



CS 스터디(NoSQL)

NoSQL이란?

NoSQL 데이터베이스는 행과 테이블을 사용하는 관계형(SQL) 데이터베이스보다 훨씬 다양한 방식으로 빠르게 바뀌는 대량의 비정형 데이터를 처리할 수 있다는 점을 강조하기 위해 “비관계형”, “NoSQL DB” 또는 “non-SQL”이라고도 한다.

지난 수십년간 SQL을 사용하는 전통적인 관계형 데이터베이스(이하 “RDB”)가 데이터베이스 시장을 점령했다. 과거에는 저장해야 할 데이터의 양이 많지도 않았고, 매출 기록이나 회원 정보 기록 등 상당히 단순하고 쉽게 정형화될 수 있는 정보만을 다루었기 때문에 RDB는 괜찮은 선택이었다. 기존의 정형화된 데이터뿐만 아니라 메신저 텍스트, 음성 등 반정형화, 비정형화된 데이터도 저장하고 다루어야 하는 수요가 생겼다. 또한 ‘클라우드’, 또는 ‘분산형 컴퓨팅’이 주목받기 시작했지만 SQL은 이에 적합하지 않다. 여전히 SQL은 강력하고 훌륭한 정보저장 수단이지만 다양하고 복잡해진 기업 환경과 분산형 컴퓨팅 추세에 따라 SQL의 약점이 드러난 것이다.

NoSQL의 특징

- 반정형화, 비정형화된 데이터에 적합하다.
- "ACID" 대신 "Eventual Consistency"를 허용한다.
 - SQL의 모든 트랜잭션은 ACID한 특성을 유지해야 한다.
 - Atomicity(원자성)
 - Consistency(일관성)
 - Isolation(고립성)
 - Durability(지속성)

NoSQL의 분산형의 특성상 일관성을 유지하기 어렵다. 수많은 다른 머신에 데이터를 분산 저장했는데 한 영역에서 업데이트가 생기면 그것을 실시간으로 전파하는 것은 쉽지 않기 때문.

→ 꼭 실제 최신은 아닐 수 있지만 "업데이트가 되기 전까지는" 가지고 있는 최신의 데이터를 반환한다.

= **Eventual Consistency**

- "BASE"
 - Basically Available
가용성 중시, 데이터는 항상 접근 가능, 다수 스토리지에 복사본 저장
 - Soft State
노드의 상태는 외부에서 전송된 정보를 통해 결정됨
 - Eventually Consistent
- 대용량 데이터 저장에 유리하다.
- 특정 도메인의 문제해결에 뛰어나다.
e.g) SNS에서 인간 관계는 모두 그래프이기 때문에 그래프 데이터베이스를 사용.
- 데이터를 질의하는 API가 다양하다.
- 데이터 모델이 다양하다.

CAP 이론

NoSQL에서 이 세 속성을 모두 가지는 것은 불가능하다.

