

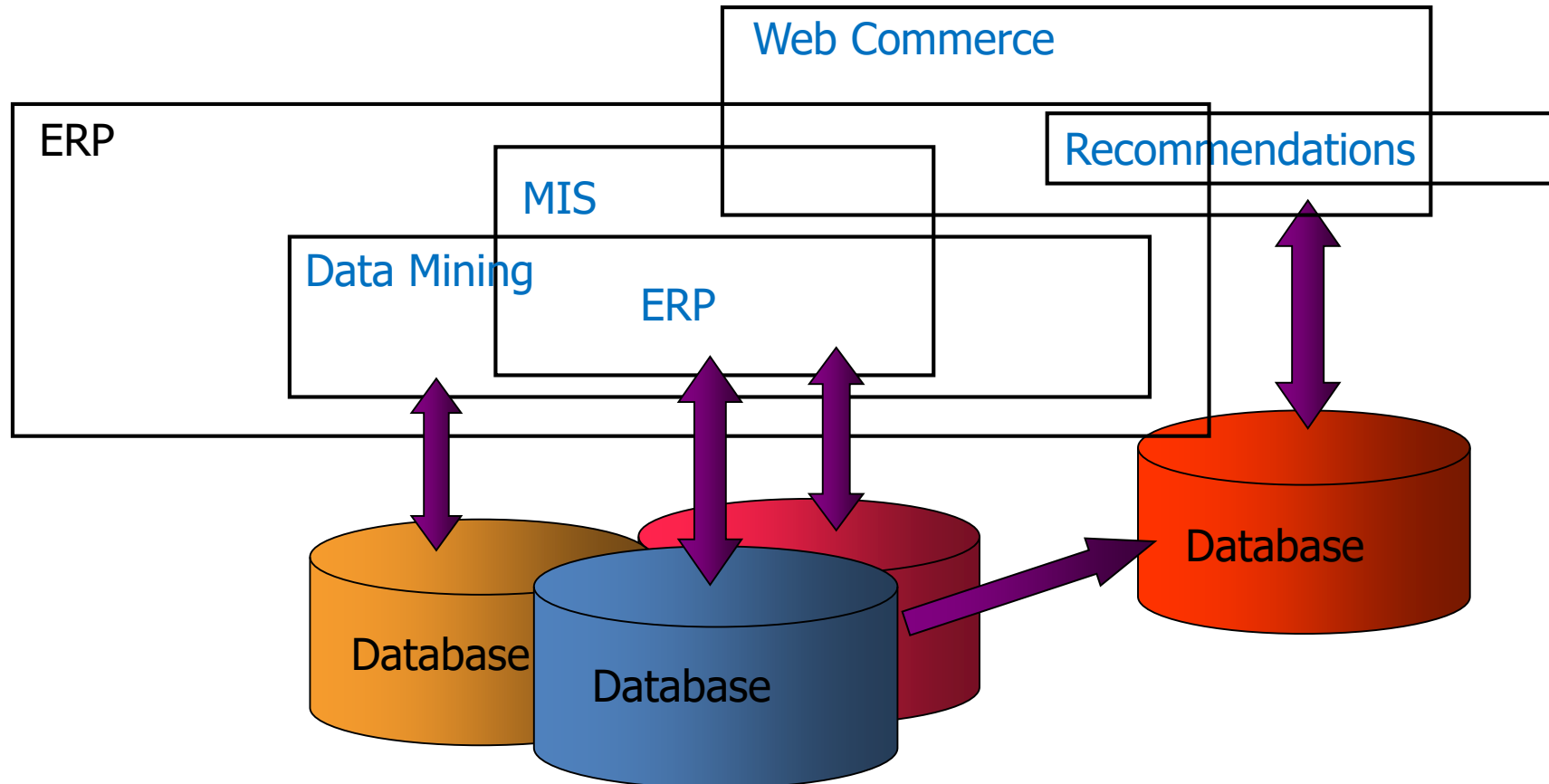
Intro to DB

CHAPTER 0

PREFACE

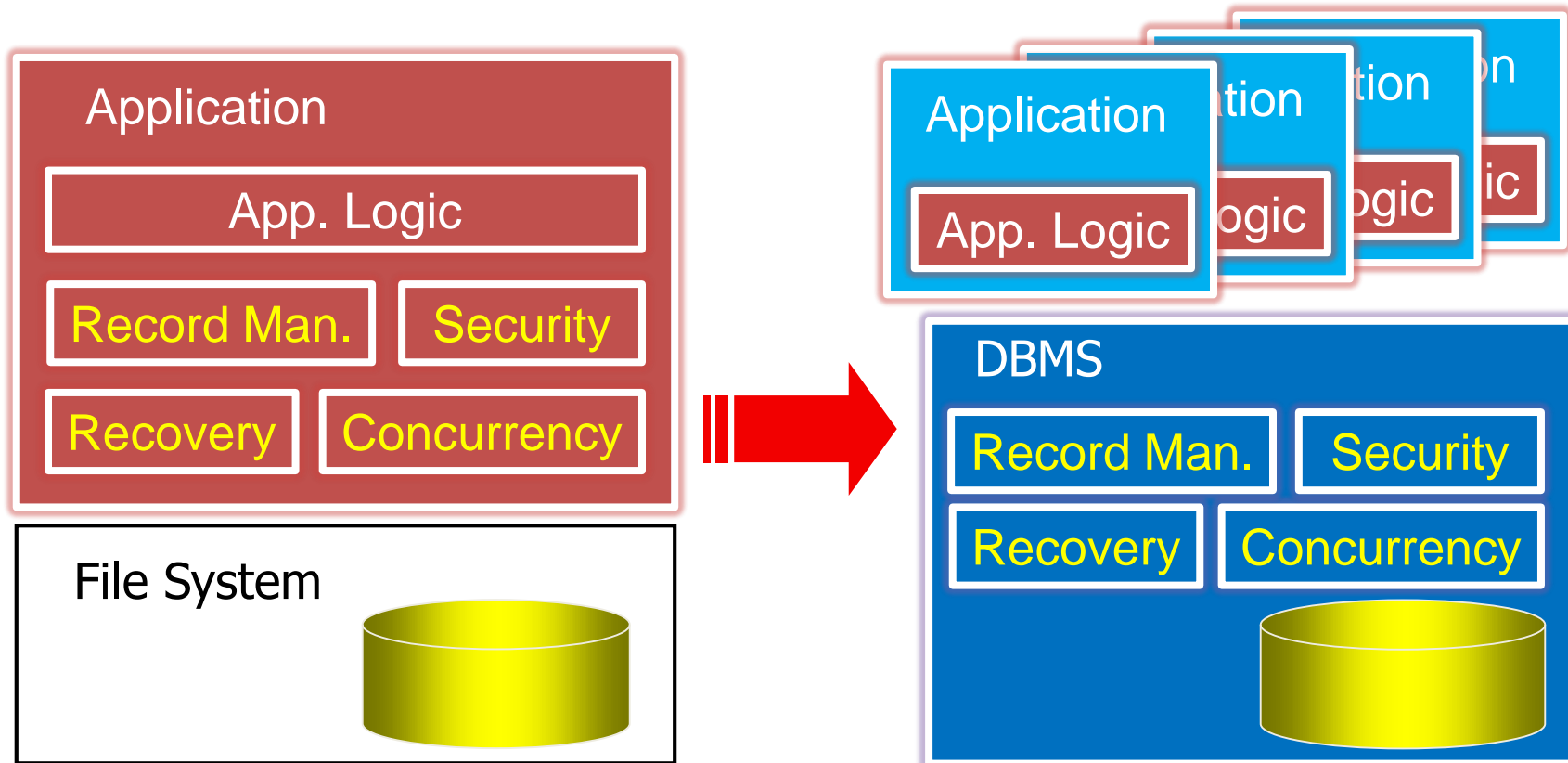
Computing = Data Processing

- 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

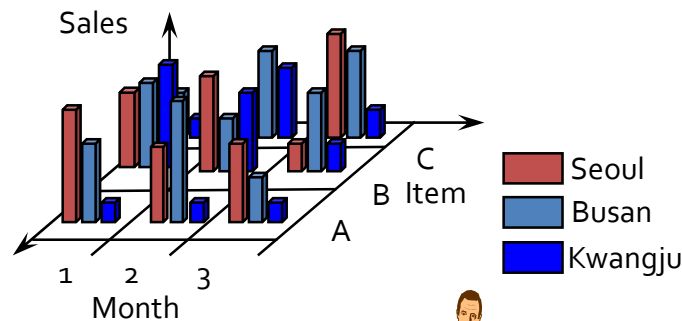
"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."



- The Adventures of Sherlock Holmes, A. Conan Doyle

- But ...

A multidimensional/analytic view



Executives

A tabular/transactional view

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000

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

course_id	title	dept_name	credits
98345	Intro. to Biology	Biology	4
76766	BIO-101	Biology	
10101	BIO-301	Genetics	
58583	BIO-399	Computational Biology	
83821	CS-101	Intro. to Computer Science	
83821	CS-190	Game Design	
15151	CS-315	Robotics	
33456	CS-319	Image Process.	
76543	CS-347	Database Syst.	

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		
15151	MU-199	1	Spring	2009		
22222	PHY-101	1	Fall	2009		
32343	HIS-351	1	Spring	2010		
45565	CS-101	1	Spring	2010		
45565	CS-101	1	Spring	2010		
45565	CS-319	1	Spring	2010		
CS-315	CS-101	76766	BIO-101	1	Summer	2009
CS-319	CS-101	76766	BIO-301	1	Summer	2010
CS-347	CS-101	83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009		
83821	CS-319	2	Spring	2010		
98345	EE-181	1	Spring	2009		

