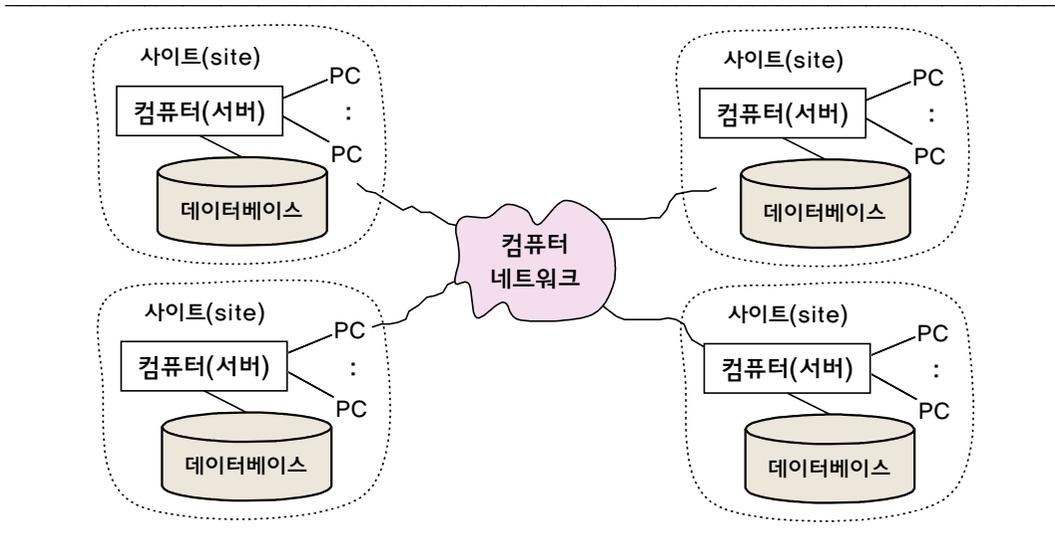


2. 분산시스템

분산시스템(distributed database system)은 지리적으로 분산되어 있는 것이다.

하지만, 사용자 측면에서는 중앙 집중 데이터베이스시스템처럼 사용할 수 있다.



// 분산 데이터베이스 구성요소

① 분산처리기(distributed processor)

- 분산처리기는 지리적으로 분산되어 있는 각 지역 컴퓨터 시스템이다.
- 분산처리기는 각 지역에서 필요한 데이터를 처리할 수 있는 DBMS를 가진다.

② 분산 데이터베이스(distributed database)

- 분산 데이터베이스는 지리적으로 분산된 지역 데이터베이스로서 각 분산처리기와 연결되어 있다.
- 지역 데이터베이스 관리 시스템이다.

③ 컴퓨터 네트워크(computer network) / 통신 네트워크(communication network)

- 통신 네트워크는 분산처리기들을 통신망으로 연결시켜 자원을 서로 공유할 수 있도록 한다.
- 즉, 분산 데이터베이스가 논리적으로 하나의 시스템 기능을 수행할 수 있게 하는 망이다.

1. 분산 데이터베이스의 투명성

(1) 위치 투명성(location transparency)

- 위치 투명성은 사용자가 접근하려는 데이터가 어느 사이트에 저장되어 있는지 알 것을 요구하지 않는 것을 말한다.
- 시스템은 데이터의 모든 위치 정보를 카탈로그에 관리한다.
- 사용자의 데이터 요구가 있으면 시스템은 카탈로그 정보를 제공한다.
- 결론적으로 사용자는 데이터가 어느 사이트에 저장되어 있는지 알 필요는 없다.

(2) 중복 투명성(replication transparency; 복사 투명성)

- 중복은 하나의 데이터를 여러 사이트에 중복 저장하는 것을 말한다.
- 중복의 이점은 시스템 성능 향상과 가용성(availability)을 높이기 위한 방편이다.
- 중복은 다른 사이트와 통신하지 않고 자신 사이트에 있는 복사본을 바로 이용할 수 있게 한다.
- 중복 투명성은 복사본은 시스템이 알아서 관리하므로 사용자는 데이터 중복에 대해서 무시해도 된다는 뜻이다.
- **중복 투명성** : 사용자는 중복 데이터 검색과 갱신에 대해서 전혀 신경 쓰지 않아도 되는 것이다.

