FIT5137 Group Assignment - Sem 2/2019 (Weight = 20%)
Due date: Week 8, Wednesday 18-Sep-2019, 11:55pm

A. General Information and Submission

- This is a group assignment. One group consists of 2 students only. You need to register your group composition in the form as soon as possible.
- *Submission method*: Submission is online through Moodle.
- *Penalty for late submission*: 10% deduction for each day.
- *Assignment Coversheet*: You will need to sign the assignment coversheet.
- *Contribution Form*: The contribution needs to be completed by all members and please sign (e-signature is acceptable) the form as an agreement between members.

B. Background – MonashBnB

MonashBnB is a new residential service that offers short time lodging to Monash students and staff around Melbourne. They have been keeping records of the accommodation listing, hosts, and reviews manually. Due to the increasing volume of people coming in and out of the accommodations listed in MonashBnB, the management team wondered to a way that could help them be in control of the environment without too much manual hassles. Therefore, they requested your team to build a new database system through the incorporation of MongoDB and Cassandra Technologies.

C. Tasks

The assignment is divided into **FOUR** main tasks:

C.1. Database Design.

Your team has been provided with the following three reports:

- **The Accommodation Host Report** (Table 1).
  This report provides some information about the hosts who are registered in MonashBnB. It consists of the host ID, name, verification details, host registration date (host since), location, response rate, and information whether the host is a super-host.
- **The Reviews Report** (Table 2).
  This report shows the reviews left by the guests after staying in a certain accommodation. The report contains the listing ID, review ID, review date, review time (sequence), reviewer ID, reviewer name, rating, satisfaction reason, and comments.

- **The Listing Report** (Table 3).
  The listing report stores information about the existing accommodation around Melbourne that are listed in MonashBnB. The report contains the information about the listing ID, listing name, host ID, neighbourhood area, listing address, longitude, latitude, accommodation room type, amenities, price per night, price for extra people, minimum booking length, and availability per 365 days.

As MonashBnB wants you to build two database systems, they have requested that the **listing** and **hosts** data to be stored in MongoDB, while all data regarding the **reviews** should be stored in Cassandra.

Task Requirement:

In this task, you are required to do the following:

MongoDB:

- Create a database called ***FIT5137_Assign***.
- Create collections for ***Listing*** and ***Host***.
- Insert the data provided in Table 1 and 3 into the collections.
- Use appropriate data types while inserting the data.
- Use appropriate naming convention for each field.

Cassandra:

- Create a keyspace called ***FIT5137_Assign***, with SimpleStrategy no replication.
- Create a table for ***Review*** if not exists.
- Insert the information provided in Table 2 into the table.
- Use appropriate data types while inserting the data.
- Use appropriate naming convention when creating the table.

The script for each task for MongoDB should be kept in a .txt file called **FIT5137_Assign_C1_MongoDB**, while the script for Cassandra should be kept in another .txt file called **FIT5137_Assign_C1_Cassandra**.

Host ID	Host Name	Host Verifications	Host Since	Host Location	Host Response Rate	Host Is Superhost
MONHOS01	Manju	email, phone, reviews	21/08/2009	Clayton, Victoria, Australia	within a day	f
MONHOS02	Lindsay	email, phone, reviews, jumio, government id	16/09/2009	Clifton Hill, Victoria, Australia	within an hour	t
MONHOS03	Adam	email, phone, google, reviews, jumio, government id, work email	31/10/2009	Port Melbourne, Victoria, Australia	within an hour	f
MONHOS04	Eleni	email, phone, facebook, reviews, jumio, offline government id, government id, work email	3/12/2009	Fitzroy, Victoria, Australia	within a day	f
MONHOS05	Colin	email, phone, facebook, reviews, jumio, offline government id, government id	22/12/2009	Saint Kilda East, Victoria, Australia	within an hour	f
MONHOS06	Daryl	email, phone, manual online, reviews, manual offline, work email	12/07/2010	Berwick, Victoria, Australia	within an hour	t
MONHOS07	Diana	email, phone, facebook, reviews, jumio, offline government id, government id, work email	27/07/2010	Parkdale, Victoria, Australia	within a day	f
MONHOS08	Belinda	email, phone, facebook, reviews, jumio, offline government id, selfie, government id, identity manual, work email	3/08/2010	Prahran, Victoria, Australia	within a few hours	f
MONHOS09	Allan	email, phone, facebook, reviews, jumio, offline government id, selfie, government id, identity manual	3/08/2010	South Melbourne, Victoria, Australia	within an hour	t
MONHOS10	Vicki	email, phone, facebook, reviews, jumio, government id	6/08/2010	Frankston, Victoria, Australia	within an hour	t

