



빅데이터 모델링과 플랫폼 = 빅모플

Big Data Modeling and Platforms

AIX30011 Spring 2025

한동대학교
HANDONG GLOBAL UNIVERSITY

School of Applied Artificial Intelligence | 전재영



Copyright Notice

These slides are available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite the author as the source of the slides.

설문조사 결과

관계형 데이터베이스

- 없다 = ~63%
- 전산전자 = 2
- 빅데이터개론 = 5 (간략)
- SQLD = 1
- 인강 = 1
- 동아리 = 1

NoSQL

- 없다 = ~95%
- 기억이 안난다 = 1
- Firebase = 1

27명



What is Database?

What is Database?



DATA

Software/Platform



STORE



ORGANIZE

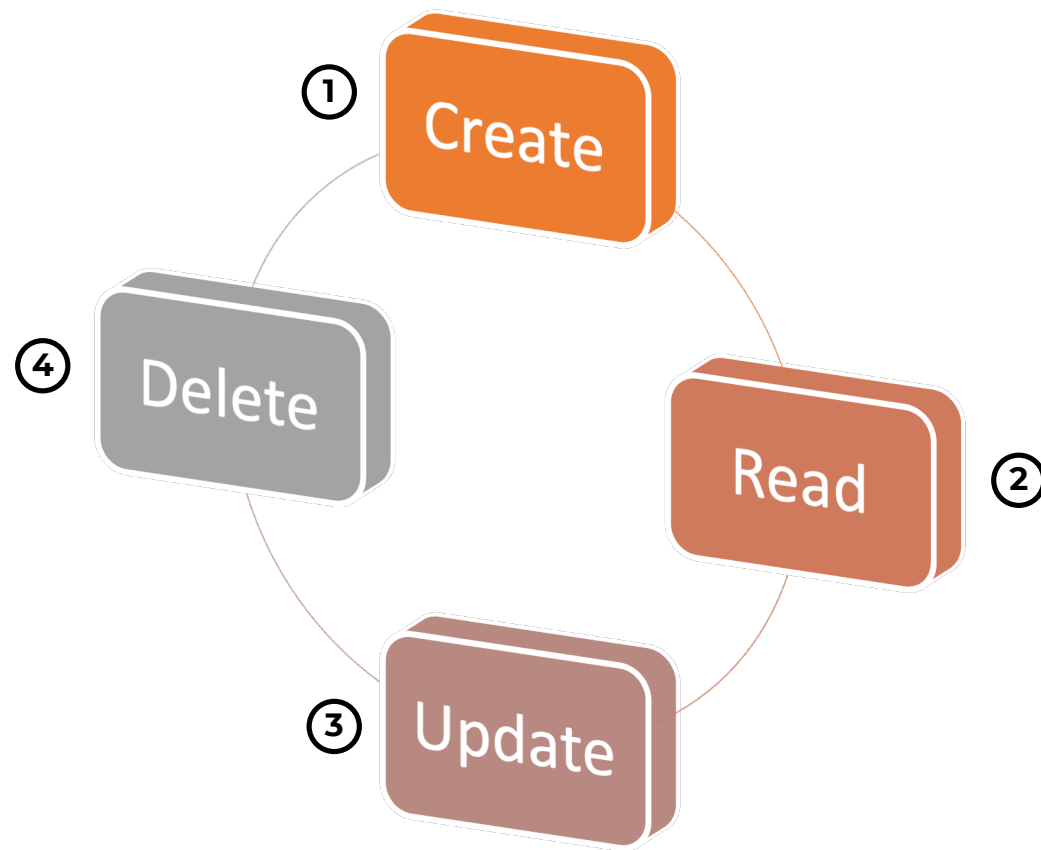


Source: freepik

Why Database?

What about files and spreadsheets?

CRUD



1. 새로운 데이터를 입력
2. 기존 데이터를 검색
3. 기존 데이터를 수정
4. 기존 데이터를 삭제

Passengers who paid \$501 for their flights

Live



booking.csv

```
booking_id  flight_id  seat  passenger_id  price
=====
...
55099780    265759    10H   34893         485.32
55099781    265759    9G     8059         343.11
55099782    265759    12C   9671         114.07
55099787    265759    8E    19595         314.48
55099788    265759    11D   26550         363.36
55099789    265759    11A   11041         398.42
55099791    265759    4G    20459         156.29
55099792    265759    8B     3621         254.24
55099796    265759    9F     2473          85.48
```

54M bookings

1.6 GB file size

1. Bash (grep, awk)
2. Text editor (Atom)
3. Spreadsheet (Microsoft Excel)
4. Pandas
5. ...

장단점이 있는 데 가장 심각한 문제는...

