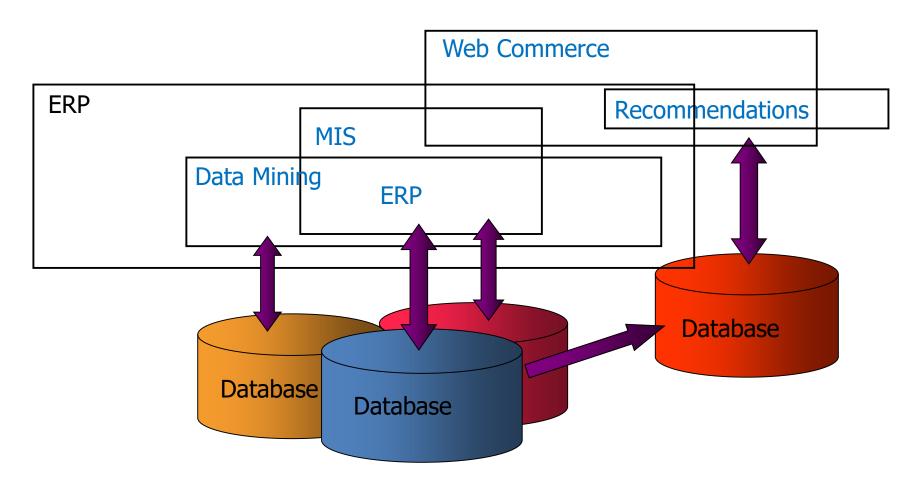
데이터베이스 개론 Chap. 00

이상구 교수



Computing = Data Processing

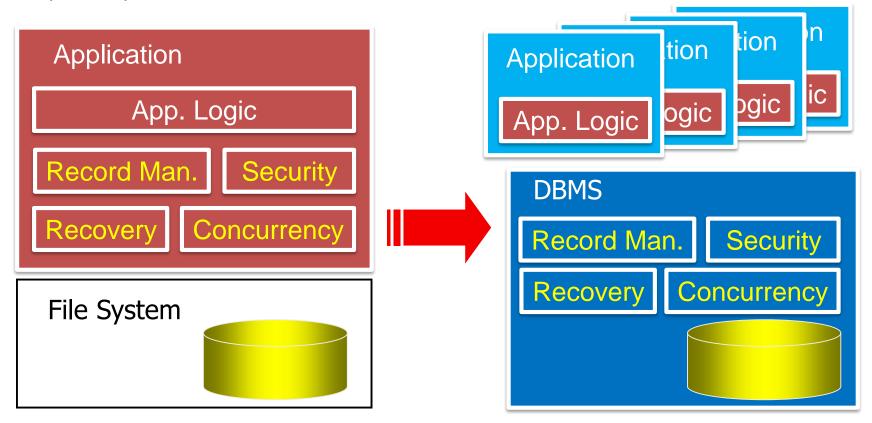
 Most (all?) computing applications use some type of a database





Ground 0: File System

- File System
 - Core part of OS
 - Stores programs, data, documents, or anything
 - (in disk)





Evidence Based Decision Making

Insights(통찰력) & foresights(예지력) through data

"It is a capital mistake to theorize before one has data.

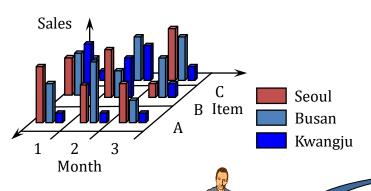
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts."



- The Adventures of Sherlock Holmes, A. Conan Doyle

But ...

