Full Marks: 75   Time: 04 hours

[Answer any five of the following. All questions are of equal value. The figures in the right margin indicate full marks. Please write parts of a question serially.]

1. a. Write the advantages and disadvantages of following two data integration models: model-first load-later and load-first model-later. **4**

   b. Business Intelligence (BI) also deals with collecting, integrating, analyzing large volume of data. Then, how Big Data is different from BI. **3**

   c. Describe the following V's considering their importance in real world use cases: Veracity, Variability, Visibility, Viability. **4**

   d. Sketch a use case to explain descriptive, predictive, and prescriptive analytics. **4**

2. a. What is linked data browser? **2**
   b. Discuss about the linked data access mechanisms. **3**
   c. Illustrate the linked data application development framework. **3**
   d. Write about the linked data resources on the web. **4**
   e. Will linked data become the semantic web? **2**
   f. What is the function of Semantic Html? **1**

3. a. "Technologies are shifting from a database centric concept to a knowledge base (or knowledge graph) centric topic"-Explain this statement considering the different components of a knowledge base and a use case. **4**

   b. We define an ontology as a model to capture the semantics of data at the schema level. RDFS and OWL are used in combination with the RDF data model to define an ontology. Write the name of the relationships that hold in the following scenarios: **4**

   1.Consider the classes of males and females. Name a relationship between them that should be included in an ontology. Which constructs will you use either from OWL or RDFS to define the relationship?

2.Consider the classes of persons, males and females. Name a relationship between all the three that should be included in an ontology. Which part of this relationship can be expressed in RDFS.

3. Suppose we declare Bob and Peter to be the father of Mary. Obviously there is a semantic error here. How should the semantic model make this error impossible?

4. What relationship exist between "is child of" and "is parent of"?

c. Give an brief overview of the inductive techniques typically applied to knowledge graphs.   2

d. What is the knowledge graph embedding? Explain how entity embeddings and relation embeddings are computed by the TransE model.   5

EMP

EMP NO INTEGER PRIMARY KEY	ENAME VARCHAR (100)	JOB VARCHAR (20)	DEPTNO INTEGER REFERENCES DEPT (DEPTNO)
7369	SMITH	CLERK	10

DEPT

DEPTNO INTEGER PRIMARY KEY	DNAME VARCHAR (30)	LOC VARCHAR (100)
10	APPSERVER	NEW YORK

Figure 1: The database schema

The desired RDF triples to be produced from this database are as follows:

```
Example output data

<http://data.example.com/employee/7369> rdf:type ex:Employee.
<http://data.example.com/employee/7369> ex:name "SMITH".
<http://data.example.com/employee/7369> ex:department
<http://data.example.com/department/10>.

<http://data.example.com/department/10> rdf:type ex:Department.
<http://data.example.com/department/10> ex:name "APPSERVER".
<http://data.example.com/department/10> ex:location "NEW YORK".
<http://data.example.com/department/10> ex:staff 1.
```

Figure 2: RDF version of the database

4. a. Figure 2 is the RDF version of the database (DB) given in Figure 1. Write the R2RML mapping document to convert the given DB to RDF. 4

 b. Draw the working flows of direct and R2RML mappings. Write the advantages of R2RML mapping over direct mapping. 4

c. Explain how to serialize RDF data (given in Figure 1) in four different formats. 3

