INTRODUCTION TO DATABASES
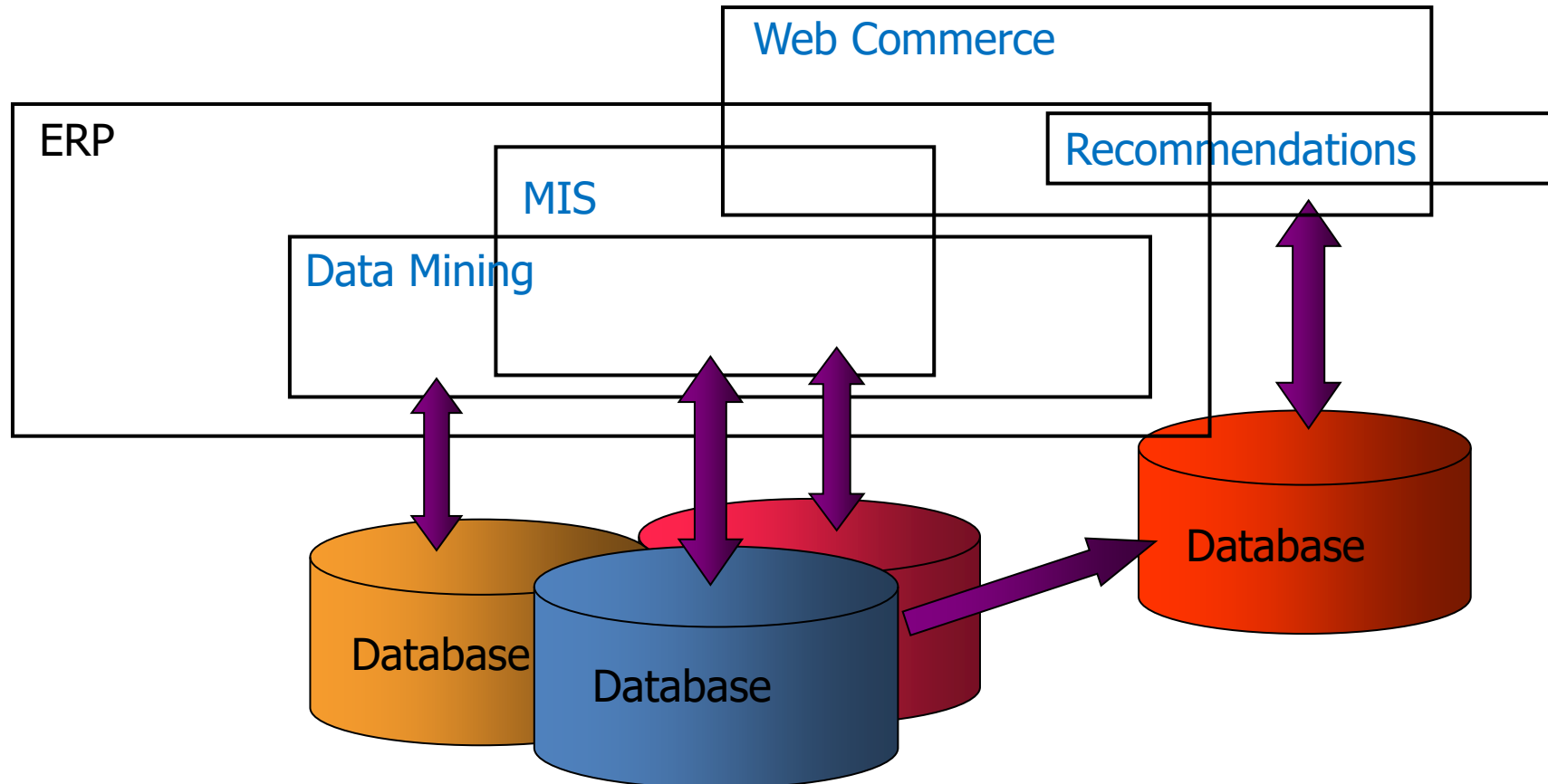
Prof. Sang-goo Lee

