

Big Data



มหาวิทยาลัยรังสิต
RANGSIT UNIVERSITY

Introduction to Big Data

Rangsit University

Date/Time : 1 February 2026 13.00-16.00

BY : DR. Rattapol Kasemrat

About Me



- Dr. Rattapol Kasemrat is a technology leader with over 20 years of experience in IT management and digital transformation. As Head of Information Technology and Core Systems at Central Food Wholesale, he drives enterprise modernization and ERP strategy across the organization.
- In parallel, he serves on EVP of PMI Thailand Chapter and has contributed to the PMBOK® Guide Thai translation and activities.
- Dr. Rattapol earned his Ph.D. in Digital Economy from Rangsit University, researching machine learning–driven hyper-personalization. His current interests extend into space deep tech, including edge AI, satellite autonomy, and inter-satellite communication.

I'M SPEAKING AT
2ND ANNUAL
FUTURECIO
CONFERENCE
THAILAND

11 FEB 2026
CROWNE PLAZA BANGKOK



cxociety
Panel Discussion:
Digital Ethics and
Responsible AI
Governance for
Asia

Dr. Rattapol Kasemrat
Head of Information
Technology
GO WHOLESALE


#CXOCIETY #CIOTH #FUTURECIO

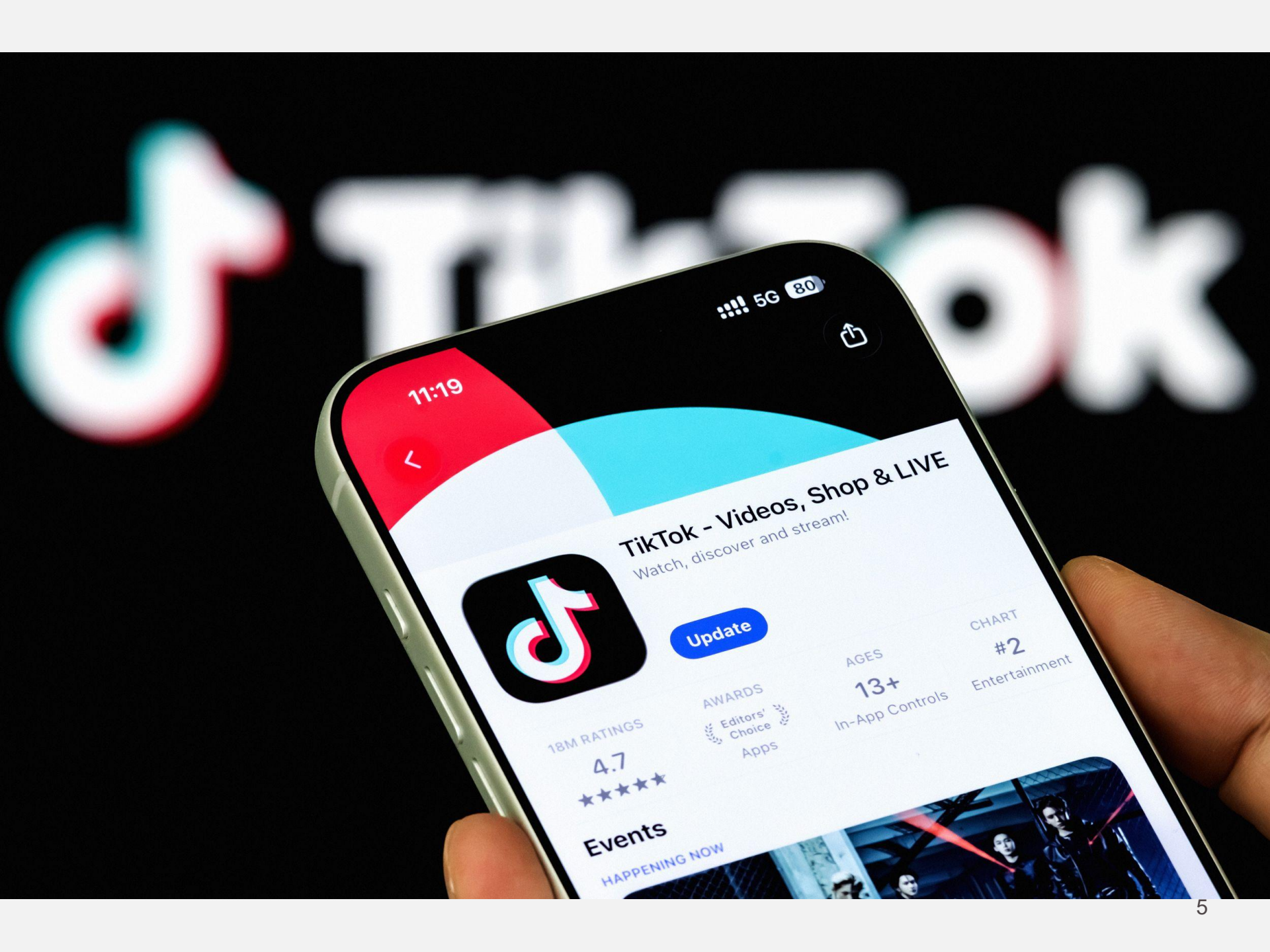


linkedin.com/in/rattapolk

Big Data คืออะไร?

Big Data หมายถึง ข้อมูลที่มีลักษณะดังนี้

- มี ปริมาณมากเกินไป จนระบบแบบเดิมจัดการไม่ไหว
- มี ความเร็วสูง เกิดและเปลี่ยนแปลงตลอดเวลา
- มี ความซับซ้อนและหลากหลาย ไม่อยู่ในรูปแบบตารางอย่างเดียว
-  จึงไม่สามารถจัดการด้วยระบบฐานข้อมูลแบบดั้งเดิมได้



11:19

5G 80



TikTok - Videos, Shop & LIVE

Watch, discover and stream!



