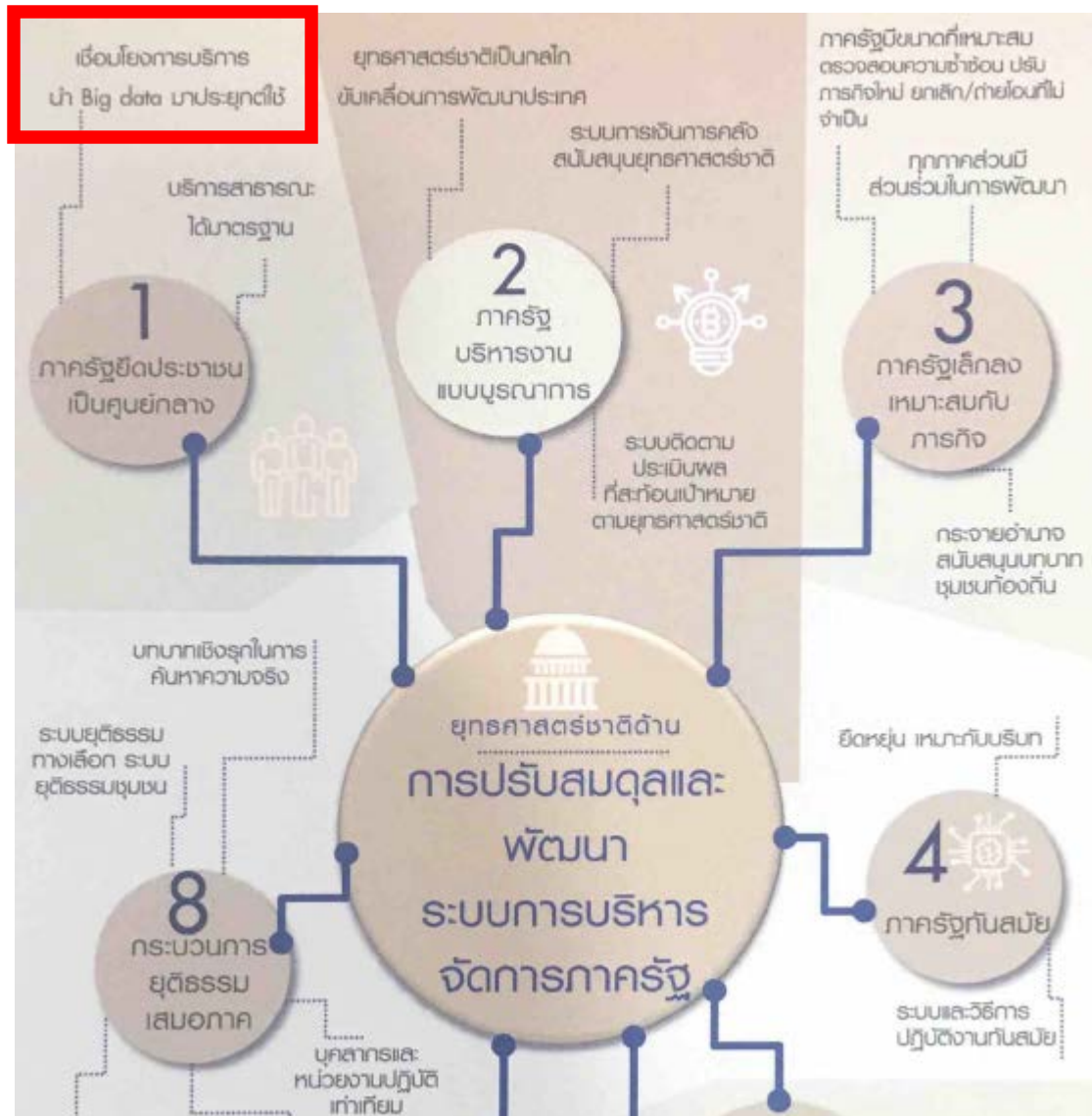




The pace of change will never be this slow again

Seize opportunity from disruption and cast a vision for the future





๔. ประเด็นยุทธศาสตร์ด้านการปรับสมดุลและพัฒนาระบบการบริหารจัดการ ภาครัฐ

๔.๑.๒ ภาครัฐมีความเชื่อมโยงในการให้บริการสาธารณะต่าง ๆ ผ่านการนำเทคโนโลยีดิจิทัลมาประยุกต์ใช้ มีระบบการบริหารจัดการข้อมูลที่มีความเชื่อมโยงระหว่างหน่วยงานและแหล่งข้อมูลต่าง ๆ นำไปสู่การวิเคราะห์การจัดการข้อมูลขนาดใหญ่ เพื่อการพัฒนานโยบายและการให้บริการภาครัฐ มีการนำเทคโนโลยีดิจิทัลมาประยุกต์ใช้ในการให้บริการสาธารณะตั้งแต่ต้นจนจบกระบวนการ เพื่อให้สามารถติดต่อราชการได้โดยง่าย สะดวก รวดเร็ว โปร่งใส เสียค่าใช้จ่ายน้อย และตรวจสอบได้

แผนพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม

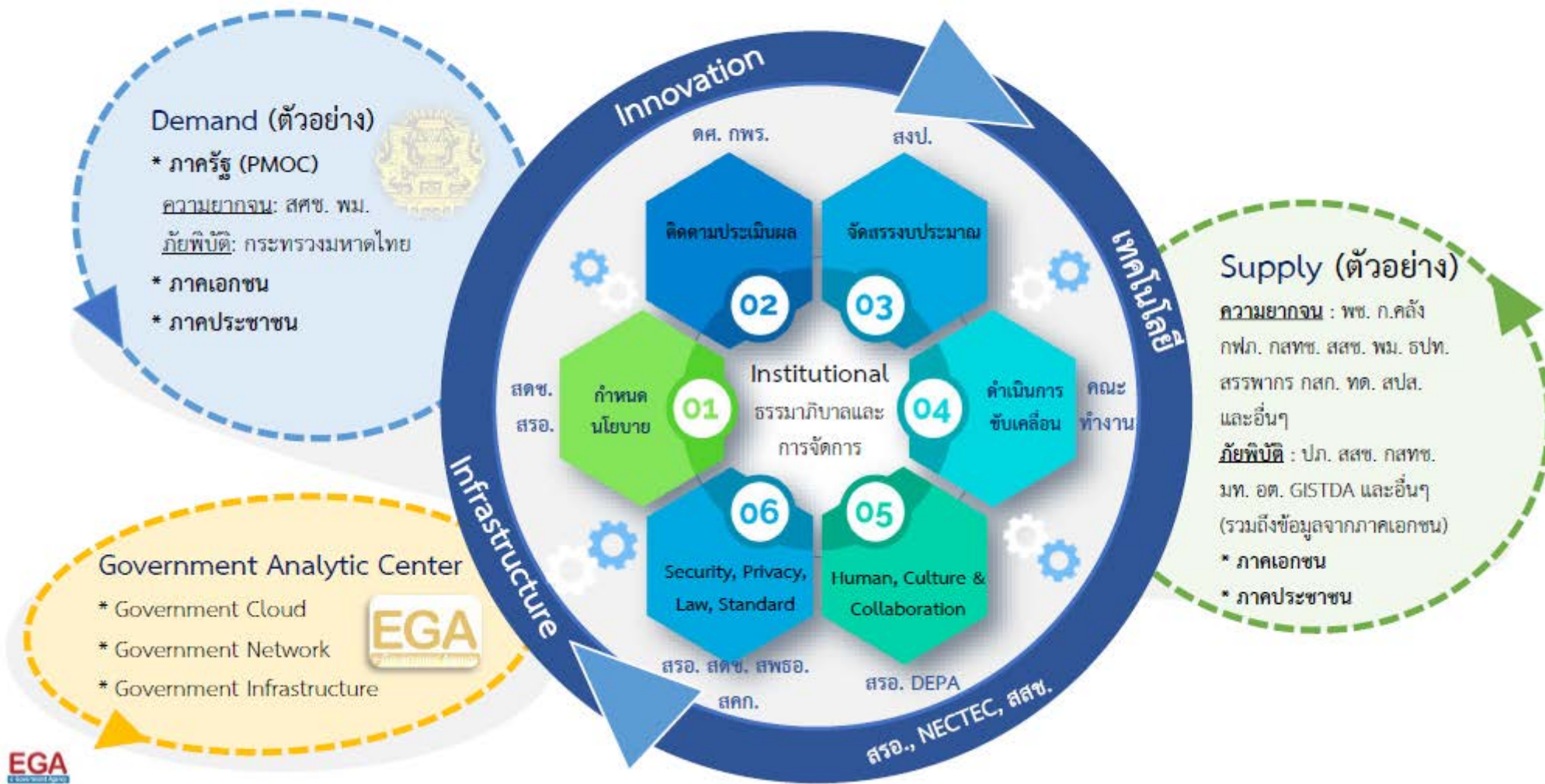
5. ภาคผนวก

อภิธานศัพท์

ข้อมูลขนาดใหญ่ (big data)

ปริมาณ**ข้อมูลที่มีขนาดใหญ่มาก** (ระดับ tera byte หรือ peta byte) เกินกว่าขีดความสามารถในการประมวลผลของระบบฐานข้อมูลธรรมดาจะรองรับได้ (volume) และข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็ว (velocity) เช่น ข้อมูลจาก social media ข้อมูลการซื้อขาย ข้อมูล transaction การเงินหรือการใช้โทรศัพท์ หรือข้อมูลจาก sensor จึงทำให้ข้อมูลมีหลากหลายรูปแบบ (variety) ทั้งที่มีรูปแบบและไม่มีรูปแบบ ซึ่งอาจจะอยู่ในรูปทั้ง RDBMS, text, XML, JSON หรือ image สำหรับ Big Data Technology คือ เทคโนโลยีในการนำข้อมูลจำนวนมากมาวิเคราะห์ ประมวลผล และแสดงผลด้วยวิธีที่เหมาะสม ส่วนการวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อให้สามารถนำข้อมูลมาใช้ได้ง่ายขึ้น

กลไกการขับเคลื่อนและหน่วยงานที่เกี่ยวข้องกับการพัฒนา Big Data ของประเทศไทย



Organization Data

ข้อมูลในองค์กรถูกแบ่งออกเป็น 3 กลุ่ม

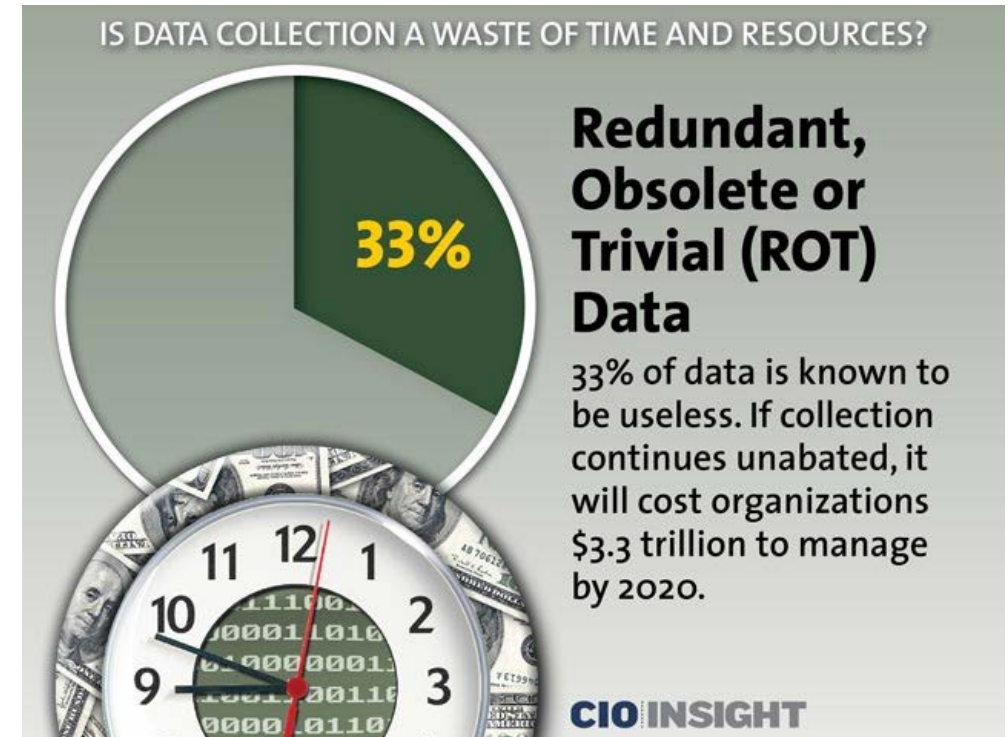
ถึงแม้ผู้ดูแลระบบภายในองค์กรหลายๆ คนอาจตระหนักดีอยู่แล้วว่า ข้อมูลภายในองค์กรนั้นมีระดับความสำคัญไม่เท่ากัน และหากข้อมูลที่มีความสำคัญสูงเกิดสูญหายไปก็อาจส่งผลเสียต่อการทำธุรกิจได้

สามารถแบ่งข้อมูลออกเป็น 3 กลุ่มตามระดับการสร้างคุณค่าให้แก่องค์กรเอาไว้ โดยเปรียบเทียบกับภูเขาน้ำแข็งและตั้งชื่อเรียกว่า Databerg



1. Business Critical Data เป็นชุดของข้อมูลที่มีความสำคัญต่อการดำเนินธุรกิจขององค์กร ข้อมูลกลุ่มนี้มีปริมาณนับเป็น 14% ของข้อมูลในองค์กรทั้งหมด

2. ROT (Redundant/Obsolete/Trivial) Data เป็นข้อมูลที่มีความซ้ำซ้อนซึ่งในอดีตนั้นองค์กรอาจเคยนำมาใช้งาน แต่ปัจจุบันข้อมูลเหล่านี้ไม่ได้สร้างคุณค่าอะไรให้กับองค์กรอีกแล้ว ข้อมูลกลุ่มนี้มีปริมาณนับเป็น 32% ของข้อมูลในองค์กรทั้งหมด



3. Dark Data or Dusty Data เป็นข้อมูลที่ถูกดูแลระบบไม่ทราบว่ามีอยู่ภายในองค์กร และไม่อาจแบ่งแยกได้ว่าข้อมูลเหล่านี้จะสามารถสร้างคุณค่าให้กับองค์กร หรือเป็น Business Critical Data หรือเป็นข้อมูลที่มีความซ้ำซ้อน หรือเป็น Rot Data ข้อมูลกลุ่มนี้มีปริมาณนับเป็น 54% ของข้อมูลในองค์กรทั้งหมด

[Gartner IT Glossary](#) > [Dark Data](#)

Dark Data

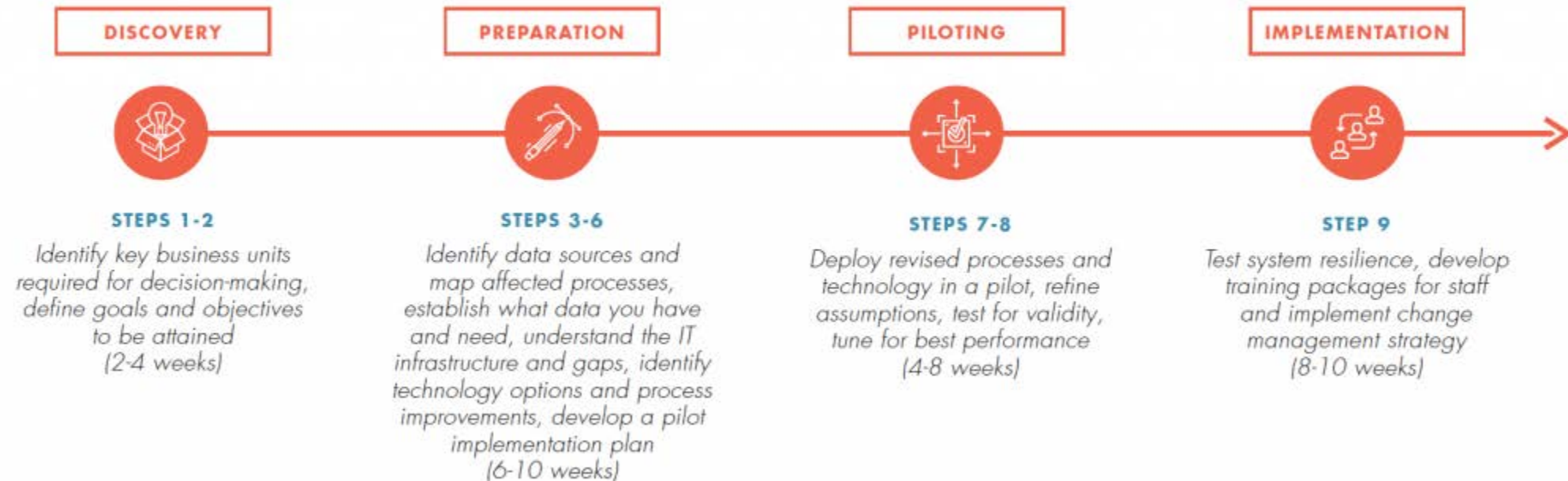


Gartner defines **dark data** as the information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes (for example, [analytics](#), business relationships and direct monetizing). Similar to dark matter in physics, dark data often comprises most organizations' universe of information assets. Thus, organizations often retain dark data for [compliance](#) purposes only. Storing and securing data typically incurs more expense (and sometimes greater risk) than value.