**Computer Science & Engineering
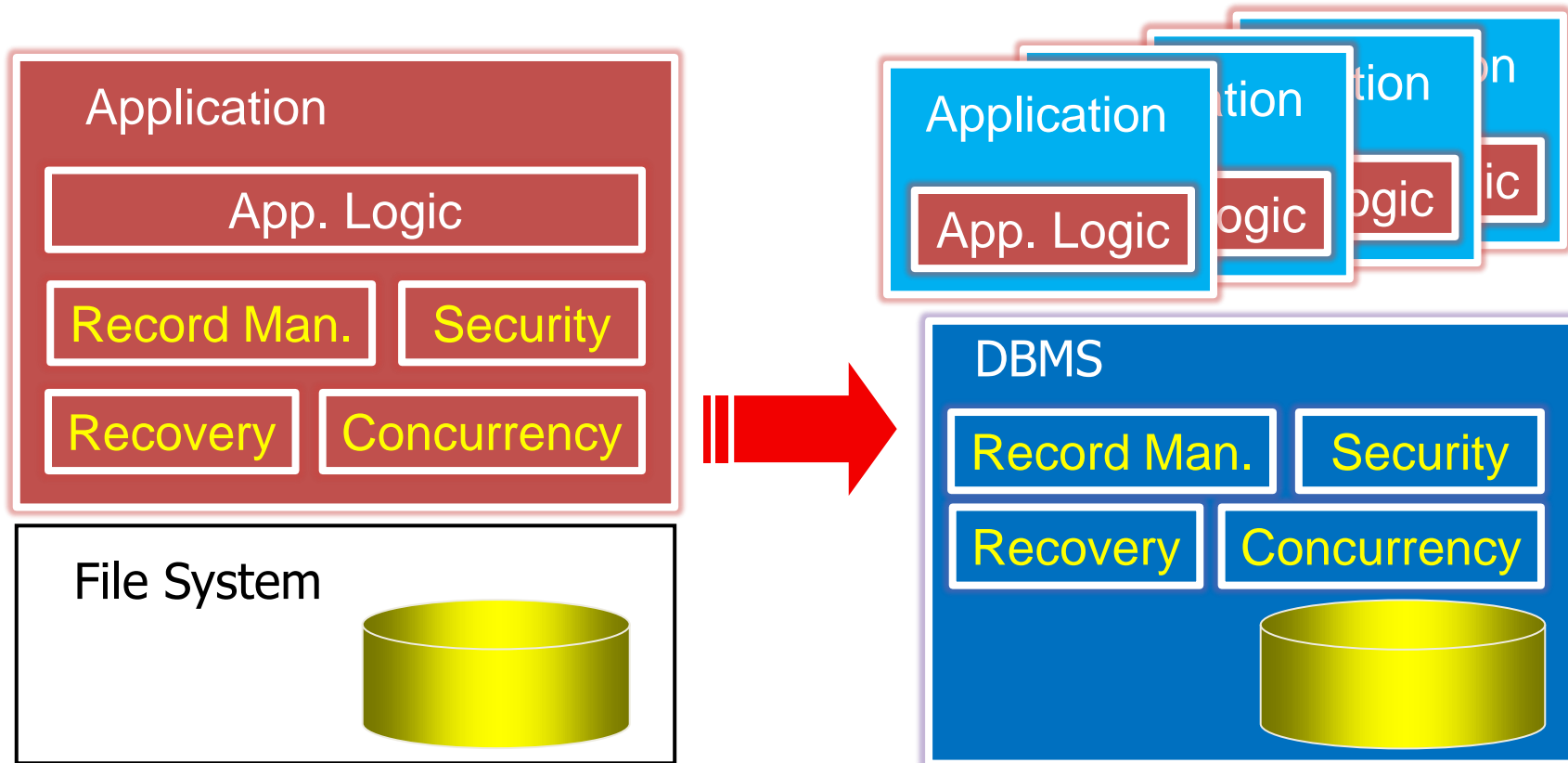Seoul National University**

Computing = Data Processing

- Most (all?) computing applications use some type of a database
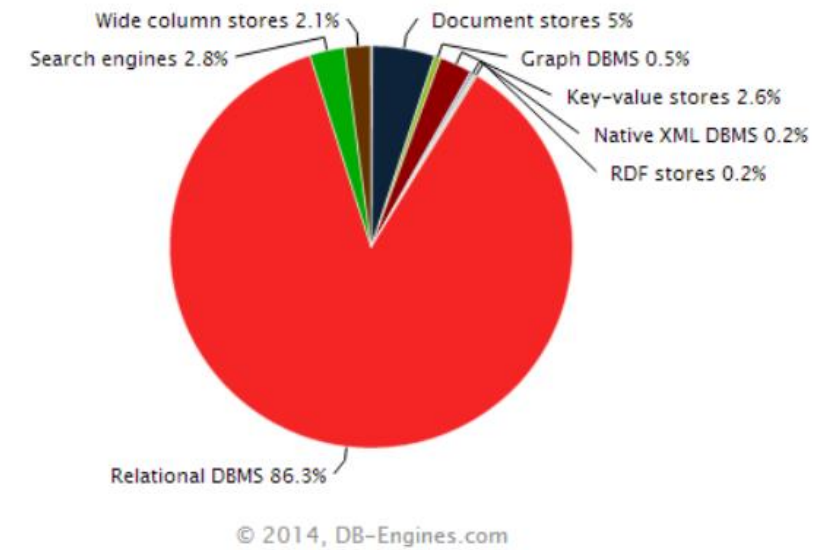
Ground 0: File System

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

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%
Search engines 2.8%
Document stores 5%
Graph DBMS 0.5%
Key–value stores 2.6%
Native XML DBMS 0.2%
RDF stores 0.2%
Relational 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

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

Evidence Based Decision Making

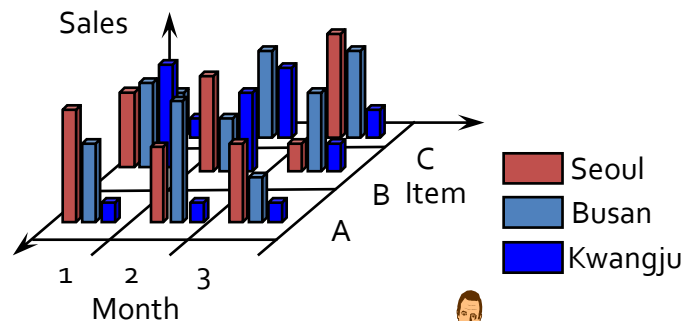- Insights(통찰력) & foresights(예지력) through data

