

Big Data

박정웅

이 문서는 나눔글꼴로 작성되었습니다. [설치하기](#)

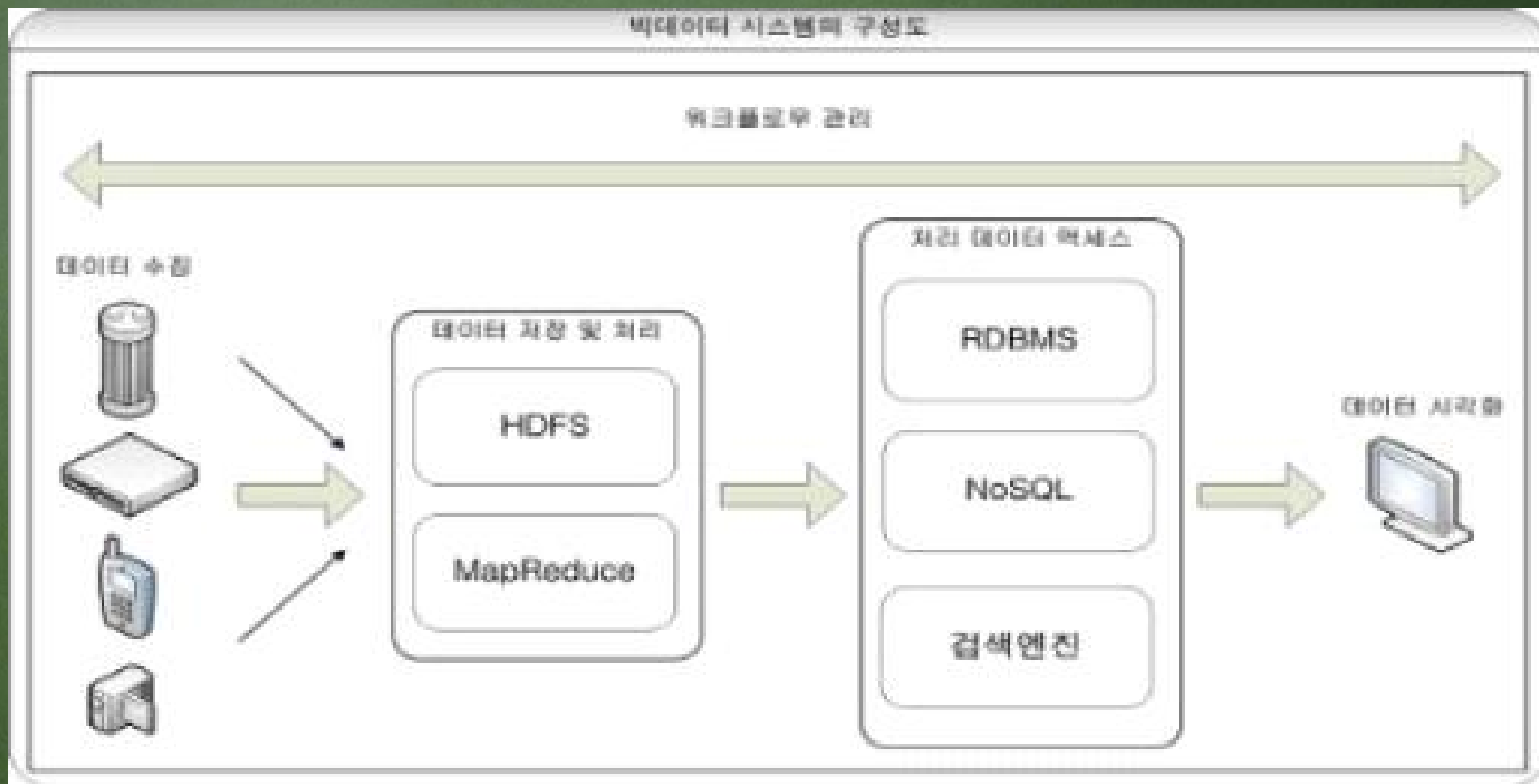
정의

- 서버 한 대로 처리할 수 없는 규모의 데이터
- 기존의 소프트웨어로는 처리할 수 없는 규모의 데이터
- 3V(Value, Velocity, Variety)
 - 주로 컨설팅회사들이 많이 사용하는 정의
 - 가. Value(규모) : 데이터의 크기가 대용량인가
 - 나. Velocity(속도) : 데이터가 얼마나 빠르게 생성되는가
 - 다. Variety(다양성) : 데이터가 구조화/비구조화된 데이터를 다 포함하는지 여부

빅데이터의 예

- 웹 검색엔진 데이터
 - 웹페이지 데이터
 - 검색어 로그와 클릭 로그 데이터
- 디바이스에서 생성되는 데이터
 - ex) 스마트폰, 스마트 tv, 보잉 제트기(매 30분마다 10TB 데이터 생성 - 예전에는 바로 없어졌지만 지금은 저장 및 처리가 가능), 스마트 미터(전기사용량 측정 - 한 달에 한 번 측정하던 것을 시간대별이나 상황별로 처리가 가능), 풍량, 강우량 등등
- 소셜미디어의 데이터
 - ex) 페이스북, 트위터, 링크드인, 포스퀘어 등등

빅데이터 시스템의 구성



빅데이터 시스템의 구성

- 데이터 수집 모듈
 - Flume
 - Chukwa
 - kafka
- 데이터 저장/처리 모듈
 - Hadoop
- 처리데이터 액세스 모듈
 - 기존 RDBMS
 - NoSQL
 - 검색엔진
 - 작업 워크플로우 관리 정의 모듈 (Cascading, Oozie, Azkaban, Ambrose ..)
- 데이터 시각화 모듈 : Matlab, R ...

대표적 빅데이터 성공 모델

- 넷플릭스 영화추천 서비스
- 이베이 쿼리로그 마이닝
- 트위터 대용량 머신러닝 시스템
- 페이스북 메시지 시스템 :
하둡은 아니지만 HDFS기반 NoSQL인 HBase
위에서 동작함

빅데이터 시스템을 도입할 때 고려해야할 문제들

- ROI(Return On Investment) 고려 - 투자자본수익률
- 개인 정보의 노출
- 오픈 소스로 구성된 시스템



이 문서는 나눔글꼴로 작성되었습니다. [설치하기](#)