[빅데이터이론] 2019.03.04(월)

빅데이터

* 빅 데이터
  + 대량의 데이터가 매 순간 새롭게 생성되고 있음
    - 데이터 양(volume), 다양한 형태(variety), 빠른 생성 속도(velocity), 가 치(value) 라는 뜻에서 4v라고 정의하기도 함.
  + 가치 있는 정보를 추출하는 것이 권력이나 부와 직결되고 있음
  + 빅 데이터는 데이터의 생성 양·주기·형식 등이 기존 데이터에 비해 매우 크기 때문에 기존의 방법으로 수집·저장·검색·분석이 어려움
* 요약하자면, 빅데이터는 데이터가 매우 크고 형태가 다양함. 따라서 기존의 방법으로는 수집하기 힘들다.
* 데이터 종류
  + 소셜 네트워크 서비스(SNS)의 발달로 과거 단순 텍스트 기반의 정형화된 데이터 이외에도, 위치정보, 이미지, 음성, 영상 등 다양한 비정형 데이터가 생성되고 있음

빅 데이터 수집(데이터 수집) - 아래 수집 방법 말고도 다른 방법들도 많음.

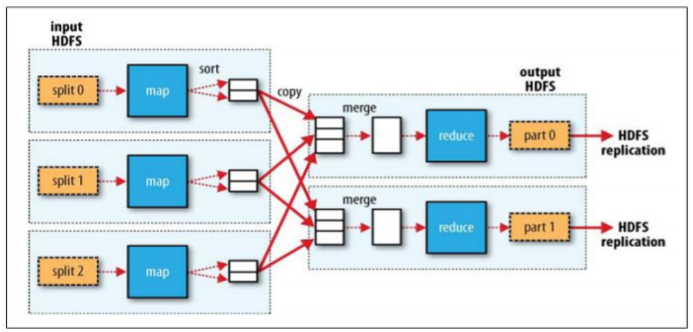
* CCTV, 적외선 카메라 등의 각종 센서를 통해 수집
* 소셜 네트워크를 통해 수집 / 모바일 기기를 통해 수집

빅 데이터 분석

* 다양한 것을 요구하지만 이중에서 가장 중요한 것은 성향을 분석하는 것임. 유비쿼터스와 클라우드
* 유비쿼터스: 언제 어디에나 존재한다는 뜻의 라틴어
  + 사용자가 컴퓨터나 네트워크를 의식하지 않고 장소와 시간에 상관없이 자유롭게 네트워크에 접속할 수 있는 환경
* 클라우드 🡪 유비쿼터스 환경을 구현하기 위해 클라우드 개념 등장
  + 하드웨어나 소프트웨어와 같은 컴퓨터 자산을 인터넷으로 연결된 다른 컴퓨터를 이용하여 처리하는 기술

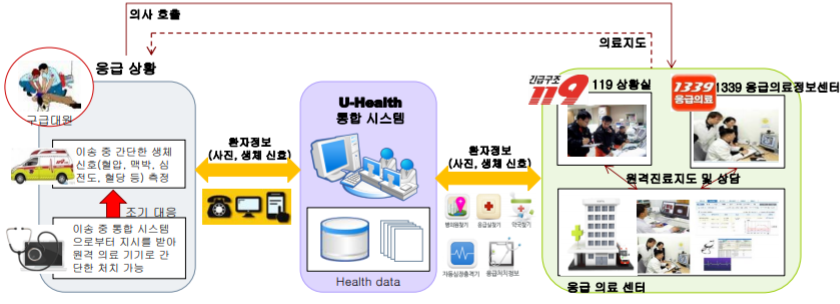
클라우드

* 클라우드 핵심기술
  + 가상화
  + 분산처리
  + 클라우드 서비스 기업: 애플, 구글, 아마존
* 클라우드 분산 처리 기술 예
  + 하둡(Hadoop): 대량의 자료를 처리할 수 있는 분산 응용프로그램을 지원하는 오픈소스 프레임워크



유헬스(u-health)