d. Write the benefits of JSON-LD over Turtle format. 2

e. What is an RDF reification? Present with an example 2

5. a. Write an RDF model in Turtle notation representing the RDF graph in Figure 3. When writing the triples separate them in two blocks: one block for terminolog- ical triples (i.e., defining concepts) and one block for asserted facts (i.e., defining instances).
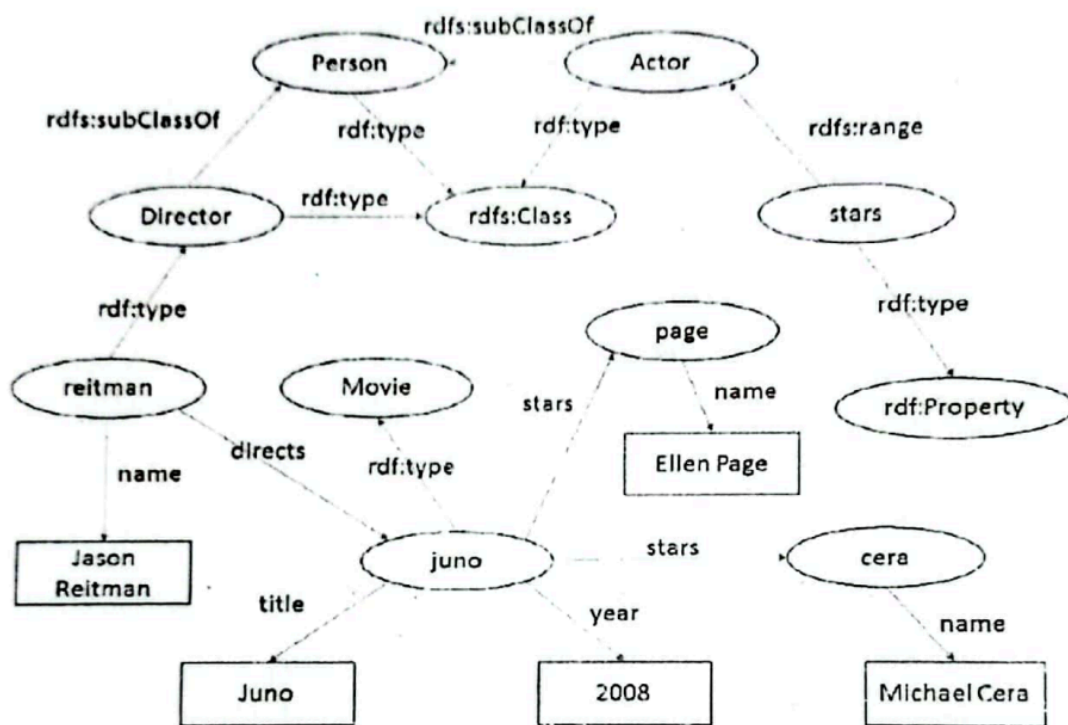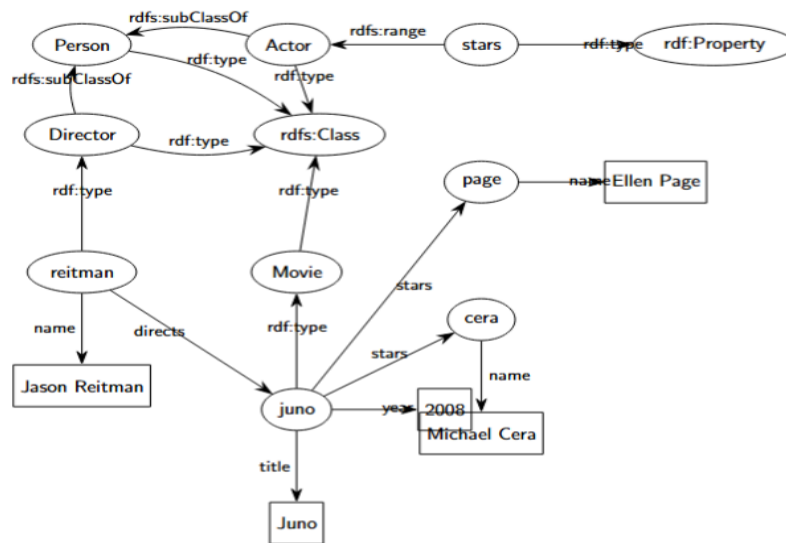
Note: Squared concepts represent literals. For the other concepts without an explicit URI assume the http://www.example.org/ namespace. Properly define any other prefix than rdf and rdfs you may want to use.

Figure 3: The database schema

b. Assume an RDFS entailment regime for the RDF graph you just created.   4
   Represent the inferred knowledge in the form of triples and for each of them
   justify from what RDFS semantic construct it is inferred. Circle those inferred
   triples that were not already represented in the graph (i.e., clearly identify the
   new knowledge inferred)

c. Write an SPARQL query to retrieve the following query: "Retrieve the   4
   number of movies starred by Ellen Page per year" The query must compile
   and include a proper definition of the prefixes used.

6. a. Consider you are asked to answer the following query from the DBpedia   15
    database. [1+2+3+3+3+3]

  i.   Write the SPARQL query to list all the DBPedia classes (i.e., instances of the owl:Class)

  ii. Write the SPARQL query to list all the DBPedia properties (i.e., either instances of the
owl:DatatypeProperty or owl:ObjectProperty properties).

  iii. Write the SPARQL query to list all the triples in DBPedia using the pop- ulationEstimate (i.e.,
http://dbpedia.org/property/populationEstimate) property:

  iv. Write the SPARQL query (i.e., a single query) to list all the pairs <SUBJECT, PROPERTY> where
<http://dbpedia.org/ontology/CyclingTeam> is the OB- JECT and all the pairs <PROPERTY OBJECT>
where <http://dbpedia.org/ontology/CyclingTeam> is the SUBJECT

  v.   Now check the results you obtained from the previous query. Briefly explain the SPARQL query
results. What do these pairs mean (you can group the re- sults per similarity and comment on each group?
You can ignore the following properties (for internal usage of DBPedia): wasDerivedFrom and
isDefinedBy as well as describedby, defines and describes (mainly used for metadata pur- poses).

  vi   Now, let us explore the data. Write the SPARQL query to list the name (i.e., rdfs:label) of all
persons (i.e., of type <http://dbpedia.org/ontology/Person>) that were born (i.e.,
<http://dbpedia.org/property/birthPlace>) in Barcelona. Assume that the range of the birthPlace property
is a literal. Also, order the result in ascending order

