

키값 저장소 설계

이번 챕터에서 키값 저장소 설계를 통해 무엇을 알기 원하는 걸까?

단일 저장소라면 우리가 아는 Map 과 다르지 않을 것.

그렇다면...?

분산 환경에서의 키값 저장소 설계가 이 챕터에서 우리가 집중해야되는 부분이 아닐까?

요구사항을 살펴보면...

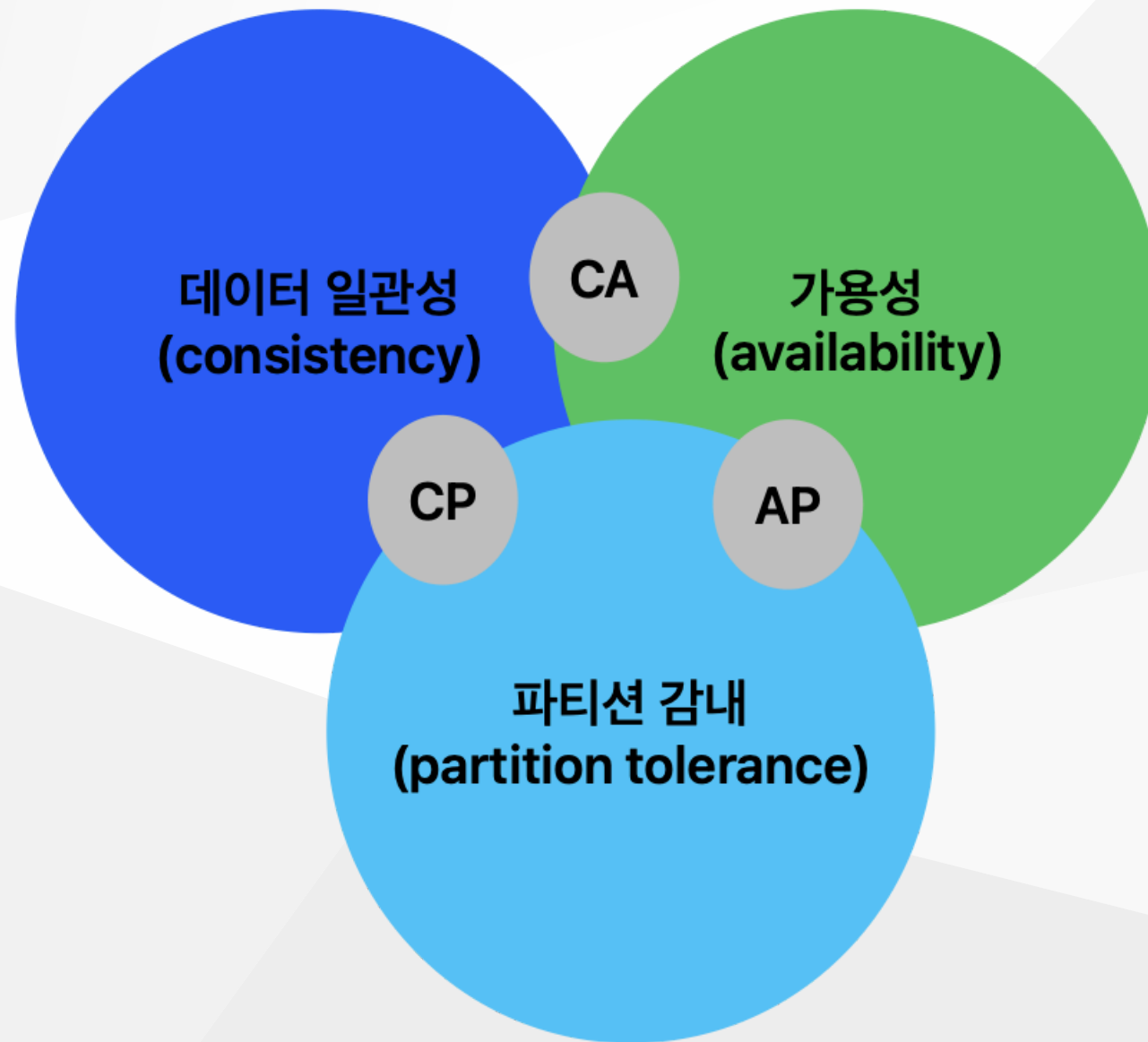
- 키-값 쌍의 크기는 10KB 이하
- 큰 데이터를 저장할 수 있어야 한다.
- ☒ 높은 가용성을 제공. 따라서 시스템은 설사 장애가 있더라도 빨리 응답
- ☒ 높은 규모 확장성 제공. 따라서 트래픽 양에 따라 자동적으로 서버 증설/삭제 이루어진다.
- ☒ 데이터 일관성 수준은 조정 가능
- ☒ 응답 지연시간(latency)이 짧아야 한다.





