

1 빅데이터의 공통적 속성 3V

빅데이터의 공통적 특징은 3V로 설명할 수 있다.

3V는 데이터의 크기(Volume), 데이터의 속도(Velocity), 데이터의 다양성(variety)을 나타내며 이러한 세 가지 요소의 측면에서 빅데이터는 기존의 데이터베이스와 차별화된다.

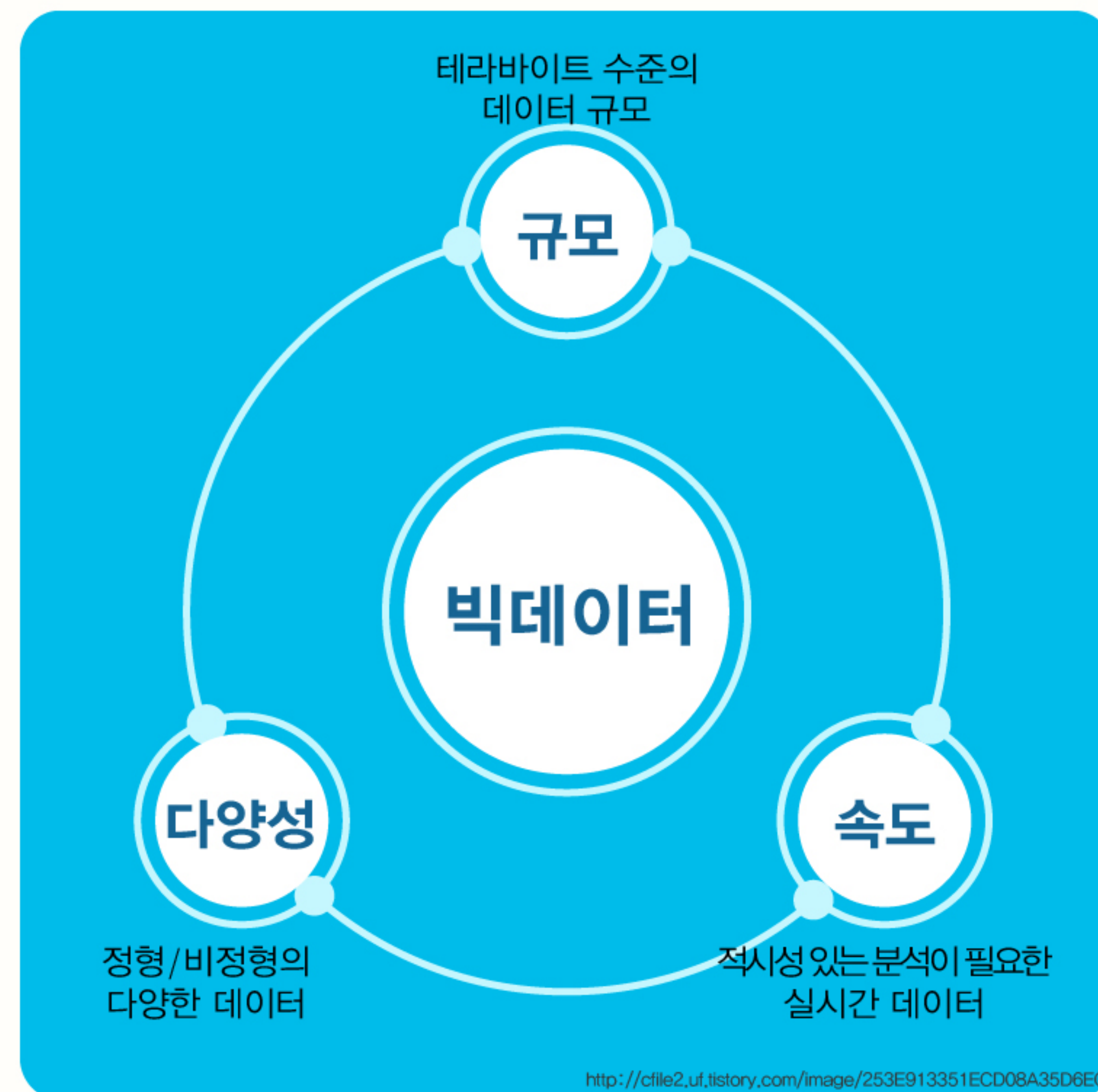
데이터 크기(Volume)는 단순 저장되는 물리적 데이터양을 나타내며 빅데이터의 가장 기본적인 특징이다.

데이터 속도(Velocity)는 데이터의 고도화된 실시간 처리를 뜻한다.