Table 1: Accommodation Hosts Report

Listing ID	Review ID	Review Date	Sequence	Reviewer ID	Reviewer Name	Review Scores Rating (out of 100)	Satisfied Reason	Comments
MONLST02	REV01	20170322	10:37:50+1300	500001	Miriam	90	{location, amenities}	Beautiful View
MONLST02	REV02	20170322	11:37:50+1300	500002	Johannes	90	{host}	Good Host
MONLST02	REV03	20170322	11:37:50+1300	500003	Camille	100	{location, view}	Nice View and Location
MONLST02	REV04	20170322	12:37:50+1300	500004	Paige	95	{price}	Excellent Price
MONLST01	REV05	20170322	15:37:50+1300	500005	Adele	93	{location, price}	Good Location
MONLST03	REV06	20170322	17:37:50+1300	500006	Greg	87	{host, view}	Very Clean House
MONLST04	REV07	20170322	19:37:50+1300	500007	Wolfgang	91	{location, price}	Nice Location
MONLST05	REV08	20170322	20:37:50+1300	500008	Klaus	96	{location, view}	Nice Building
MONLST06	REV09	20170323	11:37:50+1300	500009	Rox	100	{host, price}	Friendly Host
MONLST05	REV10	20170323	12:37:50+1300	500010	Elisabeth	98	{host, price}	Friendly Host
MONLST09	REV11	20170323	19:37:50+1300	500011	Derek	100	{space, clean}	Very Clean and comfortable
MONLST09	REV12	20170325	10:07:40+1300	500012	Joy	92	{host, clean}	Friendly and Nice Host
MONLST10	REV13	20170326	10:02:10+1300	500013	Anouck	93	{host, view}	Very Comfortable
MONLST12	REV14	20170326	10:49:40+1300	500014	Jerome	85	{location, clean}	Friendly Host
MONLST13	REV15	20170326	10:48:40+1300	500015	Jehan	98	{location, amenities}	Beautiful View
MONLST14	REV16	20170326	10:48:10+1300	500012	Joy	97	{amenities, view}	Good Location
MONLST14	REV17	20170326	10:47:40+1300	500014	Jerome	30	{price, view}	Bad Location
MONLST10	REV18	20170326	10:47:10+1300	500002	Johannes	20	{amenities, view}	Bad Service
MONLST12	REV19	20170327	09:37:50+1300	500013	Anouck	87	{space, clean}	Good Location
MONLST05	REV20	20170327	10:45:10+1300	500011	Derek	96	{host, view}	Nice Building

