- 정형데이터(Structured data/formal data)

즉시 통계적 분석에 사용될 수 있을 만한 형태로 정리되고 가공된 데이터

고정된 필드에 저장된 데이터(관계형 데이터베이스, 스프레드시트 등)

정형 데이터는 보통 데이터베이스의 정해진 규칙에 맞게 데이터를 들어간 데이터 중에 수치 만으로 의미 파악이 쉬운 데이터들을 말합니다. 표를 그려 넣고 채워 넣는 형식의 데이터로 이름, 나이, 주민등록번호, 카드번호 등 주로 숫자와 짧은 단어로 구성된 데이터입니다.

- 반정형데이터(semi-structured data)

파일 형태, 메타데이터(데이터 내부에 정형 데이터의 스키마)

반정형 데이터의 반은 Semi를 의미하는 것인데요. 즉, 완전한 정형이 아니라 약한 정형 데이터라는 뜻을 담고 있습니다. 그렇기 때문에, 고정된 양식은 없으나 어느 정도 구조가 정해져 있는 데이터로, 반정형 데이터의 종류로는 로그 데이터, HTML, XML 등이 있습니다.

- 비정형데이터(Unstructured data)

데이터 세트가 아닌 하나의 데이터가 수집 데이터로 객체화

언어 분석이 가능한 텍스트 데이터, 멀티미디어 데이터 - 동영상, 이미지, 텍스트 등

비정형 데이터는 정형 데이터와 반대되는 단어로, 정해진 규칙이 없어서 값의 의미를 쉽게 파악하기 힘든 경우 비정형 데이터로 불립니다. 통제가 힘들거나 불가능한 데이터이기도 하며, 비정형 데이터는 글이나 이미지, 동영상, 음성과 같이 멀티미디어 데이터가 대표적입니다.

최근 이러한 비정형 데이터인 스마트폰과 CCTV, 블랙박스, 드론, 인공위성, 디지털카메라 등에서 수집되는 영상 데이터의 양이 엄청나게 증가했습니다.

그리고 이러한 빅데이터의 85%가량은 형태가 정해지지 않은 비정형 데이터라고 합니다. 또한 최근에는 비정형 데이터의 수가 훨씬 많아지고 있습니다.