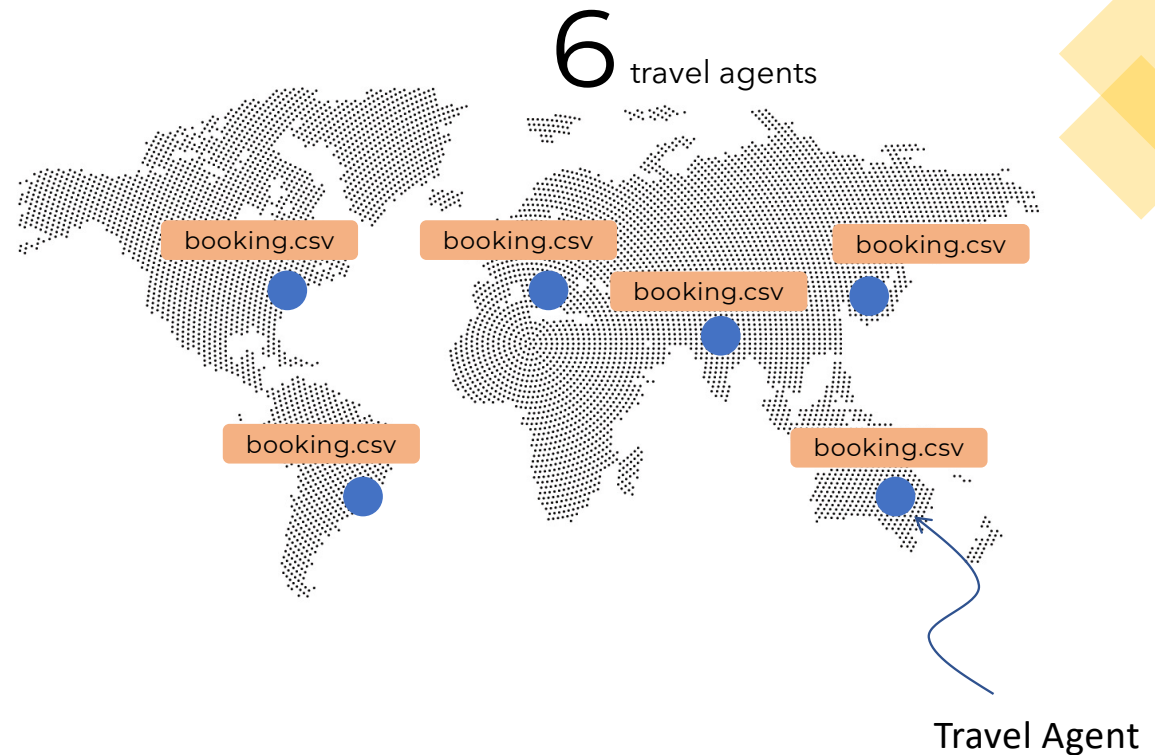
All Starting with the Same Data

booking.csv

booking_id	flight_id	seat	passenger_id	price
...
55099780	265759	10H	34893	485.32
55099781	265759	9G	8059	343.11
55099782	265759	12C	9671	114.07
55099787	265759	8E	19595	314.48
55099788	265759	11D	26550	363.36
55099789	265759	11A	11041	398.42
55099791	265759	4G	20459	156.29
55099792	265759	8B	3621	254.24
55099796	265759	9F	2473	85.48

54M bookings

1.6 GB file size



Update Problem

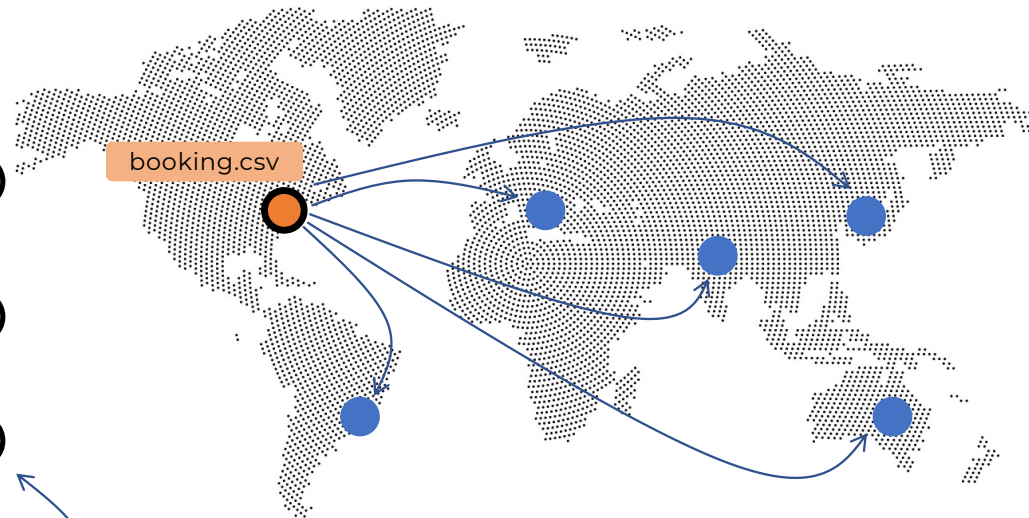
1.6 GB×5 agents = 8 GB
1.6 GB×100 agents = 160 GB
1.6 GB×201,400 agents = 322.24 TB

(*) There are roughly 201K agents in the world as of January 2023.



booking.csv

booking_id	flight_id	seat	passenger_id	price
...
55099780	265759	10H	34893	485.32
55099781	265759	9H	8059	343.11
55099782	265759	12C	9671	114.07
55099787	265759	8E	19595	314.48
55099788	265759	11D	26550	363.36
55099789	265759	11A	11041	398.42
55099791	265759	4G	20459	156.29
55099792	265759	8B	3621	254.24
55099796	265759	9F	2473	85.48
55099798	265759	5F	33987	353.56



File Transfer
Record Transfer

Updates from the travel agent
in New York City

Transfer Problem

booking.csv

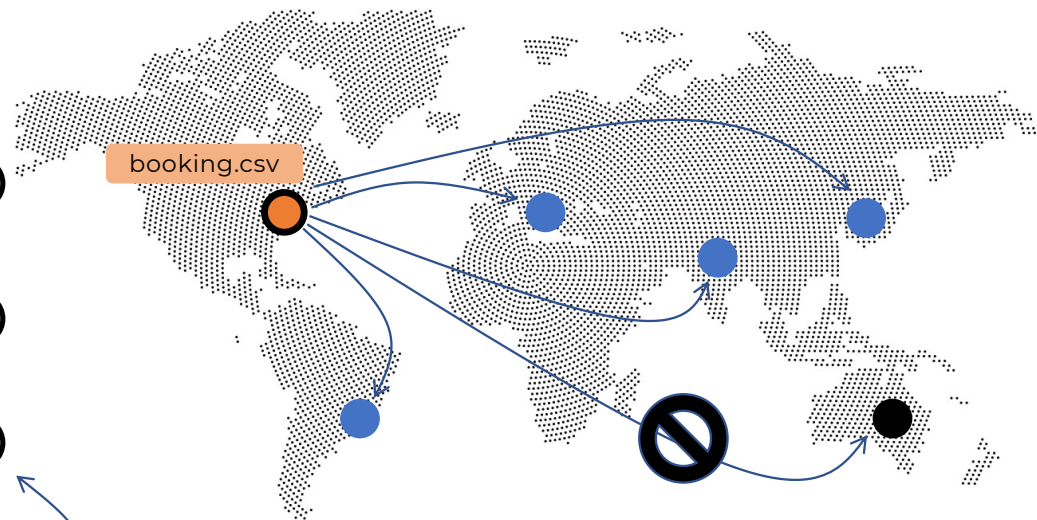
booking_id	flight_id	seat	passenger_id	price
...
55099780	265759	10H	34893	485.32
55099781	265759	9H	8059	343.11
55099782	265759	12C	9671	114.07
55099787	265759	8E	19595	314.48
55099788	265759	11D	26550	363.36
55099789	265759	11A	11041	398.42
55099791	265759	4G	20459	156.29
55099792	265759	8B	3621	254.24
55099796	265759	9F	2473	85.48
55099798	265759	5F	33987	353.56

①

②

③

Updates from a travel agent in
New York City



Computer at Sydney is off.