Table 2: Accommodation Reviews Report

Listing ID	Name	Host ID	Neighbourhood	Address	Latitude	Longitude	Room Type	Amenities	Price per night ($)	Price for extra people ($)	Minimum nights requuired for booking	Availability per 365 days
MONLST01	Monash Beautiful House	MONHOS14	Manningham	Clayton, VIC 3800	-37.773	145.09213	Entire home	TV, Wifi, Pets Allowed, Family friendly, 24-hour check-in, Self check-in	61	22	1	365
MONLST02	Monash Brunswick Deco	MONHOS08	Moreland	Brunswick East, VIC 3057	-37.767	144.98074	Private room	Kitchen, Toiletries, Hair dryer, Iron, Microwave, Coffee Maker, Refrigerator, Cooking basics, Stove, Garden	35	15	3	194
MONLST03	Monash Beachside Retreat	MONHOS01	Port Phillip	St Kilda, VIC 3182	-37.86	144.97737	Entire home	Cooking basics, Oven, Stove, Dishwasher, Balcony, Wifi, Kitchen, Washer, Toiletries, Hair dryer, Iron, Microwave, Coffee Maker, Refrigerator, Long term stay allowed, Luggage dropoff, 24-hour check-in, Self check-in	159	0	2	82
MONLST04	Monash Close2City	MONHOS04	Darebin	Thornbury, VIC 3071	-37.759	144.98923	Private room	TV, Internet, Wifi, Kitchen Free parking on premises, Breakfast, Heating, Washer, Toiletries, Hair dryer	50	20	2	0
MONLST05	Monash City and Sports	MONHOS05	Port Phillip	St Kilda East, VIC 3183	-37.865	144.99224	Private room	Internet, Wifi, Heating, Washer, Toiletries, Bed linens, Gym Access, 24-hour check-in, Self check-in	69	20	1	274
MONLST06	Monash Trafford Apartment	MONHOS06	Casey	Berwick, VIC 3806	-38.057	145.33936	Entire home	Microwave, Coffee Maker, Wifi, Kitchen,, Toiletries, Hair dryer	99	30	1	353
MONLST07	Monash Close2Airport	MONHOS07	Darebin	Reservoir, VIC 3073	-37.697	145.00082	Private room	Long term stay allowed, Luggage dropoff, 24-hour check-in, Self check-in, Airport shuttle service, Host greets you, Wifi, Kitchen, Heating, Toiletries, Microwave, Coffee Maker, Refrigerator, Cooking basics, Oven, Stove, Dishwasher, Garden,	50	20	7	0
MONLST08	Monash Home In The City	MONHOS02	Melbourne	East Melbourne, VIC 3002	-37.81	144.98592	Private room	Cooking basics, TV, Wifi, Pets Allowed, Family friendly, 24-hour check-in, Self check-in	99	25	15	62
MONLST09	Monash Japanese-Style	MONHOS11	Monash	Oakleigh East, VIC 3166	-37.9	145.11447	Entire home	Kitchen, Toiletries, Hair dryer, Iron, Microwave, Garden, Coffee Maker, Refrigerator, Cooking basics, Stove, Dishwasher,	98	0	2	219

								Balcony, Paid Parking, Long term stay allowed				
MONLST10	Beautiful Monash House	MONHOS10	Frankston	Frankston, VIC 3199	-38.149	145.14157	Entire home	TV, Wifi, AC, Kitchen, Heating, Washer, Toiletries, Hair dryer, Iron, Microwave, Coffee Maker, Refrigerator, Cooking basics, Oven, Stove, Dishwasher, Balcony, Paid Parking, Long term stay allowed, Luggage dropoff, 24-hour check-in, Self check-in	59	10	2	318
MONLST11	Fabulous Monash Richmond	MONHOS09	Yarra	Richmond, VIC 3121	-37.818	145.00442	Entire home	Bed linen, Wifi, AC, Kitchen, Heating, Washer, Toiletries, Hair dryer, Iron, Microwave, Coffee Maker, Free Parking, Long term stay allowed, Luggage dropoff, 24-hour check-in, Self check-in, Host greets you, Refrigerator, Cooking basics, Oven, Stove, Dishwasher, Balcony, Garden	98	30	14	16
MONLST12	Monash Central Lux	MONHOS12	Port Phillip	St Kilda, VIC 3182	-37.861	144.98038	Entire home	Wifi, AC, Kitchen, Heating, Toiletries, Hair dryer, Iron, Coffee Maker, Refrigerator, Stove, Dishwasher, Beachside view, Free Parking 24-hour check-in, Self check-in	189	29	2	6
MONLST13	Central Monash Warehouse Apartment	MONHOS13	Melbourne	Melbourne, VIC 3000	-37.815	144.96267	Entire home	Dishwasher, Garden, Paid Parking, Long term stay allowed, Coffee Maker, Refrigerator, Cooking basics, Oven, Stove, Wifi, AC, Kitchen, Heating, Washer, Toiletries, Hair dryer, Iron, Microwave, Luggage dropoff, 24-hour check-in, Self check-in	249	40	2	353
MONLST14	Monash Near the Park	MONHOS03	Bayside	Melbourne, VIC 3187	-37.928	145.02518	Private room	Pets Allowed, Breakfast, Garden, Family friendly, 24-hour check-in, Self check-in	40	11	2	365

Table 3: Accommodation Listing Report

C.2. Database Modifications.

