2.

a. Not exactly following here… Is this just asking for the pseudo code effectively in each loop?

M = RDD of (row, (column, value))

v = RDD of (row, value)

mmult = join v to M

flat Map to get (row, M\_value \* v\_value)

reduceByKey (add up values)

b. A stage is a unit of execution. It consists of a set of parallel tasks with 1 task per partition.

c.

M = (1, (1, 16)), (1, (2, 2)), (1, (3, 3)), (1, (4, 13));

(2, (1, 5)), (2, (2, 11)), (2, (3, 10)), (2, (4, 8));

(3, (1, 9)), (3, (2, 7)), (3, (3, 6)), (3, (4, 12));

(4, (1, 4)), (4, (2, 14)), (4, (3, 15)), (4, (4, 1));

v = (1, 1), (2, 2), (3, 3), (4, 4)

d.

e. (m \* n) + n

f.