Also see:

- [Gartner's Data & Analytics Summit](#) – This inaugural summit will offer valuable insight and help redefine the role of data and analytics in digital business.
- [Free research and webinars from Gartner on Big Data Analytics.](#)

FRAMEWORK FOR DERIVING VALUE FROM DATA



ข้อมูลขนาดใหญ่

จากวิกิพีเดีย สารานุกรมเสรี

ในสาขาเทคโนโลยีสารสนเทศ, ข้อมูลขนาดใหญ่ (Big data) คือขุมทรัพย์ของชุดข้อมูลที่มีขนาดและความซับซ้อนมาก จนมันยากที่จะประมวลผลได้ด้วยเครื่องมือจัดการฐานข้อมูลที่มีอยู่ ความท้าทายนี้รวมถึงการจับบันทึก การจัดเก็บ[4] การค้นหา การแบ่งปัน การวิเคราะห์ และการวาดภาพข้อมูล แนวโน้มของชุดข้อมูลต่างๆ ที่ใหญ่ขึ้นเป็นผลจากสารสนเทศเพิ่มเติมที่ได้มาจากการวิเคราะห์ชุดข้อมูลชุดใหญ่ชุดเดียวของข้อมูลที่เกี่ยวข้องกัน เทียบกับชุดข้อมูลย่อยๆ หลายชุดที่แยกจากกันที่มีขนาดรวมกันแล้วเท่ากัน สิ่งนี้อนุญาตให้ความเชื่อมโยงถูกค้นพบได้ เพื่อ "หาแนวโน้มทางธุรกิจ ตัดสินคุณภาพของงานวิจัย ป้องกันโรค วิเคราะห์การอ้างอิงกฎหมาย ต่อสู้กับอาชญากรรม และบอกสภาพการจราจรตามเวลาจริง"



WIKIPEDIA
The Free Encyclopedia

นิยาม

Big Data มักรวมถึงชุดข้อมูลที่มีขนาดใหญ่เกินกว่าความสามารถของซอฟต์แวร์ที่ใช้กันอยู่ทั่วไป จะจับบันทึก จัดการ และประมวลผลข้อมูลดังกล่าวได้ภายในเวลาที่ยอมรับได้ ขนาดของ big data นั้นเป็นเป้าหมายที่เคลื่อนไปเรื่อยๆ ตามเวลา ในปี 2012 ขนาดของมันอยู่ที่ตั้งแต่ไม่กี่เทราไบต์ไปจนถึงหลายๆ เพตาไบต์ในชุดข้อมูลชุดเดียว ด้วยความยากลำบากนี้ แพลตฟอร์มใหม่สำหรับ "big data" จึงได้เกิดขึ้นเพื่อจะสามารถทำจัดการกับข้อมูลจำนวนมากเช่นนั้นได้ ตัวอย่างเช่น Apache Hadoop Big Data Platform.



Big Data Definition



ทั้งหมด

คำรูป

หนังสือ

วิดีโอ

ข่าวสาร

เพิ่มเติม

การตั้งค่า

เครื่องมือ

ผลการค้นหาประมาณ 631,000,000 รายการ (0.57 วินาที)

big da·ta

noun COMPUTING

extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

"much IT investment is going towards managing and maintaining big data"

แปล Big Data เป็น

1. ข้อมูลใหญ่

คำจำกัดความเพิ่มเติมและที่มาของคำ

แสดงความคิดเห็น

“**Big data** จะมีลักษณะ ปริมาณมหาศาล (high Volume) ความเร็วสูง (high Velocity) และมีความหลากหลาย (high Variety) จำเป็นต้องการรูปแบบของ กระบวนการประมวลผล เพื่อให้สามารถเพิ่มขีดความสามารถในการตัดสินใจ (Decision-Making) ข้อมูลเชิงลึก (Insight Discovery) และ การเพิ่มประสิทธิภาพ กระบวนการ (Process Optimization)”

(Gartner 2012)



ลักษณะพิเศษของ Big Data

Volume

คือข้อมูลมหาศาลขนาดใหญ่ มีจำนวนมากเกินกว่าระบบฐานข้อมูลแบบเดิมๆ จะสามารถที่จะจัดการได้

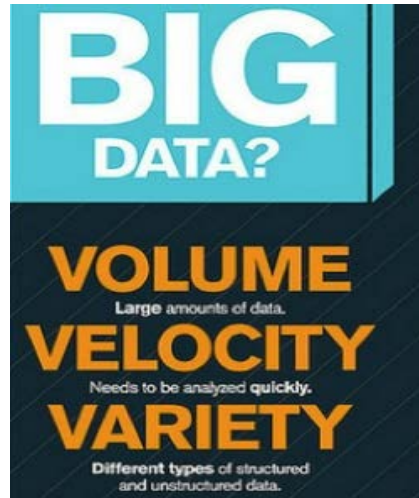
Velocity

ข้อมูลที่ต้องวิเคราะห์เข้าสู่ระบบฐานข้อมูลอย่างรวดเร็ว โดยให้ความสำคัญกับข้อมูลที่เป็น Real-time

Variety

ข้อมูลที่มีความหลากหลายทั้งที่เป็นแบบโครงสร้างหรือรูปแบบที่ไม่แน่นอน

3 V



Volume
- เกินขีดจำกัดของฐานข้อมูลดั้งเดิม (traditional column and row)
- มีการเพิ่มจำนวนข้อมูลอย่างมาก



Vertical scalability
- ความสามารถขยายแหล่งจัดเก็บสำหรับข้อมูล

Velocity
- เดินทางอย่างรวดเร็ว, โดยทันทีทันใด (real time)

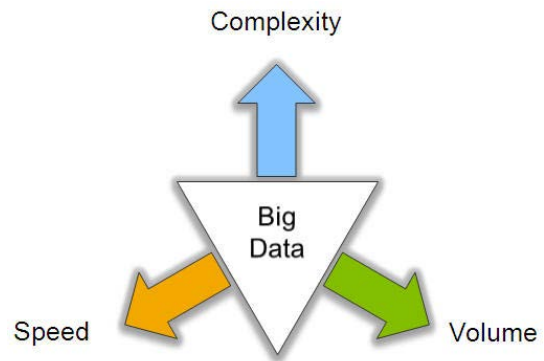


Data streaming
- การประมวลผล/การวิเคราะห์/การเปลี่ยนรูป

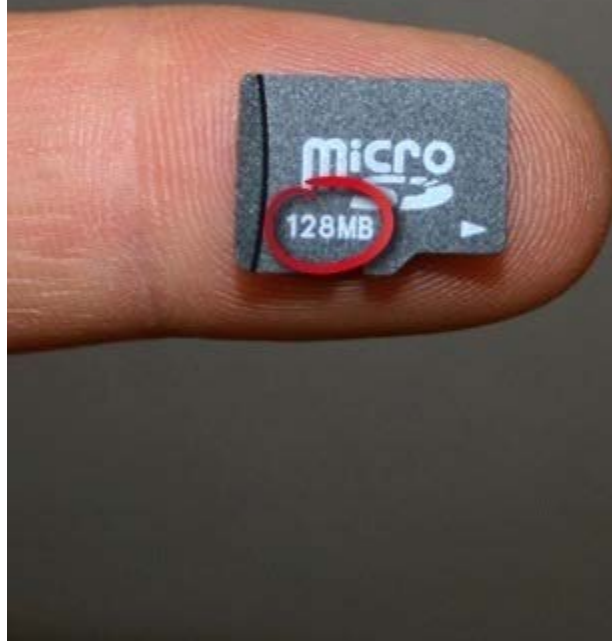
Variety
- ไม่มีโครงสร้างมาตรฐาน เช่น text, images



Horizontal scalability
- ความสามารถในการปรับเปลี่ยนโครงสร้างข้อมูล



2005

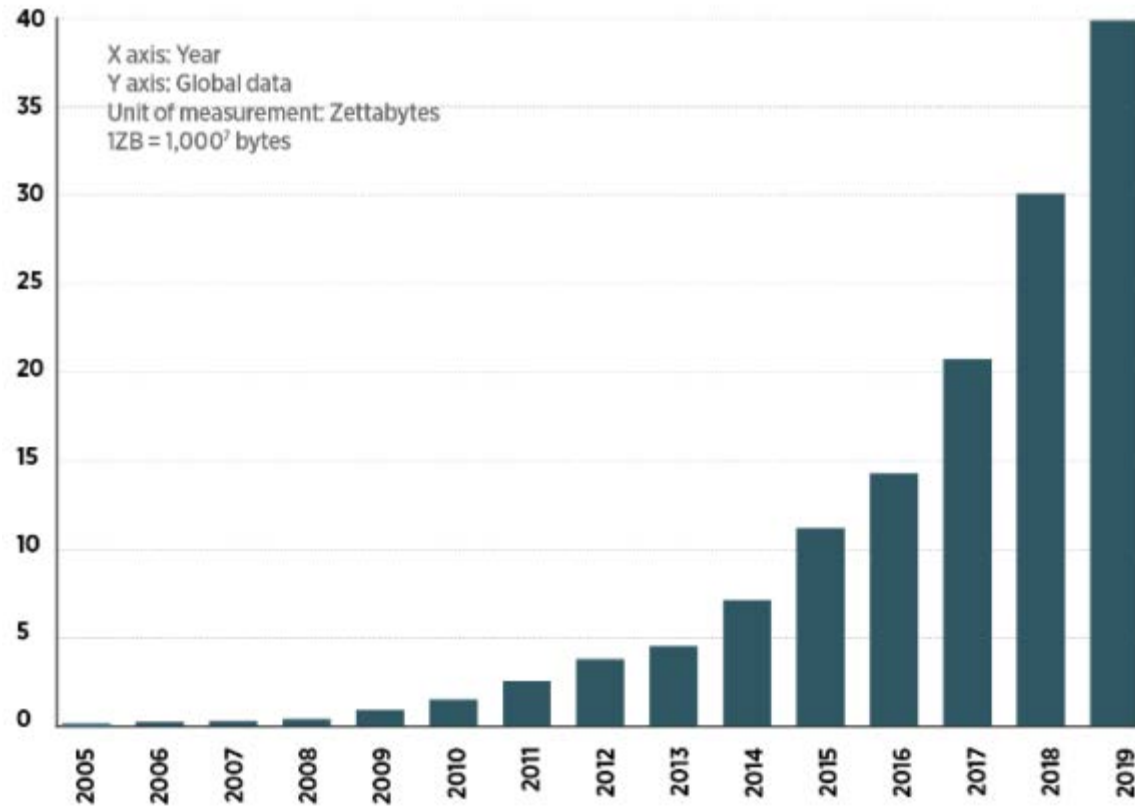


2014



	Approx. Bytes	Actual Bytes	Approx. Bits	Typical file/media
1B	1	1	8	Text email, SMS
1KB	1000B = 10^3	1024B = 2^{10}	8×10^3	Word document
1MB	1000KB = 10^6	1024KB = 2^{20}	8×10^6	Digital photo
1GB	1000MB = 10^9	1024MB = 2^{30}	8×10^9	DVD
1TB	1000GB = 10^{12}	1024GB = 2^{40}	8×10^{12}	Hard disk
1PB	1000TB = 10^{15}	1024TB = 2^{50}	8×10^{15}	Cloud?

DATA GROWTH

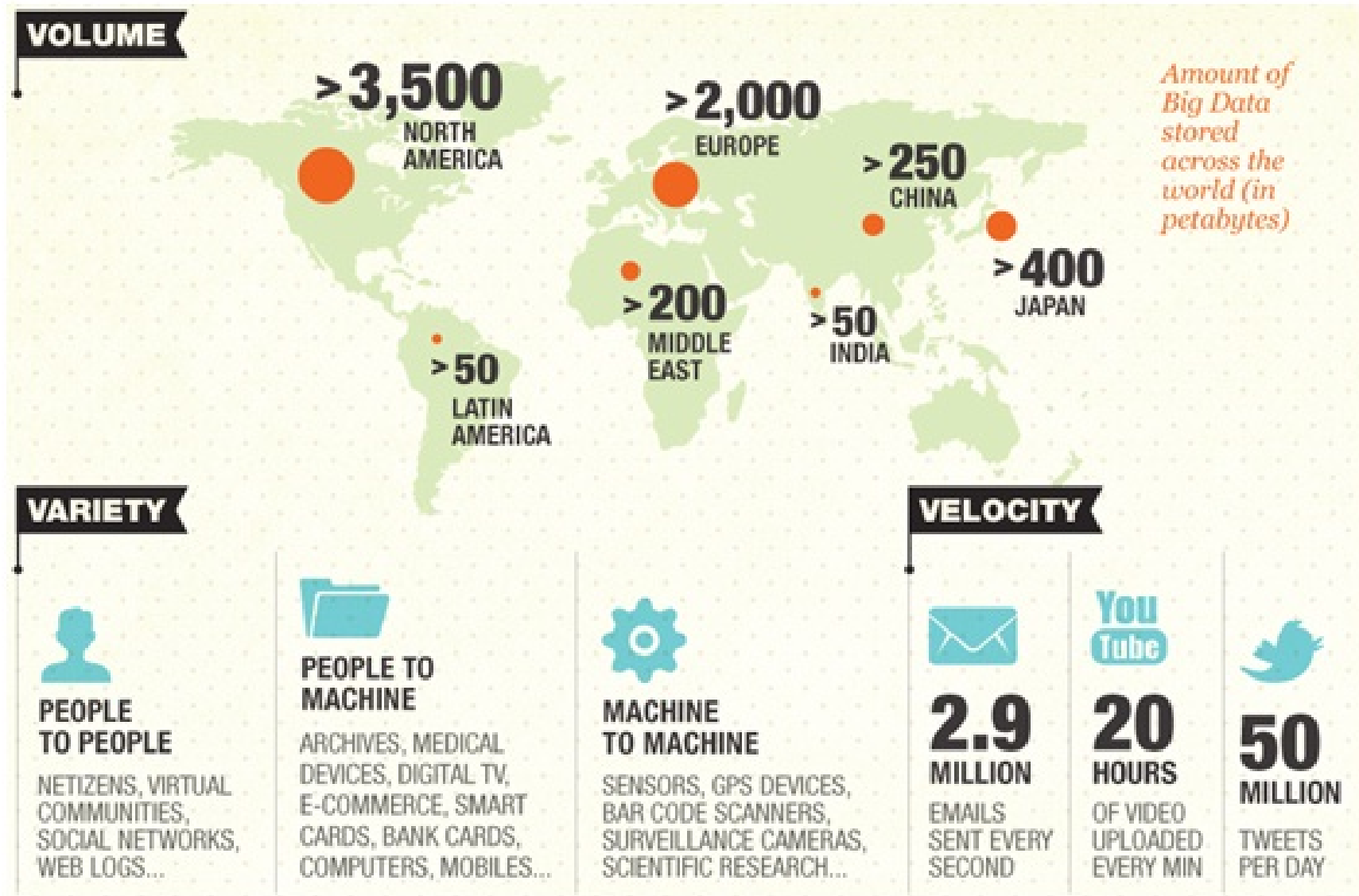


Note: Post-2013 figures are predicted. Source: UNECE

A PETABYTE IS A LOT OF DATA

1 PETABYTE	▪ 20 MILLION FOUR-DRAWER FILING CABINETS FILLED WITH TEXT
1 PETABYTE	▪ 13.3 YEARS OF HD-TV VIDEO
1.5 PETABYTES	▪ SIZE OF THE 10 BILLION PHOTOS ON → FACEBOOK
15+ PETABYTES	▪ INTERNET USER'S DATA BACKED UP ON MOZY.COM
20 PETABYTES	▪ THE AMOUNT OF DATA PER PROCESSED BY GOOGLE DAY
20 PETABYTES	▪ TOTAL HARD DRIVE SPACE MANUFACTURED IN 1995
50 PETABYTES	▪ THE ENTIRE WRITTEN WORKS OF MANKIND, FROM THE BEGIN- NING OF RECORDED HISTORY, IN ALL LANGUAGES

(all approximate)



The Three V's of Big Data (Source: Wipro)

Why is big data important now?



Every second:


43,824 searches


6,956 tweets


2,329,627 sent


82,500 videos viewed


21,393 GB of Internet Traffic

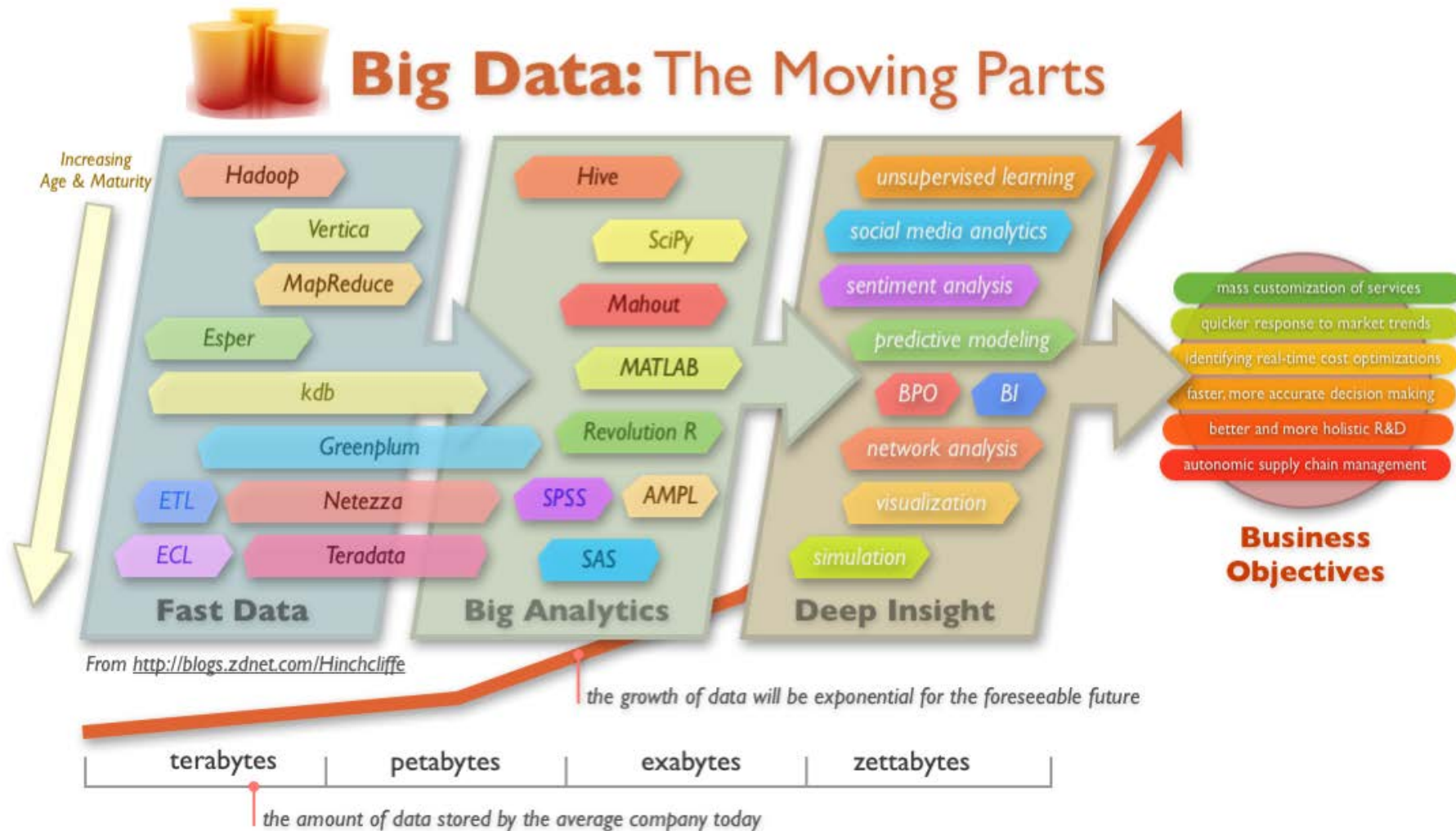


By 2020 at least one third
of all data will exist in or pass
through the cloud.

