โครงการศึกษาเพิ่มเติมด้าน

BIG DATA GOVERNANCE AND BIG ANALYTIC

ณ LUDWIGSHAFEN UNIVERSITY OF APPLIED SCIENCES (HOCHSCHULE LUDWIGSHAFEN AM RHEIN)

รองศาสตราจารย์สำรวย กมลายุตต์

บิ๊กดาต้าคืออะไร

- Big Data หมายถึงความสามารถในการวิเคราะห์และตีความหมายที่ซ่อนอยู่ในข้อมูลที่มีปริมาณมหาศาล(Big Analytic)เหล่านั้น
- ข้อมูลปริมาณมากมายมหาศาลเหล่านี้มีรูปแบบที่หลากหลายทั้งที่มีโครงสร้าง ง่ายต่อการจัดเก็บในรูปแบบที่นำไปวิเคราะห์ใช้ประโยชน์และทั้งที่ไม่มีโครงสร้าง ซึ่งยากต่อการจัดเก็บและนำไปใช้ประโยชน์

ปัจจัยที่ทำให้เกิดบิ๊กดาตา

- การประมวลผลที่มีความสมบูรณ์มากขึ้น

  1) Moore’s Law

  2) Mobile Computing

  3) Social Networking

  4) Cloud Computing

ปัจจัยที่ทำให้เกิดบิ๊กดาตา (ต่อ)

