

[빅데이터 개념II – 빅데이터 처리 인프라 및 s/w]

1. 빅데이터 프로세스: 처리 -> 분석 -> 시각화

2. 빅데이터 처리 인프라

(1) 클라우드 컴퓨팅 사용 방식

- 인터넷 상의 서버를 통해 IT관련 서비스를 사용할 수 있는 컴퓨팅 환경
- 초기 투자비용 적게 발생
- 시스템에 대한 구축 오버헤드 절감
- 향후 시스템 확장을 위한 확장성 확보 가능

(2) 개별적인 시스템 구축 방식

- 대표적인 하둡 배포판: 아파치, 맵알, 클라우데라, 호튼웍스

3. 대표적인 빅데이터 처리 s/w

하둡(Hadoop)	<ul style="list-style-type: none"> - 오픈 소스 s/w - 하둡 분산 파일 시스템(HDFS) + MapReduce / YARN - 빅데이터 처리 프레임워크 - 다양한 하둡 에코 시스템으로 구성 - 결함 허용 - 데이터 블록의 복사본 중복 저장, 유지
스파크(Spark)	<ul style="list-style-type: none"> - 인-메모리 방식의 분산 처리 시스템 - 메모리 사용으로 반복작업이나 스트리밍 데이터를 효율적으로 처리 - 배치, 스트리밍 처리, SQL쿼리 수행 기능, 머신러닝 라이브러리 - Scala Shell을 제공하여 사용자와 대화형으로 데이터 관리