

พื้นฐาน
เทคโนโลยีสารสนเทศ และ ระบบสารสนเทศ
*Introduction to
Information Technology and Information Systems*

ดร.ยรรยง เต็งอำนวย
คณบดีคณะวิศวกรรมศาสตร์
ศูนย์เชี่ยวชาญด้านวิศวกรรมซอฟต์แวร์
ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย
2553 ภาค 6 สถาปัตยกรรมข้อมูล
ลงทะเบียน 20122012

10/11/53 อ.ดร.ยรรยง Intro IT / IS Module # 6-Data Architecture 1

ความสำคัญของข้อมูล

- Computer → กล้องที่นำสนใจ
- Information Technology → กล้องหนึ่งที่ใช้ประโยชน์ได้
- Information System → ระบบงานที่ธุรกิจต้องการ

• ข้อมูล ???

- ผิด ขาด ซ้ำ - นั่นคือ Garbage in = Garbage out
- มาก หลากหลาย หมายถึง รวบรวม เก็บ แยกแยะ ค้นคืน

10/11/53 อ.ดร.ยรรยง Intro IT / IS Module # 6-Data Architecture 2

ประเด็นเกี่ยวกับข้อมูล

- สถาปัตยกรรมข้อมูล
 - ธรรมชาติของข้อมูล
 - นิยามของข้อมูล แหล่งข้อมูล ความถูกต้องของข้อมูล การเก็บข้อมูล การนำข้อมูลมาใช้งาน
- มาตรฐานข้อมูล
 - Data Model
 - Data Dictionary
- ความมั่นคงปลอดภัยของข้อมูล
 - Physical Security
 - Logical Security
- Centralized & Distributed Data

10/11/53 อ.ดร.ยรรยง Intro IT / IS Module # 6-Data Architecture 3

ข้อมูลจำนวนมาก

- Enterprise Data Architecture
- Enterprise → องค์กรขนาดใหญ่
 - ร้านโชห่วย ไม่ใช่องค์กรใหญ่
 - องค์กรแบ่งเป็นส่วนงานย่อย ทำงานเฉพาะด้าน
- Data → ข้อมูล (ดิบ) และสารสนเทศ (สุก)
- Architecture → สถาปัตยกรรม โครงสร้างในภาพรวม
 - ดีกอาการ = สถาปัตยกรรม + วิศวกรรม

10/11/10 6.01.21100 Intro IT / IS Module # 6-Data Architecture 4

ข้อมูลขนาดใหญ่ของโลก

- Google
 - กช. 2550 ประมาณว่า 20 petabyte ต่อวัน
 - โฆษณา / การค้นหา
 - 20 PB = 20,000 TB = 20 ล้าน GB
 - <http://techcrunch.com/2008/01/09/google-processing-20000-terabytes-a-day-and-growing/>
- Youtube
 - กช. 2552 ประมาณว่ามีวิดีโอ 100 ล้านเรื่อง
 - 30 MB ต่อเรื่อง = 3 PB = 3 ล้าน GB
 - http://www.associatedcontent.com/article/1927414/how_many_videos_are_on_youtube.html

10/11/10 6.01.21100 Intro IT / IS Module # 6-Data Architecture 5

สถิติต่างๆ ของ Youtube

- จำนวนวิดีโอ มากกว่า 120 ล้านเรื่อง
- Upload 200,000 เรื่องต่อวัน
- จำนวนวิดีโอที่ชมต่อวัน 200 ล้านเรื่อง
- ปริมาณวิดีโอที่ upload ต่อนาทีคือ 13 ชม.
- จำนวนผู้ชมที่สมัครสมาชิก มากกว่า 300 ล้านคน
- ประเภทของวิดีโอ -- เพลง 20% มันทึ่ง 15% ส่วนตัว 14% เบาสมอง 13% กีฬา 7% ความรู้ 6% รถยนต์ 5% ภาพยนตร์ 5%

ที่มา http://wiki.answers.com/Q/How_many_videos_are_there_on_YouTube

10/11/10 6.01.21100 Intro IT / IS Module # 6-Data Architecture 6

สถาปัตยกรรมข้อมูล

- ธรรมชาติของข้อมูล
- ลักษณะของข้อมูล
- ข้อมูลเป็นโครงสร้างพื้นฐานแบบหนึ่ง
- คุณสมบัติของข้อมูลที่ดี
- องค์ประกอบของระบบข้อมูล