- การเกิดของ 3V’s

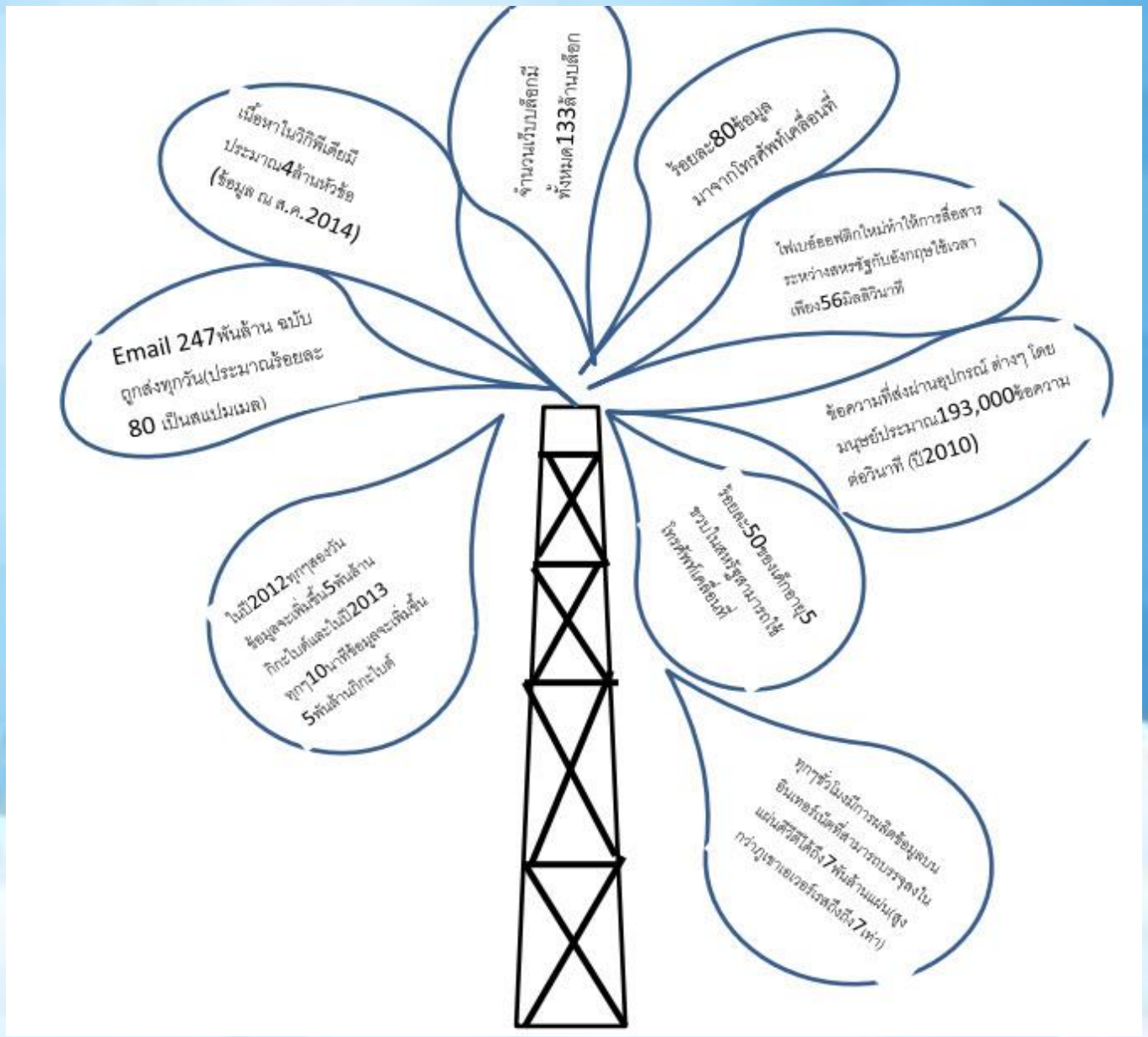
  1) Volume

  2) Velocity

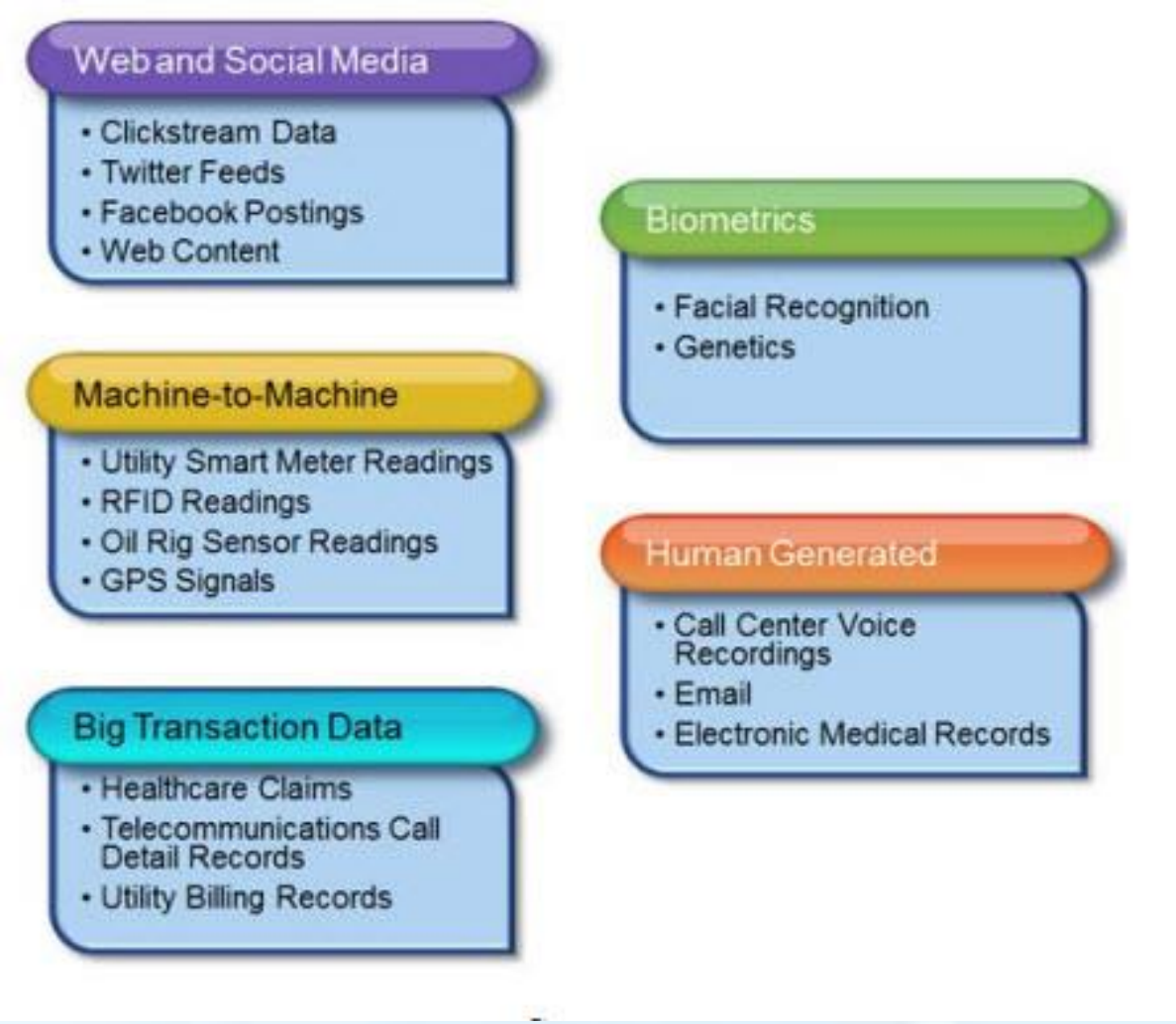
  3) Variety

- การผสมผสานของเทคโนโลยี

การเกิดขึ้นของข้อมูลจากแหล่ง/ช่องทางต่างๆ