The management team of MonashBnB is happy with your databases designed in Task C.1, however, they wish to make the following updates/deletions to your implemented MonashBnB database systems:

1. Updating the verification Adam and adding Facebook to his list of existing verifications.
2. Insert 5 new hosts with the following details using a single insert command:

Host ID	Host Name	Host Verifications	Host Since	Host Location	Host Response Rate	Host Is Superhost
MONHOS11	Alison	email, phone, facebook, reviews	9/1/2019	Caulfield, Victoria, Australia	within an hour	f
MONHOS12	Mike	email, phone	9/1/2019	Clayton, Victoria, Australia	within a day	t
MONHOS13	Robyn	facebook, reviews	9/1/2019	Berwick, Victoria, Australia	within an hour	f
MONHOS14	Daniel	email, manual offline, work email	9/1/2019	Frankston, Victoria, Australia	within a day	t
MONHOS15	Ron	facebook	9/1/2019	Caulfield, Victoria, Australia	within a day	f

3. Update a host who responds “within an hour” to a superhost. For this update you may only use the “host response time” and “host is a super host” information.
4. Delete all listings with zero availability, taking into account any future addition of data where availability maybe zero.
5. Using only the neighbourhood information, change the neighbourhood name of “Monash” to “Monash City”.
6. Change the satisfied reason to “space” and “price” for review ID “REV11”.
7. Deleting reviews commented by user “500015”.

Task Requirement:
In this task, you are required to do the following:

- Create the necessary modifications given in Task C.2.
- Make sure data which are not required to be modified are not modified in any way.

The script for each task for should be kept in a .txt file called **FIT5137_Assign_C2**.

C.3. Queries.

The management team of MonashBnB have started storing more data into the databases. However, they need your help in answering some queries for them to analyse their business. They have provided you with some JSON data and CSV file, which consists of:

- **host.json** – the list of registered hosts in MonashBnB.
- **listing.json** – the detailed listing data of Melbourne and its surrounding.
- **review.csv** – the review history of each listing in Melbourne.

Using the provided JSON data, the management team is interested in knowing if the new database system can answer the following queries:

1. How many accommodations were rented in December 2018?

2. What is the average price for the accommodations in the Port Phillip neighbourhood?
3. What are the top 10 most popular neighbourhoods based on the average reviews per month?
4. What is the range of number of houses rented (the lowest to the highest number) in Melbourne?
5. What is the most popular room type?
6. Where are the top 5 most expensive accommodations?
7. Display all listings with host name "Eleni".
8. Display the entire homes that the host can respond within a few hours.
9. Display the listing belongs to "Colin" with internet and gym access.
10. What is the price and room type for the listings in Clayton and the name contains "Beautiful"?
11. For each listing, display the listing's name, address and neighbourhood in the following format: "Monash Beautiful House, Clayton, VIC, 3800, Clayton" and sort the list in alphabetical order.
12. Which listing is available for the longest number of days? Add the attribute for report generation time and store the current time for when your output is generated.
13. Using a single query, list the unique neighbourhoods with the price per night greater than $50. Your list should display only the neighbourhoods and prices, and be sorted in reverse alphabetical order of neighbourhood names.
14. For each host, find the total number of verification methods and store the results as "number of verification methods". The output should be sorted according to descending order for number of verification method and it should show only host id, host name and their number of verification methods.
15. What is the most recent review about listing number 10803?
16. Display all reviews which rating is between 70 and 90.
17. Display the listing, reviewer, and comment that the reviewer rating is below 50.
18. How many reviews left in 2015?
19. Display the review with the highest rating on 26th March 2017.
20. Display the listing ID, reviewer name, and its highest rating.

