

01 빅데이터란 무엇인가?



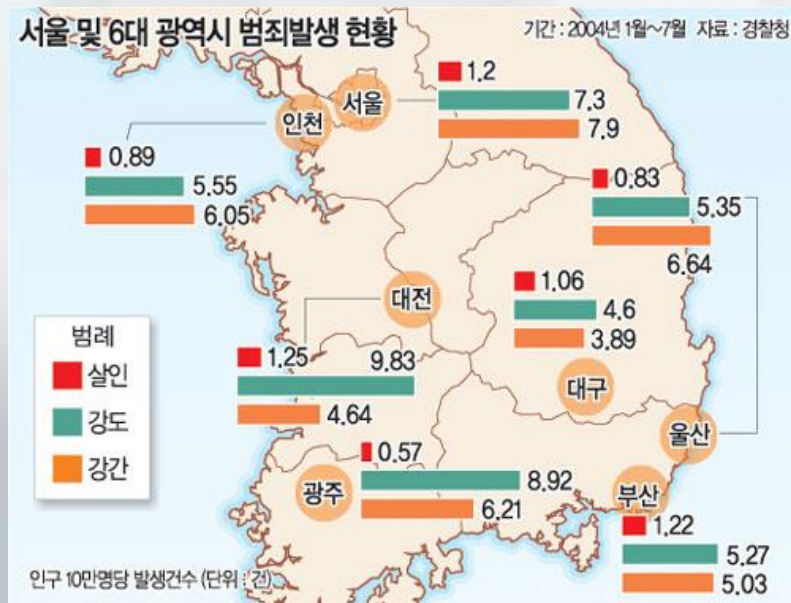
데이터 유형의 역사



- 서버이 데이터 수준 또는 간단한 정형 데이터
- 일 또는 월 단위 발생하는 데이터
- 300~10만 건의 Record로 구성
- 샘플링 된 데이터로 통계적 지식 및 처리 필요
- 국내에서는 1999년 이전에 주로 활용



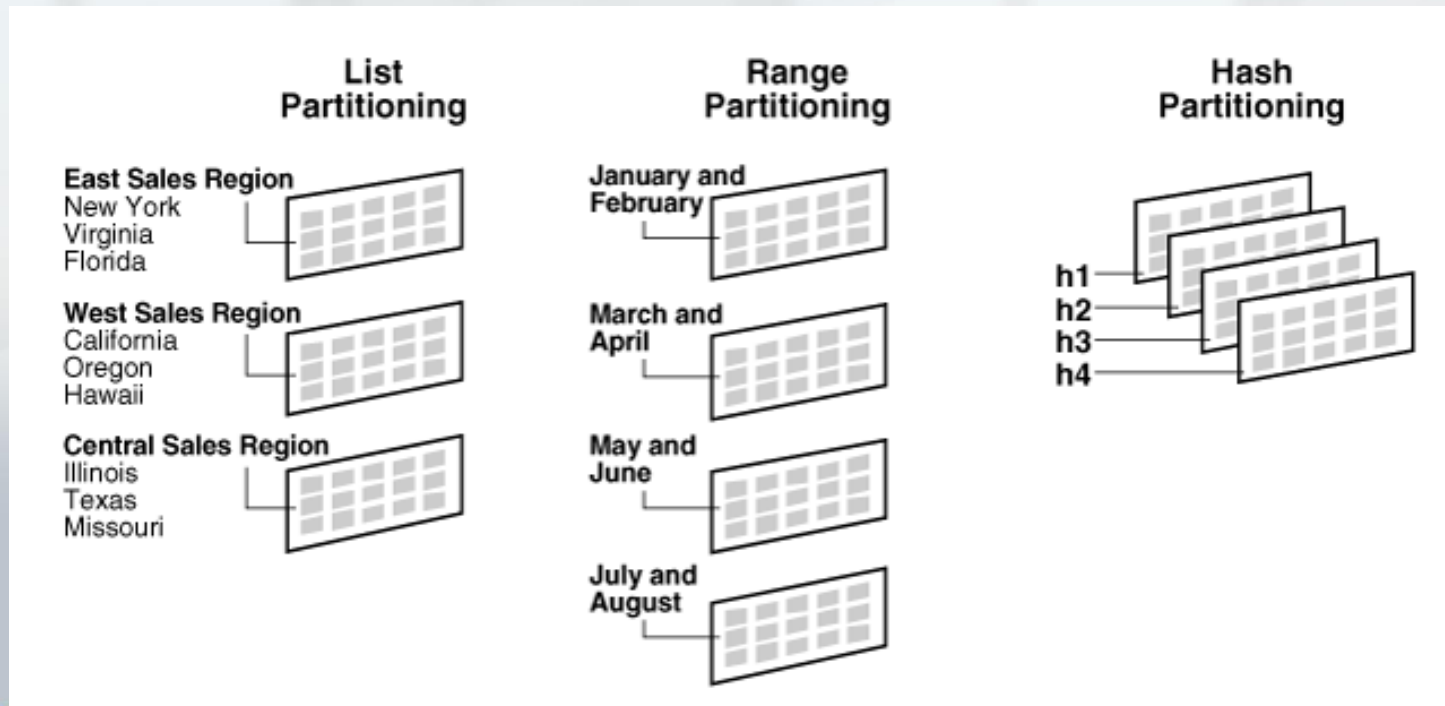
- 데이터에 대한 **가설**이 있음(예 : 정규성, 독립성 등)
- 통계분석 전문 소프트웨어, Excel 등으로 처리가능
- 활용 예 : “지역별 범죄유형 비율 및 차이”



