



**MONGODB**

박경미

1. NoSQL의 개념
2. MongoDB 설치 및 데이터 처리
3. 인덱스 생성과 관리 & 사용자 관리
4. MongoDB를 위한 Data Modeling
5. 논리적 구조 & 물리적 구조
6. Sharding System
7. Replica & ReplicaSet
8. MongoDB 성능 튜닝
9. MongoDB 백업/복구 & 유틸리티

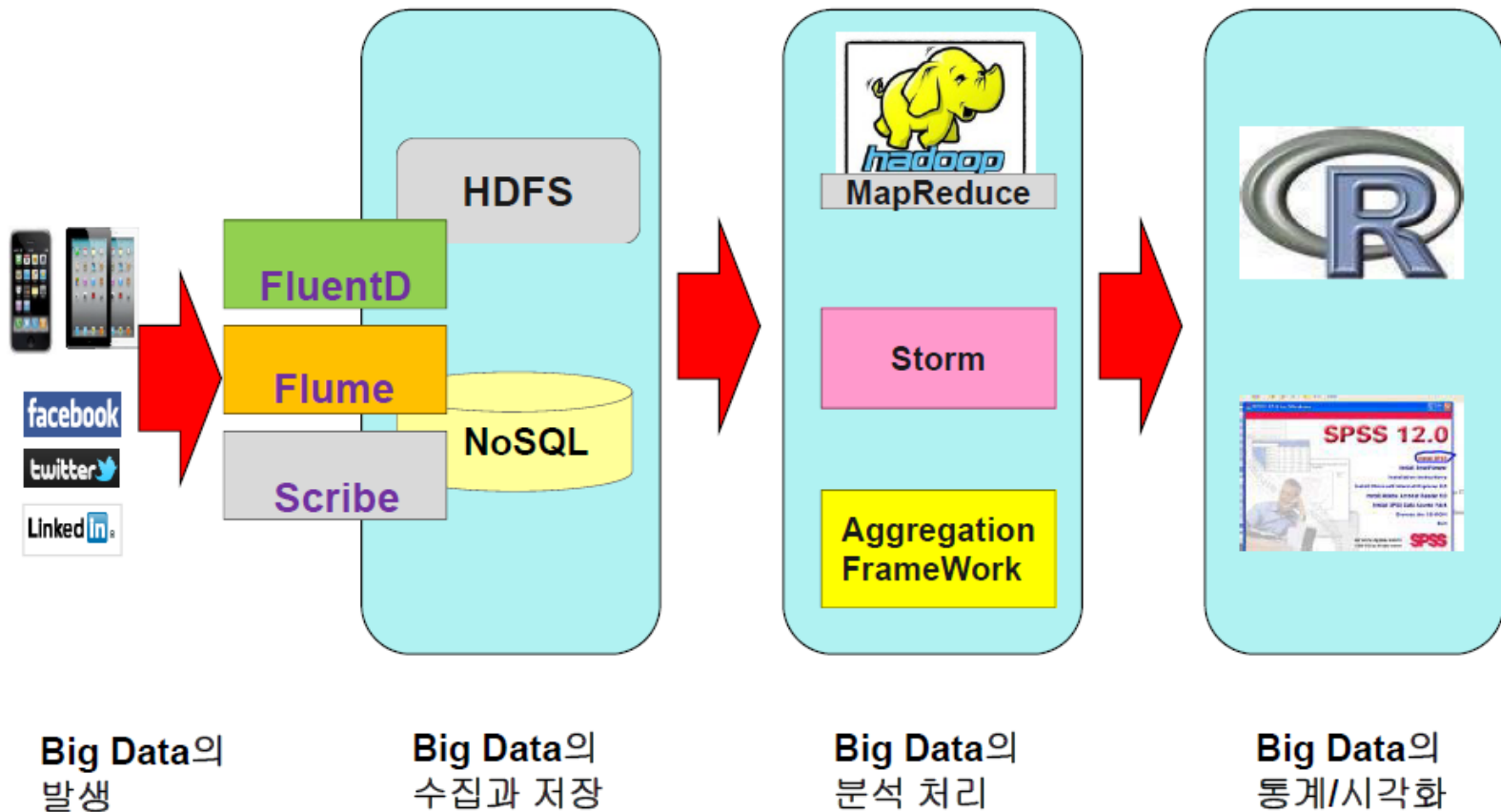


# 1. NoSQL의 개념

1. NoSQL이란
2. NoSQL의 시대적 요구
3. NoSQL의 장점
4. NoSQL의 종류
5. NoSQL의 제품군
6. MongoDB EcoSystem
7. 적용사례

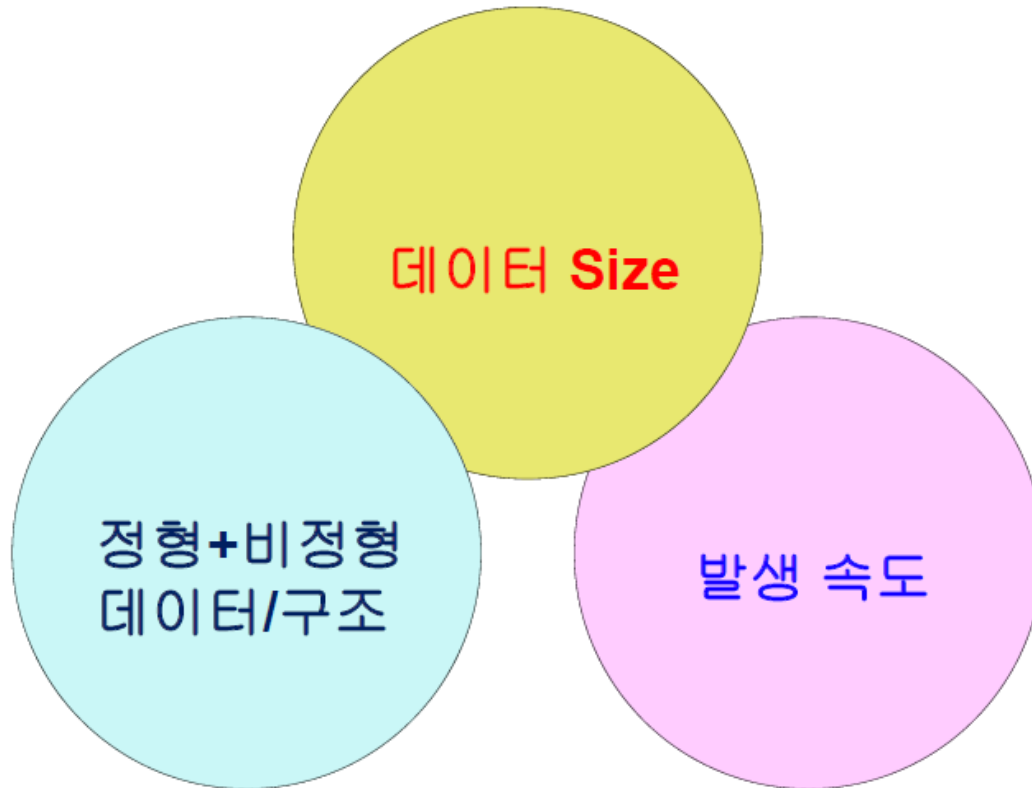
