



วิธี Export ข้อมูลขนาดใหญ่ที่รวดเร็ว

ภญ. ภัทราธร คุณาภรณ์ธรรม

งานสารสนเทศเพื่อการวิเคราะห์ข้อมูลทางสุขภาพ ฝ่ายสารสนเทศ

1. สภาพก่อนปรับปรุง

กระบวนการทำงานแบบเดิมมี ดังนี้



ปัญหา: ถ้าข้อมูลมีจำนวนแถวมากกว่า 150,000 แถว ต้องกรองข้อมูลแล้วคลิกปุ่มส่งออกทีละปีหรือทีละเดือนจนครบทั้งหมด

ตัวอย่าง: จากการเตรียมข้อมูลการบันทึกโรค พ.ศ. 2564 จำนวน 3,050,511 แถว รวมใช้เวลาส่งออกข้อมูลทั้งหมด 21 ไฟล์ นาน 40 นาที (ขั้นตอนที่ 1-3 ใช้เวลา 10 นาที และขั้นตอนที่ 4 ใช้เวลา 30 นาที)

วัตถุประสงค์: เพื่อหาวิธีส่งออกข้อมูลขนาดใหญ่ที่รวดเร็วมากขึ้น

3.ผลที่ได้จากการปรับปรุง/ผลลัพธ์/ผลดำเนินการโครงการ

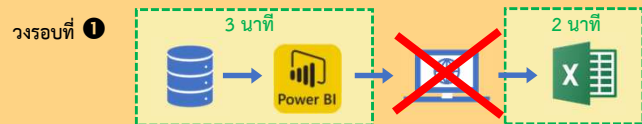
ตัวชี้วัด	เป้าหมาย	ก่อน	หลังดำเนินการ (เดือน / ปี)		
			ส.ค.64	ก.ย.64	ต.ค.64
ระยะเวลาเตรียมข้อมูล (นาที)	<40 นาที	40	5 (▼87.5%)	7 (▼82.5%)	6 (▼85%)
จำนวนขั้นตอนในการเตรียมข้อมูล (ขั้น)	≤4 ขั้น	4 ขั้น	3 (▼25%)	3 (▼25%)	3 (▼25%)
การใช้กับข้อมูลที่มีสัญลักษณ์ comma	ใช้ได้	ใช้ได้	-	ใช้ไม่ได้	ใช้ได้

สรุปได้ว่า การเขียนคิวรีเพื่อแบ่งไฟล์ข้อมูลให้เล็กลง ร่วมกับการประยุกต์ใช้ R-script ช่วยส่งออกข้อมูลเป็นไฟล์ ".txt" เป็นวิธีที่เหมาะสมที่สุด เพราะ ลดระยะเวลาการเตรียมข้อมูลเหลือ 6 นาที ลดขั้นตอนเหลือ 3 ขั้น และสามารถใช้งานข้อมูลที่มีสัญลักษณ์ comma ได้

4.การจัดทำมาตรฐาน

- ☐ ยังไม่ได้จัดทำมาตรฐานการทำงาน(WI) / มาตรฐานการใช้งาน
- ☒ มีการจัดทำมาตรฐานการทำงาน(WI) / มาตรฐานการใช้งาน

2.กระบวนการปรับปรุง/สภาพหลังปรับปรุง



P: หาวิธีปรับปรุงขั้นตอนที่ 4 เพราะเป็นขั้นตอนที่ใช้เวลานานที่สุด

D: ใช้วิธีส่งออกข้อมูลด้วย DAX Studio ซึ่งเป็นโปรแกรมช่วยส่งออกข้อมูลจาก MS Power BI Desktop โดยตรงแทนวิธีเดิม จะได้เป็นไฟล์สกุล ".csv"

C: ทดสอบเตรียมข้อมูลการบันทึกโรค พ.ศ. 2564 และส่งออกด้วย DAX Studio

A: พบว่า ช่วยลดขั้นตอนที่ 3 และใช้เวลาส่งออกข้อมูลทั้งหมดนาน 5 นาที (ลดลง 87.5%)

****แต่ไฟล์ข้อมูลมีขนาดใหญ่ ไม่สามารถเปิดด้วย MS Excel**



P: หาวิธีแบ่งไฟล์ข้อมูลให้เล็กลง เพื่อให้สามารถเปิดด้วย MS Excel ได้

D: ใช้วิธีเขียนคิวรี (query) เพิ่มในขั้นตอนที่ 2 เพื่อแบ่งไฟล์ข้อมูลให้เล็กลง

C: เตรียมข้อมูลโรค พ.ศ. 2564 แบ่งเป็น 4 ไตรมาส และส่งออกด้วย DAX Studio

A: พบว่า ใช้เวลาส่งออกข้อมูลทั้งหมด 4 ไฟล์ นาน 7 นาที (ลดลง 82.5%)

****แต่การใช้ DAX Studio ไม่เหมาะกับการส่งออกข้อมูลที่มีสัญลักษณ์ comma เมื่อเปิดไฟล์ด้วย MS Excel จะได้ข้อมูลที่ไม่ถูกต้อง ข้อมูลคลาดเคลื่อน**



P: หาวิธีส่งออกข้อมูลแบบอื่น แทนการใช้ DAX Studio

D: ใช้วิธีเขียนคิวรีขั้นสูงเพิ่มในขั้นตอนที่ 2 โดยประยุกต์ใช้ R-script ซึ่งเป็นภาษาคอมพิวเตอร์อีกชนิดหนึ่งช่วยให้ข้อมูลส่งออกจาก MS Power BI Desktop เป็นไฟล์ ".txt" แทน

C: เตรียมข้อมูลโรค พ.ศ. 2564 แบ่งข้อมูลเป็น 4 ไตรมาส และส่งออกด้วย R-script รวมถึงทดสอบกับข้อมูลที่มีสัญลักษณ์ comma

A: พบว่า ใช้เวลาส่งออกข้อมูลทั้งหมด 4 ไฟล์ นาน 6 นาที (ลดลง 85%) และสามารถเปิดไฟล์ที่มีสัญลักษณ์ comma ด้วย MS Excel ได้อย่างถูกต้อง

5.หลังเสร็จสิ้นโครงการท่านดำเนินการในการขยายผลอย่างไร

- ☐ ใช้เฉพาะตนเองเท่านั้น
- ☒ ใช้เฉพาะในหน่วยงานของท่าน
- ☐ ใช้ในหน่วยงานอื่นภายในคณะฯ