> *"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

A *tabular/transactional* view

ineffective
DSS Cycle

Executives

IS team

Tipping Point 2: Data Warehouse

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

Data Processing (Transaction) ⟸ ⟹ **Information Processing (Analytics)**
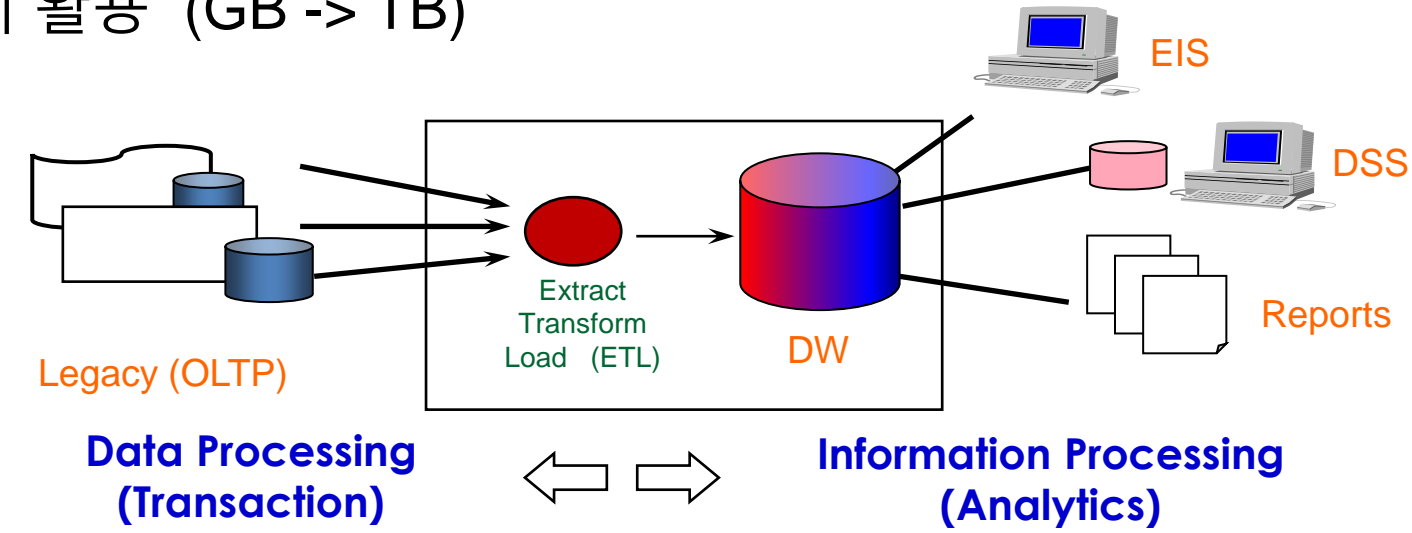
=> *데이터 분석의 전성기*
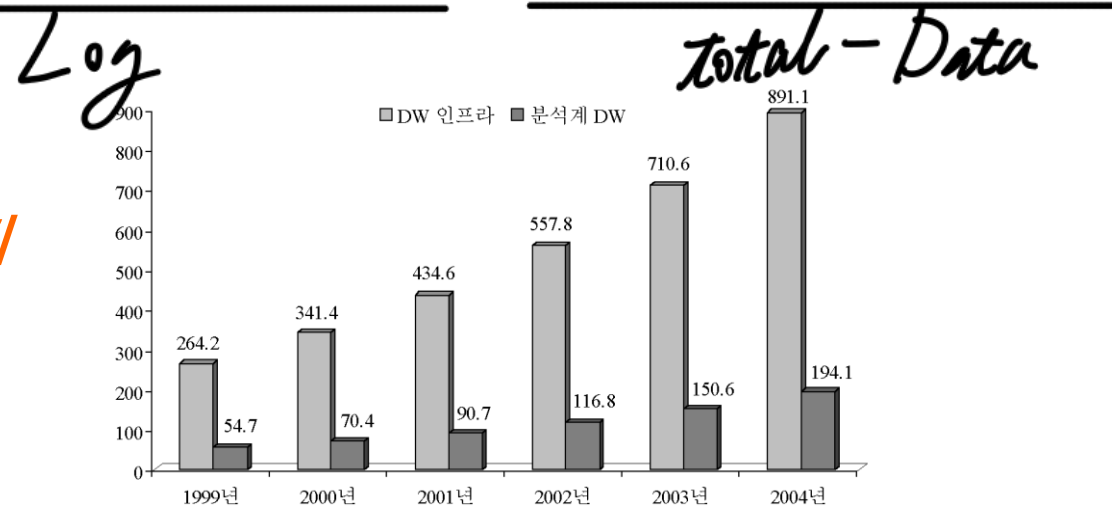
Data Explosion

- 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
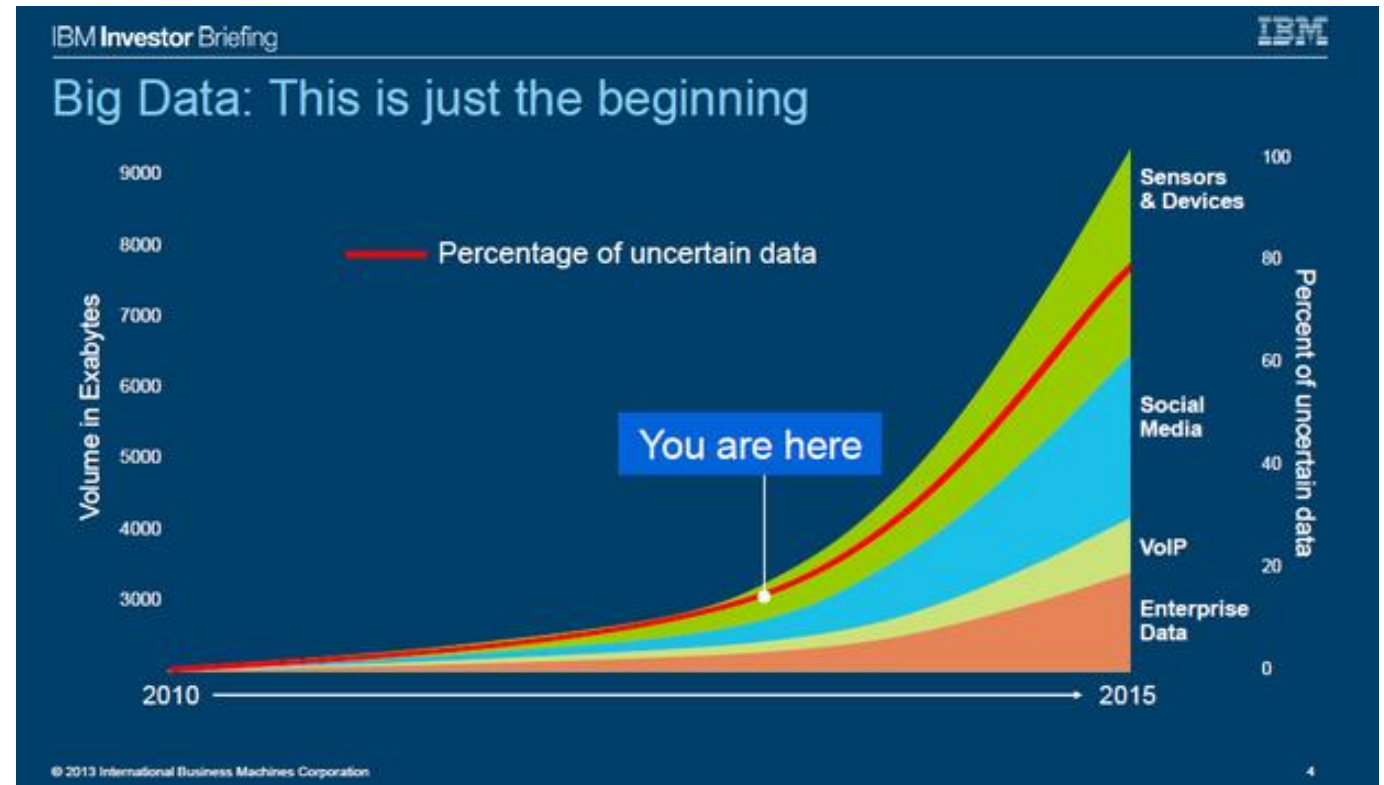
Tipping Point 3: Big Data : *Relational to Webscale*