# 1. NoSQL이란

## □ Big Data Processing Flow



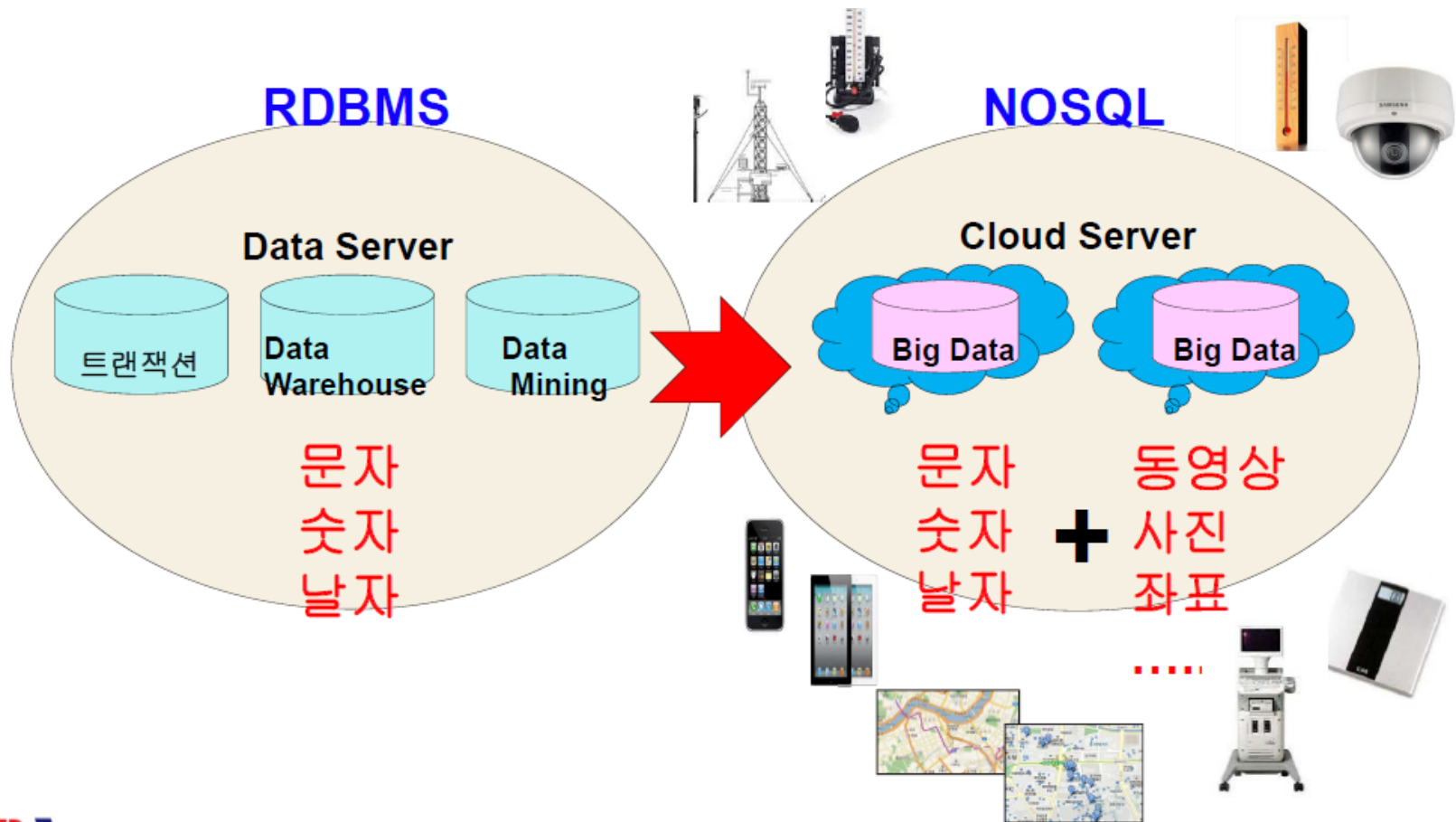
# 1. NoSQL이란

## □ Big Data 의미



# 1. NoSQL이란

## □ RDBMS VS NoSQL



# 1. NoSQL이란

## □ NoSQL ?

- NoSQL 데이터베이스는 전통적인 관계형 데이터베이스 보다 덜 제한적인 일관성 모델을 이용하는 데이터의 저장 및 검색을 위한 매커니즘을 제공

No SQL  
Not Only SQL

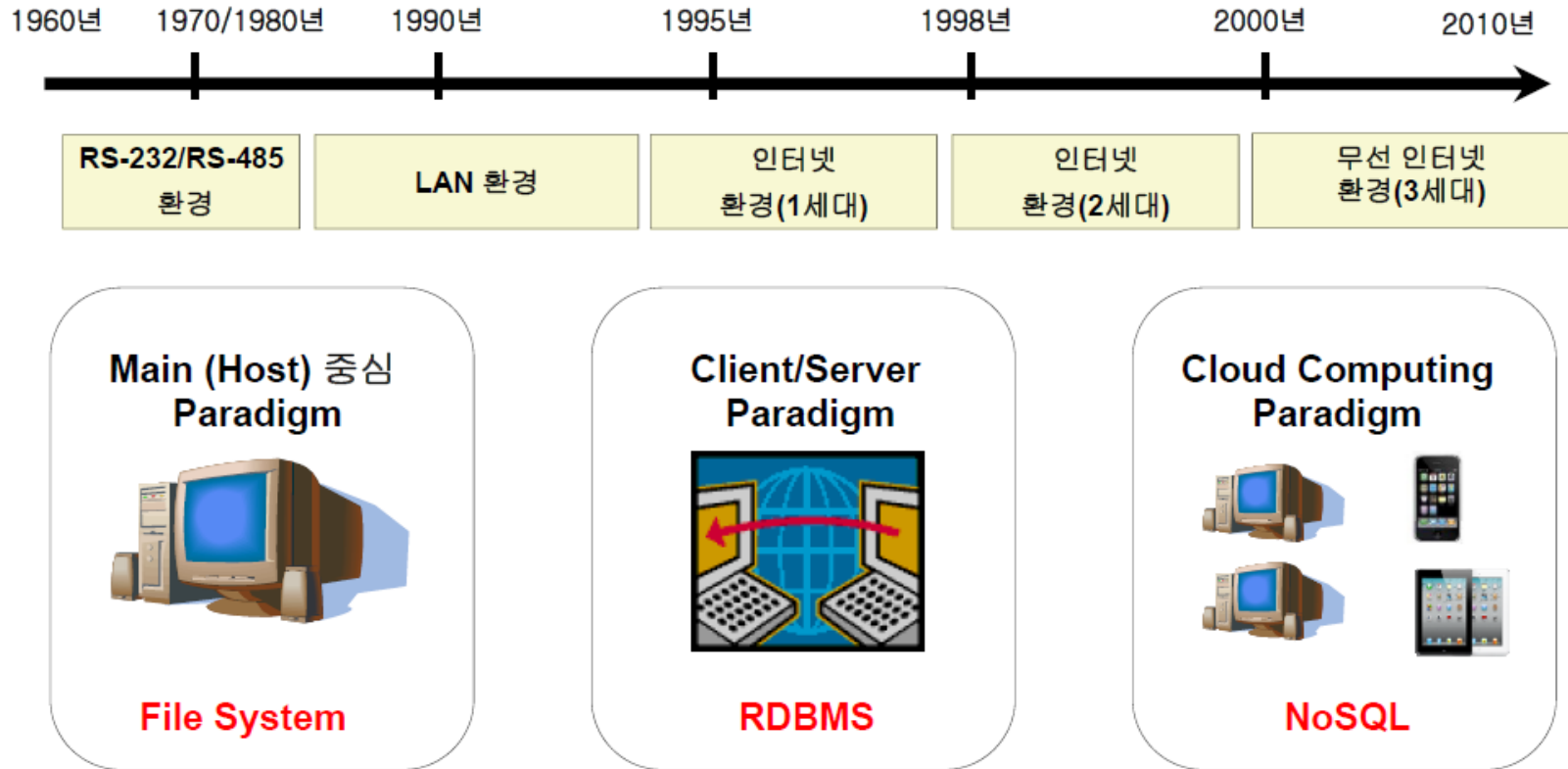
Non-Relational Operational Database SQL



mongoDB

## 2. NoSQL의 시대적 요구

### □ NoSQL의 시대적 요구





## 2. NoSQL의 시대적 요구

### □ 역사

- 카를로 스트로찌(Carlo Strozzi)는 1998년 표준 SQL 인터페이스를 채용하지 않은 자신의 경량 오픈 소스 관계형 데이터베이스를 *NoSQL*이라고 명명
- 스트로찌는 현재의 NoSQL 운동이 “전반적인 관계형 모델에서 점차 멀어지고 있으므로” NoREL로 부르는 것이 더 적절하다고 언급
- 2009년 초에 라스트 FM의 요한 오스칼손(Johan Oskarsson)이 오픈 소스 분산 데이터베이스를 논하기 위한 미트업 행사를 조직하면서, 이와 같은 데이터베이스를 NoSQL이라고 불렀다.
- 고전적인 관계형 데이터베이스 시스템의 주요 특성을 보장하는 ACID 제공을 주로 시도하지 않은 수많은 비관계형, 분산 데이터 자료 공간의 등장에 따라 이 이름이 사용