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 7

ธรรมชาติของข้อมูล

- การออกแบบระบบข้อมูลต้องเข้าใจ “ธรรมชาติ” ของข้อมูลด้วย
 - ข้อมูลคืออะไร
 - ข้อมูลมาจากไหน
 - รูปแบบของข้อมูล
 - เก็บข้อมูลไว้ที่ไหน
 - เรียกข้อมูลมาใช้ได้ไหม
 - ต้องการใช้อะไรเร็วแค่ไหน
 - ใครใช้ข้อมูลบ้าง
 - เสียหายแล้วกู้คืนได้ไหม

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 8

ข้อมูลคืออะไร

- ข้อมูล (data) มีนิยามหลากหลาย แต่สรุปได้คือ
 - **ข้อมูลคือข้อเท็จจริง (Fact)**
- ข้อมูลคืออะไรก็ตามที่อยู่ในรูปแบบที่ใช้กับคอมพิวเตอร์ได้
 - ต่างกับโปรแกรมที่เป็น “คำสั่ง” ให้คอมพิวเตอร์
 - ทุกอย่างที่ไม่ใช่โปรแกรม
 - ที่มา http://en.wikipedia.org/wiki/Data_%28computing%29
- Data / Information / Knowledge
 - อยู่ที่ “มุมมอง” (abstraction)
 - เทือกเขา Everest → ความสูง / สภาพเส้นทาง / การขึ้นสู่ยอด

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 9

ข้อมูลมาจากไหน
Data Source & Origin

- แหล่งข้อมูล
 - ภายนอก – ข้อมูลจากภายนอกระบบ
 - ภายใน – ข้อมูลถูกกำหนดไว้ในโปรแกรม เช่น 1 วันคือ 24 ชม.
- ทั่วไปมาจากเป็นพิมพ์ แต่ปัจจุบันมาได้จากหลายแหล่ง เช่น barcode และ sensor กล้อง
- ข้อมูลเป็นของ "ผู้ให้" และ "ผู้ใช้" ข้อมูล
- "ไปช่วยฝ่ายบัญชี เพราะข้อมูลไม่ถูกต้อง" ไม่อาจทำได้ นั่นคือ **ข้อมูลต้องมาจากแหล่งของมัน**

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 10

รูปแบบของข้อมูล
Data Format

- แรกก่อนเป็นตัวเลข ตัวอักษร เดียวกันเป็นสื่อประสม เสียง เพลง หนังสือ
- แรกก่อนเสียงเป็น analog เดียวกันเป็น digital
- Analog → ปromptอ่านอุณหภูมิจึงเป็นเส้น / เป็นตัวเลข
- ตัวเลข จำนวนเต็ม หรือมีจุดทศนิยม
- ระวังหน่วย ยานสำรวจดาวอังคารตกเพราะผสมฟุตกับเมตร ทำให้เกิดเป็นมาตรฐาน XML
- ตรวจสอบ วอลเปก เพราะหลากหลาย format

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 11

เก็บข้อมูลไว้ที่ไหน
การเก็บข้อมูลจำนวนมาก-ระยะเวลาานาน

- ปริมาณ
 - ข้อมูลมีแต่เพิ่มพูน แม้แต่ข้อมูลส่วนบุคคล
 - ผู้สอนมีรูปถ่ายประมาณ 30 GB
 - ไม่กล้าลบวีดีโอ เพราะไม่มีที่เก็บ
- Backup
 - ข้อมูลต้องมีการเสียบหาย hard disk พัง
 - ปกติต้องหยุดระบบ เดียวกันระบบใช้งาน 24/7
- Archive
 - กฎหมายให้เก็บ 10 ปี

10/11/10 6.01.21104 เอกสารที่มีความสำคัญทางประวัติศาสตร์(และสิ่งสำคัญ?) Architecture 12

ลักษณะการเก็บข้อมูล

- เหมาะสมต่อการใช้งาน
- เพิ่มธรรมชาติ
 - ไม่มีโครงสร้างพิเศษ ใช้กับข้อมูลจำนวนน้อย หรือเป็นชิ้นๆ เช่นรูปภาพ
- เพิ่มดัชนี มี index ช่วยการสืบค้น
- ฐานข้อมูล (database)
 - ข้อมูลจำนวนมาก หลากหลาย
 - ความสัมพันธ์ซับซ้อน
 - การใช้งานหลากหลาย

101153 6.10.21101 Intro IT / IS Module # 6-Data Architecture 13

กรณีศึกษา - การเก็บข้อมูลระยะยาว

- ต้องการวิเคราะห์ประวัติการแพร่กระจายของโรคนั้นๆ ในไทย ย้อนกลับไปที่สิบปี
- ต้องการวิเคราะห์พัฒนาการของหลักสูตรต่างๆ นับแต่จัดตั้งมหาวิทยาลัย
- ต้องการวิเคราะห์การย้ายถิ่นของประชากรไทยย้อนหลังไปยี่สิบปี เพื่อศึกษาการเติบโตของเมืองใหญ่
- ต้องการข้อมูลการเงินย้อนหลังห้าปีเพื่อใช้ในทางศาล