이는 데이터가 생성되고, 저장되며, 시각화되는 과정이 얼마나 빠르게 이뤄져야 하는지에 대한 중요성을 나타낸다.

다양성(Variety)은 다양한 형태의 데이터를 포함하는 것을 뜻한다.

정형 데이터뿐만 아니라 사진, 오디오, 비디오, 소셜 미디어 데이터, 로그 파일 등과 같은 비정형 데이터도 포함된다.



< 그림 1. 빅데이터의 3V >

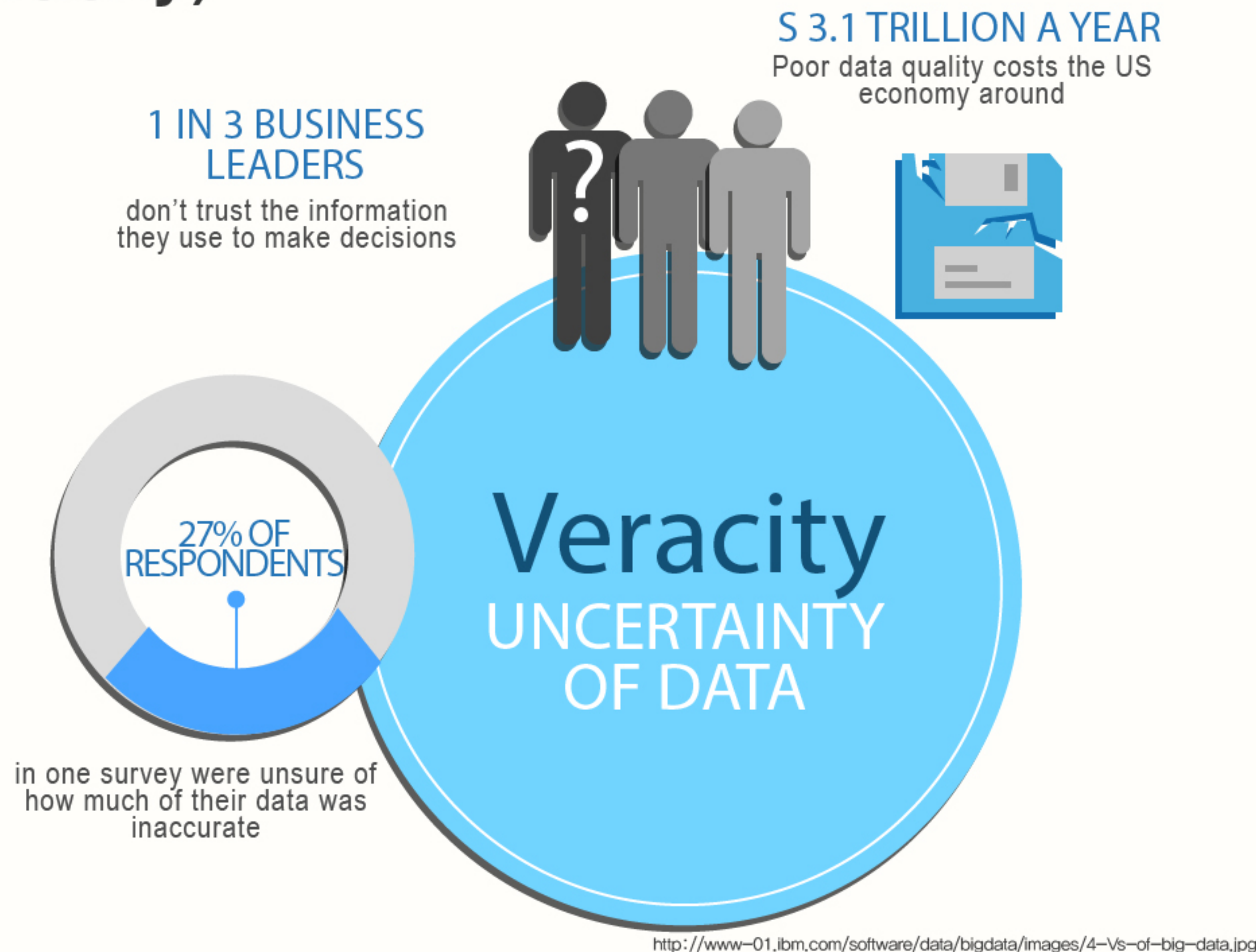
2 빅데이터의 새로운 V - 정확성(Veracity)

빅데이터 시대에는 방대한 데이터의 양을 분석하여 일정한 패턴을 추출할 수 있다.