- CRM(고객관계관리) 등에서 고객 별 수년간의 거래정보와 유사함
- 300~1500만 고객이 하루에도 여러 건 정형 데이터 발생, 1~3년간의 거래정보 다룸
- 예측(Prediction)에 주로 활용
- 국내에서는 주로 1999~2012년에 주로 활용



- DBMS에서 SQL 활용, OLAP 도구 사용, 데이터 마이닝 기법 적용
- Excel 처리 불가능



[VLDB를 위한 Oracle 파티셔닝]



3 Big Data

- IT 발전으로 보다 규모가 큰 **TB 이상의 규모**,
1천만~수십억 건의 **3~10년 정형 데이터** 활용

정형
데이터



비정형
데이터



Log성 데이터, 텍스트, 사운드, 이미지 등



3 Big Data

- 보다 정밀하고, 다양한 분석을 하고자 함
- 분석에서 **서비스로의 확장**에 관심이 큼
- 국내에서는 **2013~현재**까지 각광 받고 있는 분야



01 빅데이터란 무엇인가?

2 빅데이터의 특징





Volume

Variety

Velocity

Value

“ 조건이 아닌 특징이며, **절대적인 기준이 아님** ”



1 빅데이터의 특징 4V

1 Volume

- 정형 데이터와 비정형 데이터 다루다 보니 데이터 크기(Volume)가 **상대적으로 이전보다 커짐**
- 상대적 요소는 기업의 처리 데이터에 상대적인 것으로 차이가 큼
- 보다 빠른 처리능력 요구 → 기술적인 발전에 기반하여 가능



1 빅데이터의 특징 4V

2 Variety

- 다양한 소스(데이터 발생 장소)에서 입수된 데이터 활용
- 데이터 형태도 **정형, 비정형**(Log, 텍스트, 사운드, 이미지, 동영상)으로 다양해짐
- 기업의 내부 데이터 다양성과 외부에서 수집하는 데이터로 다양해짐



1 빅데이터의 특징 4V

3 Velocity

- 데이터 발생 및 처리속도 빨라짐
- Web Log 처럼 수많은 사람들이 웹 페이지 클릭 시 생기는 수많은 데이터가 매우 빠르게 발생
- Social Media(페이스북, 트위터 등)에서 다수 사람이 데이터 생성 → 매우 빠르게 수집하여 **준 실시간(Near Real-time)**으로 처리



1 빅데이터의 특징 4V

4 Value

- 상대적으로 매우 중요한 요소
- 전통적으로 분석을 통한 Insight에서 **실행(Action)** 통한 기업 및 고객의 가치 창출 가능
- **서비스 제공**에 활용 증대
- 만족도 증대, 비용절감, 수익증대, 신규사업 진출로 가치 발생



01 빅데이터란 무엇인가?

