빅 데이터의 3대 요소 (3V)

 크기(Volume)

비즈니스 특성에 따라 다를 수 있지만 일반 적으로 수십 테라바이트 혹은 수십 페타바이트 이상이 빅데이터에 해당합니다.

이러한 데이터는 기존 파일 시스템에 저장하기 어려울뿐더러 데이터 분석을 사용하는 기존 데이터웨어하우스 같은 솔루션에서 소화하기 어려울 정도로 데이터 양이 많습니다.

속도(Velocity)

매 순간 많은 양의 데이터가 생산되고 있기 때문에 데이터의 생산, 저장, 유통, 수집, 분석이 실시간으로 처리 되어야 합니다. 예를 들어, 게임의 채팅창에 불건전한 내용을 입력할 경우 시스템에서 이러한 문구를 빠르게 분석해 다른 사용자에게 피해가 없도록 해야 합니다.

다양성(Variety)

다양한 종류의 데이터들이 빅데이터를 구성하고 있다. 데이터 정형화의 종류에 따라 정형, 반정형, 비정형으로 나눌 수 있습니다. 정형 데이터는 문자 그대로의 데이터로 고정된 필드에 저장되는 데이터를 의미한다. 반 정형 데이터는 XML이나 HTML 같이 메타데이터나 스키마 등의 데이터를 의미합니다. 비정현 데이터는 동영상이나 사진 등 고정된 필드에 저장되 있지 않은 데이터를 의미합니다.

빅데이터는 위의 3가지 요소중 2가지 요소가 충족한다면 빅데이터라고 볼수 있습니다.