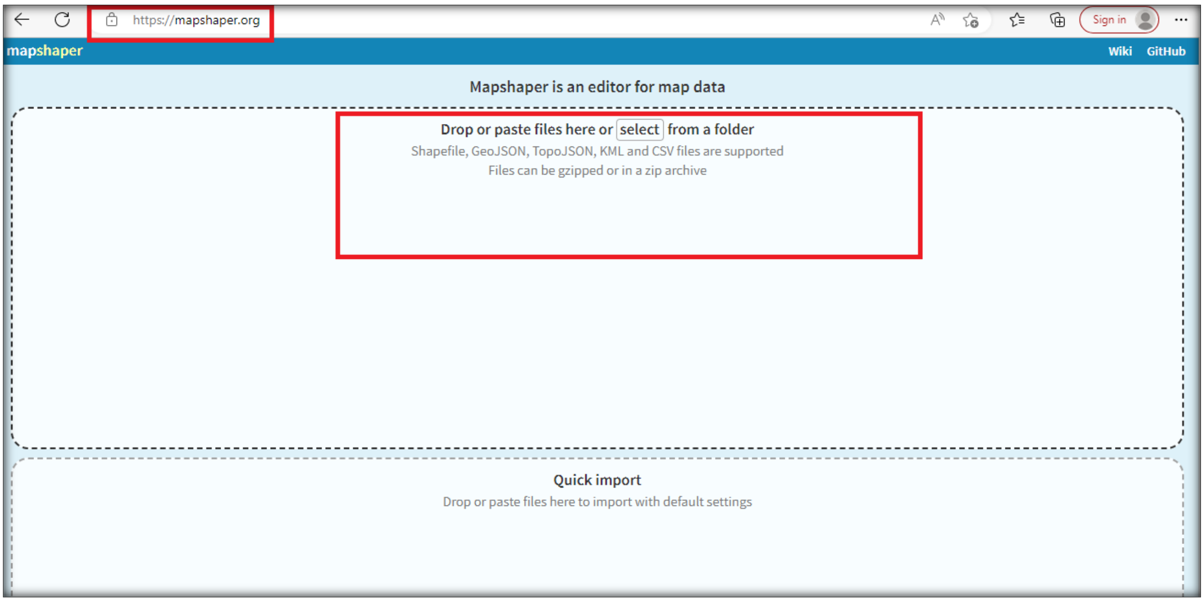
๓) กด Import

๔) กด select features

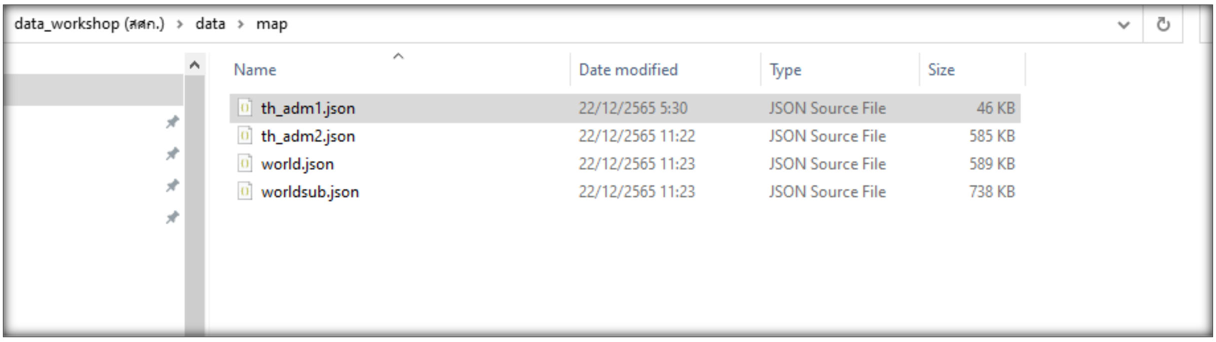
๕) คลิกเลือกพื้นที่จังหวัดที่ต้องการ และกด Split

๖) หลังจากที่กด split จะได้แผนที่เฉพาะจังหวัดที่เลือกไว้ และกด Export ที่มุมบนด้านขวา เพื่อบันทึกไฟล์สำหรับนำไปใช้ในการเปิดบนโปรแกรม Power BI

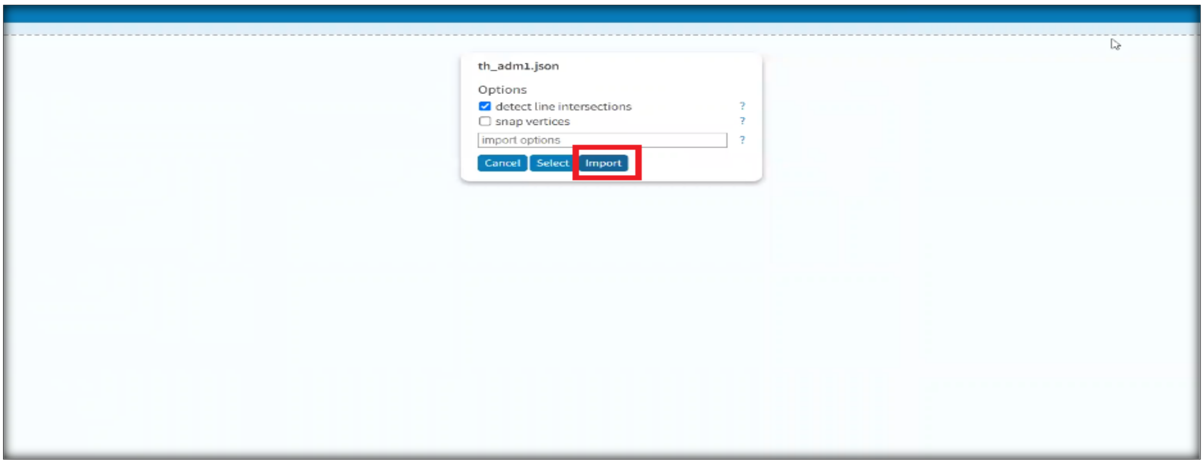
๗) เลือกบันทึกแบบ TopoJson และกด Export อีกครั้ง



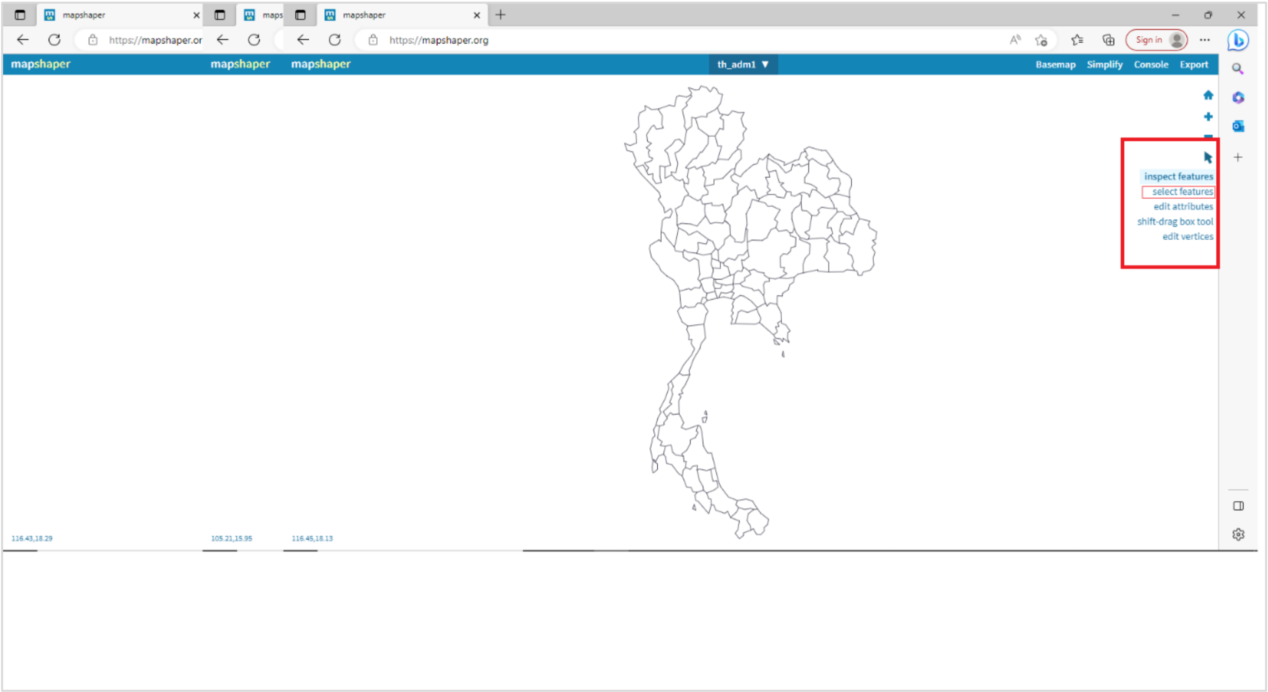
ภาพที่ ๑๔



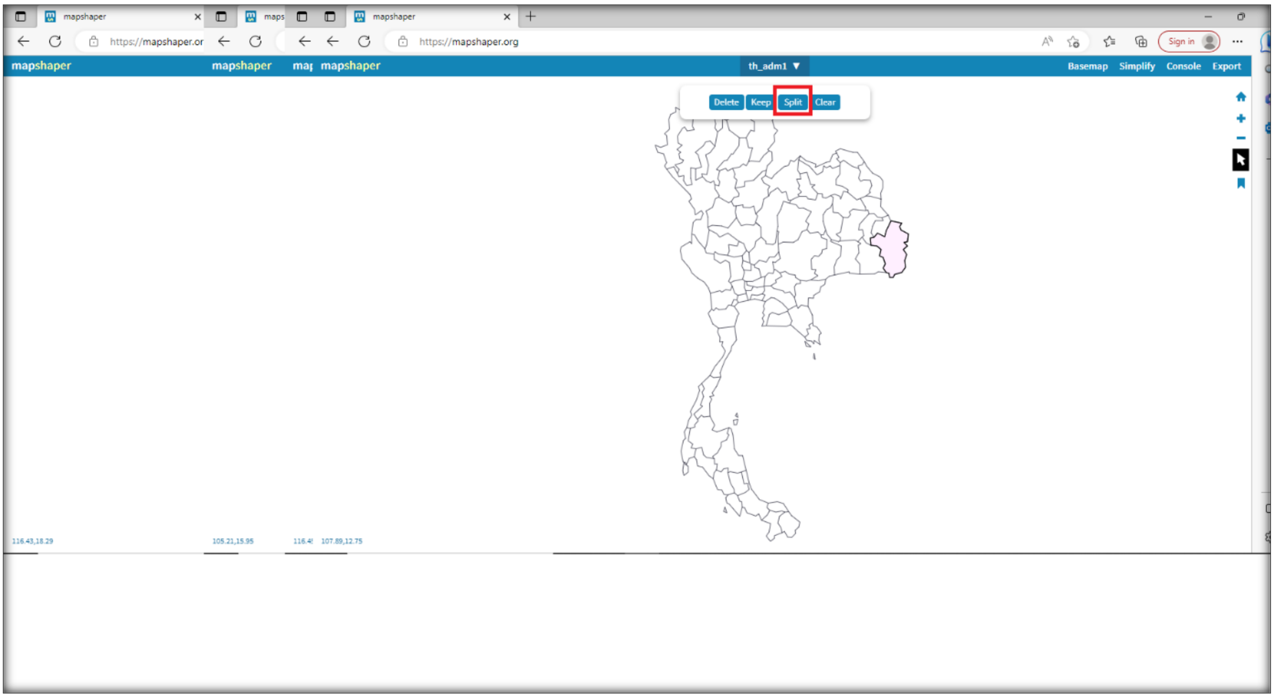
ภาพที่ ๑๕



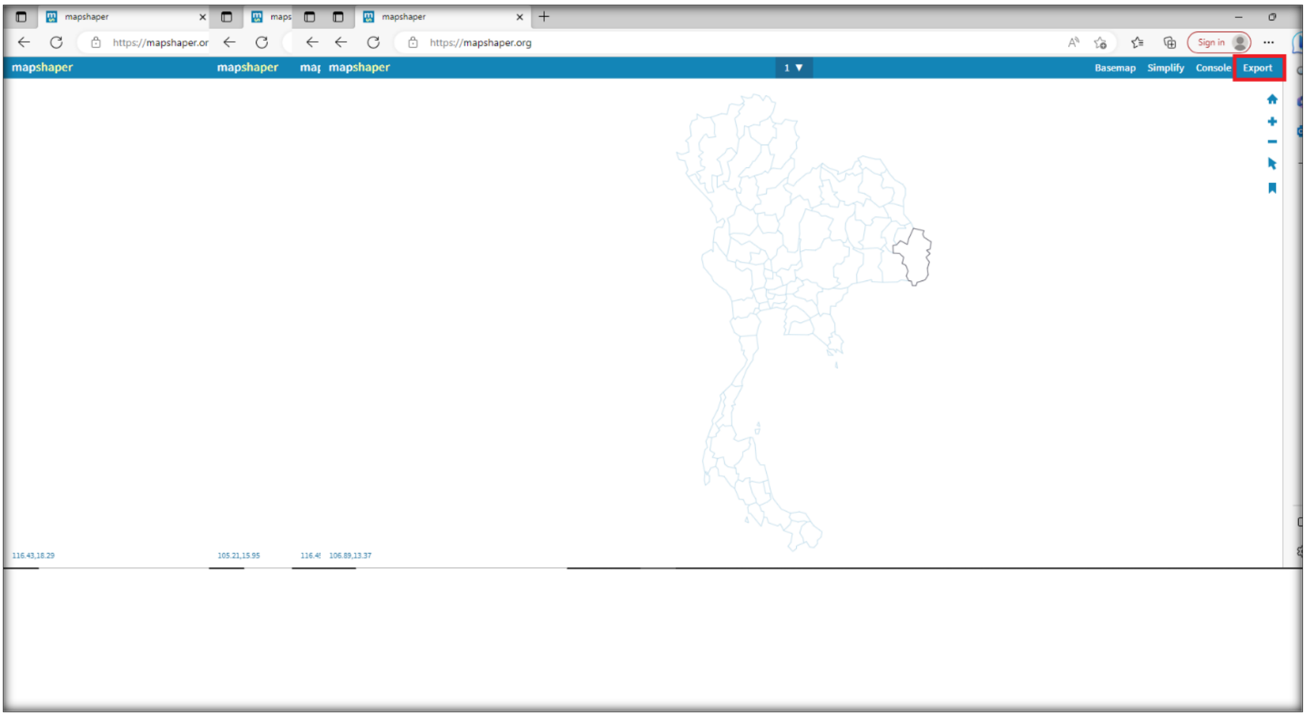
ภาพที่ ๑๖



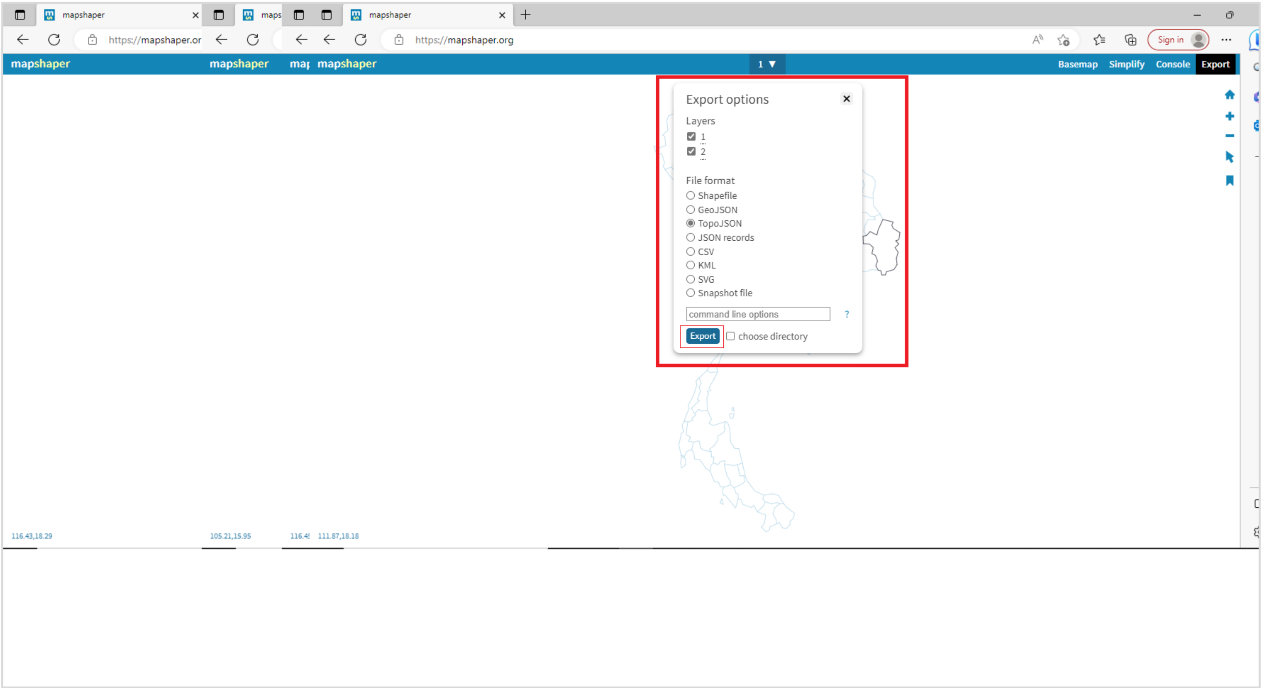
ภาพที่ ๑๗



ภาพที่ ๑๘

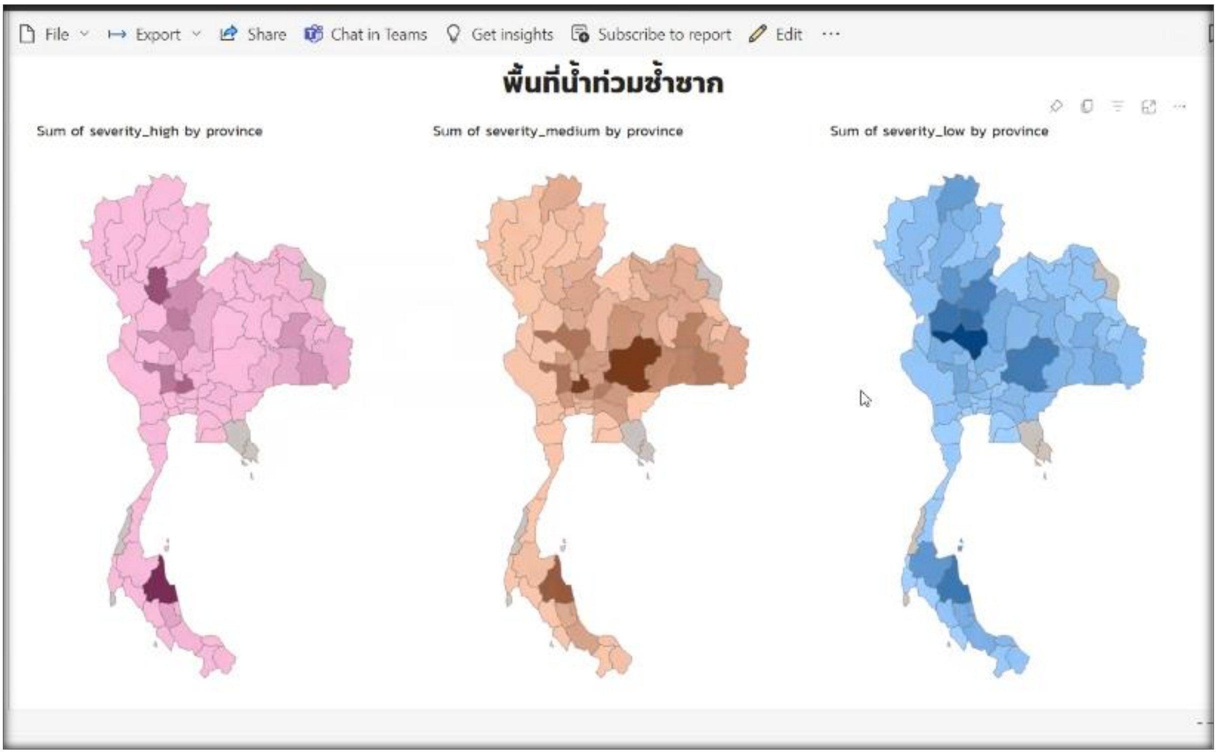


ภาพที่ ๑๙



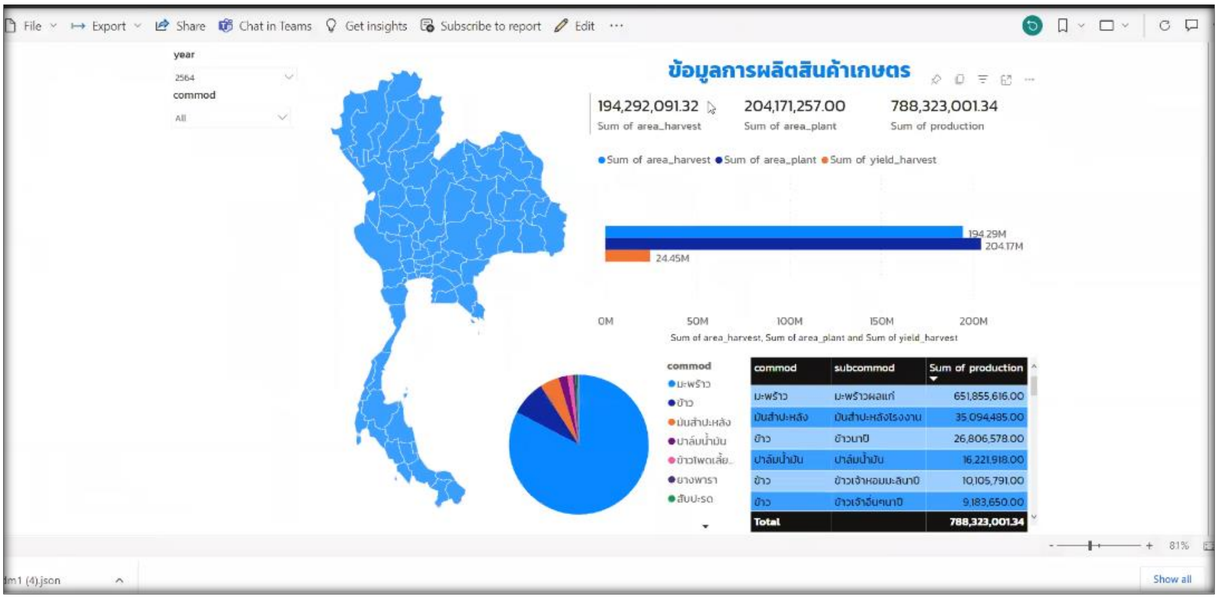
ภาพที่ ๒๐

๒. ตัวอย่างของการนำข้อมูลไปใช้เพื่อสร้างแดชบอร์ด จากชุดข้อมูลพื้นที่น้ำท่วมซ้ำซาก จากกรมพัฒนาที่ดิน โดยสามารถแสดงกราฟแผนที่ประเทศไทยเกี่ยวกับพื้นที่น้ำท่วมซ้ำซาก แบ่งเป็นระดับสูง ระดับกลาง และระดับต่ำ ดังภาพที่ ๒๑



ภาพที่ ๒๑

๓. ตัวอย่างของการนำข้อมูลไปใช้เพื่อสร้างแดชบอร์ด จากชุดข้อมูลการผลิตสินค้าเกษตร จากสำนักงานเศรษฐกิจการเกษตร โดยสามารถแสดงกราฟแผนที่ประเทศไทย เกี่ยวกับการผลิตสินค้าเกษตร ได้แก่ เนื้อที่เพาะปลูก เนื้อที่เก็บเกี่ยว ผลผลิตต่อไร่ อันดับสินค้าที่มีผลผลิตมากที่สุด เป็นต้น ดังภาพที่ ๒๒



ภาพที่ ๒๒

๒. การประเมินความพึงพอใจการเข้าร่วมอบรม

เป็นการประเมินความพึงพอใจของผู้เข้าร่วมอบรมที่มีต่อปัจจัยที่เกี่ยวข้องกับการอบรม โดยใช้แบบสอบถาม ซึ่งผลจากการประเมินจะใช้เป็นแนวทางในการพัฒนาคุณภาพการจัดอบรมให้ดียิ่งขึ้นต่อไป โดยมีเกณฑ์ในการประเมิน ดังนี้

คะแนน	การประเมิน	การแปลผล
๕	๔.๐๑ – ๕.๐๐	มากที่สุด
๔	๓.๐๑ – ๔.๐๐	มาก
๓	๒.๐๑ – ๓.๐๐	ปานกลาง
๒	๑.๐๑ - ๒.๐๐	น้อย
๑	๐.๐๐ – ๑.๐๐	น้อยที่สุด

ผู้ตอบแบบสำรวจความพึงพอใจเข้าร่วมอบรมเชิงปฏิบัติการเรื่อง การใช้งานเครื่องมือเพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร “การจัดการข้อมูลและการจัดทำแดชบอร์ดสำหรับนักวิเคราะห์ข้อมูล (Data Analytics) รุ่นที่ ๒” ระหว่างวันที่ ๑๙ – ๒๑ เมษายน ๒๕๖๖ เวลา ๐๘.๓๐ – ๑๖.๐๐ น. ผ่านระบบออนไลน์ โดย Google Form จำนวน ๓๐ คน พบว่า ผู้ใช้บริการส่วนใหญ่เป็นเพศชาย ร้อยละ ๔๖.๖๗ และเพศหญิง ร้อยละ ๕๓.๓๓ โดยมีอายุอยู่ในช่วง ๓๑ – ๔๐ ปี ร้อยละ ๔๓.๓๓ รองลงมาอายุอยู่ในช่วง ๒๐ – ๓๐ ปี ร้อยละ ๓๓.๓๓ และอายุอยู่ในช่วง ๔๑ – ๕๐ ปี ร้อยละ ๒๓.๓๓ สำหรับระดับการศึกษาส่วนใหญ่ จบการศึกษาระดับปริญญาโท ร้อยละ ๕๓.๓๓ รองลงมาระดับปริญญาตรี ร้อยละ ๔๖.๖๗ สำหรับตำแหน่งส่วนใหญ่เป็นนักวิเคราะห์นโยบายและแผน ร้อยละ ๒๐.๐๐ รองลงมาเป็นนักวิเคราะห์นโยบายและแผนชำนาญการ เศรษฐกรปฏิบัติการ และนักวิชาการคอมพิวเตอร์ ร้อยละ ๑๓.๓๓ นักวิชาการคอมพิวเตอร์ปฏิบัติการ ร้อยละ ๑๐.๐๐ นักวิชาการสถิติปฏิบัติการ และ เศรษฐกร ร้อยละ ๖.๖๗ เศรษฐกรชำนาญการ นักประชาสัมพันธ์ชำนาญการ นักวิเคราะห์นโยบายและแผนปฏิบัติการ นักวิชาการแผนที่ภาพถ่ายและเจ้าพนักงานสถิติปฏิบัติงาน ร้อยละ ๓.๓๓ โดยส่วนใหญ่อยู่ภายใต้สังกัดหน่วยงานสำนักงานเศรษฐกิจการเกษตรที่ ๑ – ๑๒ ร้อยละ ๔๐.๐๐ รองลงมาเป็นศูนย์ข้อมูลเกษตรแห่งชาติ ร้อยละ ๒๓.๓๓ สำนักวิจัยเศรษฐกิจการเกษตร กองเศรษฐกิจการเกษตรระหว่างประเทศ กองนโยบายและแผนพัฒนาการเกษตร และศูนย์สารสนเทศการเกษตร ร้อยละ ๖.๖๗ ศูนย์ประเมินผล สำนักงานเลขานุการกรม และกลุ่มพัฒนาระบบบริหาร ร้อยละ ๓.๓๓

รายละเอียด	ร้อยละ
เพศ	
- ชาย	๔๖.๖๗
- หญิง	๕๓.๓๓
อายุ	
- ๒๐ – ๓๐ ปี	๓๓.๓๓
- ๓๑ – ๔๐ ปี	๔๓.๓๓
- ๔๑ – ๕๐ ปี	๒๓.๓๓
- ๕๑ – ๖๐ ปี	๐.๐๐
รายละเอียด	**ร้อยละ**
ระดับการศึกษา	
- ต่ำกว่าปริญญาตรี	๐.๐๐
- ปริญญาตรี	๔๖.๖๗
- ปริญญาโท	๕๓.๓๓
- ปริญญาเอก	๐.๐๐
ตำแหน่ง	
- นักวิเคราะห์นโยบายและแผน	๒๐.๐๐
- นักวิชาการคอมพิวเตอร์	๑๓.๓๓
- นักวิเคราะห์นโยบายและแผนชำนาญการ	๑๓.๓๓
- เศรษฐกรปฏิบัติการ	๑๓.๓๓
- นักวิชาการคอมพิวเตอร์ปฏิบัติการ	๑๐.๐๐
- นักวิชาการสถิติปฏิบัติการ	๖.๖๗
- เศรษฐกร	๖.๖๗
- เศรษฐกรชำนาญการ	๓.๓๓
- นักประชาสัมพันธ์ชำนาญการ	๓.๓๓
- นักวิเคราะห์นโยบายและแผนปฏิบัติการ	๓.๓๓
- นักวิชาการแผนที่ภาพถ่าย	๓.๓๓
- เจ้าพนักงานสถิติปฏิบัติงาน	๓.๓๓
หน่วยงาน	
- สำนักงานเศรษฐกิจการเกษตรที่ ๑ – ๑๒	๔๐.๐๐
- ศูนย์ข้อมูลเกษตรแห่งชาติ	๒๓.๓๓
- สำนักวิจัยเศรษฐกิจการเกษตร	๖.๖๗
- กองเศรษฐกิจการเกษตรระหว่างประเทศ	๖.๖๗
- กองนโยบายและแผนพัฒนาการเกษตร	๖.๖๗
- ศูนย์สารสนเทศการเกษตร	๖.๖๗
- ศูนย์ประเมินผล	๓.๓๓
- สำนักงานเลขานุการกรม	๓.๓๓
- กลุ่มพัฒนาระบบบริหาร	๓.๓๓

๒.๑ ความพึงพอใจของผู้เข้าร่วมอบรม ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรม การจัดการข้อมูลและการจัดทำแดชบอร์ด สำหรับนักวิเคราะห์ข้อมูล (Data Analytics) ในประเด็นต่าง ๆ อาทิ เช่น ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร เทคนิคหรือวิธีการถ่ายทอดความรู้ ความเหมาะสมของเนื้อหาในการอบรม การสรุปเนื้อหาสาระได้อย่างชัดเจน ความรู้ความเข้าใจเนื้อหา ก่อน-หลัง การเข้าร่วมอบรมเชิงปฏิบัติการ ความสามารถในการนำความรู้ ความเข้าใจ และประสบการณ์ที่ได้จากการเข้าร่วมอบรมเชิงปฏิบัติการไปปรับใช้ในการปฏิบัติงานและถ่ายทอดให้เพื่อนร่วมงาน ระยะเวลาการอบรมเชิงปฏิบัติการ การประสานงานและการอำนวยความสะดวก พบว่า ในภาพรวมผู้รับบริการส่วนใหญ่มีความพึงพอใจอยู่ในระดับมากที่สุด คะแนนเฉลี่ย ๔.๕๓ ร้อยละ ๙๐.๖๗ โดยมีรายละเอียดตามประเด็นต่าง ๆ ดังนี้

