- 데이터의 정의

1. 라틴어인 dare(주다)의 과거 분사형, '주어진 것'이란 의미로 사용되었음
2. 데이터는 추론과 추정의 근거를 이루는 사실(옥스포드 대사전 정의)
3. 데이터는 단순한 객체로서의 가치뿐만 아니라, 다른 객체와의 상화관계 속에서 가치를 갖는 것

- 데이터의 특성

1. 존재적 특성 : 객관적 사실
2. 당위적 특성 : 추론, 예측, 추정을 위한 근거

존재적 특성, 당위적 특성.

- 데이터의 유형

1. 정성적 데이터 : 언어 문자 등, '회사 매출이 증가함'이 이에 대한 예시, 저장 검색 분석에 많은 비용이 소모 됨, 정성적 데이터를 빅데이터라고 할 수 있음, 통계적 분석이 어려움
2. 정량적 데이터 : 수치 도형 기호 등, 나이 몸무게 주가가 이에 대한 예시, 통계적 분석이 용이함

정성적 데이터
정량적 데이터

- 지식경영의 핵심 이슈

1. **암묵지** : 사회(조직)으로 부터 나와서 개인에게 축적된 내면화된 지식 -> 조직의 지식으로 **공통화, 내면화**
2. **형식지** : 언어, 기호, 숫자로 표출화된 지식 -> 개인의 지식으로 **연결화, 표출화**

암묵지 - 공통화, 내면화
형식지 - 연결화, 표출화
표출함으로써 개인들이 연결됨.

- **DIKW**

1. 데이터 : 개별 데이터 자체로는 의미가 중요하지 않은 객관적인 사실(A마트는 100원에, B마트는 200원에 연필을 판매한다.)
2. 정보 : 데이터의 가공, 처리와 데이터간 연관관계 속에서 의미가 도출된 것(A마트의 연필이 더 싸다)
3. 지식 : 데이터를 통해 도출된 다양한 정보를 구조화하여 유의미한 정보를 분류하고 개인적인 경험을 결합시켜 고유의 지식으로 내재화된 것, 정보 패턴을 이해하여 이를 토대로 예측한 결과물 또는 특정 결론(상대적으로 저렴한 A마트에서 연필을 사야겠다.)
4. 지혜 : 지식의 축적과 아이디어가 결합된 창의적인 산물(A마트의 다른 상품들도 B마트보다 쌀 것이라고 판단한다.)

데이터를 통해 알 수 있는 유의미한 사실
정보를 통해 내린 결론

- 데이터베이스의 일반적인 특성

1. **통합된 데이터** : 동일한 내용의 데이터가 중복되어 있지 않다는 것을 의미
2. **저장된 데이터** : 컴퓨터가 접근할 수 있는 저장 매체에 저장되는 것을 의미
3. **공용 데이터** : 여러 사용자가 서로 다른 목적으로 데이터를 공동으로 이용한다는 것을 의미
4. **변화되는 데이터** : 데이터의 삽입, 기존 데이터의 삭제, 갱신으로 항상 변화하면서도 항상 현재의 정확한 데이터를 유지해야 함

· 통합된 데이터, 저장된 데이터, 변화하는 데이터, 공용데이터.

- 데이터베이스의 다양한 측면에서의 특징

1. 정보의 축적 및 전달 측면 : 기계가독성, 검색가독성, 원격조작성
2. 정보 이용 측면 : 이용자의 정보 요구에 따라 다양한 정보를 신속하게 획득
3. 정보 관리 측면 : 정보를 일정한 질서와 구조에 따라 정리, 저장, 검색, 관리 할 수 있도록 하여 방대한 양의 정보를 체계적으로 축적하고 새로운 내용의 추가나 갱신이 용이
4. 정보 기술 발전 측면 : 네트워크 발전에 기여
5. 경제 산업 측면 : 경제·산업 발전에 기여

컴퓨터에게가 읽고 씀.

· 정보 이용 측면, 정보 관리 측면, 정보축적·전달, 경제·산업 발전, 네트워크 기술 발전.

- OLTP : 호스트 컴퓨터와 통신 회선으로 접속되어 있는 복수의 사용자 단말에서 발생한 트랜잭션을 호스트 컴퓨터에서 처리하여 그 결과를 즉석에서 사용자 단말 측으로 되돌려 보내 주는 처리 형태이다. 쉽게 말해, 기업 내 인사, 급여, 구매, 생산, 재고, 물류 등 기업 운영의 전반적인 측면을 포함하는 데이터베이스에 접근하여, 조건에 맞는 데이터를 사용자 단말에 전달하는 처리 형태이다.