분산 환경에서의 키값 저장소 설계에 관한 이론

그럼 분산 키-값 저장소를 이해할 때는 CAP(consistency, Availability, Partition Tolerance theorem)

CAP??

- 데이터 일관성
 - 어떤 노드에 접속했느냐에 관계없이 언제나 같은 데이터를 보게 되어야 한다.
- 가용성
 - 시스템이 무너지더라도 클라이언트는 항상 응답을 받아야 한다.
- 파티션 감내
 - 통신 장애가 발생하더라도 시스템은 문제가 없어야 한다.



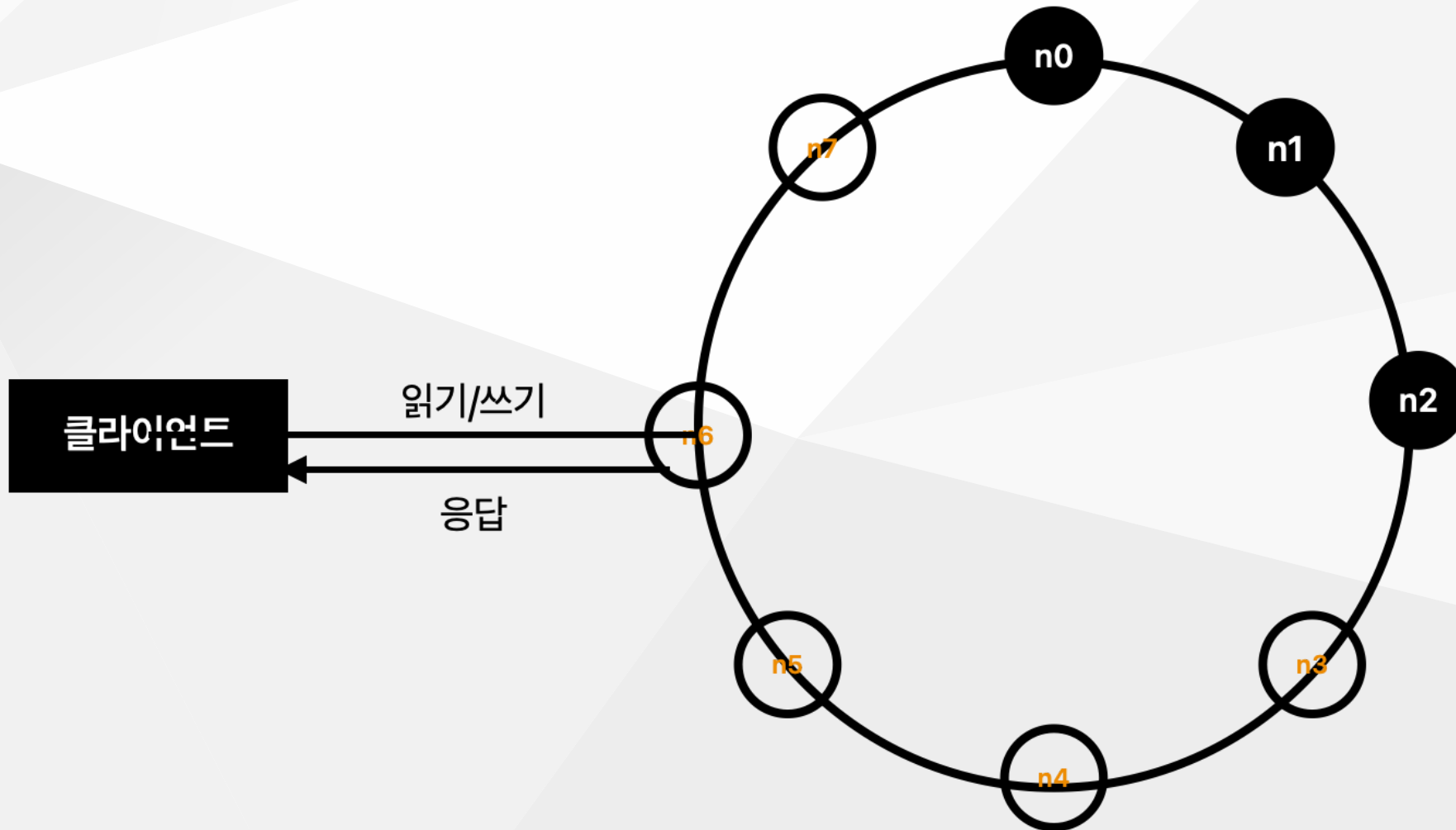
-  높은 가용성을 제공. 따라서 시스템은 설사 장애가 있더라도 빨리 응답
-  높은 규모 확장성 제공. 따라서 트래픽 양에 따라 자동적으로 서버 증설/삭제 이루어진다.
-  데이터 일관성 수준은 조정 가능
-  응답 지연시간(latency)이 짧아야 한다.