Update

18M RATINGS

4.7



Events

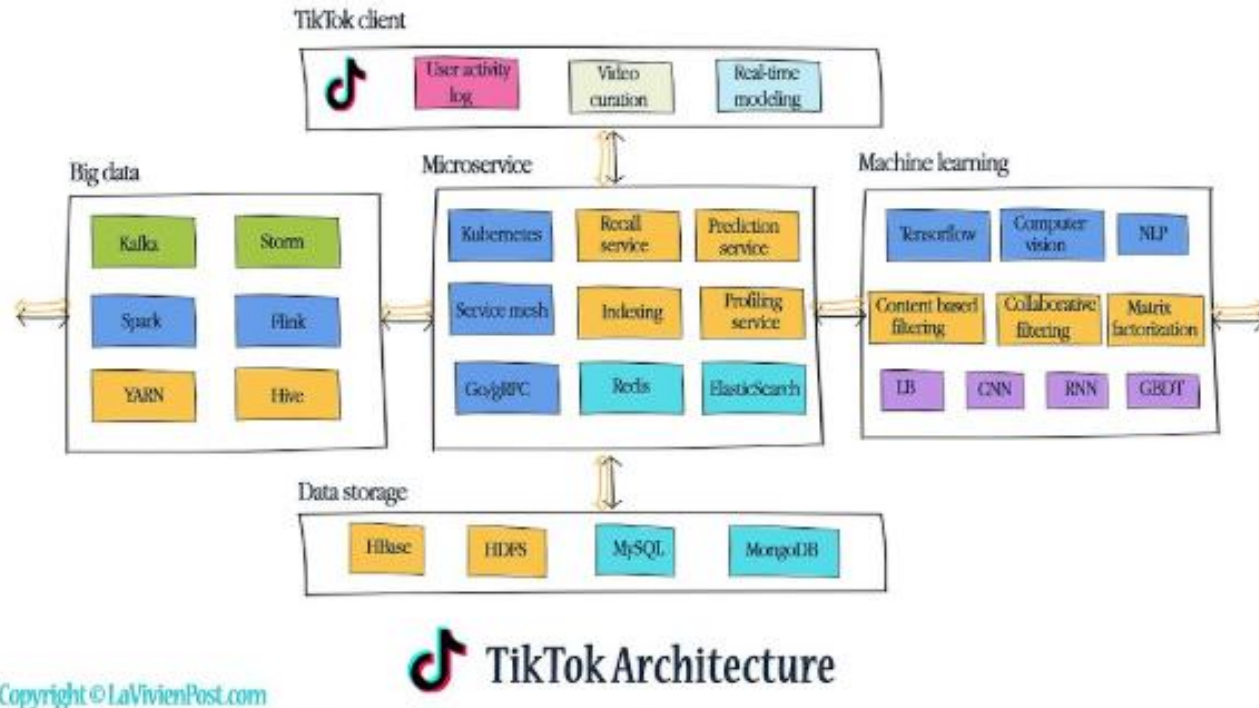
HAPPENING NOW

AWARDS
Editors' Choice
Apps

AGES
13+
In-App Controls

CHART
#2
Entertainment










ภาพนี้แสดง ภาพรวมของระบบ **Big Data** → **Machine Learning** ที่ใช้ในแพลตฟอร์มขนาดใหญ่อย่าง **TikTok** อ่านจากบนลงล่าง / ซ้ายไปขวา

1. ผู้ใช้ใช้งาน **TikTok**
2. เกิดข้อมูลจำนวนมหาศาล (**Big Data**)
3. ระบบจัดเก็บและประมวลผลข้อมูล
4. ส่งข้อมูลไปให้ **Machine Learning**
5. ระบบแนะนำวิดีโอแบบเฉพาะบุคคล

ทำไม Big Data ถึงสำคัญ