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๑. ภาพรวมของวิทยากร						**๔.๖๔**
๑.๑ ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร	๖๔.๒๙ (๑๘)	๓๕.๗๑ (๑๐)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๔
๑.๒ เทคนิคหรือวิธีการถ่ายทอดความรู้	๖๔.๒๙ (๑๘)	๓๕.๗๑ (๑๐)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๔
๑.๓ ความเหมาะสมของเนื้อหาในการอบรม	๖๐.๗๑ (๑๗)	๓๙.๒๙ (๑๑)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๑
๑.๔ การสรุปเนื้อหาสาระได้อย่างชัดเจน	๖๗.๘๖ (๑๙)	๓๒.๑๔ (๙)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๘
๒. ความรู้ความเข้าใจเกี่ยวกับเนื้อหาวิชาที่เข้าอบรมเชิงปฏิบัติการ						**๔.๓๑**
๒.๑ ก่อน การเข้าร่วมอบรมเชิงปฏิบัติการท่านมีความรู้ความเข้าใจเพียงใด	๑๗.๘๖ (๕)	๗.๑๔ (๒)	๑๔.๒๙ (๔)	๓๙.๒๙ (๑๑)	๒๑.๔๓ (๖)	๒.๖๑
๒.๒ หลัง การเข้าร่วมอบรมเชิงปฏิบัติการท่านมีความรู้ความเข้าใจเพียงใด	๓๕.๗๑ (๑๐)	๖๐.๗๑ (๑๗)	๓.๕๗ (๑)	๐.๐๐	๐.๐๐	๔.๓๒
๒.๓ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมเชิงปฏิบัติการไปปรับใช้ ในการปฏิบัติงานได้เพียงใด	๓๕.๗๑ (๑๐)	๖๐.๗๑ (๑๗)	๓.๕๗ (๑)	๐.๐๐	๐.๐๐	๔.๓๒
๒.๔ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วม	๓๙.๒๙ (๑๑)	๕๐.๐๐ (๑๔)	๑๐.๗๑ (๓)	๐.๐๐	๐.๐๐	๔.๒๙

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
อบรมเชิงปฏิบัติการไปถ่ายทอด ให้เพื่อนร่วมงานได้เพียงใด						
๓. ภาพรวมการจัดการอบรมเชิงปฏิบัติการ						๔.๖๒
๓.๑ ระยะเวลาการอบรมเชิงปฏิบัติการ	๔๖.๔๓ (๑๓)	๕๐.๐๐ (๑๔)	๓.๕๗ (๑)	๐.๐๐	๐.๐๐	๔.๔๓
๓.๒ การประสานงานและการอำนวยความสะดวก	๕๓.๕๗ (๑๕)	๔๖.๔๓ (๑๓)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๕๔
๓.๓ ความพึงพอใจในภาพรวมของการอบรมเชิงปฏิบัติการ	๘๙.๒๙ (๒๕)	๑๐.๗๑ (๓)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๘๙
รวม						**๔.๕๒**

๓. จุดเด่น ของการเข้าร่วมอบรมที่ประทับใจ

๑) สามารถนำความรู้มาใช้กับงานได้อย่างดียิ่ง
๒) สามารถนำมาใช้เป็นเครื่องมือเพื่อนำมาประมวลผลและแสดงผลข้อมูลมากขึ้น
๓) ได้ลงมือปฏิบัติจริง วิทยากรมีความสามารถ
๔) อาหารเบรกอร่อย สถานที่จัดอบรมสะดวกและวิทยากรมีความรู้
๕) ระยะเวลาสมเหตุสมผล วิทยากรสอนดี ใจเย็น ไม่กดดันผู้อบรม ถ่ายทอดความรู้ได้ดีมาก อาหารและอาหารว่าง อร่อย มีคุณภาพ น่ารับประทาน
๖) สามารถนำองค์ความรู้ที่ได้รับไปประยุกต์กับการวิเคราะห์ข้อมูลของงานและนำเสนอในรูปแบบที่แตกต่างจากเดิม
๗) ทำให้ได้รับความรู้และนำไปใช้ในการทำงานได้ดี
๘) ได้รับความรู้ใหม่เพื่อนำไปประยุกต์ใช้กับการทำงาน
๙) ช่วยทำให้รู้โปรแกรมและสามารถนำไปใช้ประโยชน์ได้มากขึ้น
๑๐) เป็นโปรแกรมที่น่าสนใจมากๆ เหมาะกับการทำงาน
๑๑) ได้ความรู้เยอะมาก
๑๒) สามารถนำความรู้ที่ได้ไปปรับปรุงรูปแบบการนำเสนอข้อมูลที่ทำให้น่าสนใจ รวมทั้งการนำเสนอข้อมูลบนเว็บไซต์ สศท. ทันสมัยและน่าสนใจมากขึ้น
๑๓) วิทยากรสอนเข้าใจง่าย
๑๔) นำไปปรับใช้กับทำงานได้
๑๕) วิทยาการมีเทคนิคการสอนได้อย่างเหมาะสม และสามารถเข้าใจได้ง่าย
๑๖) วิทยากรมีความเชี่ยวชาญ
๑๗) ได้ทักษะการนำเสนอข้อมูลในรูปแบบรูปภาพ
๑๘) ทำให้เข้าใจเรื่องการจัดการฐานข้อมูลขนาดใหญ่ ได้อย่างเหมาะสม และมีประสิทธิภาพ

๑๙) สามารถนำไปปรับใช้ในการทำงานได้ ในการออกรายงานต่าง ๆ
๒๐) นำไปใช้กับการทำงานได้จริง
๒๑) ได้เปิดโลกการจัดการข้อมูลใหม่ ๆ
๒๒) ได้ความรู้ อาหารเบรกอร่อย
๒๓) ได้รับความรู้ในการใช้โปรแกรม Power BI มากขึ้น
๒๔) ได้รับความรู้ สามารถต่อยอดการนำเสนอข้อมูลของ สศก. ได้หลากหลายและมีความน่าเชื่อถือ ประกอบกับผู้จัดอบรม วิทยากรและผู้เข้าร่วมอบรมให้ความร่วมมือกันดีมาก ๆ
๒๕) ทำให้ทราบว่ามีโปรแกรมที่สามารถจัดการข้อมูลได้อย่างรวดเร็วและสะดวกสำหรับการจัดการกับข้อมูลที่มีปริมาณมาก
๒๖) ได้รับความรู้เพื่อนำไปประยุกต์ใช้ต่อยอดในการทำงานได้

๔. จุดด้อย ที่ควรปรับปรุง

๑) ควรระบุสเปคคร่าวๆ ของ Laptop เนื่องจากบางเครื่องช้า ตามเนื้อหาไม่ทัน
๒) เอกสารประกอบการอบรมเย็บมาผิด กลับหัวกลับหาง
๓) ต้องนำคอมฯ มาเอง อยากให้จัดในห้องปฏิบัติการที่มีคอมพิวเตอร์เตรียมไว้ให้
๔) อยากให้มีไวไฟ และอุปกรณ์ให้ผู้เข้าร่วมอบรม
๕) ควรมีคอมพิวเตอร์สำรองให้ใช้กรณีที่บางท่านขัดข้องในการใช้งานโปรแกรม
๖) อยากให้คุมเวลาสอนไม่ให้เกินเที่ยง
๗) อยากให้จัดหลายวัน
๘) บางครั้งอาจารย์ไปเร็ว และเวลาเทียบกับเนื้อหายังไม่ค่อยสอดคล้องกัน
๙) ส่งแบบฟอร์มเบิกเงินให้ผู้เข้าอบรมก่อนถึงวันอบรมเพื่อจัดเตรียมเอกสารให้ถูกต้อง
๑๐) อาจจะมีการยกเคสของหน่วยงานเกษตรในการจัดทำไปทีละขั้นตอนให้ดู เพื่อให้เห็นภาพและเข้าใจมากยิ่งขึ้น
๑๑) อาหารกลางวันไม่ค่อยอร่อย
๑๒) ไม่ควรจัดติดกับเทศกาลหรือวันหยุดยาว การเดินทางลำบาก
๑๓) วิทยากรสอนไวไปนิดนึงคะ
๑๔) อยากให้เพิ่มระยะเวลาการอบรมมากขึ้น เช่น ๔-๕ วัน เพื่อให้มีความรู้ความเข้าใจมากขึ้น เนื่องจากมีรายละเอียดค่อนข้างมาก

๕. ข้อเสนอแนะเพิ่มเติม

๑) เป็นหลักสูตรที่ดีมาก
๒) บุคลากรหน่วยงานภายในกระทรวงเกษตรมีความสนใจในเครื่องมือ อาจจะเปิดโอกาสให้บุคคลภายนอกมาร่วมในการอบรมด้วย
๓) อาจจะหารือรูปแบบการนำเสนอ Dashboard ของหน่วยงานภาพรวม เพื่อให้เกิดการทำงานแบบบูรณาการร่วมกัน
๔) ขอบคุณทุกท่านที่จัดงานนี้ขึ้นมา
๕) เนื้อหาที่สอนเป็นหัวข้อเกินไป อยากให้นำงานหรือตัวอย่างข้อมูลจริงของ สศก. มาใช้ซึ่งจะทำมองเห็นภาพชัดเจนขึ้น
๖) อยากให้มี Case Study มากกว่านี้
๗) เตรียมคอมพิวเตอร์ไว้สำหรับอบรม เนื่องจากเครื่องที่นำไปเองอาจจะมี ram ที่น้อยทำให้เครื่องค้าง

๘) เห็นควรจัดให้บุคลากร สศก. อื่นๆ ได้เรียนรู้จะเป็นประโยชน์ยิ่ง

๙) เอกสารประกอบการอบรม หากจัดทำเป็นคู่มือการทำงานจะได้สะดวกกว่าเดิม และสามารถนำกลับมาปฏิบัติงานได้ บางช่วงในการอบรมนั้นตามไม่ทันบ้าง

๑๐) ควรจัดห่างจากช่วงเทศกาล

๑๑) ควรประสานส่วนเทคโนฯ ให้เพิ่มความแรงของ Wifi ให้มากขึ้น และแจ้งผู้เข้าอบรมให้เตรียม Notebook ที่สเปคของเครื่องสามารถรองรับระบบได้

๑๒) การจัดอบรมต่อยอดการใช้งาน Power BI เช่น Beginner intermediate Advanced อยากให้อาจารย์เน้นเรื่อง Power Query มากกว่า Power BI เพราะคิดว่า Power BI เมื่อเรียนพื้นฐานแล้วก็สามารถต่อยอดเองได้ง่ายกว่า

สรุปการอบรมเชิงปฏิบัติการ เรื่อง การใช้งานเครื่องมือเพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data)
หลักสูตร "การประยุกต์เทคโนโลยีการรับรู้จากระยะไกลและภูมิสารสนเทศเพื่อการเกษตร"
วันที่ ๒๗ - ๓๐ มิถุนายน ๒๕๖๖ เวลา ๐๘.๓๐ – ๑๖.๐๐ น.
ณ ห้องประชุม ๓ ชั้น ๓ อาคารนวัตกรรม
สำนักงานเศรษฐกิจการเกษตร และผ่านโปรแกรม Zoom Meeting

....

ศูนย์ข้อมูลเกษตรแห่งชาติ (ศกช.) โดยส่วนธรรมาภิบาลข้อมูล (สธข.) ได้จัดอบรมเชิงปฏิบัติการ เรื่อง การใช้งานเครื่องมือ เพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร "การประยุกต์เทคโนโลยีการรับรู้จากระยะไกลและภูมิสารสนเทศเพื่อการเกษตร" โดยมีวัตถุประสงค์เพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในการจัดการและวิเคราะห์ข้อมูลเชิงพื้นที่ โดยมีผู้เข้าร่วมอบรม ๓๐ คน รายละเอียดของการอบรมและข้อคิดเห็นต่าง ๆ สรุปได้ดังนี้

๑. ภาคบรรยาย เป็นการให้ความรู้ทางวิชาการโดยวิทยากร คุณนววิทย์ พงศ์อนันต์ ผู้เชี่ยวชาญงานพิเศษด้านข้อมูลระดับสูง สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (สวข.) โดยมีหัวข้อบรรยาย ดังนี้

๑.๑ แนะนำโปรแกรม

ภูมิสารสนเทศศาสตร์ (Geoinformatics) ศาสตร์สารสนเทศที่เน้นการบูรณาการเทคโนโลยีทางด้านการสำรวจ การทำแผนที่ และการวิเคราะห์ข้อมูลเชิงพื้นที่เข้าด้วยกัน เพื่อศึกษาเกี่ยวกับพื้นที่บนโลกประกอบด้วย ระบบสารสนเทศภูมิศาสตร์ (GIS) การรับรู้จากระยะไกล (RS) และระบบกำหนดตำแหน่งบนพื้นโลก (GNSS) เทคโนโลยีทั้งสามประเภทนี้สามารถทำงานเป็นอิสระต่อกัน หรือสามารถนำมาเชื่อมโยงร่วมกัน ทำให้ประสิทธิภาพเพิ่มมากขึ้น สามารถนำมาใช้ประโยชน์ได้หลายด้าน เช่น กิจการทหาร การจัดการทรัพยากรธรรมชาติ การจัดการภัยพิบัติต่าง ๆ การวางผังเมืองและชุมชน หรือแม้แต่ในเชิงธุรกิจก็ได้มีการนำเทคโนโลยีภูมิสารสนเทศศาสตร์มาประยุกต์ใช้และประกอบการวางแผนการตัดสินใจในเรื่องต่าง ๆ ได้อย่างถูกต้อง รวดเร็ว และมีประสิทธิภาพ

องค์ประกอบของระบบสารสนเทศภูมิศาสตร์

๑. คอมพิวเตอร์และอุปกรณ์ต่างๆ (Hardware)

๒. โปรแกรม (Software)

๓. ข้อมูล (Data) แบ่งออกเป็น ข้อมูลเชิงพื้นที่ (Spatial Data) และข้อมูลเชิงบรรยาย (Attribute Data, Non-Spatial Data)

๔. บุคลากร (People)

๕. ขั้นตอนการทำงาน (Procedures)

ข้อมูลในทางภูมิศาสตร์ แบ่งออกเป็น ๒ รูปแบบ ได้แก่

๑. ข้อมูลเชิงพื้นที่ (Spatial Data) เป็นข้อมูลที่สามารถอ้างอิงกับตำแหน่งทางภูมิศาสตร์ทางภาคพื้นดิน โดยแบ่งออกเป็น ๒ ประเภท ดังนี้

๑.๑ ข้อมูล Raster คือ จุดของเซลล์ ที่อยู่ในแต่ละช่วงสี่เหลี่ยม (Grid) โครงสร้างของ Raster ประกอบด้วยชุดของ Grid cell หรือ Pixel ข้อมูลแบบ Raster เป็นข้อมูลที่อยู่บนพิกัดรูปตารางแถวนอนและแถวตั้งแต่ละ cell อ้างอิงโดยแถวและสดมภ์ภายใน cell จะมีตัวเลขหรือภาพข้อมูล Raster ความสามารถแสดงรายละเอียดของข้อมูล Raster ขึ้นอยู่กับขนาดของเซลล์ ณ จุดพิกัดที่ประกอบขึ้นเป็นฐานข้อมูลแสดงตำแหน่งชุดนั้น ซึ่งข้อมูลประเภท Raster มีข้อได้เปรียบในการใช้ทรัพยากรระบบคอมพิวเตอร์ที่มีประสิทธิภาพดีกว่า ข้อมูล Raster เช่น ข้อมูลภาพถ่ายทางอากาศ ข้อมูลภาพถ่ายดาวเทียม ข้อมูลความสูง (DEM) เป็นต้น

๑.๒ ข้อมูล Vector คือ ลักษณะทางภูมิศาสตร์ที่เป็นตัวแทนของปรากฏการณ์ทางภูมิศาสตร์บนโลกแผนที่ กระดาษบันทึกตำแหน่งทางภูมิศาสตร์และแทนสิ่งต่าง ๆ บนโลกที่เป็นลายเส้นและพื้นที่ด้วยสัญลักษณ์แบบ จุด เส้น พื้นที่และตัวอักษร ในระบบสารสนเทศภูมิศาสตร์จะใช้ Feature ประเภทต่าง ๆ ในการแทนปรากฏการณ์โดยแบ่งออกเป็น ๓ ลักษณะ ดังนี้

- จุด (Point) เป็นลักษณะของจุดในตำแหน่งใด ๆ ซึ่งจะสังเกตได้จากขนาดของจุดนั้น ๆ โดยจะอธิบายถึงตำแหน่งที่ตั้งของข้อมูล เช่น ที่ตั้งสถานที่สำคัญต่าง ๆ เป็นต้น
- เส้น (Line) ประกอบไปด้วยลักษณะของเส้นตรงเส้นหักมุมและเสนโค้ง ซึ่งรูปร่างของเส้นเหล่านี้จะอธิบายถึงลักษณะต่าง ๆ โดยอาศัยขนาดทั้งความกว้างและความยาว เช่น ถนน หรือ แม่น้ำ เป็นต้น
- รูปปิด (Polygon) อธิบายถึงขอบเขตเนื้อที่และเส้นรอบวง และข้อมูล Polygon ลักษณะเหล่านี้จะใช้อธิบายขอบเขตของข้อมูลต่าง ๆ เช่น ขอบเขตของพื้นที่จำแนกการใช้ประโยชน์ เป็นต้น

๒. ข้อมูลเชิงบรรยาย (Attribute data) เป็นข้อมูลที่เกี่ยวข้องกับคุณลักษณะต่าง ๆ ในพื้นที่นั้น ๆ ได้แก่ ข้อมูลการถือครองที่ดิน ข้อมูลปริมาณธาตุอาหารในดิน และข้อมูลเกี่ยวกับสภาวะเศรษฐกิจและสังคม เป็นต้น

ความสามารถของ GIS

๑. การรวบรวมข้อมูล (Capturing Data) สามารถรวบรวมข้อมูลให้เป็นระบบฐานข้อมูลทางภูมิศาสตร์ ซึ่งเกิดจากการนำเข้าข้อมูลโดยวิธีต่าง ๆ เช่น การดิจิไทซ์ (Digitizing) สแกนภาพ (Scanning) On - screen Digitizing การนำเข้าจาก GPS (Global Positioning System) เป็นต้น โดยต้องคำนึงถึงความถูกต้องเป็นสำคัญ ทั้งการนำเข้าข้อมูลเชิงพื้นที่และข้อมูลตาราง

๒. การจัดเก็บข้อมูล (Storing Data) จะจัดเก็บข้อมูลในรูปของ Vector หรือ Raster รวมไปถึงข้อมูลตารางในระบบฐานข้อมูลเชิงสัมพันธ์

๓. การสืบค้นข้อมูล (Querying Data) สามารถสืบค้นข้อมูลที่สนใจตามตำแหน่งหรือเงื่อนไขที่ต้องการ โดยการสืบค้นข้อมูลเชิงบรรยายจะสืบค้นจากข้อมูลเชิงพื้นที่ และการสืบค้นข้อมูลเชิงพื้นที่จะสืบค้นจากการสร้างเงื่อนไขในข้อมูลเชิงบรรยาย เช่น การหาประเทศที่มีประชากรมากกว่า ๒๐ ล้านคน เป็นต้น

๔. การวิเคราะห์ข้อมูล (Analyzing Data) เพื่อใช้ตอบปัญหาต่าง ๆ โดยมีเทคนิคที่สำคัญ ดังนี้

- การวิเคราะห์การซ้อนทับ (Overlay Analysis)
- แบบจำลอง (Modeling)
- การกำหนดแนวกันชน (Buffer)
- การวิเคราะห์โครงข่าย (Network Analysis)

๕. การแสดงผลข้อมูล (Displaying Data) GIS มีเครื่องมือในการแสดงข้อมูลด้านภูมิศาสตร์ โดยการให้สัญลักษณ์ที่แตกต่างกัน สำหรับกระบวนการวิเคราะห์จะได้ผลลัพธ์ซึ่งสามารถสร้างเป็นแผนที่ กราฟ หรือรายงานเพื่อให้ดูเข้าใจได้ง่ายขึ้น เพื่อให้บรรลุวัตถุประสงค์ที่ต้องการแสดง ทั้งนี้รวมไปถึงการใช้มัลติมีเดีย

๖. การสร้างผลงานจากข้อมูล (Outputting Data) การเผยแพร่ผลงานด้าน GIS ทำให้เกิดการแลกเปลี่ยนทรัพยากรด้าน GIS ซึ่งอาจอยู่ในรูปของแผนที่ กราฟ หรือรายงาน เมื่อมีการเผยแพร่มากขึ้นทำให้เพิ่มศักยภาพของงานด้านนี้มากขึ้น รวมทั้งทำให้ผู้ที่สนใจได้รับข้อมูลที่มีความถูกต้องมากขึ้น

๑.๒ การเรียกค้นและวิเคราะห์ข้อมูลซ้อนทับเชิงพื้นที่ (Query and Overlay Analysis)

การซ้อนทับข้อมูล (Map overlay) คือ การวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยมากกว่าหนึ่งปัจจัย ต้องอาศัยการซ้อนทับข้อมูลเชิงพื้นที่ โดยใช้เงื่อนไขทางคณิตศาสตร์และตรรกศาสตร์ในการวิเคราะห์ เพื่อให้ได้ผลลัพธ์ตามที่ต้องการ ในการซ้อนทับข้อมูลเชิงพื้นที่ ชั้นข้อมูลแต่ละชั้นจะต้องมีระบบพิกัดเดียวกัน

กระบวนการซ้อนทับ มี ๓ รูปแบบดังนี้

๑. พ้อยต์-อิน-โพลีกอน (Point-in-polygon) คือ การเลือกข้อมูลจุดที่ตกอยู่ในรูปปิดที่ต้องการคำตอบ เช่น จำนวนโรงพยาบาลในอำเภอที่กำหนด เป็นต้น

๒. ไลน์-อิน-โพลีกอน (Line-in-polygon) คือ การเลือกข้อมูลเส้นที่ตกอยู่ในรูปปิดหรือลากผ่านรูปปิดที่ต้องการคำตอบ เช่น จำนวนถนนที่อยู่ในหรือผ่านอำเภอที่กำหนด เป็นต้น

๓. โพลีกอน-ออน-โพลีกอน (Polygon-on-polygon) คือ การซ้อนทับข้อมูลรูปปิดสองชั้นข้อมูล ผลลัพธ์คือชั้นข้อมูลรูปปิดใหม่และข้อมูลคุณลักษณะที่เกิดขึ้นมาจากการผสมกันของข้อมูลทั้งสอง

วิธีการซ้อนทับ มี ๓ รูปแบบดังนี้

๑. การซ้อนทับข้อมูลแบบยูเนียน (Union) คือ ชั้นข้อมูลนำเข้าและชั้นข้อมูลยูเนียนเป็นแบบโพลีกอน-ออน-โพลีกอน โดยใช้ตรรกศาสตร์บูลีน ‘OR’ ผลลัพธ์จะเป็นการรักษารูปปิดเดิมและรูปปิดใหม่ที่เกิดจากการผสมกันของสองชั้นข้อมูล โดยรักษาชั้นข้อมูลเดิมที่ไม่มีพื้นที่ซ้ำซ้อนและผสมข้อมูลคุณลักษณะของชั้นข้อมูลทั้งสองจากชั้นข้อมูลใหม่ที่มีพื้นที่ซ้ำซ้อนกัน

๒. การซ้อนทับข้อมูลแบบอินเตอร์เซกชัน (Intersection) คือ ชั้นข้อมูลนำเข้าและชั้นข้อมูลอินเตอร์เซกชันที่เป็นแบบ พ้อยต์-อิน-โพลีกอน ไลน์-อิน-โพลีกอน และ โพลีกอน-ออน-โพลีกอน โดยมีการทดสอบตรรกศาสตร์บูลีน ‘AND’ ผลลัพธ์จะเป็นการรักษารูปปิดและข้อมูลคุณลักษณะของชั้นข้อมูลทั้งสองที่มีพื้นที่ซ้ำซ้อนกันเท่านั้น

๓. การซ้อนทับข้อมูลแบบไอเดนทิตี้ (Identity) คือ ชั้นข้อมูลนำเข้าและชั้นข้อมูลไอเดนทิตี้เป็นแบบพ้อยต์-อิน-โพลีกอน ไลน์-อิน-โพลีกอน และ โพลีกอน-ออน-โพลีกอน โดยมีการทดสอบตรรกศาสตร์บูลีน ‘[(input) AND (identity)] OR (input)’ ผลลัพธ์จะเป็นการรักษารูปปิดและข้อมูลคุณลักษณะของข้อมูลนำเข้าเป็นเกณฑ์และส่วนที่ซ้อนทับกัน ของชั้นข้อมูลทั้งสอง ข้อมูลคุณลักษณะที่ได้เป็นการผสมกันจากชั้นข้อมูลทั้งสอง

๑.๓ หลักการใช้งาน Remote Sensing

เทคโนโลยีการสำรวจข้อมูลระยะไกล (Remote sensing) เป็นเทคโนโลยีที่ใช้คลื่นแสงในช่วงความยาวคลื่นต่าง ๆ และคลื่นแม่เหล็กไฟฟ้าเป็นสื่อในการได้มาซึ่งข้อมูล โดยส่วนใหญ่อยู่ในรูปของภาพถ่ายทางอากาศหรือภาพถ่ายดาวเทียม หลังจากนั้นข้อมูลจะถูกนำมาจำแนกและวิเคราะห์เพื่อให้เข้าใจถึงวัตถุและสภาพแวดล้อมต่าง ๆ โดยทั่วไปเทคโนโลยีการสำรวจข้อมูลจากระยะไกลจะประกอบด้วย ๒ กระบวนการหลัก คือ

๑. การรับข้อมูล (Data Acquisition) เริ่มตั้งแต่พลังงานแม่เหล็กไฟฟ้าจากแหล่งกำเนิดพลังงาน เช่น ดวงอาทิตย์ เคลื่อนที่ผ่านชั้นบรรยากาศ เกิดปฏิสัมพันธ์กับวัตถุบนพื้นผิวโลก สะท้อนเข้าสู่เครื่องวัดหรือยานสำรวจ (Platform) ที่โคจรผ่านวัตถุนั้น ๆ จากนั้นยานสำรวจจะทำหน้าที่บันทึกข้อมูลของวัตถุหรือปรากฏการณ์บนพื้นผิวโลก และแปลงข้อมูลเป็นสัญญาณอิเล็กทรอนิกส์ เพื่อส่งไปยังสถานี

รับภาคพื้นดิน (Receiving Station) โดยข้อมูลเหล่านั้น จะถูกประมวลผลแปลงเป็นข้อมูลเชิงอนุมาน (Analog Data) หรือข้อมูลเชิงตัวเลข (Digital Data) เพื่อจะนำไปวิเคราะห์ข้อมูล (Data Analysis) ต่อไป

๒. การวิเคราะห์ข้อมูล (Data Analysis) มี ๒ วิธี ดังนี้

๒.๑ การวิเคราะห์ด้วยสายตา (Visual Analysis) ให้ผลข้อมูลออกมาในเชิงคุณภาพ (Quantitative) ไม่สามารถวัดออกมาเป็นค่าตัวเลขได้

๒.๒ การวิเคราะห์ด้วยคอมพิวเตอร์ (Digital Analysis) ให้ผลข้อมูลในเชิงปริมาณ (Quantitative) ที่สามารถแสดงผลการวิเคราะห์ออกมาเป็นค่าตัวเลขได้

เทคโนโลยีการสำรวจข้อมูลระยะไกล (Remote Sensing) กับภาคการเกษตร โดยส่วนใหญ่การสำรวจจากระยะไกลในภาคการเกษตรจะใช้กล้องถ่ายภาพติดบนอากาศยานไร้คนขับ (Drone) เพื่อเก็บข้อมูลในระดับต่ำ (Aerial Photo) และใช้ภาพถ่ายจากดาวเทียมเพื่อเก็บข้อมูลในระดับสูง (Satellite Image) แล้วนำข้อมูลทั้งสองส่วนนี้มาทาการวิเคราะห์ร่วมกัน ซึ่งปัจจุบันภาพถ่ายที่ได้สามารถนำมาวิเคราะห์ข้อมูลที่เป็นประโยชน์ต่อภาคการเกษตรสำหรับการเพาะปลูกพืช เพื่อช่วยด้านการวัดและตรวจวิเคราะห์ข้อมูลด้านการเกษตรแม่นยำ (Precision Farming) ของเกษตรกร เช่น

- การจำแนกประเภทพืช จำนวนลำต้นในบริเวณที่มีการเพาะปลูกหนาแน่น
- การประเมินข้อมูลเกี่ยวกับสุขภาพของพืช เช่น การประมาณค่าคลอโรฟิลล์ ปริมาณไนโตรเจน ปริมาณความชื้นและน้ำที่ปกคลุมผิวดินจากคุณสมบัติการดูดซับในแถบสเปกตรัม เป็นต้น
- การตรวจสอบโรคพืช แมลงศัตรูพืช โดยการวัดการสะท้อนแสงและการเปลี่ยนแปลงในพฤติกรรมสเปกตรัมต่อความเข้มของ Bacterial Leaf Bright (BLB) - การติดตามสถานะการเพาะปลูกช่วงเริ่มต้น ช่วงเจริญเติบโตและสิ้นสุด ด้วยดัชนีพืชพรรณที่ใช้วัดการเปลี่ยนแปลงความเขียว ที่เปลี่ยนไปตามเวลา
- การประมาณการผลผลิตพืช ซึ่งเป็นประเด็นที่สำคัญที่สุดสำหรับการจัดการทางการเกษตร ผลผลิตมีความสัมพันธ์อย่างยิ่งกับการนำไฟฟ้าของดินซึ่งกำหนดพื้นผิวดินและลักษณะความเค็มของดินที่แสดงในแถบสเปกตรัม การสะท้อนแถบแสงของพืชแสดงถึงความเร็วในการเติบโตของพืช ช่วยให้เกษตรกรสามารถจัดการปรับปรุงที่ดินและประเมินผลผลิตก่อนการเก็บเกี่ยว