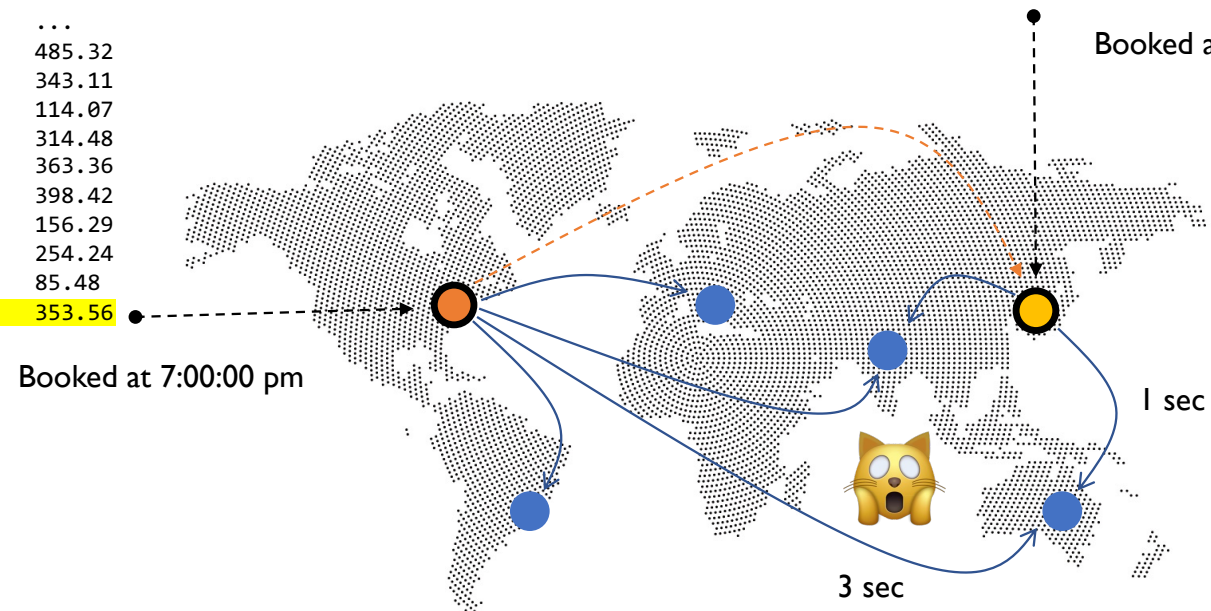
Inconsistency Problem

booking_id	flight_id	seat	passenger_id	price
...
55099780	265759	10H	34893	485.32
55099781	265759	9G	8059	343.11
55099782	265759	12C	9671	114.07
55099787	265759	8E	19595	314.48
55099788	265759	11D	26550	363.36
55099789	265759	11A	11041	398.42
55099791	265759	4G	20459	156.29
55099792	265759	8B	3621	254.24
55099796	265759	9F	2473	85.48
55099798	265759	5F	33987	353.56

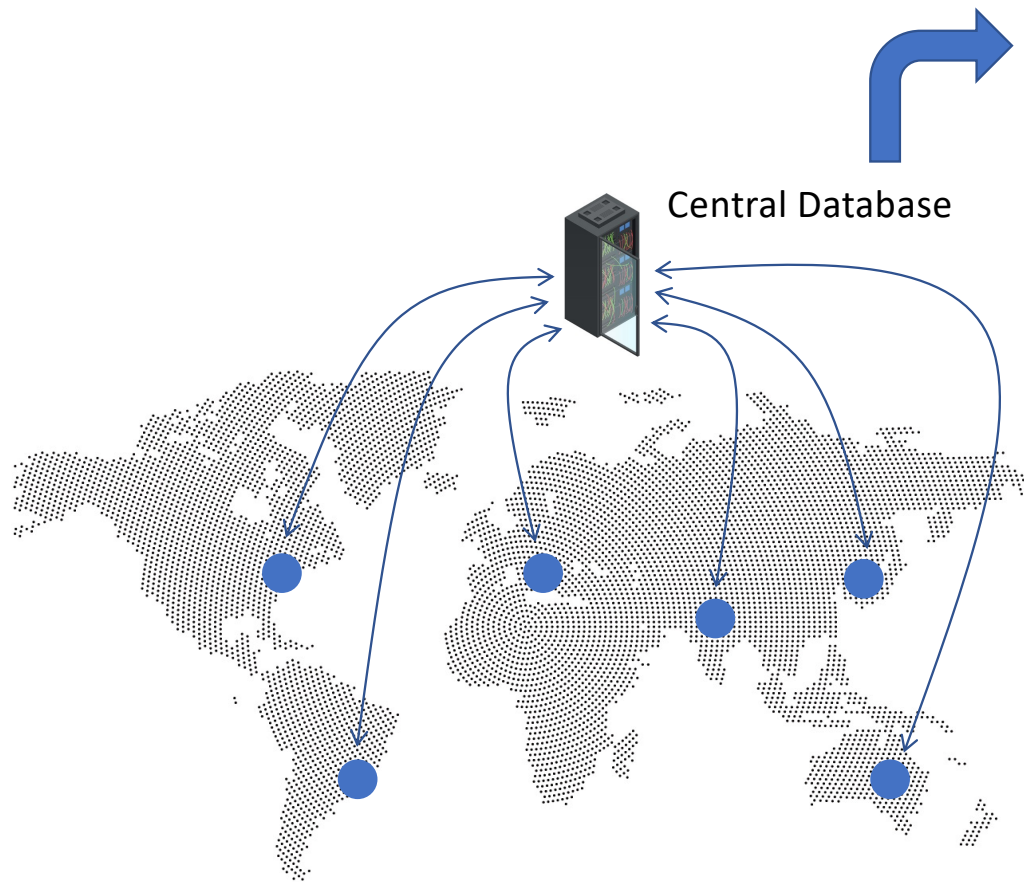
Booked at 7:00:00 pm

booking_id	flight_id	seat	passenger_id	price
...
55099780	265759	10H	34893	485.32
55099781	265759	9G	8059	343.11
55099782	265759	12C	9671	114.07
55099787	265759	8E	19595	314.48
55099788	265759	11D	26550	363.36
55099789	265759	11A	11041	398.42
55099791	265759	4G	20459	156.29
55099792	265759	8B	3621	254.24
55099796	265759	9F	2473	85.48
55099798	265759	5F	316	343.12

Booked at 7:00:01 pm



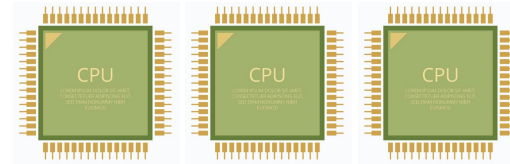
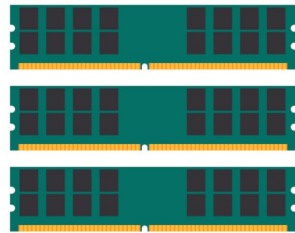
Introducing a Central Database



booking_id	flight_id	seat	passenger_id	price
...
55099780	265759	10H	34893	485.32
55099781	265759	9H	8059	343.11
55099782	265759	12C	9671	114.07
55099787	265759	8E	19595	314.48
55099788	265759	11D	26550	363.36
55099789	265759	11A	11041	398.42
55099791	265759	4G	20459	156.29
55099792	265759	8B	3621	254.24
55099796	265759	9F	2473	85.48
55099798	265759	5F	33987	353.56