그러나 과연 데이터 일정 패턴을 설명할 수 있을 만큼 신뢰성이 있느냐는 문제가 생긴다.

데이터가 많아질수록 엉터리 데이터도 커질 가능성이 높아지기 때문이다.

따라서 빅데이터를 분석하는 데 있어 기업이나 기관에 수집한 데이터가 정확한 것인지, 분석할 만한 가치가 있는지 등을 살펴야 하는 필요성이 생겼고 이러한 측면에서 빅데이터의 새로운 속성인 정확성(Veracity)이 제시되고 있다.



< 그림 2. 빅데이터의 정확성 >

3 빅데이터의 새로운 V-가변성(Variability)



< 그림 3. 빅데이터의 가변성 >

최근 소셜미디어의 확산으로
자기 의견을 웹을 통해 자유롭게 게시하는 것이
쉬워졌지만 실제로 자신의 의도와는 달리
자기 생각을 글로 표현하게 되면 맥락에 따라
자신의 의도가 다른 사람에게 오해를 불러일으킬 수 있다.
이처럼 데이터가
맥락에 따라 의미가 달라진다고 하여
빅데이터의 새로운 속성으로
가변성(Variability)이 제시되고 있다.

4 빅데이터의 새로운 V-시각화(Visualization)

빅데이터는 정형 및 비정형 데이터를 수집하여 복잡한 분석을 실행한 후 용도에 맞게 정보를 가공하는 과정을 거친다.

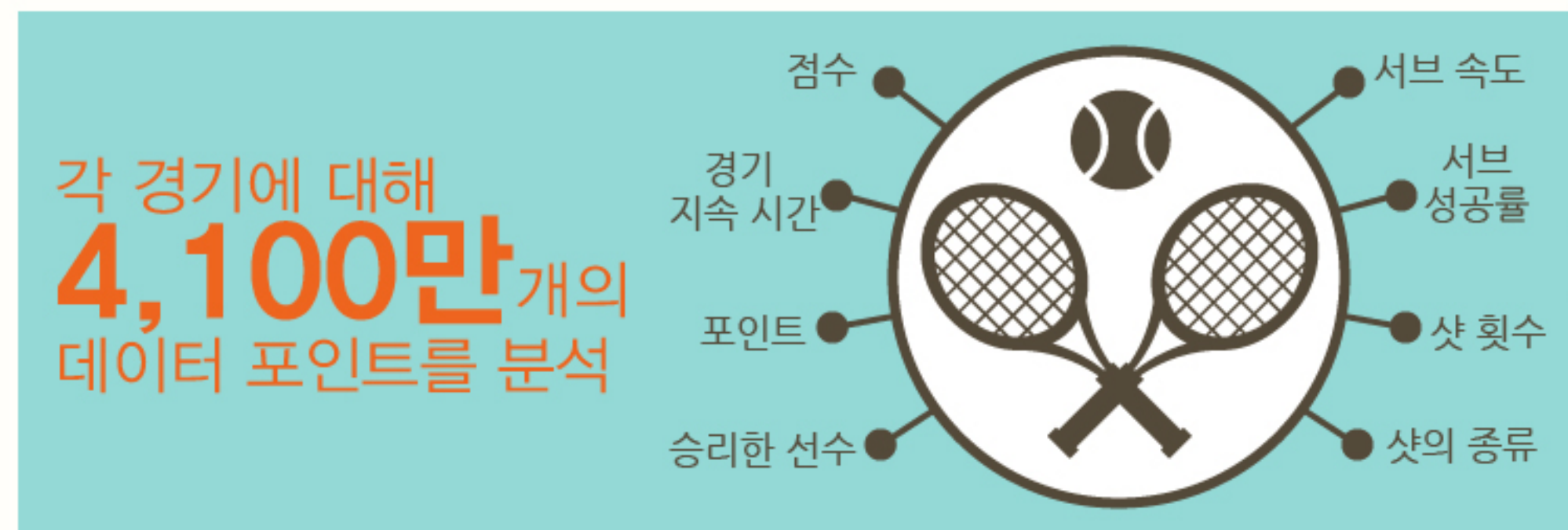
이때 중요한 것은 정보의 사용 대상자가 쉽게 이해할 수 있어야 한다.