A multidimensional/analytic view



Executives



A tabular/transactional view

ID	name	dept_nar	ne s	salary					d	ept_nan	e l	nuilding	budge	et
22222	Einsteir	n Physics		95000					В	iology	1	Watson	9000	00
12121	Wu	u Finance		90000					C	Comp. Sci.		Taylor	10000	00
32343	El Said	History		60000						lec. Eng		Taylor	8500	00
45565	Katz	Comp. 5	ici.	75000						nance		Painter	12000	
98345	course_id	title			dept_nam	e	credits			istory		Painter	5000	
76766	BIO-101			\rightarrow	Biology	\top	4			lusic		Packard		
10101	BIO-301	Genetics			Biology		course_id	500		semester	war		room_number	
58583	BIO-399	Computational	Biology		Biology		BIO-101	1		Summer	2009	Painter	514	В
83821	CS-101	Intro. to Compt	iter Scier		Comp. Sc		BIO-301	1	- 1	Summer	2010	Painter	514	A
	CS-190	Game Design			Comp. Sc		CS-101 CS-101	1 1	1	Fall Spring	2009	Packard Packard	101	H
15151	CS-315	Robotics			Comp. Sc	i.	CS-190	Ιŝ	- 1	Spring	2009	Taylor	3128	E
33456	CS-319	Image Processi	ID	cours	e id so	_id	semester		uear	pring	2009	Taylor	3128	A
76543	CS-347	Database Syste	_					_		pring	2010	Watson	120	D
	EE-181	Intro. to Digita		CS-10		1	Fall		2009		2010	Watson Taylor	100 3128	B C
	FIN-201	Investment Bar		CS-3		1	Spring		2010		2009	Taylor	3128	A
	HIS-351	World History	10101	CS-3		1	Fall		2009	pring	2009	Taylor	3128	C
	MU-199	Music Video Pi	12121	FIN-	201	1	Spring	1 2	2010	pring	2010	Packard	101	В
co	urse id	prerea_id ci	15151	MU-	199	1	Spring	- 2	2010	pring	2010	Painter	514	C
			22222	PHY	-101	1	Fall	- 13	2009	pring	2010	Packard Watson	101	D
	10-301	BIO-101	32343	HIS-	351	1	Spring		2010		2009	watson	100	_ A
E	IO-399	BIO-101	45565	CS-16	21	i	Spring		2010					
- 0	CS-190	CS-101	45565	CS-3		1	Spring		2010					
	S-315	CS-101	76766	BIO-		î	Summe		2009					
	S-319	CS-101	76766	BIO-		1	Summe		2010					
			83821	CS-19		1	Spring		2009					
	S-347	CS-101	83821	CS-19		2	Spring		2009					
	E-181	PHY-101	83821	CS-3		2	Spring		2010					
			98345	EE-12		1	Spring		2009					



IS team



Tipping Point 1: Data Base – Relational

- Simple and intuitive representation
- Powerful language (SQL)
- Performance through automatic query optimization
- Robust transaction support

Wide	column stores 2.1%	/ Document st	ores 5%
arch engi	nes 2.8%	Grap	oh DBMS 0.5% Key-value store
			Native XML D RDF stores
ery	V	/	
Rela	ational DBMS 86.3%		

© 2014, DB-Engines.com

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

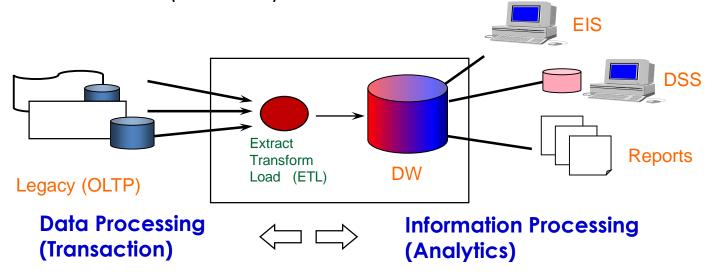
dept_name	building	budget
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

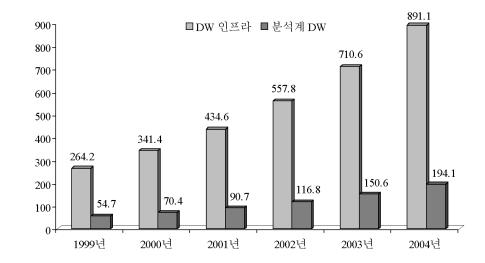


Tipping Point 2: Data Warehouse

- 업무시스템(transaction system)으로부터 쌓이는 데이터를 한 곳에 모아
- 분석적 작업에 활용 (GB -> TB)



