Robert Gordon University

CMM702 – Advanced Databases 2019/2020 – PART B

Module leader	Mr. Sheron Dinushka
Unit	Coursework – Part B
Weighting:	This is Part B of the Coursework assignment a schema−less DB development exercise contributing 50% of the total module assessment.
Learning Outcomes Covered in this Assignment:	LO3. Critically appraise relational and schema−less database offerings (e.g. NoSQL) for a significant technical problem. LO4. Design and implement a solution to a significant industry−focused problem providing insights and conclusions about challenges, opportunities and risks for big data management.
Handed Out:	5th March 2020
Due Date	21st of April 2020 (Tuesday), 1.00 PM UK, online submission.
Expected deliverables	One electronic file containing the report as specified below.
Method of Submission:	Online via Moodle – Report
Type of Feedback and Due Date:	Written feedback and marks - 10 working days (2 weeks) after the submission deadline, the week starting Monday 27th April 2020.

PART 1

In this section, you have to prepare and submit a document with answers for following three (3) questions. Follow the instructions given for each question when compiling your response document.

Question 1 (24 marks)

Describe a real-world application for which following NoSQL types would be **most suitable**:

1. Key/Value stores
2. Document stores
3. Column-Family databases
4. Graph databases

You must describe the application with sufficient details and examples, so it is clearly understood why each NoSQL type is best suited for the application you have used in your respective descriptions.

You **must** consider following aspects along with any other aspects that are applicable, when describing the suitability of each NoSQL type with respective application:

1. Critical business requirements/features in the application discussed
2. Mapping of above critical requirements/features of the application with the key characteristics of the discussed NoSQL type
3. Advantages and disadvantages of the NoSQL type in the discussed application

Instructions for documentation

Provide your descriptions/justification on, but not limited to the three aspects listed above, for all 4 NoSQL types separately. At least 500 words should be used to describe an application for ***each*** *NoSQL type.*

(Page size: A4, font name: Times New Roman, font size: 12 point, line spacing: double spaced, margins: 1.0" on all sides)

[LO3, LO4, 4 x 6 marks]

Question 2 (40 marks)

There are hundreds of "NoSQL" technologies and dozens of them are widely used. Pick any NoSQL technology that is **not** discussed in the course module (any technology excluding Riak, MongoDB, Cassandra, Neo4j). Please note that the selected technology is to be approved by the lecturer.

Your work with the selected technology should cover following areas:

- Study the functionality of the system (objectives, key features, drawbacks, etc.)
- Download, run and work with the system (practical work with the system)

You should **include** following two sections under question 2 in your document as follows.

- Section 1 **(1000 words)**: Describe the features of the system. Within this section, include following sub sections:
    - Introduction - Technology used, maturity level, commercials, support
    - Objectives
    - Key features
    - Architecture (Master-slave/peer-to-peer, consistency, availability, replication, etc.)
    - Data model
    - Advantages/disadvantages

[LO3, 20 marks]

- Section 2 **(4 pages)**: Perform practical operations with system and provide the NoSQL commands you have used and results you have obtained. You may use an appropriate dataset that you have created or publicly available online for your practical work. You should focus on at least following operations:
    - Creating/dropping databases/collections
    - Storing, modifying and deleting data
    - Retrieving data
        - Simple select
        - Filtering
        - Sorting
        - Aggregations and pipelines
    - Basic administration tasks
        - Start/stop the instance
        - Tune consistency/availability
        - Replication

[LO3, 20 marks]

Instructions for documentation

(Page size: A4, font name: Times New Roman, font size: 12 point, line spacing: double spaced, margins: 1.0" on all sides)

Question 3 (26 marks)

For Question 3, use data provided with '*restaurants.json*' file and load it to a MongoDB instance. Write MongoDB commands for following requirements and provide them in the document. Where requested, provide the results you got through the commands used.

1. Create a database called "iitdb" and a collection called 'restaurants' and load the given data set to the "restaurants" collection. **[LO4, 1 mark]**

2. List only the names of all the restaurants in borough 'Brooklyn'. **[LO4, 1 mark]**

3. Count the number of restaurants which achieved a score more than 70 – Report the number you get as the answer. **[LO4, 2 marks]**

4. Find the restaurants which do not prepare any cuisine of 'Australian' and has achieved a grade point 'A'. Display only the name and the cuisine. **[LO4, 3 marks]**

5. Find the restaurants with grade point 'A' according to the latest grading (first element in the 'grades' Array). Display only the name, cuisine and the borough. Arrange the name of the **cuisine** in **ascending** order, and **borough** for that same cuisine should be in **descending** order. **[LO4, 4 marks]**

6. Find the restaurants with more than three grade surveys ('grades' Array contains more than three elements). **[LO4, 4 marks]**

7. Write an aggregation pipeline to count the number of restaurants in borough Bronx for each cuisine type. Display the number of restaurants that prepare Caribbean cuisine. **[LO4, 5 marks]**

8. Write an aggregation pipeline to return the smallest and the highest number of restaurants by cuisine for each borough. Display the cuisine and the restaurant count for the highest number for borough 'Manhattan' (*Hint: you may sort the results and access the first and the last)***.** **[LO4, 6 marks]**

Instructions for documentation

In your document, you should include MongoDB commands for each task given above and the results you obtained where appropriate.

PART 2 (10 marks)

- Part 2 is a viva voce examination of 15 minutes' duration.
- The viva is designed to assess students' knowledge and ability to understand, apply, analyse and evaluate the knowledge
- Students will be examined on the following:
    - three questions in Part 1
    - principles and concepts related to the six topics discussed during the lectures.
- For Question 2 and Question 3, students should demonstrate a running database instance (single node) of a selected NoSQL technology and a single node MongoDB instance respectively. Students are expected to execute a few database operations on the instances.
- In cases where an examiner notes significant concerns about a candidate, the examiner opinion (provided it is well documented in the score sheets) outweighs the marks assigned to the answers provided in the report for Part 1.

[LO3, LO4, 10 marks]