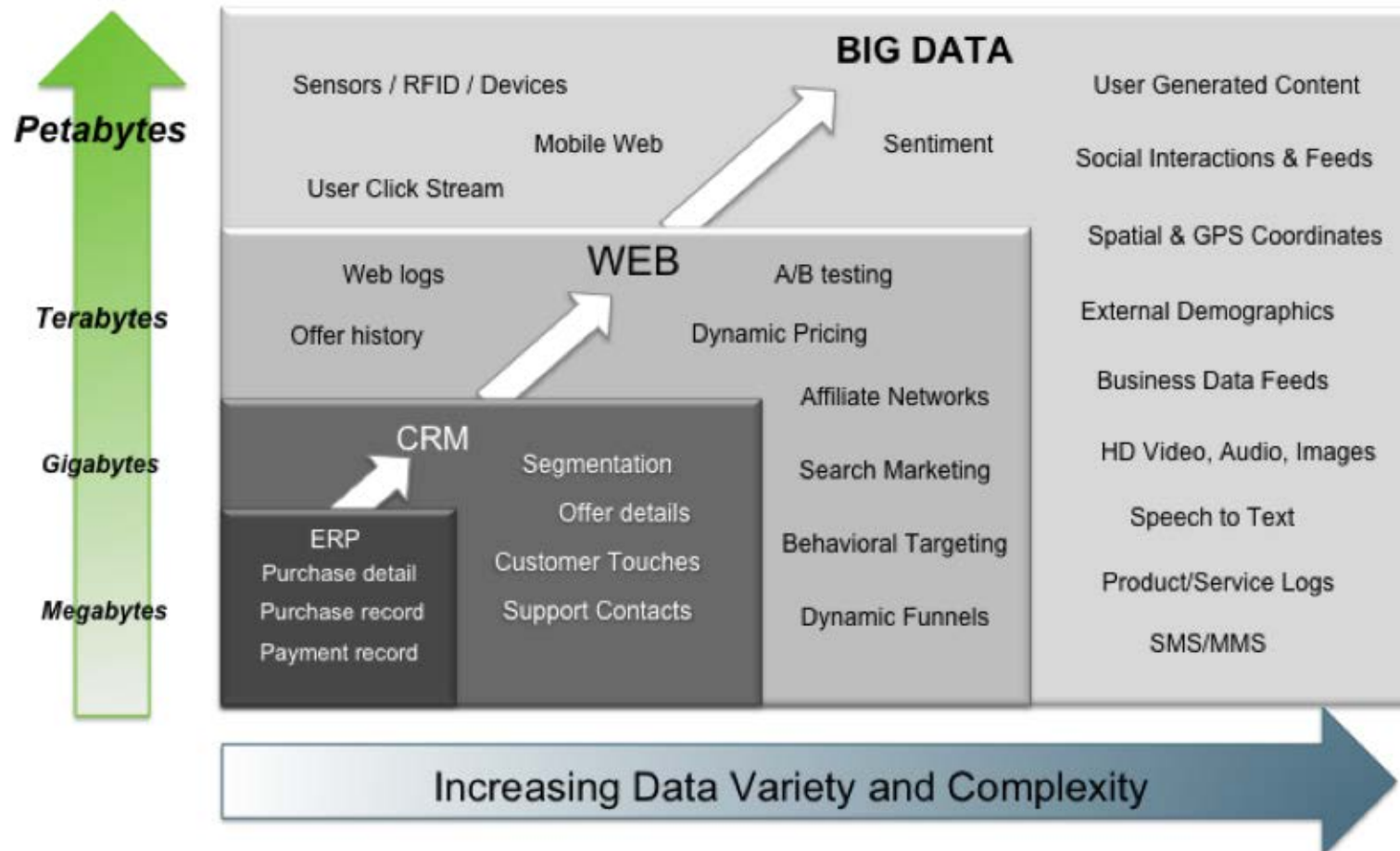


Businesses store
80% of all data.

เทคโนโลยี Big Data

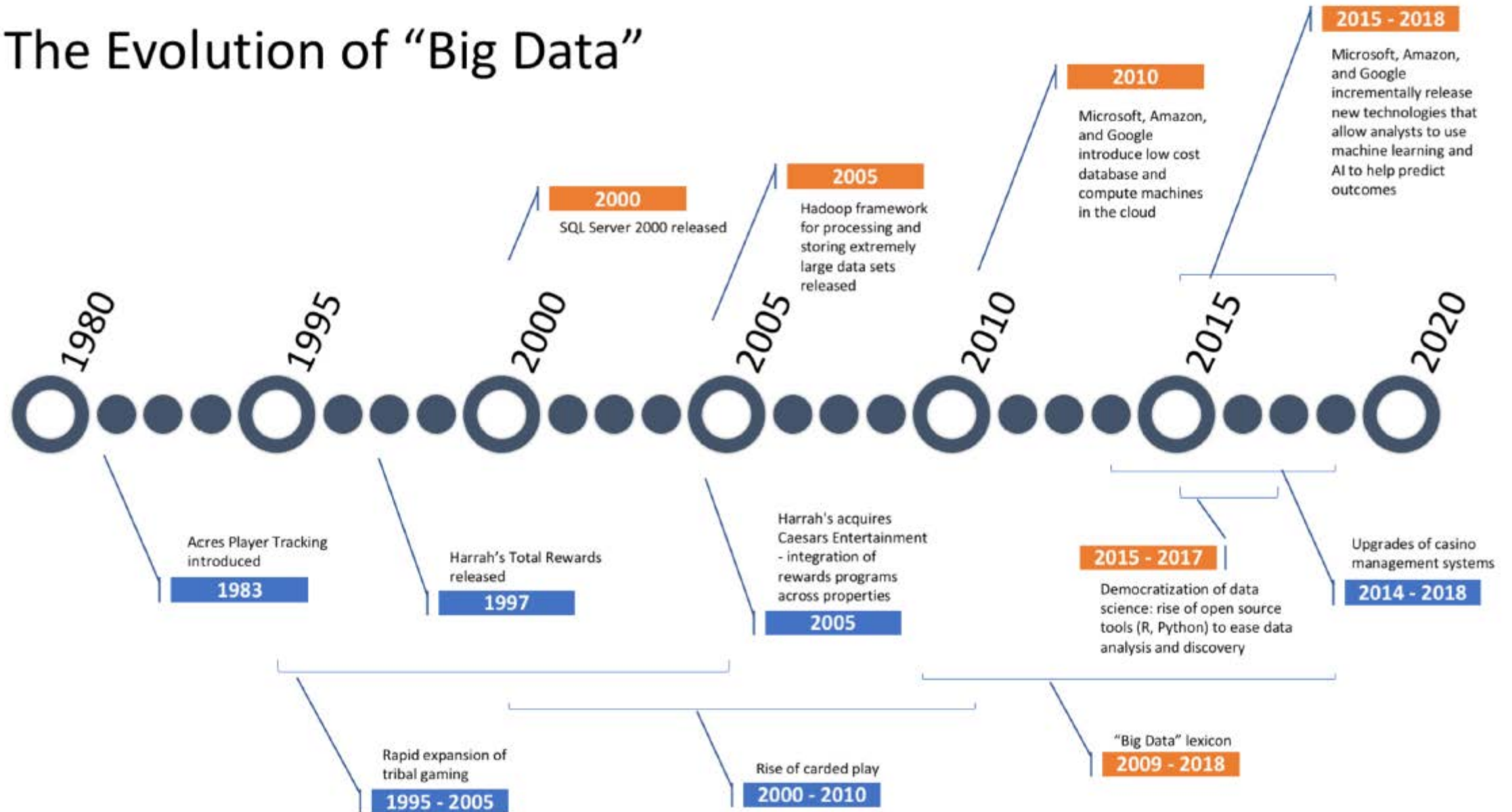


Big Data = Transactions + Interactions + Observations

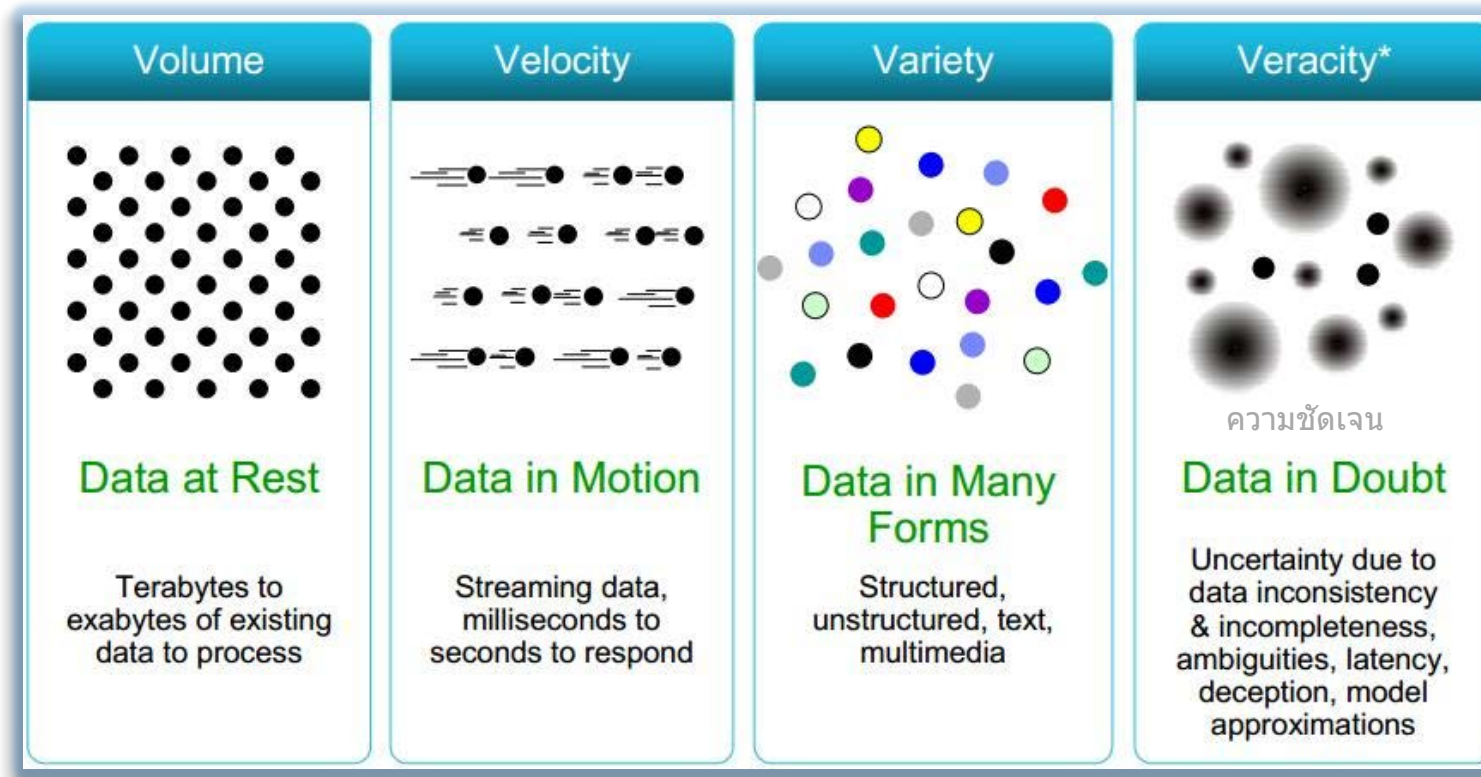


Source: Contents of above graphic created in partnership with Teradata, Inc.

The Evolution of "Big Data"



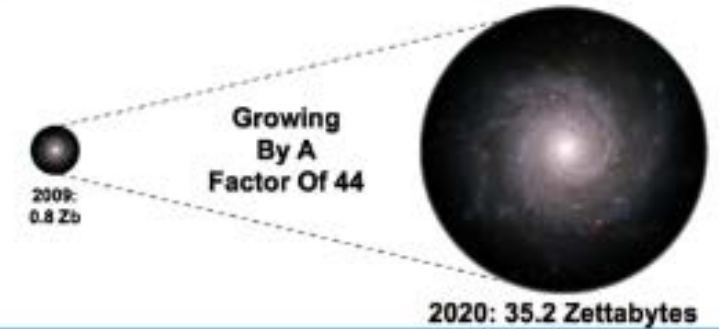
Some Make it 4V's



มีการเพิ่มข้อมูลประมาณ 44 เท่า จากปี 2009 ถึง 2020 หรือประมาณ 0.8 ZB (zettabytes) เป็น 35 ZB เพิ่มแบบ Exponentially



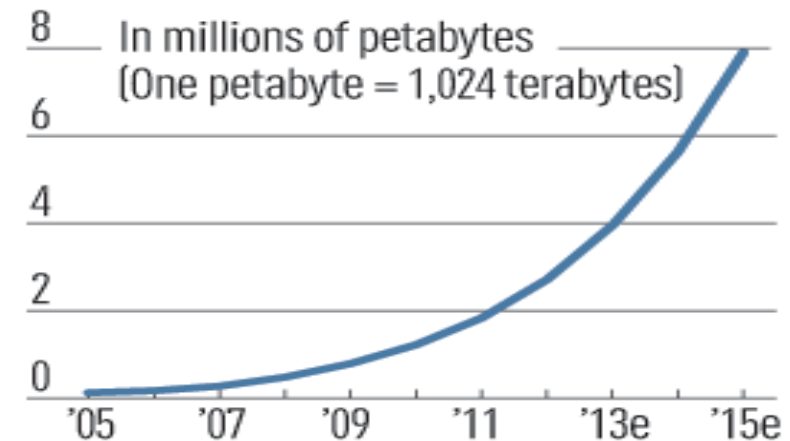
The Digital Universe 2009-2020



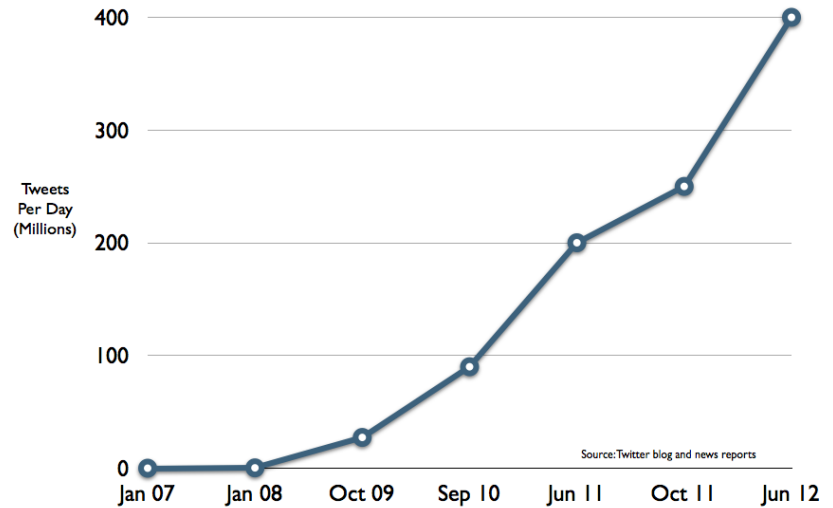
Source: EMC Digital Universe Study, sponsored by EMC, May 2010