위 요구사항을 해결하기 위해서는 우리가 알아야할 지식이 있다.

하나씩 살펴보자

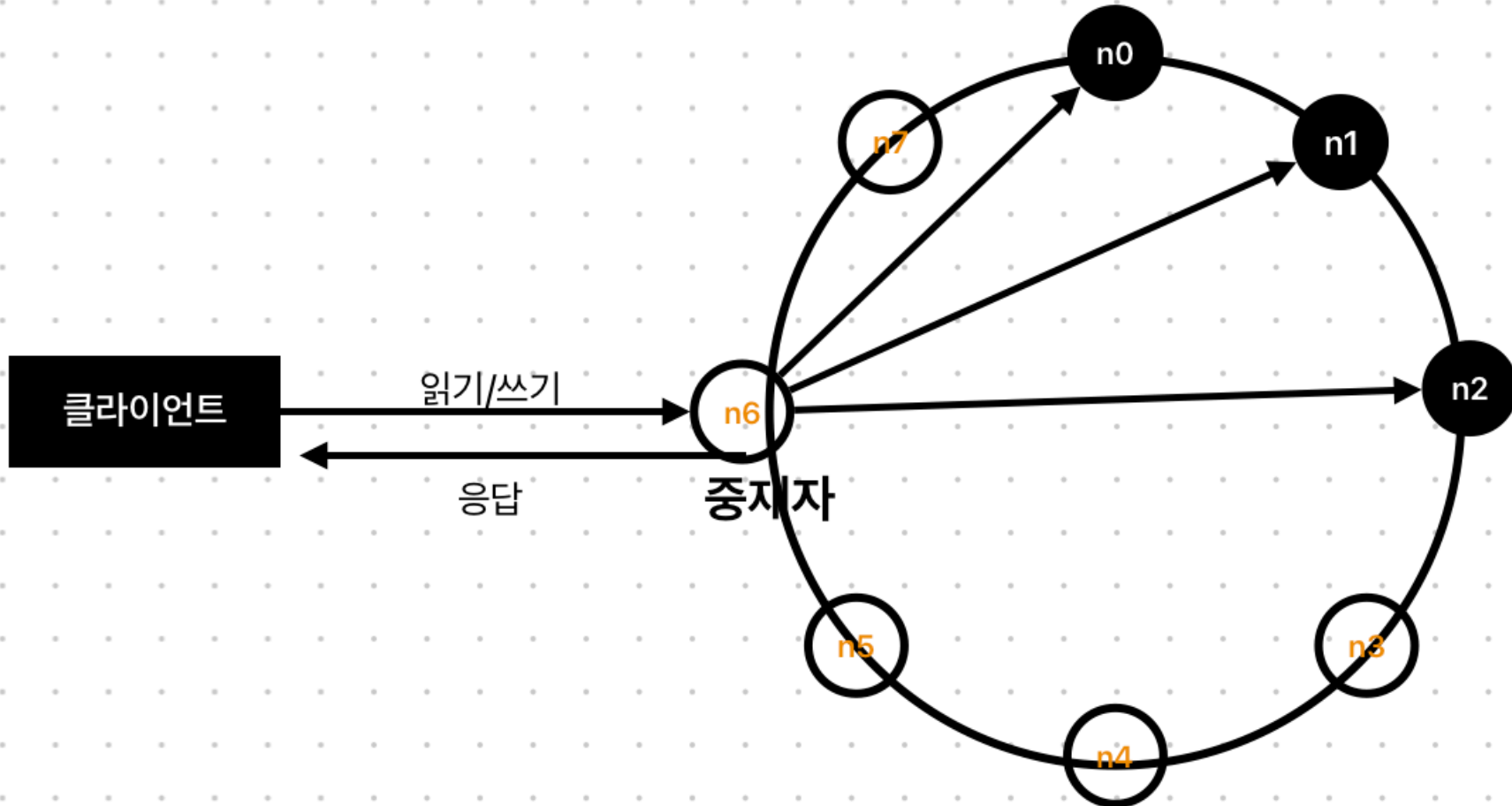
기술	무슨 목표/문제를 위해?
안정 해시를 사용해 서버들의 부하 분산	
데이터를 여러 데이터센터에 다중화	
버저닝 및 벡터 시계를 사용한 충돌 해소	
안정해시	
정속수 합의(quorum consensus)	
느슨한 정속수 프로토콜(sloppy quorum)과 단서 후 임시 위탁(hinted handoff)	
머클 트리(Merkel tree)	