- ***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
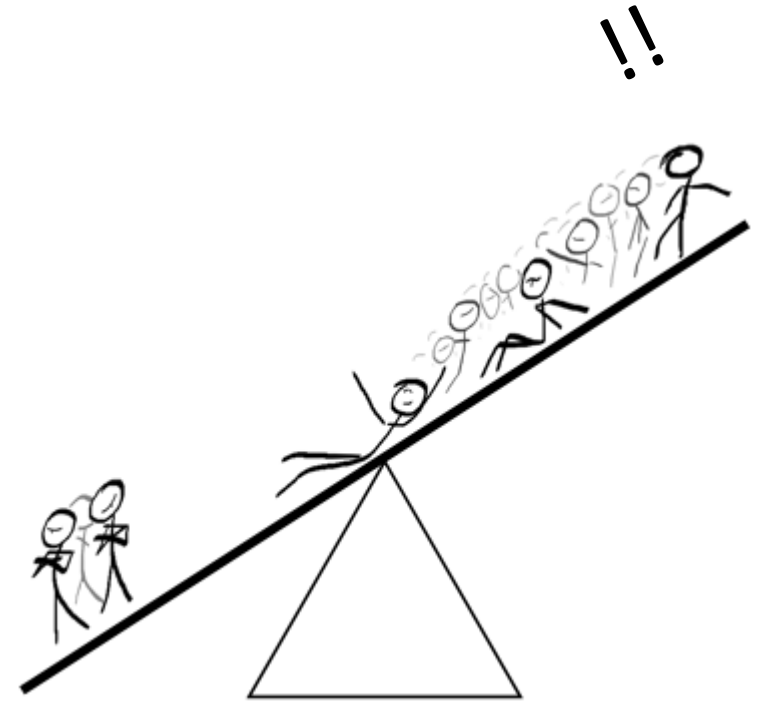
Web Scale Computing

- Different data,
- Different operations,
- Different scale!
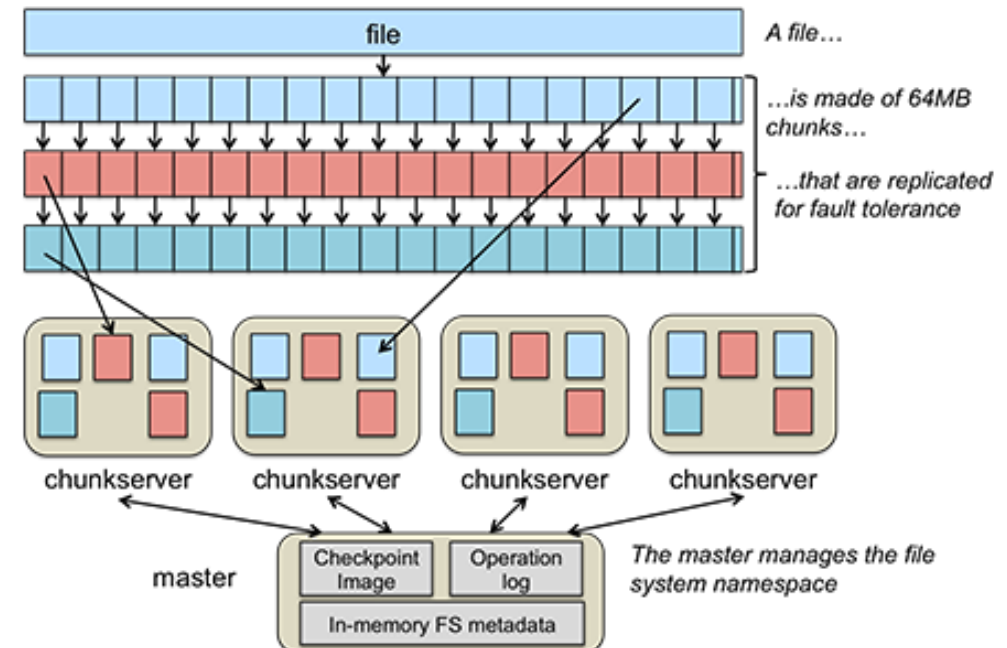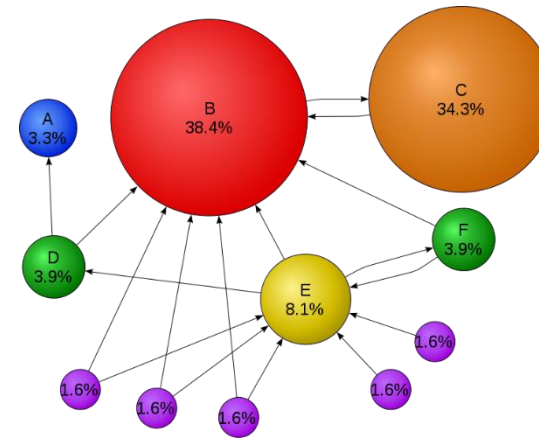
Big Data Systems
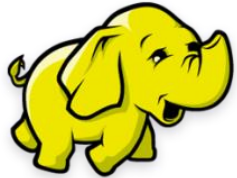
- ## Hadoop
  - Apache Open Source SW
  - MapReduce 기반의 대량 데이터 분산처리 framework
  - Yahoo!에서 시작/지원 (2006) : *Based on Google filesystem*