그렇지 않으면 정보의 가공을 위해 소모된

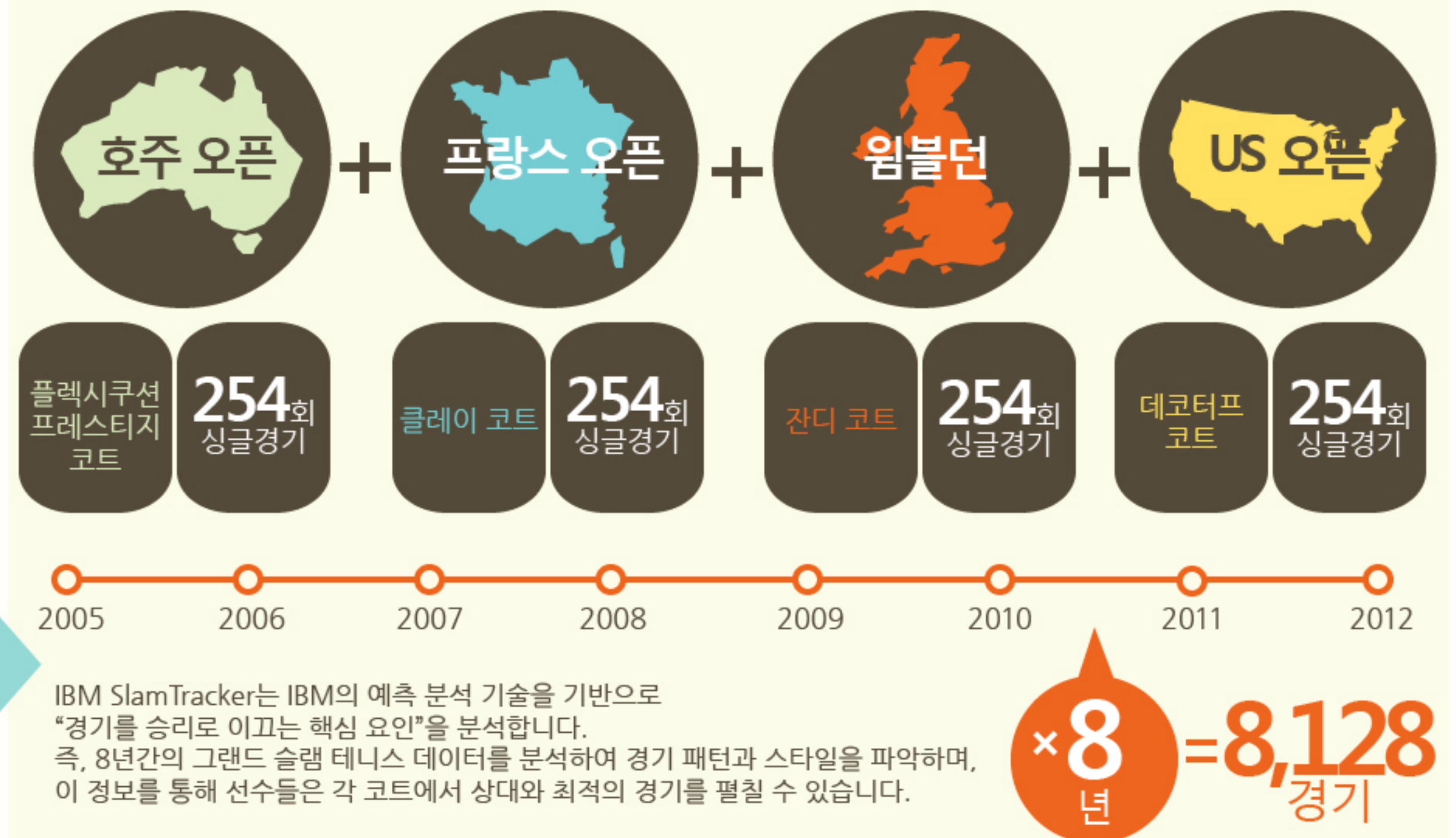
시간적, 경제적 비용이 무용지물이 될 수 있기 때문이다.

이러한 필요성으로 인해

빅데이터의 새로운 속성으로 시각화(Visualization)가 제시되고 있다.



