Your Name:

1.

If hash partitioning needed, use: h(k) = k mod 4 + 1
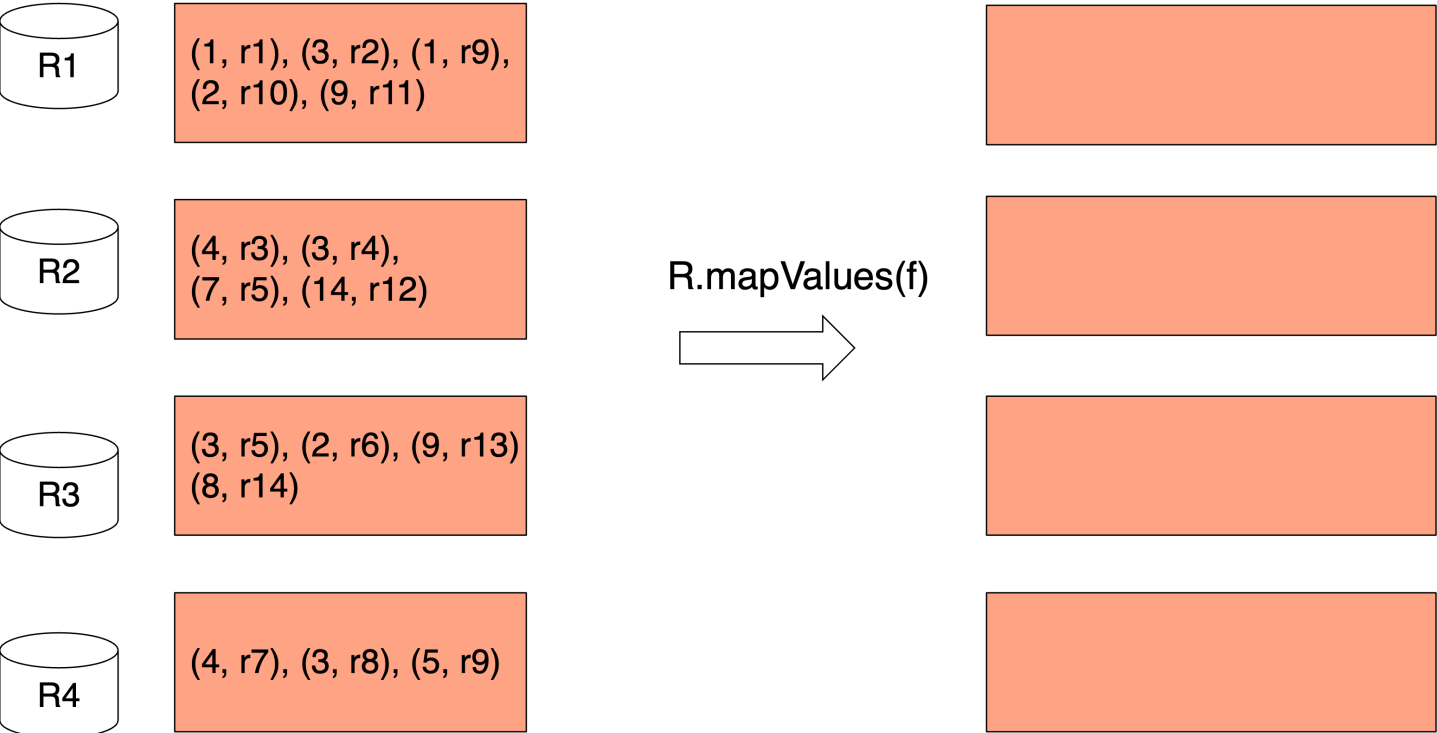
| R1 | (1, r1), (3, r2), (1, r9), (2, r10), (9, r11) |