- ## NoSQL
  - Not Only (?) SQL : *More complicated process needed*
  - 단순한 데이터 모델: Key-Value store
  - 단순한 질의: get(), put()
  - 단순한 트랜잭션 모델: BASE – Basically Available, Soft state, Eventual consistency

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

```
                          ┌─────────────────────┐
                          │   Data Processing   │
                          └─────────────────────┘
```

log **Transactional** 개별 개체나 사건 등의 상태 등을 기록, 갱신, 조회   **Analytical** 데이터 집합의 성질, 특이점, 변화추이 등을 추출 *Set*

Structured   **Unstructured**   **Structured**   **Unstructured**

Immediately

OLTP (Online)
은행업무, 재무회계
항공예약, 병원원무

웹 서비스
*Facebook, Twitter
YouTube, Flicker*

OLAP (Analytic)
*Data warehouse
기반 DSS, EIS*

미디어 별 분석
멀티미디어 검색
요약, 시각화

Batch (일괄처리)
고지서발급
결산, 졸업사정

Batch
색인, 압축, 변환
중복제거, 태그생성

Batch 분석
데이터마이닝
통계분석

그래프 분석
SNS 관계도
개체 유사도

On time
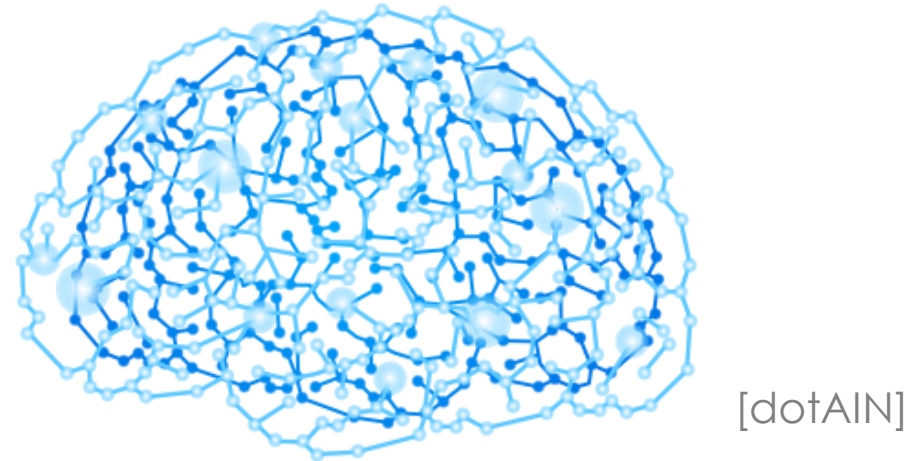
FLIPPED LEARNING

Classroom - 1233

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

Classroom of the Future?
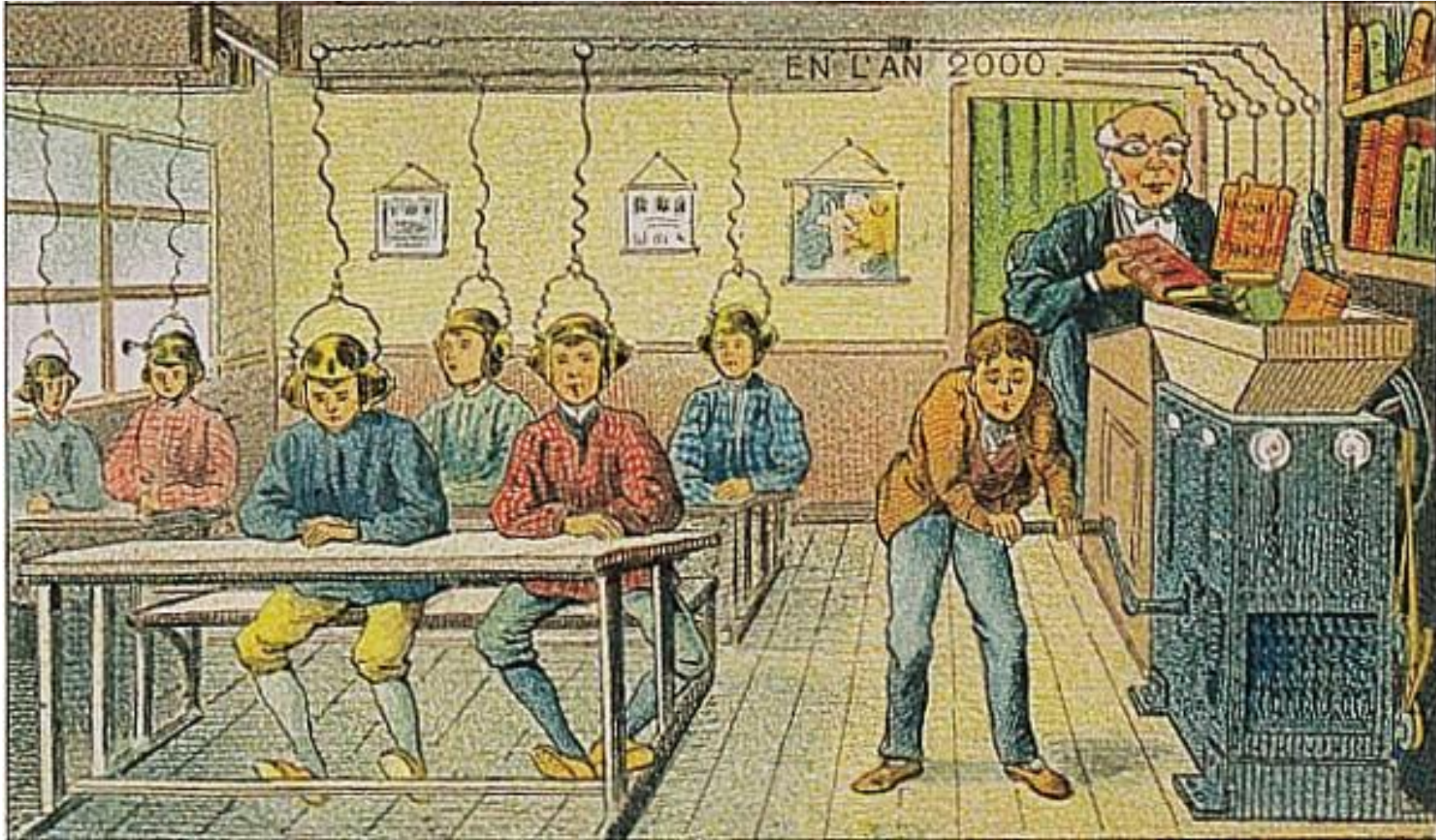
[Villemard, 1910]