ทฤษฎีสี

๑. ทฤษฎีสี

- เนื้อสี (Hue) คือ ความแตกต่างของสีบริสุทธิ์แต่ละสีมีความหมายเหมือนกันกับคำว่าสี เช่น สีแดง สีเขียว สีน้ำเงิน สีเหลือง และสีส้ม ซึ่งแตกต่างกันไปตามช่วงคลื่นในสเปกตรัม สีที่สว่างมากจะมีค่าเนื้อสีสูงและเปล่งแสงออกมากกว่า เช่น สีเหลืองสดจะสว่างกว่าสีน้ำเงินเข้ม
- ความเข้ม (Intensity) แสดงถึง ความมืด ความสว่างของสี ข้อมูลความเข้มจะคล้ายคลึงกับข้อมูลการแสดงสีดำและสีขาวของภาพสี
- การอิ่มตัว (Saturation) ในระบบสีที่กำหนดด้วยเอชไอเอส หมายถึง ความบริสุทธิ์ของสี ถ้าสีมีความอิ่มตัวสูงจะมีความบริสุทธิ์ของสีมาก เช่น ความบริสุทธิ์ของสีน้ำเงิน

๒. การถ่ายภาพ

- การถ่ายภาพ (Imaging) การแสดงค่าของวัตถุในรูปแบบอิเล็กทรอนิกส์หรือเชิงแสง ซึ่งบันทึกบนฟิล์มหรือการแสดงด้วยอุปกรณ์อิเล็กทรอนิกส์หรือสื่ออื่น ๆ

- เรดาร์ถ่ายภาพ (Imaging Radar) เครื่องรับรู้แบบ Active ที่บันทึกข้อมูลในช่วงคลื่นไมโครเวฟที่มีความถี่ระหว่าง ๓-๑๒.๕ กิกกะเฮิรตซ์ หรือ ๒.๔-๑๐๐ เซนติเมตร เครื่องรับรู้เรดาร์ทำการผลิตและส่งสัญญาณไมโครเวฟในลักษณะพัลส์ (pulse) ไปยังวัตถุเป้าหมายและรับสัญญาณพัลส์ที่สะท้อนกลับ การทำงานของระบบนี้ต้องอาศัยสายอากาศ (antenna) ที่ทำหน้าที่ส่งและรับสัญญาณสลับกันได้ด้วยสายอากาศเดียวกัน พัลส์ที่สะท้อนกลับมาจากวัตถุเป้าหมายจะถูกบันทึกเอาไว้เพื่อทำการวิเคราะห์ต่อไป ระบบเรดาร์ถ่ายภาพจะผลิตภาพในสองมิติ ได้แก่ พิสัยและแอซิมัทดู Active Sensor ประกอบ

- สเปกโทรมิเตอร์ถ่ายภาพ (Imaging Spectrometry) การใช้สเปกโทรมิเตอร์ในการถ่ายภาพเชิงแสง เพื่อวัดความเข้มแสงที่ความยาวคลื่นในย่านที่ตามองเห็น ย่านอินฟราเรดหรือย่านอัลตราไวโอเลต

- มุมตกกระทบ (Incidence Angle) มุมที่พลังงานแม่เหล็กไฟฟ้าทำกับแนวดิ่ง เมื่อตกกระทบพื้นผิวโลก สำหรับระบบเรดาร์ คือ มุมที่อยู่ระหว่างเส้นสมมุติปกติ (surface normal) กับพื้นผิวและเส้นตรงที่เชื่อมระหว่างสายอากาศกับเป้าหมาย

- พลังงานตกกระทบ (Incident Energy) พลังงานในรูปแบบต่าง ๆ เช่น พลังงานที่ได้จากแสงอาทิตย์หรือพลังงานที่สร้างขึ้นเองในระบบเรดาร์ที่ตกกระทบพื้นผิววัตถุ และมีปฏิสัมพันธ์กับพื้นผิวนั้น เช่น การส่งผ่าน การดูดกลืน และการสะท้อนออกจากพื้นผิว

- ไม่ร่วมนัย/ไม่เชื่อมกัน (Incoherent) คุณสมบัติของชุดสัญญาณหรือข้อมูล ซึ่งเฟสของส่วนประกอบไม่มีสหสัมพันธ์ที่มีนัยสำคัญเชิงสถิติหรือเชิงระบบ

- การแปลตีความทางอ้อม (Indirect Interpretation) การแปลตีความรูปลักษณ์โดยอาศัยการอนุมานจากข้อมูลของรูปลักษณ์อื่น

- อินฟราเรดแบนด์ (Infrared Band) แถบคลื่นอินฟราเรดแถบความถี่คลื่นแม่เหล็กไฟฟ้าที่อยู่ต่อจากช่วงคลื่นที่ตามองเห็น (ตั้งแต่ ๐.๗๐ ไมโครเมตร) ไปจนถึงช่วงคลื่นไมโครเวฟ (ประมาณ ๑,๐๐๐ ไมโครเมตร) ประกอบด้วย

๑) คลื่นอินฟราเรดใกล้ (Near Infrared: NIR) หรืออินฟราเรดสะท้อนอยู่ระหว่างช่วงคลื่น ๐.๗ - ๑.๓ ไมโครเมตรแบ่งออกเป็น คลื่นอินฟราเรดใกล้มาก (๐.๗ - ๑.๐ ไมโครเมตร) และคลื่นอินฟราเรดใกล้ (๑.๐ - ๒.๕ ไมโครเมตร)

๒) คลื่นอินฟราเรดกลาง (Mid Infrared: MIR) ช่วงความยาวคลื่นแม่เหล็กไฟฟ้าจาก ๑.๓ - ๓ ไมโครเมตร ซึ่งส่วนใหญ่เป็นรังสีความร้อนที่แผ่ออกจากวัตถุ เป็นส่วนหนึ่งของแถบคลื่นอินฟราเรดความร้อน

๓) คลื่นอินฟราเรดไกล (Far Infrared: FIR) เป็นแถบคลื่นอินฟราเรดที่อยู่ถัดจากอินฟราเรดกลาง หรือเป็นที่ยอมรับกันว่าเป็นแถบคลื่นไมโครเวฟความยาวสั้น ตั้งแต่ ๑๔ - ๑,๐๐๐ ไมโครเมตร ซึ่งมีข้อจำกัดอย่างมากในการใช้งานบนพื้นดินเนื่องจากชั้นบรรยากาศมีการส่งผ่านรังสีในช่วงคลื่นนี้น้อยมาก

๔) แถบคลื่นอินฟราเรดความร้อน (Thermal Infrared Band) ช่วงคลื่นแม่เหล็กไฟฟ้าที่โดยทั่วไปถือเป็นความยาว คลื่นอินฟราเรดกลางอยู่ในย่านความยาวคลื่นระหว่าง ๓ - ๑๔ ไมโครเมตร การประยุกต์ใช้ข้อมูลการรับรู้จากระยะไกล ส่วนใหญ่จะใช้ช่วงคลื่นที่มีพิสัยระหว่าง ๘ - ๑๔ ไมโครเมตร

การถ่ายภาพ

การถ่ายภาพอินฟราเรด โดยใช้ฟิล์มหรือเครื่องรับรู้ที่มีความไวต่อช่วงคลื่นอินฟราเรด ซึ่งโดยปกติฟิล์มจะมีความไวต่อช่วงคลื่นที่ตามองเห็นด้วย จึงต้องใช้ตัวกรองช่วงคลื่นที่ตามองเห็นไม่ให้ผ่านได้ ข้อดีของการถ่ายภาพอินฟราเรด คือ การทะลุผ่านการสลัวของแสงบนท้องฟ้า ทำให้ได้ภาพที่ชัดเจนกว่าภาพที่ถ่ายในช่วงคลื่นที่ตามองเห็น โดยใช้อุปกรณ์กราดภาพซึ่งจะบันทึกค่าการแผ่รังสีอินฟราเรดหรือไมโครเวฟโดยแปลงเป็นภาพที่แสดงความร้อน ดู Color Infrared (CIR) Image ประกอบ

๑.๔ การแปลภาพถ่ายดาวเทียม (Image Interpretation)

การแปลตีความข้อมูลจากดาวเทียมด้วยสายตา (Visual interpretation) ใช้องค์ประกอบหลักที่สำคัญ (Elements of interpretation) ได้แก่

๑. ความเข้มของสีและสี (Tone/Color) ระดับความแตกต่างของความเข้มของสีหนึ่ง ๆ มีความสัมพันธ์กับค่าการสะท้อนของช่วงคลื่นและการผสมสีของช่วงคลื่นต่าง ๆ เช่น น้ำในช่วงคลื่นอินฟราเรดใกล้ถูกดูดกลืนทำให้ปรากฏเป็นสีดำ ในภาพสีผสมพืชพรรณปรากฏเป็นสีแดงเมื่อกำหนดให้ช่วงคลื่นอินฟราเรดใกล้เป็นสีแดง ช่วงคลื่นสีแดงกำหนดให้เป็นสีเขียว และช่วงคลื่นสีเขียวกำหนดให้เป็นสีน้ำเงิน

๒. ขนาด (Size) ขนาดของภาพวัตถุที่ปรากฏในข้อมูลจากดาวเทียมขึ้นอยู่กับขนาดของวัตถุและมาตราส่วนของข้อมูลจากดาวเทียม เช่น ความยาว ความกว้าง หรือพื้นที่ แสดงให้เห็นความแตกต่างของขนาดระหว่างแม่น้ำและลำคลอง

๓. รูปร่าง (Shape) รูปร่างของวัตถุที่เป็นเฉพาะตัวอาจสม่ำเสมอ (Regular) หรือไม่สม่ำเสมอ(Irregular) วัตถุที่มนุษย์สร้างขึ้นมีรูปร่างส่วนใหญ่เป็นรูปทรงเรขาคณิต เช่น สนามบิน พื้นที่นาข้าว ถนนคลองชลประทาน และเขื่อนเก็บกักน้ำ เป็นต้น

๔. เนื้อภาพ (Texture) หรือความหยาบละเอียดของผิววัตถุ เป็นผลมาจากความแปรปรวนหรือความสม่ำเสมอของวัตถุ เช่น น้ำมีลักษณะเรียบ และป่าไม้มีลักษณะขรุขระ เป็นต้น

๕. รูปแบบ (Pattern) ลักษณะการจัดเรียงตัวของวัตถุ ปรากฏเด่นชัดระหว่างความแตกต่างตามธรรมชาติและสิ่งที่มนุษย์สร้างขึ้น เช่น แม่น้ำ คลอง กับคลองชลประทาน บ่อ และสระน้ำกับเขื่อน เป็นต้น

๖. ความสูงและเงา (Height and Shadow) เงาของวัตถุมีความสำคัญในการคำนวณหาความสูงและมุมสูงของดวงอาทิตย์ เช่น เงาบริเวณเขาหรือหน้าผา เงาของเมฆ เป็นต้น

๗. ที่ตั้ง (Site) หรือตำแหน่งของวัตถุที่พบตามธรรมชาติ เช่น พื้นที่ป่าชายเลนพบบริเวณชายฝั่งทะเลน้ำท่วมถึง สนามบินอยู่ใกล้แหล่งชุมชน เป็นต้น

๘. ความเกี่ยวพัน (Association) หมายถึง ความเกี่ยวพันขององค์ประกอบทั้ง ๗ ที่กล่าวมา เช่น บริเวณที่มีต้นไม้เป็นกลุ่ม ๆ มักเป็นที่ตั้งของหมู่บ้าน ไร่เลื่อนลอยอยู่ในพื้นที่ป่าไม้บนเขา นากุ้งอยู่บริเวณชายฝั่งรวมกับป่าชายเลน เป็นต้น

การแปลตีความภาพเพื่อจำแนกวัตถุได้ดีและถูกต้อง ขึ้นอยู่กับองค์ประกอบต่าง ๆ ดังกล่าวข้างต้นอย่างใดอย่างหนึ่งหรือหลายอย่างพร้อม ๆ กันไป ตามความยากง่ายและมาตราส่วนที่แตกต่างกัน ซึ่งอาจไม่แน่นอนเสมอไปรูปร่าง สี และขนาด อาจใช้เป็นองค์ประกอบในการแปลตีความภาพพื้นที่หนึ่งหรือลักษณะหนึ่ง ส่วนอีกบริเวณอื่นของพื้นที่เดียวกันอาจต้องใช้องค์ประกอบอีกอย่างก็ได้ นอกจากนี้จำเป็นต้องนำข้อมูลจากดาวเทียมอีก ๓ ลักษณะมาประกอบการพิจารณา คือ

๑) ลักษณะการสะท้อนช่วงคลื่นแม่เหล็กไฟฟ้าของวัตถุ (Spectral characteristic) ซึ่งสัมพันธ์กับความยาวช่วงคลื่นแสงในแต่ละแบนด์โดยวัตถุต่าง ๆ สะท้อนแสงในแต่ละช่วงคลื่นไม่เท่ากัน ทำให้สีของวัตถุในภาพแต่ละแบนด์แตกต่างกันในระดับสีขาว - ดำ ซึ่งทำให้สีแตกต่างในภาพสีผสมด้วย

๒) ลักษณะรูปร่างของวัตถุที่ปรากฏในภาพ (Spatial characteristic) แตกต่างตามมาตราส่วนและรายละเอียดภาพจากดาวเทียม เช่น MSS วัตถุหรือพื้นที่ขนาด ๘๐×๘๐ เมตร จึงจะปรากฏในภาพ และระบบ PLA มีขนาด ๑๐×๑๐ เมตร เมื่อคุ้นเคยกับลักษณะรูปร่างวัตถุทำให้ทราบลักษณะที่จำลองในภาพจากดาวเทียม

๓) ลักษณะการเปลี่ยนแปลงของวัตถุตามช่วงเวลา (Temporal characteristic) ซึ่งทำให้สถานะของวัตถุต่าง ๆ มีการเปลี่ยนแปลง เช่น การเปลี่ยนแปลงตามช่วงฤดูกาล การเปลี่ยนแปลงรายปีหรือรายคาบ เป็นต้น ลักษณะการเปลี่ยนแปลงดังกล่าว ทำให้มีความแตกต่างของระดับสีในภาพขาวดำ

และภาพสีผสม ทำให้เราสามารถใช้ข้อมูลดาวเทียมที่ถ่ายซ้ำที่เดิมในช่วงเวลาต่าง ๆ มาติดตามการเปลี่ยนแปลงได้ เช่น สามารถติดตามการบุกรุกทำลายป่าการเติบโตของพืชตั้งแต่ปลูกจนถึงการเก็บเกี่ยว เป็นต้น

กระบวนการก่อนการประมวลผลภาพ (Pre-processing)

ขั้นตอนที่ ๑ ขั้นตอนกระบวนการก่อนการประมวลผลภาพ (Pre-processing)

การปรับแก้ภาพ มีวัตถุประสงค์เพื่อปรับแก้ความคลาดเคลื่อนของข้อมูล (Data error) สัญญาณที่รบกวน (Noise) และความบิดเบี้ยวเชิงเรขาคณิตที่เกิดขึ้นในระหว่างกระบวนการถ่ายภาพ การบันทึกข้อมูล สัญญาณ การสะท้อนคลื่นแม่เหล็กไฟฟ้า การส่งสัญญาณ และการโคจรของดาวเทียม มีกระบวนการปรับแก้ ๒ กระบวนการที่ต้องกระทำ ดังนี้

๑) การตรวจแก้คลื่นรังสี (Radiometric correction)

ก่อนที่จะส่งข้อมูลจากระยะไกลให้กับผู้ใช้งาน ข้อมูลเหล่านี้จะต้องผ่านการตรวจแก้คลื่นรังสีจากสถานีรับสัญญาณดาวเทียมภาคพื้นดินมาแล้วระดับหนึ่ง แต่บางครั้งยังปรากฏข้อบกพร่องทางเชิงคลื่นเนื่องจากสาเหตุหลายประการ เช่น การรบกวนจากชั้นบรรยากาศ หรือจากความบกพร่องของเครื่องรับสัญญาณ ทำให้เกิดความไม่ชัดเจนพร่ามัว มีลายเส้นปะปน (Strip/Noise) ปรากฏบนข้อมูลจากดาวเทียม ปัญหาอีกด้านหนึ่งที่จำเป็นต้องมีการตรวจแก้คลื่นรังสี คือ เมื่อต้องการใช้ข้อมูลหลายช่วงเวลาเพื่อการศึกษาการเปลี่ยนแปลงของปรากฏการณ์ใดปรากฏการณ์หนึ่งต้องทำการปรับแก้ค่ามุมยกของดวงอาทิตย์ (Sun elevation correction) ที่เปลี่ยนแปลงไปตามแต่ละช่วงเวลาและแต่ละฤดูกาล การแก้ไขปัญหาดังกล่าวทำได้โดยการตรวจแก้คลื่นรังสี พร้อมทั้งต้องมีรายละเอียดตัวแปร (Parameters) เกี่ยวกับการรับสัญญาณ มุมที่ดาวเทียมกระทำกับแสงดวงอาทิตย์ (Solar illumination angles) ค่ารังสีตกกระทบ(Irradiance) การกระจายแสงในเส้นทางผ่าน (Path radiance) ค่าการสะท้อนของวัตถุเป้าหมาย (Reflectance of target) ค่าการส่งผ่านของบรรยากาศ เป็นต้น และข้อมูลสภาวะอากาศในขณะที่ทำการบันทึกข้อมูล การปรับแก้มีกรรมวิธีในการคำนวณที่ซับซ้อนมาก โดยต้องใช้ซอฟต์แวร์ที่มีโปรแกรมเฉพาะสำหรับการตรวจแก้คลื่นรังสี

๒) การตรวจแก้เชิงเรขาคณิต (Geometric correction)

ก่อนนำข้อมูลจากดาวเทียมไปใช้ประโยชน์มีความจำเป็นต้องปรับแก้เชิงเรขาคณิต เพราะพิกัดตำแหน่งของวัตถุต่าง ๆ มีความคลาดเคลื่อนไปจากความเป็นจริง เนื่องจากข้อบกพร่องทางเครื่องรับสัญญาณและรูปลักษณะของวัตถุ การตรวจแก้เชิงเรขาคณิตมีความจำเป็นมากยิ่งขึ้น เมื่อต้องการนำข้อมูลจากระยะไกลไปใช้งานร่วมกับข้อมูลเชิงแผนที่อื่น ๆ เพื่อให้สามารถซ้อนทับกันได้ หรือหากต้องการศึกษาปรากฏการณ์อย่างหนึ่งในหลายช่วงเวลาการเปรียบเทียบข้อมูลแต่ละช่วงเวลาต้องมีระบบพิกัดเดียวกัน จึงจะสามารถซ้อนข้อมูลแต่ละเวลาลงกันได้สนิทจึงจำเป็นต้องทำการปรับแก้เชิงเรขาคณิต

ขั้นตอนที่ ๒ การเน้นข้อมูลภาพ (Image enhancement)

เป็นกระบวนการปรับเปลี่ยนค่าของจุดภาพหรือค่าระดับสีเทา เพื่อเพิ่มรายละเอียด ความชัดเจนของข้อมูลภาพ หรือการเพิ่มระดับความแตกต่างระหว่างวัตถุ ทำให้สังเกตเห็นขอบเขตของวัตถุต่างชนิดกันได้ชัดเจนมากขึ้น หรือสามารถที่จะเน้นความคมชัดเฉพาะในส่วนที่ต้องการศึกษาช่วยให้การตีความประเภทวัตถุง่ายมากขึ้นแล้วนำผลลัพธ์ที่ได้ไปทำการแปลตีความ โดยทำการศึกษาแปลตีความด้วยสายตา (Visual interpretation) เพื่อกำหนดประเภทข้อมูล ก่อนที่นำไปใช้เพื่อการจำแนกประเภทข้อมูล เทคนิคต่าง ๆ ที่จะใช้ในการเน้นข้อมูลภาพนั้นอยู่กับ

- ค่าจุดภาพ ซึ่งประกอบด้วยข้อมูลจากหลายช่วงคลื่นของข้อมูลจากดาวเทียมที่ออกแบบมาเฉพาะผู้ใช้จะต้องทำความเข้าใจรูปแบบของปฏิสัมพันธ์ระหว่างพลังงานในแต่ละช่วงคลื่นกับวัตถุบนพื้นโลก
- วัตถุประสงค์ของการเน้นข้อมูล คือ ต้องเปลี่ยนแปลงข้อมูลเพื่อให้เห็นรายละเอียดต่าง ๆ ในเรื่องที่ต้องการศึกษาได้ดียิ่งขึ้น

- ผลที่คาดว่าจะได้รับการเน้นข้อมูลภาพ

- พื้นฐานของผู้วิเคราะห์ซึ่งจะต้องมีประสบการณ์ในการวิเคราะห์และเทคนิคการเน้นข้อมูลระดับความแตกต่างระหว่างวัตถุในข้อมูลภาพ ซึ่งวัดได้จากสัดส่วนของระดับความแตกต่างของความชัดเจน (Contrast ratio) โดยนำค่าความสว่างสูงที่สุดมาหารด้วยค่าการสะท้อนต่ำที่สุด

ถ้ามีสัดส่วนความแตกต่างของความชัดเจนสูง หมายความว่า ภาพมีความชัดเจนมาก ขอบเขตของวัตถุต่าง ๆ จะสามารถแยกแยะได้ชัดเจน และถ้ามีสัดส่วนของความแตกต่างของความชัดเจนต่ำ หมายความว่า ภาพลดความคมชัดลง ขอบเขตของแต่ละประเภทข้อมูลไม่ชัดเจน

ค่าความสว่างหรือค่าระดับสีเทาในแต่ละช่วงคลื่น สามารถนำมาปรับปรุงให้ค่าระดับสีเทามีความแตกต่างมากขึ้น แล้วทำให้ได้รับรายละเอียดของข้อมูลภาพมากขึ้น เทคนิคการเน้นรายละเอียดภาพมีหลายวิธีสามารถเน้นภาพให้มีลักษณะรายละเอียดได้แตกต่างกัน ดังนั้นผู้ใช้งานจึงต้องมีความรู้เรื่องวิธีการเน้นรายละเอียดภาพอย่างเพียงพอ

ขั้นตอนที่ ๓ การจำแนกประเภทข้อมูลภาพ (Image classification)

การจำแนกประเภทข้อมูลภาพเป็นการประมวลผลในทางสถิติ เพื่อแยกข้อมูลจุดภาพทั้งหมดที่ประกอบเป็นพื้นที่ศึกษาออกเป็นกลุ่มย่อย โดยใช้ลักษณะทางสถิติเป็นตัวกำหนดความแตกต่างระหว่างกลุ่มจุดภาพ โดยจุดภาพที่ถูกจัดให้อยู่กลุ่มเดียวกันจะมีลักษณะทางสถิติเฉพาะกลุ่มเป็นไปในทิศทางเดียวกันแต่ละกลุ่มจุดภาพที่จำแนกได้นั้น จะแสดงถึงสิ่งปกคลุมพื้นดินประเภทใดประเภทหนึ่งแตกต่างกันไป กล่าวอีกนัยหนึ่ง การจำแนกประเภทข้อมูลภาพ หมายถึง การแบ่งจุดภาพที่มีคุณสมบัติการสะท้อนแสงคล้าย ๆ กันออกเป็นกลุ่มหรือเป็นระดับ ซึ่งเรียกว่า ชนิดหรือประเภท (Class) เพื่อที่จะแบ่งแยกวัตถุต่าง ๆ ที่แสดงในภาพออกจากกัน ในการจำแนกประเภทข้อมูลนี้ผู้ปฏิบัติต้องใช้กฎการตัดสินใจหรือความรู้ทางสถิติเข้าช่วยเนื่องจากปริมาณจุดภาพที่ประกอบเป็นพื้นที่ศึกษา มีปริมาณจุดภาพมากการคำนวณทางสถิติเองโดยใช้เครื่องคิดเลขจึงทำได้ยากใช้เวลามากและอาจเกิดข้อผิดพลาดได้ จึงมีการนำเอาความสามารถของคอมพิวเตอร์มาช่วยในการประมวลผล ทำให้ได้ผลลัพธ์ในเวลารวดเร็วสามารถตรวจสอบความถูกต้องได้ทันที

การจำแนกประเภทข้อมูลภาพด้วยระบบคอมพิวเตอร์แบ่งออกได้เป็น ๒ วิธี คือ การจำแนกประเภทข้อมูลภาพแบบควบคุม (Supervised classification) และการจำแนกประเภทข้อมูลภาพแบบไม่ควบคุม (Unsupervised classification) การจำแนกประเภทข้อมูลภาพทั้งสองวิธีให้ได้ผลลัพธ์ที่มีประสิทธิภาพนั้น ก่อนเริ่มจำแนกประเภทข้อมูลภาพ ควรศึกษาสถิติของข้อมูลภาพในแต่ละช่วงคลื่นเสียก่อน ทั้งนี้เพื่อให้ได้ช่วงคลื่นที่เหมาะสมในการใช้จำแนกประเภทข้อมูลภาพ ค่าสถิติเบื้องต้นที่ใช้ในการเลือกช่วงคลื่นที่เหมาะสม ได้แก่

- ค่าการสะท้อนต่ำสุด - สูงสุด (Minimum - Maximum value) ของแต่ละช่วงคลื่น เป็นค่าที่แสดงการสะท้อนของข้อมูลภาพในแต่ละช่วงคลื่นว่ามีค่าการสะท้อนตกอยู่ในช่วงใด ระหว่างค่า ๐ - ๒๕๕ หากเป็นค่าที่อยู่ใกล้ ๐ มาก หมายถึง ช่วงคลื่นนั้นจะให้ข้อมูลเกี่ยวกับวัตถุที่มีการดูดกลืนพลังงานมาก หากค่าค่อนไปทาง ๒๕๕ หมายถึง ช่วงคลื่นนั้นจะให้ข้อมูลเกี่ยวกับวัตถุที่มีการสะท้อนพลังงานสูง และหากมีช่วงค่ากว้าง คือ มีทั้งค่าต่ำสุดใกล้ ๐ และมีค่าสูงสุดใกล้ ๒๕๕ หมายถึง มีข้อมูลของกลุ่มวัตถุทั้งที่ดูดกลืนพลังงานและที่สะท้อนพลังงาน ถือเป็นช่วงคลื่นที่มีความหลากหลายของข้อมูลมาก

- ค่าเฉลี่ยเลขคณิต (Mean) เป็นค่าเฉลี่ยค่าการสะท้อนทั้งหมดของแต่ละช่วงคลื่น สามารถใช้เป็นตัวแทนภาพรวมข้อมูลจุดภาพทั้งหมดของช่วงคลื่นหนึ่ง ๆ ได้ ค่าเฉลี่ยเลขคณิตสามารถคำนวณโดยนำผลบวกของค่าการสะท้อนทั้งหมดมาหารด้วยจำนวนจุดภาพทั้งหมด

- ค่าเฉลี่ยเลขคณิตนิยมใช้ในการวัดค่าเฉลี่ยค่าการสะท้อนของจุดภาพมากที่สุด โดยจะเป็นค่าเฉลี่ยที่ดีนั้นต่อเมื่อข้อมูลจุดภาพทั้งหมดมีการแจกแจงค่าการสะท้อนในลักษณะสมมาตรหรือการแจกแจงของคะแนนไม่มีความเบ้ (Skewness) ซึ่งสามารถตรวจสอบลักษณะการแจกแจงของคะแนนได้จากการนำเอาค่าการสะท้อนของทุกจุดภาพมาสร้างเป็นแผนภูมิภาพของข้อมูลภาพ

- ค่ากึ่งกลาง (Median) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีหนึ่งที่ใช้การเรียงค่าการสะท้อนของจุดภาพจากค่าน้อยที่สุดไปหาค่ามากที่สุด โดยค่ากึ่งกลางเป็นค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งชุด ค่ากึ่งกลางจึงเป็นตัวแทนค่าการสะท้อนของจำนวนจุดภาพทั้งหมดในช่วงคลื่นหนึ่ง ๆ ที่แสดงให้ทราบว่ามีจำนวนจุดภาพที่มีค่าการสะท้อนมากกว่าและน้อยกว่าค่ากึ่งกลางอยู่ประมาณร้อยละ ๕๐

- ค่าฐานนิยม (Mode) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกวิธีหนึ่ง โดยดูจากจำนวนความถี่ของค่าการสะท้อนซึ่งมีความถี่สูงที่สุด นิยมนำมาใช้กับข้อมูลที่เป็นนามบัญญัติ เช่น ค่าของประเภทข้อมูลหลังจากการจำแนกประเภทแล้ว ถือเป็นค่าการสะท้อนที่แสดงการใช้ที่ดินประเภทต่างๆ ไม่ใช่ค่าการสะท้อนของวัตถุอีกต่อไป

- ความเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) เป็นการวัดการกระจายที่นิยมใช้มากที่สุด การคำนวณใช้วิธียกกำลังสองของผลต่างระหว่างค่าการสะท้อนของทุกจุดภาพในแต่ละช่วงคลื่นกับค่าเฉลี่ยเลขคณิตของช่วงคลื่นนั้น

- สหสัมพันธ์ (Correlation) เป็นการวัดความสัมพันธ์ระหว่างข้อมูลตั้งแต่ ๒ ชุดขึ้นไป วัดได้จากการคำนวณหาสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) มีขอบเขตตั้งแต่ ๐ ถึง + ๑.๐๐ เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อมูลภาพ ๒ ช่วงคลื่น เข้าใกล้ ๑.๐๐ หมายถึง ข้อมูลทั้ง ๒ ชุดนั้นมีความสัมพันธ์ต่อกันในระดับสูงซึ่งอาจจะมีความสัมพันธ์โดยตรง (ค่าสัมประสิทธิ์สหสัมพันธ์เป็นค่าลบ) และเมื่อสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ ๐ หมายถึง ข้อมูลทั้ง ๒ ช่วงคลื่นมีความสัมพันธ์กันในระดับต่ำหรือแตกต่างกันเป็นประโยชน์ในการเลือกช่วงคลื่นในการจำแนกประเภท

ในการศึกษาความสัมพันธ์ระหว่างช่วงคลื่น ๒ ประเภท ทำได้โดยให้แสดงการกระจายของค่าการสะท้อนของจุดภาพลงบนกราฟแบบ ๒ แกน โดยแกนที่ ๑ เป็นค่าการสะท้อนของช่วงคลื่นที่ ๑ และแกนที่ ๒ เป็นค่าการสะท้อนของช่วงคลื่นที่ ๒ หรือเรียกว่า การทำตารางไขว้ (Cross tabulation) รูปแบบการกระจายของค่าการสะท้อนจะบอกถึงลักษณะความสัมพันธ์ของข้อมูลในสองช่วงคลื่นได้เช่นเดียวกับค่าสัมประสิทธิ์สหสัมพันธ์

๑.๕ การจำแนกข้อมูลแบบกำกับดูแลและไม่กำกับดูแล (Supervise and Unsupervised classification)