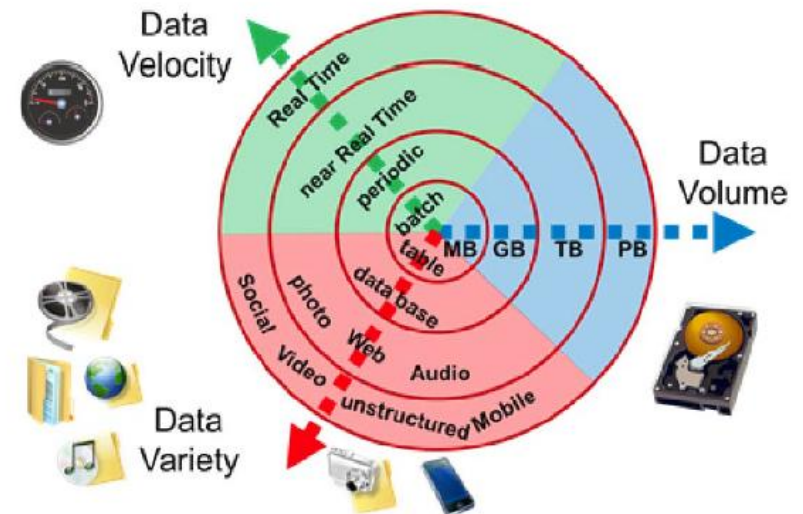
Big Data มีความสำคัญ เพราะ

-  ปริมาณข้อมูลเพิ่มขึ้นอย่างรวดเร็ว
จากโซเชียลมีเดีย อุปกรณ์ IoT และระบบดิจิทัลต่าง ๆ
-  ข้อมูลเกิดขึ้นตลอดเวลาและต้องใช้ทันที
การตัดสินใจแบบ **real-time** กลายเป็นเรื่องจำเป็น
-  ข้อมูลมีรูปแบบหลากหลายและซับซ้อน
เช่น ข้อความ รูปภาพ วิดีโอ เสียง และ log จากระบบ
-  เป็นพื้นฐานของ **Analytics, AI** และ **Machine Learning**
โมเดลอัจฉริยะต้องอาศัยข้อมูลจำนวนมากในการเรียนรู้
-  สร้างคุณค่าทางธุรกิจและความได้เปรียบในการแข่งขัน
องค์กรที่ใช้ข้อมูลได้ดีกว่าจะตัดสินใจได้แม่นยำกว่า