(3) 분할 투명성(fragmentation transparency)

- 릴레이션 분할(단편화)은 수평 또는 수직으로 이루어진다.
- 사용자는 데이터가 분할 저장되어 있어도 분할 저장되어 있지 않은 것처럼 사용하는 것이다.
- 즉, 사용자는 데이터가 분할되어 저장되어 있는 것을 알 필요가 없다는 것이다.

(4) 병행 투명성(concurrency transparency)- 트랜잭션 투명성(transaction transparency)

- 여러 트랜잭션이 동시에 실행되어도 데이터베이스 일관성이 유지되어야 한다.
- 여러 네트워크의 사이트에서 트랜잭션 실행이 정상적으로 보장되어야 한다.

(5) 장애 투명성(failure transparency)

- 장애와 무관하게 트랜잭션 원자성이 보장되어야 한다.
- 시스템에 장애가 발생해도 사용자는 이를 인지 못하고 시스템이 복구하는 것을 의미한다.

(6) 지역사상 투명성(local mapping transparency)

- **지역사상 투명성**은 지역 DBMS로부터의 독립성을 제공하는 것이다.
- **동질 분산시스템** : 지역 DBMS가 모두 같은 데이터 모델을 사용하므로 지역사회상은 간단하다.
- **이질 분산시스템** : 지역 DBMS가 서로 다른 데이터 모델을 사용하므로 지역사회상은 복잡하다.
- 지역 DBMS의 특정 데이터 모델과 무관하게 분산 데이터베이스 관리 문제를 해결한다.
- 지역 DBMS와 물리적 데이터베이스 사이에 Mapping을 보장한다.

2. 분산 데이터베이스의 장단점

// 장점

지역 자치성	<ul style="list-style-type: none"> • 중앙시스템에 의존하지 않고 각 지역 고유 데이터를 자치적으로 행사 가능 • 지역시스템은 필요한 경우에 다른 사이트를 접근할 수 있다.
자원 공유 (resource sharing)	<ul style="list-style-type: none"> • 하드웨어 및 소프트웨어 자원 공유 가능 • 디스크, 프린터, 파일, 컴파일러 등
개방성(openness)	<ul style="list-style-type: none"> • 다양한 공급자로부터의 하드웨어 및 소프트웨어를 포함할 수 있다.
병행성(concurrency)	<ul style="list-style-type: none"> • 네트워크의 서로 다른 컴퓨터에서 다양한 프로세스가 동시 처리될 수 있다.
확장성(scalability)	<ul style="list-style-type: none"> • 시스템에 대한 새로운 요구에 대해서 쉽게 새로운 자원 추가가 가능하다.
결함 내성 (fault tolerance)	<ul style="list-style-type: none"> • 여러 컴퓨터 사용이나 정보의 중복이 이루어질 수 있다. • 이는 하드웨어나 소프트웨어 결함에 대한 내성을 가질 수 있도록 한다. • 결함 내성은 일부분이 고장나도 시스템 전체에는 영향주지 않는 것이다.
신뢰성과 가용성	<ul style="list-style-type: none"> • 중앙집중시스템에 비해 신뢰성과 가용성이 우수 • 특정 사이트 고장은 다른 사이트에는 거의 무관하다.

// 단점

복잡성 (complexity)	<ul style="list-style-type: none"> • 분산시스템은 중앙처리시스템보다 훨씬 복잡하다. - 구현하기 어렵다. • 소프트웨어 설계 및 관리의 복잡성과 개발 비용이 증가한다.
오류 잠재성	<ul style="list-style-type: none"> • 여러 사이트에서 병렬적으로 운영되므로 알고리즘 정확성 확인이 어렵다. • 치명적인 오류가 잠재되어 있을 가능성도 있다.
관리 (manageability)	<ul style="list-style-type: none"> • 다양한 컴퓨터들은 다양한 운영체제, 버전 등을 가진다. • 이러한 다양한 컴퓨터 환경에 대해 고려해주어야 한다. • 관리 비용 증대 : 사이트 사이의 조정, 통신 비용이 필요
예측불가능 (unpredictability)	<ul style="list-style-type: none"> • 분산시스템의 반응(response)에 대해서는 예측불가능하다.(불규칙한 응답시간) • 시스템을 둘러싸고 있는 다양한 환경이 원인이다. • 예 : 시스템 부하, 시스템의 구성 및 네트워크의 트래픽 등 예측이 어렵다.
무결성	<ul style="list-style-type: none"> • 데이터 무결성 보장의 어려움 • 무결성을 보장하려면 상당한 네트워크 자원을 필요로 한다.
보안 (security)	<ul style="list-style-type: none"> • 보안 관리가 다른 시스템들에 비해 상대적으로 어렵다.(보안 취약점) • 분산시스템은 다양한 컴퓨터로부터 접근 가능하므로(통제의 어려움) • 네트워크의 트래픽이 노출되므로

