Big Data Technologies (Advanced Database) Course
The exam has 4 questions You must complete 3 questions.

Short Revision Questions
These questions should assist you in revising particular topics on the course.

DATABASES OBJECT RELATIONAL

1) Explain Object Types in Oracle

2) Compare and contrast Object tables and a relational tables containing Column Objects

3) By way of example, explain the use of REF's. What is a SCOPED REF?

4) What is a VArray? Why would you use it over a Nested Table Type?

5) Discuss the TWO collection types supported on Oracle. What are their differences?

6) Explain the differences of the VARRAY and NESTED TABLE object relational features of Oracle.

7) Briefly describe the use of the Object type in Oracle.

8) Discuss how inheritance is implemented in Oracle Object Relational DBMS

NoSQL

1) Compare and contrast Relational Databases with the common characteristics of NoSQL Databases.

2) Discuss the common characteristics of NoSQL Databases.

3) Domain Driven Design using Aggregates is commonly used in designing collections. Explain what this is. How does it differ to designing databases in the relational world?

4) Riak's Key-Value store architecture uses a logical ring. Discuss the structure of the ring. What are vnodes?

5) Highlight the significance of **Consistent Hashing** and the **Hinted Hand-Off** in the operation of the ring in a NoSQL db like Riak.

6) Discuss the CAP Theorem in the context of NoSql Database.

7) Explain "**Strong Consistency"** and "**Eventual Consistency".** How does a noSQL db you are familiar with provide these forms of consistency.

8) Explain the Hash Sharding and Ranges Sharding? Why is Sharding used in Aggregate Oriented NoSQL Database like MongoDB?

9) Explain Sharding and Replication in the NoSql environment.

10) Compare and contrast Key/Value stores with Document Databases.

11) Why is key design important in Key-Value Stores?

12) What is **Master/Slave** Replication? What advantages does it bring? What are its challenges?

13) Explain how a **Read Repair** is carried out in a NoSql database that you are familiar with. What is its advantage?

14) What is Peer-Peer (Master-less) Replication? What advantages does it bring? What are its challenges?

15) You have been asked to determine the design options in MongoDB for the following scenario.

- In a web development company, a software developer works on up to 10 development projects at any one time whilst a project have typically up to 6 software developers assigned. (Cardinality M:M).

As part of your answer you should discuss the design options available and under what circumstances would you choose each of the options. Is there any trade-offs with these options you have identified? What determines the optimal design?

16) Briefly explain the key factors you would take into account in choosing the correct design for documents in MongoDB collections.

Oracle have a tech blog where you can post up a technical problem and users can post up solutions and advice. If you have a 1:M relationship like Blog:BlogPosts what are the design options available? There 50,000 blogs in the database and each blog can have up to 60 blogs post for a particular blog. Which is the optimal solution and why? State your assumptions.

Blog [50K]
_id <int>
Blog Category <string>
Title <string>
Text Body<string>
User_id <string>
Poster_name <string>
Date <date>

BlogPosts [0,60]
_id <int>
Poster_name <string>
Date <date>
Post Body <string>

Data Warehousing and Design

1) Describe what is meant by the following terms, when describing the characteristics of the data in a data warehouse:
   (a) subject-oriented;
   (b) integrated;
   (c) time-variant;
   (d) non-volatile.

2) Briefly discuss how Online Transaction Processing (OLTP) systems differ from data warehousing systems.

3) Describe the characteristics and main functions of the following components of an enterprise data warehouse.

4) Metadata is important in a Data warehouse environment? Explain the different types of metadata one would expect in this environment

5) Provide a diagrammatic representation of the typical architecture and main components of a data warehouse.

6) Discuss the three types of slowly changing dimensions.

7) Briefly discuss how data marts differ from data warehouses and discuss the main reasons for implementing a data mart.

8) Describe how a dimensional model (DM) differs from an Entity–Relationship (ER) model.

9) Describe how the fact and dimensional tables of a star schema differ.

10) Discuss how star, snowflake, and starflake schemas differ. Give examples to demonstrate your answer

11) What is a Galaxy\ Fact Constellation? How do you ensure you achieve this design?

12) Why are dimensions in a star schema de-normalised?

13) What criteria must a dimension comply with to be deemed a conformed dimension?

14) What is a fact-less fact table? Give an example to demonstrate your answer

15) Explain how one would devise candidate star schemas from a corporate data model.

16) What is meant by a semi-additive measure? What other categories of measures are there?

17) What are Type 0 and Type 1 slowly changing dimensions?

18) Explain Type 2 and Type 3 slowly changing dimensions. When would you use one over the other?

19) Explain the importance of conformed dimensions.

20) Explain Component entities, Classification entities and Transaction entities in the context of identifying star schemas.

21) Discuss an approach in identifying candidate star schemas from corporate data model

Hadoop

1) Explain how a file is stored in HDFS.

2) What is the role of NameNode and Datatnodes id the Hadoop file system?

3) Compare and contrast the Secondary NameNode (2NN) with Standby NameNode

4) The Hadoop environment's philosophy is **data locality** where possible. By way of an example, explain data locality in this context.

5) Explain how data replication is carried out on a Hadoop Cluster. As part of your answer explain rack awareness.

6) What Daemons and processes are involved in YARN and what are their respective roles?

7) Explain the role of the Secondary NameNode

8) What is the sequence of events that occurs in a HDFS file system write operation? Provide a diagram to support your answer.

9) Why do master nodes (e.g. NameNode) normally require a higher degree of fault tolerance that slave nodes (Datanodes)

10) Compare and contrast Hadoop with an RDBMS.

11) Hadoop supports Block Replications and Slicing. Briefly explain this statement.

12) How is the failure of a DataNode detected in Hadoop? What happens after detection?

13) Explain the Standby NameNode operates in a Hadoop cluster?

14) Briefly explain MapReduce in the context of the Hadoop cluster.