EMC

Data storage growth



Twitter: Tweets Per Day

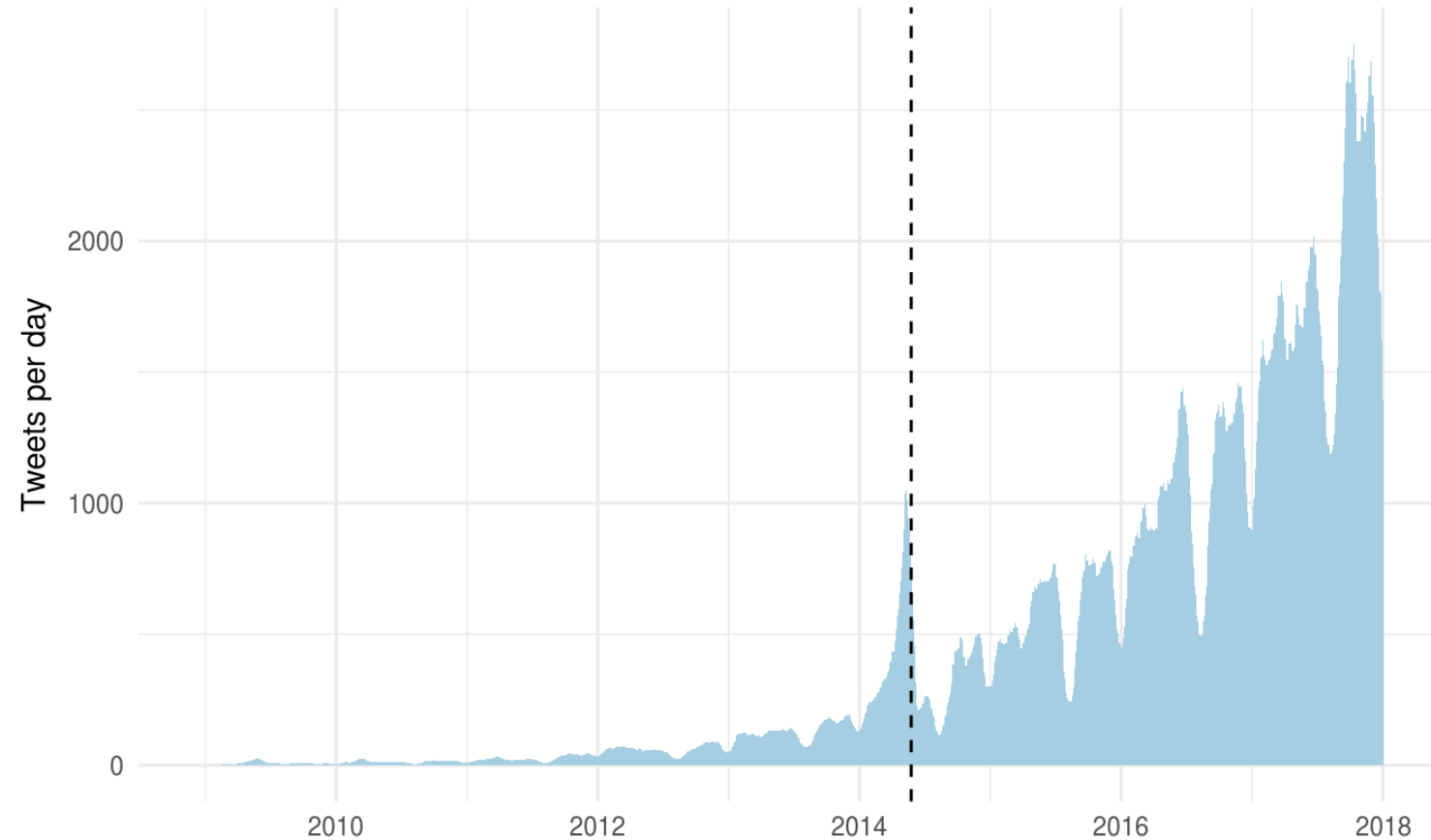


Source: Twitter blog and news reports
Copyright © 2012 Dave Feinleib
dave@vcdave.com
blogs.forbes.com/davefeinleib

16

Number of tweets posted by all current MEP per day

The dotted line presents the final day of the latest European Parliament elections



*Calculated on a 31 days rolling average for clarity

? TBs of data every day

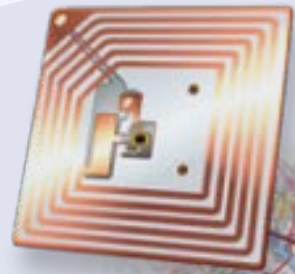


12+ TBs
of tweet data
every day



25+ TBs of
log data
every day

30 billion RFID tags
today
(1.3B in 2005)



76 million smart meters in
2009...
200M by 2014



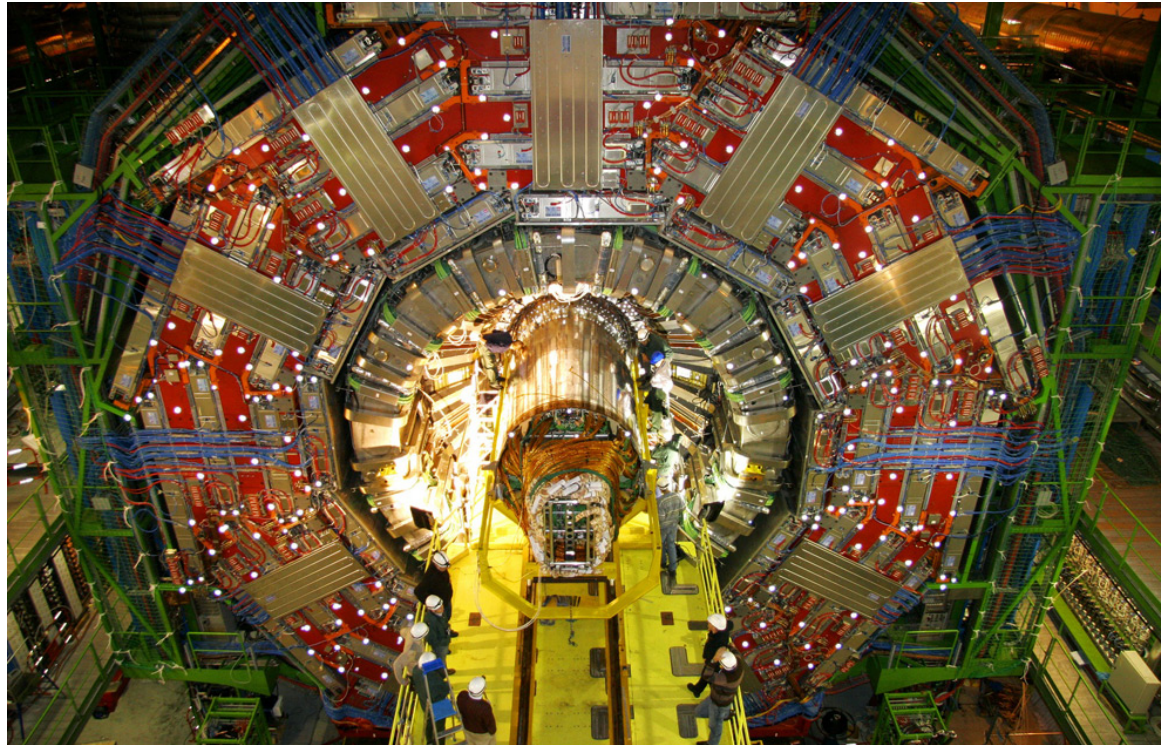
4.6 billion
camera
phones
world wide



100s of millions of
GPS enabled
devices sold
annually

http://www.

2+ billion
people on
the Web
by end
2011

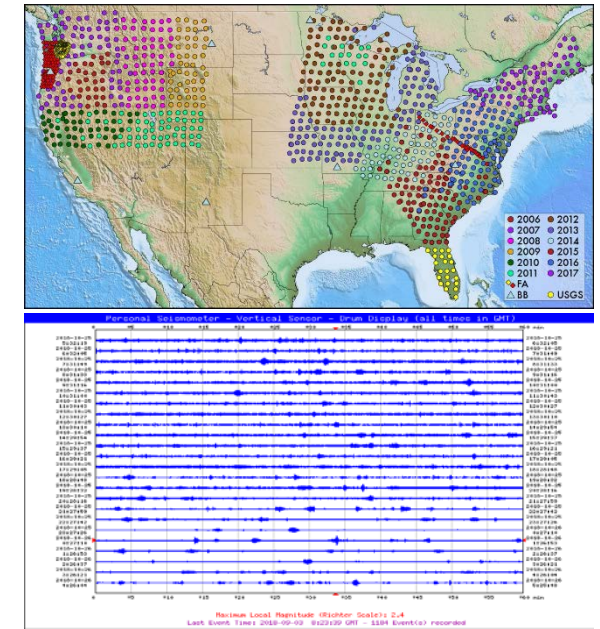


CERN's Large Hadron Collider (LHC)
สร้างข้อมูลขนาด 15 PB ต่อปี

เครื่องชนอนุภาคแฮดรอนขนาดใหญ่ (อังกฤษ: Large Hadron Collider; LHC) คือเครื่องเร่งอนุภาคที่ใหญ่ที่สุดในโลก มีเป้าหมายที่จะสร้างอนุภาคโปรตอน 7 TeV ขึ้น เพื่อพิสูจน์ข้อเท็จจริงและข้อจำกัดของทฤษฎีทางฟิสิกส์อนุภาคที่มีอยู่ในปัจจุบันอันอยู่ภายใต้กฎของแรงทั้งสี่ องค์การวิจัยนิวเคลียร์แห่งยุโรป (European Organization for Nuclear Research) หรือ ซีอีอาร์เอ็น (Conseil Européen pour la Recherche Nucléaire) เป็นผู้สร้างเครื่องนี้ขึ้นที่บริเวณเขตแดนประเทศฝรั่งเศสและสวิตเซอร์แลนด์ ใกล้กับกรุงเจนีวา เป็นท่อใต้ดินลักษณะเป็นวงแหวนขนาดความยาวเส้นรอบวง 27 กิโลเมตร

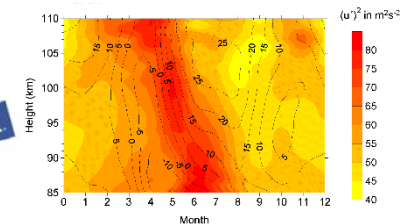
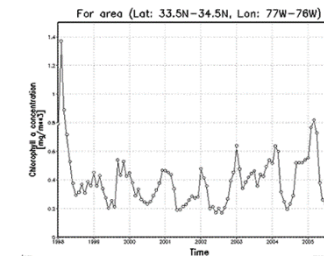
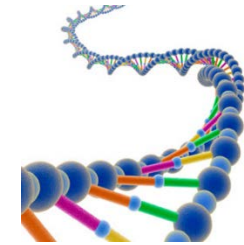
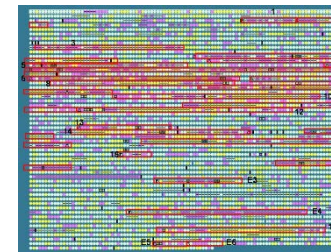
The Earthscope

เป็นโครงการที่ใหญ่ที่สุดในโลก ทางธรณีวิทยา
และธรณีฟิสิกส์เพื่อสำรวจโครงสร้างและ
วิวัฒนาการของ ทวีปอเมริกาเหนือ และเพื่อทำ
ความเข้าใจกับกระบวนการควบคุมการเกิด
แผ่นดินไหวและ ภูเขาไฟ ครอบคลุมพื้นที่กว่า
3.8 ล้านตร.ไมล์ ข้อมูลมากกว่า 67 TB



Variety (Complexity)

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
 - Social Network, Semantic Web (RDF), ...
- Streaming Data
 - Live (Facebook, Youtube, Netflix)
- A single application can be generating/collecting many types of data
- Big Public Data (online, weather, finance, etc)



To extract knowledge → all these types of data need to be linked together

Velocity (Speed)

- Data มีการสร้างขึ้นมาได้อย่างรวดเร็ว และต้องการการประมวลผลที่รวดเร็วตาม
- การวิเคราะห์ข้อมูลออนไลน์ (Online Data Analytics) ซึ่งไม่ใช่แค่การประมวลผล (Processing)
- การตัดสินใจที่ล่าช้า → การพลาดโอกาส



ตัวอย่าง

E-Promotions: ขึ้นอยู่กับตำแหน่งปัจจุบันของลูกค้า ประวัติการซื้อ สิ่งที่ลูกค้าชอบ → การส่งข้อเสนอได้อย่างถูกต้องสำหรับสถานที่ขายให้กับลูกค้า

Healthcare monitoring: อุปกรณ์ตรวจวัดกิจกรรมร่างกาย → การเปลี่ยนแปลงที่ผิดปกติใด ๆ จะได้รับการตอบสนองโดยทันที



Real-time/Fast Data



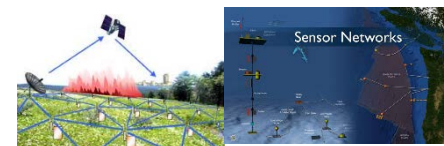
Social media and networks
(all of us are generating data)



Scientific instruments
(collecting all sorts of data)



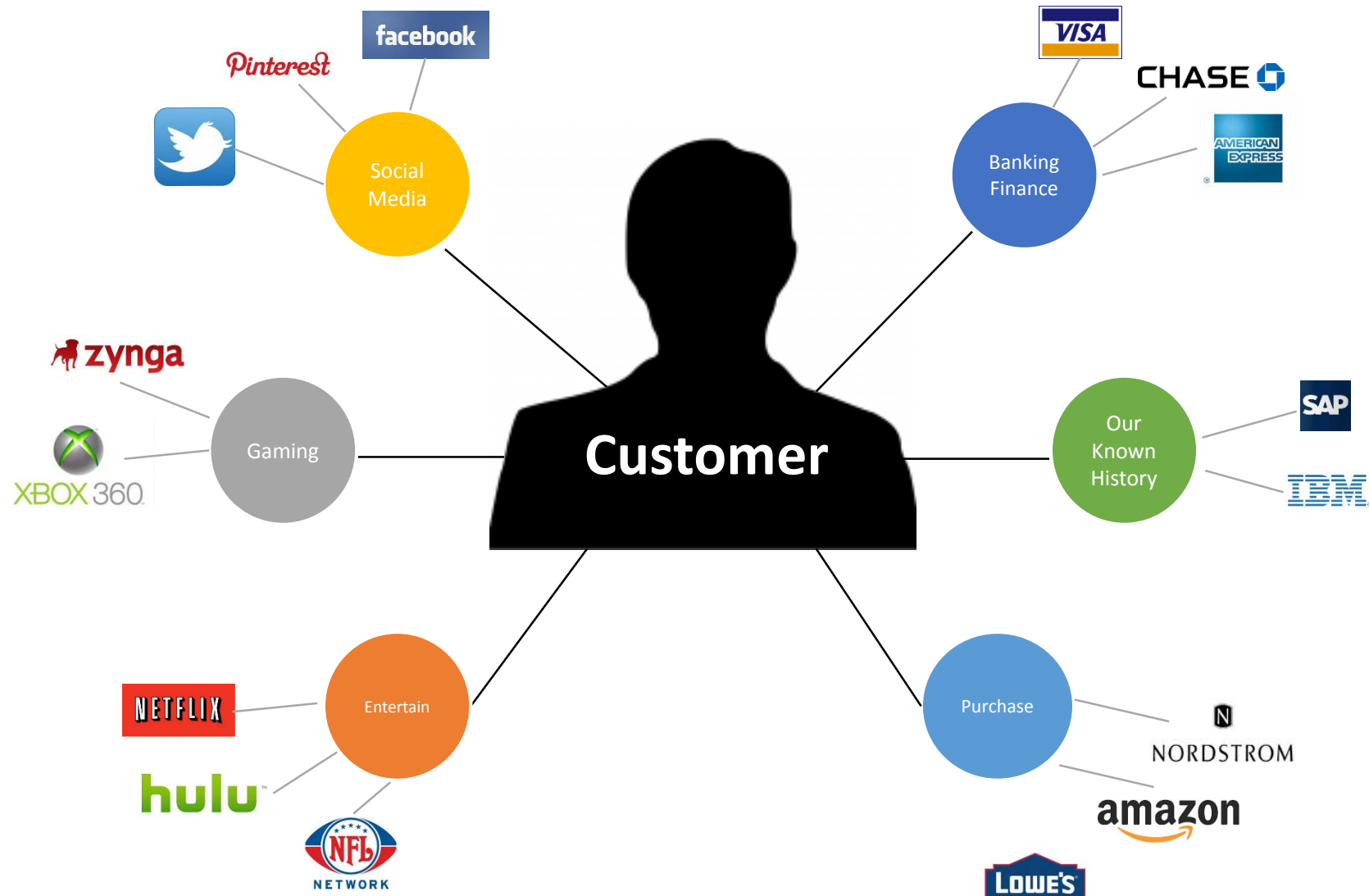
Mobile devices
(tracking all objects all the time)



Sensor technology and networks
(measuring all kinds of data)

- ความก้าวหน้าและนวัตกรรมไม่ได้เป็นอุปสรรคต่อความสามารถในการรวบรวมข้อมูล
- แต่ด้วยความสามารถในการบริหารจัดการ, วิเคราะห์, สรุป, การแสดงจินตนาภาพและค้นพบความรู้จากข้อมูลที่เกิดขึ้นรวบรวมได้ทันเวลาและในรูปแบบที่ปรับขนาดได้ ที่มีผลกระทบ

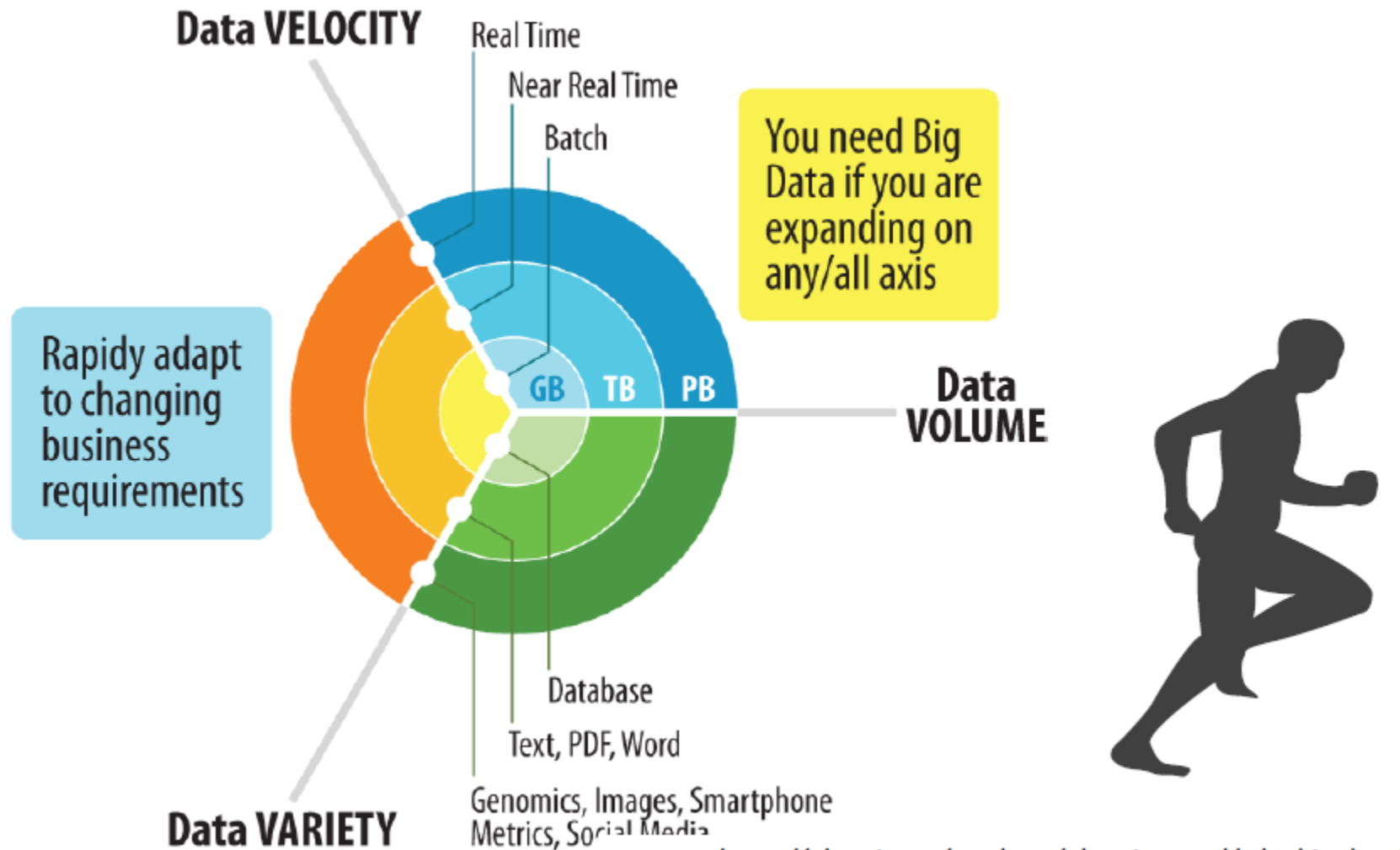
A Single View to the Customer



Real-Time Analytics/Decision Requirement



Big Data ?