Flipped Learning

The Flipped Classroom
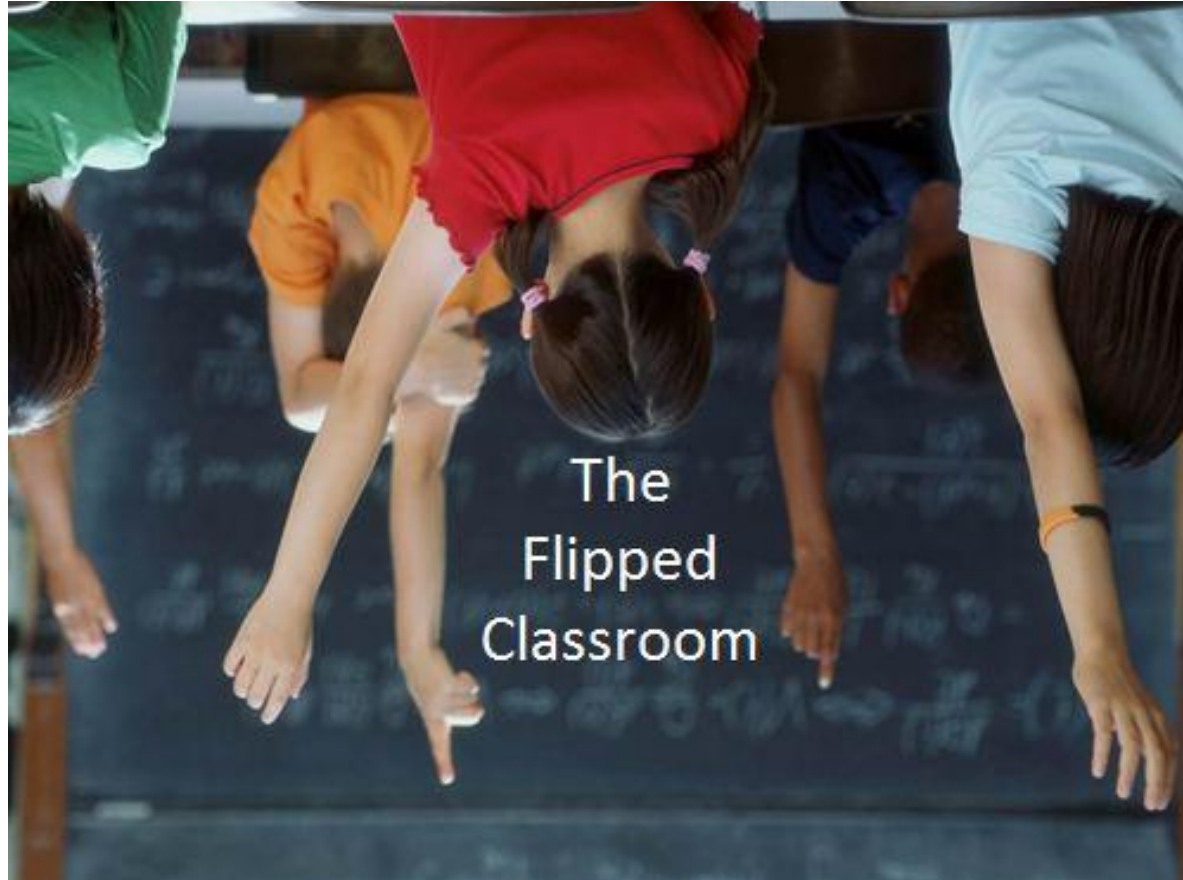
Flipped Learning

□ **Online learning as homework**

  ▪ primary delivery of content and instruction is online

□ **Face-to-face, teacher-guided practice or projects**

  ▪ quizzes, discussions and exercises in class