- 트랜잭션 : 단말에서 호스트 컴퓨터로 보내는 처리 단위 1회의 메시지로, 보통 여러 개의 데이터베이스 조작 명령을 포함하는 하나의 논리 단위이다.(예 : 특정 테이블의 데이터 값을 변경하는 경우, 해당 테이블과 연결되어 있는 다른 테이블의 데이터 값도 변경해야 하는 경우가 존재한다. 이런 경우에는 2개의 처리를 1개의 트랜잭션으로 연속해서 행해야 한다.)

- OLAP : 쉽고 빠르게 다차원적인 데이터에 접근하여 의사 결정에 활용할 수 있는 정보를 얻을 수 있게 해주는 기술, 즉 다차원으로 이루어진 데이터로부터 통계적인 요약정보를 제공할 수 있는 기술, OLTP에서 처리된 트랜잭션 데이터를 분석해 제품의 판매 추이/구매 성향 파악/재무 회계 분석 등을 프로세싱하는 것을 의미함

2000년대 기업 내부 데이터베이스

- CRM(Customer Relationship Management)

1. '고객관계관리'라고 하며, 기업이 고객과 관련된 내외부 자료를 분석 통합해 고객 중심 자원을 극대화하고, 이를 토대로 고객특성에 맞게 마케팅 활동을 계획 자원 평가하는 과정이다.
2. 고객데이터의 세분화를 실시하여 신규고객획득 및 잠재고객 활성화(즉 외부 고객에게도 관심이 있다는 뜻), 우수고객 유지, 고객가치증진, 평생고객화와 같은 사이클을 통하여 고객을 적극적으로 관리하고 유도한다.

- SCM(Supply Chain Management)

1. "공급망" 관리'를 뜻하는 말로, 기업에서 원재료의 생산 유통 등 모든 공급망 단계를 최적화해 수요자가 원하는 제품을 원하는 시간과 장소에 제공하는 것이다.
2. 부품 공급업체와 생산업체 그리고 고객에 이르기까지 거래관계에 있는 기업들간 IT를 이용한 실시간 정보공유를 통해 시장이나 수요자들의 요구에 기민하게 대응토록 지원하는 것이다.

기업 분야별 데이터베이스 소개

- 제조분야

1. ERP(Enterprise Resource Planning) : 인사 재무 생산 등 기업의 전 부문에 걸쳐 독립적으로 운영되던 각종 시스템의 경영자원을 하나의 통합 시스템으로 재구축한 것
2. BI(Business Intelligence) : 기업이 보유하고 있는 수많은 데이터를 정리하고 분석하여, 기업의 의사결정에 활용하는 일련의 프로세스
3. CRM
4. RTE(Real-Time Enterprise) : **회사의 주요 정보를 통합관리하는 실시간** 기업의 새로운 기업경영시스템이다. ERP, CRM, SCM 등 부문별 전산화에서 한발 나아가 회사 전 부문의 정보를 하나로 통합함으로써, 경영자의 빠른 의사결정을 이끌어 내려는 목적에 만들어졌다.

- 금융분야

1. **EAI**(Enterprise Application Integration) : **기업 내 상호 연관된 모든 어플리케이션(ERP, CRM, SCM 등)을 유기적으로 연동**하여 필요한 정보를 중앙 집중적으로 통합, 관리, 사용할 수 있는 환경을 구현하는 것
2. **EDW**(Enterprise Data Warehouse) : 기존의 **DW(Data Warehouse)를 전사적으로 확장한 모델**로 BPR, CRM, BSC 같은 다양한 분석 애플리케이션들을 위한 원천이 된다.

- 유통부문

1. KMS(Knowledge Management System) : **지식관리시스템**을 의미하며, 기업의 환경이 물품을 주로 생산하던 산업사회에서 **지적 재산의 중요성이 커지는 지식사회로 급격히 이동함에 따라**, 기업 경영을 지식이라는 관점에서 새롭게 조명하는 접근방식이다.
2. RFID(Radio Frequency Identification) : **주파수를 이용해 물품에 대한 ID를 식별**하는 SYSTEM으로 일명 전자태그로 불린다.