# 3. NoSQL의 장점

## 1. 클라우드 컴퓨팅 환경에 적합

- 1) Open Source
- 2) 하드웨어 확장에 유연한 대처가 가능
- 3) RDBMS에 비해 저렴한 비용으로 분산 처리와 병렬 처리가 가능

## 2. 유연한 데이터 모델

- 1) 비정형 데이터 구조 설계로 설계 비용 감소
- 2) 관계형 데이터베이스의 Relationship과 Join구조를 Linking과 Embedded로 구현하여 성능이 빠름

## 3. Big Data 처리에 효과적

- 1) Memory Mapping 기능을 통해 Read/Write가 빠름.
- 2) 전형적인 OS와 Hardware에 구축 가능.
- 3) 기존 RDB와 동일하게 데이터 처리가 가능

# 4. NoSQL의 종류

## □ DBMS for NoSQL



- 참조 사이트
- <http://www.nosql-database.org/>
- <https://www.mongodb.com/>
- <http://hbase.apache.org/>
- <http://couchdb.apache.org/>

# 5. NoSQL 제품군

## 1. Key-Value Database

- 1) Amazon's Dynamo Paper
- 2) Data Model : Collection of K-V pairs
- 3) 제품유형 : Riak, Voldemort, Tokyo\*

## 3. Document Database

- 1) Lotus Notes
- 2) Data Model : Collection of K-V collection
- 3) 제품유형 : Mongo DB, Couch DB

## 2. BigTable Database

- 1) Google's BigTable paper
- 2) Data Model : Column Families
- 3) 제품유형 : Hbase, Casandra, Hypertable

## 4. Graph Database

- 1) Euler & Graph Theory
- 2) Data Model : nodes, rels, K-V on both
- 3) 제품유형 : AllegroGraph, Sones

**Availability**(유용성), **Consistency**(일관성), **Partitioning**(지속성)에 따른 제품군 구분



# 5. NoSQL 제품군





## □ DBMS Ranking



참조자료: <https://db-engines.com/en/ranking>



Rank			DBMS	Database Model	Score		
Jan 2018	Dec 2017	Jan 2017			Jan 2018	Dec 2017	Jan 2017
1.	1.	1.	Oracle +	Relational DBMS	1341.94	+0.40	-74.78
2.	2.	2.	MySQL +	Relational DBMS	1299.71	-18.36	-66.58
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1148.07	-24.42	-72.89
4.	4.	↑ 5.	PostgreSQL +	Relational DBMS	386.18	+0.75	+55.81
5.	5.	↓ 4.	MongoDB + ★	Document store	330.95	+0.18	-0.96
6.	6.	6.	DB2 +	Relational DBMS	190.28	+0.70	+7.78
7.	7.	↑ 8.	Microsoft Access	Relational DBMS	126.70	+0.82	-0.75
8.	↑ 9.	↓ 7.	Cassandra + ★	Wide column store	123.88	+0.67	-12.57
9.	↓ 8.	9.	Redis + ★	Key-value store	123.14	-0.10	+4.44
10.	10.	↑ 11.	Elasticsearch +	Search engine	122.55	+2.77	+16.38
11.	11.	↓ 10.	SQLite +	Relational DBMS	114.25	-0.94	+1.88
12.	12.	12.	Teradata	Relational DBMS	72.63	-2.11	-1.54
13.	↑ 14.	13.	SAP Adaptive Server +	Relational DBMS	65.46	-0.22	-3.64
14.	↓ 13.	14.	Solr	Search engine	64.37	-1.93	-3.71
15.	15.	↑ 16.	Splunk ★	Search engine	64.00	+0.21	+8.51
16.	16.	↓ 15.	HBase ★	Wide column store	61.64	-1.78	+2.50
17.	17.	↑ 20.	MariaDB +	Relational DBMS	58.30	+1.56	+13.26
18.	↑ 19.	↑ 19.	Hive + ★	Relational DBMS	55.49	+0.81	+4.35
19.	↓ 18.	↓ 17.	FileMaker	Relational DBMS	55.20	+0.00	+1.72
20.	20.	↓ 18.	SAP HANA +	Relational DBMS	46.16	-0.33	-5.77