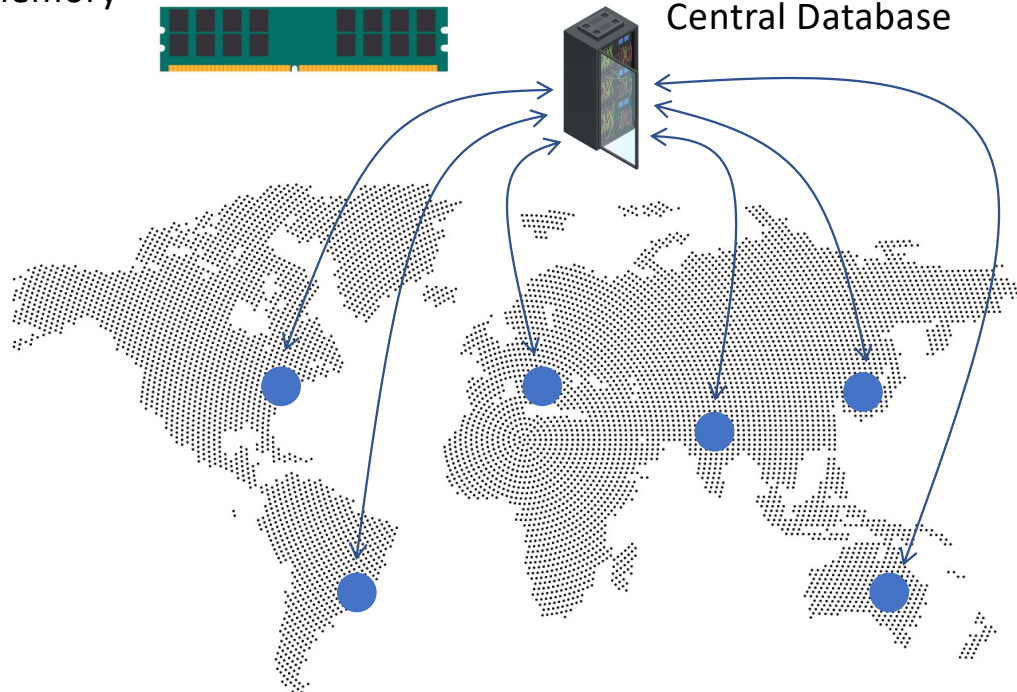
Scaling Up

Adding more memory

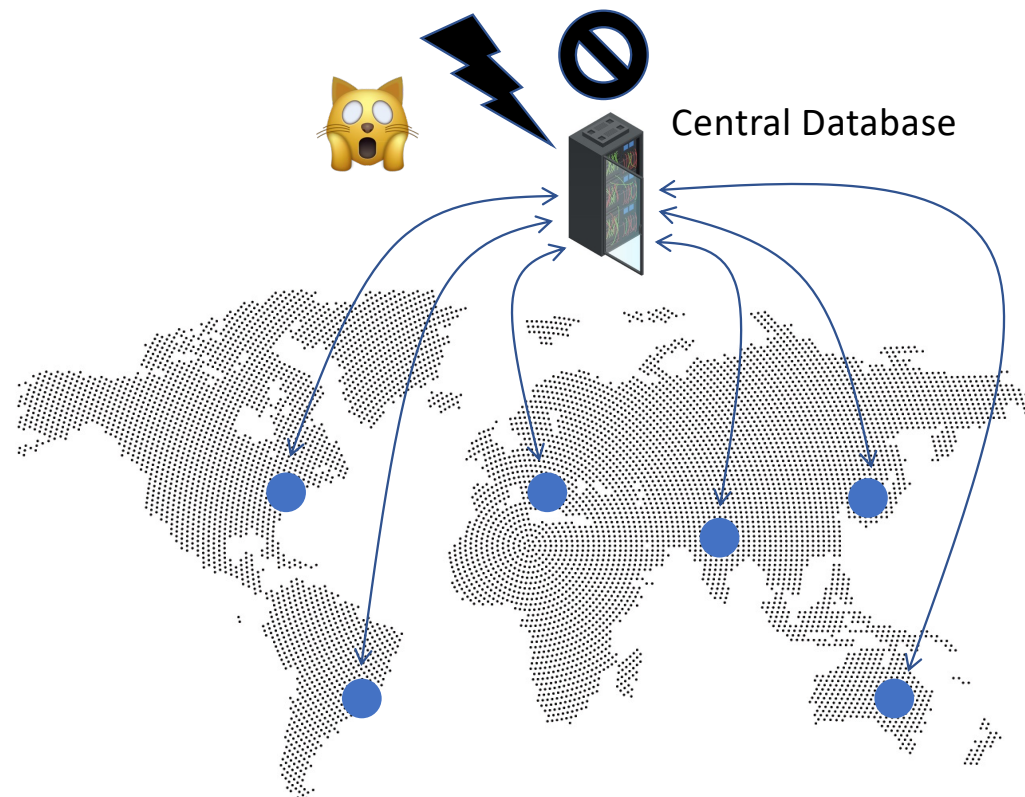


Adding more CPUs