사회기반구조로서의 데이터베이스

1. EDI(Electronic Data Interchange) : 주문서, 납품서, 청구서 등 **무역에 필요한 각종 서류를 표준화된 양식을 통해 전자적 신호로 바꿔** 컴퓨터통신망을 이용하여, 거래처에 전송하는 시스템
2. VAN(Value Added Network) : **부가가치통신망**, 부가가치가 높은 서비스를 위해 독자적인 네트워크를 형성하는 것
3. CALS(Commerce At Light Speed) : 전자상거래 구축을 위해 기업 내에서 비용 절감과 생산성 향상을 목적으로 시작된, 제품 설계 개발 생산에서 유통 폐기에 이르기까지 제품의 라이프 사이클 전반에 관련된 데이터를 통합하고 공유 교환할 수 있도록 한 경영통합정보시스템을 말한다.

- 3V 빅데이터 정의 : Volume, Variety, Velocity

- 4V 빅데이터 정의 : **Volume, Variety, Velocity + Value, Visualization, Veracity**

- 빅데이터 정의의 범주 및 효과(밑으로 갈 수록 넓은 관점의 정의이다.)

1. 데이터 변화 : 3V
2. 기술 변화 : 데이터 처리, 저장, 분석 기술 및 아키텍쳐 변화
3. 인재 조직 변화 : Data Scientist 같은 새로운 인재 필요, 데이터 중심 조직 필요

- 빅데이터 출현 배경 : 빅데이터 현상은 없었던 것이 새로 등장한 것이 아니라, 기존의 데이터, 처리기술, 다루는 사람과 조직 차원에서 일어나는 '변화'를 말한다.

- **3가지 출현 배경** : 산업계(고객 데이터 축적), 학계(거대 데이터 활용, 과학 확산, 통계 도구들의 발전), 기술발전(관련기술의 발달[ex : 저장 기술의 발달, 인터넷 보급, 모바일 혁명, 클라우드 컴퓨팅 등])
- 빅데이터가 만들어 내는 본질적인 변화 : 사전처리 -> 사후처리, 인과관계 -> 상관관계, 질 -> 양, 표본조사 -> 전수조사
- 빅데이터에 대한 가치 선정이 어려운 이유

1. **특정 데이터를 언제 어디서 누가 활용할지 알 수 없게 되었다.** 따라서 가치를 산정하는 것도 어려워졌다.

2. 빅데이터 시대에는 **데이터가 '기존에 없던 가치'를 창출**함에 따라 그 가치를 측정하기가 어려워졌다.
3. **현재는 가치가 없는 데이터일지라도, 추후에** 새로운 분석 기법이 등장한다면 **거대한 가치를 지닌 데이터가 될 수도 있다.**

- 빅데이터 시대의 위기 요인

1. **사생활 침해** : 개인정보가 포함된 데이터를 목적 외에 활용한 경우
2. **책임 원칙 훼손** : 빅데이터 기본분석과 예측기술이 발전하면서 정확도가 증가한 만큼, 분석대상이 되는 사람들은 예측 알고리즘의 희생양이 될 가능성도 증가한다. 즉, 어떤 사람이 범행을 저지르기 전에 체포, 자신의 신용도와 무관하게 부당하게 대출이 거절되는 경우가 발생할 수 있다.
3. **데이터 오용** : 빅데이터 분석 결과는 이미 일어난 일에 대한 데이터에 의존하기 때문에, 항상 맞을 수 없다. 또한 잘못된 지표를 사용하는 것도 빅데이터의 폐해가 될 수 있다.

- 위기 요인에 따른 통제 방안

1. **동의에서 책임으로** : 개인정보 제공자의 동의 -> 개인정보 사용자의 책임
2. **결과 기반 책임 원칙 고수** : 민주주의 국가의 형사 처벌은 잠재적 위협이 아닌 명확하게 행동한 결과에 대해 책임을 묻고 있다. 이 기본 원칙을 좀 더 보강하고 강화할 필요가 있다.
3. **빅데이터 분석 알고리즘에 대한 접근 허용** : 알고리즈미스트가 사용된 빅데이터 분석 알고리즘에 접근하여, 해당 예측 알고리즘의 부당함을 반증할 수 있는 방법을 명시해 공개할 것을 주문한다.

