

데이터 베이스 관리 시스템

7기 SQL 기초반

3. 파일 시스템과 DBMS의 비교



그림 1-20 파일 시스템으로 구축된 구매 및 판매 응용 프로그램

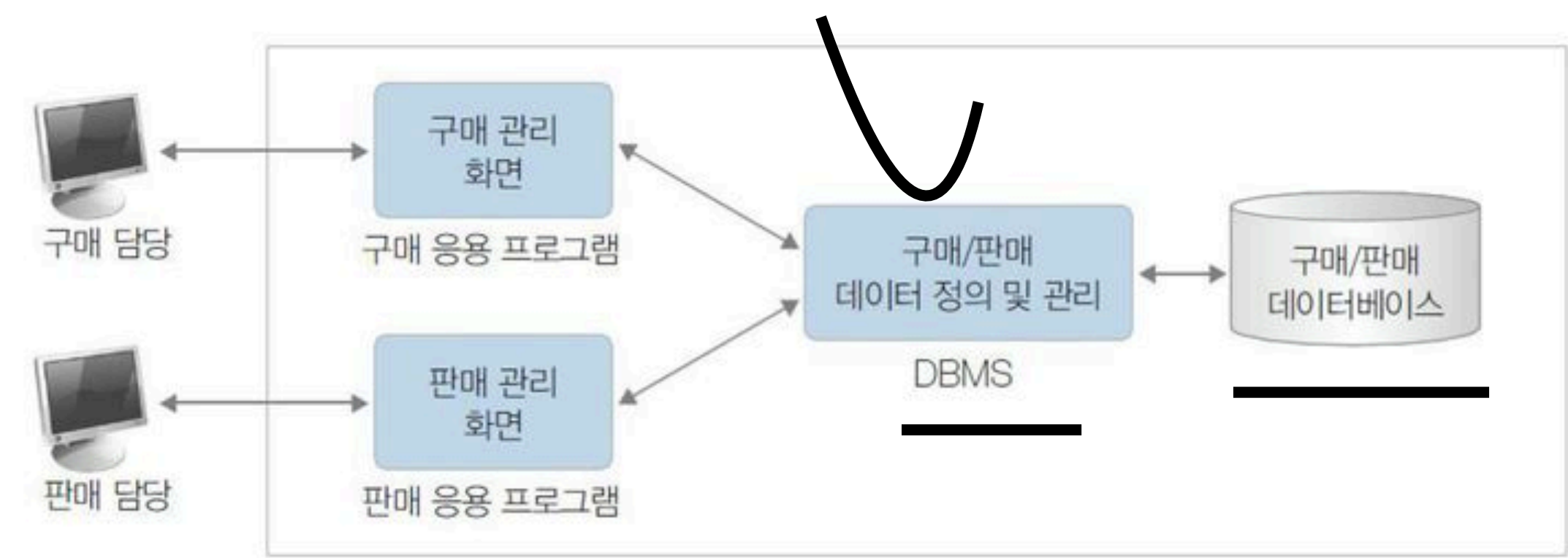


그림 1-21 DBMS로 구축된 구매 및 판매 응용 프로그램

데이터 중복성
데이터 종속성
파일 시스템의 문제

데이터 무결성 (정확성)