5Vs ของ Big Data

คุณลักษณะสำคัญของ Big Data (5Vs)

- **Volume** ปริมาณข้อมูล
ข้อมูลมีจำนวนมหาศาล ตั้งแต่ระดับ TB ถึง PB
- **Velocity** ความเร็วของข้อมูล
ข้อมูลเกิดขึ้นและต้องถูกประมวลผลอย่างรวดเร็ว
- **Variety** ความหลากหลายของข้อมูล
ข้อมูลมีหลายรูปแบบ เช่น ข้อความ รูปภาพ วิดีโอ และเสียง
- **Veracity** ความน่าเชื่อถือของข้อมูล
ข้อมูลอาจมีความผิดพลาด ซ้ำซ้อน หรือไม่สมบูรณ์
- **Value** คุณค่าของข้อมูล
การนำข้อมูลไปใช้สร้างประโยชน์และมูลค่าทางธุรกิจ



Traditional Data vs Big Data

Traditional Data

◆ การจัดการข้อมูลแบบเดิม (Traditional Data)

- ข้อมูลอยู่ในรูปแบบ ตาราง (Structured)
- ปริมาณข้อมูลระดับ GB – TB
- ใช้ ฐานข้อมูลแบบศูนย์กลาง
- ประมวลผลบน เครื่องเดียวหรือเซิร์ฟเวอร์เดียว
- ตัวอย่าง: Excel, Relational Database

Big Data

◆ การจัดการข้อมูลแบบ Big Data

- ข้อมูลมีทั้ง Structured และ Unstructured
- ปริมาณข้อมูลระดับ TB – PB
- ใช้ ระบบแบบกระจาย (Distributed System)
- ประมวลผลด้วย หลายเครื่องพร้อมกัน
- รองรับข้อมูลจากหลายแหล่งแบบ real-time

1 PB (Petabyte) = 1,000,000 GB 1 TB = 1,000 GB

Question for 1st section

- ถ้าองค์กรยังใช้ระบบแบบเดิมกับข้อมูลระดับ PB (Petabyte) จะเกิดปัญหาอะไรบ้าง?

10 Minutes Break

Let's try to experience Kahoot.