101153 6.10.21101 Intro IT / IS Module # 6-Data Architecture 14

เรียกข้อมูลมาใช้ได้ไหม

- ดูหนังสือสอบ แต่ทำสอบจำไม่ได้ มาจำได้ก่อนออกจากห้องสอบ
- ปัญหาเรื่องการเรียกใช้ข้อมูล = เตรียมข้อมูลเพื่อการเรียกใช้ = ภารกิจเด็กขายบนปากอ่าง (ถอดคณูปอ)
 - ใช้ได้ / เร็ว / ครบ / แม่นยำ
- ศาสตร์แห่งการสืบค้นสารสนเทศ
- Information Retrieval

101153 6.10.21101 Intro IT / IS Module # 6-Data Architecture 15

ต้องการใช้ข้อมูลเร็วเพียงใด
ความทันสมัยของข้อมูล

- **Batch** สมัยก่อนเทคโนโลยียังล่าสมัย
 - ป้อนข้อมูลเป็นชุด (batch) เช่นรายวัน รายเดือน แล้วป้อนให้คอมพิวเตอร์ประมวลผล
 - ได้รายงานเป็นรายวัน รายเดือน หรือรายปี
 - โบนัสประจำปีใช้โทรศัพท์ ดอกเบี้ยเงินฝากทุก 6 เดือน
- **Interactive**
 - ป้อนข้อมูลตรงเข้าระบบจากหน้าจอ ได้ผลลัพธ์ทันที
 - ถอนเงินจากตู้ ATM
- **Real-time** ข้อมูลแบบ "ทันกาล"

1011-53 ข้อมูลเบื้องต้นเกี่ยวกับ IS Module # 6-Data Architecture 16

ใครใช้ข้อมูลได้บ้าง
สิทธิ์การใช้งานและความมั่นคงปลอดภัยของข้อมูล

- สิทธิ์การใช้งาน (usage authorization)
 - ให้อาจอู่ในระบบเพื่อให้งานเรียบร้อย
- ความมั่นคงปลอดภัย (security)
 - แก้ไขเงินในบัญชีธนาคาร ตรวจสอบพบไหม
- ความเป็นส่วนตัว (privacy)
 - Spam mail ได้อีเมลของเราได้อย่างไร

• **ต้องเตรียม "ระบบข้อมูล" เพื่อไว้ก่อน**

1011-53 6.11.2010 Intro IT / IS Module # 6-Data Architecture 17

ข้อมูลเสียหายสามารถกู้คืนได้ไหม

- สมมติว่าคืนมาตอนเช้าเข้าที่ทำงานไม่ได้ มีออฟดาวน์รวมทั้งห้องเซิร์ฟเวอร์ด้วย
 - เซิร์ฟเวอร์ ฮาร์ดแวร์ เครือข่าย มีเงินซื้อทดแทนได้
 - ข้อมูลเสียหาย **ทดแทนยากมาก**

ความต่อเนื่องของธุรกิจ
Business Continuity

1011-53 6.11.2010 Intro IT / IS Module # 6-Data Architecture 18

ลักษณะของข้อมูล

Garbage in → garbage out (ป้อนขยะ ได้ขยะ)

- ตัวอย่าง ข้อมูล Admission สำหรับเข้ามหาวิทยาลัย ถ้าข้อมูลผิดพลาดถึงเข้ามหาวิทยาลัยไม่ได้
- ข้อมูล Admission ผิดได้ที่
 - เลขรหัส 13 หลักประจำตัวประชาชน
 - ชื่อ นามสกุล
 - โรงเรียน
 - คะแนนสอบ
 - ...

101153 ภาควิชาวิศวกรรมคอมพิวเตอร์
Architecture 19

โครงสร้างพื้นฐาน

- ข้อมูลถือว่าเป็น **โครงสร้างพื้นฐาน (infrastructure)** ได้แบบหนึ่ง
 - เปรียบเสมือน Server / Network / Data Center
- Google
 - ผู้คนป้อนข้อมูลเข้าไปในเครือข่าย ในเว็บ อย่างมหาศาล
 - Google ไปค้นและทำดัชนี (crawl & index)
 - ผู้คนเข้าไปสืบค้น (search) ข้อมูลจาก Google ได้