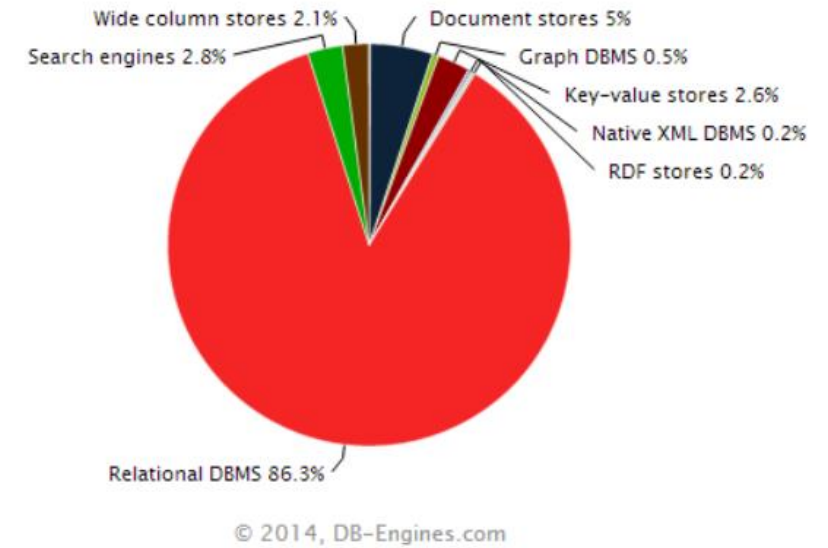


IS team



Tipping Point 1: Data Base – Relational

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



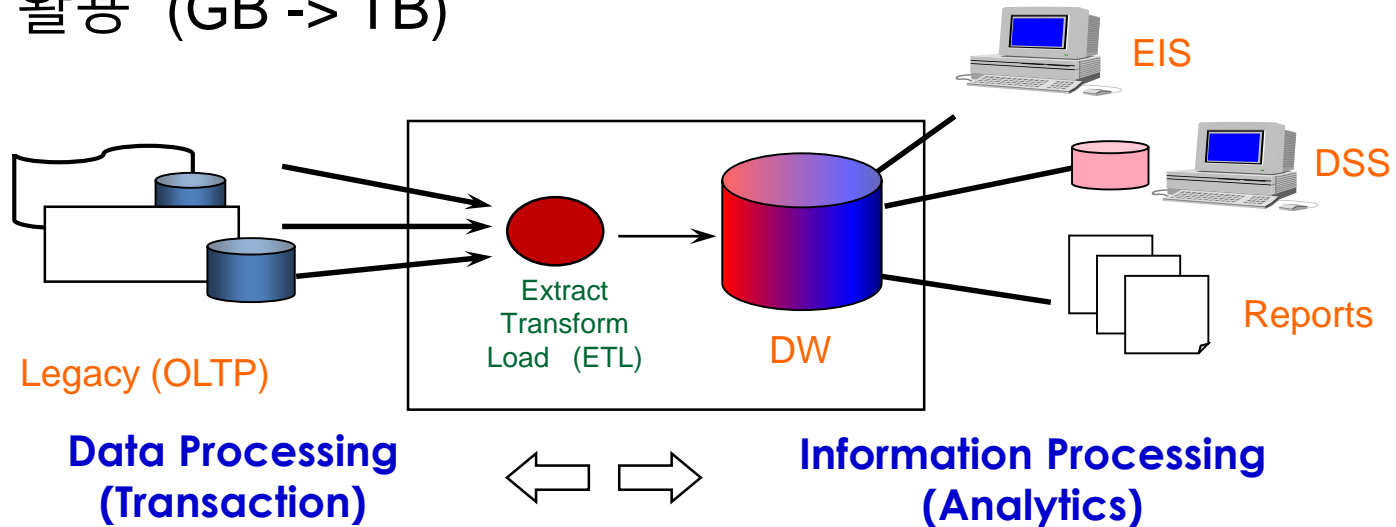
ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000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

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

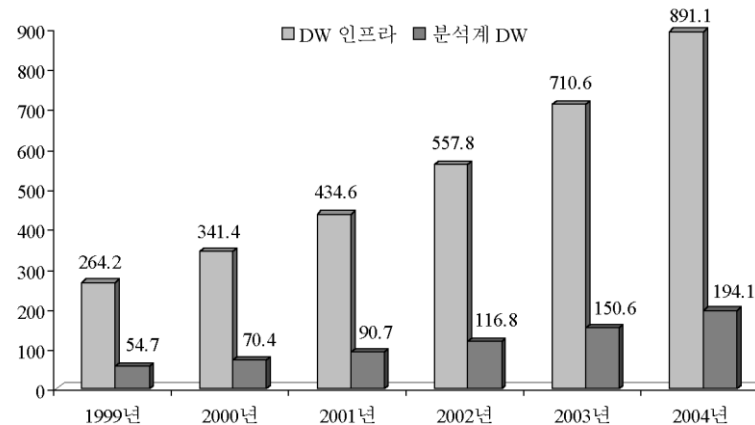
dept_name	building	budget
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