| R2 | (4, r3), (3, r4), (7, r5), (14, r12) |

R.mapValues(f) →

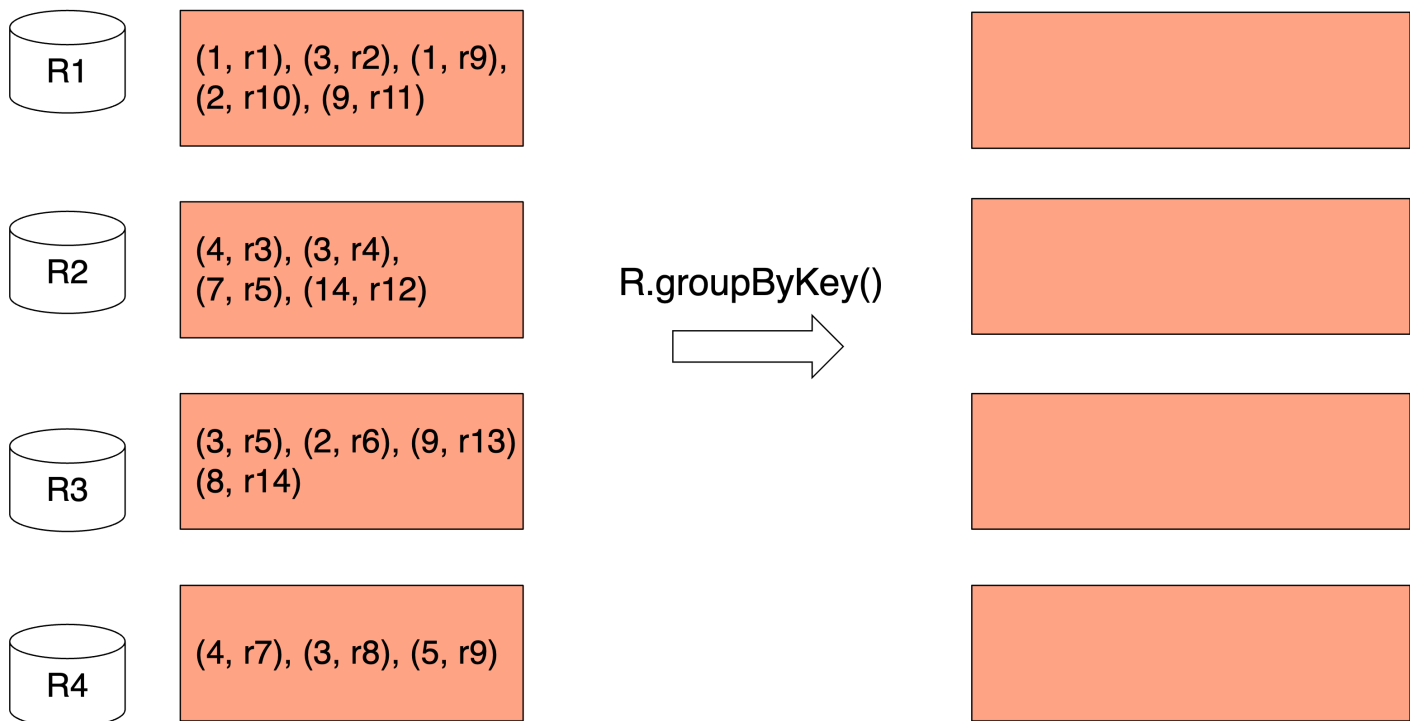| R3 | (3, r5), (2, r6), (9, r13) (8, r14) |

| R4 | (4, r7), (3, r8), (5, r9) |

2.

If hash partitioning needed, use: h(k) = k

| R1, S1 | R1: (1, r1), (3, r2) S1: (2, s1), (3, s2) |

| R2, S2 | R2: (4, r3), (3, r4) S2: (2, s3), (1, s4) |

R.join(S) →

| R3, S3 | R3: (3, r5), (2, r6) S3: (2, s5), (1, s6) |

| R4, S4 | R4: (4, r7), (3, r8) S4: (2, s7), (3, s8) |

3.

