

6. (5 pts) Describe what is computed in the following pySpark program, i.e. what is in `mtuples`?

```
# we start with records of schema (k,v) = (animal_type, weight)

cbk = records.countByKey()

# add is a built in reduction function that adds elements
sbk = records.reduceByKey(add)

# join returns (k, (v1, v2)) tuple when (k, v1) is in self and (k, v2) is in other.
# RB note: instead of the following line, I would prefer sctuples = sbk.join(cbk).
# But, cbk is a python defaultdict, not an RDD.
sctuples = sbk.join(sess.sparkContext.parallelize(cbk.items()))

mtuples = sctuples.mapValues(lambda x : x[0]/x[1])
```

7. (5 pts) In an Intel processor architecture, how much faster is it to access 16 sequential and aligned words than 16 individual words from random addresses? You should assume a cold (empty) cache. Justify your answer and give specific parameters used in the calculation.