Tipping Point 2: Data Warehouse

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

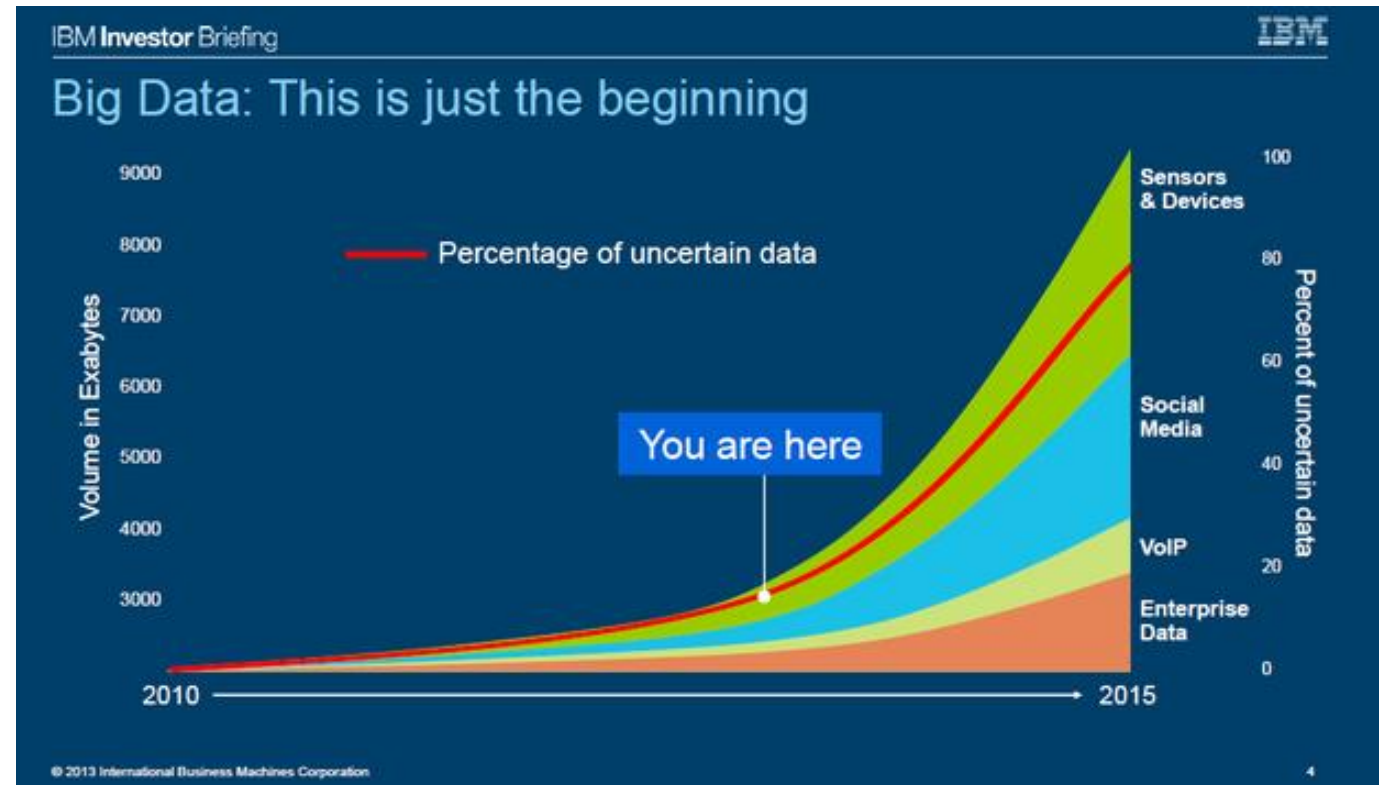


=> 데이터 분석의 전성기



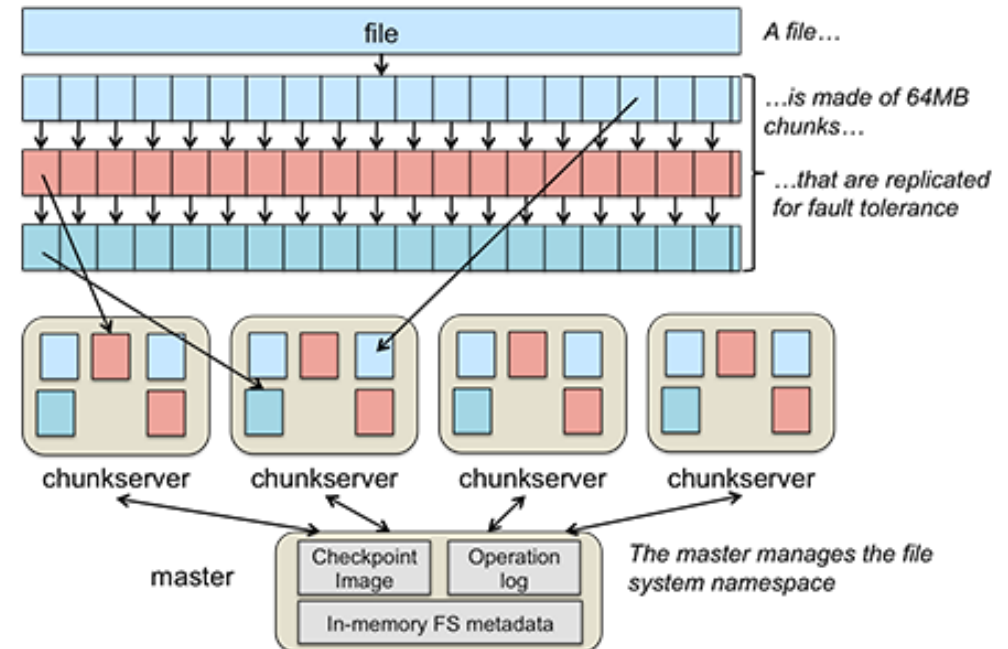
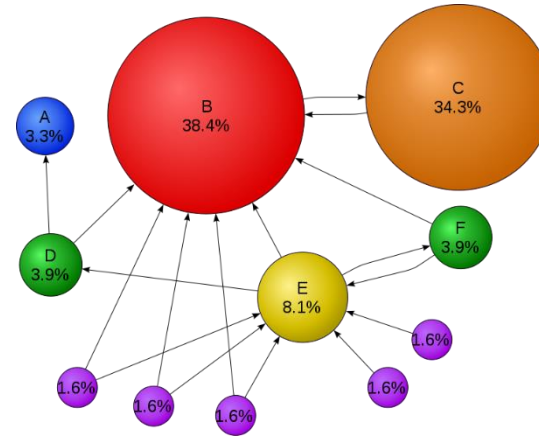
Data Explosion

- 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



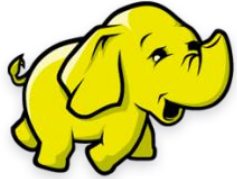
Web Scale Computing

- 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)