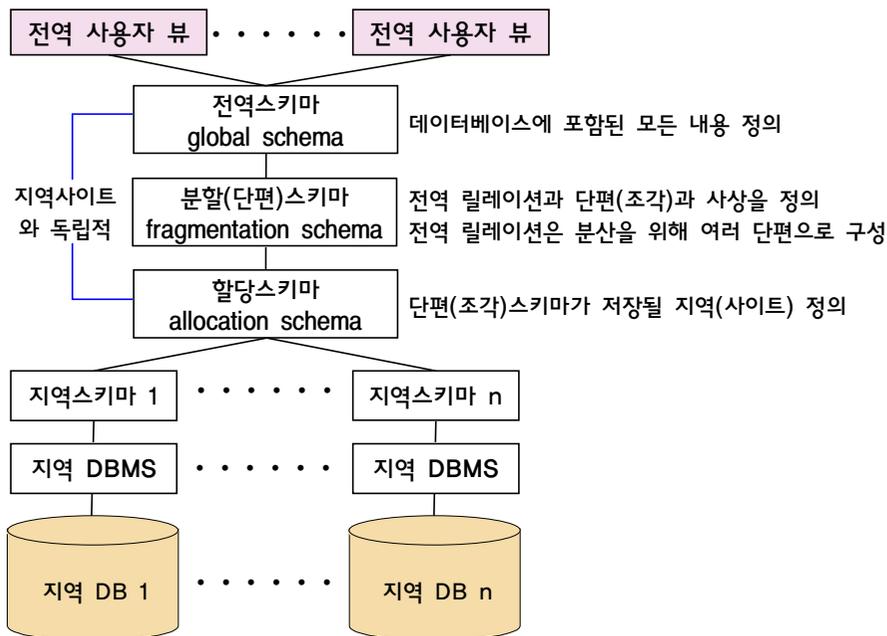
〈분산 데이터베이스에서 투명성〉

- 분산 데이터베이스 = 중앙 집중식 데이터베이스 + 분산 데이터 독립성
- 분산 데이터베이스는 중앙집중식 데이터베이스처럼 '분산 데이터 독립성'을 요구한다.
- '분산 데이터 독립성'을 지원하는 기법을 '분산 투명성'이라 한다.
- '분산 투명성'에는 단편화 투명성, 위치 투명성, 중복 투명성 등을 포함한다.

분산 데이터 독립성	분산 데이터 독립성은 실재는 데이터베이스가 분산되어 있지만 응용프로그램 작성은 중앙 집중식처럼 데이터가 분산되어 있지 않는 것과 같이 프로그래밍할 수 있는 것을 말한다.
------------	--

〈이석호 교수의 데이터베이스시스템〉

3. 분산 데이터베이스 구조



// 전역스키마(전역 개념스키마)

- 데이터베이스가 분산과 전혀 관계가 없다라고 생각하고 분산데이터베이스에 포함될 모든 것을 정의
- 해서, 전역스키마는 분산되지 않는 중앙데이터베이스와 똑같이 정의할 수 있다.
- 사용자들은 전역스키마를 통하여 마치 분산되지 데이터베이스를 사용하는 것처럼 이용할 수 있다.

기출문제 분석

1. 분산시스템 아키텍처의 장점이 아닌 것은? [2012년 국가 7급]

- ① 보안성 ② 자원 공유 ③ 병행성 ④ 결함 내성

☞ 분산시스템 아키텍처

• 분산시스템은 보안에 취약점을 가진다.

