



01



Data pipeline

Ingestion

Transformation

Storage

Analysis



Data pipeline คืออะไร

Data pipeline เป็นลำดับของกระบวนการในการจัดการข้อมูลตั้งแต่ที่เป็นข้อมูลดิบ ทำให้ได้ข้อมูลที่มีความถูกต้อง น่าเชื่อถือ และพร้อมใช้งาน ไปจนถึงการวิเคราะห์ข้อมูล การพัฒนาตัวแบบเพื่อการวิเคราะห์ข้อมูล ที่ตรงกับวัตถุประสงค์ในการใช้งาน

Data pipeline ประกอบด้วยกระบวนการ 4 ขั้นตอน ดังนี้

1. Data ingestion
2. Data transformation
3. Data storage
4. Data analysis

2. Data transformation

Data transformation เป็นการจัดการข้อมูลตั้งแต่การทำความสะอาดข้อมูล การตรวจสอบความสอดคล้องของข้อมูล การตรวจสอบความเป็นไปได้ของข้อมูล การปรับข้อมูลให้เป็นมาตรฐานเดียวกัน หรือเรียกว่ากระบวนการอีทีแอล (ETL, Extract transform loading)

3. Data storage

Data storage เป็นการจัดเก็บข้อมูลในฐานข้อมูลที่เรียกว่าคลังข้อมูล (Data warehouse) โดยคลังข้อมูลมุ่งเน้นการเก็บข้อมูลเพื่อวิเคราะห์ข้อมูล

4. Data analysis

Data analysis เป็นการเรียกใช้ข้อมูลที่เก็บไว้ในคลังข้อมูล โดยสามารถแบ่งกลุ่มการวิเคราะห์ข้อมูลได้เป็น 3 กลุ่ม คือ

1. การสร้างรายงานธุรกิจอัจฉริยะ
2. การสร้างคิวเตต้าและการวิเคราะห์ OLAP
3. เหมืองข้อมูลและการเรียนรู้ของเครื่อง

