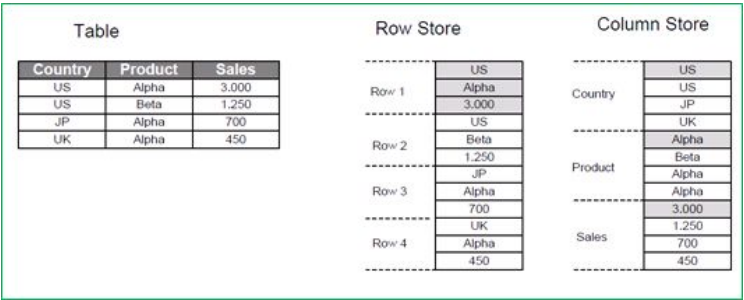
데이터의 종료 -정형/ 반정형/ 비정형

-정형데이터(Structured Data)

**스키마**에 따른 데이터 저장, 자유가 낮음

RDBMS, XML 형태의 데이터를 말함 ex) 주민등록번호



-반정형데이터(Semi-Structured Data)

명확한 스키마가 없음, 어느 정도의 자유도 허용

데이터마다 완전히 동일한 구조를 따를 필요는 없음

Object embedding 형식으로 처리

조인으로 처리해야 할 정보를 중복형식으로 관리

주로 NoSQL, JSON 형태의 데이터를 말함 ex) URL

JSON

[{"id":"123", "name":"Katie", "eyeColor":"brown"},  
 {"id":"234", "name":"Michael", "age":22}, (eyeColor이 없음!)  
 {"id":"345", "name":"Simone", "age":23, "eyeColor":"blue"}]  
 키: 밸류, 배열 [ ] , 오브젝트 { }  
 - 중첩(nesting) 가능  
 - 어떤 형태의 자료구조도 표현가능  
 Schema-less

-비정형데이터(UnStructured Data)

스키마가 없음 / 구조가 사실상 없음

비디오, 오디오 등의 멀티미디어 데이터

일반 텍스트나 웹자료 등을 말함 ex)구글 검색창, 노래감사 검색

- Inverted Index

- 비정형데이터를 처리하는 검색엔진에서 사용하는 인덱스

**- 문서ID -> 텀(Term)**

문서1 -> “하늘” “거리” “고양이”

문서2-> “고양이” “개” “자동차”

문서3-> “바다” “개” “거리”

**- 텀(Term) -> 문서ID**

“하늘” -> 문서1 “

거리” -> 문서1, 문서3

“고양이” -> 문서1, 문서2

“개” -> 문서2, 문서3

“자동차” -> 문서2

“바다” -> 문서3