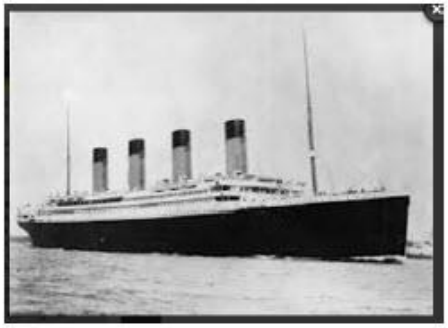
ประเภทของบิ๊กดาตา



Big Analytics

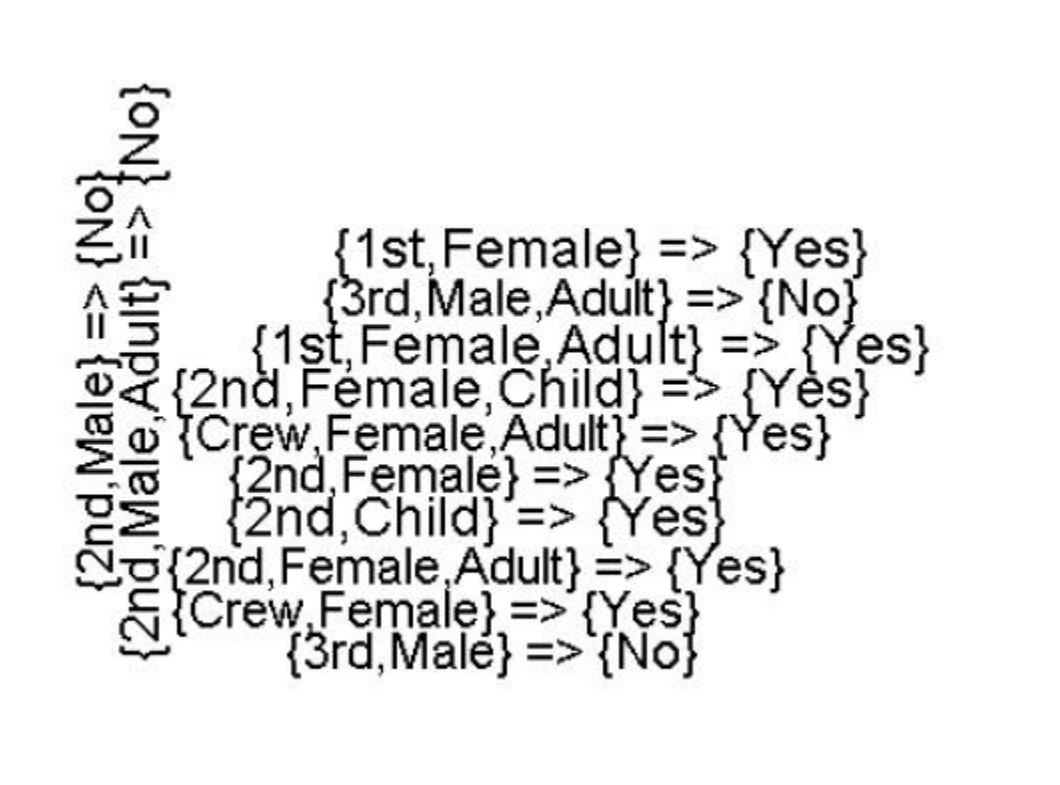
- การวิเคราะห์ Big Data ด้วยเทคโนโลยีที่พัฒนาขึ้นมาใหม่ เพื่อค้นหาแพทเทิร์นความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในข้อมูลปริมาณมหาศาลนั้น เช่น นำข้อมูลบนเฟซบุ๊กขององค์การมาวิเคราะห์ว่าลูกค้าหรือบุคคลภายนอกมององค์การอย่างไรทั้งในเชิงบวกและเชิงลบ