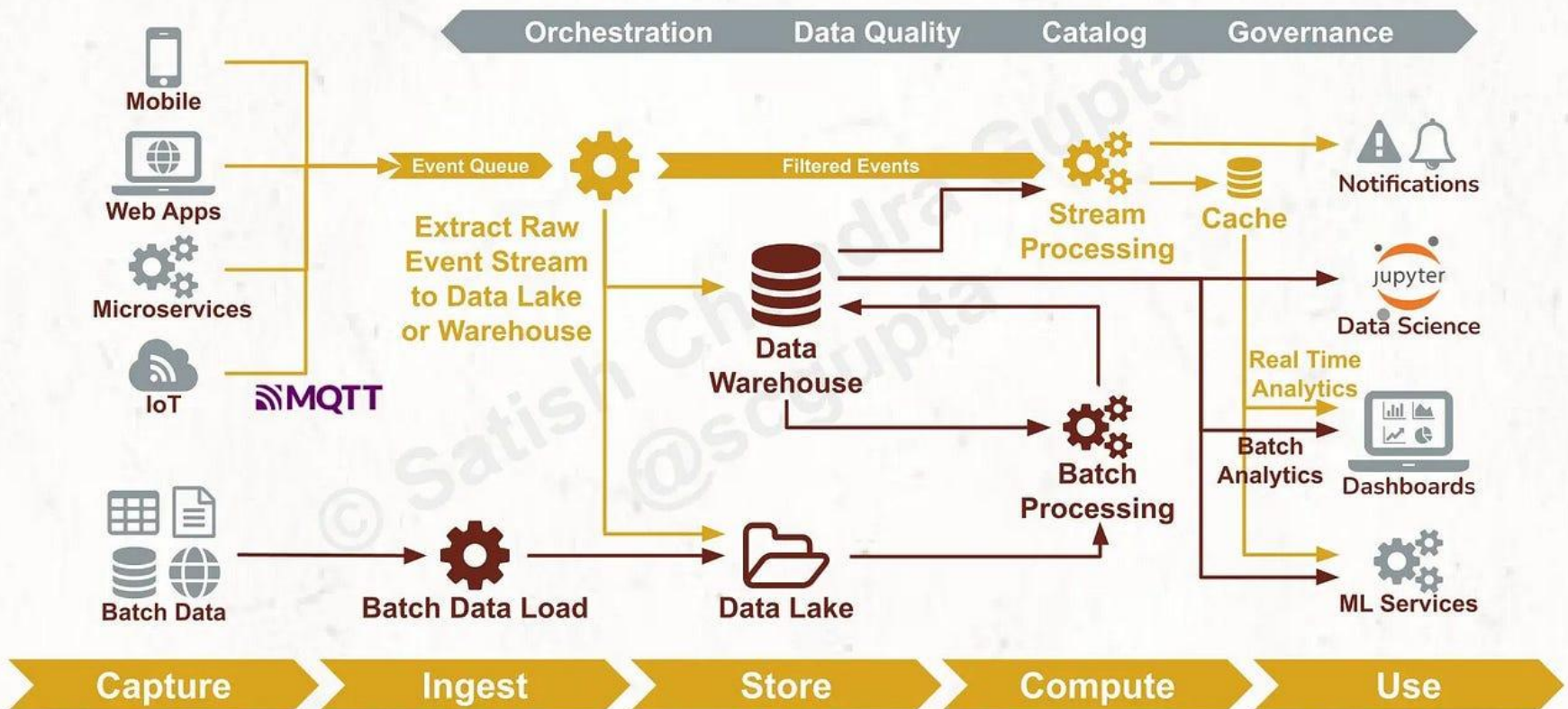
Big Data Pipeline (กระบวนการทำงานของ Big Data)

■ Data Source → Storage → Processing → Analytics → Insight

- แหล่งข้อมูล (Data Sources)
เช่น ผู้ใช้งาน ระบบธุรกิจ โซเชียลมีเดีย อุปกรณ์ IoT
- การจัดเก็บข้อมูล (Storage)
เก็บข้อมูลจำนวนมากในระบบแบบกระจาย
- การประมวลผล (Processing)
ทำความสะอาด รวม และแปลงข้อมูล
- การวิเคราะห์ (Analytics)
ค้นหารูปแบบ แนวโน้ม และ Insight
- การนำไปใช้ (Decision / Action)
ใช้ข้อมูลช่วยตัดสินใจหรือสร้างบริการอัจฉริยะ

Big Data Pipeline Architecture

ml4devs.com/big-data-pipeline



© Satish Chandra Gupta, Creative Commons BY-NC-ND 4.0 International License.