- **빅데이터 활용의 3요소**

1. 데이터 : 모든 것에 대한 데이터화
2. 기술 : 새로운 분석 기법 개발, 인공지능 발전
3. 인력 : 데이터 사이언티스트, **알고리즈미스트**(데이터 사이언티스트가 한 일로 인해 부당하게 피해가 발생하는 것을 막는 역할을 하며, 알고리즘 코딩 해석을 통해 빅데이터 알고리즘에 의해 부당하게 피해를 입은 사람을 구제하는 전문인력)

데이터 세트(레파지토리) 종류

- **Data Lake** : 데이터 웨어하우스와는 달리 별도로 "정형화"나 정규화 등을 하지 않고, 데이터를 있는 그대로 원시 데이터 상태를 저장한다는 것이다. 원시데이터를 사용자가 원하는대로 가공하여 본다.
- **Data Warehouse** : "오랜기간동안 기업의 업무 과정을 통해 **수집된 데이터**"를 **주제 중심적으로 통합하고 정제해서**, 가장 필요한 데이터들만 최소 비용으로 "**최적화되어 있지 않고 비효율적으로 배치된 상태**"이다. **정형화된 데이터**를 사용자가 약간의 변화를 통해서 동일하게 본다.
- **Data Mart** : "데이터 웨어하우스에 있는 일부 데이터를 가지고 특정 사용자를 대상으로 한다". 그래서 사용하기 쉽게 시스템에 최적화하고 사람이 알기 쉽게 변환하고 성능 면에서 효율적으로 모아놓는 곳이 데이터 마트이다.

(데이터베이스와 레파지토리 차이점 : 일반적으로 레파지토리가 더 넓은 개념이다. **레파지토리는 단순히 무엇인가 모여서 저장해둔 일반적인 개념이다.** 데이터베이스는 특정한 데이터 모델에 따르는 구조화된 데이터를 저장한다. **레파지토리라는 용어를 사용할 때는 구조화라는 제약이 반드시 수반되지 않는다. 쉽게 말해서, 구조화 되었거나 되지 않았거나 무조건 모아 놓으면 단순히 '레파지토리'라는 용어를 사용한다.**)

- 빅데이터를 활용한 기본 테크닉

1. **연관규칙학습** : 변인들 간에 주목할 만한 상관관계 분석
2. **유형분석** : 데이터들을 그룹으로 나누는 분석
3. 유전자 알고리즘 : 생명의 진화를 모방하여 최적해를 구하는 알고리즘
4. 기계학습 : train dataset으로부터 학습한 알려진 특성을 활용해 예측하는 방법
5. 회귀분석 : 독립변수에 따라 종속변수가 어떻게 변하는지를 보면서 두 변인의 관계를 파악할 때 사용
6. **감정분석** : 특정 주제에 대해 말하거나 글을 쓴 사람의 감정을 분석
7. **소셜네트워크분석** : 특정 사람들 간의 관계가 어떤지 파악할 때 사용

빅데이터 분석의 가치

1. **빅데이터 분석도 기존의 분석과 마찬가지로, 데이터 분석을 통한 가치를 만드는 것에 집중해야 한다**. 단순히 '빅데이터'자체에 중점을 두면 안된다.(즉, 데이터의 크기에 중점을 두면 안된다.)
2. 전략과 비즈니스의 핵심 가치에 집중하고 이와 관련된 분석 평가지표를 개발하고 이를 통해 효과적으로 시장과 고객 변화에 대응할 수 있을 때 빅데이터 분석은 가치있게 여겨진다.

- 빅데이터 분석을 본격적으로 실시하기 전에 도메인 지식(해당 분야의 전문지식)에 대해 철저히 파악해야 하고, 해당 분석 결과물을 통해 어떤 새로운 가치를 창출할 것인지를 명확히 정해야한다.

- **전략적 통찰력 없이 실시하는 데이터 분석의 함정**

1. 단순히 분석을 많이 사용하는 것이 곧바로 경쟁우위를 가져다 주지는 않는다
2. (호기심과 비판적 시각을 통해)명확한 전략을 세우지 않거나 비즈니스의 핵심 가치에 집중하지 않은 상태로 분석을 실시하게 되면, 쓸모없는 분석 결과들만 잔뜩 쏟아내게 된다.