การจำแนกข้อมูลแบบไม่กำกับดูแล (Unsupervised Classification) เป็นการจำแนกข้อมูลจากการหาความสว่างของจุดภาพที่ใกล้เคียงกันซึ่งไม่ต้องใช้ความรู้ในการจำรูปแบบเลยทำให้การแบ่งแยกรายละเอียดเป็นกลุ่มๆ (Cluster) จึงขึ้นอยู่กับค่าความสว่างของภาพ ซึ่งแบ่งออกเป็น ๒ แบบ คือ ๑) การรวมกลุ่มแบบลำดับชั้น (Hierarchical clustering) วิธีนี้จุดภาพจะถูกจัดรวมเป็นกลุ่มที่คล้ายกันโดยใช้ระยะห่างเป็นเครื่องวัด เริ่มต้นด้วยการสมมติว่าแต่ละจุดภาพเป็น ๑ กลุ่ม จุดภาพที่มีระยะห่างกันน้อยที่สุดก็จะรวมตัวเข้าด้วยกัน ถัดจากนั้นจะเป็นการรวมกลุ่มจุดภาพไปเรื่อย ๆ จนกระทั่งได้กลุ่มตามจำนวนที่กำหนดไว้จึงหยุด และ ๒) การรวมกลุ่มแบบไม่เป็นลำดับชั้น (Non-hierarchical clustering) เริ่มต้นด้วยการแบ่งข้อมูลออกเป็นกลุ่มชั่วคราวจำนวนหนึ่ง หลังจากนั้นสมาชิกในแต่ละกลุ่มจะถูกตรวจสอบโดยใช้ตัวแปรหรือระยะห่างที่เลือกมาเพื่อทำการจัดตำแหน่งใหม่ให้อยู่ในกลุ่มที่เหมาะสมกว่าโดยมีการแบ่งกลุ่มชัดเจนดีขึ้น ตัวอย่างของการรวมกลุ่มวิธีนี้ได้แก่ วิธี ISODATA และวิธี K-mean

การจำแนกข้อมูลแบบกำกับดูแล (Supervised Classification) เป็นการจำแนกรายละเอียดที่ต้องการรู้คุณสมบัติของวัตถุว่าเป็นเช่นใดเพื่อนำมาใช้เลือกกลุ่มข้อมูลตัวอย่างที่เป็นค่าความสว่างของจุดภาพซึ่งเป็นตัวแทนของรายละเอียดนั้นในการจำแนกข้อมูล โดยจุดภาพที่เป็นตัวแทนของกลุ่มตัวอย่างนั้นเรียกว่า กลุ่มตัวอย่าง (Training Fields, Training Areas หรือ Spectral Signatures) ดังนั้นจะเห็นได้ว่าวิธีการจำแนกข้อมูลแบบกำกับดูแลจะเป็นการกำหนดกลุ่มตัวอย่างที่เป็นตัวแทนของรายละเอียดก่อนการจำแนก

๑.๖ การวิเคราะห์ศักยภาพพื้นที่ (Potential Surface Analysis: PSA)

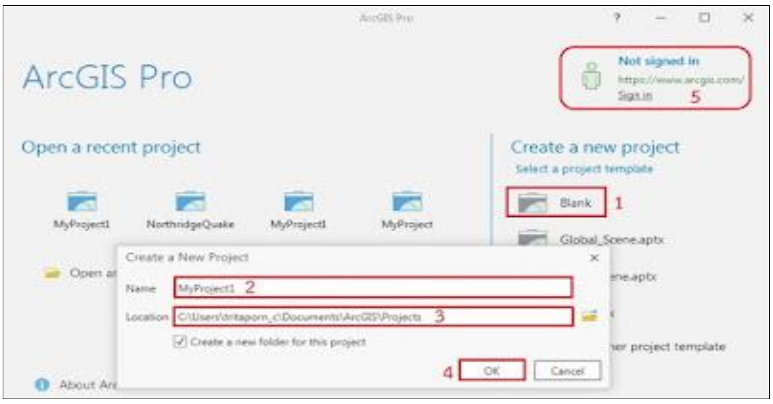
การวิเคราะห์ศักยภาพพื้นที่ (Potential Surface Analysis: PSA) เป็นกระบวนการวิเคราะห์ข้อมูลทางกายภาพของพื้นที่ศึกษา เพื่อพิจารณาว่าบริเวณใดของพื้นที่ที่มีศักยภาพในการพัฒนาหรือมีข้อจำกัดที่ส่งผลต่อการพัฒนาเชิงกายภาพ ใช้ประเมินและวิเคราะห์ทำเลที่ตั้งที่เหมาะสมสำหรับการพัฒนาที่อยู่อาศัย พาณิชยกรรม และแหล่งท่องเที่ยว ทั้งนี้การวิเคราะห์ความเหมาะสมเชิงพื้นที่มีความสำคัญอยู่ที่การให้ค่าน้ำหนักแก่กลุ่มปัจจัยต่าง ๆ ซึ่งเป็นการจัดลำดับความเหมาะสมเพื่อรองรับการพัฒนาพื้นที่ การวิเคราะห์ศักยภาพเชิงพื้นที่ได้พัฒนามาจากหลักการของเทคนิคการวิเคราะห์แบบซ้อนทับแผนที่ (overlay mapping) ซึ่งต้องใช้ข้อมูลเชิงพื้นที่ (spatial data) และข้อมูลรายละเอียด (attribute data) จากระบบสารสนเทศทางภูมิศาสตร์ (geographic information system) มาวิเคราะห์และสังเคราะห์ร่วมกัน

๑.๗ ตัวอย่างการใช้งานโปรแกรม

การสร้าง Project

เมื่อเปิดโปรแกรม ArcGIS Pro จะปรากฏหน้าต่างของโปรแกรม โดยมีขั้นตอนการสร้าง Project ใหม่ดังภาพที่ ๑ ดังนี้

๑) คลิกเลือก Blank เพื่อเลือก Project เปล่า ที่ยังไม่ได้ทำอะไร
๒) จะปรากฏหน้าต่าง Create a New Project กำหนดชื่อ Project
๓) กำหนดสถานที่จัดเก็บ Project
๔) คลิก OK เพื่อไปยังขั้นตอนถัดไป
๕) แสดงถึงการใช้งานแบบ Authorize Use ซึ่งไม่จำเป็นต้อง Sign In เข้าโปรแกรม



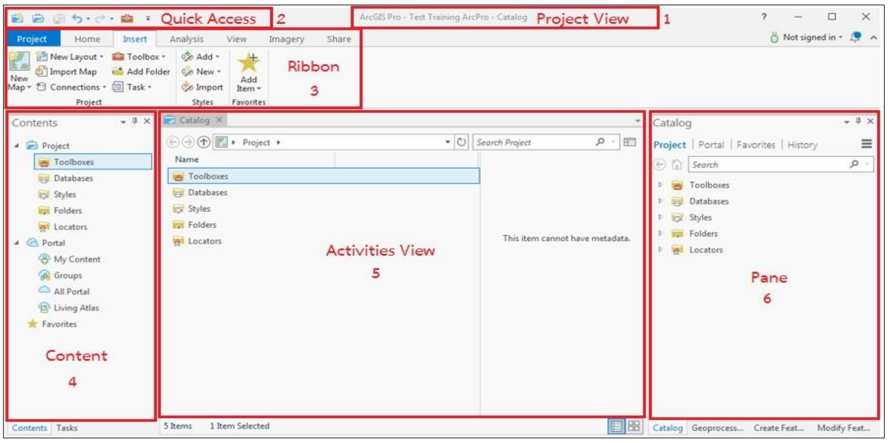
ภาพที่ ๑

จากนั้นโปรแกรมจะเข้าสู่ Project ที่ได้สร้างขึ้น ดังภาพที่ ๒ ดังนี้

๑) ส่วนแสดงชื่อ Project และแหล่งที่จัดเก็บ รวมถึงมุมมองของกิจกรรมที่กำลังทำ
๒) แถบเครื่องมือลัด โดยสามารถปรับแต่งหรือเพิ่มเครื่องมือที่ใช้เป็นประจำได้
๓) ส่วนเครื่องมือต่าง ๆ โดยจะเรียกส่วนนี้รวม ๆ ว่า Ribbon
๔) ส่วนรายละเอียดและมุมมอง Content ทั้งหมดที่ใช้งานอยู่

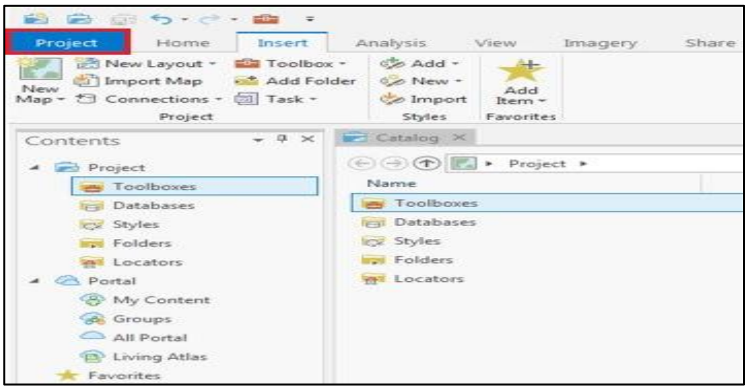
๕) ส่วนมุมมองกิจกรรมที่กำลังทำงาน ซึ่งจะเปลี่ยนไปตามแถบมุมมองด้านบน

๖) ส่วนหน้าต่างที่ใช้สำหรับการจัดการ ซึ่งจะเปลี่ยนไปสิ่งที่จัดการตามแถบจัดการด้านล่าง



ภาพที่ ๒

เนื่องจาก ArcGIS Pro ถูก Set Default ไว้ให้ดึง Base map จาก Web Map Service ของ ESRI เข้ามาเสมอ ซึ่งจะทำให้ใช้เวลาเปิดค่อนข้างนาน ถ้าหากเครื่องไม่ค่อยแรง แนะนำให้ไปปลด Base map ออกเสียก่อน โดยการเลือกคลิกที่ แถบ Ribbon Project ดังภาพที่ ๓



ภาพที่ ๓

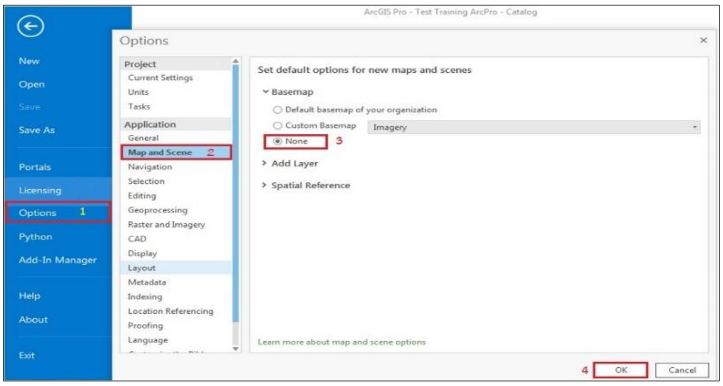
จากนั้นจะปรากฏหน้ามุมมอง Project ดังภาพที่ ๔ ดังนี้

๑) เลือก Option

๒) เลือก Map and Scene

๓) ใน Set default option for new maps and scenes ส่วน Basemap เลือก None

๔) คลิก OK เพื่อบันทึกการตั้งค่าเริ่มต้น



ภาพที่ ๔

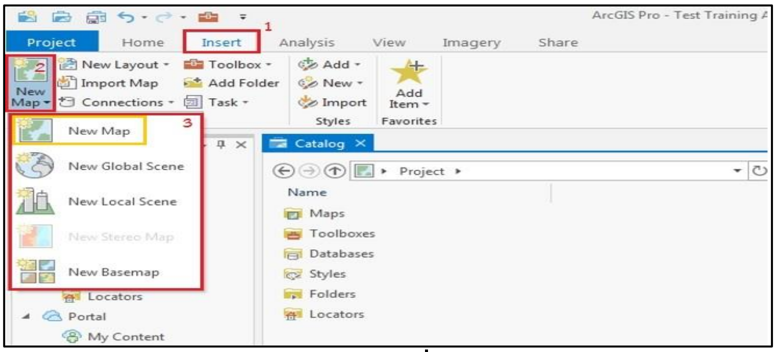
จากนั้นคลิกที่ ลูกศรด้านบนสุด เพื่อกลับไปยังหน้ากิจกรรม จะทำการเรียกข้อมูลแผนที่ ดังภาพที่ ๕ ดังนี้

๑) คลิกแถบ Ribbon Insert

๒) คลิกเครื่องมือ New Map ใน Ribbon

๓) เลือกรูปแบบ Map ที่ต้องการทำงาน ซึ่งจะมีให้เลือกอยู่ ๔ แบบ ได้แก่

- New Map สำหรับแผนที่เปล่าแบบกำหนดเอง
- New Global Scense สำหรับแผนที่ในระดับใหญ่ เช่น ระดับประเทศ ภูมิภาค เป็นต้น
- New Local Map สำหรับแผนที่ระดับเล็ก ระดับพื้นที่อาคาร ถึงระดับจังหวัด เป็นต้น
- New Basemap เป็นการเรียก Basemap จาก Web Map Service ของ ESRI มาใช้งาน



ภาพที่ ๕

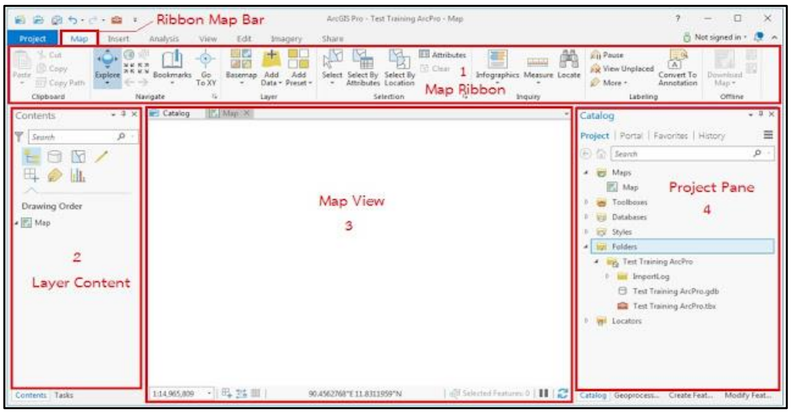
เลือก New Map เพื่อกำหนดเอง เมื่อคลิกแล้วจะได้ตามภาพที่ ๖ ดังนี้

๑) สังเกตว่า เครื่องมือในส่วน Ribbon จะเปลี่ยนไปตาม แถบ Ribbon ที่เลือก ในที่นี้เลือกแถบ Map

๒) ใน contents ก็จะเปลี่ยนไป จะ Focus ในระดับ Layer แทน

๓) ใน Activities View จะเปลี่ยนเป็น Map View เพื่อจัดการข้อมูลแผนที่ คล้าย ArcMap

๔) ใน Pane ก็จะเป็นมุมมองรายละเอียดของการจัดการกับสิ่งที่ทำ



ภาพที่ ๖

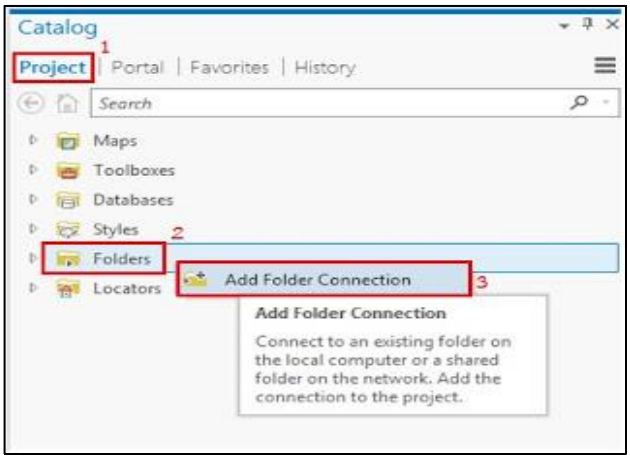
การเรียกชั้นข้อมูลหรือสร้างชั้นข้อมูล

การเรียกชั้นข้อมูลหรือสร้างชั้นข้อมูลเข้ามาใน Project มีขั้นตอนตามภาพที่ ๗ ดังนี้

๑) คลิกแถบ Project ใน Catalog Pane ด้านขวามือ

๒) คลิกขวาที่กล่อง Folders

๓) เลือก Add Folder Connection เพื่อนำเข้าชั้นข้อมูลที่มีอยู่แล้ว



ภาพที่ ๗

จากนั้นนำเข้าชั้นข้อมูลที่อยู่ใน DataBase เข้าไปยัง Map View เพื่อจัดการข้อมูล ซึ่งมีวิธีการนำเข้า ๓ วิธี ดังนี้

๑) คลิกเลือกที่เครื่องมือ Add Data (จะคล้ายกับ ArcMap) ระบบจะนำไปยังแหล่งข้อมูล (Folder) ที่ได้นำเข้าไว้แล้วโดยอัตโนมัติ

๒) คลิกขวาที่ชั้นข้อมูลที่ต้องการ ใน Catalog Pane แล้วเลือก Add To Current Map

๓) คลิกเลือกชั้นข้อมูลที่ต้องการ ใน Catalog Pane ค้างไว้ แล้วจับลากเข้าไปยัง Map View

จากทั้ง ๓ วิธีจะทำให้ได้ชั้นข้อมูลที่ต้องการแสดงบนหน้า Map View เพื่อให้สามารถจัดการกับชั้นข้อมูลนั้นได้

การจัดการกับข้อมูลตาราง (Field) ของชั้นข้อมูลที่นำเข้า

การเพิ่ม field ในตารางที่มีการสร้างชั้นข้อมูลไว้แล้ว มีขั้นตอน ดังภาพที่ ๘ - ๑๐ ดังนี้

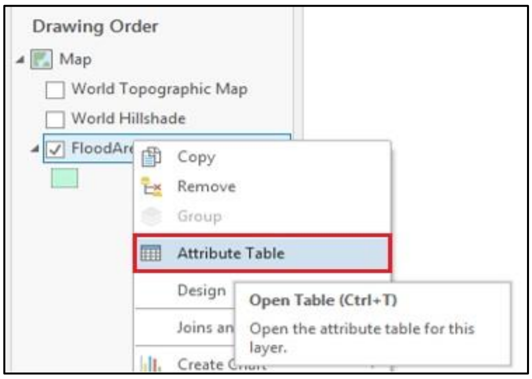
๑) คลิกขวา เลือก Attribute Table ของชั้นข้อมูลที่จะจัดการ

๒) ปรากฏหน้าต่างของตาราง FloodArea ที่มี field เดิมจากที่ได้ตั้งไว้

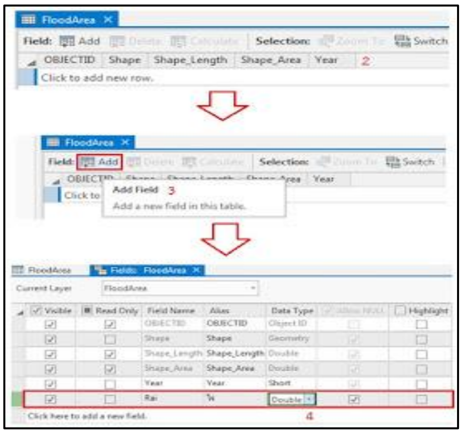
๓) คลิกเลือก Add field เพื่อเพิ่มตาราง

๔) กำหนดค่าคุณสมบัติของ field เช่น เพิ่มฟิลด์ Rai และ Data type เป็น Double

๕) เมื่อเพิ่ม field เสร็จแล้ว ที่ Tab Field ให้กดปุ่ม Save เพื่อบันทึกการเพิ่ม field



ภาพที่ ๘



ภาพที่ ๙

ภาพที่ ๑๐

การนำเข้าข้อมูลจากข้อมูลตารางที่มีค่าพิกัด

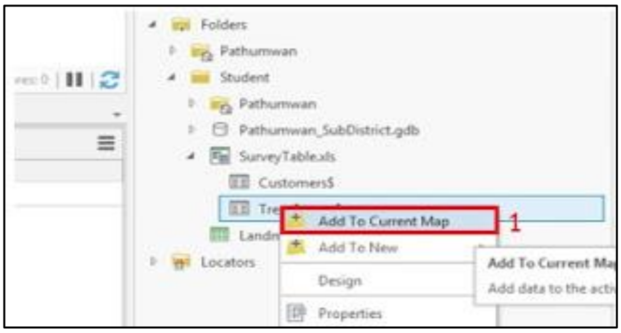
การนำเข้าข้อมูลจากข้อมูลตารางที่มีค่าพิกัด มีขั้นตอน ดังภาพที่ ๑๑ – ๑๓ ดังนี้

๑) นำเข้าข้อมูลตารางที่มีค่าพิกัด ตัวอย่าง เลือก TreeSurvey ให้เข้ามาอยู่ในหน้า Map View โดยเลือก Add to Current Map หรือ ลากเข้าไปยัง Map View โดยตรง

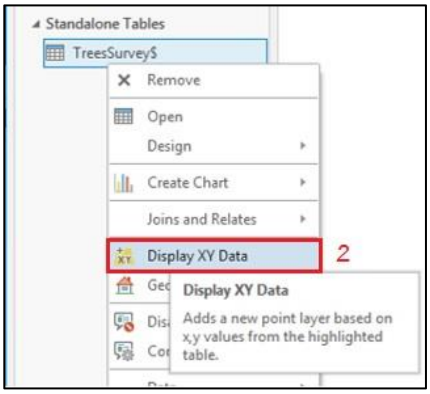
๒) คลิกขวาที่ TreesSurvey เลือก Display XY Data

๓) จากนั้น กำหนดค่าคุณสมบัติต่างๆ

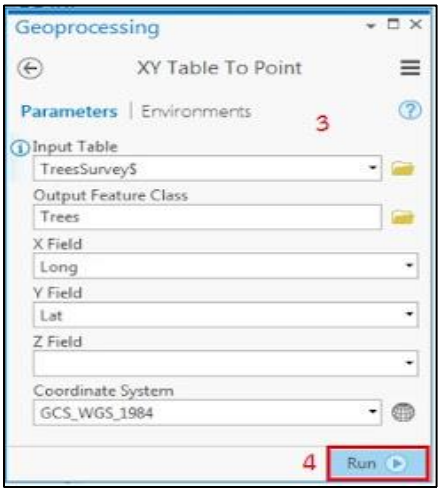
๔) กด Run เพื่อเพิ่มชั้นข้อมูลใหม่เข้ามาอยู่ใน Map View



ภาพที่ ๑๑



ภาพที่ ๑๒



ภาพที่ ๑๓

๒. การประเมินความพึงพอใจการเข้าร่วมอบรม

๒.๑ คุณสมบัติของผู้ตอบแบบสำรวจความพึงพอใจการเข้าร่วมอบรม (ตารางที่ ๑)

ผู้ตอบแบบสำรวจความพึงพอใจการเข้าร่วมอบรม ตอบแบบสำรวจผ่าน Google Form จำนวน ๒๙ คน ซึ่งประกอบด้วย

๒.๑.๑ เพศ ผู้ใช้บริการส่วนใหญ่เป็นเพศหญิง ร้อยละ ๕๑.๗๒ และเพศชาย ร้อยละ ๔๘.๒๘

๒.๑.๒ อายุ ส่วนใหญ่อยู่ในช่วง ๓๑ – ๔๐ ปี ร้อยละ ๔๔.๘๓ รองลงมาอายุอยู่ในช่วง ๒๐ – ๓๐ ปี ร้อยละ ๓๗.๙๓ และอายุอยู่ในช่วง ๔๑ – ๕๐ ปี ร้อยละ ๑๗.๒๔

๒.๑.๓ ระดับการศึกษา ส่วนใหญ่จบการศึกษาระดับปริญญาโท ร้อยละ ๕๕.๑๗ รองลงมาระดับปริญญาตรี ร้อยละ ๔๔.๘๓

๒.๑.๔ ตำแหน่ง ส่วนใหญ่เป็นเศรษฐกรปฏิบัติการ ร้อยละ ๒๐.๖๙ รองลงมาเป็นนักวิเคราะห์นโยบายและแผนปฏิบัติการ ร้อยละ ๑๗.๒๔ เศรษฐกรชำนาญการ ร้อยละ ๑๐.๓๔ นักวิเคราะห์นโยบายและแผนชำนาญการ นักวิชาการคอมพิวเตอร์ปฏิบัติการ เศรษฐกร เจ้าพนักงานสถิติปฏิบัติงาน และนักวิเคราะห์นโยบายและแผน ร้อยละ ๖.๙๐ นักวิชาการสถิติปฏิบัติการ นักวิชาการสถิติชำนาญการพิเศษ นักวิชาการคอมพิวเตอร์ เศรษฐกรชำนาญการพิเศษ และเจ้าพนักงานสถิติชำนาญงาน ร้อยละ ๓.๔๕

๒.๑.๕ หน่วยงาน ส่วนใหญ่อยู่ภายใต้สังกัดศูนย์ข้อมูลเกษตรแห่งชาติ ร้อยละ ๒๗.๕๙ รองลงมาเป็นศูนย์สารสนเทศการเกษตร ร้อยละ ๒๔.๑๔ สำนักวิจัยเศรษฐกิจการเกษตร ร้อยละ ๑๗.๒๔ กองเศรษฐกิจการเกษตรระหว่างประเทศ กองนโยบายและแผนพัฒนาการเกษตร และศูนย์ประเมินผล ร้อยละ ๑๐.๓๔

ตารางที่ ๑ คุณสมบัติของผู้ตอบแบบสำรวจความพึงพอใจการเข้าร่วมอบรม

รายละเอียด	ร้อยละ
เพศ	
- ชาย	๔๘.๒๘
- หญิง	๕๑.๗๒
อายุ	
- ๒๐ – ๓๐ ปี	๓๗.๙๓
- ๓๑ – ๔๐ ปี	๔๔.๘๓
- ๔๑ – ๕๐ ปี	๑๗.๒๔
- ๕๑ – ๖๐ ปี	๐.๐๐
รายละเอียด	**ร้อยละ**
ระดับการศึกษา	
- ต่ำกว่าปริญญาตรี	๐.๐๐
- ปริญญาตรี	๔๔.๘๓
- ปริญญาโท	๕๕.๑๗
- ปริญญาเอก	๐.๐๐
ตำแหน่ง	
- เศรษฐกรปฏิบัติการ	๒๐.๖๘
- นักวิเคราะห์นโยบายและแผนปฏิบัติการ	๑๗.๒๓
- เศรษฐกรชำนาญการ	๑๐.๓๔
- นักวิเคราะห์นโยบายและแผนชำนาญการ	๖.๙๐
- นักวิชาการคอมพิวเตอร์ปฏิบัติการ	๖.๙๐
- เศรษฐกร	๖.๙๐
- เจ้าพนักงานสถิติปฏิบัติงาน	๖.๙๐
- นักวิเคราะห์นโยบายและแผน	๖.๙๐
- นักวิชาการสถิติปฏิบัติการ	๓.๔๕
- นักวิชาการสถิติชำนาญการพิเศษ	๓.๔๕
- นักวิชาการคอมพิวเตอร์	๓.๔๕
- เศรษฐกรชำนาญการพิเศษ	๓.๔๕
- เจ้าพนักงานสถิติชำนาญงาน	๓.๔๕

หน่วยงาน	
- ศูนย์ข้อมูลเกษตรแห่งชาติ	๒๗.๕๙
- ศูนย์สารสนเทศการเกษตร	๒๔.๑๕
- สำนักวิจัยเศรษฐกิจการเกษตร	๑๗.๒๔
- กองเศรษฐกิจการเกษตรระหว่างประเทศ	๑๐.๓๔
- กองนโยบายและแผนพัฒนาการเกษตร	๑๐.๓๔
- ศูนย์ประเมินผล	๑๐.๓๔

๒.๒ ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรม (ตารางที่ ๒)

การประเมินความพึงพอใจของผู้เข้าร่วมอบรมที่มีต่อปัจจัยที่เกี่ยวข้องกับการอบรม โดยใช้แบบสอบถาม ซึ่งผลจากการประเมินจะใช้เป็นแนวทางในการพัฒนาคุณภาพการจัดสัมมนาให้ดียิ่งขึ้นต่อไป โดยมีเกณฑ์ในการประเมิน ดังนี้

คะแนน	การประเมิน	การแปลผล
๕	๔.๐๑ – ๕.๐๐	มากที่สุด
๔	๓.๐๑ – ๔.๐๐	มาก
๓	๒.๐๑ – ๓.๐๐	ปานกลาง
๒	๑.๐๑ - ๒.๐๐	น้อย
๑	๐.๐๐ – ๑.๐๐	น้อยที่สุด

ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรมฯ ในประเด็นต่าง ๆ อาทิ เช่น ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร เทคนิคหรือวิธีการถ่ายทอดความรู้ ความเหมาะสมของเนื้อหาในการอบรม การสรุปเนื้อหาสาระได้อย่างชัดเจน ความรู้ความเข้าใจเนื้อหา ก่อน-หลัง ความสามารถในการนำความรู้ ความเข้าใจ และประสบการณ์ที่ได้จากการเข้าร่วมการอบรม ไปปรับใช้ในการปฏิบัติงานและถ่ายทอดให้เพื่อนร่วมงาน ระยะเวลาการอบรม การประสานงานและการอำนวยความสะดวก พบว่า ในภาพรวมผู้รับบริการส่วนใหญ่มีความพึงพอใจอยู่ในระดับมาก คะแนนเฉลี่ย ๔.๑๒ ร้อยละ ๘๒.๓๒ โดยมีรายละเอียดตามประเด็นต่าง ๆ ดังนี้

ตารางที่ ๒ ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรม

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๑. ภาพรวมของวิทยากร						**๔.๓๔**
๑.๑ ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร	๕๕.๑๗ (๑๖)	๓๗.๙๓ (๑๑)	๖.๙๐ (๒)	๐.๐๐	๐.๐๐	๔.๔๘
๑.๒ เทคนิคหรือวิธีการถ่ายทอดความรู้	๔๘.๒๘ (๑๔)	๓๗.๙๓ (๑๑)	๑๐.๓๔ (๓)	๓.๔๕ (๑)	๐.๐๐	๔.๓๑
๑.๓ ความเหมาะสมของเนื้อหาในการอบรม	๓๗.๙๓ (๑๑)	๔๘.๒๘ (๑๔)	๑๓.๗๙ (๔)	๐.๐๐	๐.๐๐	๔.๒๔

๑.๔ การสรุปเนื้อหาสาระได้อย่างชัดเจน	๓๗.๙๓ (๑๑)	๕๕.๑๗ (๑๖)	๖.๙๐ (๒)	๐.๐๐	๐.๐๐	๔.๓๑
๒. ความรู้ความเข้าใจเกี่ยวกับเนื้อหาวิชาที่เข้าร่วมอบรม						**๓.๖๑**
๒.๑ ก่อน การเข้าร่วมอบรม ท่านมีความรู้ความเข้าใจเพียงใด	๖.๙๐ (๒)	๑๐.๓๔ (๓)	๑๓.๗๙ (๔)	๒๔.๑๔ (๗)	๔๔.๘๓ (๑๓)	๒.๑๐
๒.๒ หลัง การเข้าร่วมอบรม ท่านมีความรู้ความเข้าใจเพียงใด	๑๓.๗๙ (๔)	๕๑.๗๒ (๑๕)	๓๑.๐๓ (๙)	๓.๔๕ (๑)	๐.๐๐	๓.๗๖
๒.๓ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมไปปรับใช้ในการปฏิบัติงานได้เพียงใด	๑๐.๓๔ (๓)	๔๔.๘๓ (๑๓)	๓๔.๔๘ (๑๐)	๑๐.๓๔ (๓)	๐.๐๐	๓.๕๕
๒.๔ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมไปถ่ายทอดให้เพื่อนร่วมงานได้เพียงใด	๑๐.๓๔ (๓)	๓๗.๙๓ (๑๑)	๔๔.๘๓ (๑๓)	๖.๙๐ (๒)	๐.๐๐	๓.๕๒
๓. ภาพรวมการจัดการอบรม						**๔.๔๐**
๓.๑ ระยะเวลาการอบรม	๔๘.๒๘ (๑๔)	๓๑.๐๓ (๙)	๑๗.๒๔ (๕)	๓.๔๕ (๑)	๐.๐๐	๔.๒๔
๓.๒ การประสานงานและการอำนวยความสะดวก	๕๑.๗๒ (๑๕)	๔๔.๘๓ (๑๓)	๓.๔๕ (๑)	๐.๐๐	๐.๐๐	๔.๔๘
๓.๓ ความพึงพอใจในภาพรวมของการอบรม	๕๑.๗๒ (๑๕)	๔๔.๘๓ (๑๓)	๓.๔๕ (๑)	๐.๐๐	๐.๐๐	๔.๔๘

๓. ข้อคิดเห็น

๓.๑ ข้อดีในการอบรมฯ

๑) ได้ความรู้เกี่ยวกับ GIS มากขึ้น
๒) สามารถนำไปใช้ในการปฏิบัติงานได้
๓) ทำให้ทราบการใช้งานโปรมแกรม ArcGIS pro และ ERDAS imagine เพื่อใช้ต่อยอดในการทำงาน
๔) อาหารอร่อย ขนมเบรคอร่อยมาก
๕) ทำให้มีความรู้ใหม่ ๆ เกี่ยวกับทางด้านระบบแผนที่
๖) เนื้อหาเรื่องภูมิศาสตร์เป็นข้อมูลเชิงลึกดี
๗) เป็นเนื้อหาเบื้องต้นสำหรับเจ้าหน้าที่ที่ไม่มีประสบการณ์

๓.๒ ข้อควรปรับปรุงในการอบรมฯ

๑) ผู้เข้าร่วมทาง Zoom ตามไม่ค่อยทัน
๒) ควรจัดในห้องคอมพิวเตอร์ของหน่วยงาน
๓) ควรมีปลั๊กพ่วงให้ด้วย
๔) อยากให้เพิ่มวันจัดอบรม
๕) การติดตั้งโปรแกรมค่อนข้างลำบาก

๖) วิทยากรสอนเร็วไป
๗) อยากให้เป็นโปรแกรม Open source เพราะหลังการอบรมจะได้นำไปใช้ได้จริง
๘) ควรมีการปูพื้นฐานก่อน เพราะคนที่ไม่ได้ทำงานเกี่ยวกับ GIS จะตามไม่ค่อยทัน
๙) ความพร้อมของข้อมูลและโปรแกรม
๑๐) เนื้อหาเรื่องภูมิศาสตร์เป็นเชิงลึก จึงไม่รู้ว่าจะนำไปประยุกต์ใช้กับงานอย่างไร
๑๑) ควรแจ้งวิทยากรถึงพื้นฐานความรู้ของผู้เข้าร่วมอบรม และควรเชิญหน่วยงานที่เกี่ยวข้องมาเข้าร่วมอบรม

๔. ข้อเสนอแนะอื่น ๆ

๑) อยากให้มี Case study เพิ่มเติม
๒) ควรเพิ่มการสอนเกี่ยวกับการดึงข้อมูลดาวเทียมด้านเกษตรด้วย
๓) ควรให้มีการอบรมเพิ่มเติมเกี่ยวกับ Spatial econometrics เพื่อสามารถนำข้อมูล GIS มาใช้ประโยชน์ได้มากขึ้น
๔) ควรกำหนดกรอบการอบรมที่เกี่ยวข้องกับภารกิจของ สศก. เท่านั้น เพื่อให้สามารถนำไปประยุกต์ใช้ได้จริง

....