3 비정형 데이터 예시



비정형 데이터(Unstructured Data)

Log

Web Log, Machine Log, 계량기

Text

소셜 미디어, 콜센터 통화 음성을
텍스트로 변환한 데이터

Sound

일상생활 현장/공장/건설현장/
기계에서 나오는 진동 및 소리

Image

사진, 얼굴 표정

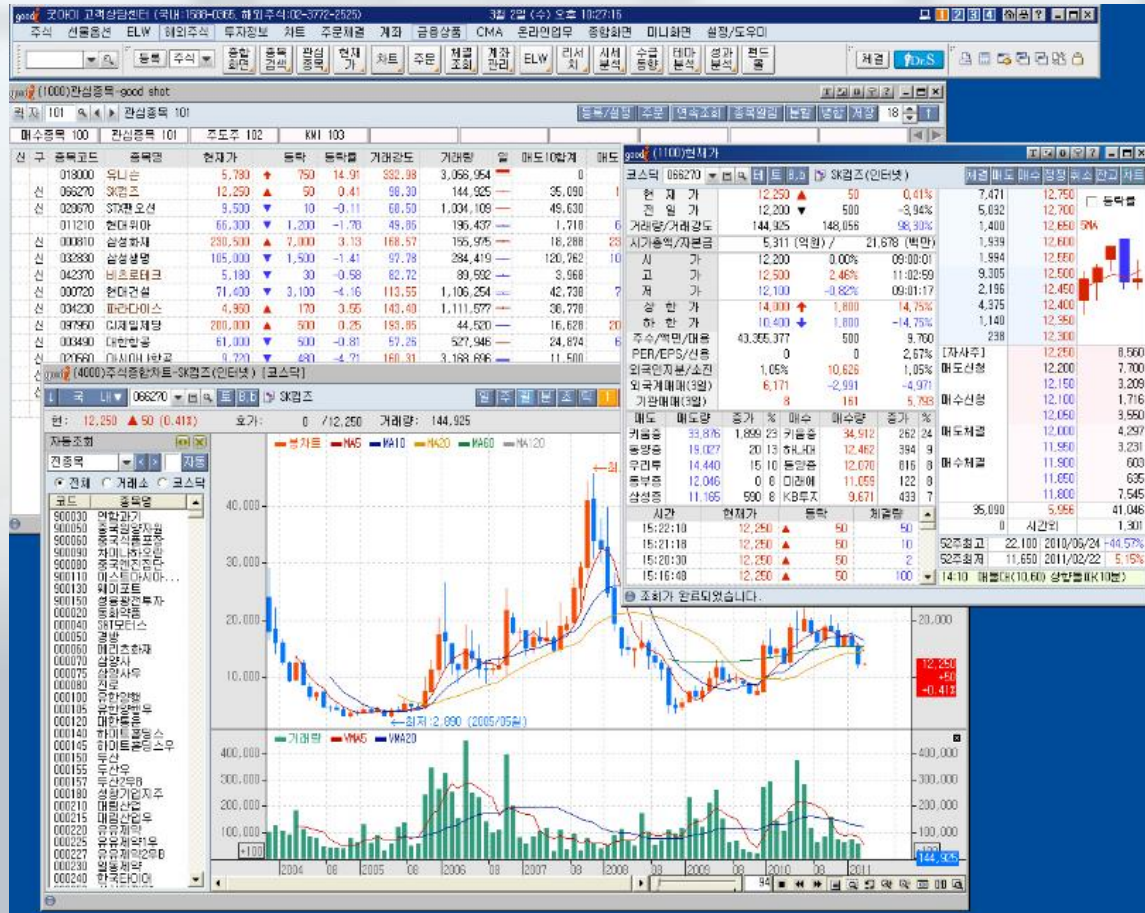
Video

CCTV



1 비정형 데이터(Unstructured Data)

1 Log(Semi-Structured)



1 비정형 데이터(Unstructured Data)

