

6장 파티셔닝

- 샤딩
- 파티션
- 부하 분산
- 파티션을 나누는 방법
 1. 수평
 2. 수직
 3. 해시 분할
- 파티셔닝을 지원하는 DB
- 대용량 데이터셋을 파티셔닝하는 방법
- 데이터 색인과 파티셔닝이 상호작용하는 방법
- 클러스터에서 노드를 제거할때 필요한 재균형화와 분산 질의에 대한 내용을 소개한다.
 - 파티셔닝과 복제
 - 키-값 데이터 파티셔닝
 - 장점
 - 데이터가 고르게 분포될 수 있다.
 - 키 범위 기준 파티셔닝
 - 장점
 - 색인 처럼 사용할 수 있음
 - 정렬
 - 단점
 - 핫스팟 유발 가능성
 - 해시값 기본 파티셔닝

- 장점
 - 쏠림, 핫스팟 없이 균일하게 질의 분산
 - 단점
 - 범위 파티션 및 질의 불가
- 핫스팟 완화
 - 키 분산방법
- 파티셔닝과 보조색인
 - 용어 파티셔닝 색인(전역 색인)
 - 색인된 값을 사용해서 보조 색인을 별도로 파티셔닝
 - 보조 색인 항목은 기본키의 모든 파티션에 있는 레코드를 포함할 수 있음
 - 문서를 쓸 때는 여러개를 갱신해야함
 - 읽을 때는 단일 파티션에서 실행할 수 있음
 - 문서 파티셔닝 색인(지역 색인)
 - 보조 색인을 기본키와 값이 저장된 파티션에 저장
 - 쓸 때는 파티션 하나만 갱신하면 됨
 - 읽을 때는 모든 파티션에 걸쳐scatter getter를 실행해야함
- 파티션 재균형화
 1. 부하가 노드들 사이에서 균일하게 분배
 2. 재균형화 도중에도 R/W가능
 - 재균형화 전략
 - 해시값에 mod N연산
 - 파티션 개수 고정
 - 동적파티셔닝

- 노드비례 파티셔닝
- 운영
 - 자동 재균형화
 - 수동 재균형화
- 요청 라우팅
 - 아무노드에 접속
 - 라우팅 계층에서 처리
 - 클라이언트에 미리 전달