

2과목 (데이터 입출력 구현)

Chapter 01. 논리 데이터 저장소 확인

1. 데이터 모델

↳ 현실 세계의 정보를 인간과 컴퓨터가 이해할 수 있도록 추상화 하여 표현한 모델

2. 데이터 모델 절차

↳ 요구사항 분석 - 개념적 설계 - 논리적 설계 - 물리적 설계

3. 논리 데이터 모델링

↳ 업무의 모습을 모델링 표기법으로 형상화하여 사람이 이해하기 쉽게 표현하는 프로세스

1. 관계 데이터 모델: 테이블 형태, 1:1, 1:N, N:M

2. 계층 데이터 모델: 트리 형태, 1:N

3. 네트워크 데이터 모델: 그래프 형태, N:M

4. 관계 대수

↳ 관계형 데이터베이스에서 원하는 정보와 그 정보를 어떻게 유도하는가를 기술하는 절차적 정형 언어

- 관계대수 연산자 종류

1. 일반 집합 연산자

- 합집합(Union): \cup
- 교집합(Intersection): \cap
- 차집합(Difference): $-$
- 카티션 프로덕트(CARTESIAN Product): \times

2. 순수 관계 연산자

- 선택(Select): σ
- 프로젝트(Project): π
- 조인(Join): \bowtie
- 디비전(Division): \div

5. 관계 해석

↳ 튜플 관계 해석과 도메인을 해석 하는 비절차적 언어

6. 논리 데이터 모델링 속성

↳ 개체 (Entity), 속성 (Attributes), 관계(Relationship)

7. 개체-관계 (E-R) 모델

↳ 데이터와 그들간의 관계를 사람이 이해할 수 있는 형태로 표현한 모델

8. 정규화

↳ 데이터 모델에서 데이터의 중복성을 제거하여 이상현상을 방지하고, 데이터의 일관성과 정확성을 유지하기 위해 무손실 분해하는 과정

9. 이상현상 (Anomaly) (삽삭갱)

↳ 데이터 중복성으로 인해 릴레이션을 조작할때 발생하는 비합리적 현상

- 삽입 이상: 불필요 정보 입력

- 삭제 이상: 원치 않은 다른 정보 같이 삭제
- 갱신 이상: 특정 부분 수정 중복된 값 모순

10. 정규화 단계 (도부이결다조)

- 1NF : 도메인이 원자값
- 2NF : 부분 함수 종속 제거
- 3NF : 이행적 함수 종속 제거
- BCNF : 결정자 후보키가 아닌 함수 종속 제거
- 4NF : 다중값 종속 제거
- 5NF : 조인 종속 제거

11. 반정규화

↳ 정규화 일부러 위배

Chapter 02. 물리 데이터 저장소 설계

1. 참조 무결성 제약 조건

↳ 릴레이션과 릴레이션 사이에 대한 참조의 일관성을 보장하기 위한 조건

- 제한(RESTRICIT) : 다른 테이블이 삭제할 테이블을 참조중이면 제거하지 않는 옵션
- 연쇄(CASCADE) : 참조하는 테이블까지 연쇄적으로 제거하는 옵션
- 널값(SET NULL) : 참조되는 릴레이션에서 튜플을 삭제하고, 참조하는 튜플들의 외래값에 NULL 값을 넣는 옵션

2. 인덱스(Index)

↳ 데이터 레코드를 빠르게 접근하기 위해 '키값, 포인터' 쌍으로 구성되는 데이터 구조

3. 뷰 (view)

↳ 접근이 허용된 자료만을 제한적으로 보여주기 위해 하나 이상의 기본 테이블로 구성된 가상 테이블

4. 클러스터 (Cluster)

↳ 데이터 액세스 효율을 향상시키기 위해 동일한 성격의 데이터를 동일한 데이터 블록에 저장하는 물리적 저장 방법

5. 파티션 (Partition)

↳ 대용량의 테이블이나 인덱스를 작은 논리적 단위인 파티션으로 나누는 것

1. 범위 분할 (Range Partitioning) : 지정한 열의 값을 기준으로 분할 함
2. 해시 분할 (Hash Partitioning) : 해시 함수를 적용한 결과 값에 따라 데이터를 분할
3. 리스트 분할 (List Partitioning) : 특정 파티션에 저장 될 데이터에 대한 명시적 제어가 가능한 분할 기법
4. 조합 분할(Composite Partitioning) : 범위, 해시, 리스트 분할 중 2개의 이상의 파티셔닝을 결합하는 방식. 파티션이 너무 클때 사용