ตัวอย่างผลการทำนายการรอดชีวิตของผู้โดยสารบนเรือไททานิก

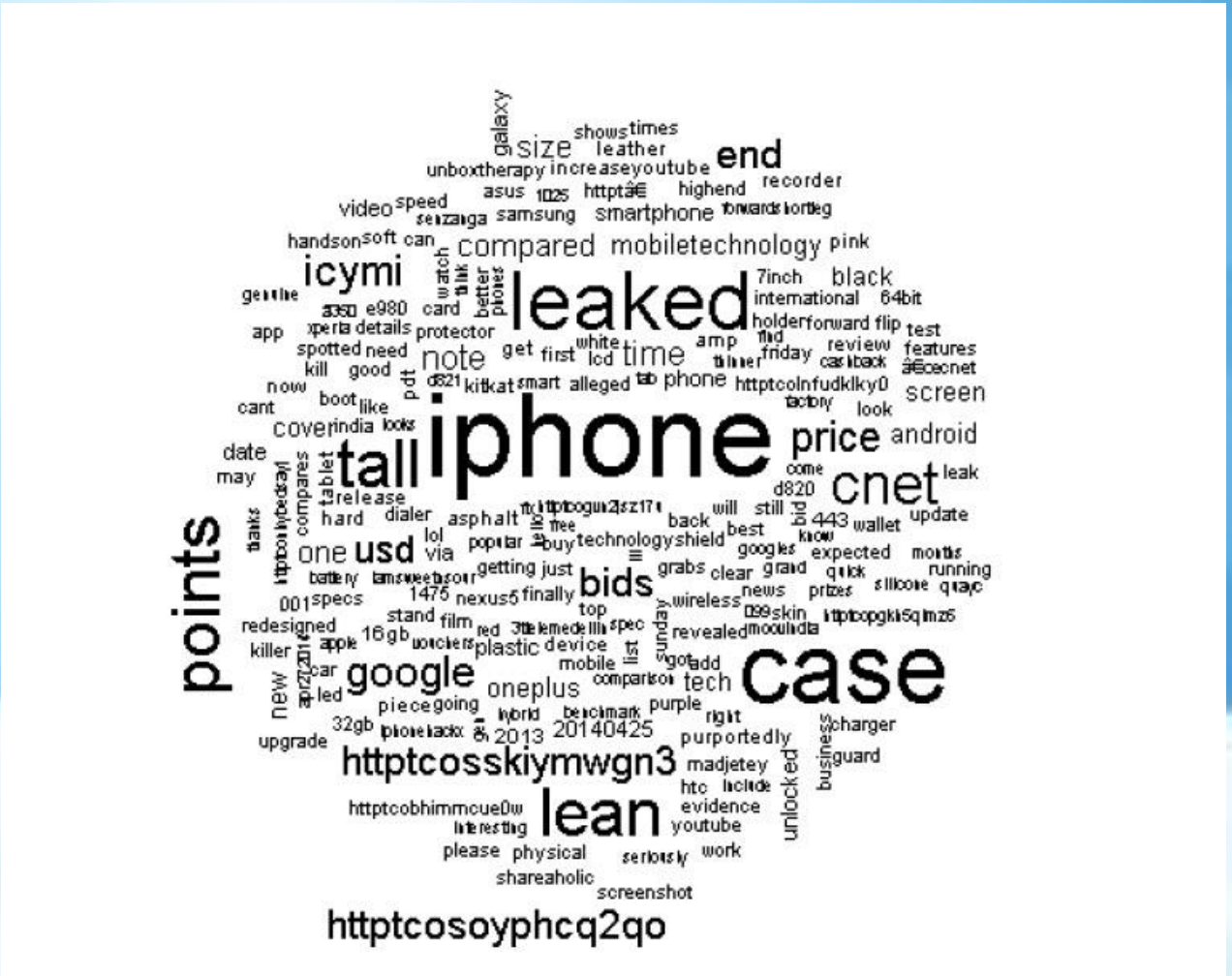


Apply a priori algorithm

ผลการทำนายการรอดชีวิตของผู้โดยสารบนเรือไททานิก



ตัวอย่างผลการวิเคราะห์ข้อมูลจากเฟซบุ๊ก



เทคโนโลยีที่พัฒนาสำหรับBig analytics

- ฮาร์ดแวร์ มีการพัฒนานวัตกรรมที่สนับสนุนการประมวลผลบิ๊กดาตาขึ้นมาหลายอย่าง เช่น
- แรมที่มีความจุสูง (High-capacity RAM)
- การประมวลผลแบบคู่ขนาน(Massively parallel processing --MPP)
- มัลติโพรเซสเซอร์สมมาตรที่มีขนาดใหญ่(Large symmetric multiprocessors --SMP)
- สถาปัตยกรรมการประมวลผลแบบมัลติคอร์ (Multi-core processor architectures)

เทคโนโลยีที่พัฒนาสำหรับBig analytics

- ซอฟต์แวร์ที่จะช่วยให้การประมวลผลมีความรวดเร็วมากขึ้น เช่น การประมวลผลแบบแบ่งส่วน (Partitioning) ไม่นำตารางข้อมูลในฐานข้อมูลมารวมกัน (No aggregate tables) การบีบอัดข้อมูล(Compression) การจัดเก็บข้อมูลทั้งแบบโรว์และแบบคอลัมน์(Row and Column storing)
- เหตุผลที่จัดเก็บข้อมูลในฐานข้อมูลแบบคอลัมน์เพราะมีข้อดีคือ มีการใช้แบนด์วิธการรับและการส่งข้อมูลที่ดีขึ้น หน่วยความจาแคชสามารถทางานได้อย่างมีประสิทธิภาพสูงกว่าเดิม การดึงข้อมูลจากหลายๆ ตารางรวดเร็ว

