		Question 1
/12	(a)	**1 point** per type of bias named **2 points** per correct definition
/4	(b)	**1 point** per correct pairing
/4	(c)	**2 points** for a reasonable solution **2 points** for naming a reasonable stage
		Question 2
/6	(a)	**3 points** per valid benefit [productivity/iteration/data in memory/good for EDA/etc]
/2	(b)	**2 points** for mentioning a SQL, queries, tables, etc.
/4	(ci)	**4 points** for a mentioning in some way replication of data
/4	(cii)	**4 points** for mentioning storing lineage graphs, or recomputation, or something similar
/4	(d)	**2 points** for stating that transformations are lazy, not evaluated immediately, etc **2 points** for stating that actions are evaluated immediately, trigger computation, etc
/5	(e)	**1 point** for naming a valid transformation [map, filter, union, etc] **2 points** for a DAG which shows at least 2 RDDs **2 points** for having multiple partitions per RDD, with one child per parent
/5	(f)	**1 point** for naming a valid transformation [reduceByKey, groupByKey, etc] **2 points** for a DAG which shows at least 2 RDDs **2 points** for having multiple partitions per RDD, with multiple children per parent
		Question 3
/10	(a)	**1 point** per RDD in one iteration [start, unlabeled, unlabeled, middle] **1 point** per correct transformation label in one iteration [flatMap, map, reduceByKey] **2 points** for having the 2nd iteration represented **1 point** for having the fullOuterJoin between middle and result
/10	(b)	**4 points** for correctly stating 3 stages (**2 points** for stating 2 stages) **3 points** per boundary drawn through the reduceByKey transformations
/12	(c)	**4 points** for identifying .reduceByKey on line 6 **4 points** for identifying .flapMap (line 4) or .map (line 5) **4 points** for identifying .collect on line 12
/4	(d)	**4 points** for stating that it runs 1 job
/4	(e)	**4 points** for stating that it implements Word Count

		Question 4
/6	(a)	**1 point** per classifier mentioned [k-NN, Naive Bayes, Logistic Regression] **1 point** per correct categorization [structural, statistical, statistical]
/4	(b)	**1 point** for stating that words are the features in spam classification **1 point** for stating that urls visited are the features in ad click probability **2 points** for stating that k-NN works better for low dimensionality (or Curse of Dim)
/2	(c)	**2 points** for mentioning that it works well with large amounts of data
/3	(d)	**3 points** for mentioning we run on each class, and pick the most likely
		Question 5
/10	(a)	**10 points** for getting the correct answer (or close with just math errors) **5 points** for having some correct steps/formulas
/10	(b)	**10 points** for getting the correct answer (or close with just math errors) **5 points** for having some correct steps/formulas
/6	(c)	**6 points** for getting the correct answer (or close with just math errors) **3 points** for having some correct steps/formulas
/4	(d)	**2 points** for mentioning Laplace Smoothing **2 points** for mentioning getting more data (or other reasonable answer)
/5	(e)	**5 points** for a reasonable explanation that points to a specific type of bias, or issue **2 points** for just mentioning bias but not anything specific
		Question 6
/2	(a)	**0.5 points** (rounded down) per correct answer [volume, velocity, variety, veracity]
/2	(b)	**1 point** per correct answer [random process generating data, random sampling]
/2	(c)	**2 points** for stating that odds of machine going down approaches 1.0 as number of machines increases
/2	(d)	**2 points** for any feasible answer
/2	(e)	**1 points** per correct answer [sample, distribute computation, Spark, MR]