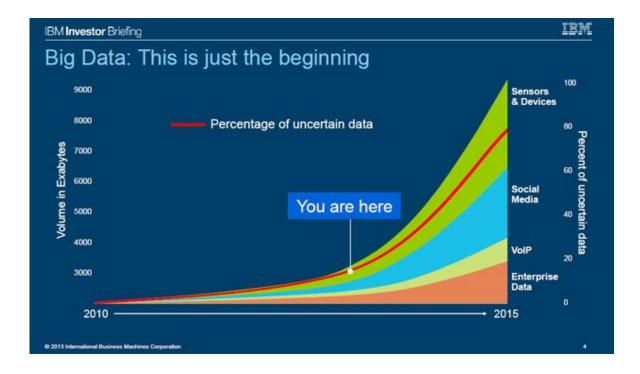
=> 데이터 분석의 전성기





Data Explosion

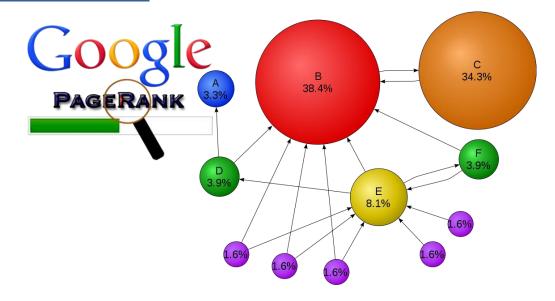
- DW 는 빙산의 일각 Enterprise data growth 도 따라가지 못함
- 1.5년마다 2배로 증가!
- Data growth 요인
 - "SW is eating the world" 모든 곳의 전산화/정보화
 - Mobile & social networks
 - Sensors & smart devices



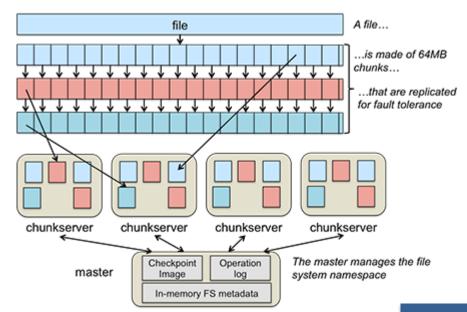


Web Scale Computing

- Different data,
- Different operations,
- Different scale!



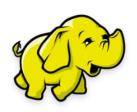






Big Data Systems

Hadoop



- Apache Open Source SW
- MapReduce 기반의 대량 데이터 분산처리 framework
- Yahoo!에서 시작/지원 (2006)

NoSQL



- Not Only (?) SQL
- 단순한 데이터 모델: Key-Value store
- 단순한 질의: get(), put()
- 단순한 트랜잭션 모델: BASE Basically Available, Soft state, Eventual consistency

Tipping Point 3: Big Data

- Ubiquitous 모든 분야에 일어나고 있는 현상
 - 생산, 유통, 의료, 공공, 문화, 언론, 역사, ...
 - 정보화/자동화, 모바일, 소셜, 센서!!
 - Impact 있는 사례
- Feasible 효과적으로 대응할 수 있는 환경
 - 풍부한 데이터
 - 강력한 컴퓨팅 자원
 - 효과적인 분석 기술
- Virtuous Cycle 데이터 기반 해결책의 가치 인정
 - 분위기 전환 more and more success stories
 - Data가 핵심 자산이라는 인식 확산
 - 연계/통합/융합으로 새로운 기회 발굴