YouTube มิกซ์ เพลย์ลิสต์ต่อเนื่องที่ปรับเปลี่ยนให้เหมาะกับคุณ

50+
((-))

มิกซ์ - Alan Walker – Lost Control (Lyrics) ft. Sorana
YouTube

50+
((-))

มิกซ์ของฉัน
YouTube

50+
((-))

มิกซ์ - The Chainsmokers & Coldplay - Something Just...
YouTube

50+
((-))

มิกซ์ - วินาทีเดียว เท่านั้น - เก้า จิรายุ Feat.เบนซ์ | I Can See...
YouTube

50+
((-))

มิกซ์ - ป๊อปปี้
YouTube

ขอแนะนำ

53:47

ทีวีวู้ตี่ ปราศรัยปล่อยมุขสุดฮา จ.ชัยนาท

3:07

Everton 0-2 Manchester City Match Highlights

4:10:03

RoV Pro League Season 3 Presented by TrueMove H ...
Garena RoV Thailand

30:16

ทษช ย่อมมาจากอะไร ทีวีวู้ตี่ ปราศรัย 29มค62
เลือกตั้งนอกเขต ติดต่อเขตปลาย...

4:10

Shayne Ward - Until You (Audio)
shaynewardtv

<https://www.youtube.com/watch?v=YfDqONbzYPc>

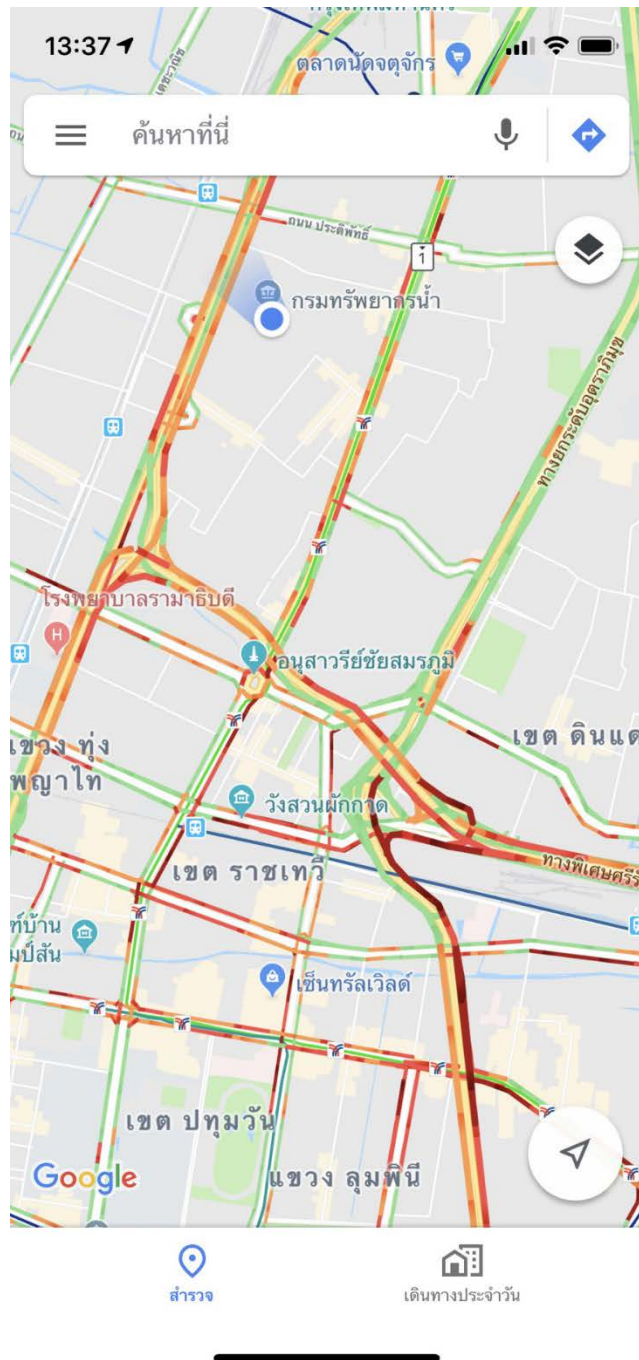
Tanasak Prasertsri : IT Center , Department of Water Resources



(+) ดูเพิ่ม

สวัสดี
 30 บทความสำหรับคุณ

- 1  **เพราะว่ารัก! เปิดใจเจ้าของ "ไอนายน้อย" นำถ้ำกระดูก เก็บในวัด เผยผ...**
 Khaosod | [ทั่วไป](#)
- 2  **สุดเศร้า ว่าที่เจ้าบ่าว โพสต์ขอโทษทุกคนที่ต้องยกเลิกงานแต่งงาน หลังว...**
 อัจฉิน | [ทั่วไป](#)
- 3  **ยอดฝีมือ! คนไทยพา "ปลากัดไทย" คว่ำแชมป์แกะสลักหิมะสมัย 8**
 SpringNews | [ทั่วไป](#)
- 4  **โจรใจบาป ขโมยดวงตาหลวงปู่เหล็ง ทำเครื่องราง ชาวบ้านขอให้เอามา...**
 Khaosod | [ทั่วไป](#)
- 5  **หลายเครือข่าย 'บ้านนาหาร้อยล้าน' พบทองคำ-เงินเป็นพอน ยึดทรัพย์ ...**
 Khaosod | [ทั่วไป](#)



การใช้ Big Data เพื่อแนะนำการจราจรแบบเรียลไทม์

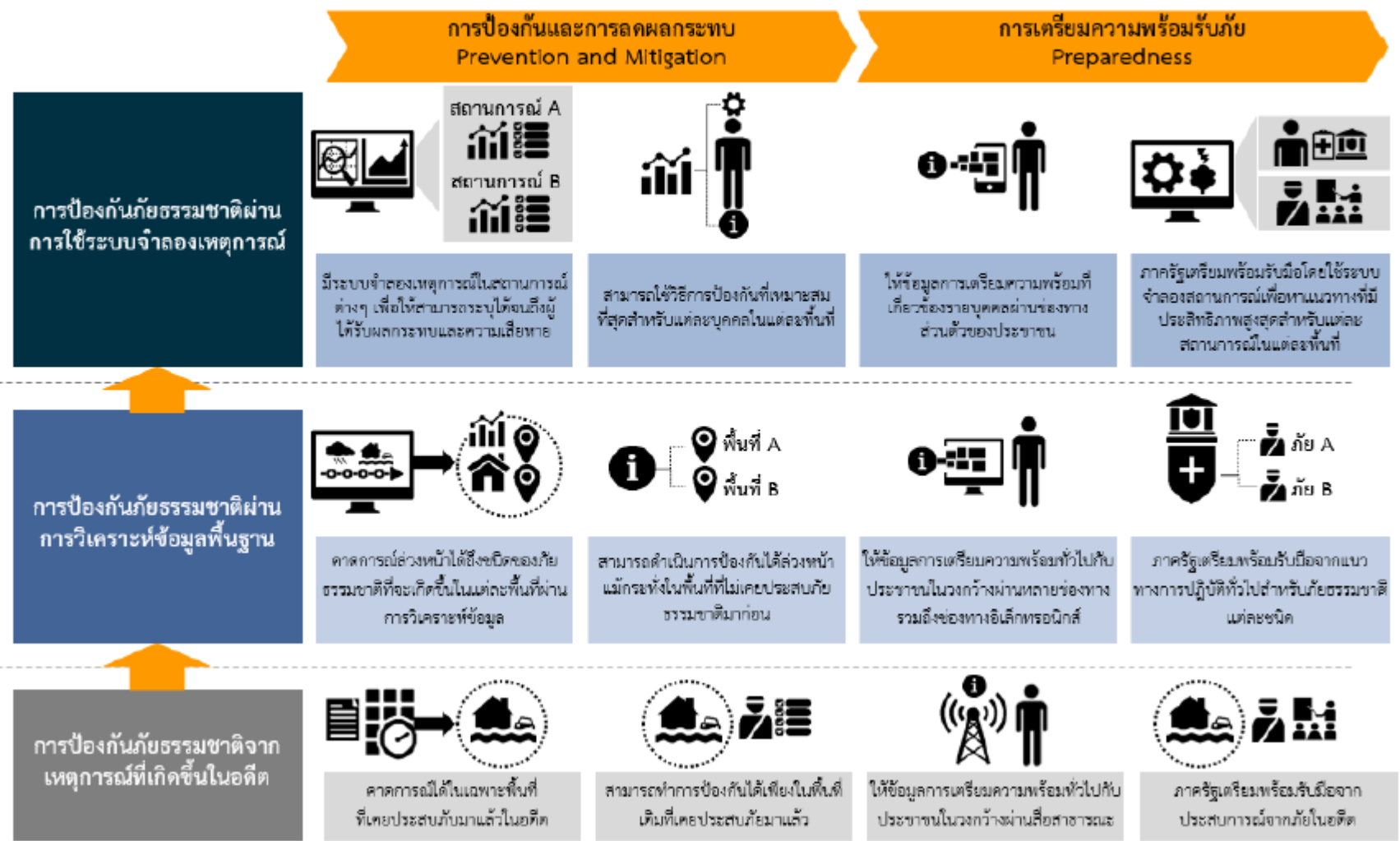
ข้อมูลที่ได้จากระบบคือ

- ความหนาแน่นของการจราจร
- ข้อมูลการใช้รถ

ข้อมูลที่ใช้ในการวิเคราะห์

- ข้อมูลพิกัดผู้เดินทางจากโทรศัพท์มือถือ
- ข้อมูลเส้นทางการจราจร
- ข้อมูลจาก Social Network

ตัวอย่างขีดความสามารถการบูรณาการข้อมูลด้าน: การบริหารจัดการภัยพิบัติ



ตัวอย่างชี้ความสามารถการบูรณาการข้อมูลด้าน: การท่องเที่ยว





Data Structures: Characteristics of Big Data

<ul style="list-style-type: none">■ Structured – defined data type, format, structure<ul style="list-style-type: none">■ Transactional data, OLAP cubes, RDBMS, CVS files, spreadsheets	20 %
<ul style="list-style-type: none">■ Semi-structured<ul style="list-style-type: none">■ Text data with discernable patterns – e.g., XML data■ Quasi-structured<ul style="list-style-type: none">■ Text data with erratic data formats – e.g., clickstream data■ Unstructured<ul style="list-style-type: none">■ Data with no inherent structure – text docs, PDF's, images, video	80 %

แบบจำลอง (Model) ของการสร้างและบริโภคข้อมูลที่เปลี่ยนแปลงไป ...

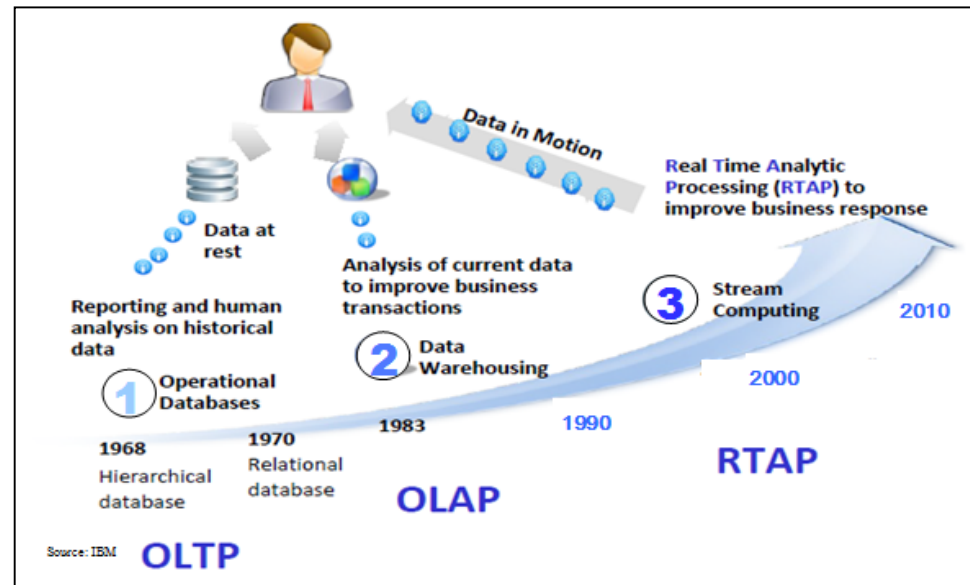
Old Model: Few companies are generating data, all others are consuming data