안정 해시를 사용해서 서버들의 부하 분산으로 무엇을 할 수 있나?



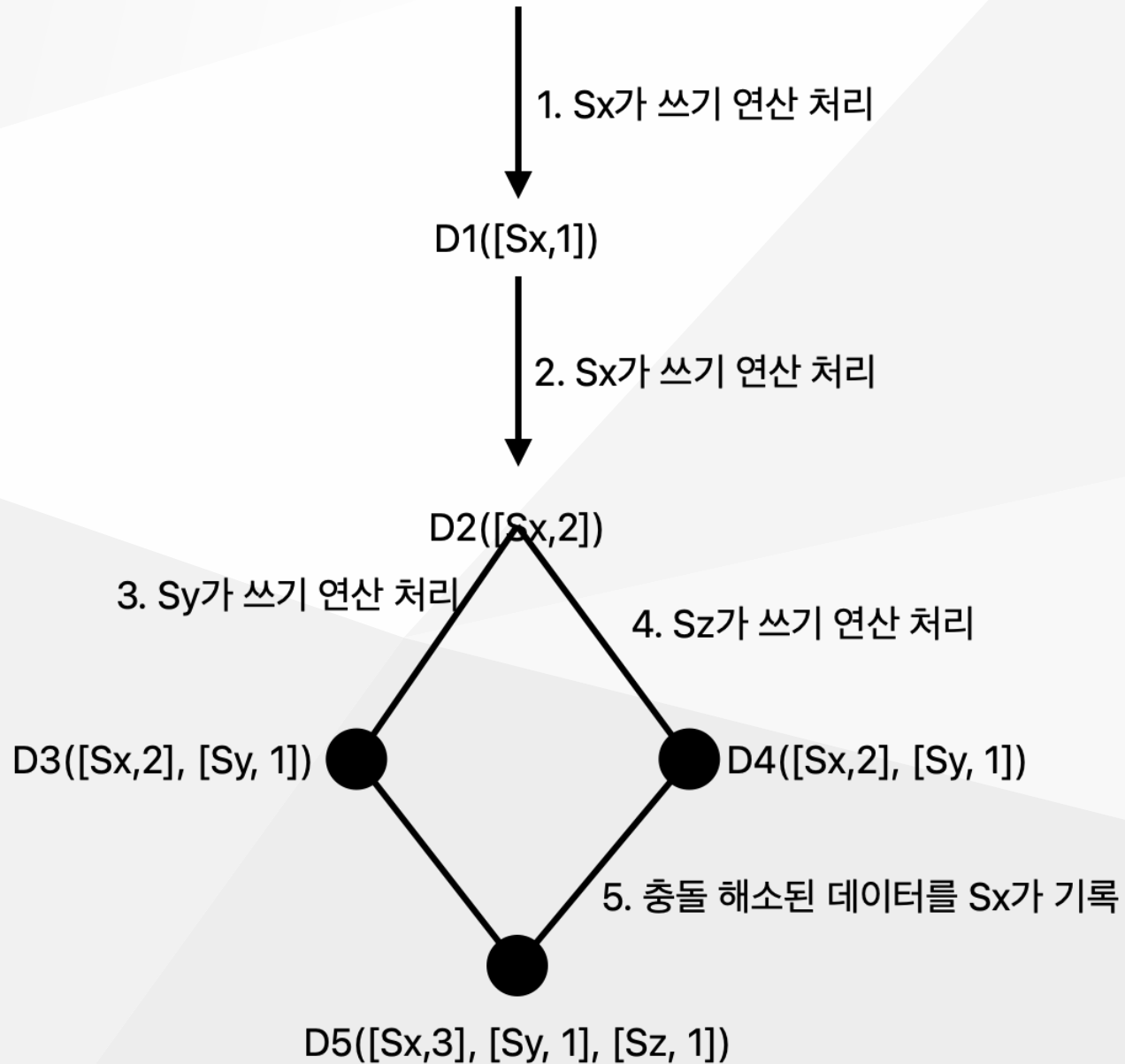
기술	무슨 목표/문제를 위해?
안정 해시를 사용해 서버들의 부하 분산	✓ 대규모 데이터 저장
데이터를 여러 데이터센터에 다중화	
버저닝 및 벡터 시계를 사용한 충돌 해소	
안정해시	
정속수 합의(quorum consensus)	
느슨한 정속수 프로토콜(sloppy quorum)과 단서 후 임시 위탁(hinted handoff)	
머클 트리(Merkel tree)	