สรุปการอบรมเชิงปฏิบัติการ เรื่อง การใช้งานเครื่องมือเพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data)
หลักสูตร "การจัดการข้อมูลขนาดใหญ่ สำหรับวิศวกรข้อมูล (Data Engineering)
ด้วย Huawei Cloud และ Microsoft SQL Server"
วันที่ ๒๘ – ๓๐ สิงหาคม ๒๕๖๖ เวลา ๐๘.๓๐ – ๑๖.๐๐น.
ณ ห้องประชุม AEOC ชั้น ๓ อาคารศูนย์ปฏิบัติการเศรษฐกิจการเกษตร สำนักงานเศรษฐกิจการเกษตร

....

ศูนย์ข้อมูลเกษตรแห่งชาติ (ศกช.) โดยส่วนธรรมาภิบาลข้อมูล (สธข.) ได้จัดอบรมเชิงปฏิบัติการ เรื่อง การใช้งานเครื่องมือเพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) หลักสูตร "การจัดการข้อมูลขนาดใหญ่ สำหรับวิศวกรข้อมูล (Data Engineering)" มีวัตถุประสงค์สำคัญ ๒ ประการ คือ ประการแรก เพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในการบริหารจัดการข้อมูลขนาดใหญ่ (Big Data) ประการที่สอง เพื่อพัฒนาศักยภาพของบุคลากรสำนักงานเศรษฐกิจการเกษตรให้มีทักษะในการพัฒนา การเชื่อมโยงข้อมูลและจัดเตรียมข้อมูลเพื่อนำไปสู่การวิเคราะห์ข้อมูลทางการเกษตร โดยมีผู้เข้าร่วมอบรม ๑๐ คน รายละเอียดของการอบรมและข้อคิดเห็นต่าง ๆ สรุปได้ดังนี้

๑. ภาคบรรยาย เป็นการให้ความรู้ทางวิชาการโดยวิทยากร ดร.รัฐศิลป์ รานอกภานุวัชร์ อาจารย์ประจำหลักสูตรบัณฑิตศึกษา วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ (CITE) มหาวิทยาลัยธุรกิจบัณฑิตย์ โดยมีหัวข้อบรรยาย ดังนี้

๑.๑ Big Data Fundamental และระบบนิเวศของ Apache Hadoop

ข้อมูลขนาดใหญ่ (Big data) คือ ข้อมูลขนาดใหญ่ที่มีความซับซ้อนมาก การที่จะใช้วิธีการประมวลผลแบบเดิม ไม่สามารถจัดเก็บและประมวลผลได้อย่างมีประสิทธิภาพ ซึ่งการจัดเก็บข้อมูลแบบ Big Data สามารถนำไปวิเคราะห์ข้อมูลในด้านต่าง ๆ เพื่อนำไปวางแผน และตัดสินใจ ในการดำเนินธุรกิจ หรือช่วยเพิ่มโอกาสในการทำธุรกิจให้ก้าวหน้ามากขึ้น ตัวอย่างการประยุกต์ใช้ Big Data ในธุรกิจต่าง ๆ เช่น Customer Segmentation ใช้การแบ่งกลุ่มลูกค้า (Segmentation) เลือกกลุ่มลูกค้าเป้าหมาย (Target) และวางตำแหน่งทางการตลาดของสินค้าและบริการ (Positioning) , Churn Prevention แนวโน้มจะยกเลิกการใช้บริการ (Churn) ในอนาคต เป็นต้น

คุณลักษณะของข้อมูลขนาดใหญ่ ตามนิยาม ๔V

๑. ปริมาณ (Volume): ข้อมูลมีขนาดใหญ่และปริมาณมาก

๒. ความหลากหลาย (Variety): ความหลากหลายของประเภทของข้อมูล

๓. ความเร็ว (Velocity): ความเร็วในการประมวลผล

๔. ความถูกต้อง (Veracity): คุณภาพของข้อมูล

ในการจัดการข้อมูลขนาดใหญ่จำเป็นต้องใช้เครื่องมือและเทคโนโลยีพิเศษ เช่น Apache Hadoop, Apache Spark, Apache Kafka, ฐานข้อมูล NoSQL และเทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning) เพื่อให้สามารถเพิ่มประสิทธิภาพในการจัดการและวิเคราะห์ข้อมูลขนาดใหญ่

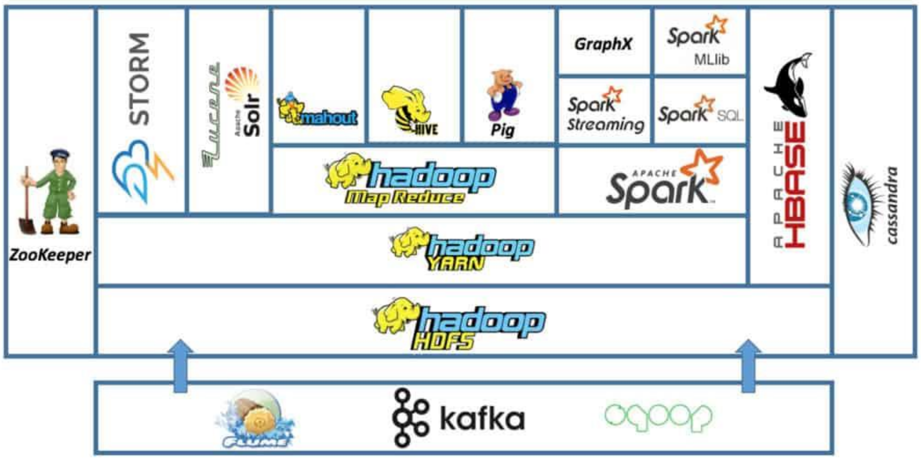
ระบบนิเวศของ Apache Hadoop เป็น open-source framework ที่ออกแบบมาเพื่อการจัดการและประมวลผลข้อมูลขนาดใหญ่และข้อมูลที่มีโครงสร้างแตกต่างกัน ซึ่งอยู่ในรูปแบบของระบบนิเวศ (ecosystem) ที่มีความหลากหลาย โดยระบบนิเวศของ Apache Hadoop ประกอบด้วยคอมโพเนนต์ (components) หลายอย่างที่มีความสำคัญ ซึ่งส่วนประกอบที่สำคัญ ได้แก่

๑. Hadoop Distributed File System (HDFS) เป็นระบบจัดเก็บข้อมูลแบบกระจายที่สามารถจัดการข้อมูลขนาดใหญ่ โดยที่แต่ละส่วนถูกคัดลอกและกระจายไปยังคอมพิวเตอร์หลายเครื่องเพื่อความปลอดภัยและความพร้อมใช้งาน

๒. Yet another resource negotiator (YARN) เป็นระบบจัดการทรัพยากรและกำหนดลำดับการทำงานในระบบ Hadoop โดยที่ YARN ช่วยให้สามารถใช้งานทรัพยากรที่มีอยู่ในระบบอย่างมีประสิทธิภาพ

๓. MapReduce เป็นโมเดลการประมวลผลแบบกระจายที่ใช้ในการนำเข้าข้อมูลและประมวลผลข้อมูลในระบบ Hadoop ซึ่ง MapReduce มีฟังก์ชันการแบ่งข้อมูลออกเป็นส่วนย่อยและการรวมผลลัพธ์จากส่วนย่อยเพื่อให้เกิดการประมวลผลแบบกระจาย

๔. Hive เป็นฐานข้อมูลคิวรี่และฐานข้อมูลหลักสำหรับการวิเคราะห์ข้อมูลในรูปแบบ MapReduce โดยใช้ภาษา SQL เป็นหลัก



ภาพที่ ๑ ภาพระบบนิเวศของ Apache Hadoop

NoSQL (Not Only SQL) เป็นแนวทางหนึ่งสำหรับจัดการข้อมูลและการออกแบบฐานข้อมูล สำหรับข้อมูลขนาดใหญ่ ซึ่งอยู่อย่างกระจัดกระจาย หลากหลายรูปแบบ ซึ่งแบ่งออกเป็น ๔ ประเภท คือ Key-Value Store, Document Store, Columnar Store, Graph

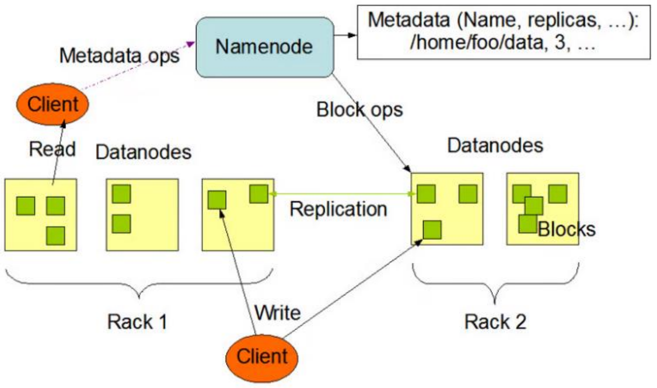
ตัวอย่างของระบบฐานข้อมูล NoSQL ที่นิยม ได้แก่ MongoDB (Document Store), Cassandra (Wide-Column Store), Redis (Key-Value Store), Neo๔j (Graph Database), และ Amazon DynamoDB (Managed NoSQL Database)

๑.๒ การจัดการข้อมูลด้วย HDFS และ การประมวลผลด้วย Hadoop MapReduce

Hadoop Distributed File System (HDFS) เป็นหนึ่งในองค์ประกอบพื้นฐานของ framework ใน Apache Hadoop ซึ่งเป็นระบบไฟล์แบบกระจาย (Distributed File System) เพื่อรองรับการจัดเก็บและประมวลผลข้อมูลขนาดใหญ่ สามารถปรับขนาดได้ให้เหมาะสมกับการจัดการข้อมูลที่มีปริมาณมากในคลัสเตอร์ของโหนดคอมพิวเตอร์ ซึ่ง HDFS ถูกออกแบบมาเพื่อรับรองการประมวลผลข้อมูลแบบเป็นชุดและยังเหมาะสมกับการอ่านข้อมูลแบบสตรีมมิ่งขนาดใหญ่ โดย HDFS จะจัดเก็บข้อมูลในหลาย ๆ โหนดภายในคลัสเตอร์ ซึ่งข้อมูลจะถูกแบ่งออกเป็นบล็อก (โดยค่าเริ่มต้น มีขนาดเท่ากับ ๑๒๘ MB) และกระจายไปทั่วคลัสเตอร์ โดยองค์ประกอบหลักของ HDFS มี ๒ องค์ประกอบ คือ

๑. NameNode เป็นองค์ประกอบหลักในการจัดเก็บและจัดการข้อมูล ให้เกิดความสมบูรณ์ของระบบไฟล์ โดยจะรักษาโครงสร้างระบบไฟล์และข้อมูลเมตาสำหรับไฟล์และไดเร็กทอรีทั้งหมด NameNode มีหน้าที่จัดการแมปและติดตามโครงสร้างของไฟล์และไดเรกทอรีใน HDFS รวมถึงบันทึกข้อมูลของแต่ละบล็อกข้อมูลและสำเนาข้อมูล (replication) ที่เก็บอยู่ใน DataNode

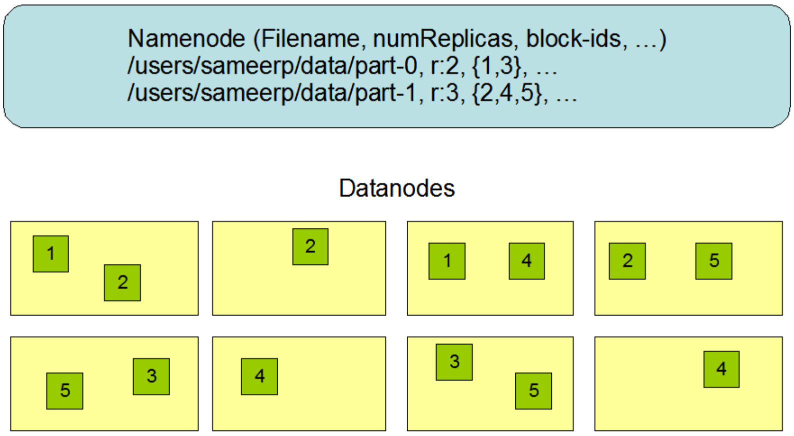
๒. DataNode เป็นโหนดของผู้ปฏิบัติงานภายใน HDFS ซึ่งมีหน้าที่จัดเก็บและจัดการบล็อกข้อมูลบนอุปกรณ์จัดเก็บข้อมูลในตัวเครื่อง โดย DataNodes จะสื่อสารกับ NameNode เพื่อจัดการพื้นที่จัดเก็บ และจะส่งรายงานสถานะของข้อมูลไปยัง NameNode ซึ่งข้อมูลที่จัดเก็บไว้ใน DataNodes จะกระจายไปทั่วดิสก์หลาย ๆ ตัว ทำให้สามารถดำเนินการข้อมูลแบบขนานได้สูง



ภาพที่ ๒ องค์ประกอบหลักของ HDFS

HDFS Data Replication เป็นกระบวนการสร้างสำเนาข้อมูล (replication) ของข้อมูลใน Hadoop Distributed File System (HDFS) ซึ่งจะทำสำเนาเป็น Block ถ้าไฟล์ไหนมีขนาดใหญ่จะถูกหั่นแบ่งออกเป็นก้อน ๆ และกระจายกัน โดยปกติแล้วจะมีสำเนาข้อมูลสองหรือสามชุด คือสำเนาข้อมูลหลักและสำเนาข้อมูลสำรอง ซึ่งเวลาประมวลผลก็จะแบ่งส่วนกันประมวลผล

การสร้างสำเนาข้อมูล (replication) เป็นคุณลักษณะสำคัญใน HDFS เพื่อให้ข้อมูลมีความทนทานและคงอยู่ แม้ว่า DataNode หนึ่งจะเสียหาย ข้อมูลยังมีสำเนาอยู่ในระบบ ซึ่งจะช่วยเพิ่มประสิทธิภาพในการอ่านข้อมูลและลดการขาดหายของข้อมูลในระบบ HDFS ได้



ภาพที่ ๓ Block Replication

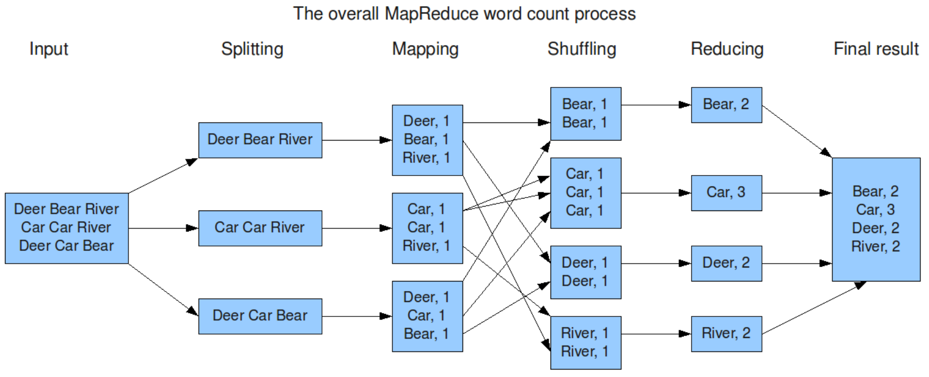
Hadoop MapReduce เป็นส่วนที่ใช้ในการสั่งการประมวลผลข้อมูลขนาดใหญ่ โดย Hadoop MapReduce ถูกออกแบบมาเพื่อการประมวลผลข้อมูลในรูปแบบ batch processing ซึ่งสามารถประมวลผลข้อมูลขนาดใหญ่มากและแยกออกเป็นงานย่อย ๆ ที่ประมวลผลพร้อมกันบนคลัสเตอร์คอมพิวเตอร์หลายเครื่อง

MapReduce แบ่งการทำงานเป็น ๓ ส่วนหลัก คือ

๑. Mapping มีหน้าที่ในการรับข้อมูลเข้ามาแล้วแปลงข้อมูลเหล่านี้เป็นคู่ (key, value) ซึ่งจะทำการแบ่งข้อมูลและทำการจับคู่ข้อมูล โดยเริ่มจากการที่มีข้อมูลชุดใหญ่ชุดหนึ่ง แล้วนำมาแบ่งเป็นหลายส่วน จากนั้นแตกข้อมูลให้อยู่ในรูปแบบ key-value

๒. Shuffling ส่วนนี้จะทำการจัดเรียงและรวบรวมข้อมูลที่มี key เหมือนกันให้อยู่ด้วยกัน ผลลัพธ์จากขั้นตอนนี้ก็ยังอยู่ในรูปแบบ key-value เหมือนเดิม

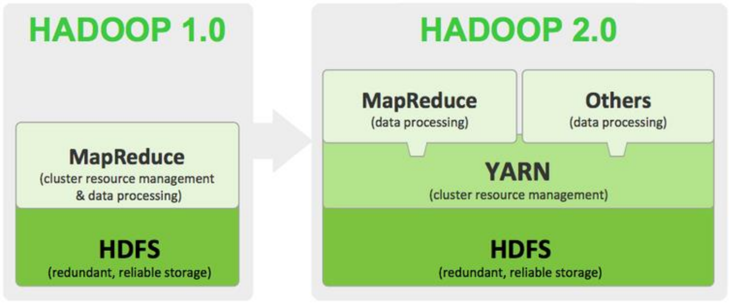
๓. Reducing จะรับข้อมูลที่ถูกจัดเรียงและจัดเก็บจาก Shuffle และทำการประมวลผลข้อมูลเหล่านี้ต่อเนื่องจากข้อมูลที่มีคีย์เดียวกัน ซึ่งสามารถทำงานพร้อมกันบนคลัสเตอร์คอมพิวเตอร์หลายเครื่อง ผลลัพธ์จากการ Reduce จะถูกเขียนลงในระบบจัดเก็บข้อมูล (เช่น Hadoop HDFS) หรือถูกส่งกลับไปยังโปรแกรมที่ร้องขอเพื่อการประมวลผล



ภาพที่ ๔ ตัวอย่างการทำงานของ MapReduce

Yet Another Resource Negotiator (YARN) เป็น Resource Manager ทำหน้าที่จัดตารางเวลาให้งาน (Task) และบริหารจัดการทรัพยากร (Resource) ต่าง ๆ เช่น หน่วยเก็บข้อมูล และหน่วยประมวลผลบนระบบ ให้ทำงานได้ลื่นไหล เป็นเหมือนผู้คุมงานที่คอยจัดสรรงานและคนงาน (Worker)

ความแตกต่างระหว่าง Hadoop ๑.๐ และ Hadoop ๒.๐



ภาพที่ ๕

Hadoop ๑.๐...

Hadoop ๑.๐ ใช้ ResourceManager และ TaskTracker ในการจัดการทรัพยากรและงานประมวลผลข้อมูล ในขณะที่ Hadoop ๒.๐ (YARN) ใช้ ResourceManager และ NodeManager เพื่อแยกการจัดการทรัพยากรและงานประมวลผล

ความสามารถของ Apache Hadoop คือ

๑. ความเร็ว (Speed) สามารถประมวลผลข้อมูลได้ในเวลาอันสั้น

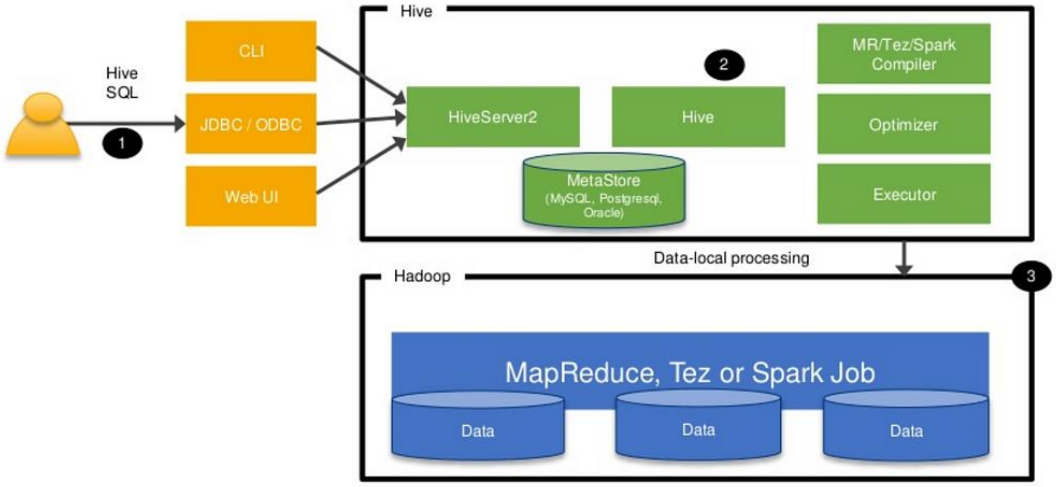
๒. ทนทานต่อความเสียหาย (Fault-tolerance) สามารถทำงานต่อได้ แม้ระบบขัดข้อง

๓. ข้อมูลพร้อมใช้งานตลอดเวลา (Data Availability) ข้อมูลจะมีการทำสำเนาและกระจายไปยัง Node อื่น ๆ ซึ่งมั่นใจว่าข้อมูลจะไม่หายไปไหน

๔. ประมวลผลแบบขนาน (Parallel Processing) ทำให้ทำงานได้ไวขึ้น

๑.๓ การจัดการข้อมูลและ Query ข้อมูลด้วย Hive

Apache Hive ถูกพัฒนาขึ้นโดย Facebook เป็นระบบคิวรี่และการประมวลผลข้อมูลในรูปแบบแถวและคอลัมน์ เพื่อการ Query ข้อมูลขนาดใหญ่ในระดับ Petabyte โดยใช้ภาษา Query ที่เรียกว่า Hive SQL ซึ่งคล้ายกับภาษา SQL และใช้ Database ที่เรียกว่า Metastore ในการเก็บ Metadata ในการทำงาน โดย ซึ่ง Hive จะแปลงคำสั่ง Hive SQL ไปเป็น MapReduce Jobs โดยทั่วไปแล้วจะเหมาะกับงานประเภท Batch Processing กับข้อมูลที่มีปริมาณมาก ๆ และเปลี่ยนแปลงไม่บ่อย แต่ไม่เหมาะกับงานประเภท Online Transaction Processing (OLTP) และ Real-Time Queries



ภาพที่ ๖ สถาปัตยกรรมของ Hive

Hive ประกอบด้วย ๓ องค์ประกอบหลัก คือ

๑. Hive Client จะมี Driver ต่าง ๆ สำหรับการติดต่อสื่อสารกับ Application

๑) Thrift Server เป็นแพลตฟอร์มผู้ให้บริการที่รองรับโปรแกรมทั้งหมดที่ support thrift เพื่ออนุญาตให้ผู้ใช้ระยะไกล (remote user) ติดต่อกับ HiveServer ๒ (the thrift server) เพื่อเชื่อมต่อและส่ง query

๒) JDBC Driver ใช้สำหรับเชื่อมต่อระหว่าง Hive และ Java

๓) Driver ODBC เป็นแอปพลิเคชันสนับสนุน Protocol ODBC เอาไว้เชื่อมต่อกับ Hive

๒. Hive Services เป็น Application สามารถติดต่อกับ Hive ได้โดยผ่าน Hive Server นอกจากนี้ Hive ก็ยังมีช่องทางอื่น ๆ ที่สามารถติดต่อได้ คือ Hive Web Interface และ Command Line Interface ซึ่งเมื่อมีการติดต่อแล้ว จะส่งต่อ Request ไปยัง Driver ซึ่ง Driver จะประมวลผล Request ให้เป็นสิ่งที่เข้าใจกันภายใน เช่น MetaStore, File System หรือ Job Client

๓. Hive Storage and Compute ผลลัพธ์จาก Query จะถูกจัดเก็บไว้ใน Hadoop Cluster บน HDFS

Hive สามารถสร้าง Table ได้ ๒ ประเภท คือ

๑. Internal Table เป็น Default Table ของ Hive ซึ่งเมื่อผู้ใช้งานทำการสร้าง Table ใน Hive โดยไม่ได้กำหนดประเภทของตาราง Hive จะถือว่า Table นั้นเป็น Internal Table ตามค่า Default ซึ่ง Table จะถูกสร้างขึ้นในตำแหน่งเฉพาะใน HDFS การทำงานของ Internal Table คือ สร้างตารางและ Metadata ก่อน จากนั้นทำการย้ายข้อมูลจาก HDFS มาที่ไว้ที่ Hive ดังนั้น หากเรา Drop Table จะทำให้ทั้ง Metadata และข้อมูลถูกลบไปด้วย

๒. External Table เป็นประเภทของตารางที่ Hive ไม่ต้องรับผิดชอบในด้านการจัดการข้อมูล การทำงานของ External Table คือ สร้างตารางมาครอบข้อมูลที่ถูกวางไว้ที่ HDFS ตามโครงสร้างของข้อมูลนั้น ๆ ดังนั้น หากเรา Drop Table จะมีเพียงแค่ Metadata ที่ถูกลบไปด้วย แต่ข้อมูลจะยังคงอยู่

Apache Impala เป็นเครื่องมือเพิ่มเติมที่ใช้ในการสืบค้นข้อมูลขนาดใหญ่ Impala ไม่ได้แทนที่การประมวลผลที่สร้างขึ้นบน MapReduce แต่ Impala จะช่วยในการสืบค้น SQL ที่รวดเร็วและการสืบค้นโดยตรงบนข้อมูลที่จัดเก็บใน HDFS โดย Impala สามารถ read และ write ลงใน Hive tables และการเปิดใช้งานการแลกเปลี่ยนข้อมูลอย่างง่ายโดยใช้ Impala สำหรับการวิเคราะห์ข้อมูลโดย Hive

๑.๔ ศึกษาการรวบรวมข้อมูลจากแหล่งข้อมูลต่าง ๆ

Data Ingestion คือ กระบวนการที่ใช้ในการนำเข้าข้อมูลจากแหล่งต่าง ๆ เข้าสู่ระบบหรือพื้นที่เก็บข้อมูล เพื่อทำให้ข้อมูลเหล่านั้นสามารถใช้งานได้ในกระบวนการวิเคราะห์หรือการประมวลผลข้อมูลต่อไป สามารถแบ่งการทำงานได้เป็น ๒ แบบ ได้แก่ แบบกลุ่ม (Batch) และเรียลไทม์หรือกึ่งเรียลไทม์ (Real-time/Near Real-time)

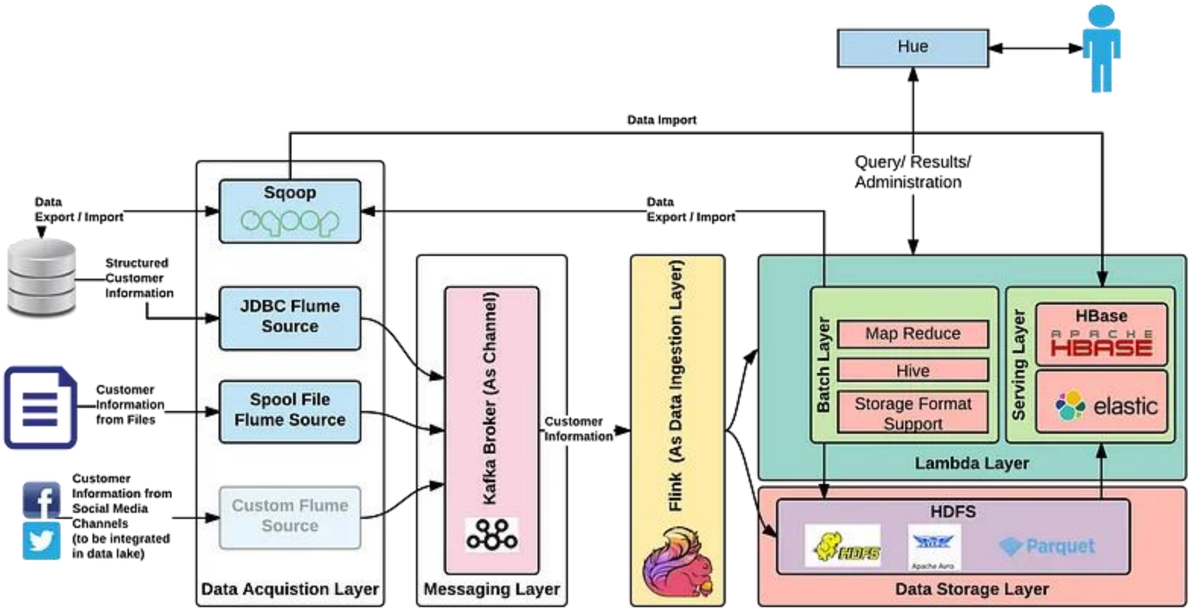
Apache Sqoop มีวัตถุประสงค์ในกระบวนการการวิเคราะห์ข้อมูลและการโอนข้อมูลระหว่างฐานข้อมูลที่เป็นโครงสร้างและระบบ Hadoop HDFS (Hadoop Distributed File System) ซึ่งมีการรองรับหลายฐานข้อมูล เช่น MySQL, Oracle, PostgreSQL, Microsoft SQL Server เป็นต้น โดยสามารถโอนถ่ายข้อมูลแบบ batch คือ โอนข้อมูลทั้งหมดในรอบเดียว และการโอนถ่ายข้อมูลแบบเรียลไทม์ คือ ติดตามการเปลี่ยนแปลงและโอนข้อมูลเมื่อมีการเปลี่ยนแปลงในฐานข้อมูลต้นทาง โดย Sqoop ใช้เมื่อต้องการรวมข้อมูลจากหลายแหล่งข้อมูลหรือโอนข้อมูลจากฐานข้อมูลเข้าสู่โครงสร้างข้อมูลขนาดใหญ่ในระบบ Hadoop

Apache Flume เป็นระบบที่ใช้สำหรับการทำ Data Ingestion จากแหล่งต่าง ๆ ไปยังแหล่งเก็บข้อมูลหรือระบบประมวลผลข้อมูลอื่น ๆ อย่างมีประสิทธิภาพ โดยทั่วไปมักใช้ในระบบ Big Data และการวิเคราะห์ข้อมูลขนาดใหญ่ (data analytics) ซึ่งเหมาะกับข้อมูลประเภท streaming ที่เป็นข้อมูลที่ไหลมาเรื่อย ๆ และรวดเร็ว เช่น logging ที่มาจาก application หรือ action จากผู้ใช้ โดยมีลักษณะการทำงาน คือรับ streaming data มาจากแหล่งข้อมูลต่าง ๆ แล้วไหล ไปยัง hadoop ecosystem ซึ่งวิธีการใช้งาน Flume ต้องระบุ source และรูปแบบของการดึงข้อมูล ซึ่งสามารถทำได้ทั้ง event-driven และวิธี polling จากนั้นก็ระบุ channel ว่าจะเก็บข้อมูลชั่วคราวก่อนส่งไปยัง target อย่างไร เช่น เก็บไว้ใน memory หรือเป็น file หรือจะเก็บไว้ใน messaging queue ก็ได้ จากนั้นระบุ target ให้เอาไปเก็บที่ไหน

Apache....