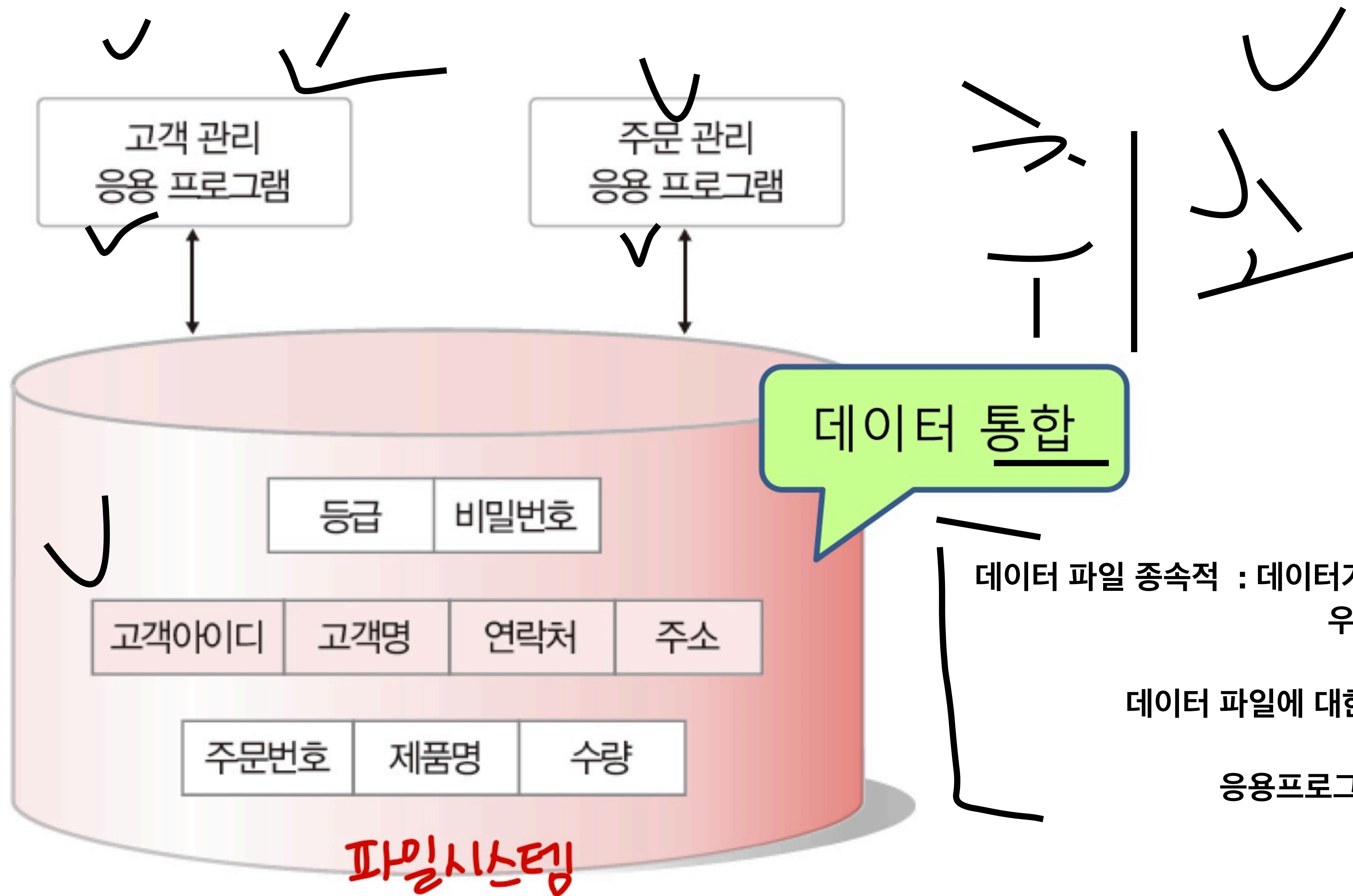