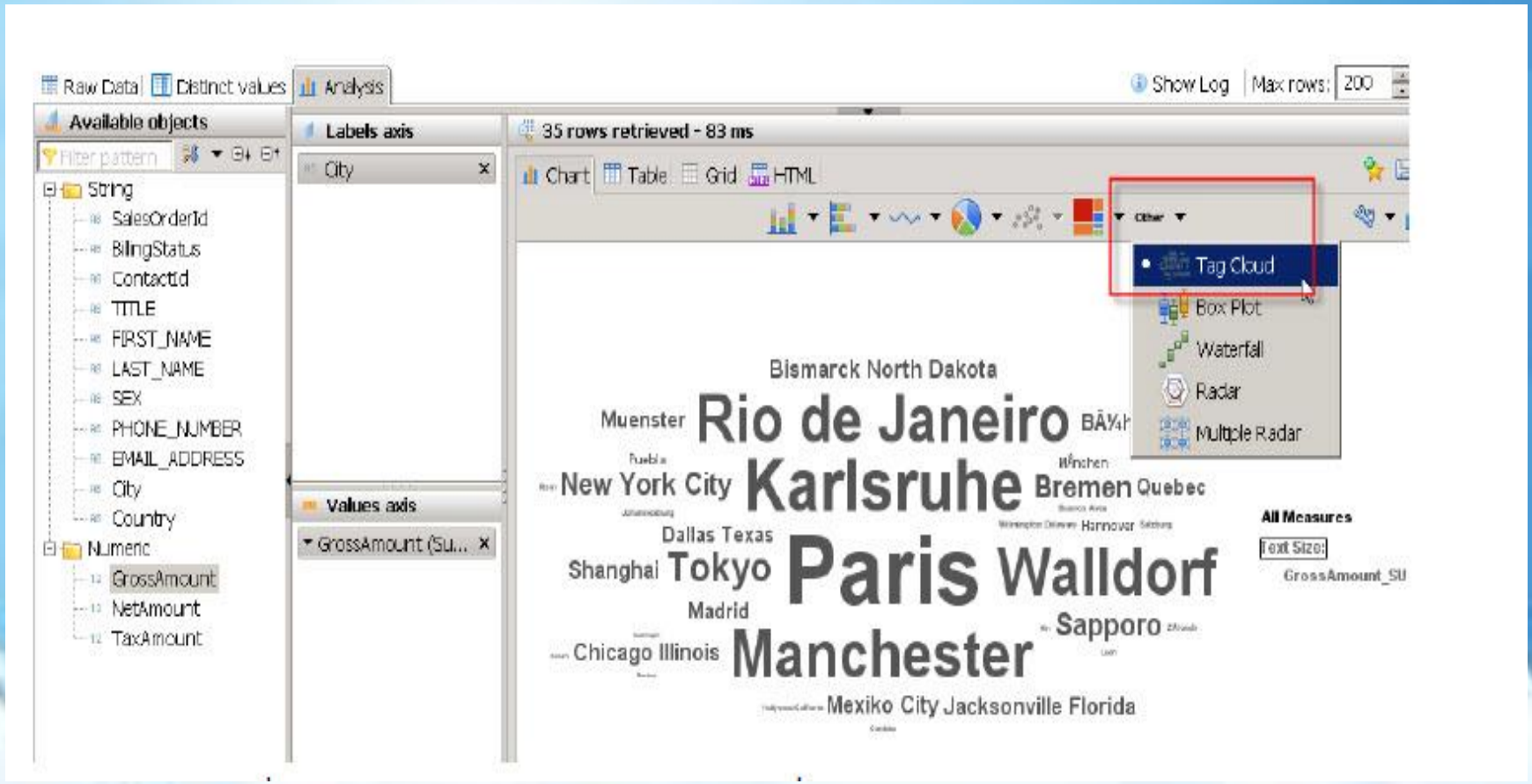
HADOOP

- ซอฟตร์แวร์โอเพนซอร์สที่พัฒนาขึ้นมาเพื่อทำหน้าที่เป็นที่จัดเก็บข้อมูลปริมาณมหาศาลที่ไม่มีโครงสร้างในลักษณะแบบกระจายกันจัดเก็บ และนำมาประมวลผลร่วมกันได้
- องค์ประกอบหลักๆของ Hadoop จะประกอบด้วย Hadoop Dustributed File System (HDFS) เป็นระบบงานที่ทาหน้าที่เป็นที่จัดเก็บข้อมูล (Storage) และแม็บ รีดิวซ์(MapReduce) เป็นซอฟต์แวร์ที่ใช้ในการพัฒนาโปรแกรมเพื่อประมวลผลข้อมูลที่จัดเก็บ ทั้งนี้โครงสร้างด้านฮาร์ดแวร์ของฮาดูป จะใช้เครื่องเซิร์ฟเวอร์ จำนวนมากต่อเป็นคลัสเตอร์เดียวกัน

SAP HANA

- สำหรับซอฟต์แวร์ที่พัฒนาโดยบริษัทSAPใช้ในการประมวลผลบิ๊กดาตาที่รู้จักกันดี คือ SAP HANA
- จุดเด่นของSAP HANA ใช้ฐานข้อมูลที่จัดเก็บข้อมูลในหน่วยความจำหลัก(In-Memory Database) หรือเรียกว่า Main Memory Database system(MMDB/IMDB)