New Model: all of us are generating data, and all of us are consuming data

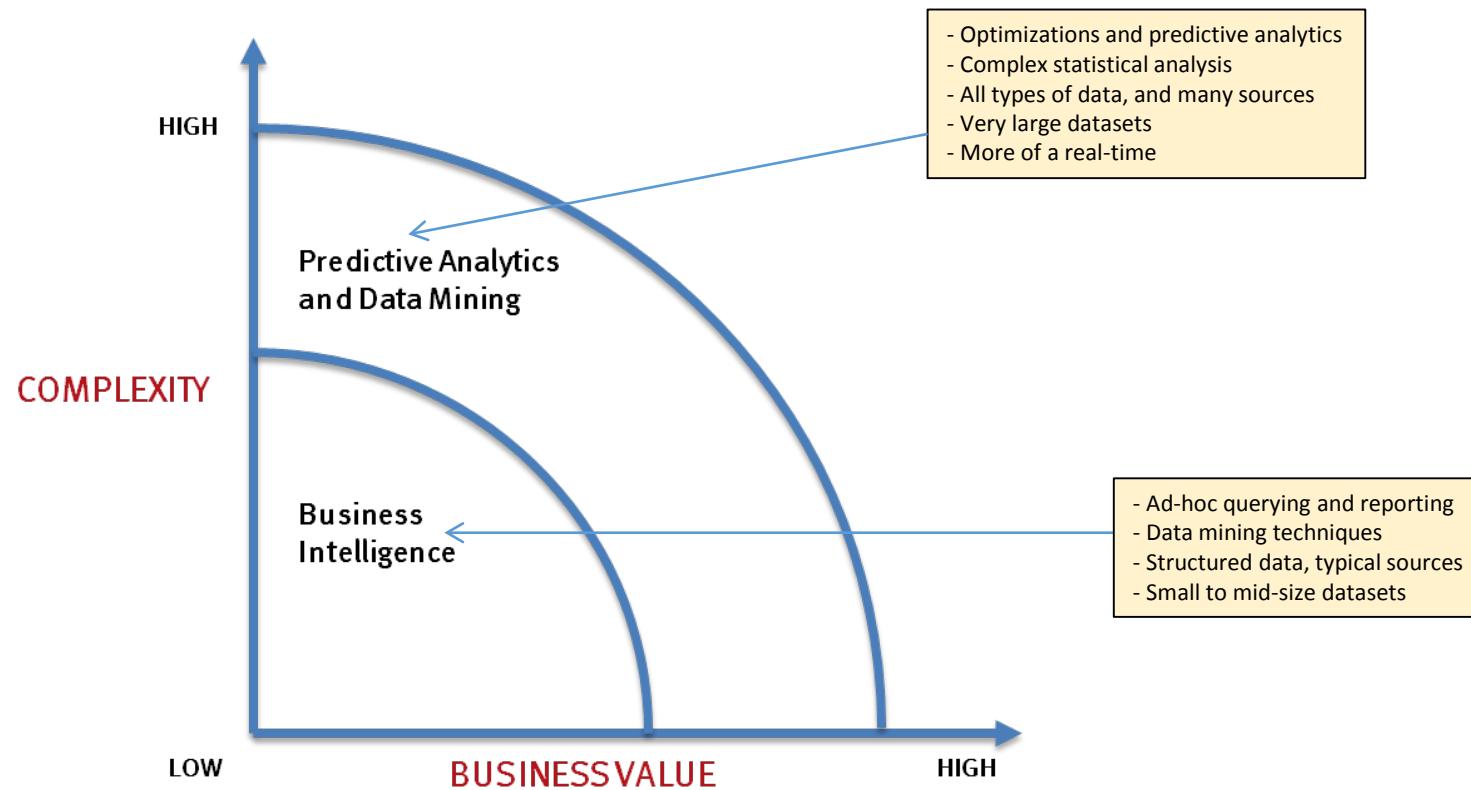


การควบคุม Big Data



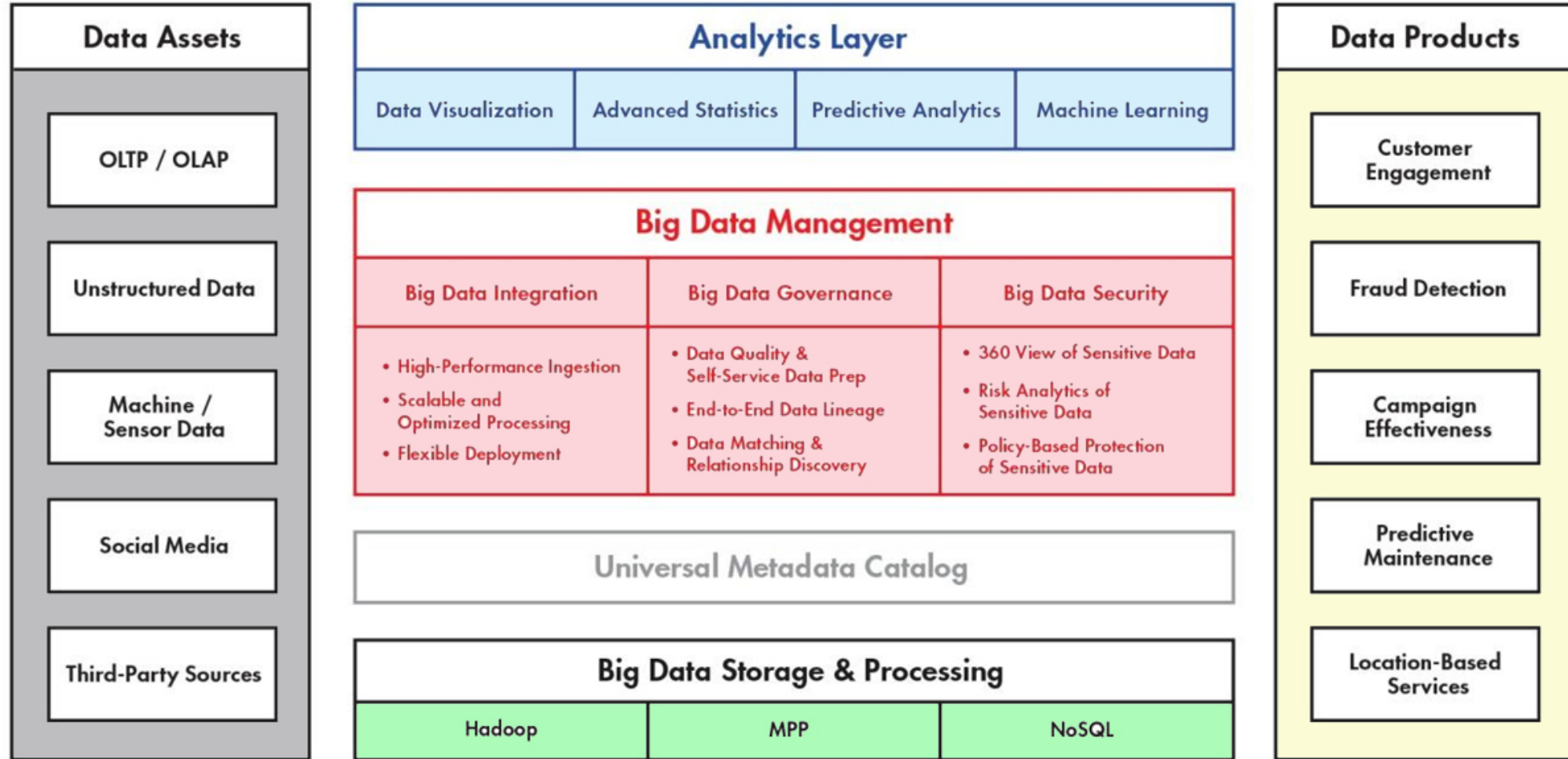
- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data Warehousing)
- **RTAP:** Real-Time Analytics Processing (Big Data Architecture & technology)

สิ่งที่ขับเคลื่อน Big Data

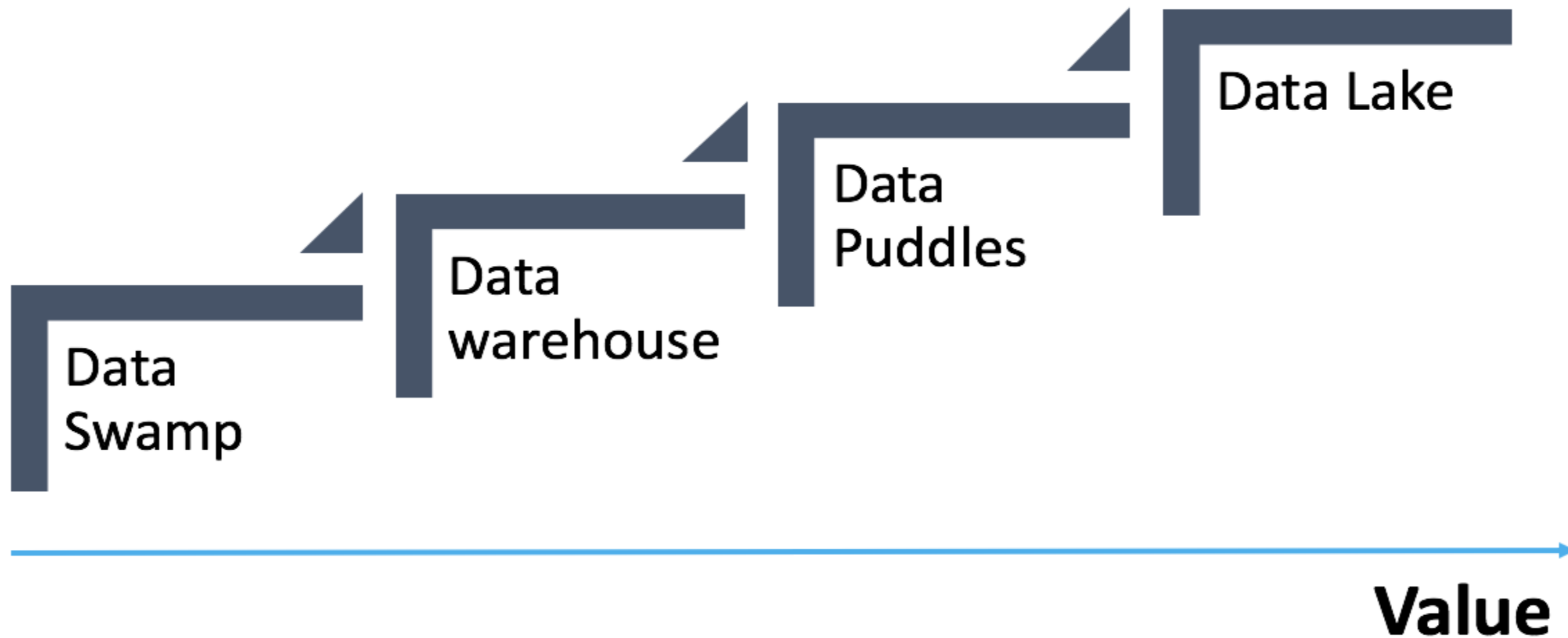




Big Data Analytics



การจัดประเภทของการจัดการ Data ออกตามระดับของ Business Value

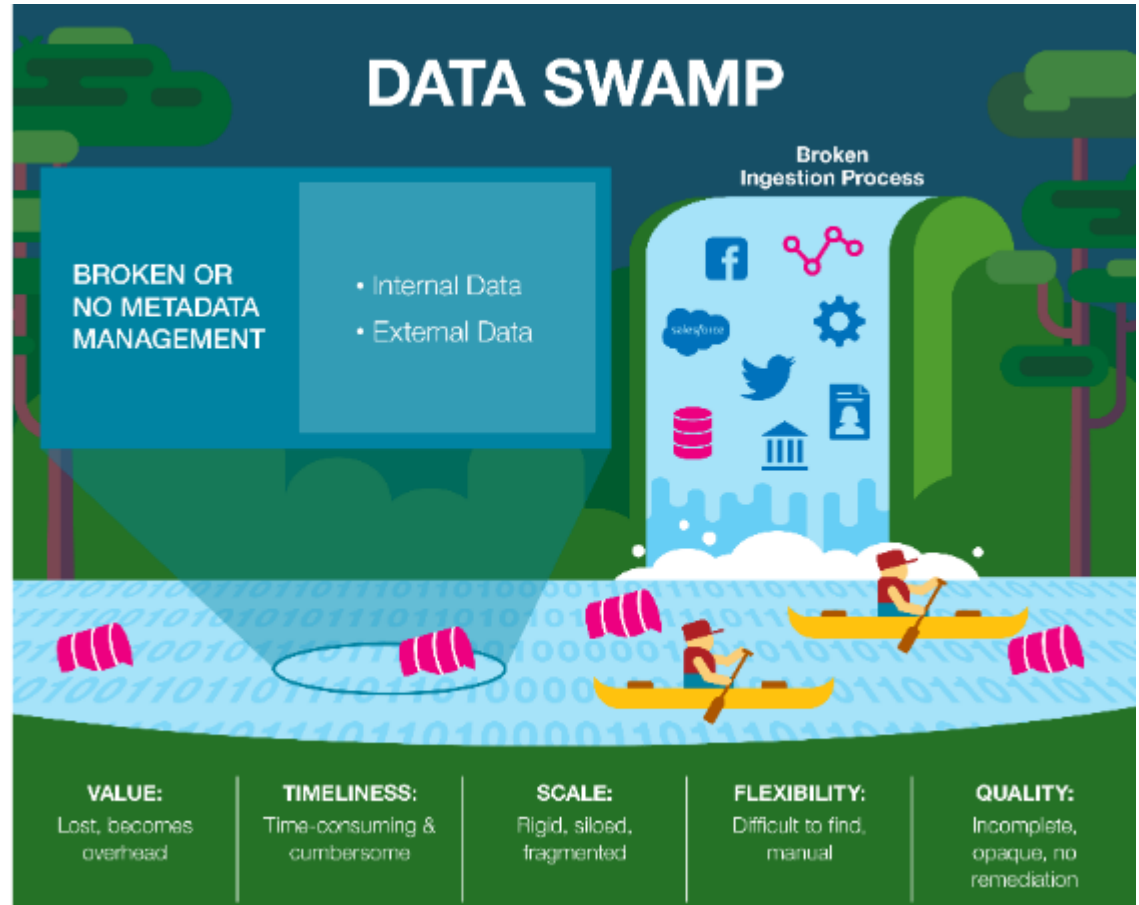


Data Swamp

ยังไม่มีคุณค่าในการทำงาน ข้อมูลจะเป็นเพียง Raw Data

- ไม่สามารถค้นหาประโยชน์ได้
- ไม่สามารถกำหนดสิทธิ์ความปลอดภัยในการเข้าถึง
- ไม่มีระบบป้องกันใด ๆ

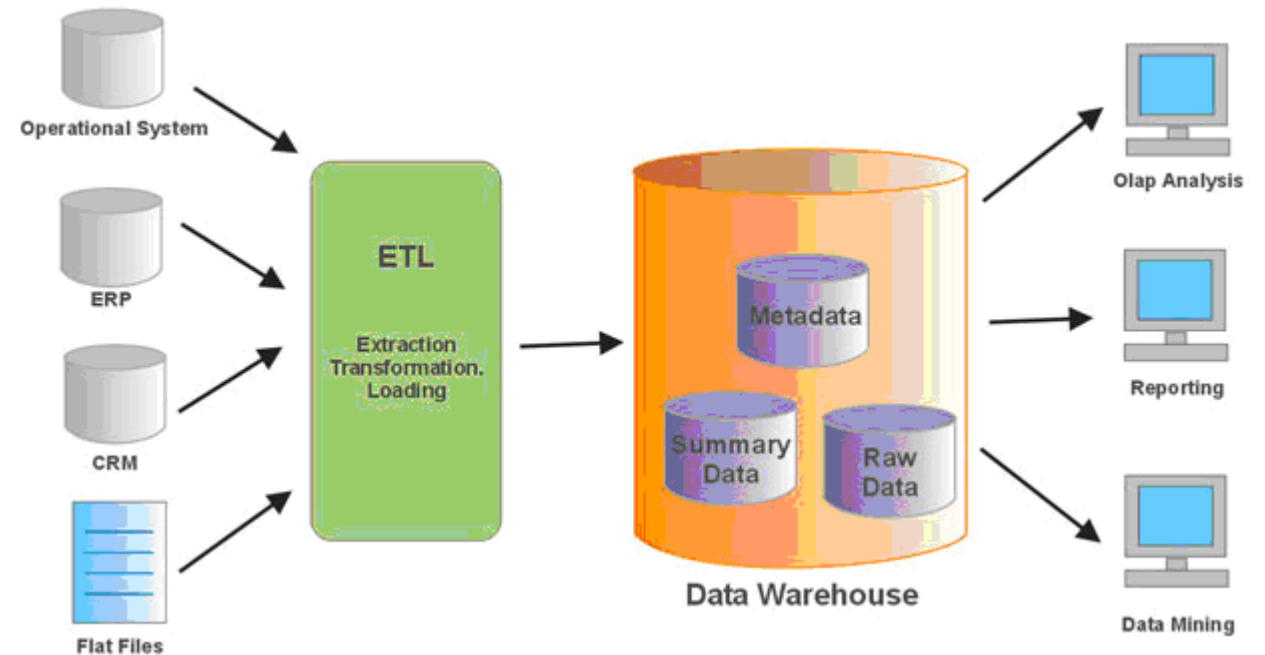
ข้อมูลที่เก็บอยู่ในระบบตัวเอง แยกกันเก็บ ยังไม่มีระบบการจัดการ



Data Warehouse Off-loading

สามารถในการทำคาดการณ์ได้ โดยการแปลงข้อมูล (ETL) จากต้นฉบับ จึงไม่จำเป็นต้องมีการจัดเก็บข้อมูลต้นฉบับ

แต่เมื่อระบบมีการขยายมากขึ้น อาจประสบปัญหา ในกรณีที่ต้องการใช้ข้อมูลต้นฉบับเดิม และไม่เหมาะในการใช้งานกับข้อมูลประเภทที่ไม่มีโครงสร้างได้



Data Puddles



การจัดการข้อมูลของกลุ่มของข้อมูลไม่
มากนัก ความหลากหลายน้อย (Low
variety) การยอมรับต่ำ (Low adoption)

- ดำเนินการในเฉพาะเรื่อง เช่น การทำ
Fraud Detection
- การทำงานแบบอัตโนมัติ
- เป็นระบบงานเฉพาะด้าน