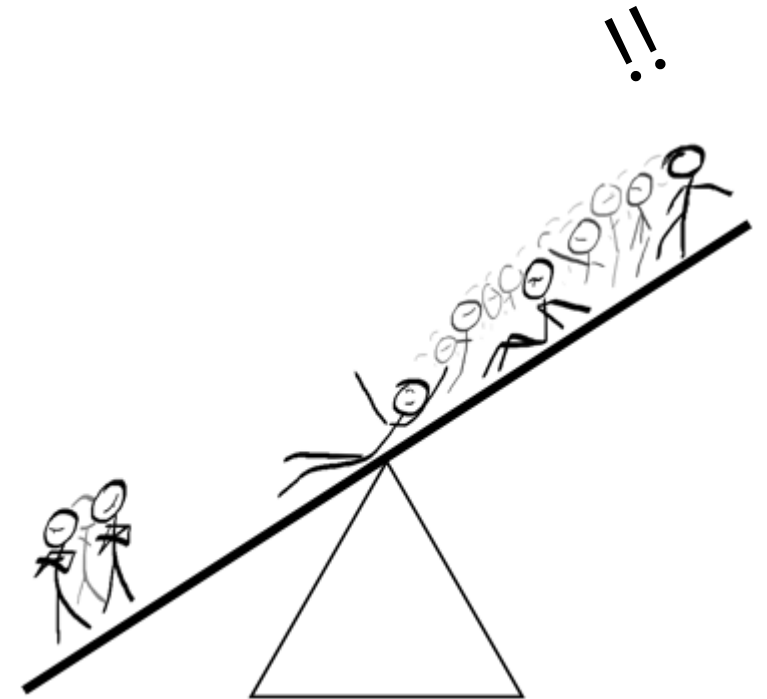
- NoSQL



- 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

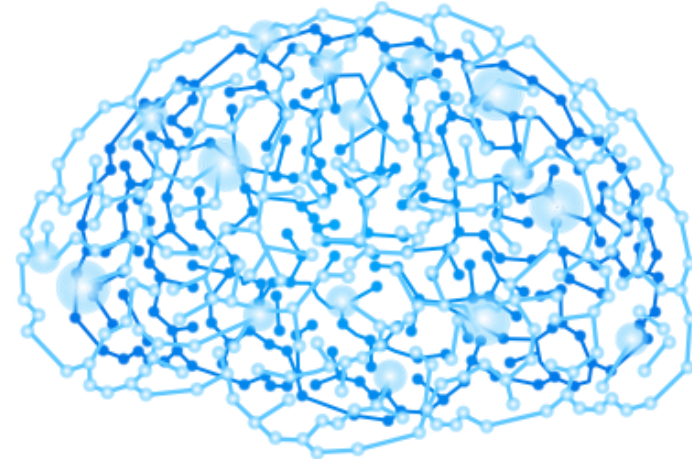


Tipping Point 4: AI – Machine Learning

