Dynamo Question

The ring formed by virtual nodes and their tokens are

n_{0a}	n_{1a}	n_{2a}	n_{0b}	n_{2b}	n_3	n_{1b}	n_4
5	20	35	50	60	75	85	95

1. [**4 points**] What is the preference list of key "**Science**"?

   Answer: (n2,n3,n1,n4)

   Explain: key 53 is managed by $n_{2b,}$ the three other nodes following it are n3,n1 and n4.

2. [**4 points**] Which node has the least number of keys? What are the keys on this node?

   Answer: n4 has the least data, with "Business" and "Law" stored on it

   Explain:

Node data allocation would look like:

Node	data
n0	Arts, Business, Education, Engineering, Law, Medicine(coordinator node)
n1	Business, Engineering(coordinator node), Science
n2	Arts(coordinator), Education(Coor. node), Engineering, Medicine, Science(coor. node)
n3	Arts, Education, Law(coor),Medicine,Science
n4	Business(coordinator node), Law

3. [**2 points]** Suppose all versions of the object with key "**Law**" have the same vector clock ([n_3, 10]), what do we know about the update history of this key?

   Answer: This suggests that the key has been updated 9 times after the initial insert, all updates are handled by n3;

4. [**5 points**] Now suppose all other nodes except $\mathbf{n_3}$ are available during the next update of key "**Law"**, what would be the vector clock of the new version? Which nodes would have the new version?

Answer: The preference list of "Law" is {n3,n1,n4 and n0} The node next in the preference list of key "Law" is n1. So the new version's vector clock will be ([n3,10],(n1,1]). All three nodes n1,n4 and n0 will have a copy of it

Neo4j Question

1. [**3 points**] Write a query to find the average passing mark of each course offered in 2018. Here passing mark means a mark that is 50 or above.

Sample Answer:

```
MATCH (t:Timetable{year:2018}) <-[e:ENROLLED]- (s:Student)
WHERE e.score >=50
RETURN t, avg(e.score)
```

2. [**6 points**] The problem domain has many constraints for nodes and their relationships. Maintaining such constraint is largely the responsibility of developers. One option is to write queries to periodically check if there is any violation. This part asks you to develop queries to test the following constraints.

a) A student cannot be the tutor of any course he/she is also currently enrolled in as a student.

Sample Answer:

```
MATCH (t:Timetable)<-[:ENROLLED]-(s:Student) -[:TUTORS]-> (:Lab)<-[:LABSCHEDULE]-(t:Timetable)
RETURN s
```

This query searches for students that are tutors and students of the same course. If the query returns any result, it indicates a violation of the rule. Note there are other ways to answer this question.

b) A student cannot enrol in more than one lab of the same course.

NO sample answer provided. There are many options, one would be searching for possible violations.

3. [**3 points**] We want to use the data to find candidate tutors of a given course. A candidate tutor is a student achieved HD (85 or above) in previous offerings of the same course. Now write a query to find candidate tutors for 2019 COMP5222 offerings.

NO sample answer provided

4. [**4 points**] Assuming no node property index has been set. Describe the execution plan of the following query

NO sample answer provided

5. [**6 points**] Write down the content of the following byte ranges in the relationship record at byte offset 34:

- byte 1~4: **1**
- byte 5~8: **10**
- byte 13~16: **0**
- byte 17~20: **2**
- byte 21~24: **Null or no data**
- type 25~28: **3**

6. [**3 points**] Which node has its record at byte offset ~~340~~ 300? Which records(s) are included in this node's doubly linked list of relationship records?

Sample Answer: Node st1's record has a byte offset 300; relationship records e1 and e2 are in its doubly linked list.

9-intersection Model Question

Answer: A covers B

Spatial index Question

Answer: The indexing method is Grid File

The index structure contains

- a 2 dimension array for bucket location: assume the buckets are named as 1~6, the actual array is [[1,2],[3,4],[5,6]]
- linear scales for grid line location: assume the range of x and y are both [0,10], the scales are
    - x:[0,5,10]
    - y:[0,4,8,10]