// 분산 데이터베이스 장단점

장점	<ul style="list-style-type: none"> • 자원 공유(resource sharing) 하드웨어 및 소프트웨어 자원 공유 가능(디스크, 프린터, 파일, 컴파일러 등) • 개방성(openness) 다양한 공급자(vendors)로부터의 하드웨어 및 소프트웨어를 포함할 수 있다. • 병행성(concurrency) 네트워크상의 서로 다른 컴퓨터에서 다양한 프로세스가 동시에 처리될 수 있다. • 확장성(scalability) 시스템에 대한 새로운 요구에 대해서는 쉽게 새로운 자원의 추가가 가능하다. • 결함 내성(fault tolerance) 여러 컴퓨터 사용이나 정보의 중복이 이루어질 수 있다. 이는 하드웨어나 소프트웨어 결함에 대한 내성을 가질 수 있도록 한다. • 중앙집중시스템에 비해 신뢰성과 가용성이 우수 특정 사이트 고장은 다른 사이트에는 거의 무관하다.
단점	<ul style="list-style-type: none"> • 보안(security) 보안관리가 다른 시스템들에 비해 상대적으로 어렵다.(보안 취약점) 분산시스템은 다양한 컴퓨터로부터 접근 가능하므로(통제의 어려움) • 복잡성(complexity) 분산시스템은 중앙처리시스템보다 훨씬 복잡하다. 소프트웨어 설계 및 관리의 복잡성과 구축 비용이 증가한다. • 관리(manageability) 다양한 컴퓨터들은 다양한 운영체제, 버전 등을 가진다. 이러한 다양한 컴퓨터 환경에 대해 고려해주어야 한다. • 예측불가능(unpredictability) 분산시스템의 반응(response)에 대해서는 예측불가능하다.(불규칙한 응답시간) 예 : 시스템 부하, 시스템의 구성 및 네트워크의 트래픽 등 예측이 어렵다. • 데이터 무결성 보장의 어려움 무결성을 보장하려면 상당한 네트워크 자원을 필요로 한다.

정답 : ①

2. 분산 데이터베이스(distributed database)에 대한 설명으로 가장 옳지 않은 것은? [2019년 서울 7급]

- ① 분산된 데이터베이스의 양이 줄어들어 데이터베이스 설계 및 구축의 비용이 감소한다.
- ② 동일한 데이터가 저장된 여러 지역에서 병렬처리를 수행할 수 있어 데이터 처리 성능이 향상된다.
- ③ 중복되는 데이터의 일치성을 보장하기 위한 추가적인 부하가 발생한다.
- ④ 데이터 처리 요청이 한 지역에 집중되지 않고 여러 지역에 분산되므로 처리 부담을 줄일 수 있다.

☞ 분산 데이터베이스

- 분산 데이터베이스의 양이 줄어들어 데이터베이스 설계 및 구축의 비용이 감소한다.(×)
- 분산 데이터베이스는 데이터베이스 설계가 복잡하고 및 구축의 비용이 증가한다.

// 분산 데이터베이스의 장단점

장점	<ul style="list-style-type: none"> • 자원 공유(resource sharing) 하드웨어 및 소프트웨어 자원 공유 가능(디스크, 프린터, 파일, 컴파일러 등) • 개방성(openness) 다양한 공급자(vendors)로부터의 하드웨어 및 소프트웨어를 포함할 수 있다. • 병행성(concurrency) 네트워크상의 서로 다른 컴퓨터에서 다양한 프로세스가 동시에 처리될 수 있다. • 확장성(scalability) 시스템에 대한 새로운 요구에 대해서는 쉽게 새로운 자원의 추가가 가능하다. • 결함 내성(fault tolerance) 여러 컴퓨터 사용이나 정보의 중복이 이루어질 수 있다. 이는 하드웨어나 소프트웨어 결함에 대한 내성을 가질 수 있도록 한다. • 중앙집중시스템에 비해 신뢰성과 가용성이 우수 특정 사이트 고장은 다른 사이트에는 거의 무관하다.
단점	<ul style="list-style-type: none"> • 보안(security) 보안관리가 다른 시스템들에 비해 상대적으로 어렵다.(보안 취약점) 분산시스템은 다양한 컴퓨터로부터 접근 가능하므로(통제의 어려움) • 복잡성(complexity) 분산시스템은 중앙처리시스템보다 훨씬 복잡하다. 소프트웨어 설계 및 관리의 복잡성과 구축 비용이 증가한다. • 관리(manageability) 다양한 컴퓨터들은 다양한 운영체제, 버전 등을 가진다. 이러한 다양한 컴퓨터 환경에 대해 고려해주어야 한다. • 예측불가능(unpredictability) 분산시스템의 반응(response)에 대해서는 예측불가능하다.(불규칙한 응답시간) 예 : 시스템 부하, 시스템의 구성 및 네트워크의 트래픽 등 예측이 어렵다. • 데이터 무결성 보장의 어려움 무결성을 보장하려면 상당한 네트워크 자원을 필요로 한다.