전통적 인공지능 분야 문제 해결

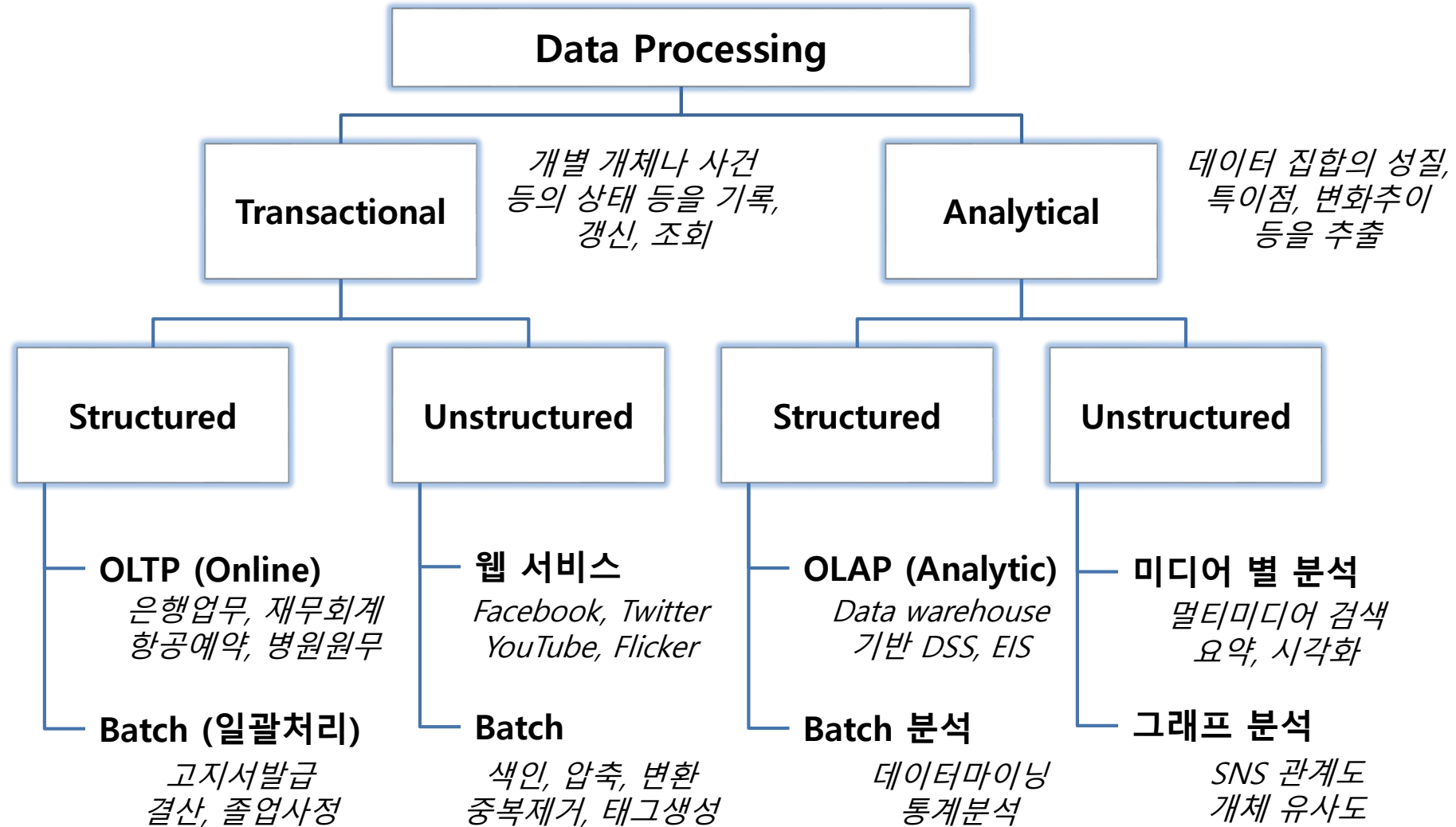
- Computer Vision
- Natural Language Processing
 - Google Translate, Narrative Science
- Q&A system
 - IBM Watson, Apple Siri
- Autonomous navigation