2 Text

- 일상생활에 대한 내용을 진실/거짓을 구분하여 입수하여 거짓을 예측하는 방법에 활용



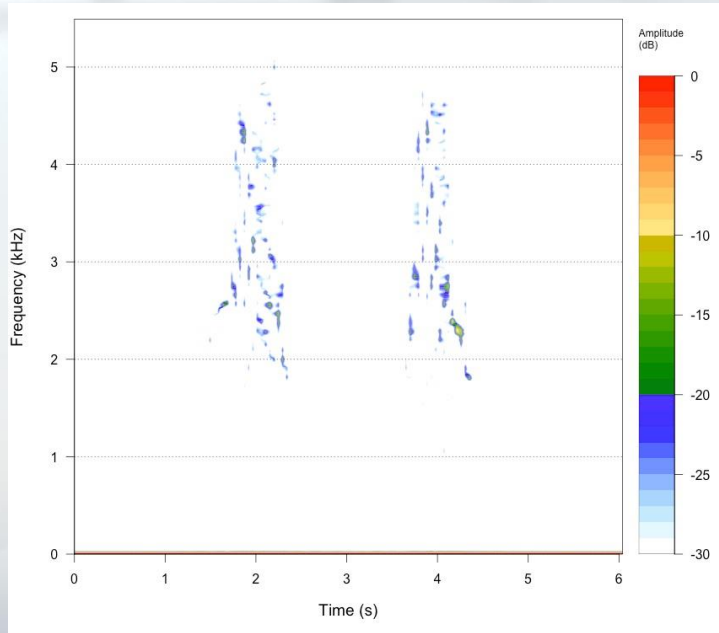
1 나는 모든 사람의 이야기를 들어 줄 수 있다. 그들의 이야기를 들어주며 그들에게 도움이 될 수 있는 방법을 찾든지 도움이 될 만한 말을 생각한다.

2 나는 절대 화를 내지 않는다. 화를 내어서 나의 감정을 상대방에게 들어내지 않기 위해 최대한 조절한다. 상대방에게도 상처를 줄 수 있다.

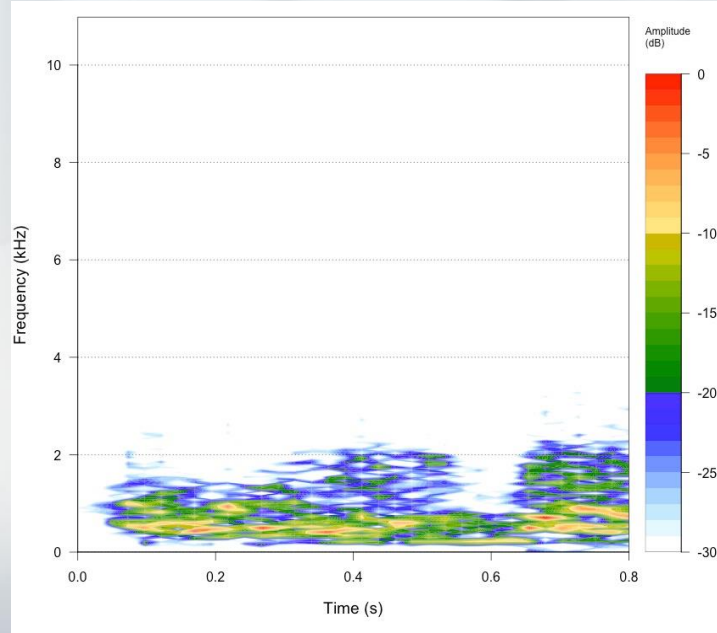


1 비정형 데이터(Unstructured Data)

3 Sound



강아지 화난 경우



강아지 배고픈 경우



1 비정형 데이터(Unstructured Data)

4 Image

- 손금 패턴 매칭을 통한 보안
- 손금 이용한 성향분석



■ 생명선			
표준형	강생명선	약 생명선	
■ 두뇌선			
표준형	직선형	짧은 형	
■ 성격선			
약간 짧은 형	아주 짧은 형	약간 긴 형	아주 긴 형
■ 생명선 & 두뇌선의 조합			
표준형	신중형	적극형	배짱형

1 비정형 데이터(Unstructured Data)

5 Video

- 장소 및 복장, 행동을 통해 범죄예방 경고
- 범인 키, 체중, 혈관 등 프로필 데이터 추출



2 빅데이터 활용 접근방법

“ 빅데이터의 가치는 활용에 대한 **창의적 접근법** ”

- 1 데이터화 시킬 수 있는 항목 파악
- 2 활용할 수 있는 방안 고민 및 리서치
- 3 분석을 통한 팩트 및 패턴 파악
- 4 데이터 마이닝 적용을 통한 패턴파악 자동화
- 5 서비스 적용