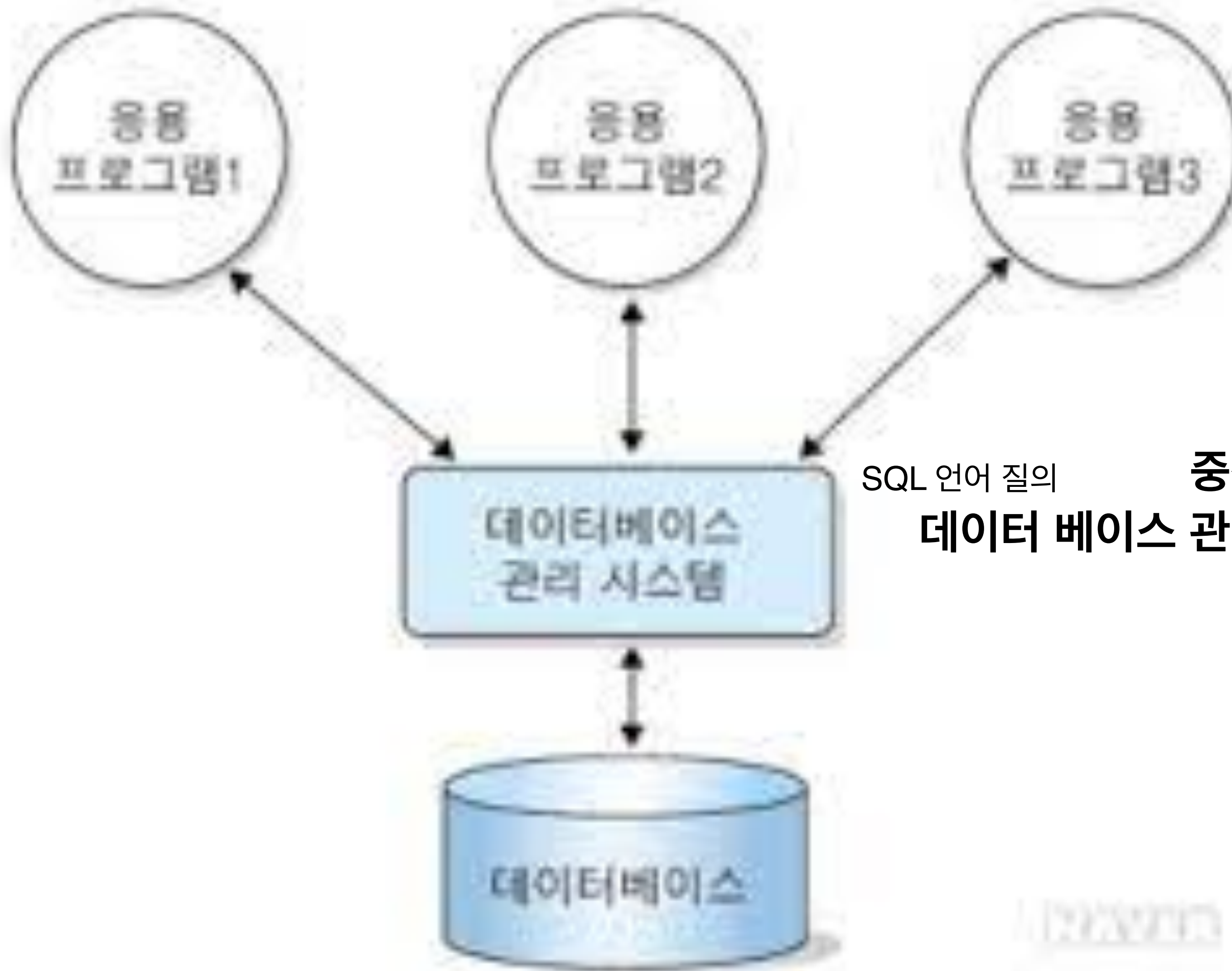