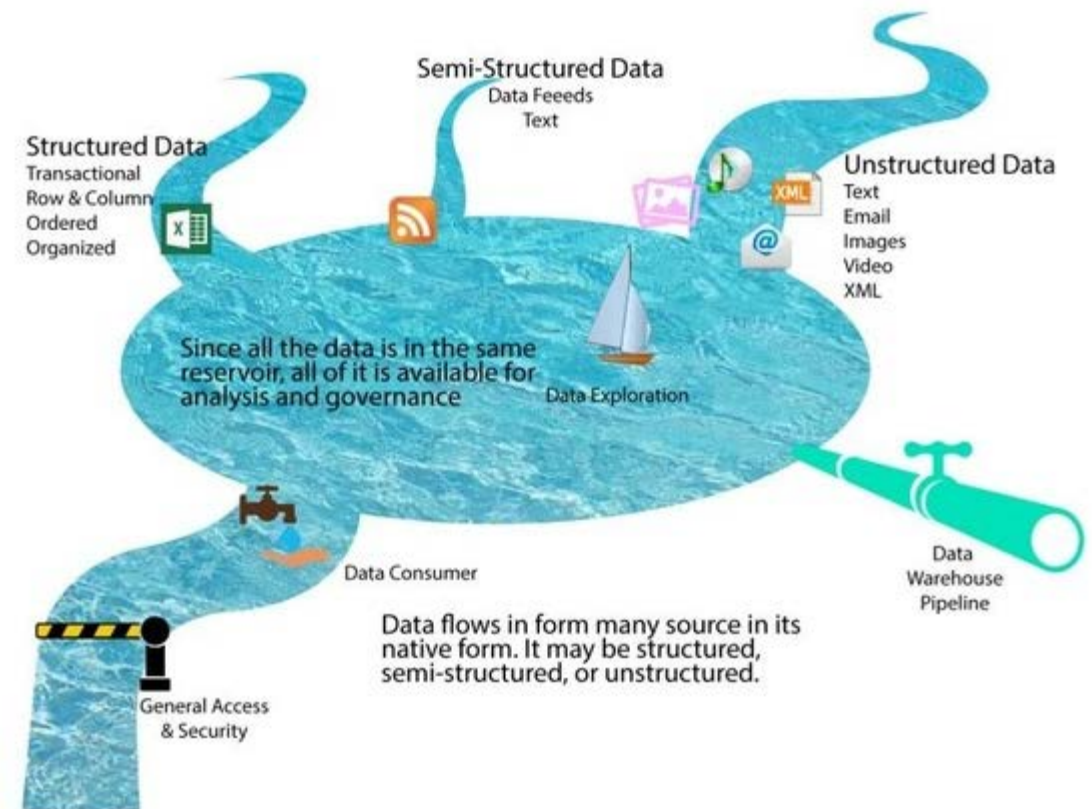
คนใช้งานต้องมีความเชี่ยวชาญสูง

Data Lake

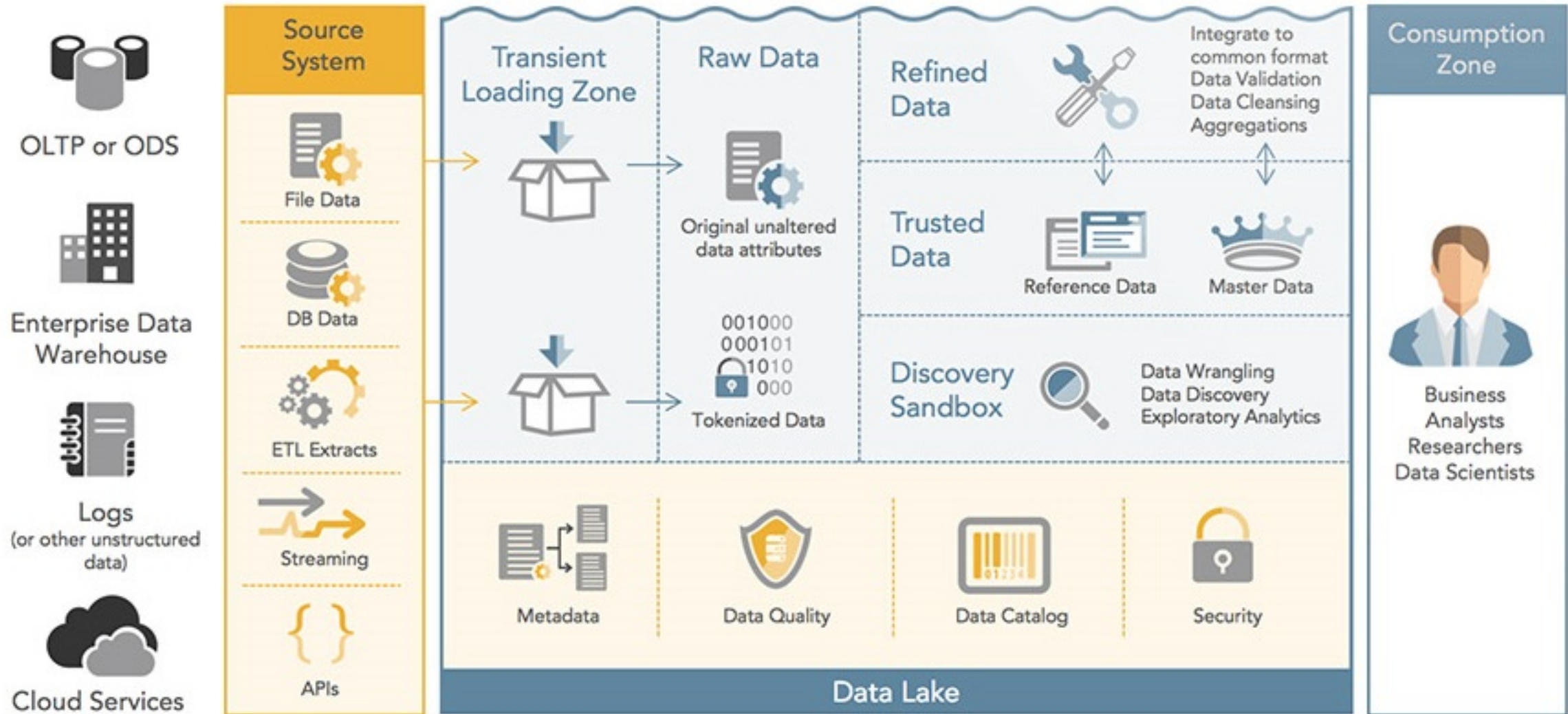
คุณค่าจะสูงสุด ซึ่งระบบนี้จะครอบคลุมการจัดการข้อมูลทุกชนิดทั้งหมดที่กล่าวมา

- การจัดเก็บข้อมูลที่มีและไม่มีโครงสร้างในทุุกขนาดได้ โดยไม่ต้องวางโครงสร้าง
- ใช้การวิเคราะห์ประเภทต่างๆ ได้ ตั้งแต่แดชบอร์ดและการแสดงภาพ
- การประมวลผล Big Data
- วิเคราะห์แบบเรียลไทม์
- Machine Learning

เพื่อสร้างแนวทางการตัดสินใจที่ดีขึ้น



Data lake reference architecture



OLTP (Online Transaction Processing) การแก้ไขข้อมูลพร้อมๆกันอยู่ตลอดเวลา
 ODS (Operational Data Store) แหล่งพักข้อมูลจากระบบงานหลัก ก่อนจะนำข้อมูลเข้าไปเก็บไว้ในคลังข้อมูล
 ETL (Extract-Transform-Load) ดึงข้อมูลออกมาจากหลายๆ ที่, ตรวจสอบคุณภาพ เชื่อมโยงและปรับข้อมูลให้เป็นไปในรูปแบบเดียวกัน

Knowledge Discovery

Data Analytics

Big Data Analytics

Big Data

Business Intelligence

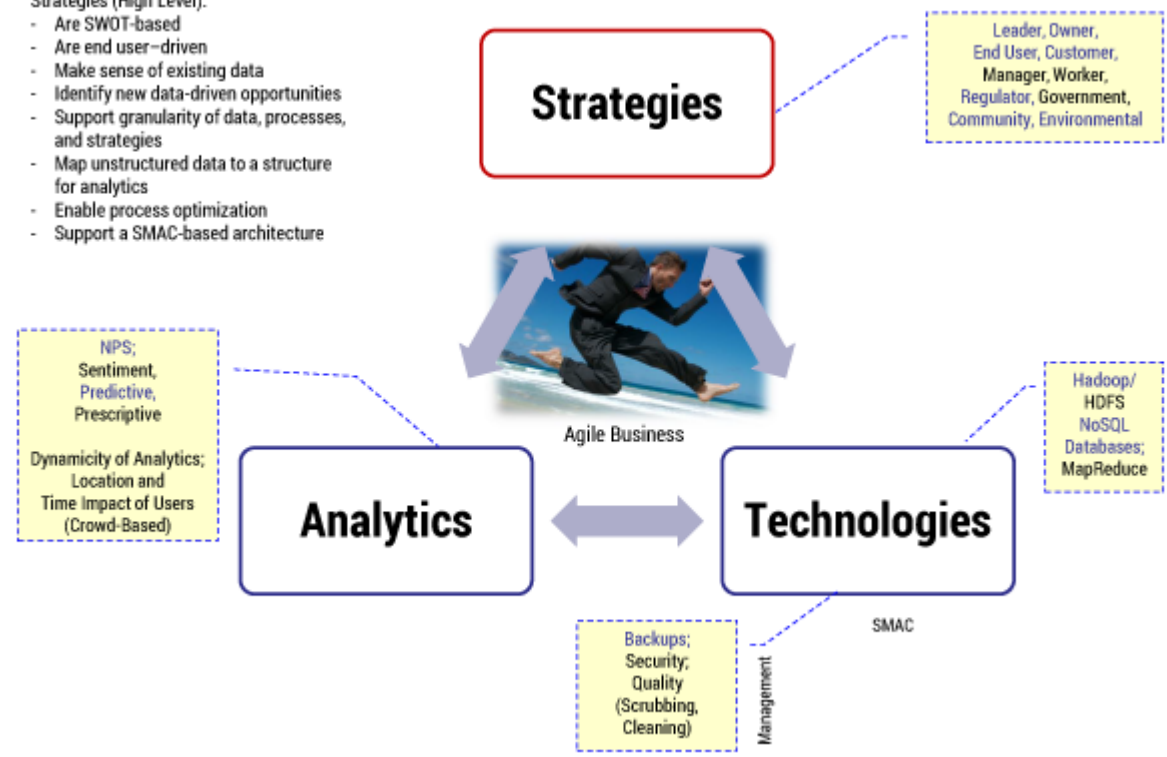
Structured Data



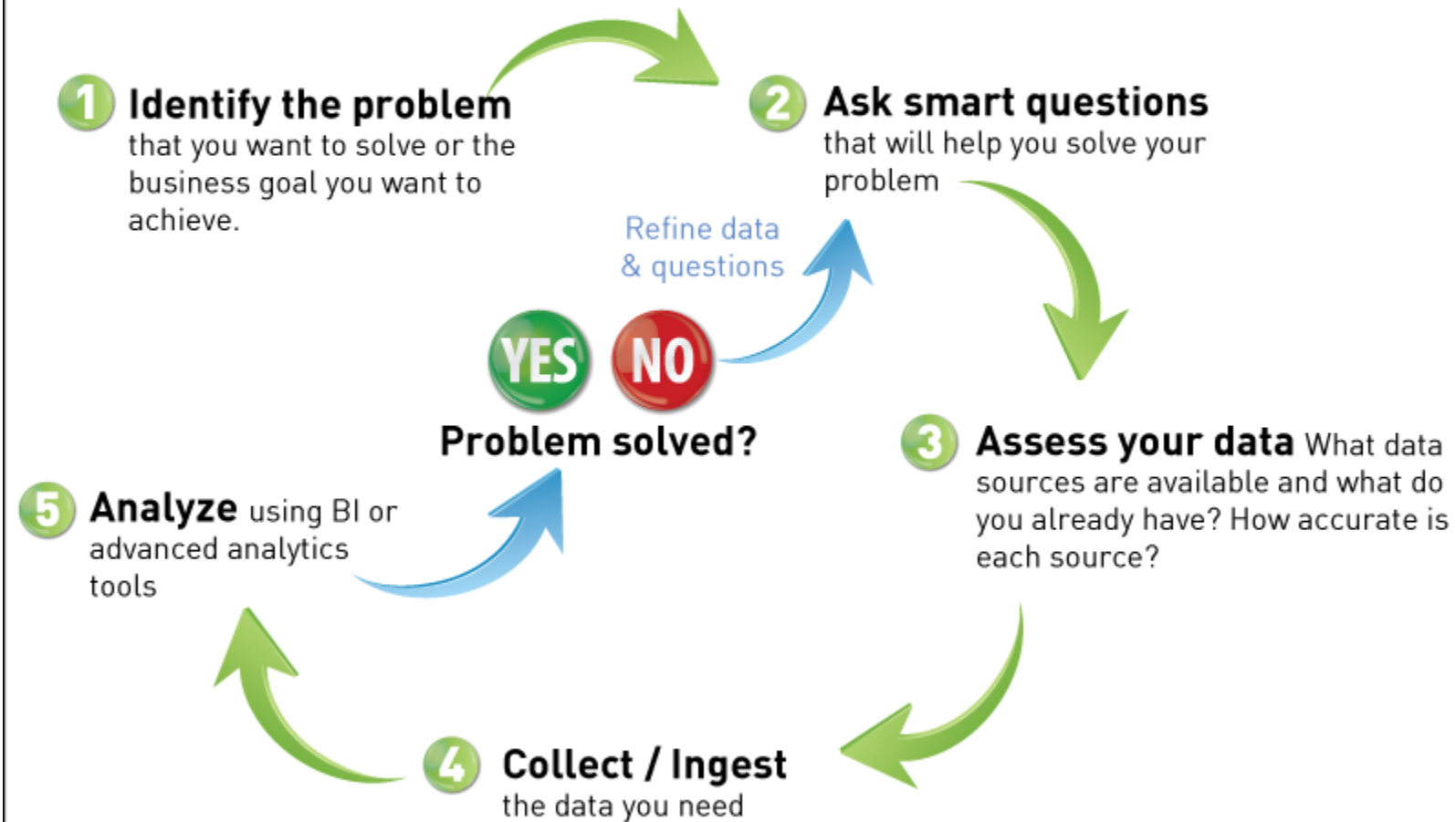
Data Warehouse
and/or Data Marts

Unstructured Data
Semi-Structured Data

- Strategies (High Level):
- Are SWOT-based
 - Are end user-driven
 - Make sense of existing data
 - Identify new data-driven opportunities
 - Support granularity of data, processes, and strategies
 - Map unstructured data to a structure for analytics
 - Enable process optimization
 - Support a SMAC-based architecture



How to Make the Most of Your Big Data





Decision Maker

ผู้กำหนดนโยบาย
ระเบียบ วิธีปฏิบัติ

Primary Custodians

เจ้าของข้อมูล

Data Scientists

ผู้วิเคราะห์ข้อมูลเพื่อ
สนองต่อความ
ต้องการใช้ข้อมูล

Domain Experts

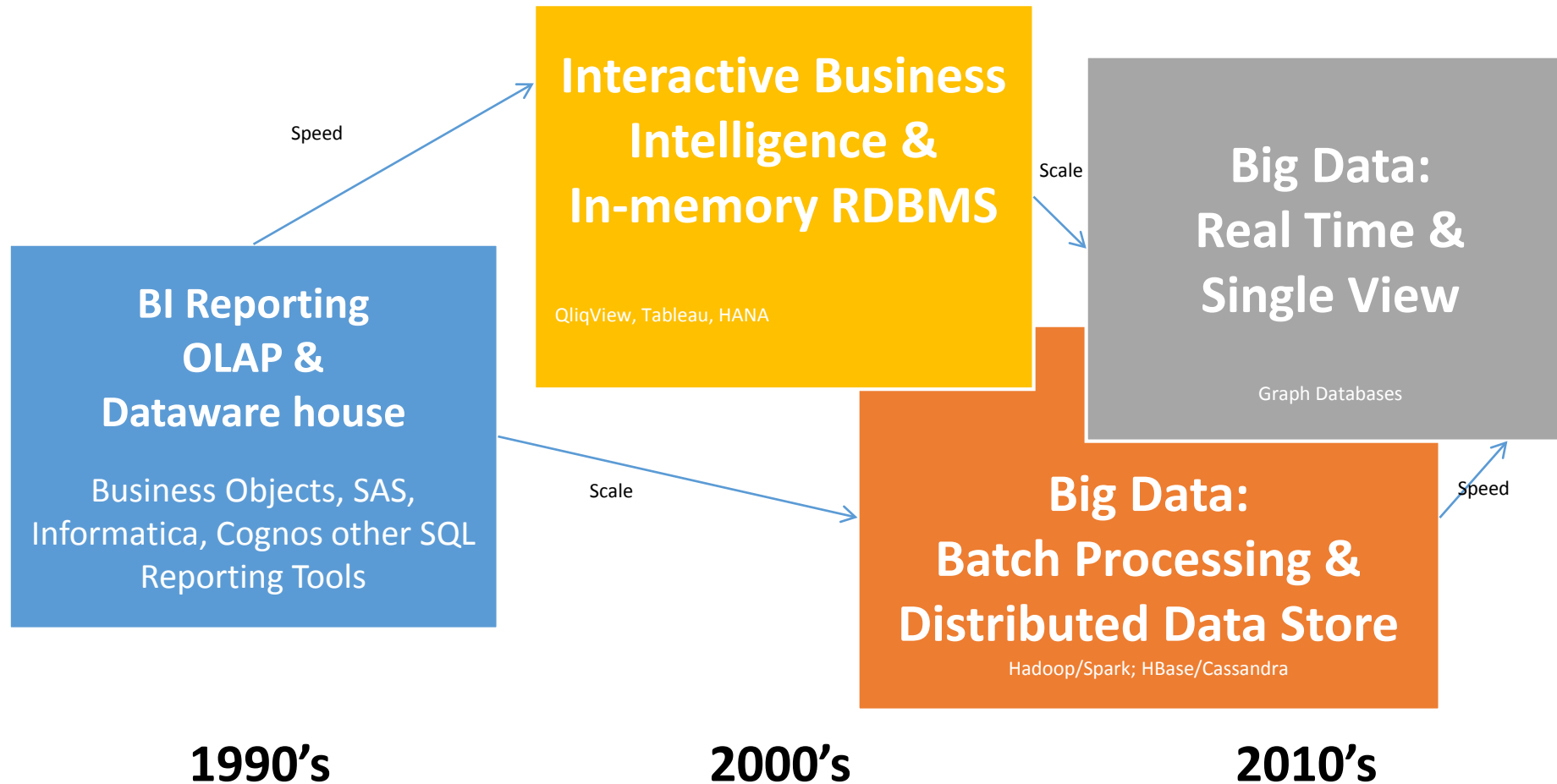
ผู้เชี่ยวชาญในการนำ
ข้อมูลไปวินิจฉัยและ
ประมวลผลเฉพาะทาง
เพื่อการคาดการณ์

IT Specialists

ผู้บริหารจัดการ
เทคโนโลยีสารสนเทศ

Approach	Live Exploration	Direct Batch Reporting	Indirect Batch Analysis
Use Case	For data analysts and data scientists who want to discover real-time patterns as they emerge from their Big Data content	For executives and operational managers who want summarized, pre-built daily reports on Big Data content	For data analysts and operational managers who want to analyze historical trends based upon pre-defined questions in their Big Data content
Latency	Low	Medium	High
Big Data	HBase, NoSQL, Analytic DBMS	Hive, NoSQL, Analytic DBMS	Hadoop, NoSQL, Analytic DBMS
Connectivity	Native	Native, SQL	ETL
Architecture			

วิวัฒนาการของ Business Intelligence



Big Data Analytics

- ข้อมูลขนาดใหญ่มีลักษณะเป็นเรียลไทม์มากกว่าแอปพลิเคชัน DW แบบเดิม
- สถาปัตยกรรมแบบดั้งเดิมของ DW (เช่น Exadata, Teradata) ไม่เหมาะสมกับแอปพลิเคชันข้อมูลขนาดใหญ่
- สามารถแบ่งปันข้อมูลกันได้, วิเคราะห์ข้อมูลแบบเรียลไทม์, สถาปัตยกรรมขนาดใหญ่เหมาะสำหรับแอปพลิเคชันข้อมูลขนาดใหญ่

