**งาน คำศัพท์**

1. **ให้นักศึกษา อธิบาย ความหมายของ keywords ที่กำหนด พร้อมยกตัวอย่าง  
   a. Big Data:** คือชุดข้อมูลที่มีปริมาณมากถึงขนาดที่เกือบไม่สามารถจัดการได้โดยเครื่องมือและวิธีการที่มีอยู่ในปัจจุบัน ประกอบไปด้วยข้อมูลที่มาจากแหล่งต่าง ๆ เช่นเครือข่ายสังคมออนไลน์ ระบบเซ็นเซอร์ ข้อมูลการซื้อขาย และอื่น ๆ  
   **ตัวอย่าง Big Data:** บริษัทเครือข่ายสังคมออนไลน์ (เช่น Facebook, Twitter) มีข้อมูลจำนวนมหาศาลเกี่ยวกับการกระทำของผู้ใช้งานแต่ละคน ซึ่งถูกนำมาวิเคราะห์เพื่อค้นหาแนวโน้มและพฤติกรรมของผู้ใช้งาน  
    **b. Data Warehouse:** เป็นฐานข้อมูลที่ถูกออกแบบมาเพื่อเก็บรวบรวมและจัดเรียงข้อมูลจากแหล่งต่าง ๆ ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์และรายงาน โดยมักใช้ในการรวมข้อมูลที่มาจากหลายแหล่งเพื่อให้เกิดความเข้าใจในมิติต่าง ๆ  
   **ตัวอย่าง Data Warehouse:** บริษัทค้าปลีกใช้คลังข้อมูลเพื่อเก็บข้อมูลการขายและสต็อกสินค้าจากสาขาทั่วประเทศ เพื่อใช้ในการวิเคราะห์การขายและวางแผนการสั่งซื้อสินค้า  
    **c. Data Lake:** เป็นพื้นที่ที่ใช้เก็บข้อมูลที่มีความหลากหลายและมาจากแหล่งต่าง ๆ ในรูปแบบเดียวกัน ไม่จำเป็นต้องทำการจัดเรียงหรือแปลงข้อมูลก่อน มักจะถูกใช้สำหรับการเก็บข้อมูลที่ไม่แน่นอนว่าจะใช้ในการวิเคราะห์หรือกระบวนการไหน  
   **ตัวอย่าง Data Lake:** บริษัทที่พัฒนาแอปพลิเคชันมือถือสะสมข้อมูลการใช้งานของผู้ใช้งาน ได้จัดเก็บข้อมูลเหล่านี้ในรูปแบบ Data Lake เพื่อให้ง่ายต่อการวิเคราะห์และค้นหาข้อมูลเชิงลึก  
    **d. Lakehouse:** คือแนวคิดที่รวมคุณสมบัติของ Data Warehouse และ Data Lake โดยเน้นการจัดเก็บข้อมูลในรูปแบบเหมือน Data Lake พร้อมทั้งให้ความสำคัญกับความมั่นคงและความเป็นระเบียบเพื่อให้สามารถนำข้อมูลมาใช้งานได้หลากหลาย  
   **ตัวอย่าง Lakehouse:** องค์กรที่พัฒนาแพลตฟอร์มการซื้อขายหลายแห่งใช้แนวคิด Lakehouse เพื่อรวมข้อมูลการซื้อขายในรูปแบบของ Data Lake แต่ยังรักษาคุณสมบัติ ACID สำหรับความมั่นคงและความถูกต้องของข้อมูล **e. ACID Transaction:** เป็นแนวคิดในการจัดการข้อมูลในฐานข้อมูลที่เน้นความมั่นคงและความเรียงระเบียบ มาจากคำว่า Atomicity, Consistency, Isolation, และ Durability (ACID) ที่เป็นคุณสมบัติหลักในการจัดการธุรกรรมที่มั่นคง  
   **ตัวอย่าง ACID Transaction:** ในระบบการทำธุรกรรมออนไลน์ เมื่อลูกค้าทำการสั่งซื้อสินค้า เว็บไซต์จะทำการตัดสต็อกสินค้าและบันทึกการทำธุรกรรมในฐานข้อมูล โดยใช้ระบบธุรกรรม ACID เพื่อให้แน่ใจว่าข้อมูลถูกบันทึกและจัดการอย่างถูกต้อง  
    **f. Data Scientist:** คือผู้ที่มีความเชี่ยวชาญในการวิเคราะห์ข้อมูลเพื่อหาความรู้และแนวโน้ม มักใช้เครื่องมือทางสถิติ คณิตศาสตร์ และการเรียนรู้ของเครื่องเพื่อดึงความหมายออกมาจากข้อมูล  
   **ตัวอย่าง Data Scientist:** นักวิทยาศาสตร์ข้อมูลที่ทำงานในบริษัทการเงินอาจทำการวิเคราะห์ข้อมูลการซื้อ-ขายหุ้นเพื่อค้นหาแนวโน้มและโอกาสในการลงทุน  
     