101153 ภาควิชาวิศวกรรมคอมพิวเตอร์
Intro IT / IS Module # 6-Data
Architecture 20

คุณสมบัติของข้อมูลที่ดี

ถูกต้อง / ครบถ้วน / รวดเร็ว

- **ถูกต้อง (accuracy / consistency)**
 - ข้อมูลตรงกับความเป็นจริง กลุ่มข้อมูลที่เกี่ยวข้องกันต้องสอดคล้องกัน หากมีหลายส่วนก็ต้องตรงกัน
- **ครบถ้วน (completeness)**
 - ข้อมูลต้องมีองค์ประกอบครบถ้วนตามความต้องการขององค์กร ไม่ใช่ตามความต้องการของระบบงาน
- **รวดเร็ว (in time / up to date / real-time)**
 - เมื่อต้องการใช้ต้องมีข้อมูลที่ต้องการ ข้อมูลควรทันต่อเหตุการณ์ หรืออาจเป็นข้อมูลที่ตรงกับเหตุการณ์ที่เกิดขึ้นจริง

101153 ภาควิชาวิศวกรรมคอมพิวเตอร์
Intro IT / IS Module # 6-Data
Architecture 21

ข้อมูลต้องถูกต้อง
Accuracy / Consistency

- ข้อมูลตรงกับความเป็นจริง
 - ลูกค้าระบุอายุในใบสมัครบัญชีเงินฝาก อายุไม่เปลี่ยนตามปีที่ผ่านไป ข้อมูลไม่ตรงกับความเป็นจริง
- กลุ่มข้อมูลที่เกี่ยวข้องกันต้องสอดคล้องกัน
 - อายุในใบสมัครไม่สอดคล้องกับวันเกิดของลูกค้าเมื่อเวลาผ่านไป
- หากมีหลายสำเนา ต้องตรงกัน
 - ที่อยู่ของลูกค้าในบัญชีฝากประจำไม่ตรงกับบัญชีเดินสะพัด

09/11/03 6.01.0100 Intro IT / IS Module # 6-Data Architecture 22

กรณีศึกษา - ความถูกต้องของข้อมูล

- นักเรียนสอบเข้า Admission แต่ละปีกรอกชื่อ-นามสกุลผิดพลาดแสนคน
- กระจายค่าตอบสองด้านใบจะมีนักเรียนฝนเลขประจำตัวสอบคิดประมาณ 5,000-10,000 ใบ ทุกปี
- โพล (poll) สามารถผิดพลาดได้เพราะเป็นการสำรวจ
- ข้อมูล Admission ผิดพลาดไม่ได้

ความผิดพลาด = นักเรียนเสียอนาคต

• ปัญหาใหญ่ 2 ประการ

09/11/03 6.01.0100 Intro IT / IS Module # 6-Data Architecture 23

ข้อมูลต้องครบถ้วน
Completeness

- ข้อมูลต้องมื่อองค์ประกอบครบถ้วนตามความต้องการขององค์กร ไม่ใช่ตามความต้องการของระบบงาน
 - ระบบกักกันโรค ต้องการข้อมูลการแพทย์ แต่ไม่ต้องการเลขสวัสดิการสังคมของผู้ป่วย
 - ระบบสวัสดิการผู้ป่วย ต้องการเลขสวัสดิการสังคมของผู้ป่วย แต่ไม่ต้องการข้อมูลการแพทย์

หากไม่มีสถาปัตยกรรม แต่ละระบบจะเก็บข้อมูลเองและไม่ครบถ้วน หากต้องการข้อมูลเพิ่มเติมจะทำไม่ได้ และจะเกิดข้อมูลไม่สอดคล้องกันคือข้อมูลที่ซ้ำซ้อน

ไม่จำเป็นต้องเก็บข้อมูลทุกอย่างของคนไข้
แต่ต้องเก็บให้เพียงพอต่อการใช้งานขององค์กร

09/11/03 6.01.0100 Intro IT / IS Module # 6-Data Architecture 24

กรณีศึกษา - ความครบถ้วน

- ต้องการหาความสูงเฉลี่ยของนักเรียนในชั้น - เก็บข้อมูลความสูงของนักเรียนแต่ละคน
- ต้องการหาความสูงเฉลี่ยแยกตามเพศ - เก็บเพศเพิ่ม
- ต้องการหาความสูงเฉลี่ยตามภูมิภาค - เก็บที่อยู่เพิ่ม

