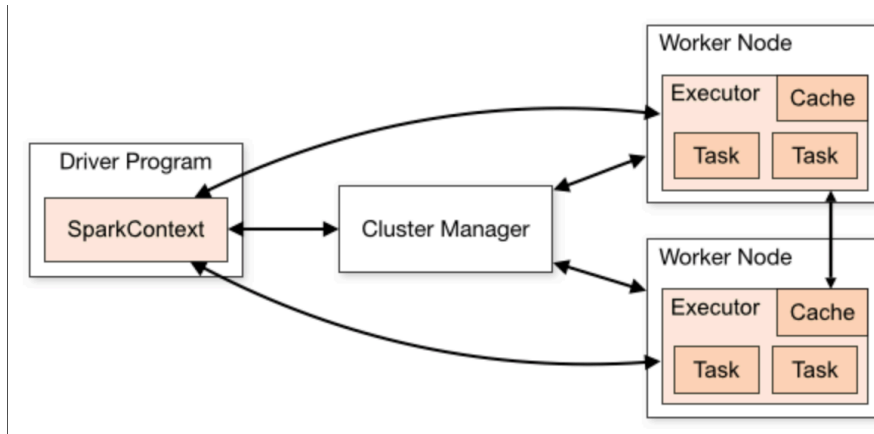
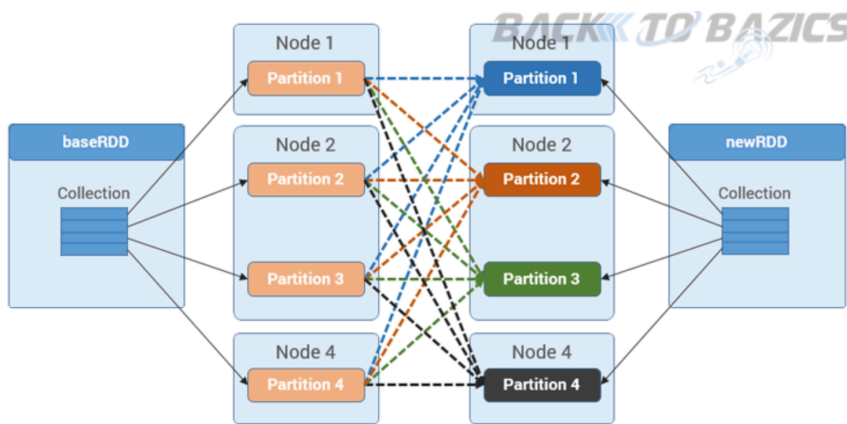


1. RDD : Resilient Distributed Datasets , SPARK 기본 DATA 의 집합 요소

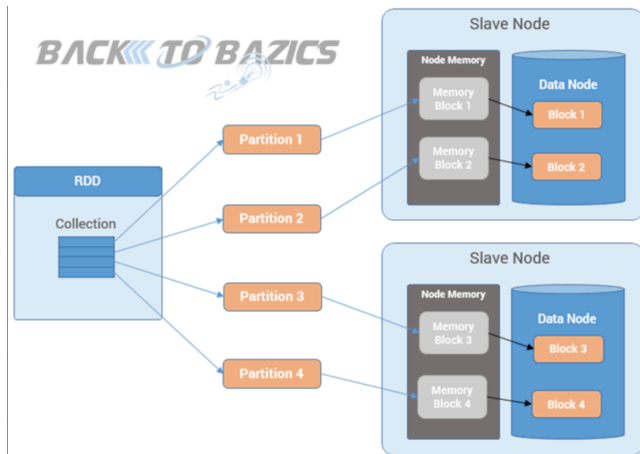
2. 클러스터 / JOB / Executor / 드라이버 프로그램



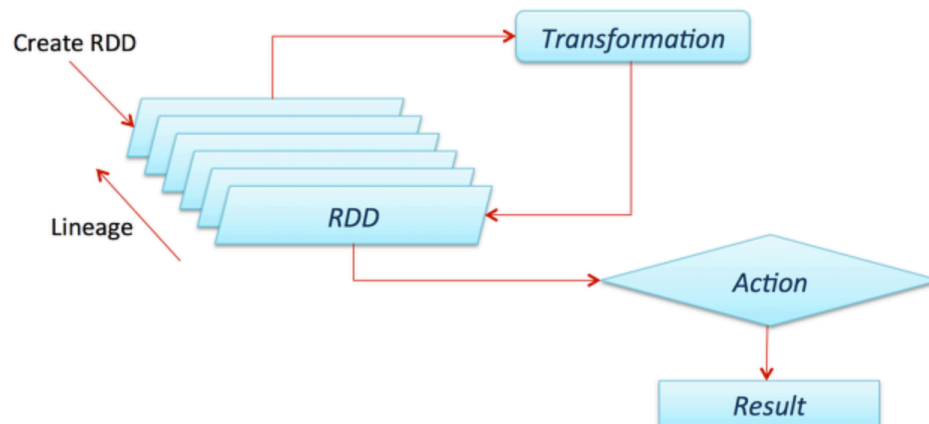
3. RDD의 불변성 : RDD 생성 -> Data 변경시 변경 방식 기억 --> 복구 시, 변경방식 역순으로 동일 RDD로 재생가능



4. 파티션 : RDD는 기본적으로 여러개의 파티션으로 나눌 수 있음



5. 트랜스포메이션, 액션, 지연 동작 최적화 : RDD자체를 변형하면 트랜스포메이션, 그 RDD 에서 다른 형태의 값을 취합하면 액션



6. 스파크 컨텍스트 생성 : `sc = SparkContext(master="local", appName="RDDCreateTest", conf=conf)`