   **g. Data Engineer:** คือผู้ที่มีความเชี่ยวชาญในการออกแบบ สร้าง และบำรุงรักษาโครงสร้างข้อมูล รวมถึงการจัดการข้อมูลที่จำเป็นสำหรับการวิเคราะห์และการใช้งาน  
   **ตัวอย่าง Data Engineer:** วิศวกรข้อมูลที่ทำงานในบริษัทโซฟต์แวร์อาจพัฒนาและบำรุงรักษาระบบฐานข้อมูลสำหรับแอปพลิเคชันที่มีผู้ใช้งานจำนวนมาก  
    **h. Apache Spark:** เป็นโครงการโอเพนซอร์สที่ใช้สำหรับการประมวลผลข้อมูลขนาดใหญ่แบบพร้อมกัน (parallel processing) โดยรวมความสามารถในการทำงานกับข้อมูลโครงสร้างและข้อมูลที่ไม่มีโครงสร้าง นอกจากนี้ยังมีคลังข้อมูลชื่อว่า Resilient Distributed Dataset (RDD) เพื่อจัดการข้อมูลที่ถูกแบ่งแยกและจัดการในแต่ละโหนดของระบบ  
   **ตัวอย่าง Apache Spark:** บริษัทที่ให้บริการออนไลน์สตรีมมิ่งวิดีโออาจใช้ Apache Spark ในการประมวลผลและวิเคราะห์ข้อมูลการรับชมแบบเรียลไทม์  
    **i. Spark SQL:** เป็นโมดูลใน Apache Spark ที่ใช้ในการคิวรีและการจัดการข้อมูลโครงสร้าง (structured data) โดยใช้ภาษา SQL สามารถทำการคิวรีข้อมูลในรูปแบบตารางและแบบโครงสร้างได้  
   **ตัวอย่าง Spark SQL:** นักวิเคราะห์ข้อมูลในบริษัทการเงินอาจใช้ Spark SQL เพื่อคิวรีและวิเคราะห์ข้อมูลทางการเงินในรูปแบบของตารางเพื่อตัดสินใจในการลงทุนหรือการจัดการทางการเงิน
2. **อธิบายความหมายของคุณลักษณะของ Big Data  
   รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ขาว

   คำอธิบายที่สร้างโดยอัตโนมัติ**

**Volume:** หมายถึงปริมาณข้อมูลที่มากมายและเพิ่มขึ้นอย่างรวดเร็ว ข้อมูลที่ถูกสร้างและเก็บรวบรวมจากแหล่งต่าง ๆ เช่นเครือข่ายสังคมออนไลน์ อุปกรณ์ IoT (Internet of Things) และระบบเซ็นเซอร์ ในยุคดิจิตอลปัจจุบันเรามีข้อมูลที่มากเหลืออย่างมากที่สามารถใช้ในการวิเคราะห์และตัดสินใจได้

**Velocity:** หมายถึงการเกิดข้อมูลและการเคลื่อนไหวของข้อมูลอย่างรวดเร็ว ข้อมูลที่เกิดขึ้นในปัจจุบันมีการเปลี่ยนแปลงอย่างรวดเร็วและต้องการการประมวลผลทันในเวลาเรียลไทม์หรือใกล้เคียง

**Variety:** หมายถึงความหลากหลายในรูปแบบของข้อมูล ไม่เพียงแค่ข้อมูลที่มีโครงสร้างเช่นฐานข้อมูลที่เป็นตาราง แต่ยังรวมถึงข้อมูลที่ไม่มีโครงสร้างเช่นข้อความ ภาพถ่าย วิดีโอ และข้อมูลชนิดอื่น ๆ

**Veracity:** หมายถึงความถูกต้องและความน่าเชื่อถือของข้อมูล ข้อมูลใน Big Data มีความน่าเชื่อถือที่ต้องพิจารณาเนื่องจากมาจากแหล่งต่าง ๆ และอาจมีความไม่แน่นอนหรือข้อผิดพลาด

**Value:** หมายถึงความคุ้มค่าที่เราสามารถดึงเอาออกมาจากการวิเคราะห์และการจัดการข้อมูล Big Data ข้อมูลที่มากมายและหลากหลายนี้สามารถนำมาเพิ่มความรู้และค่าให้กับธุรกิจ ช่วยในการตัดสินใจทางธุรกิจ การวิเคราะห์แนวโน้ม หรือการสร้างความเข้าใจเกี่ยวกับผู้ใช้งาน