scgupta

linkedin.com/in/scgupta

Distributed Computing

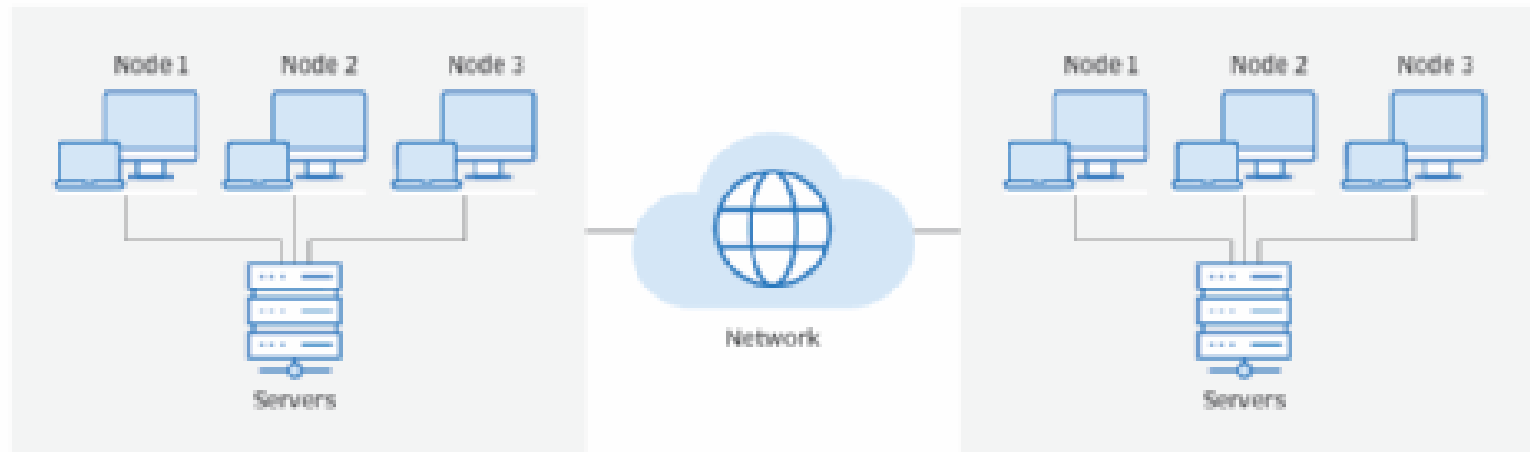
Distributed Computing คือแนวคิดที่

- ใช้ คอมพิวเตอร์หลายเครื่องทำงานร่วมกัน
- แบ่งข้อมูลขนาดใหญ่ออกเป็นส่วนย่อย
- ประมวลผลข้อมูล **พร้อมกัน**หลายเครื่อง
- รวมผลลัพธ์กลับมาเป็นคำตอบเดียว



แทนที่จะใช้เครื่องเดียว → ใช้หลายเครื่องช่วยกัน

The distributed computing process



Big Data Technology

เทคโนโลยี Big Data ถูกออกแบบมาเพื่อ

- จัดเก็บข้อมูลขนาดใหญ่มาก
- ประมวลผลข้อมูลจากหลายเครื่องพร้อมกัน
- รองรับข้อมูลที่มาเร็วและหลากหลาย

◆ กลุ่มเทคโนโลยีหลักของ Big Data

- ระบบจัดเก็บข้อมูลแบบกระจาย (Distributed Storage)
ใช้เก็บข้อมูลจำนวนมากหลายอย่างมีประสิทธิภาพ
ตัวอย่าง: Apache Hadoop (HDFS)
- ระบบประมวลผลข้อมูล (Processing Engine)
ใช้ประมวลผลข้อมูลจากหลายเครื่องพร้อมกัน
ตัวอย่าง: Apache Spark
- ฐานข้อมูลแบบ NoSQL
รองรับข้อมูลที่ไม่อยู่ในรูปแบบตาราง
เช่น ข้อมูลข้อความ รูปภาพ และ log

DATA SOURCES

Internal data sources such as data from CRM system, ERP system, sales reports, etc.

External data sources such as government statistics and media channels



DATA STORAGE

Big data storage software tools store, manage and retrieve massive amounts of data.



DATA MINING

Data mining tools allow businesses to extract usable data from a huge set of raw data to find relationships, patterns, and anomalies.