Apache Kafka เป็น distributed message queue มีการประมวลผลข้อมูลแบบสตรีมและ message broker ถูกออกแบบมาเพื่อจัดการกับปริมาณข้อมูลแบบ real-time ใหญ่ ๆ และให้บริการโครงสร้างระบบกระจายที่มีความทนทานและสามารถขยายขนาดได้ ซึ่งใช้โมเดลการส่งข้อความแบบ publish-subscribe โดยผู้สร้างข้อมูล (publishers) จะส่งข้อมูลไปยังหัวข้อ (topics) และผู้บริโภคข้อมูล (consumers) สามารถสมัครใช้และรับข้อมูลจากหัวข้อที่ต้องการได้ โดย Kafka สามารถใช้งานร่วมกับหลายภาษาโปรแกรมมิ่ง เช่น Java, Python, .NET, และมีการพัฒนา Client Library ที่หลากหลาย ซึ่ง Apache Kafka มีการใช้งานในด้านการประมวลผลข้อมูลแบบสตรีม (stream processing) เช่น การวิเคราะห์ข้อมูลที่เป็นลำดับเวลา (time-series data analysis) การตรวจจับข้อมูลที่เรียบเรียงลำดับ (event sourcing) และการสร้างแพลตฟอร์มสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ (big data analytics) เป็นต้น

Apache Flink เป็นระบบการประมวลผลข้อมูลแบบสตรีมและแบบ batch processing แต่เน้นการใช้งานในการประมวลผลข้อมูลแบบ real-time ระบบ Cluster ข้อมูลใหญ่ ๆ ซึ่งสามารถนำมาใช้ในการวิเคราะห์ข้อมูลทางธุรกิจ, การตรวจจับความผิดปกติในระบบ, การคาดการณ์และการจัดการข้อมูลแบบ real-time และอื่น ๆ อีกมากมาย



ภาพที่ ๗ Data lake implementation

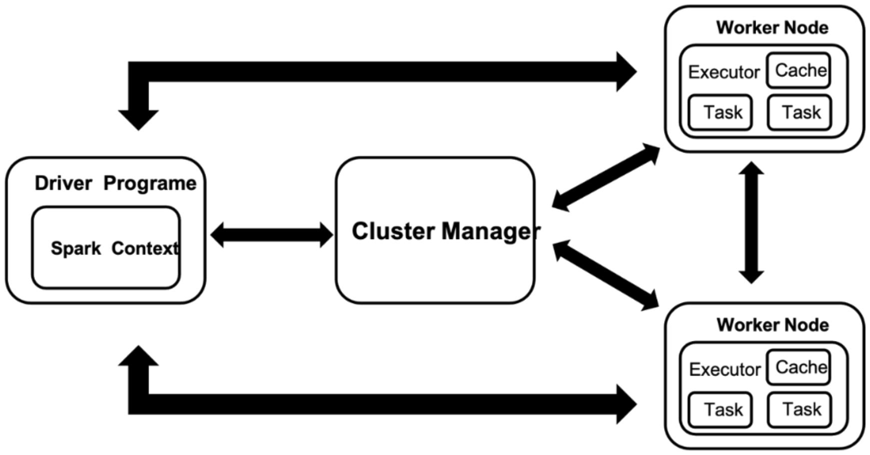
๑.๕ การวิเคราะห์ข้อมูลขนาดใหญ่โดยใช้ Apache Spark

Apache Spark เป็นหนึ่งในเทคโนโลยีด้าน Big data ที่ได้รับความนิยมสูง ซึ่งเป็น Open Source ที่สามารถใช้งานได้ร่วมกันหลายภาษา เช่น Python, R หรือ Scala โดยที่เชื่อมต่อระบบการทำงานของคอมพิวเตอร์เข้าด้วยกัน หรือเรียกว่า Cluster computing platform ซึ่งสามารถกระจายงานที่ต้องทำไปยังเครื่องอื่น ๆ ภายในระบบ ทำให้เราสามารถประมวลผลข้อมูลขนาดใหญ่แบบเต็มประสิทธิภาพ หรือแบบ real-time ไปพร้อม ๆ กันได้ นอกจากนี้ Spark ยังสามารถใช้งานกับฐานข้อมูลได้หลากหลาย ไม่ว่าจะเป็น Cassandra, AWS S๓, HDFS ฯลฯ รวมถึงทำงานกับ Disk Storage ได้อีกด้วย ความแตกต่างระหว่าง Spark กับ MapReduce คือ MapReduce จะต้องให้ไฟล์จัดเก็บในระบบ HDFS แต่ Spark ไม่จำเป็น และ Spark สามารถใช้งานได้รวดเร็วกว่า MapReduce ถึง ๑๐๐ เท่า ซึ่งเหตุผลที่เร็วกว่านั้น

คือ MapReduce...

คือ MapReduce จะเขียนข้อมูลหลังจากการทำ Map และ Reduce ลงใน Disk Storage แต่ Spark จะเก็บข้อมูลทั้งหมดหลังจากการทำแต่ละ Transformation ไว้ที่หน่วยความจำ แต่ถ้าหน่วยความจำที่เก็บเต็ม Spark ก็ยังสามารถกระจายข้อมูลไปสู่ Disk Storage ได้

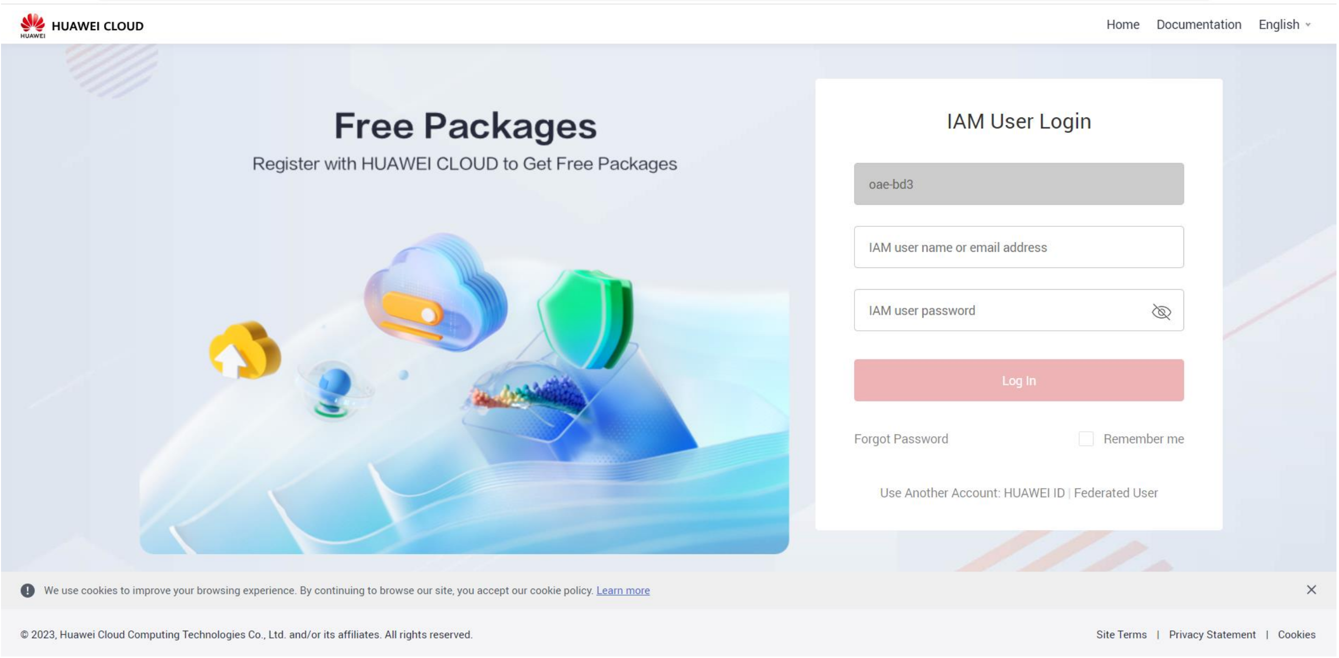
Resilient Distributed Dataset (RDD) โดยหลัก ๆ แล้ว Spark จะทำงานแบบ Cluster computing โดยมีลักษณะสำคัญคือ Distributed Collection of Data การกระจายข้อมูลอยู่ในหลาย ๆ node ของ cluster นอกจากนี้ยังมีความสามารถในการประมวลแบบขนาน (Parallel Operation-portioned) และ Fault-tolerant อีกด้วย



ภาพที่ ๘ Spark RDDs

ตัวอย่างการใช้งาน Service Apache Hadoop บน Huawei Cloud

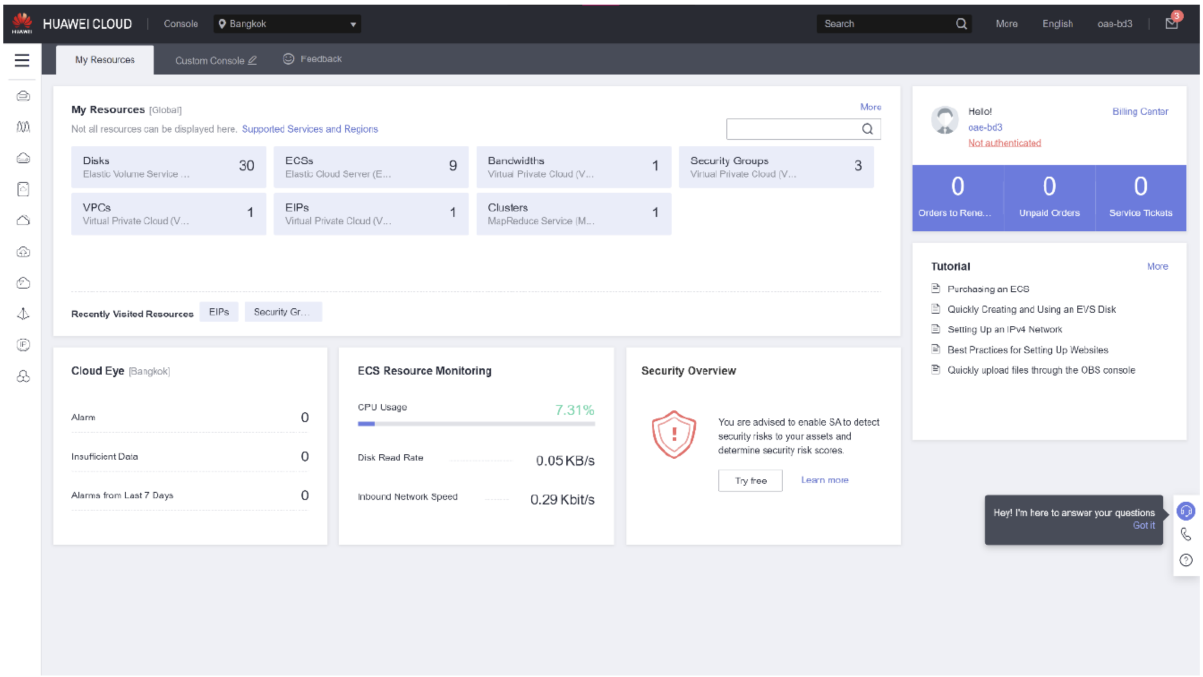
๑. ทำการ Log in เข้าใช้งาน Huawei Cloud



ภาพที่ ๙

๒. หน้าจอ

๒. หน้าจอควบคุมภายใน Huawei Cloud



ภาพที่ ๑๐ console ของ Huawei Cloud MapReduce Service (MRS)

ซึ่งประกอบด้วย Core Service สำคัญดังภาพที่ ๑๑ ดังนี้

๑) การคำนวณ (Compute)

- Elastic Cloud Server (ECS) เป็นบริการเครื่องเสมือน (Virtual Machines) ที่ช่วยสร้างและจัดการเซิร์ฟเวอร์เสมือนได้อย่างยืดหยุ่น โดยสามารถปรับขนาดและกำหนดคุณสมบัติของเครื่องเสมือนตามความต้องการ

๒) การจัดเก็บ (Storage)

- Object Storage Service (OBS) เป็นการจัดการการเก็บข้อมูลแบบอ็อบเจกต์ OBS และอัพโหลดหรือดาวน์โหลดข้อมูลได้
- Block Storage เป็นการจัดการการจัดเก็บข้อมูลแบบบล็อกและการตั้งค่าดิสก์ที่เชื่อมต่อกับ ECS

๓) Network (เครือข่าย)

- Virtual Private Cloud (VPC): เป็นการสร้างและจัดการเครือข่ายคลาวด์เสมือนส่วนตัว (VPC) เพื่อควบคุมการสื่อสารระหว่างทรัพยากรในคลาวด์
- Elastic IP: เป็นการจัดการ Elastic IP สำหรับการกำหนดที่อยู่ IP สำหรับเครื่องเสมือน ECS

๔) ฐานข้อมูล (Database)

- Relational Database Service (RDS): เป็นการจัดการฐานข้อมูล SQL แบบความสัมพันธ์โดยใช้บริการ RDS
- NoSQL Database: เป็นการจัดการฐานข้อมูล NoSQL และแคชข้อมูลแบบกำหนดเองได้

๕) ความปลอดภัย...

๕) ความปลอดภัยและการรับรองตัวตน (Security & Identity)

- Identity and Access Management (IAM): เป็นการจัดการผู้ใช้และสิทธิ์การเข้าถึงทรัพยากรต่าง ๆ ในคลาวด์
- Firewall: เป็นการกำหนดกฎและกำหนดความปลอดภัยสำหรับคลาวด์

๖) การตรวจสอบและการจัดการ (Monitoring & Management)

- Cloud Eye: เป็นบริการการตรวจสอบและจัดการคลาวด์ เพื่อตรวจสอบสถานะและประสิทธิภาพของทรัพยากร
- Cloud Container Engine (CCE): เป็นการจัดการคลัสเตอร์ Docker และ Kubernetes

๗) บริการแอปพลิเคชัน (Application Service)

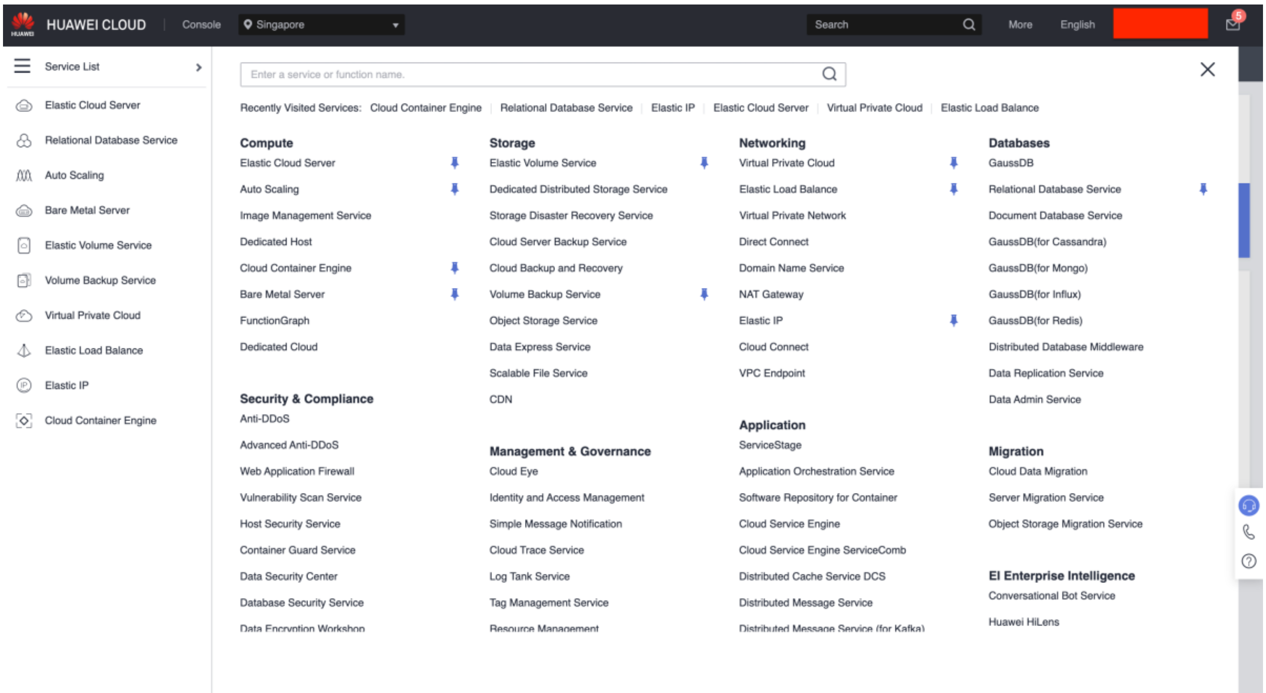
- Elastic Load Balancer (ELB): เป็นการจัดการโหลดแบลานซิ่งและการตั้งค่าการแจกแจงการเข้าถึงสำหรับแอปพลิเคชัน

๘) ข้อมูลและ AI (Data & AI)

- DataWorks: เป็นบริการสำหรับจัดการและวิเคราะห์ข้อมูล.
- ModelArts: เป็นการสร้างและจัดการโมเดลเรียนรู้ของเครื่องและการใช้งานข้อมูล AI

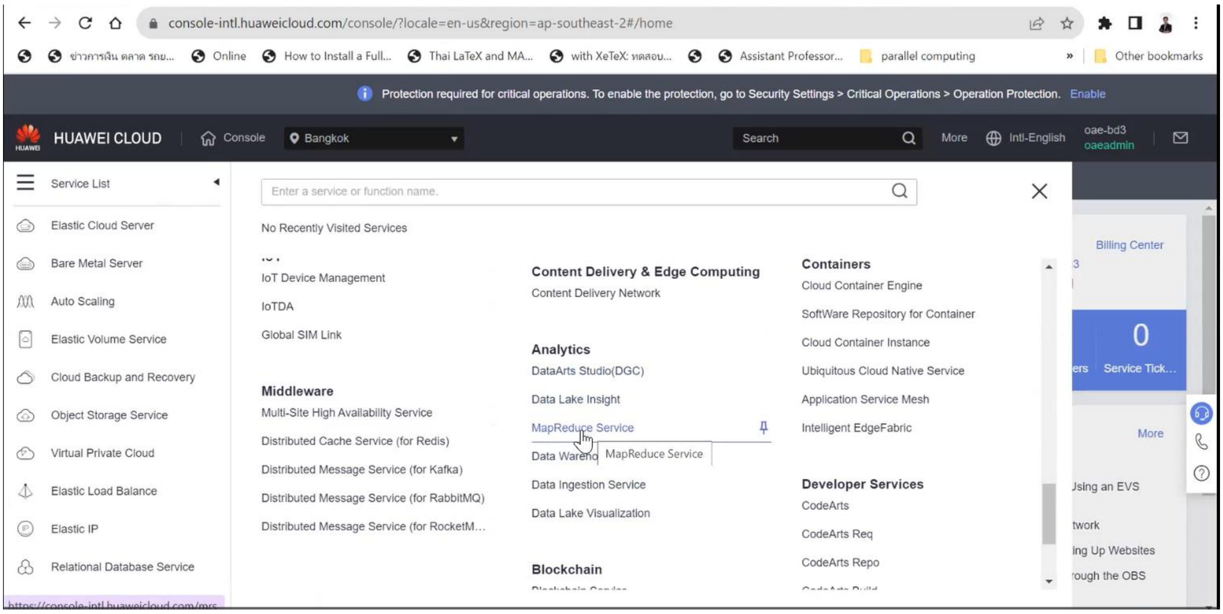
๙) บล็อกเชน (Blockchain) เป็นการจัดการและใช้งานเทคโนโลยีบล็อกเชนของ Huawei Cloud

๑๐) IoT (Internet of Things) เป็นบริการเฉพาะที่ช่วยในการเชื่อมต่ออุปกรณ์ IoT และการจัดการข้อมูล IoT



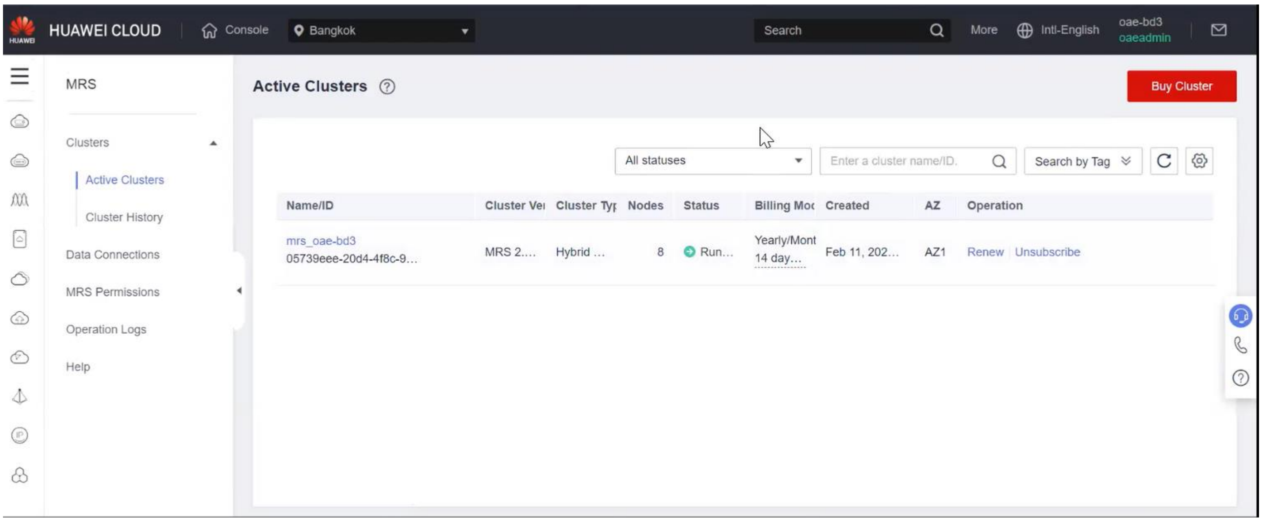
ภาพที่ ๑๑

๓. จากนั้นเข้าใช้งานที่ MapReduce Service เพื่อใช้ Service ที่เกี่ยวกับ Big data ดังภาพที่ ๑๒



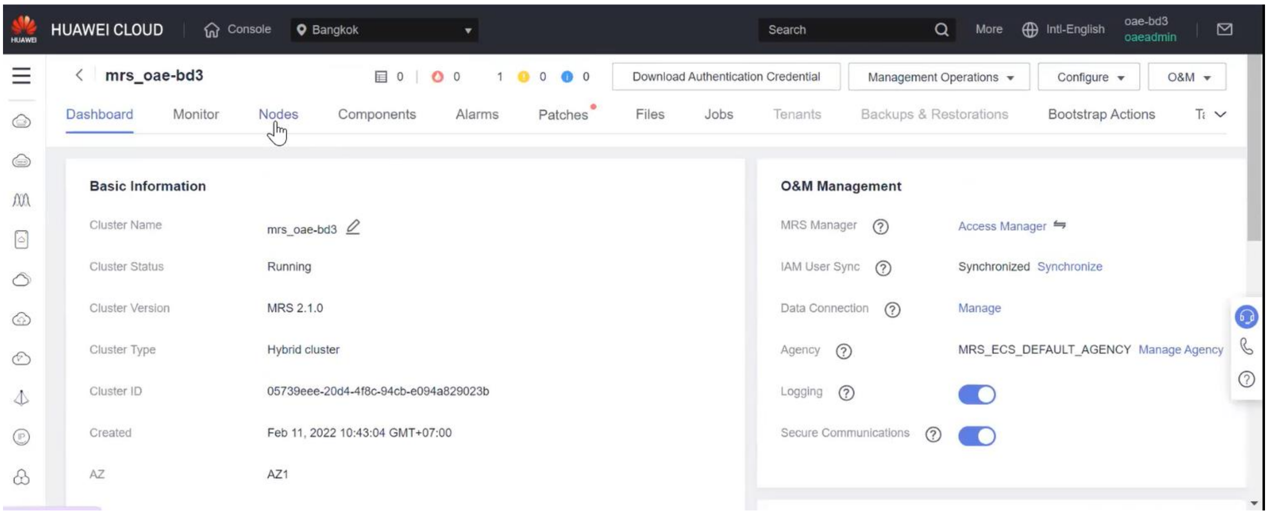
ภาพที่ ๑๒

๔. จะพบว่า มี cluster ที่สร้างอยู่แล้ว จำนวน ๑ cluster ประกอบด้วย node จำนวน ๘ node



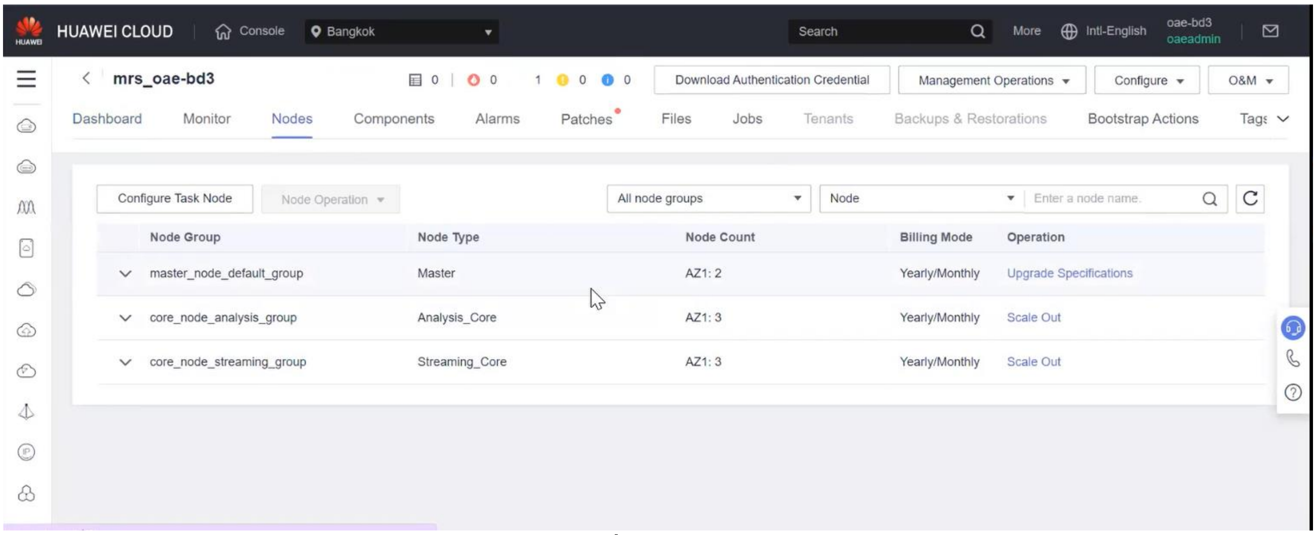
ภาพที่ ๑๓

๕. จากนั้นไปที่ Nodes



ภาพที่ ๑๔

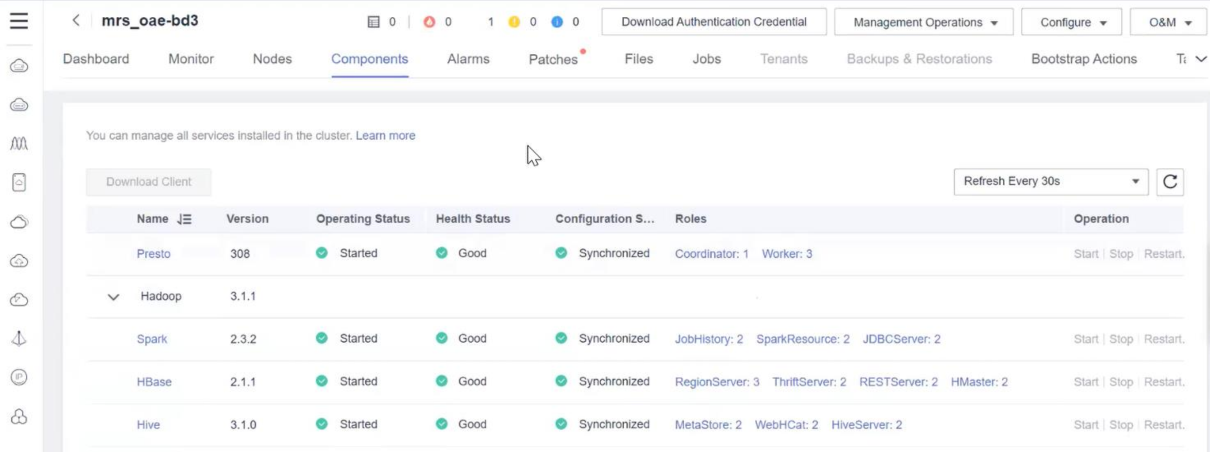
๖. จากนั้น จะพบว่า ๘ node ถูกแบ่งออกเป็น Master node จำนวน ๒ node ซึ่งเป็นแบบ Active-Stand by มี Analysis core node ทั้งหมด ๓ node และ Steaming core node อีก ๓ node รวมทั้งหมด ๘ node



ภาพที่ ๑๕

๗. จากนั้น...