The management team also wants you to provide **five additional queries** that you consider to be useful to the MonashBnB management team's operation.

Task Requirement:

In this task, you are required to do the following:

- **MongoDB**

  In MongoDB, we understand that there are two data modelling methods, which are embedding and referencing. Please create an embedding model called *listing_host* using existing two collections. For the queries above, you are required to answer them using **both embedding and referencing**.

  Thus, in this task, you have to:

- Import the JSON files and add them into a new database called ***FIT5137_Assign_C3***. This database is a different database from the one you created in Task C.1.
- Provide the appropriate read operations for each query using both embedding and referencing models. For your answer you may follow where needed the format similar to that shown in the tutorials.
- When creating each query, you have to take into consideration the efficiency of the query operation.
- Create at least **two indices** including compound index for queries that are frequently used and justify why you have chosen the fields to be the index.
- Provide the necessary code for each query.

The script for this task should be kept in a .txt file called **FIT5137_Assign_C3_MongoDB**.

- **Cassandra**

In this task, you have to:

- Import the data from the CSV file to a new keyspace called ***FIT5137_Assign_C3***. This keyspace is a different keyspace from the one you created in Task C.1.
- Provide the appropriate read operations for each query.
- When creating each query, you have to take into consideration the efficiency of the query operation.

The script for this task should be kept in a .txt file called **FIT5137_Assign_C3_Cassandra**.

Note:

1. When answering the queries above, you have to ensure that you are querying the correct database as each query can only be answered by using one database system (either MongoDB or Cassandra).
2. The provided JSON and CSV files may contain some raw information and some information are different from the Tables provided in Task C.1.
3. The five additional queries should utilise both MongoDB and Cassandra. The marking criteria for this task will be based on your query variation and complexity in doing the read operation. Just attempting five additional queries does not mean that you could get full mark for this section.

C.4. Database Comparison.

MonashBnB is happy with your work. However, they have to cut down their budget and they have decided to use only one database instead of two. Therefore, the management team wants to consult this matter with you before they proceed to only one database.

Task Requirement:

In this task, you are required to do the following:

- Decide which database to use (either MongoDB or Cassandra).
- Provide a list of the steps (including an explanation and a flowchart) on how the merging process could be implemented.
- Provide a comparison in a tabular format with details on the main strengths and weaknesses of each database.
- Provide a valid explanation on why you have selected one database over the other.
- This comparison report should only contain a maximum of 750 words.

Note: You are NOT required to implement the merged database, but your explanations must be valid for MonashBnB.

D. Submission Checklist

1. One **combined pdf file** containing all tasks mentioned above:
   - □ Cover page
   - □ A signed cover sheet
   - □ A contribution declaration form, e.g.:

     Each student must state the parts of the assignment that he/she did.
     An example is as follows:

     Percentage of contribution:
     I. Name: Adam, ID: 210008, Contribution: 60%
     II. Name: Ben, ID: 230933, Contribution: 40%

     List of parts that each student did:
     I. Adam: list the parts that Adam did
        - □ Task C.1 (MongoDB create collections, Cassandra insert,…)
        - □ Task C.2
        - □ Task C.3

     II. Ben: list the parts that Ben did
        - □ Task C.1 (Cassandra create table, Cassandra insert, …)
        - □ Task C.3
        - □ Task C.4

   - □ A **report** that combines all the tasks, including some screenshots of the output for each task.

2. **txt files** from the following tasks:
   - □ Task C.1 Database Design (**FIT5137_Assign_C1_MongoDB, FIT5137_Assign_C1_Cassandra**)
   - □ Task C.2 Database Modifications (**FIT5137_Assign_C2**)
   - □ Task C.3 Queries (**FIT5137_Assign_C3_MongoDB, FIT5137_Assign_C3_Cassandra**)

   All of the above txt files must contain scripts that are run-able in MongoDB and Cassandra.

3. Zip all the files above (pdf from #1 above, and txt files from #2 above), and upload this zip file to Moodle.

Reference

http://insideairbnb.com/get-the-data.html

THE END