

[빅데이터 개념I – 빅데이터 처리 시스템]

1. 빅데이터 시스템이란?

: 대규모 양의 데이터를 수집, 관리, 유통 분석 처리하는 일련의 **분산 병렬 처리** 프레임워크

2. 빅데이터의 특성(3V)

- (1) 데이터 볼륨 증가(Volume)
- (2) 데이터 발생 속도 증가(Velocity)
- (3) 데이터 포맷 다양성 증가(Variety)

- 기존 데이터웨어 하우스: 텍스트 위주의 정형화 데이터

- 최근 데이터: 반정형 및 비정형 데이터 (그림, 동영상, 음성, 로그, 센서 데이터 등)

=> 다양한 데이터들이 생성됨에 따라 기존 관계형 데이터베이스로 처리하기에는 어려워짐

3. 빅데이터의 유형

유형	특징	종류	예시
정형 데이터	정형화된 스키마를 가진 데이터	RDB, File	재고 관리, 인사 정보.....
반정형 데이터	메타 구조를 가진 데이터	HTML, XML, JSON, 웹로그, 센서 데이터	마케팅 정보, 보안 및 특성 센서.....
비정형 데이터	이미지나 동영상으로 존재하는 데이터	이진파일, 동영상, 이미지, 텍스트	기상 데이터, SNS 데이터

4. 빅데이터 처리 시스템의 목표

- (1) **결함 허용 시스템(Fault Tolerance)**
: 장애가 발생해도 대체 시스템이나 고장 대응체계를 통해 시스템이 계속 운영되도록 한다.
- (2) **저비용 시스템(Cost Effective System)**
: 작업에 적합한 시스템을 적절하게 선택하여 비용을 절감한다.
- (3) **기존 시스템과 연계**
- (4) **다양한 도구 지원**