# 5. NoSQL 제품군

## □ 유형별 Ranking

Rank			DBMS	Database Model	Score		
Aug 2015	Jul 2015	Aug 2014			Aug 2015	Jul 2015	Aug 2014
1.	1.	1.	MongoDB 	Document store	294.65	+7.26	+57.30
2.	2.	2.	CouchDB	Document store	27.28	+0.35	+3.15
3.	3.	3.	Couchbase	Document store	26.16	-0.11	+8.62
4.	4.	4.	Amazon DynamoDB	Multi-model 	18.45	+1.93	+8.34
5.	5.	5.	MarkLogic	Multi-model 	11.71	+0.58	+3.63
6.	6.	6.	RavenDB	Document store	6.28	+0.17	+1.71
7.	↑9.	↑8.	Cloudant	Document store	4.90	+0.74	+3.05
8.	↓7.	↑9.	OrientDB	Multi-model 	4.65	+0.19	+2.87
9.	↓8.	↓7.	GemFire	Document store	4.35	+0.04	+2.27
10.	10.	↑12.	RethinkDB	Document store	2.77	+0.17	+2.14

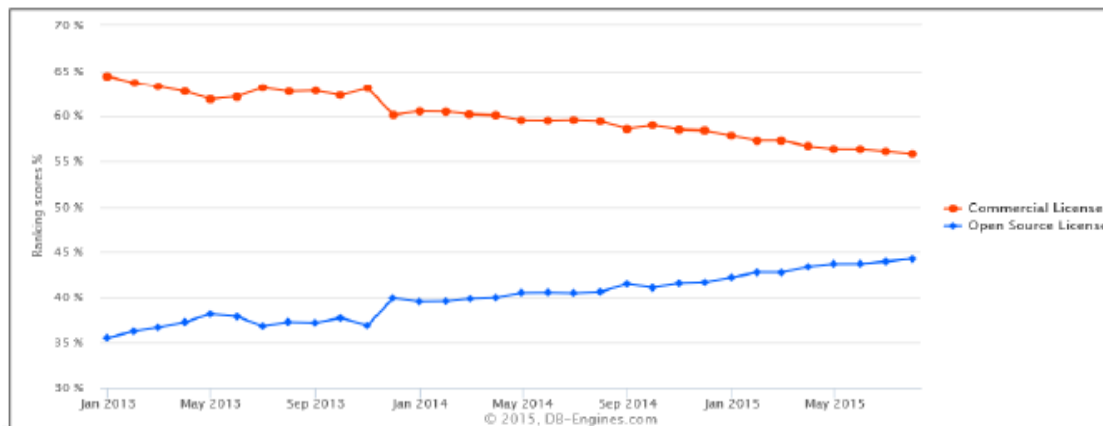
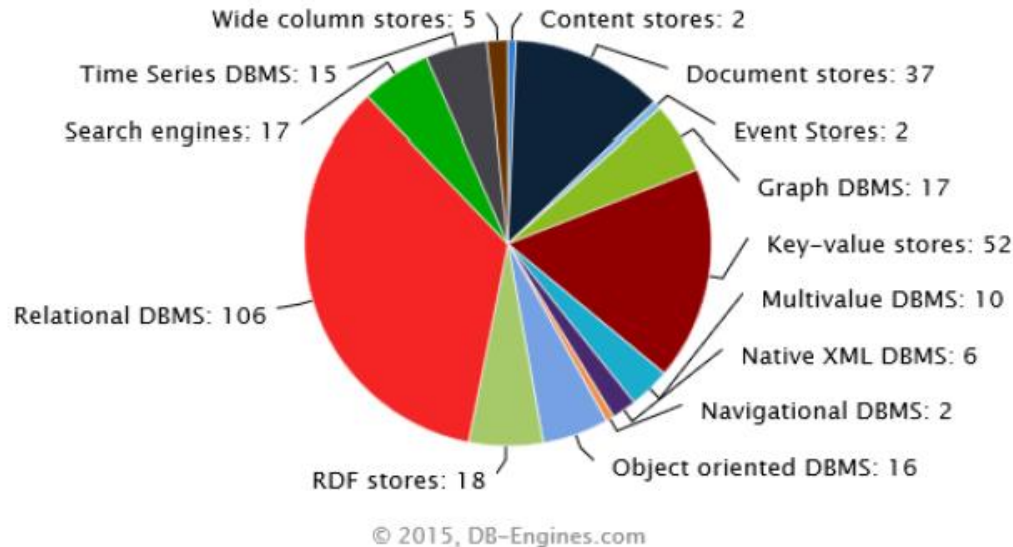
Rank			DBMS	Database Model	Score		
Aug 2015	Jul 2015	Aug 2014			Aug 2015	Jul 2015	Aug 2014
1.	1.	1.	Redis 	Key-value store	98.81	+3.73	+28.01
2.	2.	2.	Memcached	Key-value store	33.38	+0.25	+2.39
3.	3.	↑4.	Amazon DynamoDB	Multi-model 	18.45	+1.93	+8.34
4.	4.	↓3.	Riak 	Key-value store	14.81	+1.16	+3.14
5.	5.	5.	Ehcache	Key-value store	8.07	+0.31	+1.25
6.	6.	6.	Hazelcast	Key-value store	6.27	+0.19	+1.60
7.	7.	↑10.	OrientDB	Multi-model 	4.65	+0.19	+2.87
8.	8.	↓7.	Berkeley DB	Key-value store	4.06	+0.22	+0.96
9.	9.	9.	Oracle Coherence	Key-value store	3.71	+0.21	+1.18
10.	10.	↓8.	Amazon SimpleDB	Key-value store	3.13	+0.14	+0.13