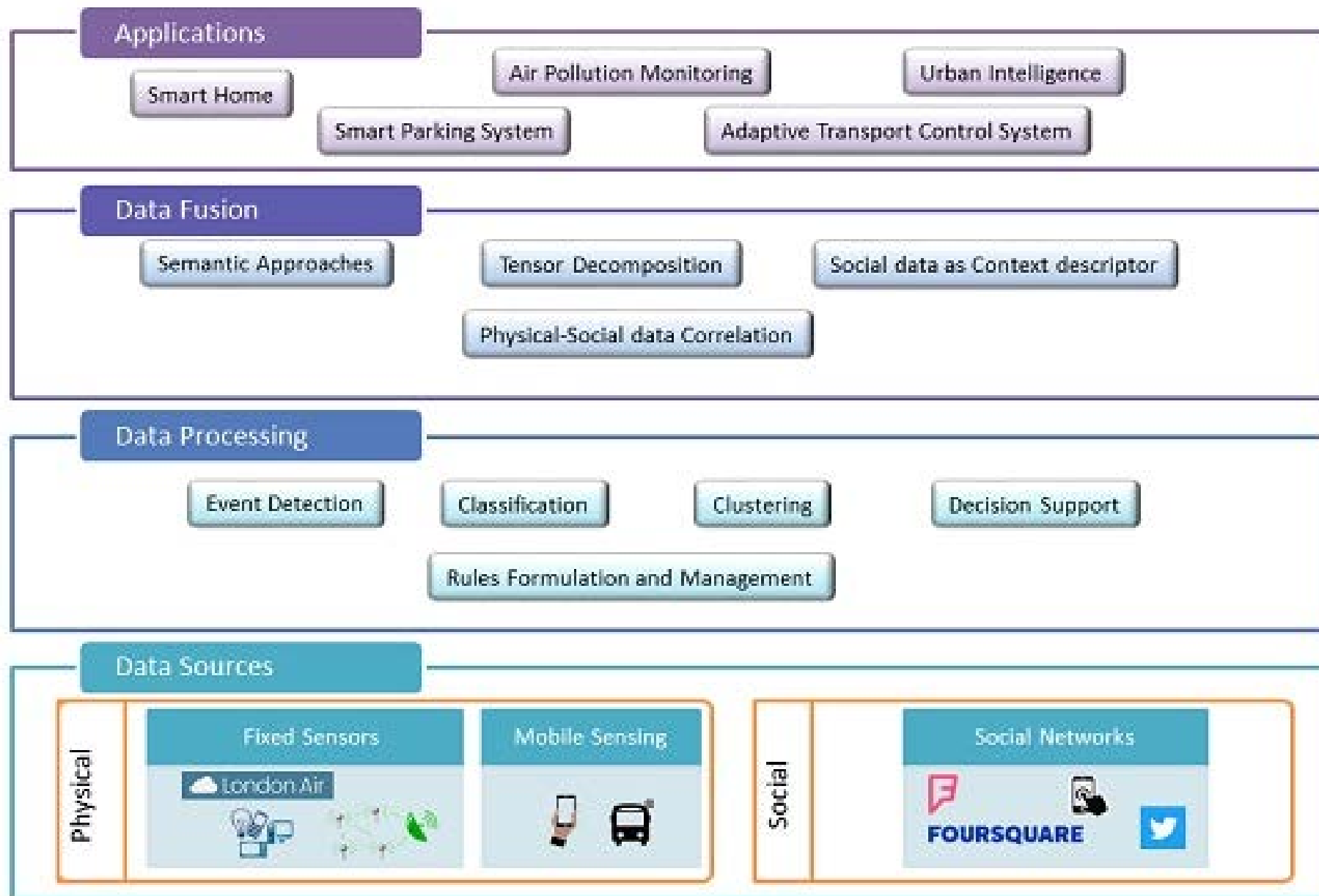
The Progression of Analytics



#BDWMeetup

@joe_Caserta

caserta



The Big Data Landscape

Apps

Vertical Apps



Operational Intelligence



Ad / Media Apps



Business Intelligence



Analytics And Visualization



Data As A Service



Infrastructure

Analytics Infrastructure



Operational Infrastructure



Infrastructure As A Service



Structured Databases



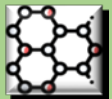
Technologies



สถานการณ์ปัจจุบันและความจำเป็นในการพัฒนาด้านข้อมูลขนาดใหญ่ จำแนกตามมิติตัวขับเคลื่อน Digital Evolution 4 มิติดังนี้

Supply	Demand	Institutional	Innovation
<ul style="list-style-type: none"> • INEFFICIENCY 	<ul style="list-style-type: none"> • HIGH VALUE DATASETS 	<ul style="list-style-type: none"> • LAW & REGULATION CONSTRAINTS 	<ul style="list-style-type: none"> • OLD TECHNOLOGY
<ul style="list-style-type: none"> • INCOMPLETENESS 	<ul style="list-style-type: none"> • BETTER GOVERNMENT SERVICE 	<ul style="list-style-type: none"> • CONCERN ON SECUREITY & PRIVACY 	<ul style="list-style-type: none"> • BUSINESS INTELLIGENCE & DATA ANALYTICS
<ul style="list-style-type: none"> • NO SINGLE SOURCE OF TRUTH 	<ul style="list-style-type: none"> • GOVERNMENT TRANSPARENCY 	<ul style="list-style-type: none"> • DATA-DRIVEN CULTURE 	<ul style="list-style-type: none"> • NO SHARING PLAT FORM
<ul style="list-style-type: none"> • INCOMPATABILITY 	<ul style="list-style-type: none"> • UNFULFILLED BUSINESS POTENTIAL 	<ul style="list-style-type: none"> • INCOMPETENCY 	-
<ul style="list-style-type: none"> • NO DATA STANDARD 	-	-	-

ประเภทข้อมูล



ข้อมูลมีโครงสร้าง Structured Data

- Transaction Data
- รายการซื้อขาย
ธุรกรรมต่างๆ
- ข้อมูลหมู่บ้าน
- ข้อมูลคลัง, พัสดุ,
งบประมาณ
- ข้อมูลการใช้บริการ,
การให้บริการ



ฐานข้อมูล Relational database



ข้อมูลกึ่งโครงสร้าง Semi - Structured Data

- key/value
- Document (Json, xml)
- Graph
- Column-family
- ข้อมูลการรักษา (เช่น
ตาม HL7)



NoSQL database



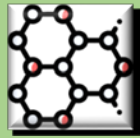
ข้อมูลไร้โครงสร้าง Semi - Structured Data

- E-mail, Logs
- Social Feed, Facebook,
Twitter
- Full Text, Web Page,
Pantip
- Instant, Messaging, Line
- Sensor, IoT, Real-time
- Image, Video
- ระดับการเตือนหัวใจทุก 30
วินาที
- เอกสาร, ไฟล์ Word, Excel



HDFS, BLOB

เทคโนโลยีในการบริหารจัดการข้อมูล



ข้อมูลมีโครงสร้าง
Structured Data



คลังข้อมูล
Data Warehouse



ข้อมูลกึ่งโครงสร้าง
Semi - Structured Data



อภิมหาข้อมูล
Big Data



ข้อมูลไร้โครงสร้าง
Semi - Structured Data



Data Lake

ข้อมูลนำเข้า



EDW



Social/
Clickstream



Emails



Documents



Images



Wearables



ODS



Big Data Platform

Hadoop Data Lake

- + Organize analytics-ready data
- + Protect sensitive data
- + Share data responsibly



Data Mart



Data Scientist



Applications



Knowledge
Workers

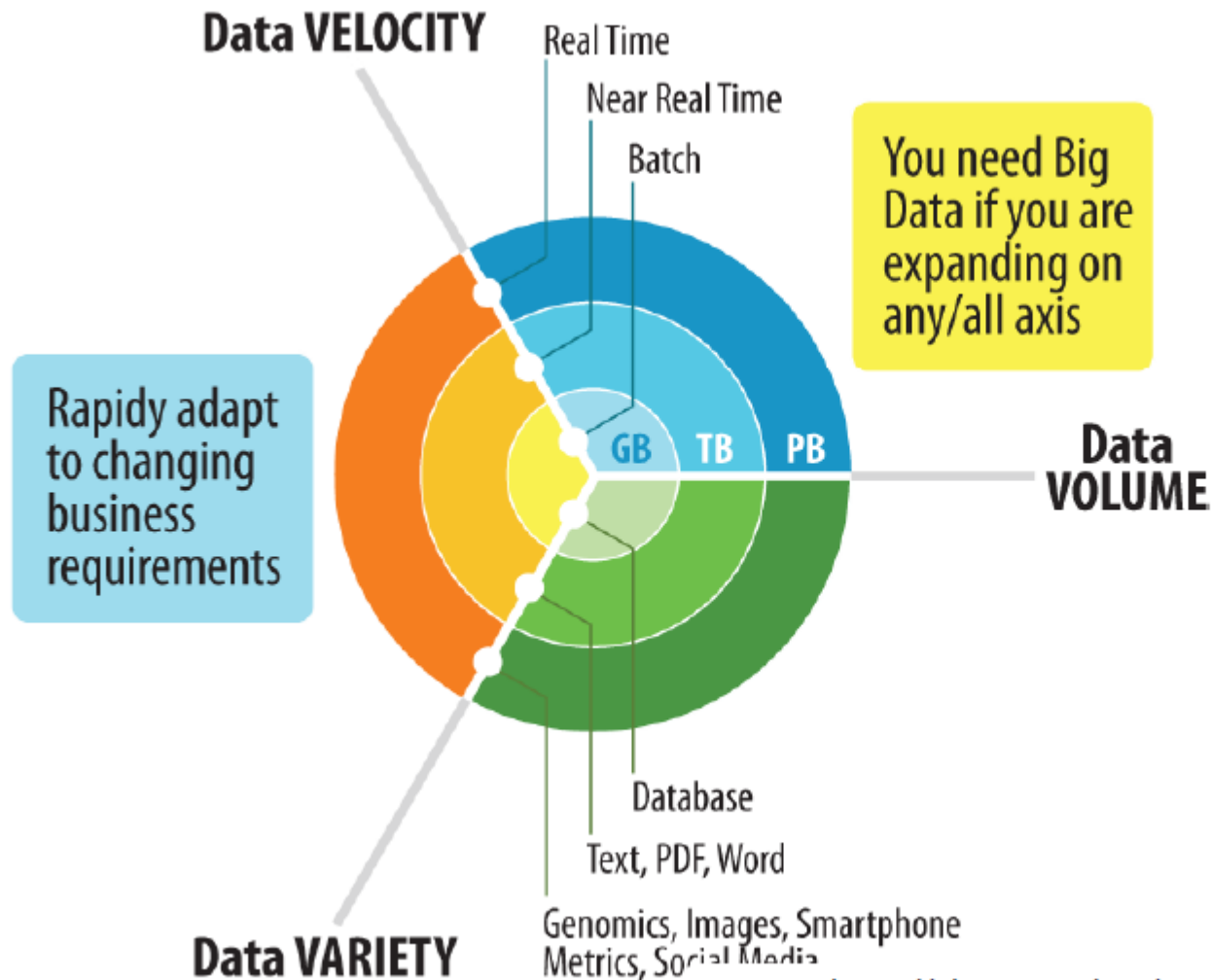


Analyst



Customers

Big Data ?



ข้อมูลจากการสำรวจกรมทรัพยากรน้ำ

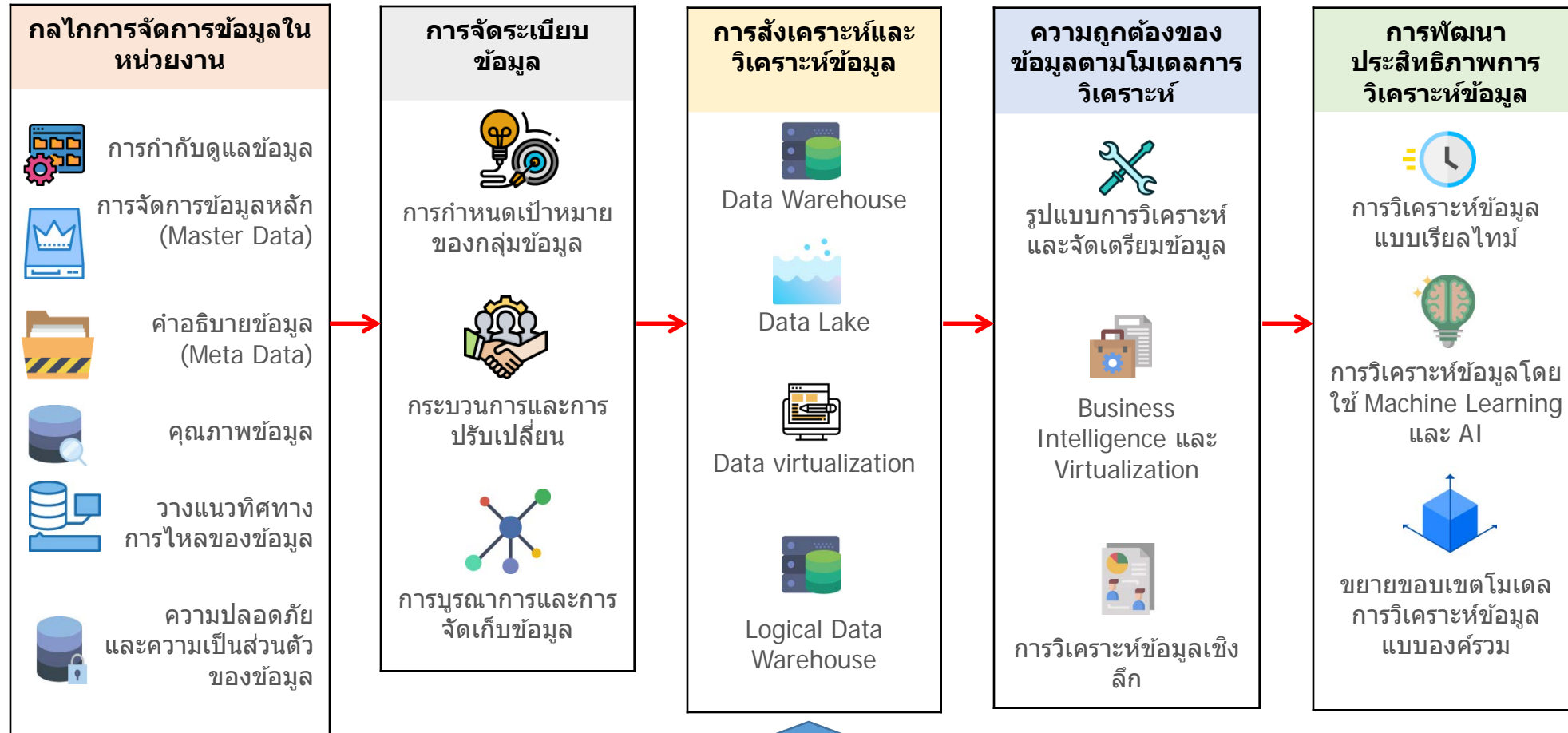
1. ข้อมูลมีขนาด <10 TB ระดับที่ 2 ปานกลาง (200 GB – 100 TB)
2. ไม่มีความต้องการใช้ข้อมูลแบบที่ไม่มีโครงสร้าง (Unstructured Data)
3. ไม่มีภารกิจที่ต้องการความเร็วในการประมวลผลสูงเพื่อวิเคราะห์แบบเรียลไทม์



เกณฑ์จาก [สำนักงานพัฒนารัฐบาลดิจิทัล \(DGA\)](#)

<https://phemi.com/products/phemi-central/why-big-data/>

แนวทางขั้นตอนการดำเนินการด้านข้อมูลสำหรับการทำ Data/Big Data Analytics สำหรับกรมทรัพยากรน้ำ



(ร่าง) ภูมิทัศน์ข้อมูลขนาดใหญ่ (Big Data Landscape) ในระยะเวลา 20 ปี

ภูมิทัศน์ข้อมูลขนาดใหญ่		1 ปี 6 เดือน	5 ปี	10 ปี	10-20 ปี
ยุทธศาสตร์	ระยะที่ 1 Big Data Foundation	ระยะที่ 2 Transformation and Exchange	ระยะที่ 3 Fusion	ระยะที่ 4 Smart	
	ข้อมูล	ข้อมูลพื้นฐานที่จำเป็นในการบริหารจัดการของหน่วยงาน	ข้อมูลที่สามารถแลกเปลี่ยนและเชื่อมโยงในรูปแบบของข้อมูลขนาดใหญ่	ข้อมูลขนาดใหญ่ที่สามารถรองรับการหลอมรวม	ข้อมูลขนาดใหญ่ที่สามารถเข้าถึงได้ทุกอุปกรณ์ อย่างสมบูรณ์แบบ
	โครงสร้างพื้นฐาน	แหล่งเก็บ คั่นคืนข้อมูลพื้นฐานในรูปแบบดิจิทัล	Data Centre และ Data Lake	Data Analytic Centre	Big Data Center
	สังคม	เข้าถึงข้อมูลพื้นฐานอย่างทั่วถึงและเท่าเทียม	เข้าถึงข้อมูลขนาดใหญ่อย่างทั่วถึงและเท่าเทียม	สามารถนำข้อมูลขนาดใหญ่ไปใช้ประโยชน์และต่อยอดในกิจกรรมที่ต้องการได้อย่างทั่วถึงและเท่าเทียม	สนับสนุนและใช้ประโยชน์ได้ในทุกสาขา
	หน่วยงาน	เชื่อมโยงและบูรณาการในระดับกระทรวง	เชื่อมโยงและบูรณาการระหว่างภาครัฐเหมือนเป็นองค์กรเดียว	มีบริการที่ขับเคลื่อนโดยความต้องการและมีส่วนร่วมของทุกองค์กร	บริหารจัดการและบริการข้อมูลขนาดใหญ่ได้อย่างเป็นเลิศ
	ทุนมนุษย์	มีทักษะด้านข้อมูลขนาดใหญ่	สามารถทำงานผ่านระบบแลกเปลี่ยนและเชื่อมโยงในรูปแบบของข้อมูลขนาดใหญ่	หน่วยงานเกิดงานคุณค่าสูง และกำลังคนที่มีความเชี่ยวชาญข้อมูลขนาดใหญ่เพียงพอต่อความต้องการ	เป็นหนึ่งในศูนย์กลางด้านกำลังคนข้อมูลขนาดใหญ่

Thank
You