파일처리 2-1

레코드

데이타 필드 -> 특정개체의 한 성질값

레코드 타입 -> 논리적으로 연관된 데이타 필드들의 구조

엔티티 타입

레코드 어커런스 -> 레코드 타입의 인스턴스

화일

보조기억장치에 저장된 같은 종류의 레코드 집합

disk(RA Mathod) tape(SAM) -> 보조기억장치 vs RAM

공통 응용 목적을 위해 함께 저장된 데이타

ex 급여 계산, 인사기록, 재고 관리, 과학기술 계산

데이타베이스 DB

1.다수의 사용자가 Matisuer

2. 정보를 공유하기 위하여 Share

3. 체계적으로 통합된 데이터 집합 intergrate <-> Dump

데이타의 집합을 화일로 구성하는 이유

1. 주기억장치에 적재하기에는 너무 많은 양

2. 프로그램은 특정시간에 데이타의 일부만 접근

-> 데이타 전부를 주기억장치에 한꺼번에 저장시킨 필요X

3. 데이타를 특정 프로그램의 수행과 독립적으로 보관

-> 데이타의 독립성 유지

-> 응용 프로그램에 영향을 주지 않고, 데이터를 변경시킬 수 있는 성질

파일의 분류(1)

* 기능에 따라

마스터 파일(maste file)

한 시스템에서 조직체(시스템)의 업무에 관한 정적인 면을 나타내는 데이터 집합

-ex(제조 회사): 급여 마스터 파일, 고객 마스터 파일, 인사 마스터 파일, -> 테이블

비교적 영구적인 데이터 또는 역사적 데이터를 포함.

사전 파일

* 마스터 파일의 특수한 종류
* 데이터의 기술 – 타입, 크기, 이름, 사용
* 메타데이터(Meta data)

데이터의 대한 또다른 데이터

트랜잭션 파일(transaction file)

마스터 파일의 변경 내용을 저장

마스터 파일에

* 새 레코드 추가
* 현존 레코드를 제거
* 수정하기 위한 데이터

트랜잭션

* 논리적인 작업 단위
* 하나의 건수로 처리하는 작업

보고서 파일(report file)

사용자에게 보고서로 보이기 위해 일정한 형식을 갖춘 데이터를 저장

* 하드카피, 단말 장치 화면 -> 알집 압축, 인코딩

작업 파일(work file)

시스템에 있는 임시 파일(temporary file)

Data 입력 -> 처리A -> 워크파일{작업파일} (데이터 가공과정에 생기는 임시파일) -> 처리 B -> 출력 정보

프로그램 파일(program file)

데이터를 처리하는 명령어들을 포함

컴퓨터 언어

-고급언어(C, C++ Java emd)

-저급(어셈블리어. 기계어)

- 작업 제어 언어(job control language)등으로 작성

파일의 분류(2)

프로그램의 파일 접근 목적에 따라

입력 파일

* 프로그램이 읽기만 함

출력 파일

* 프로그램이 기록만 하기 위해 사용
* 프로그램에 의해 작성

파일의 분류(3)

1. 순차파일(sequential file) 배열
2. 인덱스된 순차 파일
3. 작접 파일(direct file) 해싱

파일 조직의 기본 개념(키와 인덱스)

키(key):

* 레코드를 식별하는데 사용되는 레코드 필드
* 기본키(primary key): 데이터 레코드를 유일하게 식별하고 저장하는 기억장소를 결정하는데 사용되는 레코드 필드
* 보조키(secondary key): 나머지 레코드 필드 중에서 레코드를 접근하는데 사용되는 레코드 필드

인덱스(index)

* 파일의 특정 필드에 대한 접근 효율을 높이기 위해서 만드는 보조적인 구조
* 레코드의 저장 위치를 기억해 두었다가 접근 시 이용함

레코드 접근 우선순위에 따라

* 기본 인덱스(primary index): 기본키를 포함한 필드들에 대한 인덱스
* 보조 인덱스: 기본 인덱스 이외의 인덱스

데이터 순서와의 일치도에 따라

* 집중 인덱스(clustered index) 데이터 레코드의 순서가 인덱스의 엔트리 순서와 동일하거나 유사하도록 유지하는 인덱스
* 비집중 인덱스: 집중 형태가 아닌 인덱스

인덱스엔트리의 구성에 따라

* 밀집 인덱스 (dense index) 데이터 레코드 하나에 대해 적어도 하나의 인덱스 엔트리를 구성해놓은 인덱스 = 역인덱스(inverted index)
* 희소 인덱스(sparse index): 레코드 그룹 또는 데이터 블록별로 하나씩 인덱스를 만들어 두는 인덱스

파일의 조직

파일의 데이터 레코드를 표현, 저장하는 기법

1. 다중 키 파일

* 여러 개의 키에 의한 접근경로
* 인덱스를 통해서만 데이터 접근
* 역화일, 다중리스트 파일 등

1. 다차원 파일

* 인덱스의 탐색키가 여러 개의 필드를 포함하는 복합키에 대한 인덱스를 지원하는 파일

Ex} CAD 등 일차원인 아닌 선, 면, 위치 등 다차원 데이터 처리

여러 개의 필드가 모여 하나의 키 역할 수행