빅데이터와 예측 분석을 이용한 스포츠 중계 [수치로 알아보는 그랜드 슬램]



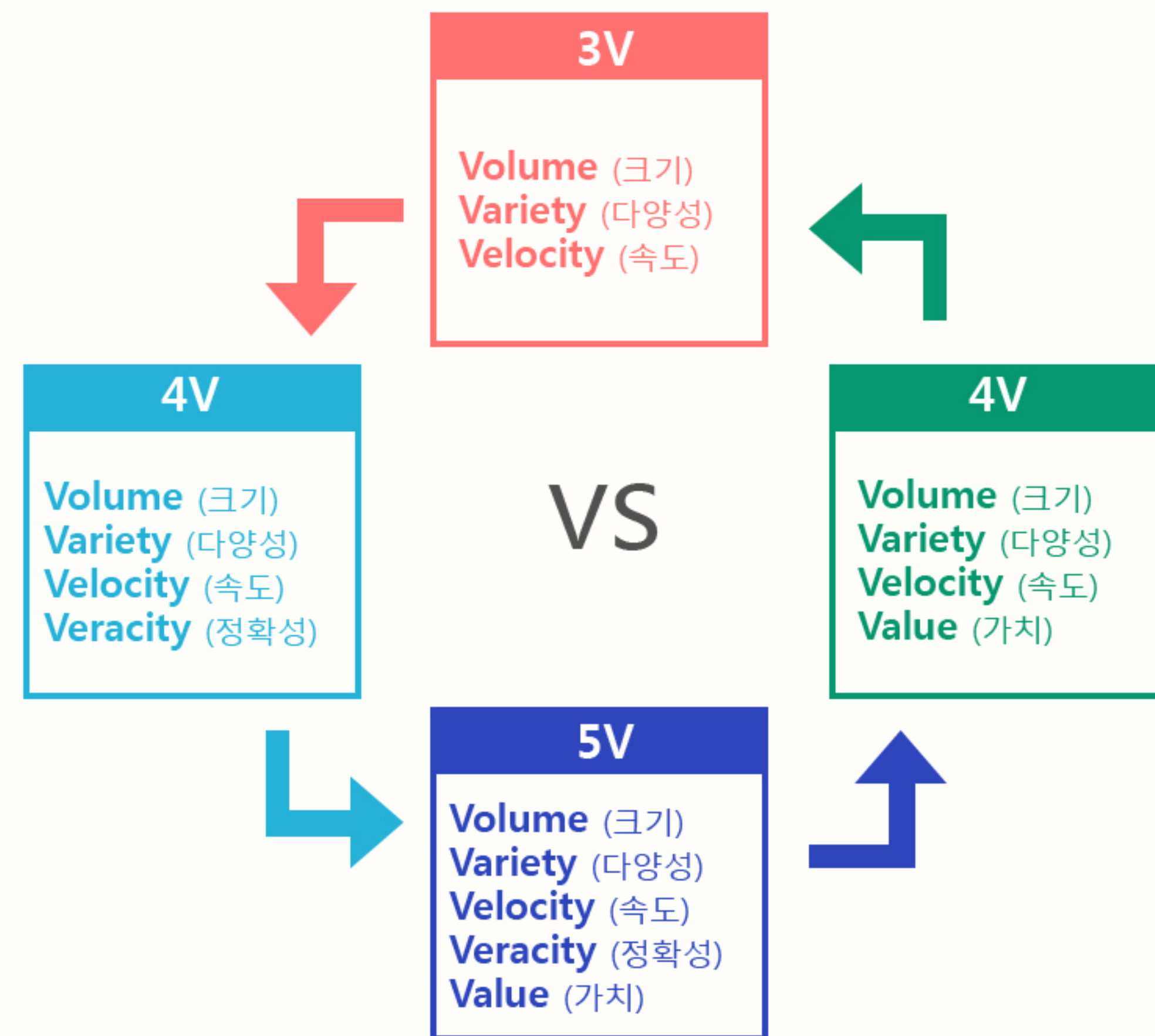
< 그림 4. 빅데이터의 시각화 >

5 아직 제대로 정의되지 않은 빅데이터

살펴본 바와 같이 빅데이터는 어떠한 한 분야를 나타내거나 한 가지 특징을 개념으로 사용하기 어려운 측면이 있다. 분석과 활용 과정에 필요한 사항을 설명할 수는 있겠지만 사용 분야마다 중요하게 생각하는 부분은 다를 수 있고, 필요 없는 것도 있을 수 있다.

또한, 3V, 4V, 5V 등 시간이 지나면서 기존 빅데이터의 특징인 3V에 새로운 속성들이 추가되고 있는 추세이다.

이러한 점이 빅데이터의 정의를 어렵게 만들고 있다. 앞으로도 빅데이터의 특징은 새롭게 추가될 것으로 보인다. 그러나 빅데이터의 가장 기본적인 속성인 3V는 변하지 않고 빅데이터의 대표적인 특징으로 남을 것이다.



<http://cf1e29.uf.tistory.com/image/275B1B3752899A892D8C8F>
< 그림 5. 빅데이터의 다양한 V >