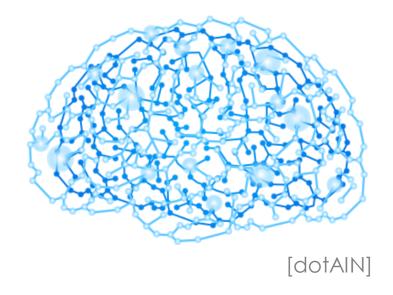




Tipping Point 4: AI – Machine Learning

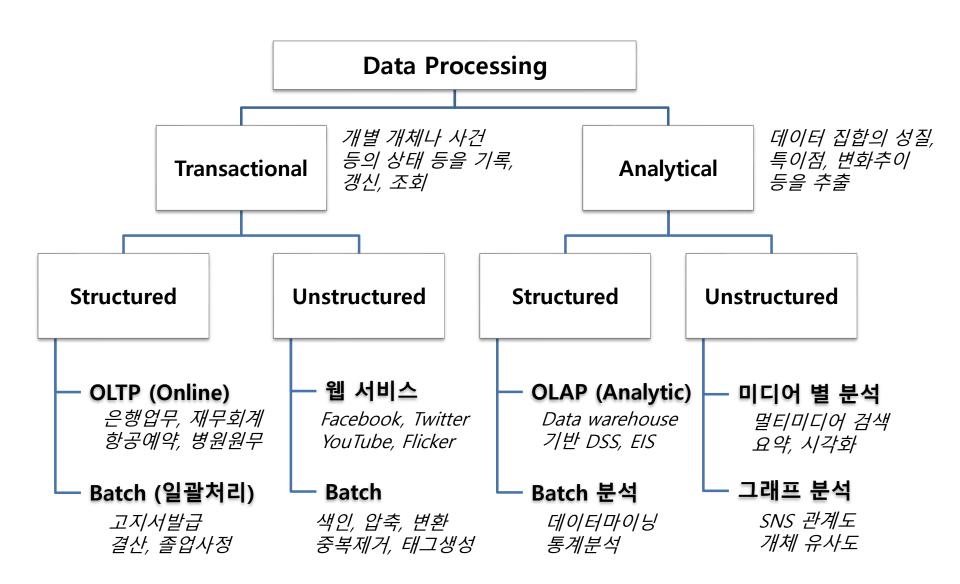
전통적 인공지능 분야 문제 해결

- Computer Vision
- Natural Language Processing
 - Google Translate, Narrative Science
- Q&A system
 - IBM Watson, Apple Siri
- Autonomous navigation
- 기계학습(machine learning)의 기반





Data Processing Tasks





Flipped Classroom

Classroom - 1233



Henry of Germany delivering a lecture to university students in Bologna, Italy, in 1233. - Artist: Laurentius de Voltolina;

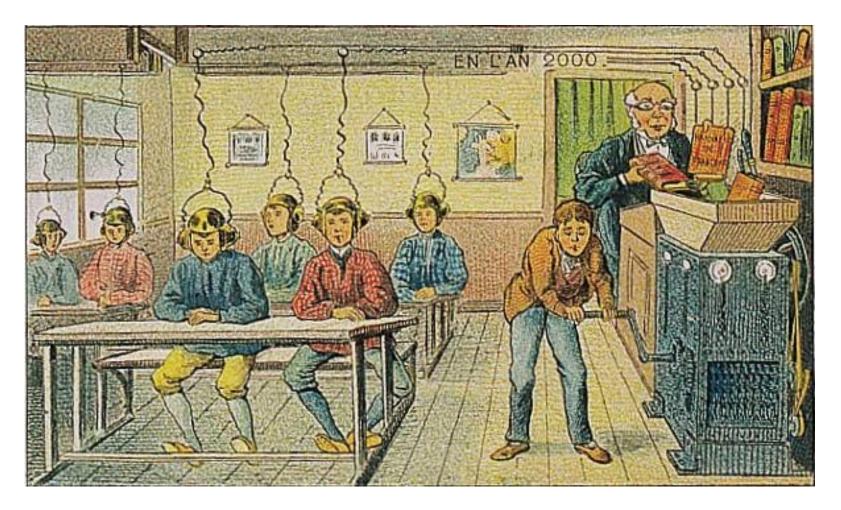


Classroom - 2016





Classroom of the Future?



[Villemard, 1910]



MOOC

Massive Open Online Course





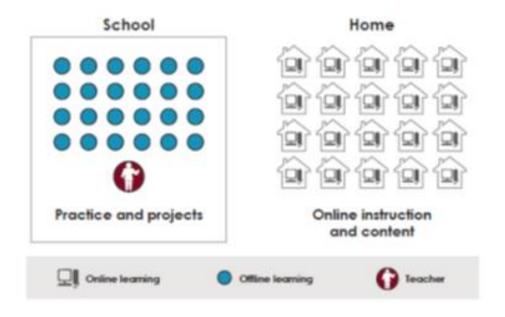
Flipped Classroom





Flipped Classroom

- Online learning off-site as homework
 - primary delivery of content and instruction is online
- Face-to-face, teacher-guided practice or projects
 - quizzes, discussions and exercises in class





In This Class

Homework & Quizzes (20%)

- Watch lecture video according to the lecture schedule
- Verification quizzes will be given at the start of each class
 - 5~10 questions (10~20min)
 - multiple choice clicker (app) problems
- Discussions and Q&A

Exams (50%)

Midterm & final: 25% each

Projects (30%)

SQL processor & DB application

F will be given for

- a score of 0 in one of the following
 - any one of the exams, or
 - over 50% of your projects, or
 - over 50% of your assignments/quizzes, or
- any type of Plagiarism!



In This Class

Text Book

Database System Concepts, 6th Ed., Silberschatz, et al, McGraw Hill, 2010

Lecture Notes & Video Links

- will be posted a few days before each class at http://ids.snu.ac.kr/site/lectures/
 - ✓ Password required
- Please use only for personal use

TA

lecture@europa.snu.ac.kr

