Semi-Structured Interview

Q1. Explain your understanding of the SCHEMA (in general). What is it and what is it used for?

Q2. What is the use of COLLECT keyword in Cypher?

Q3. In which other programming language/tool is there something similar to COLLECT? Does it remind you of something you've used before?

Q4. What is the use of WITH keyword in Cypher?

Q5. In which other programming language/tool is there something similar to WITH? Does it remind you of something you've used before?

Q6. Given this question : Write a Cypher query that returns each distinct node label together with its corresponding set of properties. Which of the following query is correct?

    1 . MATCH (n) RETURN DISTINCT labels(n) AS labels, keys(n) AS properties

    2 . MATCH (n) RETURN DISTINCT labels(n) AS labels, collect(distinct keys(n)) AS properties

    3. MATCH (n) RETURN DISTINCT labels(n) AS labels, collect(keys(n)) AS properties

Q7. Given this question : Write a Cypher query that returns each distinct node label together with its corresponding set of properties.
Given the following queries, explain the difference between them.

    1 . MATCH (n) RETURN DISTINCT labels(n) AS labels, keys(n) AS properties

    2 . MATCH (n) RETURN DISTINCT labels(n) AS labels, collect(distinct keys(n)) AS properties

    3. MATCH (n) RETURN DISTINCT labels(n) AS labels, collect(keys(n)) AS properties

Q8. Given this question : Write a Cypher query that returns each distinct node label together with its corresponding set of properties. Which of the following query is correct?

    1. MATCH (n) RETURN DISTINCT labels(n), collect(keys(n)) AS properties

    2. MATCH (n) WITH DISTINCT labels(n) AS labels, n RETURN labels, collect(keys(n)) AS properties

Q9. Given this question : Write a Cypher query that returns each distinct node label together with its corresponding set of properties.

Given the following queries, explain the difference between them.

    1. MATCH (n) RETURN DISTINCT labels(n), collect(keys(n)) AS properties

    2. MATCH (n) WITH DISTINCT labels(n) AS labels, n RETURN labels, collect(keys(n)) AS properties

Q10. Given this question : Write a Cypher query to return each pangenome, together with its associated genes.
Which of the following query is correct?

    1 . MATCH (p:Pangenome)--(f:Family)--(g:Gene) RETURN distinct p.name, g.name;

    2 . MATCH (g:Gene)-[:IS_IN_FAMILY]->(:Family)-[:IS_IN_PANGENOME]->
    (p:Pangenome)
    RETURN p.name as pangenomeName, COLLECT(g.name) AS genes

    3. MATCH (gene:Gene)-[:IS_IN_FAMILY]->(family:Family)-[:IS_IN_PANGENOME]->
    (pangenome:Pangenome) RETURN pangenome, gene

Q11. Given this question : Write a Cypher query to return each pangenome, together with its associated genes.
Given the following queries, explain the difference between them.

    1 . MATCH (p:Pangenome)--(f:Family)--(g:Gene) RETURN distinct p.name, g.name;

    2 . MATCH (g:Gene)-[:IS_IN_FAMILY]->(:Family)-[:IS_IN_PANGENOME]->
    (p:Pangenome)
    RETURN p.name as pangenomeName, COLLECT(g.name) AS genes

    3. MATCH (gene:Gene)-[:IS_IN_FAMILY]->(family:Family)-[:IS_IN_PANGENOME]->
    (pangenome:Pangenome)
    RETURN pangenome, gene

Q12. Given this question: Write a Cypher query to return each pangenome, together with its associated spots. State also how many answers it returns.
Are both of the following queries correct?
    1. MATCH (s:Spot)<-[:IS_IN_SPOT]-(:RGP)<-[:IS_IN_RGP]-(g:Gene)
    -[:IS_IN_FAMILY]->(:Family)-[:IS_IN_PANGENOME]->(p:Pangenome)
    RETURN p.name as pangenome, COLLECT(s.name) AS spots

    2. MATCH (sp:Spot)<-[:IS_IN_SPOT]-()<-[:IS_IN_RGP]-()-[:IS_IN_FAMILY]->
    ()-[:IS_IN_PANGENOME]->(p:Pangenome)
    RETURN distinct p.name, sp.name

Q13. Explain the answer of the previous question

Q14. If a query has the following return statements:

    RETURN a, COLLECT (b)

Would you assume based on this formulation that the names of b unique? Please motivate your answer.