6. 파티션 장점

↳ 성능 향상, 가용성 향상, 백업 가능, 경합 감소

Chapter03 데이터베이스 기초 활용하기

1. 데이터베이스

↳ 다수의 인원, 시스템 또는 프로그램이 사용할 목적으로 통합하여 관리되는 데이터의 집합

2. 데이터베이스 정의

- 통합된 데이터 : 자료의 중복을 배제한 데이터의 모임
- 저장된 데이터 : 저장 매체에 저장된 데이터
- 운영 데이터 : 조직의 업무를 수행하는데 필요한 데이터
- 공용 데이터 : 여러 애플리케이션, 시스템들의 공동으로 사용하는 데이터

3. 데이터베이스 특성

- 실시간 접근성 : 쿼리에 대하여 실시간 응답이 가능해야 함
- 지속적인 변화 : 새로운 데이터의 삽입, 삭제, 갱신으로 항상 최신의 데이터를 유지
- 동시공용 : 다수의 사용자가 동시에 같은 데이터를 이용할 수 있어야 함
- 내용참조 : 사용자가 요구하는 데이터 내용으로 데이터를 찾을 수 있음

4. 데이터베이스 종류

- **파일 시스템(File System)** : 파일에 이름을 부여하고 저장이나 검색을 위해 논리적으로 그것들을 어디에 위치시켜야 하는지 등을 정의한 뒤 관리하는 데이터베이스 전 단계의 데이터 관리 방식
- **관계 데이터베이스 관리시스템(RDBMS)** : **관계형 모델**을 기반
 - ↳ 종류 : Oracle, MySQL 등..
- **계층형 데이터베이스 관리시스템(HDBMS)** : **데이터를 상하 종속적인 관계로 계층화** 하여 관리
 - ↳ 종류 : IMS, System2000 등..
- **네트워크 데이터베이스 관리시스템 (NDBMS)** : **데이터를 네트워크상의 망상 형태로 표현한 데이터 모델**
 - ↳ 종류 : IDS, IDMS 등..

5. DBMS (DataBase Management System)

↳ **데이터 관리의 복잡성을 해결**하는 동시에 데이터 추가, 변경, 검색, 삭제 및 백업, 복구, 보안 등의 기능을 지원하는 소프트웨어

6. DBMS 유형

- 키-값 (Key-Values) DBMS : Unique 한 키에 하나의 값을 가지고 있는 형태
- 컬럼 기반 데이터 저장 DBMS : Key안에 (Column, Value) 조합으로된 여러개의 필드를 갖는 DBMS
- 문서저장 DBMS : 값(Value)의 데이터 타입이 문서라는 타입을 사용하는 DBMS
- 그래프 DBMS : 시맨틱웹과 온톨로지 분야에서 활용되는 그래프로 데이터를 표현하는 DBMS

7. DBMS 특징

↳ **무결성, 일관성, 회복성, 보안성, 효율성**

8. 빅데이터

↳ 시스템, 서비스, 조직등에서 주어진 비용, 시간 내에 처리 가능한 수십 페타바이트(PB) 크기의 비정형 데이터

- 빅데이터 특성
 - ↳ 데이터의 양, 데이터의 다양성, 데이터의 속도
- 빅데이터 수집, 저장, 처리 기술
 - 비정형/ 반정형 데이터 수집 : 내/외부 정제되지 않은 데이터를 확보하여 수집 및 전송하는 기술
 - 정형 데이터 수집 : 내/외부 정제된 대용량 데이터의 수집 및 전송 기술

- 분산 데이터 저장/처리 : 대용량 파일의 효과적인 분산 저장 및 분산 처리 기술
- 분산 데이터베이스 : HDFS 칼럼 기반 데이터베이스로 실시간 랜덤 조회 및 업데이트 가능

9. NoSQL (Not Only SQL)

↳ 데이터 저장에 고정된 테이블 스키마가 필요하지 않고 **조인 연산을 사용할 수 없으며, 수평적으로 확장이 가능한 DBMS**

10. 데이터 마이닝 (Data Mining)

↳ 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아내는 기술

- 데이터 마이닝 주요기법
 - 분류 규칙 : 과거 데이터를 토대로 새로운 레코드의 결과 값을 예측하는 기법
 - 연관 규칙 : 데이터 안에 항목들 간의 종속관계를 찾아내는 기법
 - 연속 규칙 : 연관 규칙에 시간 관련 정보가 포함된 형태의 기법
 - 데이터 군집화 : 대상 레코드들을 유사한 특성을 지닌 몇 개의 소그룹으로 분할하는 작업