1. **Data warehouse กับ data lake ต่างกันอย่างไร พร้อมยกตัวอย่าง  
   Data Warehouse (คลังข้อมูล):**
   * Data Warehouse เป็นฐานข้อมูลที่ถูกสร้างขึ้นเพื่อเก็บข้อมูลที่ถูกดึงมาจากแหล่งต่าง ๆ เช่นระบบฐานข้อมูลอื่น ๆ และแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์และรายงาน
   * มักจะมีโครงสร้างและรูปแบบที่มีการกำหนดไว้ล่วงหน้า และใช้ข้อมูลที่มีโครงสร้างเช่นตาราง คอลัมน์ และแถว
   * มีการจัดระบบการจัดเก็บข้อมูลที่แยกจากแหล่งต้นทาง และมีการควบคุมความเป็นระเบียบเพื่อให้ข้อมูลมีความน่าเชื่อถือและคุณภาพ

**ตัวอย่าง:** บริษัทขายสินค้าทางออนไลน์มี Data Warehouse เพื่อเก็บข้อมูลการขายและสต็อกสินค้าจากสาขาทั่วประเทศ เพื่อวิเคราะห์แนวโน้มการขายและสร้างรายงานเพื่อการตัดสินใจทางธุรกิจ

**Data Lake (ทะเลข้อมูล):**

* Data Lake เป็นพื้นที่ที่ใช้เก็บข้อมูลที่มีความหลากหลายและมาจากแหล่งต่าง ๆ โดยไม่จำเป็นต้องแปลงข้อมูลให้อยู่ในรูปแบบที่ตามมาตรฐานล่วงหน้า
* ข้อมูลใน Data Lake มีความยืดหยุ่นมาก และสามารถรับมือกับข้อมูลที่ไม่มีโครงสร้างเช่นข้อความ ภาพถ่าย วิดีโอ และข้อมูลชนิดอื่น ๆ
* สามารถเป็นอันตรายเพราะข้อมูลที่เก็บอาจมีความไม่แน่นอน หรือความไม่เชื่อถือ

**ตัวอย่าง:** บริษัทที่พัฒนาแอปพลิเคชันมือถือมี Data Lake เพื่อเก็บข้อมูลการใช้งานของผู้ใช้งาน ซึ่งอาจรวมข้อมูลจากแอปพลิเคชันต่าง ๆ และที่มาจากแหล่งต่าง ๆ เพื่อวิเคราะห์และค้นหาข้อมูลเชิงลึกในภายหลัง

**<---------------------------------------------------- Credit by ChatGPT --------------------------------------------------->**

**งาน SQL for Data Science**

**งานที่1 Download & Install SQLite Tools (save ภาพ มาส่ง)**

**1. ให้นักศึกษาไปที่** [**https://www.sqlite.org/download.html**](https://www.sqlite.org/download.html)

**2. จากนั้นทำการ download  
 3. Files ที่ download เป็น zip file จากนั้น สร้าง folder C:\sqlite  
 4. ทำการแตก files ที่ download มา และนำมาไว้ที่** C:\sqlite

**5. จากนั้น เปิด command prompt (cmd)**

**พิมพ์ คำสั่ง cd\ เพื่อต้องการไปสู่ C:\**

**คำสั่ง cd sqlite เพื่อต้องการเข้าสู่** C:\sqlite **ดังภาพ**

**รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ภาพหน้าจอ, ออกแบบ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**6. จากนั้นพิมพ์คำสั่ง sqlite3 และ enter**

**รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ภาพหน้าจอ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**7. พิมพ์คำสั่ง .help เพื่อดู คำสั่งต่าง ๆ ได้**

**8. ในกรณีที่ต้องการออกจากโปรแกรม พิมพ์คำสั่ง .quit ดังภาพ**

**9. \*\*\*ในกรณ๊ที่ต้องการใช้ GUI Tool สามารถใช้ SQLiteStudio หรือ DBeaver (**[**https://sqlitebrowser.org/dl/**](https://sqlitebrowser.org/dl/)) **หรือ อื่น ๆ ได้**

**งานที่ 2 How to connect to SQLite sample database (save ภาพ มาส่ง)  
 1. ให้นักศึกษา สร้าง folder C:\sqlite\db**

**2. จากนั้น นำ file ของฐานข้อมูล chinook.db เก็บใน folder ดังกล่าว**

**3. เข้าถึงข้อมูลด้วย C:\sqlite>sqlite3 C:\sqlite\db\chinook.db**

**รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ภาพหน้าจอ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**4. พิมพ์คำสั่ง .tables เพื่อดูตารางทั้งหมดที่มีอยู่ในฐานข้อมูล**

**รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ภาพหน้าจอ, สีดำ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**5. พิมพ์คำสั่ง .database เพื่อดูฐานข้อมูล**

**รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ภาพหน้าจอ, สีดำ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**6. เปิด SQLiteStudio จากนั้น เลือก Add a database ตามภาพ**

**7. พิมพ์คำสั่ง** รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ซอฟต์แวร์, แสดง

คำอธิบายที่สร้างโดยอัตโนมัติ