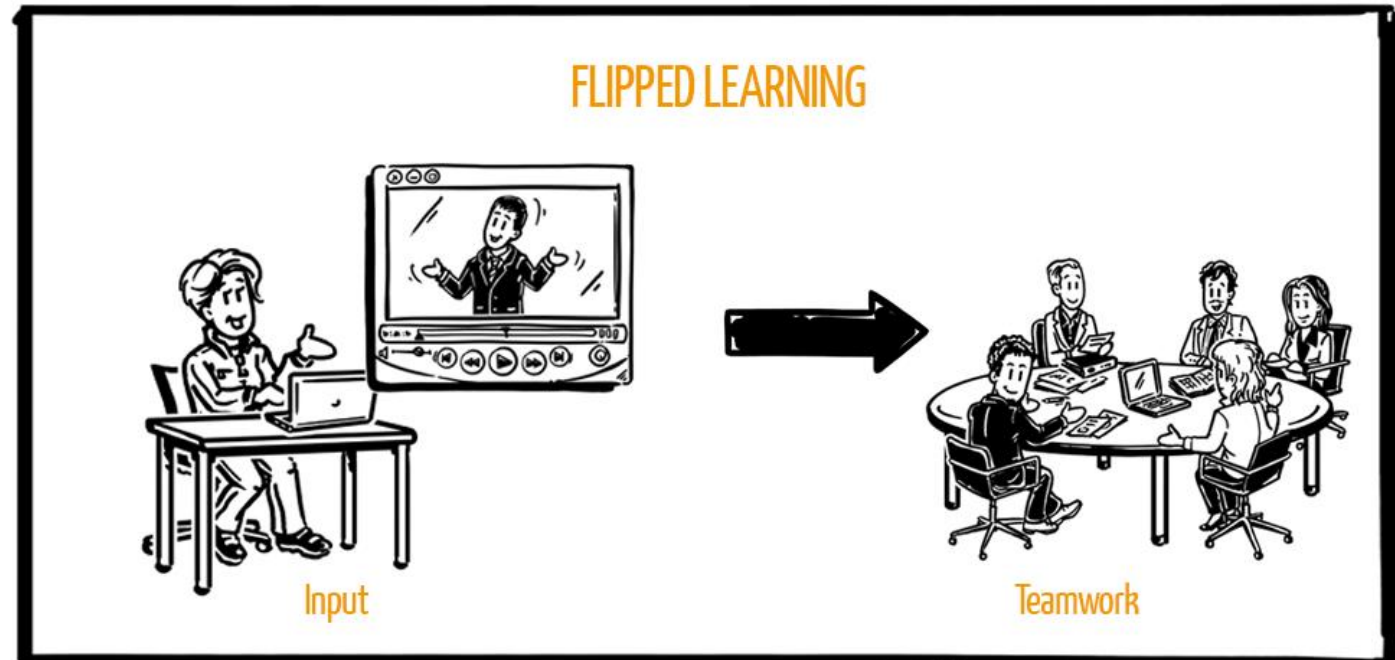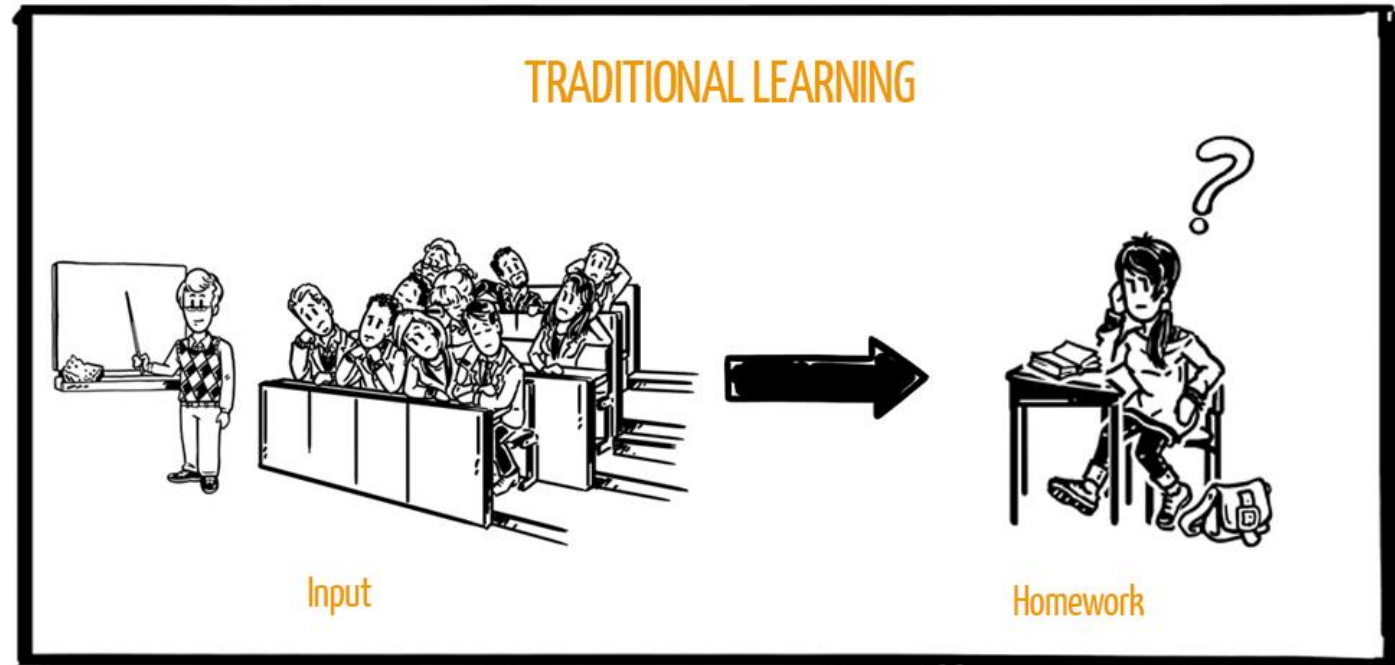
In This Class

- **Homework & Quizzes (20%)**
  - **Watch lecture video** according to the lecture schedule
  - Verification quizzes will be given at the start of each class (online)
    - 5~10 multiple choice questions (10~20min)
  - 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*!