- ถ้าข้อมูลมีอยู่แล้ว สมองความต้องการได้เร็ว

101153 6.01.21104 Intro IT / IS Module # 6-Data Architecture 25

กรณีศึกษา - ความครบถ้วน

- ผู้ป่วยเปลี่ยนชื่อ-นามสกุล ไม่อาจตรวจหาประวัติได้

- การเปลี่ยนแปลงข้อมูลดิจิทัล *ไม่ทิ้งร่องรอย*

- ต้องสร้างร่องรอยขึ้นเอง

101153 6.01.21104 Intro IT / IS Module # 6-Data Architecture 26

ข้อมูลต้องรวดเร็ว
In time / up to date / real time

- ข้อมูลที่ต้องการต้องมีให้ใช้เมื่อต้องใช้ข้อมูลนั้น
 - ระบบโอนเงินเรียกเก็บเงินจากข้อมูลบัตรเครดิต
- ข้อมูลที่มีให้ใช้ต้องทันสมัย ไม่เป็นข้อมูลเก่า
 - ยอดเงินในบัญชีที่ถอนจาก ATM
- อาจต้องเป็นข้อมูลแบบทันที (real time)
 - ข้อมูลจากลูกออกบัตรระดับสารเคมีในถัง เพื่อคุมปริมาณสารเคมีเข้าถึง
 - ข้อมูลจากตัวควบคุมการจุดระเบิดในกระบอกสูบเครื่องยนต์

101153 6.01.21104 Intro IT / IS Module # 6-Data Architecture 27

องค์ประกอบสถาปัตยกรรมข้อมูลองค์กร

- มาตรฐาน (Standardization)
- การผนวกรวม (Integration)
- การใช้ร่วมกัน (Sharing)

- **ต้องออกแบบระบบข้อมูลให้ครอบคลุมสามประเด็นนี้**
 - ไม่เช่นนั้นจะเกิดปัญหายุ่งยากตามมา

001153 6.01.2553 Intro IT / IS Module # 6-Data Architecture 28

มาตรฐานข้อมูล
Data Standardization

- สมัยก่อนรหัสนิติศตป.ศร. ต่างจากป.โท ใช้ร่วมกันไม่ได้
- นิติศตป.ศร.เข้าต่อโท กลายเป็นคนละคน บางครั้งชื่อไม่ตรง
- ระบบอาจแยกกันแต่ข้อมูลต้องใช้มาตรฐานเดียวกัน จึงถืออำนาจการใช้งานร่วมกัน การแบ่งปันข้อมูล
- ปี 2517 จุฬาฯ ใช้เวลา 2 ปีเพื่อทำมาตรฐานข้อมูลห้าด้าน นิสิต / บุคลากร / ราชวิชา / การเงิน / อาคารสถานที่ ให้ไอทีจุฬาฯ พูดภาษาเดียวกัน ไม่มีใครใช้!!! ตอนนี้อยากใช้ ทำ GDx (Government Data Exchange) ซึ่งไม่สำเร็จ

001153 6.01.2553 Intro IT / IS Module # 6-Data Architecture 29

การผนวกรวมข้อมูล
Data Integration

- ฝ่ายต่างๆ ทำงานเป็นเอกเทศ ต่างเก็บข้อมูล "ของตน" คือของส่วนงาน
- ข้อมูลบางส่วนจะซ้ำกัน เช่น ชื่อพนักงาน ในฝ่ายบุคคล เพื่อคิดเงินเดือน กับฝ่ายผลิต เพื่อคิดเวลาทำงานและต้นทุนการผลิต
- พนักงานเปลี่ยนชื่อ/สกุล ต้องตามเปลี่ยน - ฝ่ายบุคคล ฝ่ายผลิต ฝ่ายสวัสดิการพนักงาน กรมการปกครอง กรมแรงงาน ธนาคาร กรมการขนส่ง บัตรเครดิต 3 ใบ สหกรณ์
- นิติศตป.ศร.เข้าต่อโท กลายเป็นคนละคน

001153 6.01.2553 Intro IT / IS Module # 6-Data Architecture 30

การใช้ข้อมูลร่วมกัน
Data Sharing

- ข้อมูลในหน่วยงานย่อยเป็นข้อมูลใช้ร่วมกันสำหรับหน่วยงานต่างๆ ขององค์กร
- นิสิตจบตรีเข้าต่อโท กลายเป็นคนละคน
- ไม่อาจทำการเรียนข้ามคณะได้ แก้ไขโดยรวมศูนย์เป็นทะเบียนกลาง
- ทะเบียน / บุคลากร / การเงิน ใช้ข้อมูลร่วมกันไม่ได้ แก้ไขโดย SAP ราคาหลายสิบล้าน (ลดแล้ว)

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 31

ขอบคุณครับ

10/11/10 6.01.21104 Intro IT / IS Module # 6-Data Architecture 32