데이터를 여러 데이터센터에 다중화 한다. 라는 것은 무슨 의미일까?



기술	무슨 목표/문제를 위해?
안정 해시를 사용해 서버들의 부하 분산	대규모 데이터 저장
데이터를 여러 데이터센터에 다중화	✓ 읽기 연산에 대한 높은 가용성 보장
버저닝 및 벡터 시계를 사용한 충돌 해소	
안정해시	
정속수 합의(quorum consensus)	
느슨한 정속수 프로토콜(sloppy quorum)과 단서 후임시 위탁(hinted handoff)	
머클 트리(Merkel tree)	

버저닝 및 벡터 시계를 사용한 충돌 해소



이게 무슨 말이야?

어렵게 생각하기 보다는 한 노드에 쓰기 연산 실행시 충돌이 발생할 경우 이것을 어떻게 해결할 것인지에 대한 해결 기법일 뿐... 자세히 알면 좋지만 몰라도 뭐..

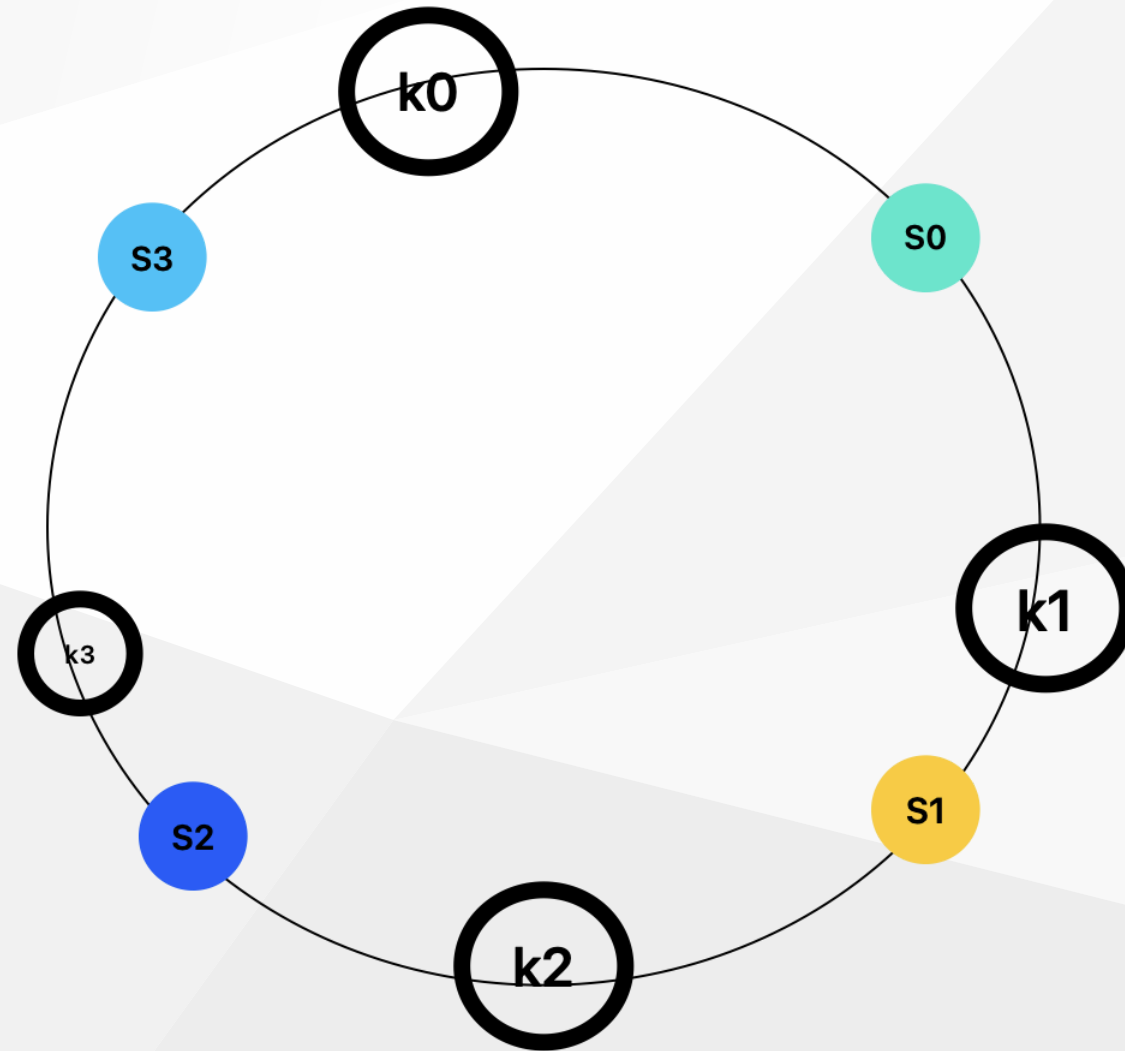
다만, 이런 쓰기연산에 대비한 기법이 무엇이 있는지 알면 좋다.

낙관적 잠금 기법과 비관적 락도 같은 내용일 수 있다고 생각한다.

즉, 버저닝 및 벡터 시계를 사용한 충돌 해소 기술은 쓰기 연산을 해결하기 위해 사용.

기술	무슨 목표/문제를 위해?
안정 해시를 사용해 서버들의 부하 분산	대규모 데이터 저장
데이터를 여러 데이터센터에 다중화	읽기 연산에 대한 높은 가용성 보장
버저닝 및 벡터 시계를 사용한 충돌 해소	✓ 쓰기 연산에 대한 높은 가용성 보장
안정해시	
정속수 합의(quorum consensus)	
느슨한 정속수 프로토콜(sloppy quorum)과 단서 후 임시 위탁(hinted handoff)	
미크로 트러거(Micro-trigger)	

안정해시는 분산 키-값 저장소에서 어떻게 활용될 수 있을까?