๗. จากนั้นไปที่ Components เพื่อดู Service ทั้งหมดที่อยู่ใน Big Data Service ใน Huawei Cloud



ภาพที่ ๑๖

๑.๖ การจัดการข้อมูลโดยใช้ SQL Server Integration Service (SSIS)

SQL Server Integration Services (SSIS) เป็นเครื่องมือและบริการใน Microsoft SQL Server ที่ใช้ในการจัดการและบริหารงานนำเข้า (ETL - Extract, Transform, Load) ข้อมูล แม้จะมีการใช้งานกับ SQL Server แต่ SSIS สามารถทำงานกับแหล่งข้อมูลและปลายทางที่หลากหลายอย่างที่รองรับ OLE DB, ODBC, XML, Flat Files, Excel และอื่น ๆ ซึ่งทำให้มีความยืดหยุ่นในการดึงข้อมูลและนำเข้าข้อมูลจากแหล่งต่าง ๆ และนำเข้าข้อมูลลงในฐานข้อมูลหรือวิเคราะห์ข้อมูลตามความต้องการ

กระบวนการ Extract, Transform and Load (ETL) เป็นกระบวนการสำคัญในการจัดการข้อมูลในระบบคอมพิวเตอร์และฐานข้อมูล โดยมักนำมาใช้ในบริษัทหรือองค์กรเพื่อนำเข้าข้อมูลจากแหล่งต่าง ๆ และนำข้อมูลเหล่านั้นเข้าสู่ระบบฐานข้อมูลหรือระบบวิเคราะห์ข้อมูลเพื่อให้ข้อมูลนั้นสามารถนำไปใช้งานและวิเคราะห์ได้อย่างมีประสิทธิภาพ. กระบวนการ ETL ประกอบด้วยขั้นตอนหลัก ๆ คือ Extract (ดึงข้อมูล), Transform (แปลงข้อมูล), และ Load (นำเข้าข้อมูล) ดังนี้

๑. Extract (ดึงข้อมูล) เป็นขั้นตอนแรกในกระบวนการ ETL คือการดึงข้อมูล (Extract) จากแหล่งต้นทาง ซึ่งอาจเป็นฐานข้อมูล, ไฟล์ข้อมูล, บริการเว็บ, และแหล่งข้อมูลอื่น ๆ ซึ่งข้อมูลที่ถูกดึงมามักเป็นข้อมูลที่จำเป็นสำหรับการวิเคราะห์หรือรายงาน โดยในขั้นตอนนี้ต้องระบุแหล่งข้อมูล, เลือกข้อมูลที่ต้องการดึง, และเข้าถึงข้อมูลในรูปแบบที่เหมาะสม (เช่น SQL queries, APIs, หรือการอ่านไฟล์) ซึ่งข้อมูลที่ถูกดึงมาจะถูกเก็บในรูปแบบชั่วคราวหรือโครงสร้างในระบบสตอเรจชั่วคราว (Staging Area) โดยยังไม่ได้รับการแปลงหรือวิเคราะห์เพิ่มเติม

๒. Transform (แปลงข้อมูล) โดยใช้การกระทำต่าง ๆ บนข้อมูลที่ถูกดึงมาเพื่อให้ข้อมูลเหมาะสมและมีคุณค่าสำหรับวิเคราะห์และรายงาน โดยการแปลงข้อมูลสามารถรวมการเรียงลำดับข้อมูล, การคำนวณ, การรวมข้อมูลจากแหล่งต่าง ๆ, การแปลงรูปแบบข้อมูล, และการทำความสะอาดข้อมูล (Data Cleansing) เพื่อแก้ไขข้อมูลที่ไม่ถูกต้องหรือไม่สมบูรณ์ ซึ่ง การแปลงข้อมูลสามารถทำได้โดยใช้ภาษาโปรแกรมหรือเครื่องมือ ETL อื่น ๆ ที่รองรับการประมวลผลข้อมูล

๓. Load...

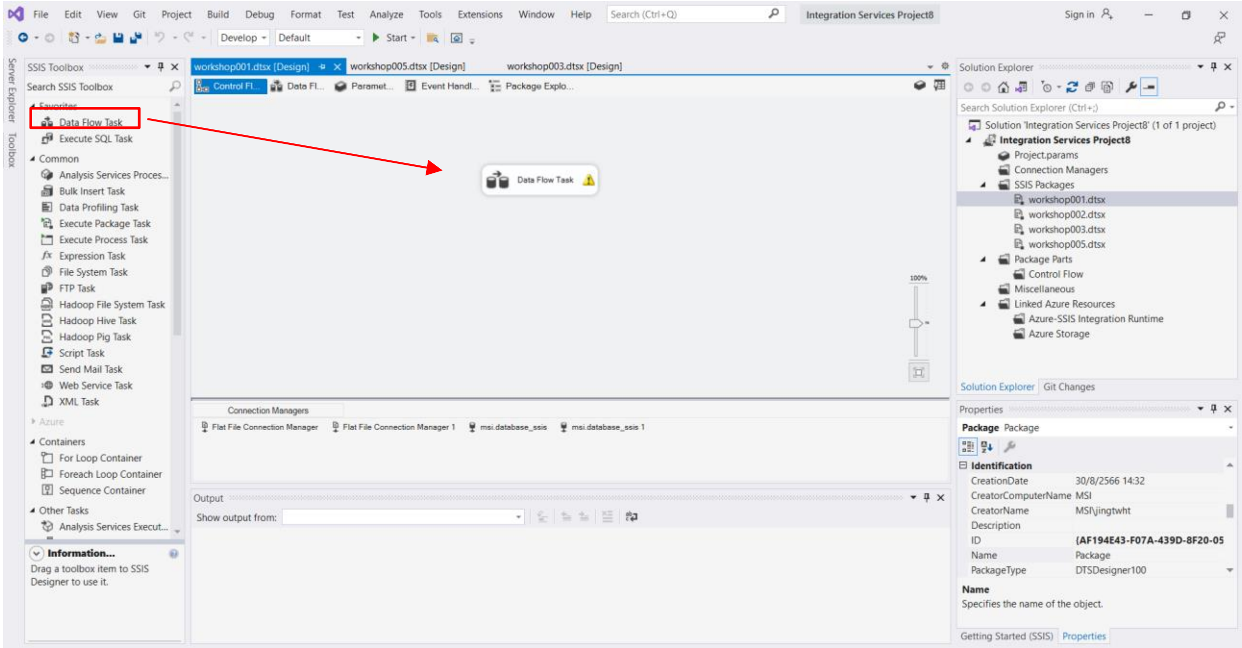
๓. Load (นำเข้าข้อมูล) เป็นขั้นตอนสุดท้าย โดยนำข้อมูลที่ผ่านการดึงและแปลงมาใส่ในระบบปลายทาง เช่น ฐานข้อมูล Data Warehouse, Data Mart, หรือระบบวิเคราะห์ข้อมูล การนำเข้าข้อมูลสามารถทำโดยเขียนข้อมูลลงในฐานข้อมูลต้นทางหรือใช้ข้อมูลของระบบ ETL ในการนำเข้า (Load) ข้อมูล โดยในกระบวนการนี้ ต้องกำหนดโครงสร้างข้อมูลในระบบปลายทางและการตั้งค่าการเข้าถึงฐานข้อมูลหรือระบบปลายทาง

ตัวอย่างการใช้งานโปรแกรม SQL Server Integration Service (SSIS)

๑. การนำข้อมูลไฟล์ประเภท Flat File ไปใส่ยัง MS SQL Server

๑) สร้าง Project SSIS ตั้งชื่อ workshop_๐๐๑.dtsx

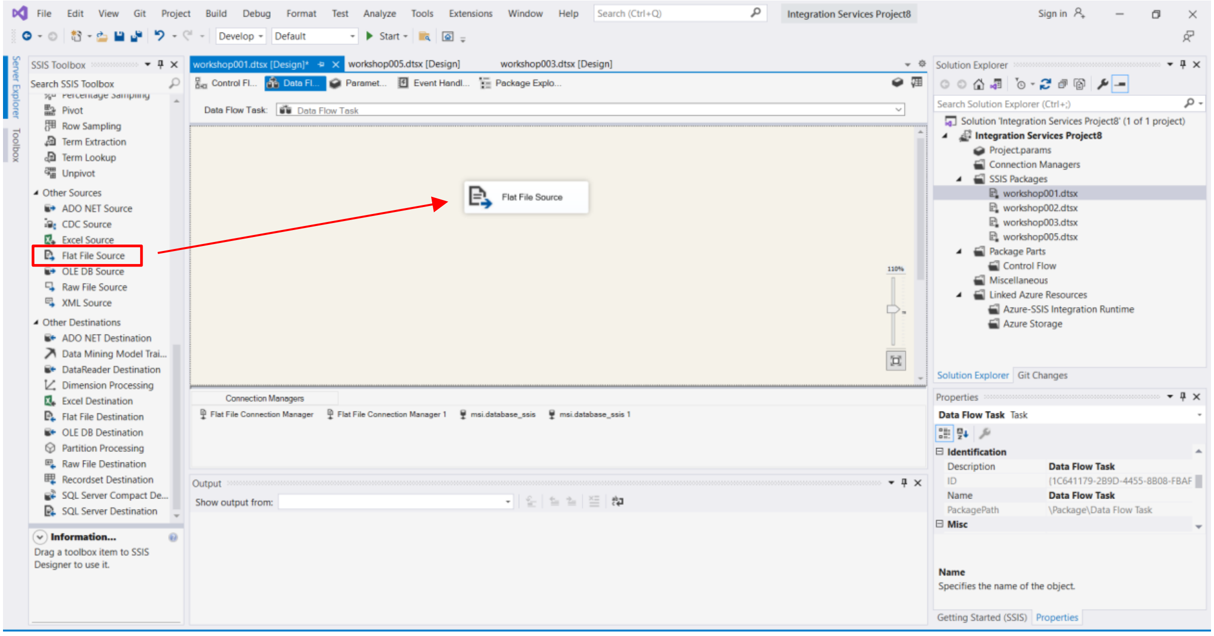
๒) ลากเครื่องมือ Data Flow จาก SSIS Toolbox มายัง Workshop Design



ภาพที่ ๑๗

๓) ไปที่ Data

๓) ไปที่ Data Flow กำหนด Source File ต้นทางเป็น Flat File



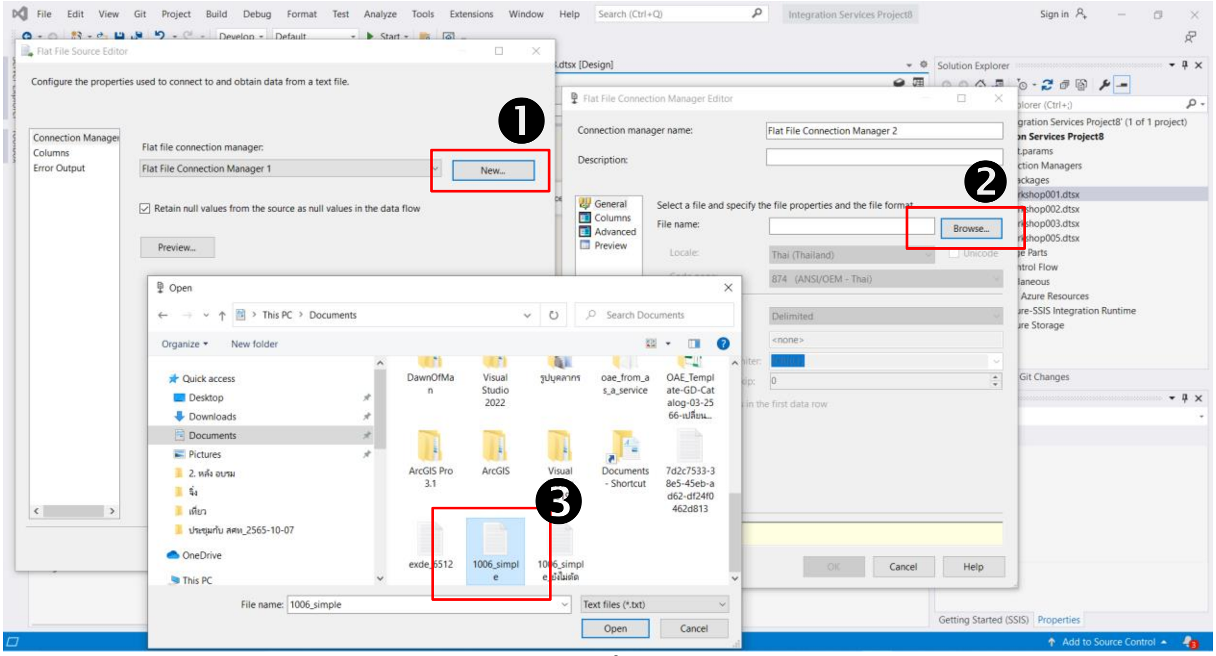
ภาพที่ ๑๘

๔) ทำการตั้งค่า Flat File Source ดังภาพที่ ๑๙ ดังนี้

❶ กด New เพื่อเพิ่มไฟล์ใหม่

❷ กด Browse เพื่อเลือกไฟล์ต้นทางที่ต้องการนำเข้า

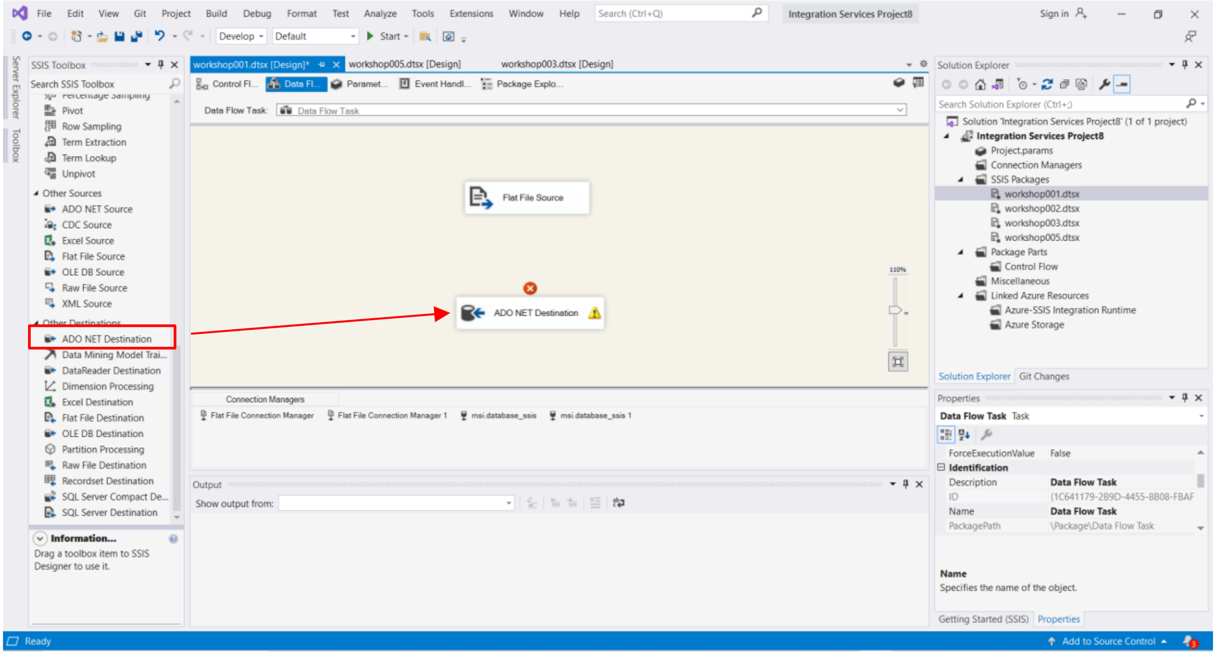
❸ จากนั้น เลือกไฟล์ที่ต้องการ



ภาพที่ ๑๙

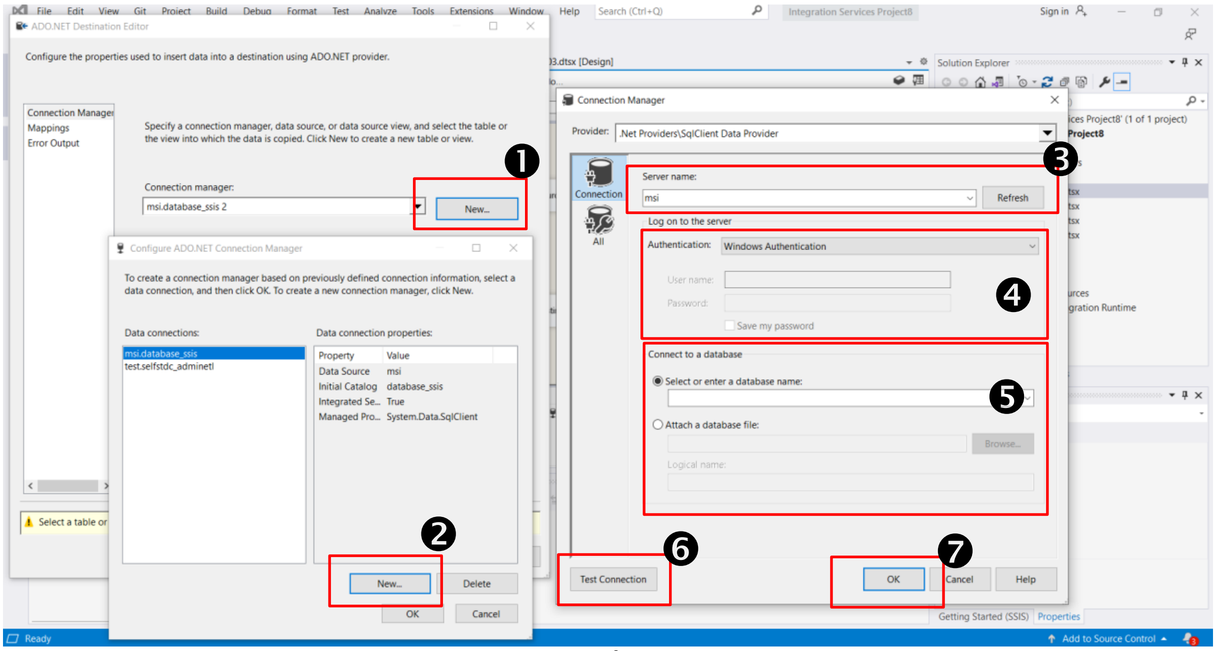
๕) กำหนด

๕) กำหนด Destination Source เป็น ADO NET Destination เนื่องจากต้องการเก็บลง Microsoft SQL Server จากนั้นลากเครื่องมือมายัง Workshop Design



ภาพที่ ๒๐

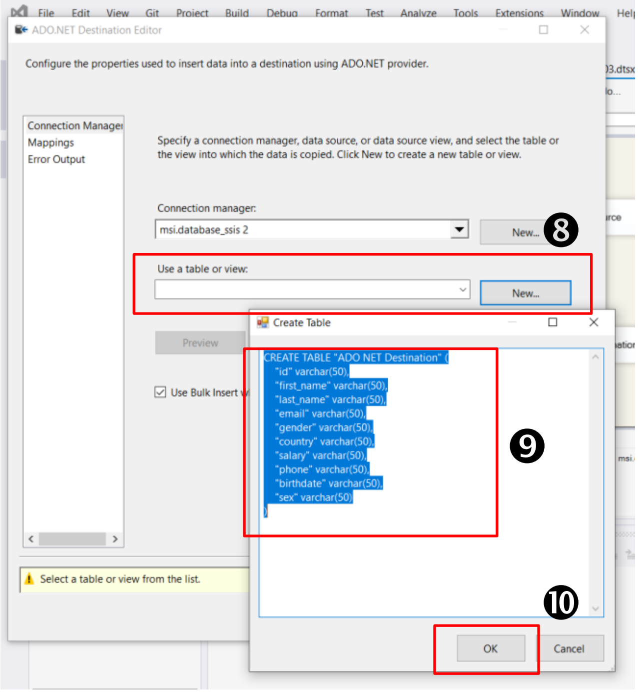
๖) ทำการตั้งค่า ADO NET Destination ดังภาพที่ ๒๑ และ ภาพที่ ๒๒ ดังนี้



ภาพที่ ๒๑

❶ เลือก New

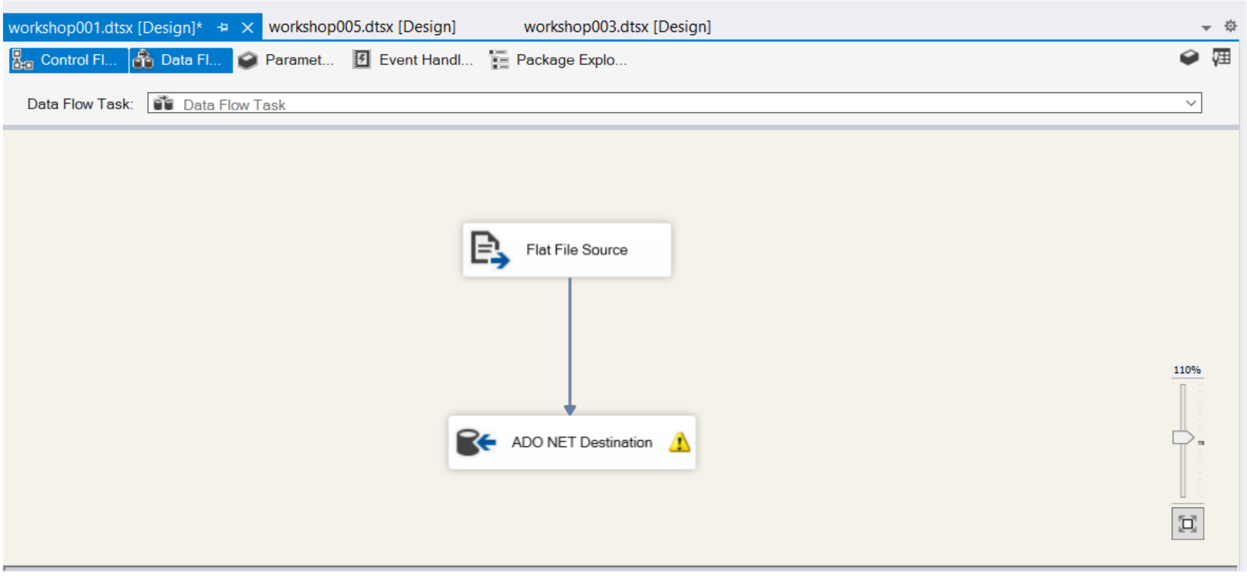
❶ เลือก New เพื่อเพิ่มการเชื่อมต่อใหม่
❷ จากนั้น เลือก New เพื่อเพิ่ม Database ใหม่
❸ ใส่ Server Name ที่ต้องการเชื่อมต่อ
❹ ใส่ Authentication ตามด้วย Username และ Password (หากมี)
❺ เลือก Database name ที่ต้องการเชื่อมต่อ เพื่อให้ข้อมูลไปเก็บที่ Database นั้น
❻ กด Test Connection เพื่อทดลองเชื่อมต่อ
❼ เมื่อลองทดสอบ หากเชื่อมต่อสำเร็จให้กด OK



❽ ทำการเลือก Table ที่ต้องการเก็บข้อมูล หากไม่มีให้กด New เพื่อเพิ่ม Table ใหม่
❾ เขียนคำสั่ง SQL เพื่อเพิ่ม Table ใหม่
❿ เมื่อเสร็จเรียบร้อยแล้ว ให้กด OK

ภาพที่ ๒๒

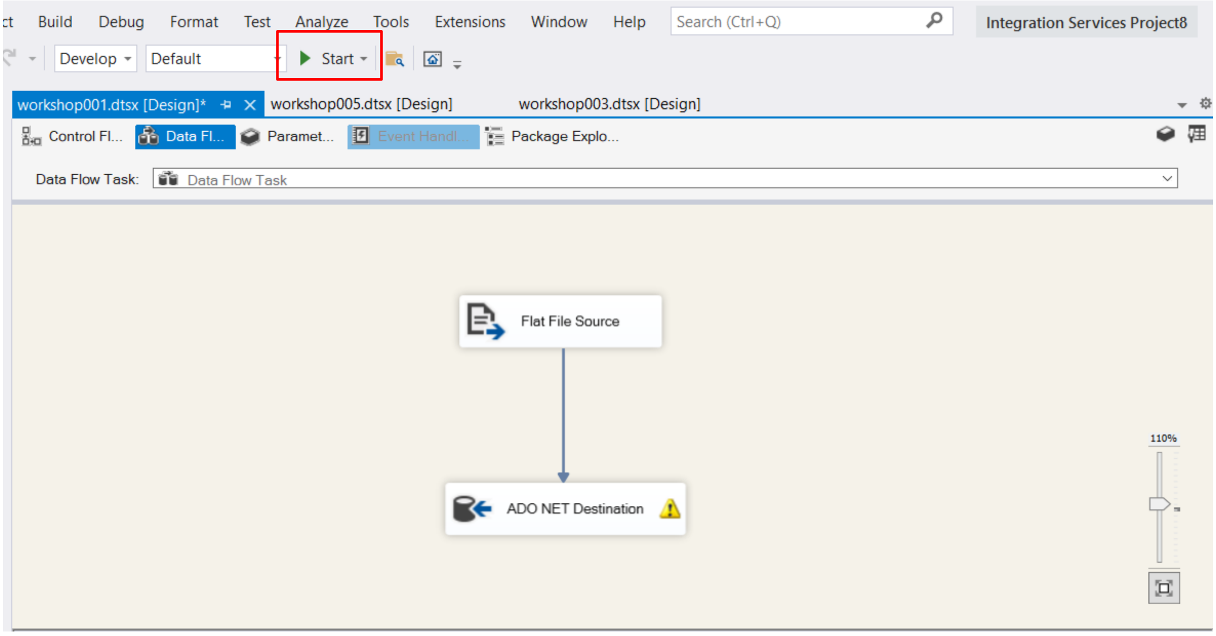
๗) จากนั้นลากเส้นเพื่อเชื่อม Workflow ให้สมบูรณ์ ดังภาพที่ ๒๓



ภาพที่ ๒๓

๘) กด Run...