- 산업별 일차원적 분석 애플리케이션

1. **금융 서비스** : 신용점수 산정, 사기 탐지, (금융 상품)가격 책정, 프로그램트레이딩, 클레임분석, 고객 수익성 분석
2. **병원** : (진료)가격 책정, 고객 로열티, 수익 관리
3. 에너지 : 트레이딩, 공급/수요 예측
4. 정부 : 사기 탐지, 사례 관리, 범죄 방지, 수익 최적화

- 일차적인 분석(간단한 분석, 이미 업계에 알려져 있는 분석)의 문제점

1. 일차적인 분석을 통해서도 해당 부서나 업무 영역에서는 상당한 효과를 얻을 수 있지만, **일차적인 분석만으로는 환경변화와 같은 큰 변화에 제대로 대응하거나 고객 환경의 변화를 파악하고 새로운 기회를 포착하기 어렵다.**

2. 특히, 급변하는 환경에서는 분석을 일차적 차원에서 점증적, 전술적으로 사용하면 성과는 미미할 수 있다.

- 전략도출 가치기반 분석

1. **전략적인 통찰력 창출에 포커스를 뒀을 때, 분석은 해당 사업에 중요한 기회를 발굴**하고, 주요 경영진의 지원을 얻어낼 수 있으며, 이를 통해 강력한 모멘텀을 만들어 낼 수 있다.
2. **최고가 되기 위해서는 일차원적인 분석을 통해 점점 분석 경험을 쌓아야하고** 작은 성공을 거두면 **분석의 활용 범위를 더 넓고 전략적으로 변화시켜야 한다.**
3. 사업성과를 견인하는 요소들과 차별화를 꾀할 기회에 대해 **전략적 통찰력을 주는 가치기반 분석단계로 나아가**

야 한다.

- 데이터 사이언스의 의미와 역할'

1. 의미 : 데이터 사이언스란 **데이터 공학, 수학, 통계학, 컴퓨터공학, 시각화, 해커의 사고방식, 해당 분야의 전문 지식을 종합한 학문**이다. **데이터로부터 의미있는 정보를 추출해내는 학문**으로, 정형 또는 비정형을 막론하고 인터넷, 휴대전화, 감시용 카메라 등에서 생성되는 숫자와 문자, 영상 정보 등 다양한 유형의 **데이터를 대상으로 분석 뿐만 아니라 이를 효과적으로 구현하고 전달하는 과정까지를 포함한 포괄적 개념**이다.

- 데이터 사이언스의 영역

1. Analytics(분석적 영역) : 수학, 확률모델, 머신러닝, 분석학, 패턴 인식과 학습, 불확실성 모델링 등
2. IT(데이터 처리와 관련된 IT 영역) : 프로그래밍, 데이터 엔지니어링, 데이터 웨어하우징, 고성능 컴퓨팅
3. 비지니스 분석(비즈니스 컨설팅 영역) : 커뮤니케이션, 데이터 시각화, 스토리텔링

- 데이터 사이언티스트의 요구 역량

1. Hard skill : 빅데이터에 대한 이론적 지식, 분석 기술에 대한 숙련
2. Soft skill : (전략적) 통찰력 있는 분석, 설득력 있는 전달, 다분야간의 협력

- 데이터 사이언스 시대에 인문학의 열풍이 도래한 이유

1. **컨버전스**(단말기가 멀티 기능을 가지게 되는 현상, 기능 융합) **-> 디버전스**(본래의 기능에만 충실하자는 개념, 컨버전스와 반대되는 개념)
2. **생산 -> 서비스**
3. **생산 -> 시장창조**

- 빅데이터에 거는 기대를 표현한 비유

1. 산업혁명의 석탄, 철 : 제조업 뿐만 아니라 서비스 분야의 생산성을 획기적으로 끌어올려 사회 경제 문화 생활 전반에 혁명적 변화를 가져올 것으로 기대된다.
2. 21세기의 원유 : 경제 성장에 필요한 정보를 제공함으로써 산업 전반의 생산성을 한 단계 향상시키고, 기존에 없던 새로운 범주의 산업을 만들어낼 것으로 전망된다.
3. 렌즈 : 렌즈를 통해 현미경이 생물학 발전에 미쳤던 영향만큼이나 데이터가 산업 발전에 영향을 미칠 것으로 기대된다.
4. 플랫폼 : 공동 활용의 목적으로 구축된 유무형의 구조물

- DMBS의 종류

1. 관계형 DBMS : 테이블 간의 관계로 이루어져 있고, 고유키가 각 튜플(레코드)을 식별한다.
2. 객체지향 DBMS : 정보를 '객체' 형태로 표현하는 데이터베이스 모델, 객체지향 프로그래밍에서 자주 사용되는 DBMS, **사용자 정의 데이터 및 멀티미디어 데이터 등 복잡한 데이터 구조를 표현 관리할 수 있는 DBMS**
3. 네트워크 DBMS : **레코드들이 노드로, 레코드들 사이의 관계가 간선으로 표현되는 그래프를 기반**으로 하는 데이터베이스 모델이다. 계층형 DBMS와 비슷하지만, 계층형 DBMS의 데이터 중복 문제를 개선한 DBMS이다.
4. 계층형 DBMS : 데이터의 관계를 트리 구조로 정의하고, 부모, 자식 형태를 갖는 구조이다.

