

셔플조인

Mapper - tableA

Mapper - tableB

shuffle –
tableA=tableB

Reducer –
join key 1

Reducer –
join key 1

맵조인

Mapper – tableA

In Memory
- tableB

셔플조인

- ① 매퍼에서 각 테이블을 읽고 ② 셔플 단계에서 조인되는 키를 기준으로 파티셔닝후 ③ 셔플을 진행하여 각 리듀서에서 조인을 수행
- 어떤형태의 데이터 크기와 구성에도 사용 가능
- 가장 자원을 많이 사용하고 느린 조인 방식

맵조인 (브로드캐스트 조인, 맵사이드 조인)

- 작은 사이즈의 테이블이 메모리에 올라가고, 각 매퍼에서 조인을 수행후 결과를 반환하는 방식
- 가장 큰 테이블에서 가장 빠른 단일 스캔
- 작은 테이블은 메모리에 들어갈 정도로 작아야 함

-- 맵조인 사용여부 설정, 3개 이상의 테이블을 조인할 때 맵조인 사용여부 설정하는 옵션

```
hive> set hive.auto.convert.join=true;
```

```
hive> set hive.auto.convert.join.noconditionaltask=true;
```

```
hive> set hive.auto.convert.join.noconditionaltask.size=10000000;
```

-- 맵조인에 사용되는 테이블의 사이즈 설정 30MB 이하면 맵조인

```
hive> set hive.mapjoin.smalltable.filesize=30000000;
```

-- 맵조인의 기준 테이블 설정 힌트

```
Select /*+ MAPJOIN(b) */ a.key, a.value  
from a join b on a.key = b.key
```

정렬-병합-버킷 조인

- 테이블이 버킷팅 되어 있을 때 사용할 수 있는 조인
- 매퍼는 각 키가 인접한 특성을 이용해 효과적인 조인을 수행
- 어떤 크기의 테이블에서도 굉장히 빠름