ZL	UZ	UN/WA	MZ/MH	VIL	RMO	Administrative Unit - *Residence Community*	Area in Acres	Total Households	Population Total	Population in Households	Floating Population	Population Density (sq. km)
15	37	29	253		1	*Enayetpur*		1310	6781	6771	10	
15	37	29	799		1	*Paschim Dhalai*		5725	30247	30247	0	
15	37	35				**Fatehpur Union Total**	5668	7562	43929	43911	18	1915
15	37	35	370		1	*Jobra*		1182	5935	5935	0	
15	37	35	448		1	*Jungle Paschim Patty*		428	4804	4804	0	
15	37	35	643		1	*Maizpatty (Fatehpur)*		2108	11697	11691	6	
15	37	35	702		1	*Mithachara (Part)*		585	3203	3203	0	
15	37	35	721		1	*Mithanala*		291	1564	1564	0	
15	37	35	838		1	*Paschim Patty*		2968	16726	16714	12	

(Table c-01: Area, Households, Population, Density by Residence and Community)

Figure 4: Data to be multidimensionally analyzed.

7. a. Figure 5 shows the demo data to be published in the semantic and analytical format. Complete the following task: [5+6+2+2]
   1. Sketch a schematic diagram by analyzing the data given in Figure 5.
   2. Represent your schema using the RDF model. Use the QB4OLAP vocabulary given in Figure 4 to annotate the schema with multi dimensional semantics.
   3. Show a sample level member in RDF.
   4. Show a sample observation in RDF.

15

8. a. Present a comparative analysis among the following triple stores: JenaTDB, Vir- tuoso, GraphDB.

3

   b. What is a property graph? How does it differ from an RDF graph? Describe with appropriate examples.
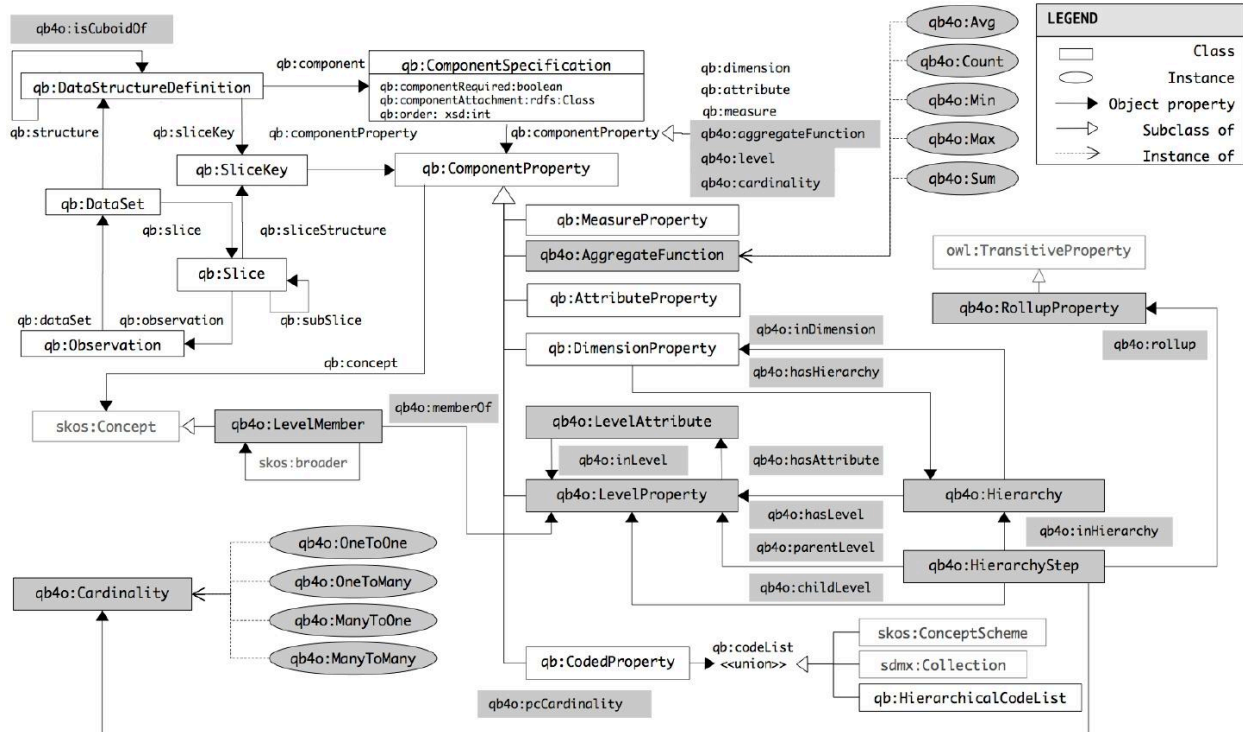
3

Figure 5: QB4OLAP vocabulary

c. Considering a resource name ex:Albert Einstein explain following data com- pleteness quality metrics: Schema , Property, Population, Interlinking, Currency, Metadata, and Labelling completeness.

4

d. You are asked to build a vocabulary to capture all hotel reviews given in the tex- tual format. Sketch the process you will follow. How will a large language model facilitate you in this direction.

5