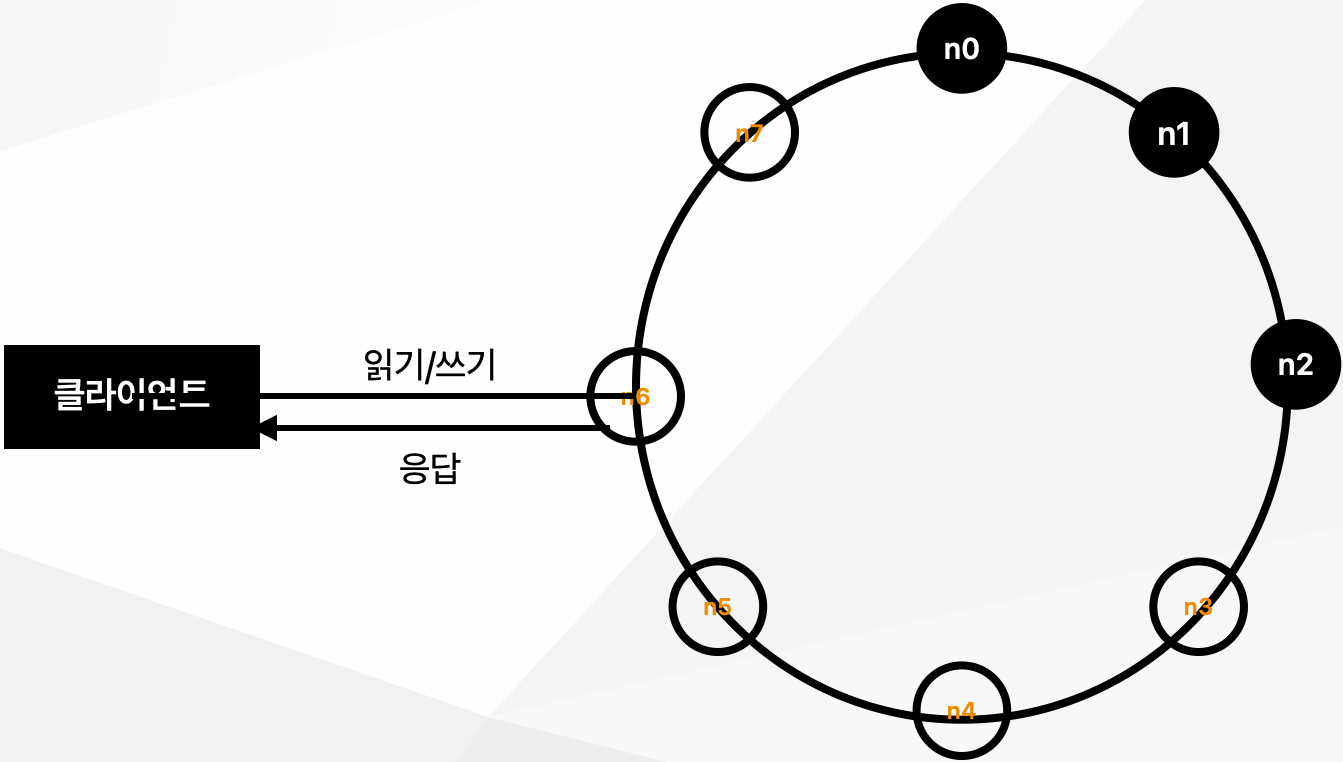
만약 1T(foo.mp4) 크기를 데이터 구조라면 어떻게 저장할 수 있을까?

각각의 노드에 분할해서 저장할 수 있을 것 - **데이터 파티션**

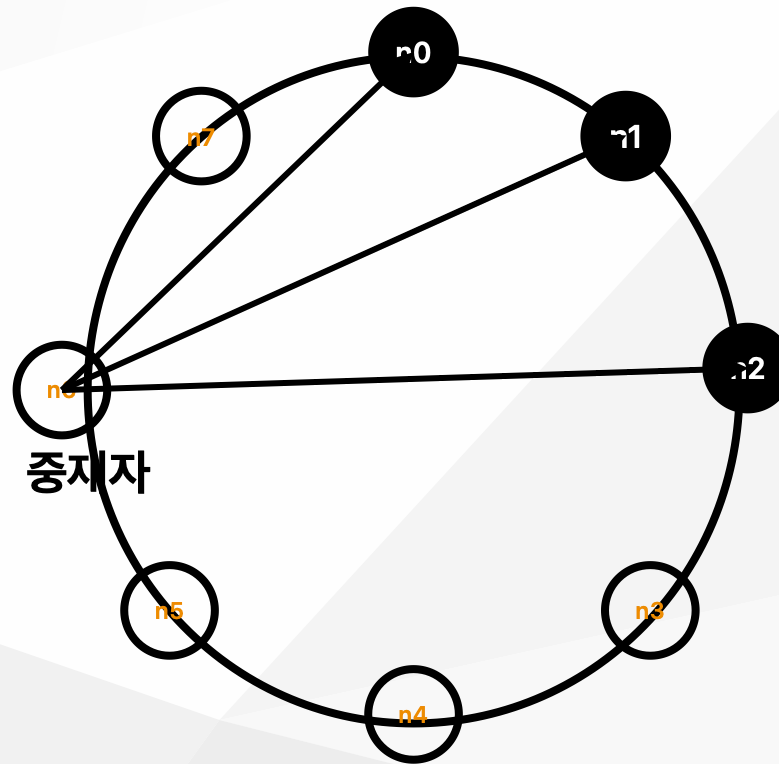
(foo.mp4-1, 분할된 데이터 100G), (foo.mp4-2, 분할된 데이터 100G) ...