If hash partitioning needed, use: h(k) = k mod 4 + 1

R1

(1, r1), (3, r2), (1, r9), (2, r10), (9, r11)

R2

(4, r3), (3, r4), (7, r5), (14, r12)

R.groupByKey()

R3

(3, r5), (2, r6), (9, r13) (8, r14)

R4

(4, r7), (3, r8), (5, r9)
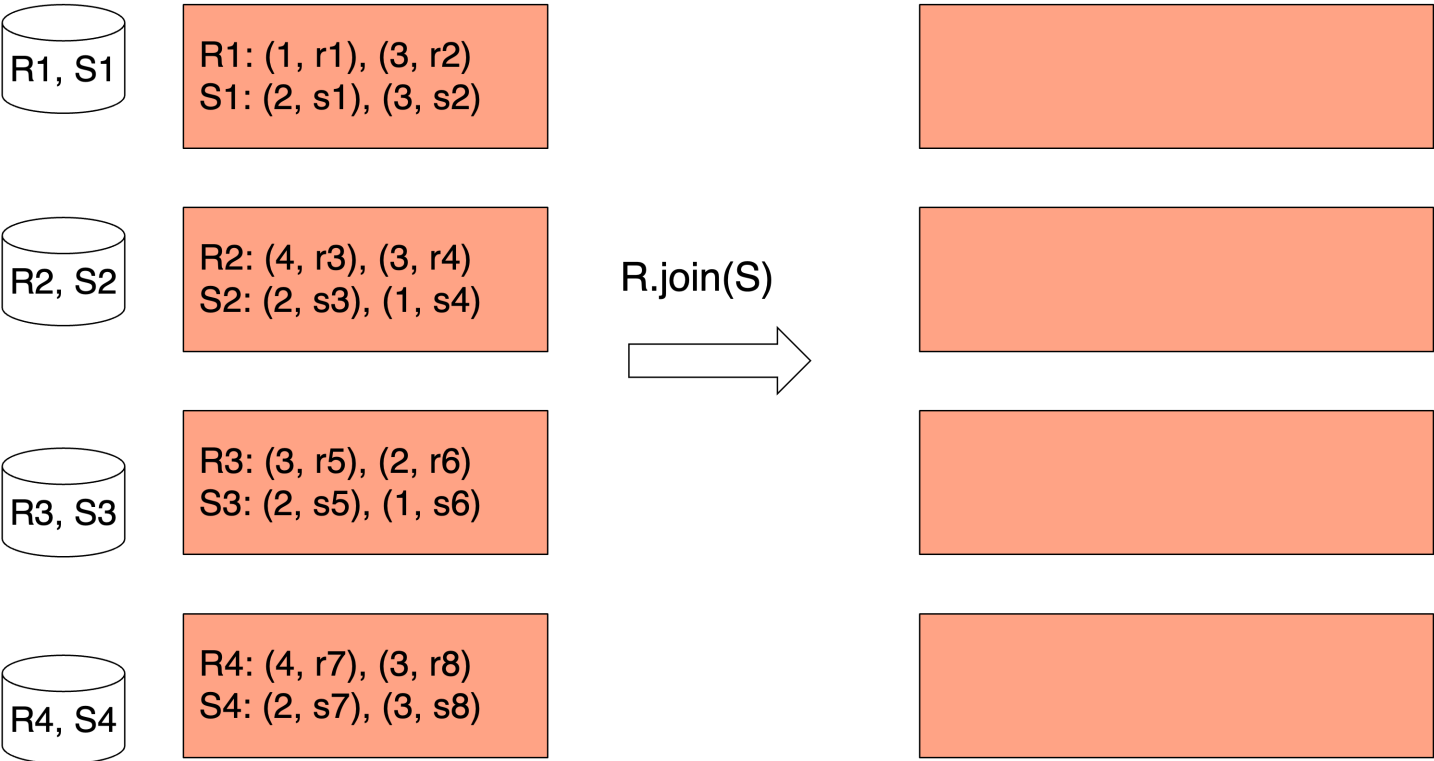
4.

lb takes an image file (e.g., 'img0.png') as input, and outputs:
[('cat', 'img0.png'), ('outside', 'img0.png), …]

R.flatmap(lb)

groupByKey()