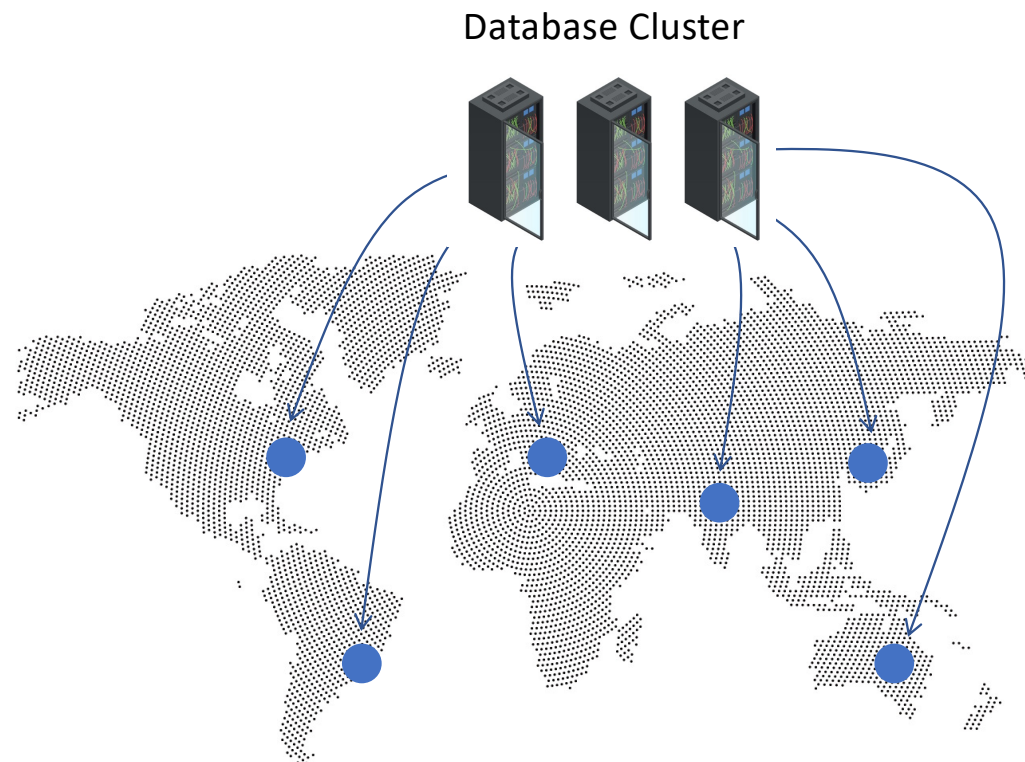
Central Database



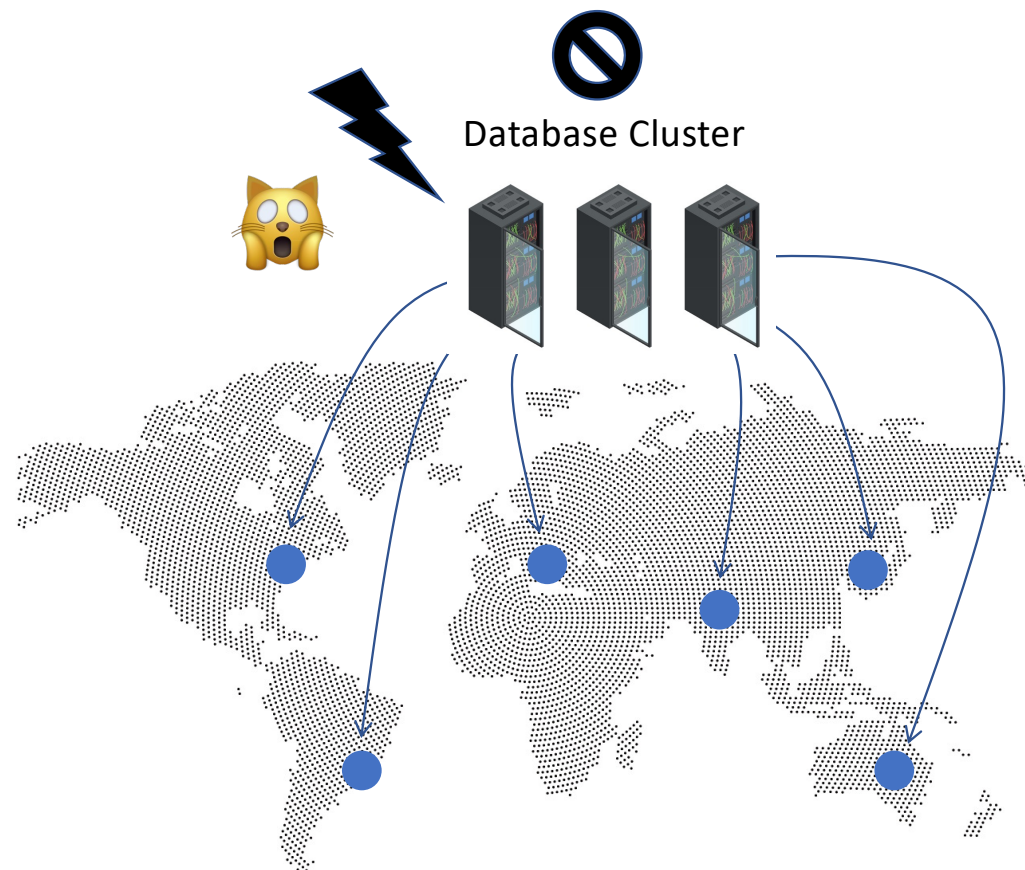
Single Point of Failure (SPOF)