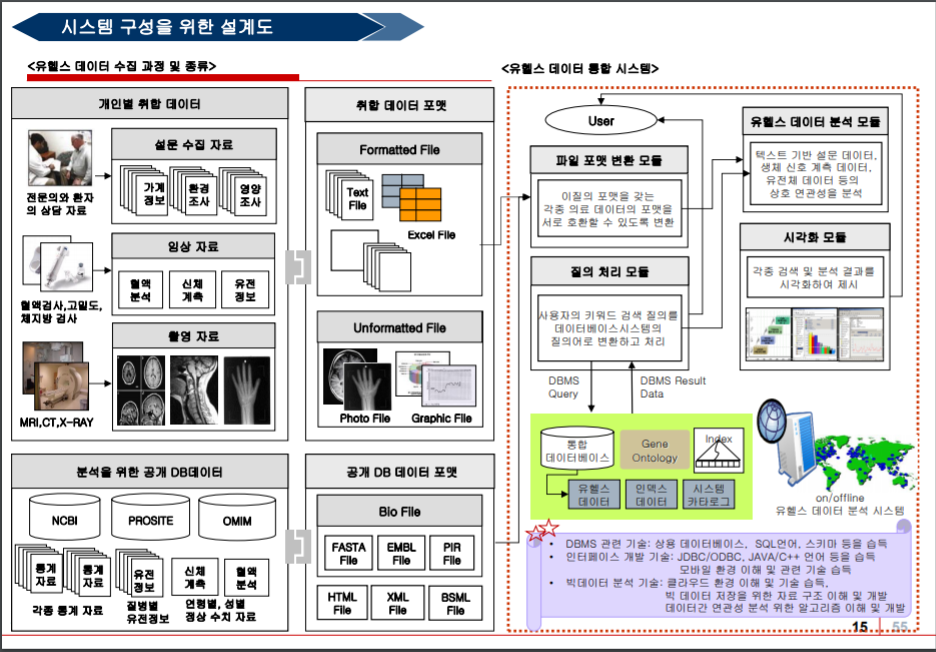
* 유헬스(ubiquitous health; u-health)
  + 정보통신과 의료산업이 융합하여 시간적, 공간적 제약 없이 환자를 진료하는 원격 진료 시스템
  + 의료 접근성이 떨어지는 취약 계층에 다양한 혜택 제공 가능
  + 의료 관광 분야의 활성화에 기여하여 새로운 경제 성장의 동력으로 주목 받고 있음.



* 유헬스 시장의 발달
  + 전세계적으로 고령화 사회에 진입
  + 개별 맞춤형 의료 서비스 시장에 대한 요구 급증
  + 유베쿼터스/클라우드 관련 시장의 급성장

유헬스 데이터

* 데이터 수집 및 종류
  + 의사와의 상담을 통해 수집
  + 의료 장비를 통해 수집
* 유헬스 데이터의 특징
  + 각종 의료 장비를 통해 얻어지는 데이터로 데이터 형태가 다양함.
    - 서로 다른 포맷을 가지는 이질 데이터를 통합 분석하기 위한 기술 필요  
      (텍스트 데이터, 이미지 데이터, 시계열 데이터 등)
  + 오랜 시간 동안 축적되는 데이터로 대용량 데이터임
    - 대용량 데이터를 분산 저장/관리하기 위한 기술 필요
  + 질병의 예측/진단을 위해서는 데이터 간의 연관성을 파악해야 함.
    - 분산 데이터를 통합/연계/분석하기 위한 기술 필요



데이터베이스 시스템(DBMS)

* 데이터베이스: 특정 조직에 필요한 정보를 얻기 위해 논리적으로 연관된 데이터를 모아 놓은 것
* 데이터 베이스 시스템
  + 데이터를 통합 관리하는 시스템
  + 데이터들과 이를 액세스하기 위한 프로그램들의 집합

데이터베이스 응용분야

* 응용 분야: 대학, 은행, 병원, 항공사 등 거의 모든 곳에서 사용된다고 보면 됨.
* 일상 생활 전반에서 데이터베이스가 구축/활용되고 있음
* 데이터베이스 시스템의 주요 작업
  + 데이터의 검색과 변경 작업을 주로 수행
  + 데이터베이스는 크기가 매우 크기 때문에 검색과 변경의 빈도에 따라 시스템 구축의 난의도가 결정됨.



* + 데이터베이스는 데이터의 양이 많고, 실시간 검색이 많은 분야일수록 활용도가 높음.
* 유헬스 데이터베이스 특징: 유형 4에 속함
  + 오랜 시간 축적되는 대용량 데이터이고, 데이터 간의 상호 연관성이 있음.
  + 일반 텍스트 데이터 뿐만 아니라 이미지 데이터, 동영상 데이터 등 형태가 다양함.
  + 사용자 수가 많고, 동시에 데이터베이스에 접근함.
  + 검색도 많고, 데이터 변경도 많음
  + 조기 대응 및 진료를 위한 실시간 검색 및 변경이 중요
* 유헬스 데이터베이스 시스템 구축은 매우 복잡하고, 어려운 작업임.

교수님이 언급하신 키워드 및 용어들

* 지형 밀착형 서비스: 특정 지역에 서비스를 발행하는 것을 말함.
* 분산처리
* 데이터의 시각화가 중요함.
* 분석한 결과를 어디에 활용할 것인지 찾아보는 것이 매우 중요한 쟁점!
* 이미지 등의 시각화 데이터, 시계열 데이터 🡪 비정형 데이터
* 데이터를 가공하여 나온 것이 정보
* 데이터베이스 응용분야 🡪 거의 모든 곳에서 사용된다고 보면 됨.
* 데이터의 성격 파악하기!!