**빅데이터의 종류 3가지**

**1. 정형 데이터 (Structured Data)**

고정된 필드에 저장된 데이터로 ex) 관계형 데이터베이스(RDB, Related Database)와 스프레드 시트가 있다. 데이터베이스를 설계한 이에 따라 수집되는 정보의 형태가 정해져 있어, 한정된 데이터( 고객의 정보, 상품, 품목 등)의 분석을 할 수 있다.

**2. 반정형 데이터 (Semi-structured Data)**

 고정된 필드의 데이터는 아니지만 XML, HTML 텍스트 등 메타데이터(Meta Data), 스키마(Schema)를 포함하는 데이터.

웹 문서의 보다 정확하고 의미있는 정보를 추출하기 위해 HTML5에서 시멘틱 태그(Semantic Tag)는 문서의 구조와 영역, 범위를 명확히 하기 때문에 검색시에 보다 정확한 정보를 추출 할 수 있도록 도움을 준다.

**3. 비정형 데이터 (Unstructured Data)**

고정된 필드에 저장되어 있지 않은 데이터를 의미한다. ex) 유튜브, 페이스북, 인스타그램, 트위터, 이미지 파일, 음원 파일, 워드 문서, PDF 문서 등.

구글, 네이버, 인스타그램 등으로 형태가 정해지지 않은, 분석 방향에 따라 특정 내용에 대해 실시간으로 업데이트 되는 정보들을 통해서 더 많은 정보를 수집하고 분석 할 수 있다. ex) 날씨, 도로교통상황, 맛집 등

빅데이터는 데이터 양이 방대하고 비정형 데이터의 비중이 높기 때문에 정확한 정보 추출을 위해서는 인프라 및 분석 기술력 확보가 중요한데, 그러한 기법에는 텍스트 마이닝(Text Mining), 오피니언 마이닝(Opinion Mining), 소셜 네트워크 분석(Social network Analytics), 클러스터 분석(Cluster Analysis)등이 있고, 빅데이터의 엄청난 규모의 데이터를 분석하기 위해 클라우데라(Cloudera), 야후(Yahoo!), 아마존(Amazon), 구글(Google) 같은 대기업들이 기술을 개발하고 오픈소스화를 진두지휘 하고 있다.