SPSS Modeler



DATA ANALYTICS

Although data mining tools incorporate data analysis, there are software designed specifically with advanced analytical capabilities.



DATA VISUALIZATION

Data visualization software is also a type of data analytics tool. However, they are specifically designed to take the raw data and presenting it with beautiful and easy digestible visuals like graphs and charts.



Batch Processing vs Streaming Processing

Batch Processing

◆ Batch Processing (ประมวลผลเป็นรอบ)

- ประมวลผลข้อมูล ทีละชุด ตามช่วงเวลา
- ใช้กับข้อมูล ย้อนหลัง / สรุปรายงาน
- ไม่ต้องการผลลัพธ์ทันที
- ตัวอย่าง: รายงานยอดขายรายวัน, Payroll

Streaming Processing

◆ Streaming Processing (ประมวลผลแบบเรียลไทม์)

- ประมวลผลข้อมูล ทันทีที่ข้อมูลเข้ามา
- ใช้กับข้อมูลที่ต้องการ ความรวดเร็ว
- รองรับการเหตุการณ์ต่อเนื่อง
- ตัวอย่าง: ตรวจสอบการโกง, ระบบแนะนำแบบสด

Big Data Analytics

คือกระบวนการ

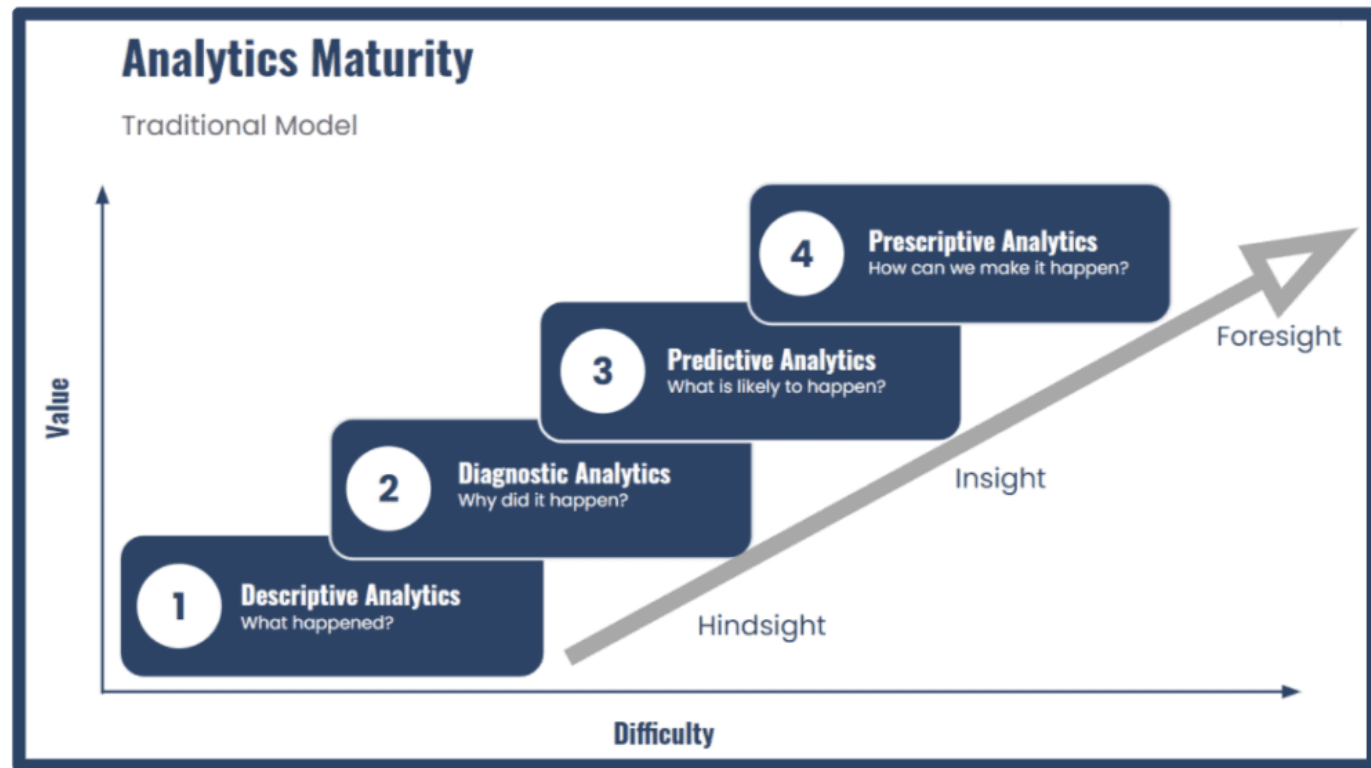
- วิเคราะห์ข้อมูลจำนวนมากจากหลายแหล่ง
- ค้นหารูปแบบ แนวโน้ม และความสัมพันธ์
- แปลงข้อมูลให้เป็น **Insight** ที่นำไปใช้ได้จริง
- สนับสนุนการตัดสินใจขององค์กร