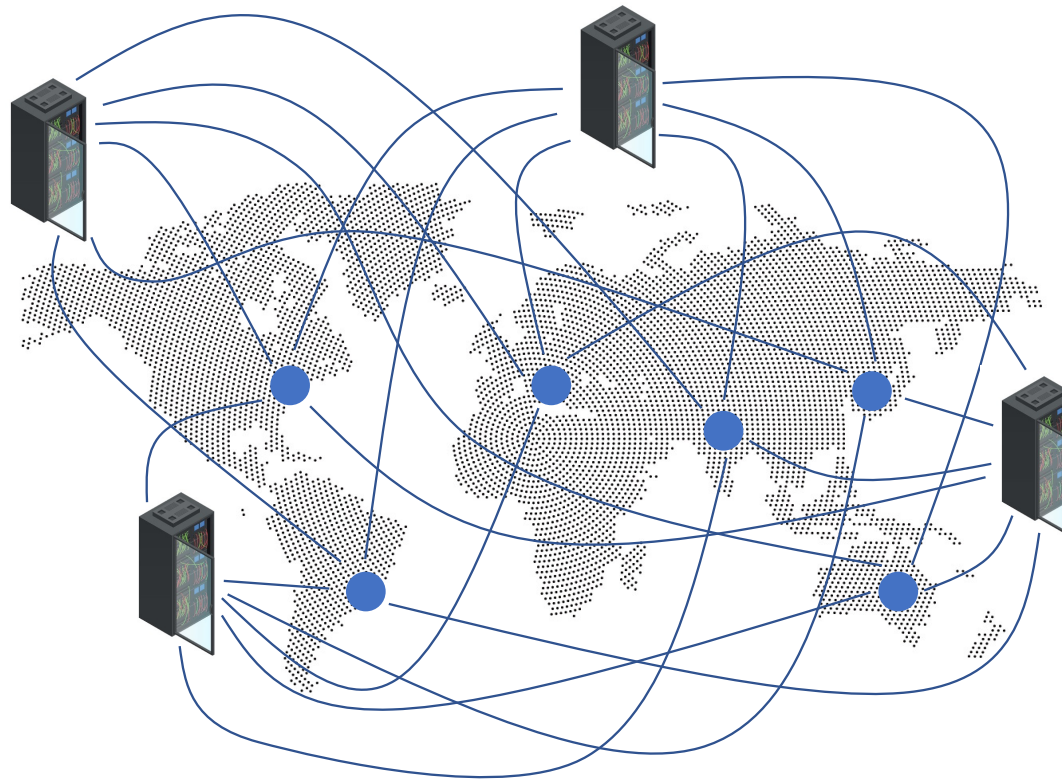
Scaling Out



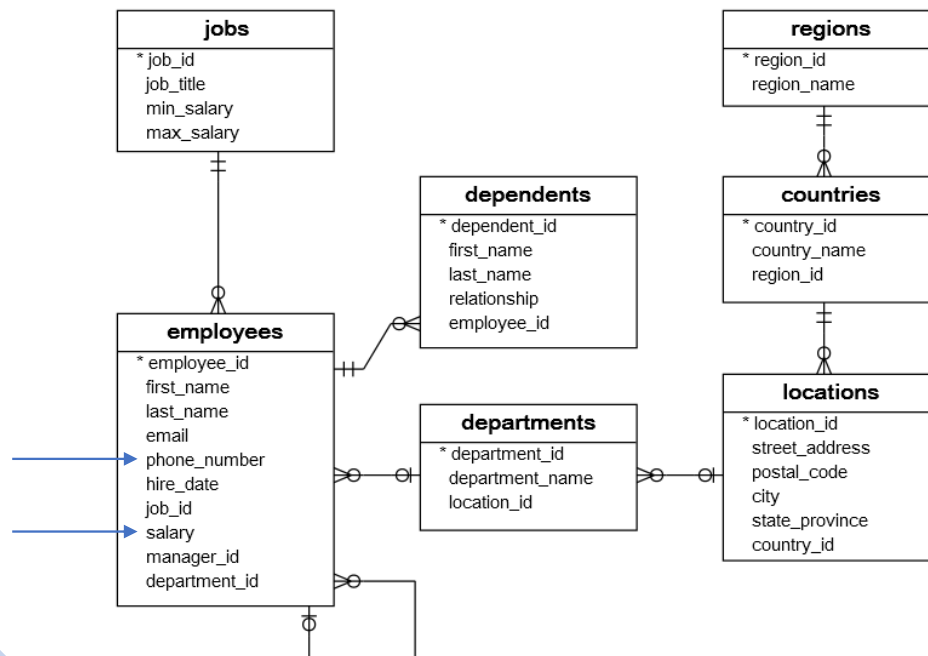
Region Failure



Distributed System



Security



email	phone_number	hire_date	job_id	salary
steven.king@sqltutorial.org	515.123.4567	1987-06-17	4	24000
neena.kochhar@sqltutorial.org	515.123.4568	1989-09-21	5	17000
lex.de haan@sqltutorial.org	515.123.4569	1993-01-13	5	17000
alexander.hunold@sqltutorial.org	590.423.4567	1990-01-03	9	9000
bruce.ernst@sqltutorial.org	590.423.4568	1991-05-21	9	6000
david.austin@sqltutorial.org	590.423.4569	1997-06-25	9	4800
valli.pataballa@sqltutorial.org	590.423.4560	1998-02-05	9	4800
diana.lorentz@sqltutorial.org	590.423.5567	1999-02-07	9	4200
nancy.greenberg@sqltutorial.org	515.124.4569	1994-08-17	7	12000
daniel.faviet@sqltutorial.org	515.124.4169	1994-08-16	6	9000
john.chen@sqltutorial.org	515.124.4269	1997-09-28	6	8200
ismael.sciarra@sqltutorial.org	515.124.4369	1997-09-30	6	7700
jose manuel.urman@sqltutorial.org	515.124.4469	1998-03-07	6	7800
luis.popp@sqltutorial.org	515.124.4567	1999-12-07	6	6900

같이 일하는 동료 전화번호는 찾을 수
있지만 그 동료의 급여 정보는 숨기려면?
아니 사장 급여는 숨기려면?;;



DBMS (Database Management System)

- 저장
- 검색
- 동시접속
- 확장
- 백업
- 보안
- ...