Rank			DBMS	Database Model	Score		
Aug 2015	Jul 2015	Aug 2014			Aug 2015	Jul 2015	Aug 2014
1.	1.	1.	Cassandra 	Wide column store	113.99	+1.28	+32.09
2.	2.	2.	HBase	Wide column store	59.95	-0.97	+18.03
3.	3.	3.	Accumulo	Wide column store	3.73	+0.25	+1.11
4.	4.	4.	Hypertable	Wide column store	0.71	+0.02	+0.05
5.	5.	5.	Sqrrl	Multi-model 	0.41	+0.03	+0.26

Rank			DBMS	Database Model	Score		
Aug 2015	Jul 2015	Aug 2014			Aug 2015	Jul 2015	Aug 2014
1.	1.	1.	Neo4j 	Graph DBMS	33.16	+1.81	+10.25
2.	2.	↑3.	OrientDB	Multi-model 	4.65	+0.19	+2.87
3.	3.	↓2.	Titan	Graph DBMS	4.23	+0.34	+2.17
4.	4.	↑6.	ArangoDB 	Multi-model 	1.39	+0.09	+1.13
5.	5.	5.	Giraph	Graph DBMS	0.98	-0.05	+0.58

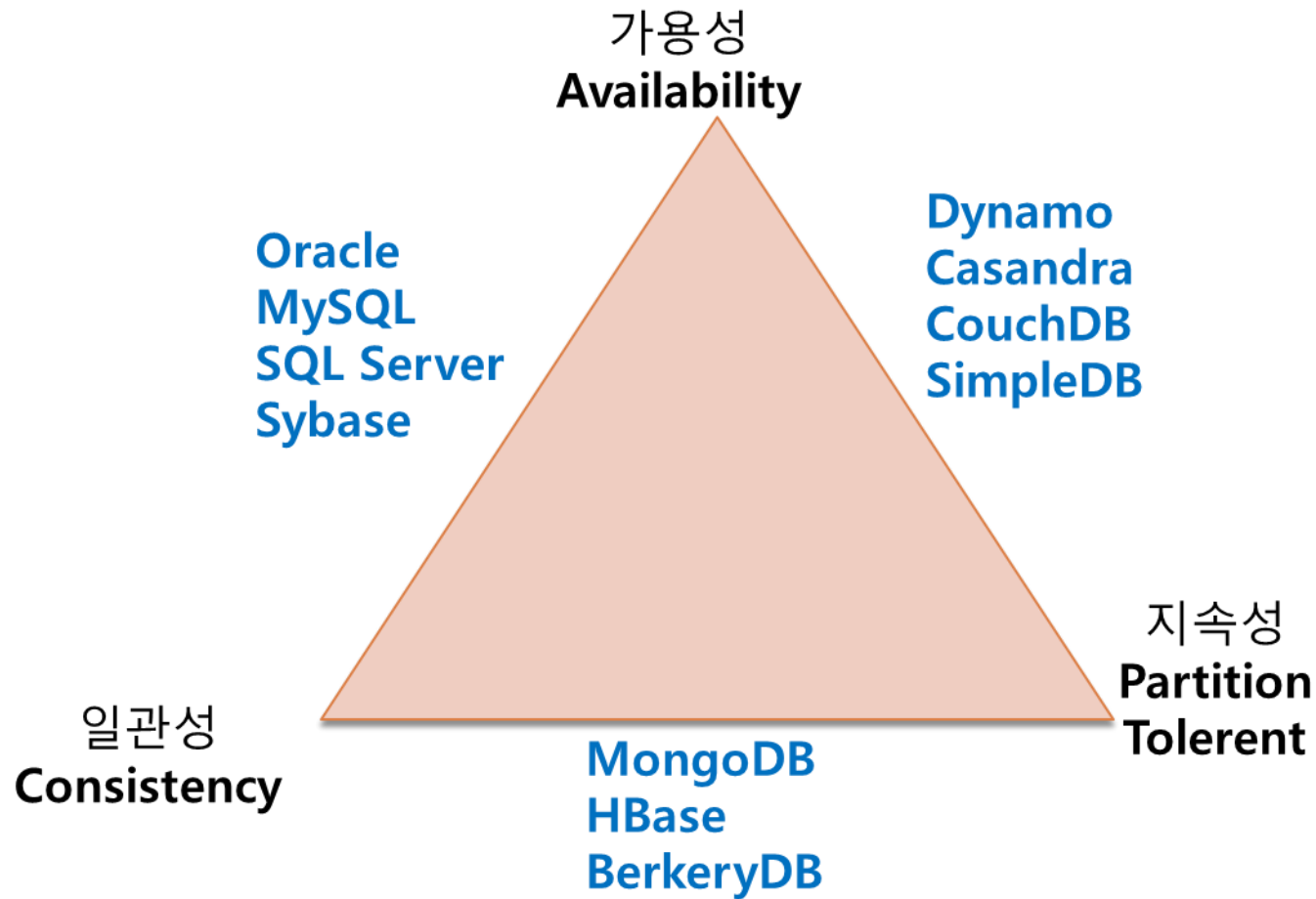
# 5. NoSQL 제품군

## □ 유형별 증가 추이와 상용 & OpenSource별 증가 추이



# 5. NoSQL 제품군

## □ CAP 이론



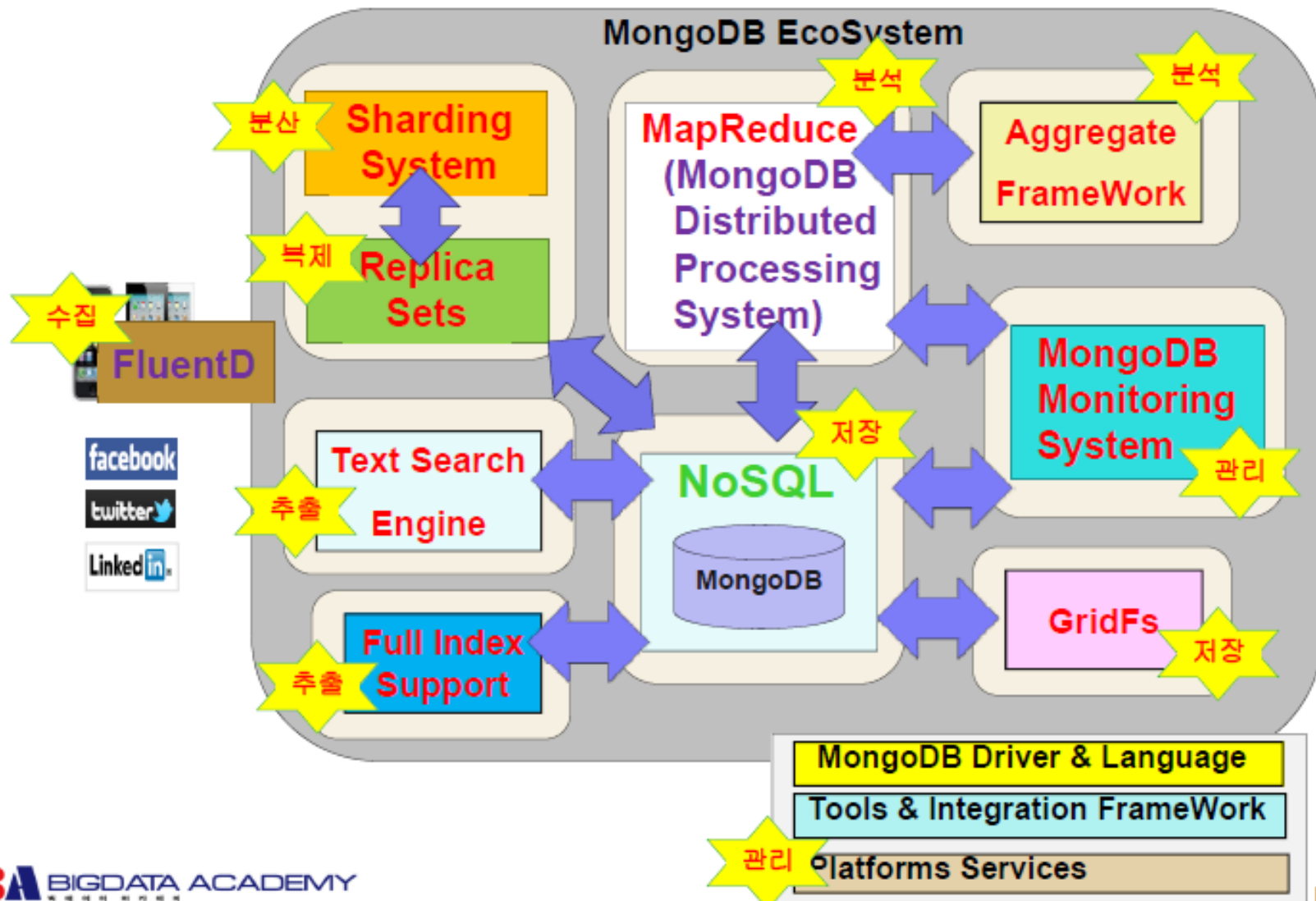


# 5. NoSQL 제품군

## □ NoSQL & RDBMS

	Dynamo	Hbase	MongoDB	RDBMS (Oracle, MySQL 등)
기능	Key/Value 데이터 저장 기술 -Amazon co.	ColumnFamily 데이터 저장기술 -Apache 재단	Document 데이터 저장 기술 -10gen co.	테이블 데이터 저장 -Oracle co, MS, IBM 등