เป้าหมายหลัก

- เข้าใจสิ่งที่เกิดขึ้น
- คาดการณ์สิ่งที่เกิดขึ้น
- ตัดสินใจได้แม่นยำขึ้น

ประเภท Analytics

- Descriptive
- Diagnostic
- Predictive
- Prescriptive



Question for 2nd section

- ทำไม Big Data ต้องใช้ Distributed Computing?

Case Study #1

- ร้านกาแฟขนาดกลาง มีสาขา 10 แห่ง มีข้อมูลยอดขายรายวัน
มีข้อมูลสมาชิกประมาณ 5,000 คน เจ้าของร้านต้องการเข้าใจพฤติกรรมลูกค้า
และเพิ่มยอดขาย

Case Study #2


บริษัทโลจิสติกส์ขนาดใหญ่ ให้บริการจัดส่งสินค้าทั่วประเทศ บริษัทมีข้อมูลดังนี้:

- ข้อมูลการจัดส่งหลายล้านรายการต่อเดือน
- ข้อมูลตำแหน่งรถ (GPS) แบบ real-time
- ข้อมูลเวลาการจัดส่งล่าช้า / ตรงเวลา
- ข้อมูลร้องเรียนจากลูกค้า

ผู้บริหารต้องการ:

- ลดเวลาการจัดส่ง
- ลดต้นทุนการขนส่ง
- คาดการณ์ปัญหาล่วงหน้า

Challenges ของ Big Data

- แม้ **Big Data** จะมีประโยชน์มาก
แต่ก็มีความท้าทายที่องค์กรต้องพิจารณา
-  ความเป็นส่วนตัวและความปลอดภัยของข้อมูล
ข้อมูลลูกค้าต้องได้รับการปกป้องตามกฎหมายและจริยธรรม
-  ต้นทุนในการลงทุนและดูแลระบบ
โครงสร้างพื้นฐาน บุคลากร และการบำรุงรักษามีค่าใช้จ่ายสูง
-  คุณภาพของข้อมูล **Data Quality**
ข้อมูลที่ไม่ถูกต้องหรือไม่สมบูรณ์
อาจนำไปสู่การตัดสินใจที่ผิดพลาด
-  ขาดบุคลากรที่มีความเชี่ยวชาญ
ต้องใช้ทักษะเฉพาะด้านในการจัดการและวิเคราะห์ข้อมูล
-  จริยธรรมและความโปร่งใส
การใช้ข้อมูลต้องคำนึงถึงผลกระทบต่อผู้ใช้งาน

สรุป

- Big Data \neq แค่ข้อมูลใหญ่
- คุณค่าอยู่ที่การวิเคราะห์

Q&A Session , Kahoot Time!