๘) กด Run เพื่อดูผลลัพธ์ ดังภาพที่ ๒๔

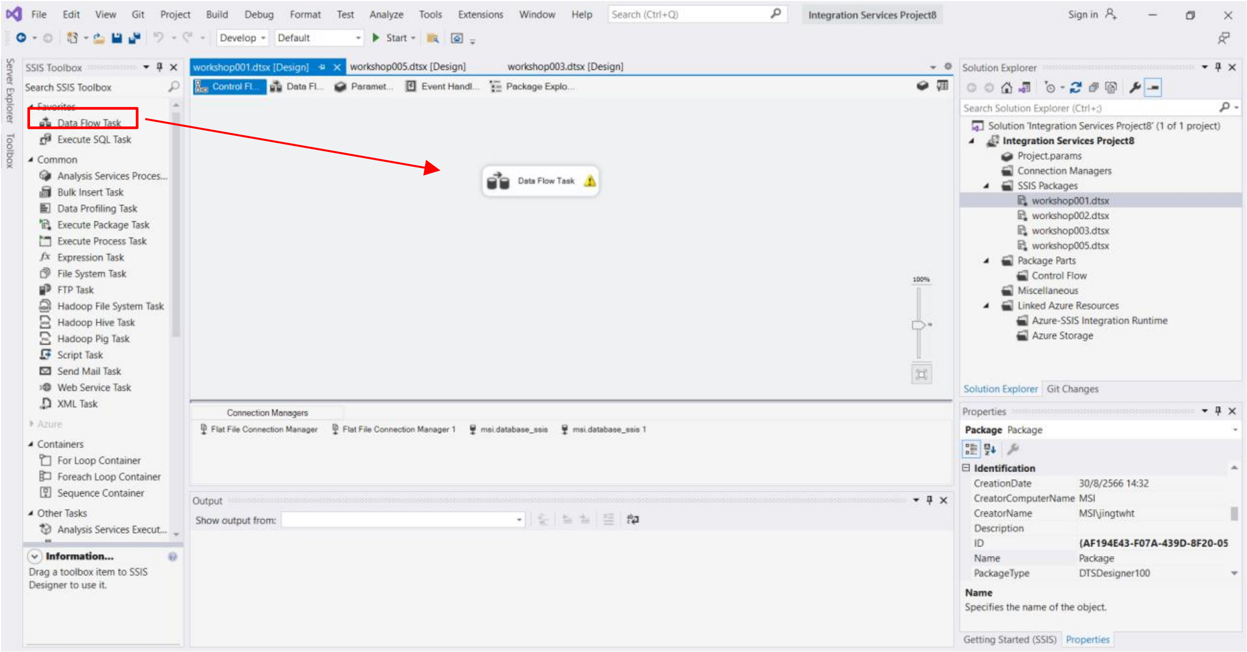


ภาพที่ ๒๔

๒. การนำข้อมูลจาก Table ในฐานข้อมูล MS SQL Server ส่งออกไปเป็น Flat File

๑) สร้าง Project SSIS ตั้งชื่อ workshop_oo๒.dtsx

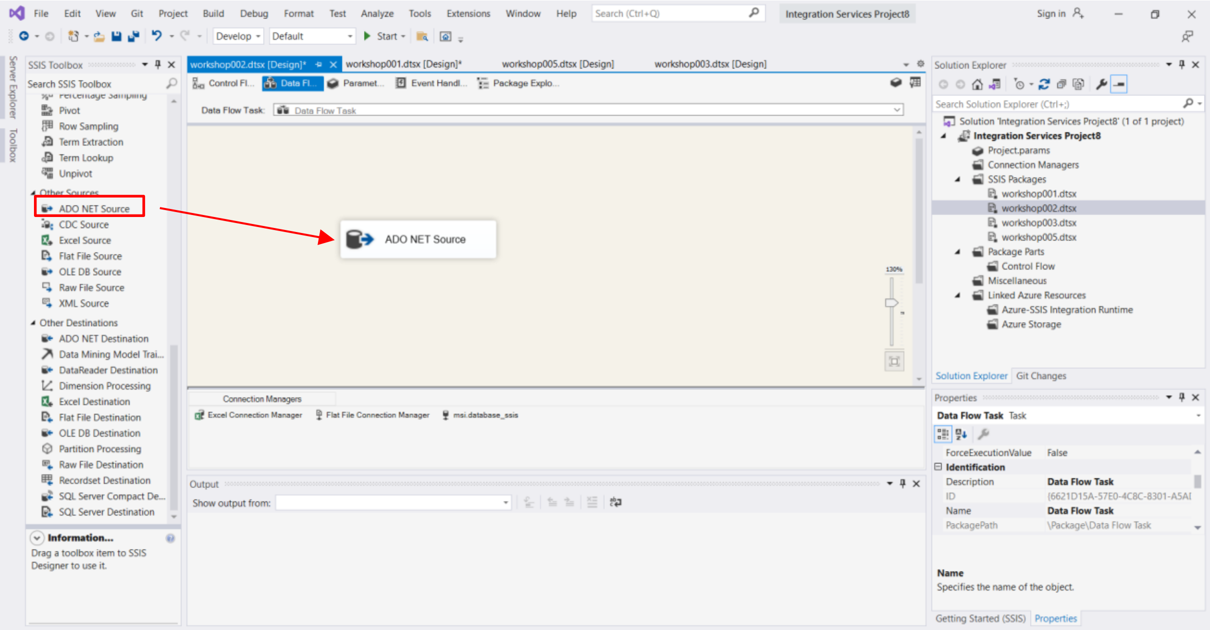
๒) ลากเครื่องมือ Data Flow จาก SSIS Toolbox มายัง Workshop Design



ภาพที่ ๒๕

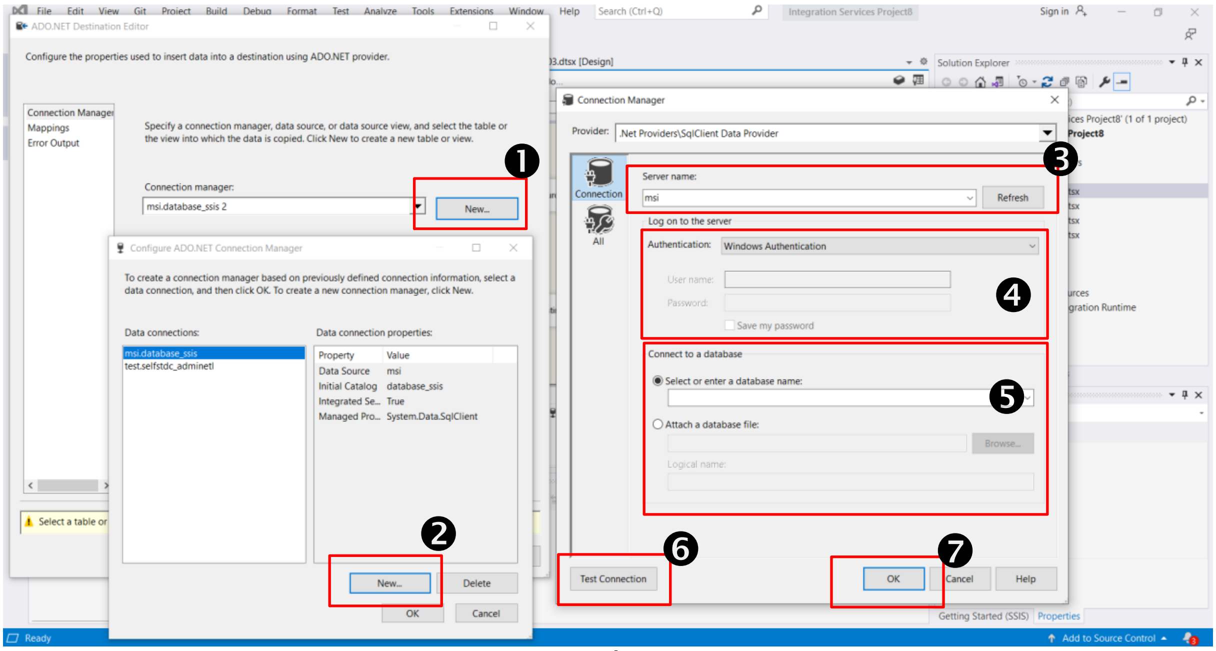
๓) ไปที่ Data...

๓) ไปที่ Data Flow กำหนด Source File ต้นทางเป็น ADO NET Source



ภาพที่ ๒๖

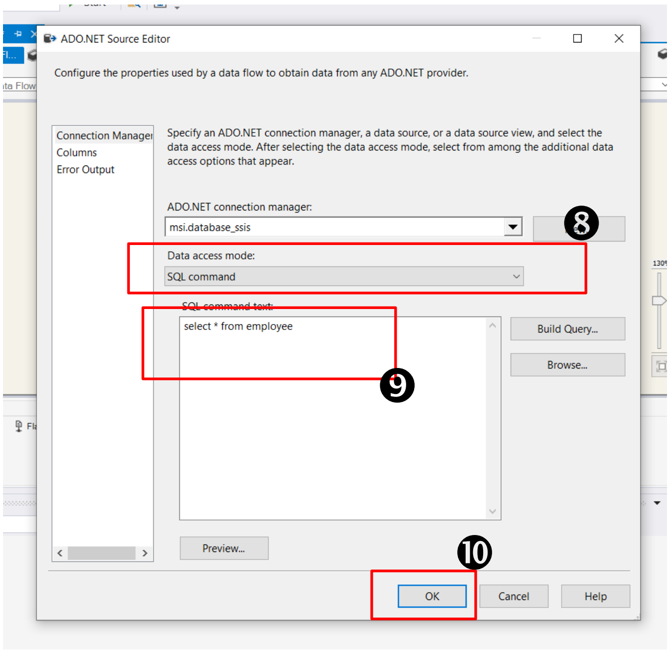
๔) ทำการตั้งค่า ADO NET Source ดังภาพที่ ๒๗ และ ภาพที่ ๒๘ ดังนี้



ภาพที่ ๒๗

❶ เลือก New เพื่อเพิ่มการเชื่อมต่อใหม่
❷ จากนั้น เลือก New เพื่อเพิ่ม Database ใหม่
❸ ใส่ Server Name ที่ต้องการเชื่อมต่อ
❹ ใส่ Authentication ตามด้วย Username และ Password (หากมี)
❺ เลือก Database name ที่ต้องการเชื่อมต่อ เพื่อไปดึงข้อมูลจาก Database นั้น
❻ กด Test Connection เพื่อทดลองเชื่อมต่อ
❼ เมื่อลองทดสอบ หากเชื่อมต่อสำเร็จให้กด OK

❽ ทำการเลือก...



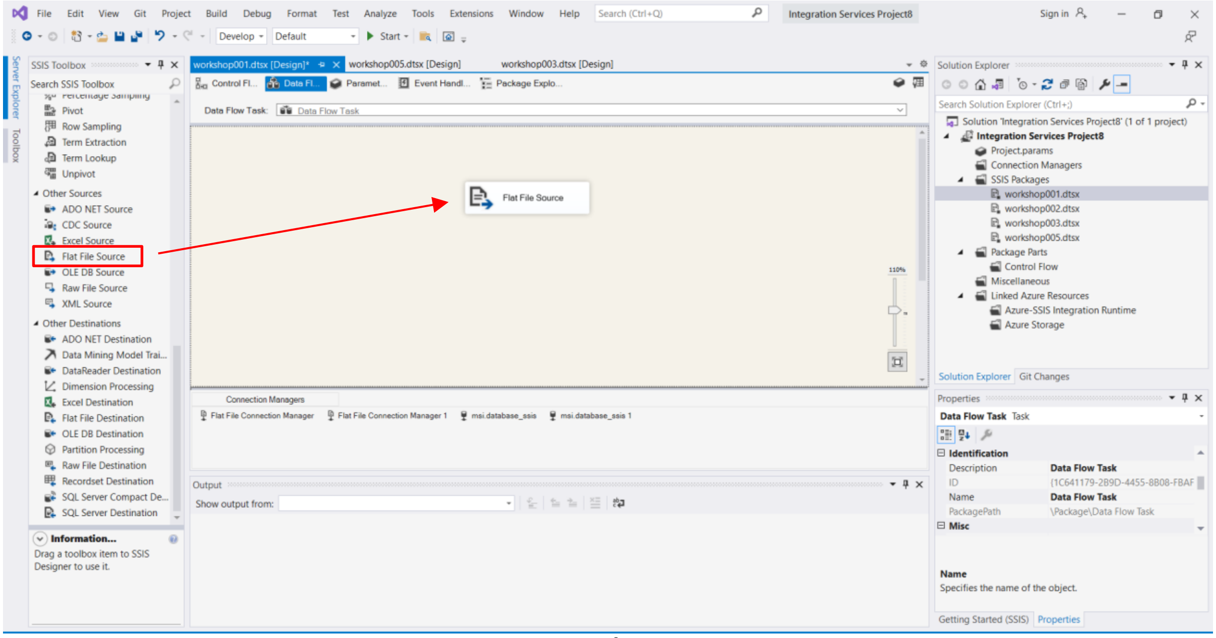
❽ ทำการเลือก Data access mode ในกรณีนี้เลือก SQL command

❾ เขียนคำสั่ง SQL เพื่อเลือกข้อมูลที่ต้องการ

❿ เมื่อเสร็จเรียบร้อยแล้ว ให้กด OK

ภาพที่ ๒๘

๕) จากนั้นกำหนด Destination File เป็น Flat File เนื่องจากต้องการไฟล์ชนิด CSV



ภาพที่ ๒๙

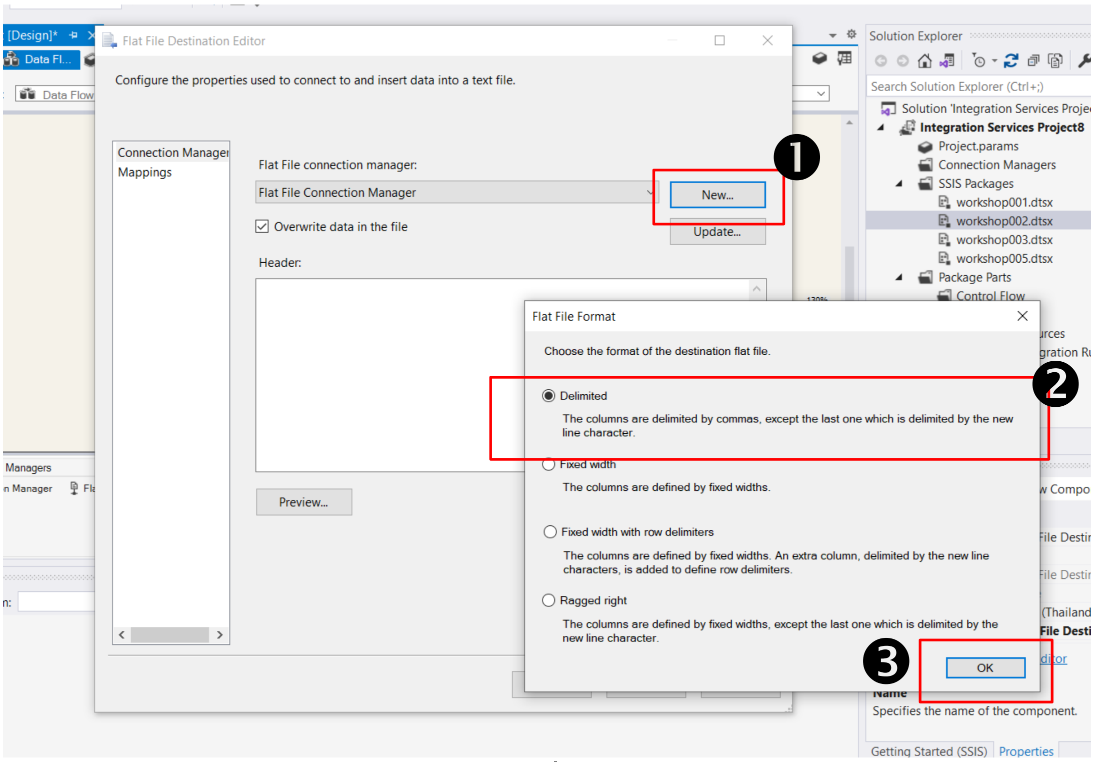
๖) ทำการตั้งค่า Flat File Source ดังภาพที่ ๓๐ และภาพที่ ๓๑ ดังนี้

❶ กด New เพื่อเพิ่มไฟล์ใหม่

❷ เลือก Delimited (มีตัวคั่น) เนื่องจากต้องการไฟล์ CSV

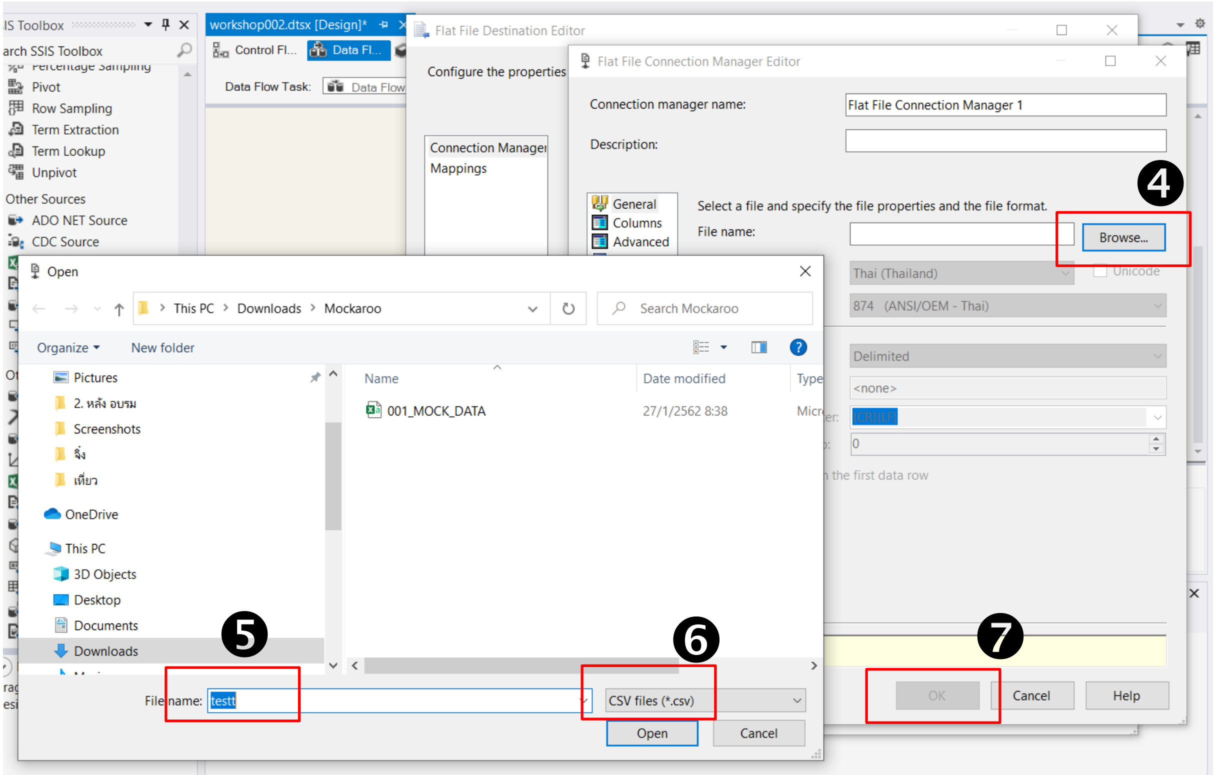
❸ จากนั้น กด OK

❹ กด Browse



ภาพที่ ๓๐

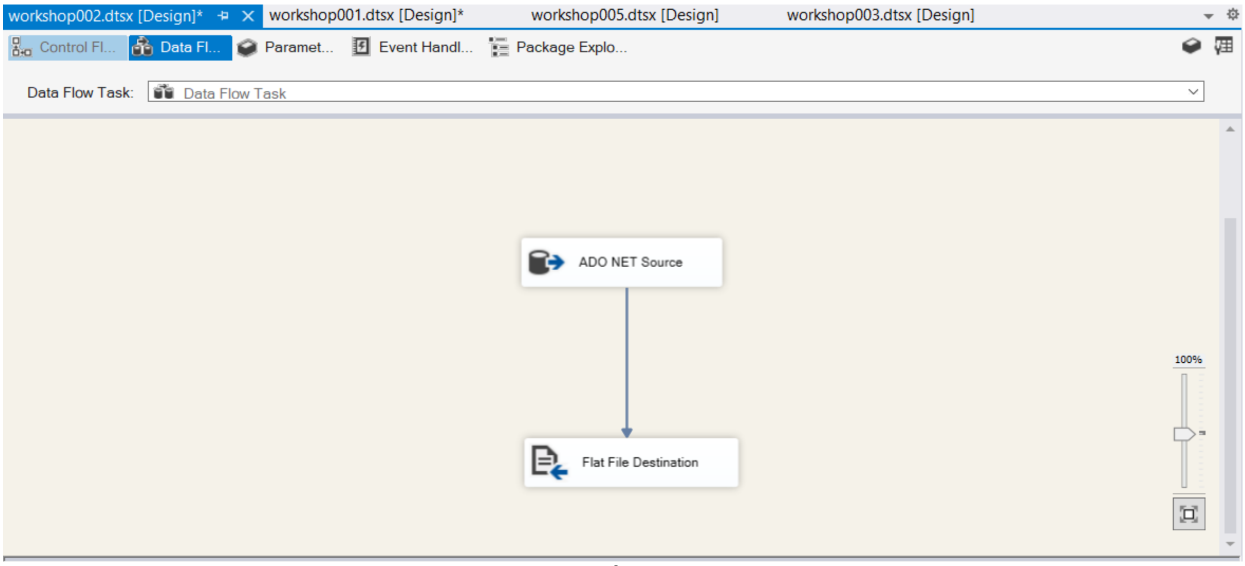
❹ กด Browse เพื่อ Save File
❺ ตั้งชื่อไฟล์
❻ เลือกชนิดไฟล์ที่ต้องการ Save ในกรณีนี้เลือกเป็น CSV
❼ จากนั้น กด OK



ภาพที่ ๓๑

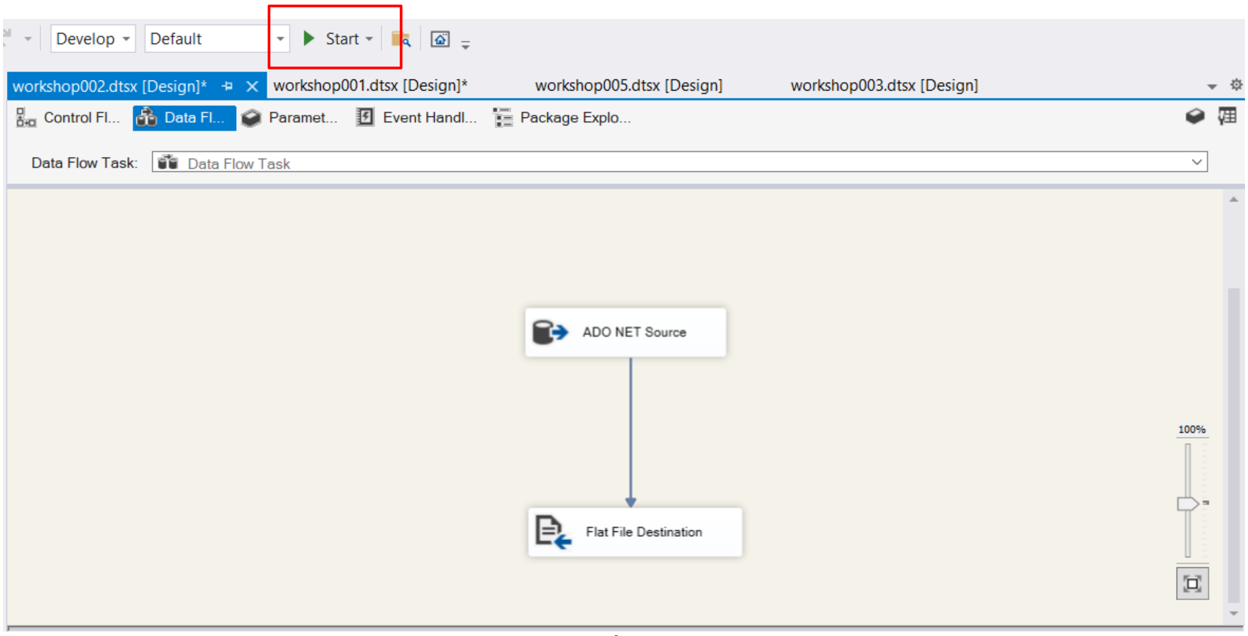
๗) จากนั้น...

๗) จากนั้นลากเส้นเพื่อเชื่อม Workflow ให้สมบูรณ์ ดังภาพที่ ๓๒



ภาพที่ ๓๒

๘) กด Run เพื่อดูผลลัพธ์ ดังภาพที่ ๓๓



ภาพที่ ๓๓

๒. การประเมินความพึงพอใจการเข้าร่วมอบรม

๒.๑ คุณสมบัติของผู้ตอบแบบสำรวจความพึงพอใจการเข้าร่วมอบรม (ตารางที่ ๑)

ผู้ตอบแบบสำรวจความพึงพอใจการเข้าร่วมอบรม ตอบแบบสำรวจผ่าน Google Form จำนวน ๑๐ คน ซึ่งประกอบด้วย

๒.๑.๑ เพศ ผู้ใช้บริการส่วนใหญ่เป็นเพศหญิง ร้อยละ ๕๐.๐๐ และเพศชาย ร้อยละ ๕๐.๐๐

๒.๑.๒ อายุ ส่วนใหญ่อยู่ในช่วง ๒๐ - ๓๐ ปี ร้อยละ ๖๐.๐๐ รองลงมาอายุอยู่ในช่วง ๓๑ – ๔๐ ปี ร้อยละ ๓๐.๐๐ และอายุอยู่ในช่วง ๔๑ – ๕๐ ปี ร้อยละ ๑๐.๐๐

๒.๑.๓ ระดับการศึกษา ส่วนใหญ่จบการศึกษาระดับปริญญาตรี ร้อยละ ๙๐.๐๐ รองลงมาระดับปริญญาโท ร้อยละ ๑๐.๐๐

๒.๑.๔ ตำแหน่ง ส่วนใหญ่เป็นนักวิชาการคอมพิวเตอร์ปฏิบัติการ ร้อยละ ๗๐.๐๐ และรองลงมาเป็นนักวิชาการคอมพิวเตอร์ ร้อยละ ๓๐.๐๐

๒.๑.๕ หน่วยงาน ภายใต้สังกัดศูนย์ข้อมูลเกษตรแห่งชาติ ร้อยละ ๕๐.๐๐ และสังกัดศูนย์สารสนเทศการเกษตร ร้อยละ ๕๐.๐๐

ตารางที่ ๑ คุณสมบัติของผู้ตอบแบบสำรวจความพึงพอใจการเข้าร่วมอบรม

รายละเอียด	ร้อยละ
เพศ	
- ชาย	๕๐.๐๐
- หญิง	๕๐.๐๐
อายุ	
- ๒๐ – ๓๐ ปี	๖๐.๐๐
- ๓๑ – ๔๐ ปี	๓๐.๐๐
- ๔๑ – ๕๐ ปี	๑๐.๐๐
- ๕๑ – ๖๐ ปี	๐.๐๐
ระดับการศึกษา	
- ต่ำกว่าปริญญาตรี	๐.๐๐
- ปริญญาตรี	๙๐.๐๐
- ปริญญาโท	๑๐.๐๐
- ปริญญาเอก	๐.๐๐
ตำแหน่ง	
- นักวิชาการคอมพิวเตอร์ปฏิบัติการ	๗๐.๐๐
- นักวิชาการคอมพิวเตอร์	๓๐.๐๐
หน่วยงาน	
- ศูนย์ข้อมูลเกษตรแห่งชาติ	๕๐.๐๐
- ศูนย์สารสนเทศการเกษตร	๕๐.๐๐

๒.๒ ผลการ...

๒.๒ ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรม (ตารางที่ ๒)

การประเมินความพึงพอใจของผู้เข้าร่วมอบรมที่มีต่อปัจจัยที่เกี่ยวข้องกับการอบรม โดยใช้แบบสอบถาม ซึ่งผลจากการประเมินจะใช้เป็นแนวทางในการพัฒนาคุณภาพการจัดอบรมให้ดียิ่งขึ้นต่อไป โดยมีเกณฑ์ในการประเมิน ดังนี้

คะแนน	การประเมิน	การแปลผล
๕	๔.๐๑ – ๕.๐๐	มากที่สุด
๔	๓.๐๑ – ๔.๐๐	มาก
๓	๒.๐๑ – ๓.๐๐	ปานกลาง
๒	๑.๐๑ - ๒.๐๐	น้อย
๑	๐.๐๐ – ๑.๐๐	น้อยที่สุด

ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรมฯ ในประเด็นต่าง ๆ อาทิ เช่น ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร เทคนิคหรือวิธีการถ่ายทอดความรู้ ความเหมาะสมของเนื้อหาในการอบรม การสรุปเนื้อหาสาระได้อย่างชัดเจน ความรู้ความเข้าใจเนื้อหา ก่อน-หลัง ความสามารถในการนำความรู้ ความเข้าใจ และประสบการณ์ที่ได้จากการเข้าร่วมการอบรม ไปปรับใช้ในการปฏิบัติงานและถ่ายทอดให้เพื่อนร่วมงาน ระยะเวลาการอบรม การประสานงานและการอำนวยความสะดวก พบว่า ในภาพรวมผู้รับบริการส่วนใหญ่มีความพึงพอใจอยู่ในระดับมาก คะแนนเฉลี่ย ๔.๒๖ ร้อยละ ๘๕.๒๐ โดยมีรายละเอียดตามประเด็นต่าง ๆ ดังนี้

ตารางที่ ๒ ผลการประเมินความพึงพอใจของผู้เข้าร่วมอบรม

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๑. ภาพรวมของวิทยากร						**๔.๖๐**
๑.๑ ความรอบรู้ในประเด็นเนื้อหาวิชาของวิทยากร	๖๐.๐๐ (๖)	๔๐.๐๐ (๔)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๐
๑.๒ เทคนิคหรือวิธีการถ่ายทอดความรู้	๖๐.๐๐ (๖)	๔๐.๐๐ (๔)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๐
๑.๓ ความเหมาะสมของเนื้อหาในการอบรม	๖๐.๐๐ (๖)	๔๐.๐๐ (๔)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๐
๑.๔ การสรุปเนื้อหาสาระได้อย่างชัดเจน	๖๐.๐๐ (๖)	๔๐.๐๐ (๔)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๖๐

ประเด็น...

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๒. ความรู้ความเข้าใจเกี่ยวกับเนื้อหาวิชาที่เข้าร่วมอบรม						**๓.๘๐**
๒.๑ ก่อน การเข้าร่วมอบรม ท่านมีความรู้ความเข้าใจเพียงใด	๑๐.๐๐ (๑)	๑๐.๐๐ (๑)	๓๐.๐๐ (๓)	๓๐.๐๐ (๓)	๒๐.๐๐ (๒)	๒.๖๐
๒.๒ หลัง การเข้าร่วมอบรม ท่านมีความรู้ความเข้าใจเพียงใด	๑๐.๐๐ (๑)	๖๐.๐๐ (๖)	๓๐.๐๐ (๓)	๐.๐๐	๐.๐๐	๓.๘๐
๒.๓ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมไปปรับใช้ในการปฏิบัติงานได้เพียงใด	๐.๐๐	๗๐.๐๐ (๗)	๓๐.๐๐ (๓)	๐.๐๐	๐.๐๐	๓.๗๐
๒.๔ ท่านสามารถนำความรู้ ความเข้าใจและประสบการณ์ที่ได้จากการเข้าร่วมอบรมไปถ่ายทอดให้เพื่อนร่วมงานได้เพียงใด	๑๐.๐๐ (๑)	๗๐.๐๐ (๗)	๒๐.๐๐ (๒)	๐.๐๐	๐.๐๐	๓.๙๐

ประเด็น	ระดับความพึงพอใจ (ร้อยละ)					
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	คะแนนเฉลี่ย
๓. ภาพรวมการจัดการอบรม						**๔.๓๗**
๓.๑ ระยะเวลาการอบรม	๓๐.๐๐ (๓)	๗๐.๐๐ (๗)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๓๐
๓.๒ การประสานงานและการอำนวยความสะดวก	๔๐.๐๐ (๔)	๖๐.๐๐ (๖)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๔๐
๓.๓ ความพึงพอใจในภาพรวมของการอบรม	๔๐.๐๐ (๔)	๖๐.๐๐ (๖)	๐.๐๐	๐.๐๐	๐.๐๐	๔.๔๐
รวม						**๔.๒๖**

๓. ข้อคิดเห็น...

๓. ข้อคิดเห็น

๓.๑ ข้อดีในการอบรมฯ

๑) ได้รับความรู้ใหม่ ๆ

๒) สามารถนำความรู้ที่ได้จากการอบรมไปประยุกต์ใช้ในการทำงาน

๓) ได้รับความรู้ต่าง ๆ ทางด้าน data engineer และการใช้เครื่องมือ และเทคนิค เพื่อใช้ในการจัดการข้อมูล Big Data

๔) อาจารย์สอนดีมาก

๕) ได้นำความรู้จากการอบรมฯ มาใช้ปรับใช้ในการทำงาน

๖) ได้เรียนรู้เทคโนโลยี Cloud และได้ความรู้มากขึ้น

๗) ได้เรียนรู้เทคโนโลยีและเครื่องมือใหม่ ๆ ที่มาช่วยในเรื่องการบริหารจัดการข้อมูล

๘) การประยุกต์ใช้เครื่องมือในการจัดการเกี่ยวกบข้อมูลขนาดใหญ่

๓.๒ ข้อควรปรับปรุงในการอบรมฯ

๑) ต้องการให้เพิ่มระยะเวลาการอบรมในช่วงการทำ LAB ให้มากกว่านี้

๔. ข้อเสนอแนะอื่น ๆ

ไม่มี

....