Q15. Given this question : Write a Cypher query to return each pangenome and its associated modules that contain at least a pair of gene families that are more than 80% identical.
Which of the following query is correct?

    1. MATCH (p:Pangenome)<--()-->(m:Module)<--(f1:Family)<-[r:IS_SIMILAR]-
    (f2:Family) WHERE f1 <> f2 AND r.coverage >= 0.8 RETURN p.name

    2 .MATCH (p:Pangenome)<-[:IS_IN_PANGENOME]-(f:Family)-[:IS_IN_MODULE]->
    (m:Module),(p:Pangenome)<-[:IS_IN_PANGENOME]-(f:Family)-[sim:IS_SIMILAR]->
    (f2:Family) WHERE sim.coverage >= 0.8 return p.name as Pan-genome

Q16. Given this question : Write a Cypher query to return each pangenome and its associated modules that contain at least a pair of gene families that are more than 80% identical.
What is the difference between the following queries?

    1. MATCH (p:Pangenome)<--()-->(m:Module)<--(f1:Family)<-[r:IS_SIMILAR]-
    (f2:Family) WHERE f1 <> f2 AND r.identity >= 0.8 RETURN p.name

    2 . MATCH (p:Pangenome)<-[:IS_IN_PANGENOME]-(f:Family)-[:IS_IN_MODULE]->
    (m:Module),(p:Pangenome)<-[:IS_IN_PANGENOME]-(f:Family)-[sim:IS_SIMILAR]->
    (f2:Family) WHERE sim.identity >= 0.8 return p.name as Pan-genome

Q17. Given a query in which the pattern contains two nodes with the same label, for example:

    1 MATCH (p:Pangenome)<--()-->(m:Module)<--(f1:Family)<-[r:IS_SIMILAR]-
    (f2:Family)

    2 .MATCH (p:Pangenome)<-[:IS_IN_PANGENOME]-(f1:Family)-[:IS_IN_MODULE]->
    (m:Module),(p:Pangenome)<-[:IS_IN_PANGENOME]-(f1:Family)-[sim:IS_SIMILAR]->
    (f2:Family) WHERE sim.identity >= 0.8 return p.name as Pangenome

The two nodes f1 and f2 are different? Could they be the same node? Motivate your answer.

Q18. Given this question : Write a Cypher query that returns pairs of inter-pangenome families that both contain annotations and that are more than 80% identical.
Which of the following query is correct?

1. MATCH (p:Pangenome)<-[:IS_IN_PANGENOME]-(f:Family)
MATCH (p2:Pangenome)<-[:IS_IN_PANGENOME]-(f2:Family)
MATCH (f)-[s:IS_SIMILAR]->(f2) WHERE s.identity >= 0.8 and p <> p2
and f is not null and f2 is not null RETURN f.name, f2.name

2. MATCH(p1:Pangenome)-[:IS_IN_PANGENOME]-(f1:Family)-[r1:IS_SIMILAR]-
(f2:Family)-[:IS_IN_PANGENOME]-(p2:Pangenome)
WHERE r1.identity >= 0.8
AND p1.taxid <> p2.taxid
AND f1.annotation IS NOT NULL
AND f2.annotation IS NOT NULL
RETURN f1.name, f2.name

Q19. Given this question : Write a Cypher query that returns pairs of inter-pangenome
families that both contain annotations and that are more than 80% identical.
What is the difference between the following queries?

1. MATCH (p:Pangenome)<-[:IS_IN_PANGENOME]-(f:Family)
MATCH (p2:Pangenome)<-[:IS_IN_PANGENOME]-(f2:Family)
MATCH (f)-[s:IS_SIMILAR]->(f2) WHERE s.identity >= 0.8 and p <> p2
and f is not null and f2 is not null RETURN f.name, f2.name

2. MATCH (p1:Pangenome)-[:IS_IN_PANGENOME]-(f1:Family)-[r1:IS_SIMILAR]-
(f2:Family)-[:IS_IN_PANGENOME]-(p2:Pangenome)
WHERE r1.identity >= 0.8
AND p1.taxid <> p2.taxid
AND f1.annotation IS NOT NULL
AND f2.annotation IS NOT NULL
RETURN f1.name, f2.name"