1. Consistency(일관성)

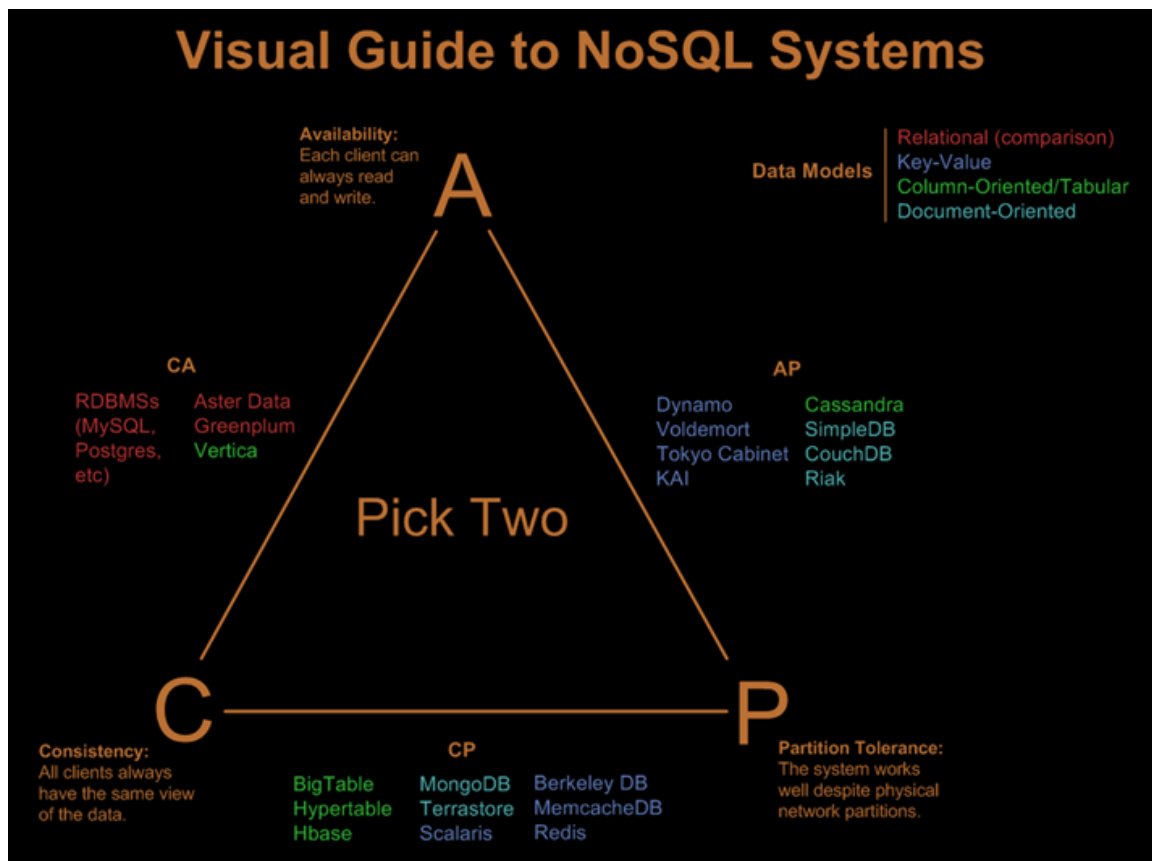
모든 노드가 같은 시간에 같은 데이터를 보여줘야 한다.

2. Availability(가용성)

특정 노드가 장애가 나도 서비스가 가능해야 한다.

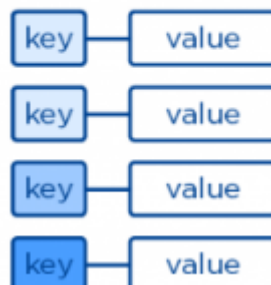
3. Partition Tolerance(분산 허용)

노드간에 통신 문제가 생겨서 메시지를 주고받지 못하는 상황이라도 동작해야 한다.



NoSQL의 유형

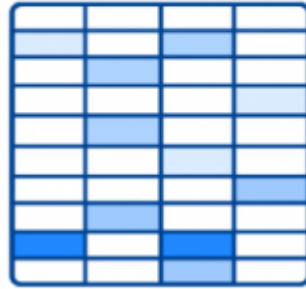
Key-Value



- Key-value store

연관배열 (Associative array)을 데이터 모델로 사용한다.

Wide-column



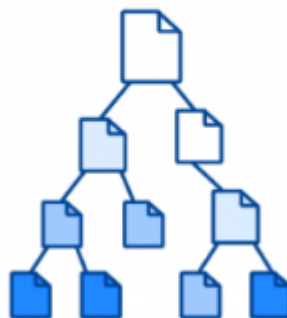
- Column-base store

기존 SQL은 행 단위로 데이터를 저장하는 반면 열 단위로 저장한다.

데이터베이스에서 어느 레코드의 특정 한 부분만을 수정해야 할 때, 기존의 데이터베이스는 필요하지 않은 모든 열을 다 질의한다. 반면 열 기반에서는 필요한 열의 데이터만 로드하면 되기 때문에 행의 수가 같을 경우 필요한 I/O 작업이 줄어든다.

AWS Redshift, Accumulo, Cassandra

Document

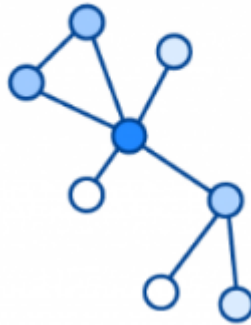


- Document-oriented store

Key/Value Store의 확장된 형태로, value에 Document라는 타입을 저장한다.

Document는 구조화된 문서 데이터(XML, JSON, YAML 등)을 말한다.

Graph



- Graph store

node들과 relationship들로 구성된 개념이다. 역시 Key/Value Store이며 모든 노드는 끊기지 않고 연결되어 있다