ตัวอย่างการวิเคราะห์ยอดขายแต่ละเมืองโดยSAP HANA



การประยุกต์ Big Analytics

- *ด้านการแพทย์และสาธารณสุข*
- การรักษาผู้ป่วยโรคหัวใจเรื้อรังนั้น แพทย์ผู้รักษาต้องหมั่นตรวจสอบการเต้นของหัวใจโดยอ่านจากข้อมูลการตรวจEKG (Electrocardiogram) ที่พิมพ์ออกมาและหาข้อมูลที่มีความผิดปกติในกราฟที่พิมพ์ออกมานั้น ซึ่งข้อมูลกราฟที่พิมพ์ออกมาหากมีการพิมพ์อย่างต่อเนื่อง10ชั่วโมง ความยาวของกระดาษกราฟที่พิมพ์ออกมาจะยาวประมาณ 2 ไมล์ ซึ่งข้อมูลที่ต้องการทราบก็คือข้อมูลกราฟที่ผิดปกติเท่านั้น

- นักวิจัยสามคนจากสถาบันการศึกษาMIT (Massachusetts Institute of Technology)คือ John Guttag นักวิทยาการคอมพิวเตอร์(comport scientist)และนักหัวใจวิทยา(cardiologist) Colln Stultz นักชีววิทยาจากสถาบันเดียวกัน และZeeshan Syed นักวิทยาการคอมพิวเตอร์จากมหาวิทยาลัยมิชิแกน
- ร่วมกันพัฒนาแบบจำลองคอมพิวเตอร์เพื่อวิเคราะห์ข้อมูลEKGที่มีปริมาณมากมายดังกล่าว โดยใช้เทคนิคการทำเหมืองข้อมูล(data mining)และการเรียนรู้ของเครื่องจักร(machine learning)ในการวิเคราะห์กลั่นกรองข้อมูลที่ไม่ได้แสดงความผิดปกติหรือข้อมูลที่ไม่มีผลต่อการเกิดหัวใจวายออกไป

การประยุกต์ Big Analytics

- *ด้านธุรกิจ*
- ข้อมูลที่บริษัทFedEXต้องจัดการในแต่ละวันมีปริมาณ2.2พันล้านรายการ (transaction)
- ปัจจุบันบริษัทได้พัฒนาระบบต้นแบบที่สามารถติดตามการส่งพัสดุภัณฑ์แบบเรียลไทม์ด้วยการสัมผัสหน้าจอคอมพิวเตอร์แทนการป้อนข้อมูล
- อนาคตพัฒนาระบบที่ชาญฉลาดกว่านั้นโดยนำข้อมูลมาวิเคราะห์ให้ได้ว่าขณะนั้นผู้รับพัสดุภัณฑ์กำลังอยู่ที่ไหน ที่บ้าน ที่ทางาน หรือสถานที่อื่นๆ บริษัทก็จะนำส่งพัสดุภัณฑ์ไปยังสถานที่ที่ผู้รับอยู่ทั้งนี้เพื่อให้พัสดุภัณฑ์ส่งถึงมือผู้รับโดยเร็วที่สุด

การแลกเปลี่ยนเรียนรู้

- ซักถามประเด็น ข้อสงสัย
- เสนอความคิดเห็น แลกเปลี่ยนเรียนรู้ร่วมกัน


thank you
tenki
takk
спасибо
kam sah hamnida
dank je
hvala
dziękuję
bedankt
dhanyavadagalu
дзякуй
gracias
xièxie
mèsi
blagodaram
tanemirt
rahmet
mochchakkeram
trugarez
arigatô
danke
kop khun krap
faafetai lava
ありがとう
dëkuji
shukriya
grácie
spas
ačiū
diolch
barka
शुक्रिया
gràcies
tau
teşekkür ederim
obrigada
mahalo
sukriya
obrigado
taiku
sagolun
choukrane
rahmat
dakujem
terima kasih
najis tuke
mercé
謝謝
köszönöm
mersi
asante
grazie
nanni
mauruuru
merci
vinaka
go raibh maith agat
paldies
ngiyabonga