그림 2-2 파일 시스템의 데이터 중복성 문제를 해결하는 1차 방안 → 파일 통합

동시 공유, 보안 회복, 복잡한 기능 여전히 필요 = ?

데이터 베이스 관리 시스템



SQL 언어 질의

중간에 데이터베이스 관리 시스템이 막아줌

데이터 베이스 관리 시스템이 유저의 요구사항을 반영하여 데이터 보여줌



데이터베이스 관리 시스템

SQL DDL .DML,DCL 이런 관리시스템에서 정의하는 언어

상황에 따라 데이터 베이스 관리시스템을 사용하는 것

정의 기능 : 데이터 베이스 구조 정의, 수정-> 테이블을 만들고 데이터 베이스 설계하기 등

데이터 검색, 추출 기타등등 데이터 분석가들이 가장 많이 사용하는 경우

데이터를 추출해서 어떤 결과를 봐야하는데 이런 경우 방대한 데이터를 SQL 언어를 이용해서 추출한다. Select * from where ... and 등등 문법을 이용해서 데이터를 추출한다.

SQL문법은 다른 언어들에 비해서는 목적도 명확하고 상대적으로 기능이 엄청 많은 건 아닌 것 같다.

Python, R 비교했을 때 SQL 자체에서도 요즘은 ML도 가능하고 다양한 기능들이 사용된다.

새로운 테이블을 추가 하고 싶은 경우는 SQL 문법을 통해 테이블 만들고 -> 그것을 데이터베이스 관리시스템에 업데이트 하는 것

조작 기능 : 데이터베이스 내용 수정, 추가, 삭제, 검색, 삽입 등 연산 가능함

SQL 문법은 결국 데이터베이스를 가지고 쿼리를 날리는 경우 데이터베이스 즉 테이블, 데이터베이스 내의 환경이나 이런 부분을 정확히 알고 있어야 한다. 컬럼과 컬럼의 관계, 테이블간 관계

제어 기능 : 데이터의 무결성을 유지하게 되고, 장애 발생해도 회복 가능, 보안 유지, 동시 접근

SQL -> 질의만 하고 추출 이런 식에서 더 확장하여 ML 패키지도 사용할 수 있다.

ML 패키지는 Python이나 다른 언어로 만들어서 해당 패키지를 불러오는 형태로 사용합니다.

데이터 분석가들 중 ML에대한 전문가적인 지식이 없을 수도 있고, 실제 코드를 이해하거나 사용하기 어려운 경우 간단하게 SQL에서도 사용할 수 있어서 편리하게 바로바로 예측이나 기타 ML 기능들을 사용 가능하다.

빅쿼리, Snowflake, 하둡 기타등등 대부분 SQL내 ML 기능을 사용할 수 있다고 많이 이야기 한다.

생각보다 전문적인 건 아니지만 그래도 다양한 기능 존재한다. 미래예측 하는 정도

Python 가지고 SQL과 함께 엔지니어링 업무 가능하다. 두 가지를 잘 알면 강점은 확실하다.

데이터 베이스 장단점

데이터는 절대로 중복되면 안 된다.

실제 중복이 되거나 문제가 발생하면 큰일난다.

중복이라는 건 있을 수 없는 일

- 데이터 중복 통제 중복되면-> 고객이 중복되는 문제가 발생 ->매출, 손익라는 것이 문제 -> 조직 전사의 문제 -> 조직 전체가 잘못된 정보를 가지고 의사결정

- 데이터 독립성 확보
 - 여러 사람들이 데이터를 추가하고, 삭제하고 변경하고 이러다보면 문제가 발생한다.
 - 동시 공유 -> 데이터베이스 관리 시스템이 만들어진 것.

데이터 베이스를 가지고 쿼리만 날리는 건 아니다.

- **데이터 공유** : 데이터를 공유할 수 있고 여러가지로 데이터 베이스로 활용할 수 있다.
대시보드를 만들어서 모든 직원들이 (태블로 대시보드) 데이터를 보고 의사결정을 하는데 동시공유가 되어 있지 않으면 문제가 크다.

- 단점

$$\nabla \cdot \frac{\partial \mathbf{H}}{\partial t}$$

데이터 양

쿼리 질의하는 시간 (데이터를 추출하는 로직)

쿼리 응답하는 시간을 비용으로 측정

동시에 사용하는 사람

함께 사용하는 다른 프로그램 ML

- **데이터 보안 향상**
관리시스템에서 데이터를 관리하니 데이터베이스 직접적으로 수정작업하긴 어렵다.
정의기능 등 SQL을 사용해야 하므로 (권한 같은 것을 줄 수 있다.) 보안 향상이 된다.
- 고

- **고비용**

- 데이터 무결성 유지
무결성 -> 데이터 값 정확한지 확인하는 것 (숫자 정확, 신규 고객 회원 가입수, 탈퇴 회원수 정확한지)
- 백업과 회복 방법

- 백업과 회복 방법 복잡

당일 매출 금액이 정확한지? / 고객의 로그데이터를 데이터 베이스에 넣는다. 고객이 메인페이지-> 서브페이지 체류시간 데이터로 쌓는다.