성능	빠른 쓰기/읽기 성능 우수 -In Memory 기술 적용 -모니터링 툴 제공 -Map/Reduce 사용 가능 -Hash Key 및 Secondary 인덱스 기능 제공	빠른 쓰기/읽기 성능 우수 -In Memory 기술 적용 -Hadoop Sub-System 활용가능 -Hadoop Map/Reduce 사용 가능 -Row Key 인덱스 만 제공	빠른 쓰기/읽기 성능 우수 -In Memory 기술 적용 -다양한 INDEX 기능 제공 -Map/Reduce 제공 -다양한 데이터 추출함수 제공 -모니터링 툴 기본 제공	트랜잭션 위주 데이터 처리 우수 -다양한 INDEX 제공 -다양한 함수 제공
확장성	.Scale Out 가능 -다수의 노드 확장 용이	.Scale Out 가능 -다수의 노드 확장 용이	.Scale Out 가능 -다수의 노드 확장 용이	-Scale Up 가능 -Scale Up도 가능하지만 추가 라이선스로 시간/비용증가
안정성	.복제 기능 제공 -다수의 노드로 복제 용이 -빠른 패치 제공	.복제 기능 제공 -다수의 노드로 복제 용이	.복제 기능 제공 -다수의 노드로 복제 용이 -빠른 패치 제공	.오랜 세월 시스템 안정성 확보 -다수의 노드로 복제 가능 -빠른 패치 제공
단점	.데이터 scan 처리량 1mb제한 .사용자 Interface 불편 .트랜잭션 처리가 매우 낮음 (Auto-Commit 만 지원)	.비영리 단체 Apache 재단 운영 .Hadoop System 연계 가능하지만 여러 프로젝트에서 만들어짐 .트랜잭션 처리가 매우 낮음 (Auto-Commit 만 지원) .기술지원 업체가 없음	.트랜잭션 처리가 낮음 (Auto-Commit/Rollback가능)	-Scale Out 가능하지만 고가비용 -대용량 데이터 처리에 적절하지 않는 메모리 구조 -조인으로 인한 성능 지연유발 -분석/설계 시 시간 및 비용 증가 -유지보수 비용 증기 -고 사양의 시스템 요구 -지속성 보장이 안됨

