Chapter 10 – Hash Tables and Dictionaries

TASK QUESTIONS (WITH ANSWERS)

1. What is a hash table data structure?
   A hash table is a data structure made up of two parts, a table (or array) of data, and a key, which identifies the location of the data within the table. A ‘hashing algorithm’ is carried out on the key, which then acts as an index to the specific location of that data item within the array.

2. Suggest a suitable hashing algorithm that could be used to store the names of everyone in your class. Write code to implement your solution.
   Any suitable method could be used, for example adding up the digits in a date of birth. The idea is to select a method that will result in hash keys that are in an appropriate range with the minimal amount of collisions.

3. Identify three possible applications for hashing algorithms.
   Databases, Memory addressing, Encryption, Checksums

4. What are the main features of a good hashing algorithm?
   They generate a suitable number of hash keys that are in a usable range with a uniform spread and the minimum amount of collisions.

5. What can you do to minimise the likelihood of collisions when creating a hash table?
   Write an algorithm that is suitably complex to generate unique hash keys.

6. What is load factor?
   The number of keys divided by the number of slots. It is a measure of how many unique keys are being generated, based on the amount of data items.

7. What is clustering and how is it caused?
   Clustering is when the hashing algorithm generates hash keys that are all around the same value. It is usually the sign of a poor hashing algorithm.

8. Explain in detail how a hashing algorithm can deal with collisions.
   There are two main methods: chaining and rehashing:
   Chaining: in this case if a collision occurs, a list is created in that slot and the key/value pair
become elements of the list. If another collision occurs, that key/value pair becomes the next element in the list and so on.

Rehashing: in this case if a collision occurs, the same algorithm is run again, or an alternative algorithm is run until a unique key is created. This normally uses a technique called probing, which means that the algorithm probes or searches for an empty slot. It could do this by simply looking for the next available slot to the index where there was a clash.

9  What is a dictionary data structure and how does it differ from a hash table?
A dictionary is an abstract data type that maps keys to data. It is called an associative array in that it deals with two sets of data that are associated with each other.

10 What are the main actions that you might want to carry out on data stored in a dictionary?
Add, retrieve and delete.