- **표준화** 표준화 -> 정해진 규칙으로 만드는 것

- 중양 집중화 관리로 인한 취약점

데이터가 타입도 다르고, 스키마 구조도 다르고 여러가지로 다른 것

고객 컬럼은 문자열, 고객 주문금액은 수치형 정하는 것, 고객 고유 ID번호 12345678 숫자? 문자로 할 것인지?

- 장애발생시 회복

장애시 발생시 회복 -> 엑셀로 작업을 한다면 큰일나죠 다 날라가면

데이터베이스관리시스템은 -> 시스템이 문제가 발생해도 데이터 베이스 장애 발생하기 전에 그전상태로 바로 복구 가능하다.

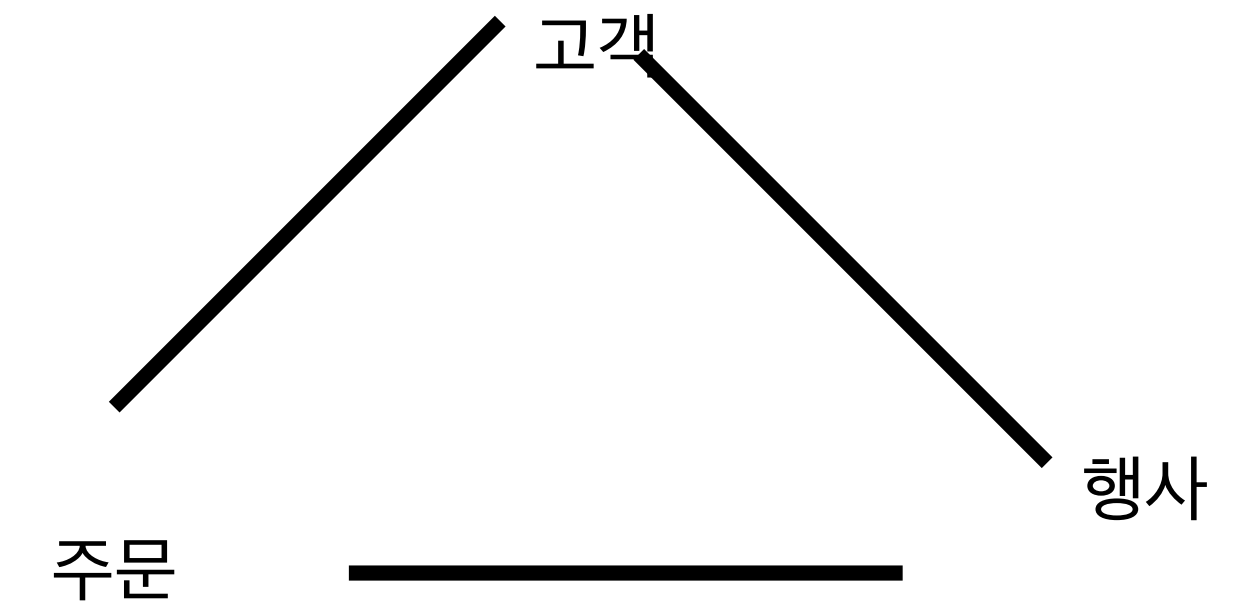
- 응용프로그램 개발 비용 줄어듦

데이터베이스 관리 시스템 발전과정

고객 : 고객과 주문과의 관계

주문 : 주문과 고객과의 관계

행사 : 주문과 행사와의 관계



- 1세대 데이터 베이스 관리 시스템 : 네트워크 계층 DBMS

- 2세대 데이터베이스 관리 시스템 : 관계 DBMS

기업에서 가장 많이 사용하는 건
관계형 데이터베이스

- 3세대 데이터베이스 관리 시스템: 객체지향, 객체관계 DBMS

- 4세대 이후 데이터베이스 관리 시스템: NoSQL, NewSQL DBMS

데이터들은 정형데이터가 아닌 비정형 수집할 수 있다.
사진, 음성이나, 이런 것들

그럼 이런 비정형데이터는 어떤 식으로 데이터베이스에 적재해야하는가?
이런 새로운 데이터베이스 관리 시스템 나온 것