## 6. MongoDB EcoSystem



# 7. 적용사례

## □ Disney Interactive Media Group

- ▣ Mysql-> MongoDB
- ▣ Mysql 바이너리 데이터 저장 한계 및 성능 문제
- ▣ 다양한 Game, Media Data 관리 시스템에 적용
- ▣ ReplicaSets & Auto Sharding 유연성과 확장성 활용

## □ Music Television

- ▣ 비디오/오디오 Content Management System에 적용
- ▣ MySQL->NoSQL로 전환
- ▣ MTV의 계층적 데이터 구조에 적합한 데이터 모델 활용
- ▣ 쉬운 Query와 Index를 이용한 빠른 검색기능 활용

# 7. 적용사례

## □ Business Media Company(Forbes)

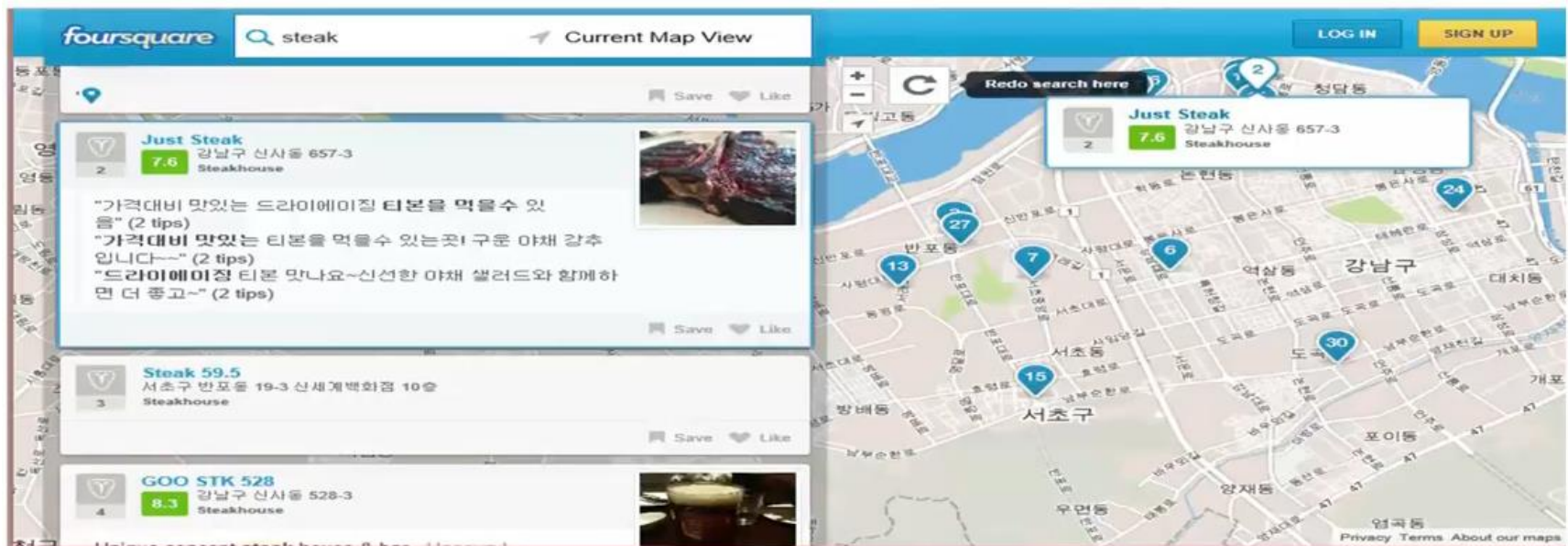
- 원고 자동 수집 및 발행 시스템에 적용
- Oracle 유를 NoSQL로 전환
- 정형적인 Static Data관리에서 Dynamic Data 관리로 전환하면서 발생하는 재 설계 및 구축 비용 절감 목적으로 활용

## □ Shutterfly

- 인터넷기반 사진 정보 및 개인 출판 서비스 사이트
- Oracle DB를 NoSQL로 전환(20TB)
- 100만명의 고객/60억개의 이미지/초당 10,000개 트랜잭션 처리에서 발생하는 구축/관리 비용 및 성능 문제가 이슈

# 7. 적용사례

- Foursquare(<https://ko.foursquare.com/>)
  - 위치기반 Socical Network 사이트
  - RDB기반의 시스템 확장 비용 및 관리 문제가 이슈
  - GeoSpatial Index 기능활용
  - ReplicaSet&Auto Sharding System 활용



## 7. 적용사례

### 적용 사례(국내)