- Machine learning: computer version of evidence-based decision making



[dotAIN]

Data Processing Tasks



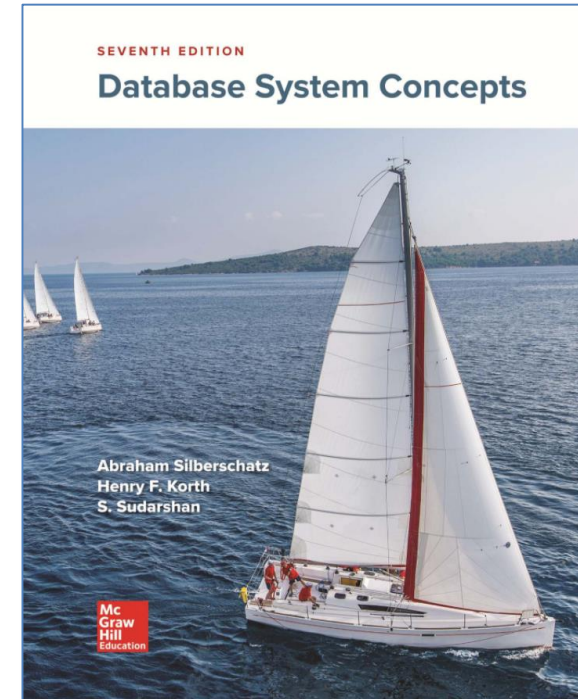
Materials

- **Text Book**

Database System Concepts, 7th Ed., Silberschatz, et al,
McGraw Hill, 2020

- **Lecture Notes & Materials**

- will be posted at <http://ids.snu.ac.kr/site/lectures/>
 - ✓ Password required
- Please use only for personal use



Evaluation

- **Quizzes, Assignments, and others (20%)**
 - **Verification quizzes** given at the beginning of each day of lecture
 - On materials covered on the previous lecture day
 - Discussion problems (HW & in-class)
 - Computer exercises – SQL
- **Exams (50%)**
 - Midterm & Final Exam: 25% each
- **Projects (30%)**
 - SQL processor
 - DB application with Python & SQL
- **Minimum**
 - A score of 0 in any one of the exams will result in F
 - A score of 0 in 50% of the projects, assignments, quizzes will result in F