- 비식별 기술의 종류와 예

1. Anonymization(익명화) : 'one way de-identification'방법을 의미한다. 즉, 익명화 이후 개인정보를 복원할

수 없게 된다.(De-identification - 개인과 관련된 정보를 제거하는 과정을 통칭)

2. Pseudonymization(가명처리) : 개인식별정보를 제거하고, 임의의 코드나 번호(가명)을 부여하는 것을 의미한다. 즉, 이름과 관련한 개인정보를 다른 이름으로 변경하는 기술이다.(ex : 홍길동, 35세, 한국대 재학 -> 임꺽정, 30세, 국내대 재학) • 가명처리 : 다른 이름 부여
3. 난수화 : **사생활 침해를 막기 위해 개인정보를 무작위 처리**하는 등, 데이터가 본래 목적 외에 가공되고 처리되는 것을 방지하는 기술 • 난수화 : 무작위(랜덤) 처리
4. 데이터 마스킹 : **데이터의 길이, 유형, 형식과 같은 속성을 유지한 채**, 새롭게 읽기 쉬운 데이터를 익명으로 생성하는 기술(ex : 홍길동, 35세, 한국대 재학 -> 홍**, 35세, **대학 재학) • 데이터 마스킹 : 다른 기호로 데이터를 가림
5. 총계처리 : **데이터의 총합 값을 보임**으로서 개별 데이터의 값을 보이지 않도록 함 • 총계처리 : 총합 처리
6. 데이터 값 삭제 : 데이터 공유, 개방 목적에 따라 데이터 셋에 구성된 값 중 필요 없는 값 또는 개인식별에 중요한 값을 삭제
7. 데이터 범주화 : 데이터의 값을 범주의 값으로 변환하여 값을 숨김(ex : 홍길동, 35세 -> 홍씨, 30~40세) • 데이터 범주화

- 데이터의 유형

1. **정형데이터** : 형태(스키마, 메타데이터)가 있으며 연산(데이터베이스 연산)이 가능하다. 주로 관계형 데이터베이스에 저장된다. 데이터 수집 난이도가 낮고 형식이 정해져 있어 처리가 쉽다.**(ex : ERP, CRM, SCM, 정보시스템 등)** ← 스키마 O, 연산 O
2. **반정형데이터** : 형태(스키마, 메타데이터)가 있으며, **연산이 불가능하다.**(코딩으로 처리해줘야 연산이 가능해진다.) 주로 파일(JSON, XML 등)로 저장된다. 데이터 수집 난이도가 중간이다. 보통 API 형태로 제공되기 때문에, 데이터처리 기술(파싱)이 요구된다.**(ex : 로그데이터, 모바일데이터, 센싱데이터)** ← 스키마 O, 연산 X
3. **비정형데이터** : **형태가 없으며, 연산이 불가능하다**. 주로 NoSQL에 저장된다. 데이터 수집 난이도가 높으며, 텍스트 마이닝 혹은 파일일 경우 파일을 데이터 형태로 파싱해야 하기 때문에, 수집 데이터 처리가 어렵다.**(ex : SNS 데이터, 영상, 이미지, 음성, 텍스트 등)** 스키마 X, 연산 X

- 데이터 무결성 : 데이터베이스 내 데이터에 대한 정확한 일관성, 유효성, 신뢰성을 보장하기 위해 데이터 변경/수정 시 여러가지 제한을 두어 데이터의 정확성을 보증하는 것을 말한다.

- 데이터 베이스 역사

1. 1950년 : 미국 군대의 군비상황을 쉽게 관리하기 위해 만들어진 data(데이터) base(기지)
2. 1975년 : 우리나라에서 데이터베이스 연구의 시작
3. 1980년 중반 : 국내 데이터베이스 기술 연구, 개발.