3. 분산 데이터베이스시스템에 대한 설명으로 옳지 않은 것은? [2011년 국가 7급]

- ① 여러 사이트에 분산된 데이터를 동기화하고 전체 데이터베이스의 무결성을 유지하기 위해 분산 트랜잭션 기능을 사용한다.
- ② 중앙집중형 데이터베이스시스템보다 구현이 간단하기 때문에 개발비용을 더 많이 줄일 수 있다.
- ③ 중앙집중형 데이터베이스시스템과는 달리 각 사이트의 붕괴와 통신회선의 고장과 같은 새로운 유형의 고장으로부터 회복하는 기능을 부가적으로 제공해야 한다.
- ④ 임의의 사이트에 존재하는 데이터베이스 카탈로그(디렉터리) 정보는 전체 분산 데이터베이스에 걸친 전역적인 정보를 유지할 수도 있고 지역적인 정보만을 유지할 수도 있다.

☞ 분산 데이터베이스시스템

- 분산 데이터베이스시스템은 중앙집중형 데이터베이스시스템보다
 - 구현이 복잡하고, 개발비용은 더 많이 든다.
-

정답 : ②

4. 분산 데이터베이스 시스템에 대한 설명으로 옳지 않은 것은? [2022년 국가 7급]

- ① 사용자들은 원하는 데이터가 어느 지역에 저장되어 있는지를 명확하게 알 수 있다.
- ② 데이터베이스를 여러 지역에 중복하여 분산 저장하면 가용성(availability)이 높아진다.
- ③ 지역 사이트들도 어느 정도 자치성(autonomy)을 가지고 독립적으로 동작할 수 있다.
- ④ 수평적 단편화(horizontal fragmentation)는 릴레이션 튜플들을 몇 개의 부분집합(단편)으로 분할한 후, 각 부분집합을 서로 다른 사이트에 저장하는 것이다.

☞ 분산 데이터베이스 시스템

- 사용자들은 원하는 데이터가 어느 지역에 저장되어 있는지를 명확하게 알 수 있다.(×)
 - 사용자는 자료가 어느 사이트에 저장되어 있는지 알 필요는 없다.(위치 투명성)
 - 자료의 모든 위치 정보는 시스템 카탈로그에서 관리
-

정답 : ①

5. 분산 데이터베이스(distributed database)의 특징에 대한 설명으로 가장 적절한 것은? [2023년
군무 7급]

- ① 인터넷 연결상태를 항상 유지해야 한다.
- ② 항상 세 개 이상의 컴퓨터들과의 연결상태를 유지해야 한다.
- ③ 사용자는 하나의 글로벌 스키마에서 데이터에 접근한다.
- ④ 갱신을 수행할 때 해당 데이터의 네트워크상의 물리적인 위치를 명기해야 한다.

♣ 분산 데이터베이스

- ① 인터넷 연결상태를 **항상 유지**해야 한다.(×)
→ 인터넷 연결상태를 항상 유지할 필요는 없다.
 - ② 항상 **세 개 이상**의 컴퓨터들과의 연결상태를 유지해야 한다.(×)
→ **두 개**의 컴퓨터들과 연결상태를 유지할 수도 있다.
 - ④ 갱신을 수행할 때 해당 데이터의 네트워크상의 **물리적인 위치**를 명기해야 한다.(×)
→ 사용자는 자료가 어느 사이트에 저장되어 있는지 알 필요는 없다.(위치 투명성)
-

정답 : ③