ACID Properties

ACID

Atomicity

원자성

Consistency

일관성

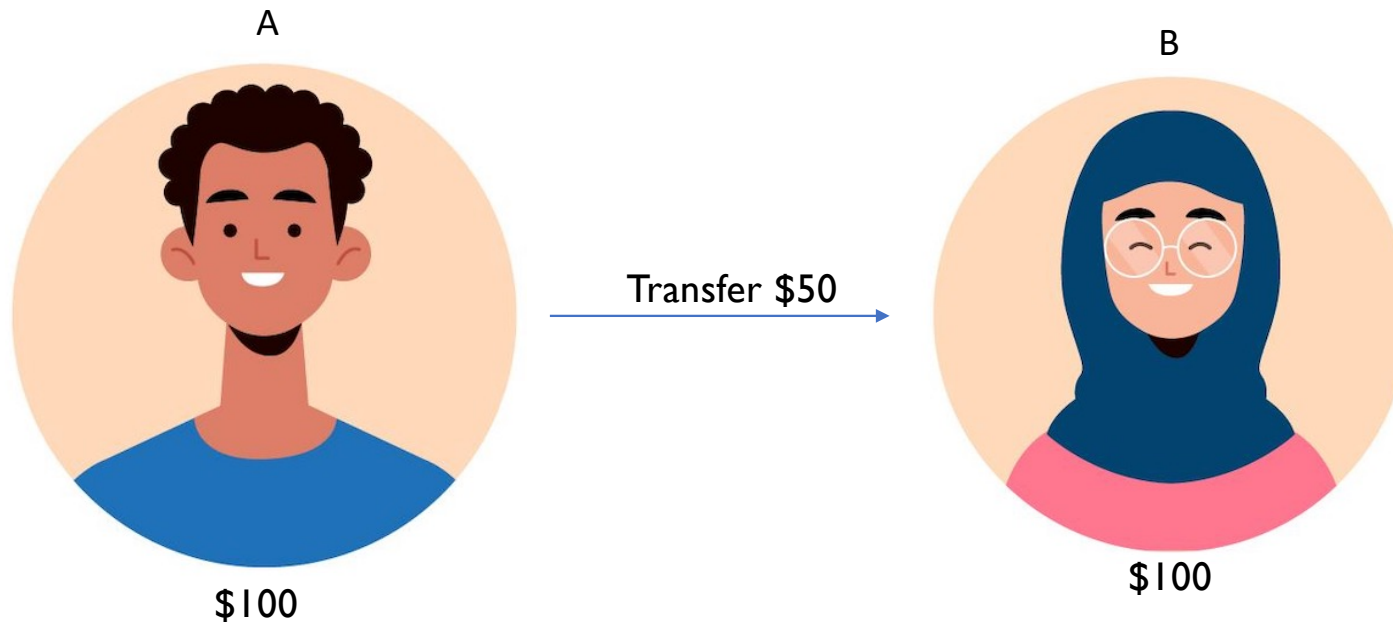
Isolation

독립성

Durability

지속성

Atomicity

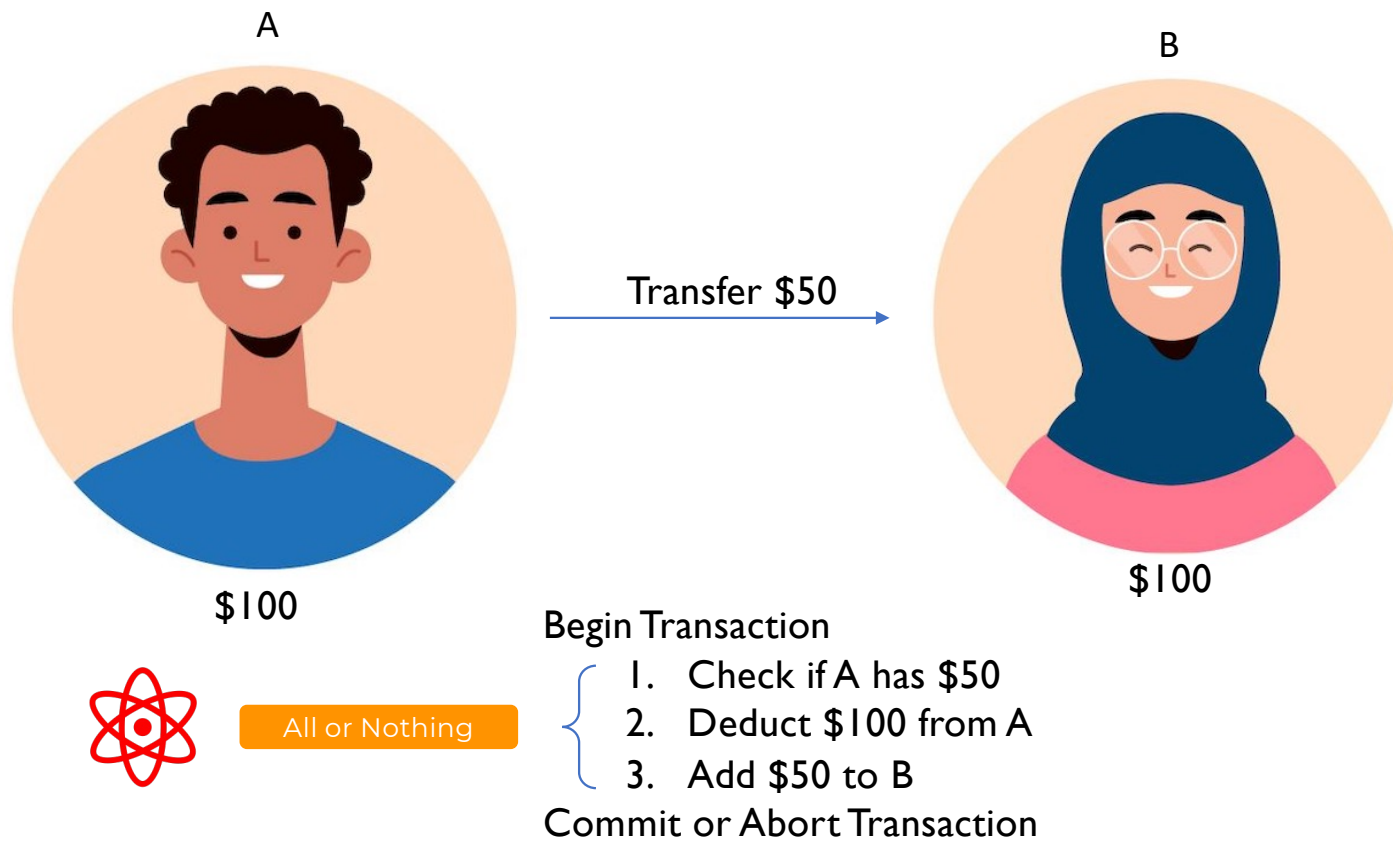


1. Check if A has \$50
2. Deduct \$50 from A
3. Add \$50 to B

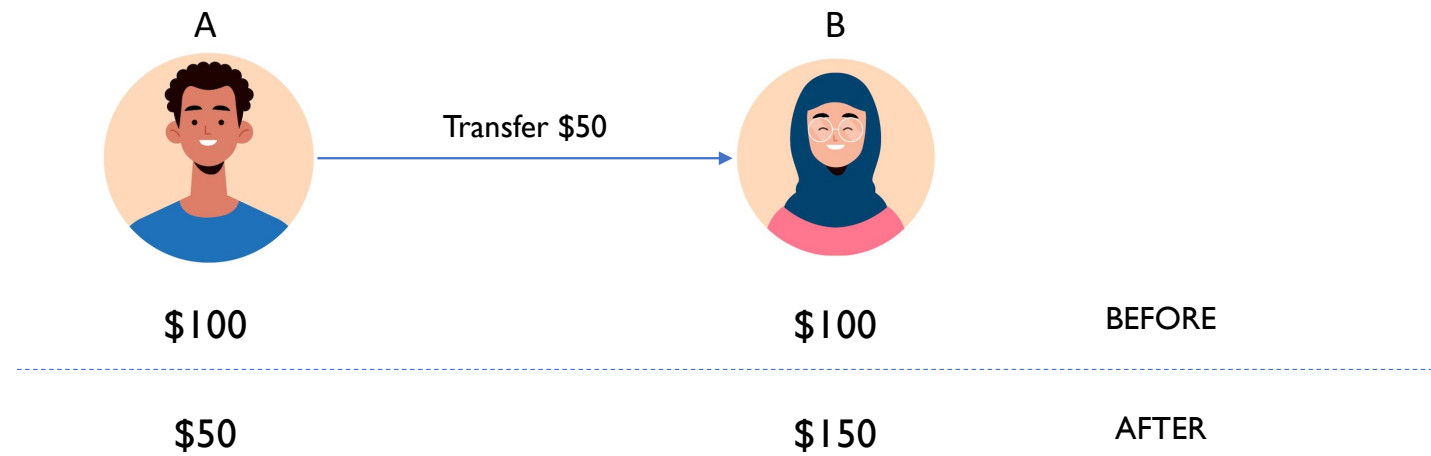
← Failed to execute



Atomicity



Consistency



Consistency

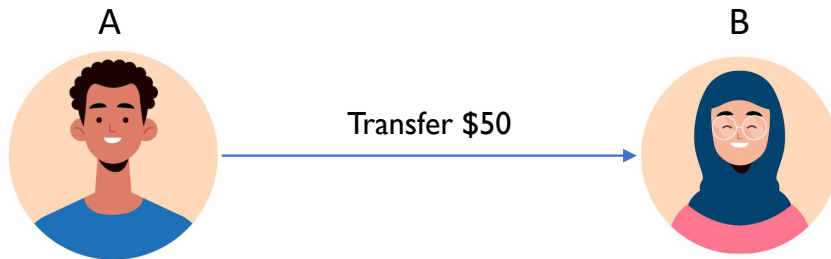
employee_id	first_name	last_name	email	phone_number	hire_date	job_id
100	Steven	King	steven.king@sqltutorial.org	515.123.4567	1987-06-17	4
101	Neena	Kochhar	neena.kochhar@sqltutorial.org	515.123.4568	1989-09-21	5
102	Lex	De Haan	lex.de haan@sqltutorial.org	515.123.4569	1993-01-13	5

5?



job_id	job_title	min_salary	max_salary
1	Public Accountant	4200	9000
2	Accounting Manager	8200	16000
3	Administration Assistant	3000	6000
4	President	20000	40000