Materials

- **Textbook**

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

- **Lecture Notes & Materials**
  - will be posted at class site
    - ✓ Password required
  - Please use only for personal use

Schedule

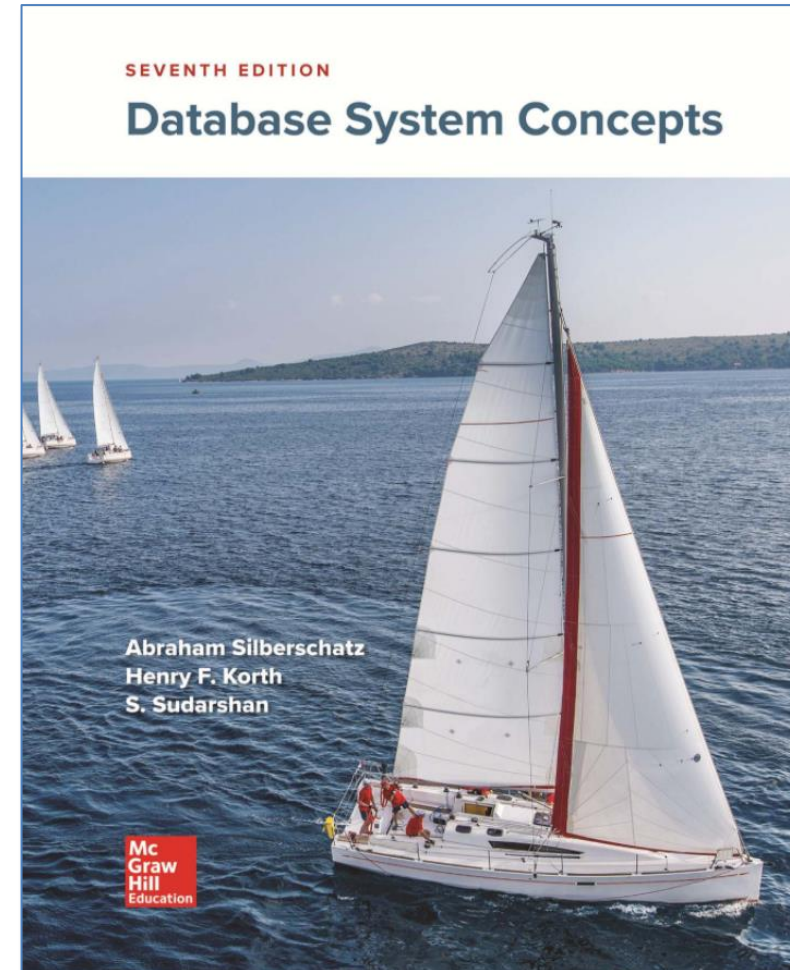| Week	Date	Chapter	Video (mins)		Subjects	Proj
WK 1	3/6 Mon	Chap 00	v1	34:02	DB History	
	3/8 Wed	Chap 01	v1	10:31	Data, Database, DBMS	
			v2	15:06	File System, Data Indep., Instance & Schema	
			v3	11:56	Data Models, DB Languages, DB Users, Architecture	
WK 2	3/13 Mon	Chap 02	v1	12:05	Structure/Definition of Relational Model, Attributes	
			v2	19:47	Relationa Schema/Instance, Keys, Schema Diagram	
			v3	12:58	Select / Project / Union / Set Difference	
	3/15 Wed		v4	14:21	Cartesian Product / Rename	
			v5	20:32	Formal Def, Intersection / Nat Join / Assignment	
WK 3	3/20 Mon	Chap 03	v1	16:42	SQL, Create/Drop/Alter Table	Proj 1-1
			v2	19:07	Select / From / Where	
	3/22 Wed		v3	18:16	Rename / Strings / Nulls	
			v4	13:55	Aggregate Functions	
			v5	19:38	Nested Query	
WK 4	3/27 Mon				**SQL Lab 1 (실습)**	
	3/29 Wed		v6	12:43	Correlation Variables	
			v7	10:56	Insert, Delete, Update	
WK 5	4/3 Mon	Chap 04	v1	27:31	Joins	
			v2	23:07	Integrity Constraints / Authorization	
	4/5 Wed				**SQL Lab 2 (실습)**	Proj 1-2
WK 6	4/10 Mon	Chap 06	v1	22:28	Design Process	
			v2	14:12	Entity, Relationship, E-R Diagram	
			v3	24:45	Mapping Cardinality	
	4/12 Wed		v4	27:36	Extended E-R Features, Weak Entity Sets	
			v5	26:16	Logical Design, Design Issues	
WK 7	4/17 Mon				**Design Lab**	
	4/19 Wed				**Midterm Exam**	

Schedule

Week	Date	Chapter	Video	Time	Topic	Project
WK 8	4/24 Mon	Chap 07	v1	42:02	1NF, Functional Dependency	Proj 1-3
			v2	27:08	BCNF, Decomposition	
	4/26 Wed		v3	18:11	Dependency Preservation	
			v4	32:18	Algorithims	
			v5	14:08	Overall Database Design Process	
WK 9	5/1 Mon	Chap 12	v1	34:51	Storage Systems	
			v2	23:10	RAID	
	5/3 Wed	Chap 13	v1	23:28	File Organization	
			v2	12:15	Database Buffer	
WK 10	5/8 Mon	Chap 14	v1	28:08	Ordered Index, Dense/Sparse, Primary/Secondary	
			v2	25:30	B+-tree: Structure, Search	
	5/10 Wed		v3	18:52	B+-tree: Insertion	
			v4	17:52	B+-tree: Deletion	
WK 11	5/15 Mon	Chap 15	v1	20:35	Query Processing, Cost Measures	
			v2	13:20	Selection operations	
	5/17 Wed		v3	12:20	External Sort Merge	
			v4	17:26	Nested Loop Join	
			v5	11:15	Hash Join	
WK 12	5/22 Mon	Chap 17	v1	31:11	Transactions, ACID	Proj 2
			v2	20:36	Transaction States, Schedules	
	5/24 Wed		v3	29:08	Serializablility, Recoverability	
WK 13	5/29 Mon	Chap 18	v1	27:57	Concurrency control, 2PL	
		Chap 19	v1	17:05	Atomicity & Durability, Data Access	
	5/31 Wed		v2	21:36	Log-based Recovery	
			v3	35:55	Rollback,Concurrent Transactions, WAL	
WK 14	6/5 Mon				General Q&A	
	6/7 Wed				Final Exam	
WK 15	6/12 Mon				Project Q&A	
	6/14 Wed					Proj 2 Due