Assignment 8

Due 12 Nov at 23:59   **Points** 100   **Questions** 23

Available 30 Oct at 6:00 - 15 Nov at 23:59   **Time limit** None

Allowed attempts 2

Instructions

(no late submission accepted) -- Complete the assessment covering all materials covered thus far with a focus on NoSQL databases. There is no time limit for the test and you may use any available resources. However, you may not collaborate with anyone. You will have two attempts and the highest grade counts.

Attempt history

	Attempt	**Time**	**Score**
KEPT	**Attempt 2**	40 minutes	91.25 out of 100
LATEST	**Attempt 2**	40 minutes	91.25 out of 100
	Attempt 1	8,174 minutes	90.4 out of 100

ⓘ Correct answers are hidden.

Score for this attempt: **91.25** out of 100

Submitted 12 Nov at 16:32

This attempt took 40 minutes.

Question 1	**4 / 4 pts**

What should be done during EDA?

☑ Generate a large quantity of questions

☑ Visualize distributions, means, variances, etc

☑ Identify typical values and outliers (unusual values)

☑ Transform data for visualization and statistical modeling

☑ Refine the questions you asked

☐ Build a sophisticated model to answer the questions

Question 2   **4 / 4 pts**

Which of these is not one of the original 3 V's proposed by Normadeau in 2003?

◉ Veracity

◯ Volume

◯ Variety

◯ Velocity

Question 3   **5 / 5 pts**

Voldemort is an open source _____ database.

◉ key-value

◯ relational

◯ graph

◯ columnar

◯ document

Question 4   **5 / 5 pts**

Which of the following is not a relational database management system?

○ Cassandra

○ Postgres

○ Microsoft Access

○ HSQLDB

○ MySQL

○ H2

Incorrect

Question 5

0 / 5 pts

Both MongoDB and CouchDB primarily use _____ as its native query language.

> JSON

Question 6

5 / 5 pts

Which of the databases below will work with R? Check all that apply.

☑ filehash

☑ MySQL

☑ HBase

☑ Cassandra

☑ Redis

☑ Riak

Question 7

5 / 5 pts

True or False? MemcacheDB is a key-value database.

◉ True

○ False

Question 8

5 / 5 pts

Storing content such as blogs is best suited for a _____ database.

◉ document

○ relational

○ graph

○ columnar

○ key-value

Question 9

5 / 5 pts

True or False? Key-Value databases are excellent for complex queries that require aggregation.

○ True

◉ False

Question 10   **5 / 5 pts**

_____ is a database built on Hadoop and is designed for scaling across clusters of inexpensive computers with support for versioning and data compression.

◉ HBase

○ MapReduce

○ Cassandra

○ MongoDB

Question 11   **5 / 5 pts**

_____ is a database that is known for its ruggedness and high availability even when hardware or network connectivity fails.

◉ CouchDB

○ MongoDB

○ Hadoop

○ Redis

Question 12   **5 / 5 pts**

A columnar database is most like an R _____ object.

- ○ list

- ◉ data frame

- ○ matrix

- ○ array

- ○ mode

- ○ factor

Question 13   **4 / 4 pts**

A table has a field "Major" that can have multiple values separated by commas. This violates _____.

- ◉ 1st Normal Form

- ○ 2nd Normal Form

- ○ 3rd Normal Form

- ○ referential integrity

Question 14   **4 / 4 pts**

Text data is encoded in SQL using the _____ or _____ data type.

    ◉   VARCHAR, TEXT

    ○   STRING, BLOB

    ○   TEXT, STRING

    ○   CHARACTER, CHARSET

Question 15   **4 / 4 pts**

All potential primary keys for a table are called its _____ keys.

    ◉   candidate

    ○   foreign

    ○   reference

    ○   unique

Question 16   **4 / 4 pts**

In the relationship, " helicopter is a particular type of aircraft", helicopter is a _____.

    ◉   specialization

    ○   generalization

    ○   actor

    ○   entity

Question 17   4 / 4 pts

Every row in a table is uniquely identified by its _____.

- ◉ primary key

- ○ key

- ○ alternate key

- ○ id

Question 18   5 / 5 pts

Which of the following are web scraping tools that would allow a data scientist to extract data from a web page and save into a data file loadable into R for analysis and visualization? Check all that apply.

- ☑ grepsr

- ☑ Mozenda

- ☐ MongoDB

- ☐ Hadoop

- ☑ Web Scraper for Google Chrome

- ☑ Mozenda

- ☐ Samurei

Question 19   5 / 5 pts

Which of these packages is required to retrieve documents via HTTP?

◉ RCurl

○ Rhttp

○ libURL

○ XML

Partial

Question 20   **2.25 / 3 pts**

For a dataset to be tidy:

☑ Each variable must have its own column

☑ Each observation must have its own row

☑ Each value has its own cell.

☐ Each type of observational unit forms a table

Question 21   **3 / 3 pts**

The function separate accepts the convert argument. When the convert argument is set equal to TRUE:

○ separate with convert the new columns to character strings.

○ separate will convert the new columns to integers

○

separate will convert the new columns to the data type that is most appropriate for the data values.

○

separate will convert the original column to the most appropriate data type.

Question 22   3 / 3 pts

When the string vocabulary is a non-empty character string, the R commands !str_detect(vocabulary, "[aeiou]") is equivalent to str_detect(vocabulary, "^[^aeiou]+$")

◉ True

○ False

Incorrect

Question 23   0 / 3 pts

The R code my_months <- factor (c("Jan", "Mar", "Feb","Mar"), month_level("Jan","Feb","Mar"))

will convert the first vector element of my_months to NA.

○ True

◉ False

Quiz score: **91.25** out of 100