Isolation



The total amount that this bank manages

	A	B	Total
BEFORE	\$100	\$100	\$200
AFTER	\$50	\$150	\$200



Transaction 1

- t_1 1. Check if A has \$50
- t_2 2. Deduct \$50 from A
- t_3 3. Add \$50 to B

	A	B	Total
INTERMEDIATE	\$50	\$100	\$150

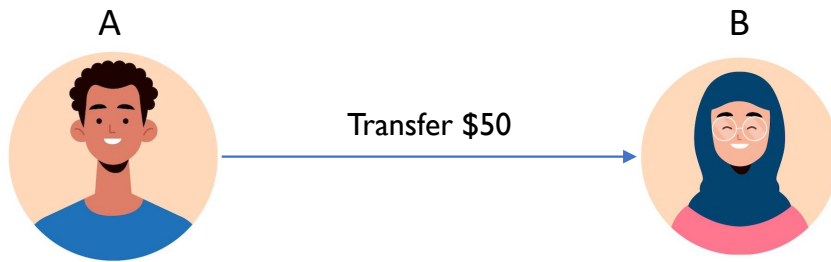


Transaction 2

Check balance A + balance B



Isolation



	A	B	Total
BEFORE	\$100	\$100	\$200
AFTER	\$50	\$150	\$200

option 1

Transaction 1

- t_1 1. Check if A has \$50
- t_2 2. Deduct \$50 from A
- t_3 3. Add \$50 to B

Transaction 2

- t_4 Check balance A + balance B

or

option 2

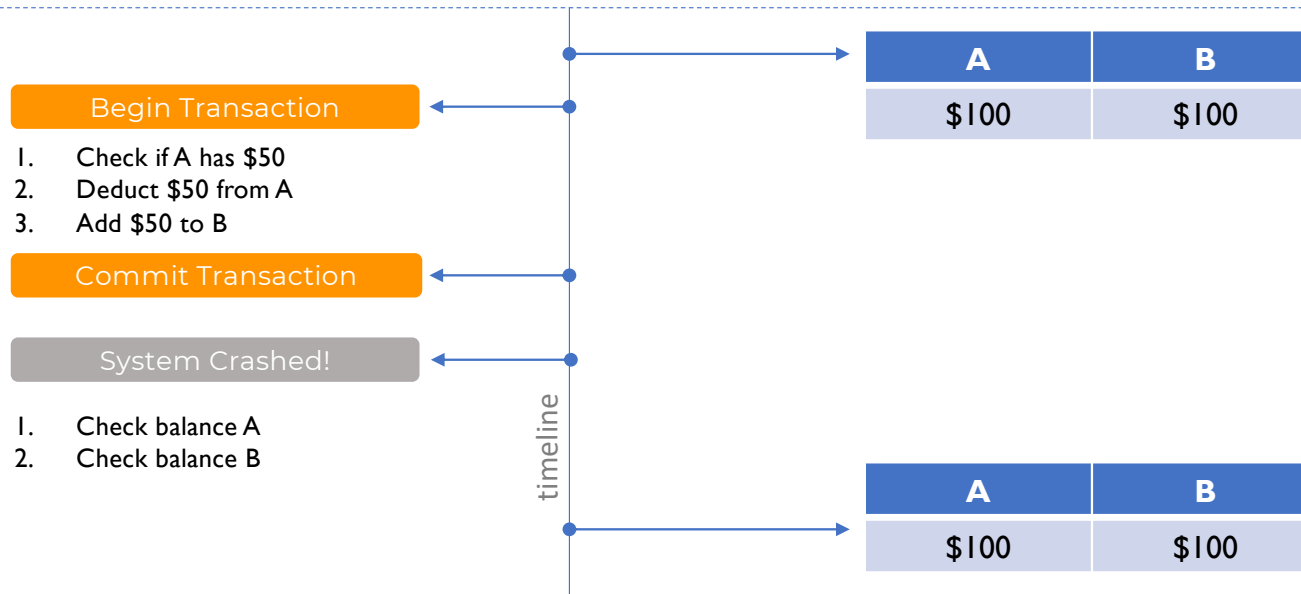
Transaction 2

- t_1 Check balance A + balance B

Transaction 1

- t_2 1. Check if A has \$50
- t_3 2. Deduct \$50 from A
- t_4 3. Add \$50 to B

Durability

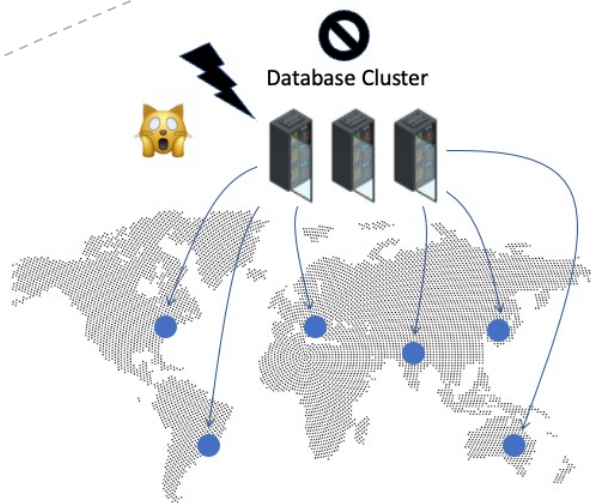
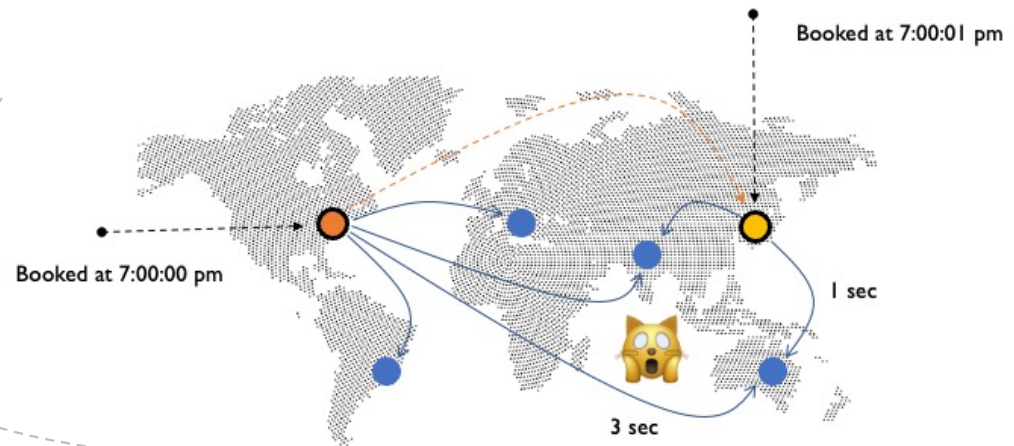


Write to memory vs. disk
Transactional Log



Perspective

- 시공간
- 혼자 vs. 우리



Transaction 1

- t_1 1. Check if A has \$50
- t_2 2. Deduct \$50 from A
- t_3 3. Add \$50 to B



Transaction 2

Check balance A + balance B