또는..

데이터 다중화



데이터 일관성



정족수 합의(quorum consensus) 라는 용어가 등장하는데...

$W=1$ 이라는 것에 대한 의미.

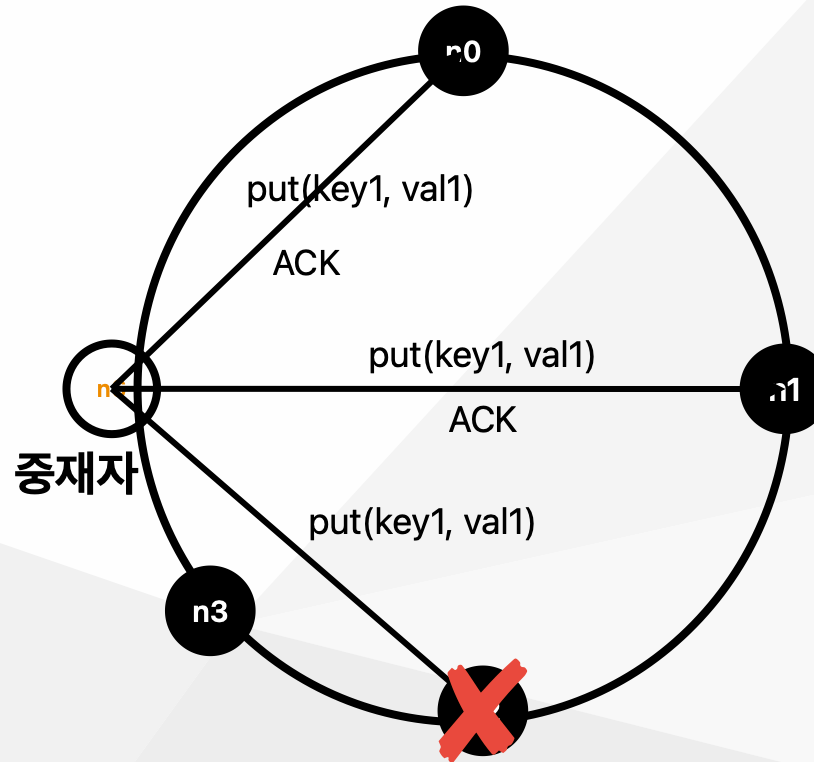
$R=1$ 이라는 것에 대한 의미를 이해할 필요가 있다.

안정해시를 사용함으로써

데이터를 파티션할 수 있고, 다중화 할 수 있고, 일관성을 지킬 수 있다.

기술	무슨 목표/문제를 위해?
안정해시	✅ 데이터 파티션, 다중화, 일관성
느슨한 정족수 프로토콜(sloppy quorum)과 단서 후임시 위탁(hinted handoff)	
머클 트리(Merkel tree)	

느슨한 정족수 프로토콜(sloppy quorum)과 단서 후 임시 위탁(hinted handoff) ?



만약 n2 가 장애가 발생하면, 임시로 n3가 대신 처리해주고, n2가 복구되면 일관 반영하여 데이터 일관성 보장

기술	무슨 목표/문제를 위해?
안정해시	데이터 파티션, 다중화, 일관성
느슨한 정족수 프로토콜(sloppy quorum)과 단서 후 임시 위탁(hinted handoff)	✅ 일시적 장애 처리
머클 트리(Merkel tree)	

Bloom filter

<https://namu.wiki/w/블룸 필터>

SSTable

<https://www.igvita.com/2012/02/06/sstable-and-log-structured-storage-leveldb/>