

Homework project: handling and reporting pattern calls in-memory data storage

Developer:	Mostafa Jabaroutimoghaddam
Version:	1.0.0
Date	7/29/2020

Introduction

the program should handle and manage the tuples of pattern calls with unique Identification number.

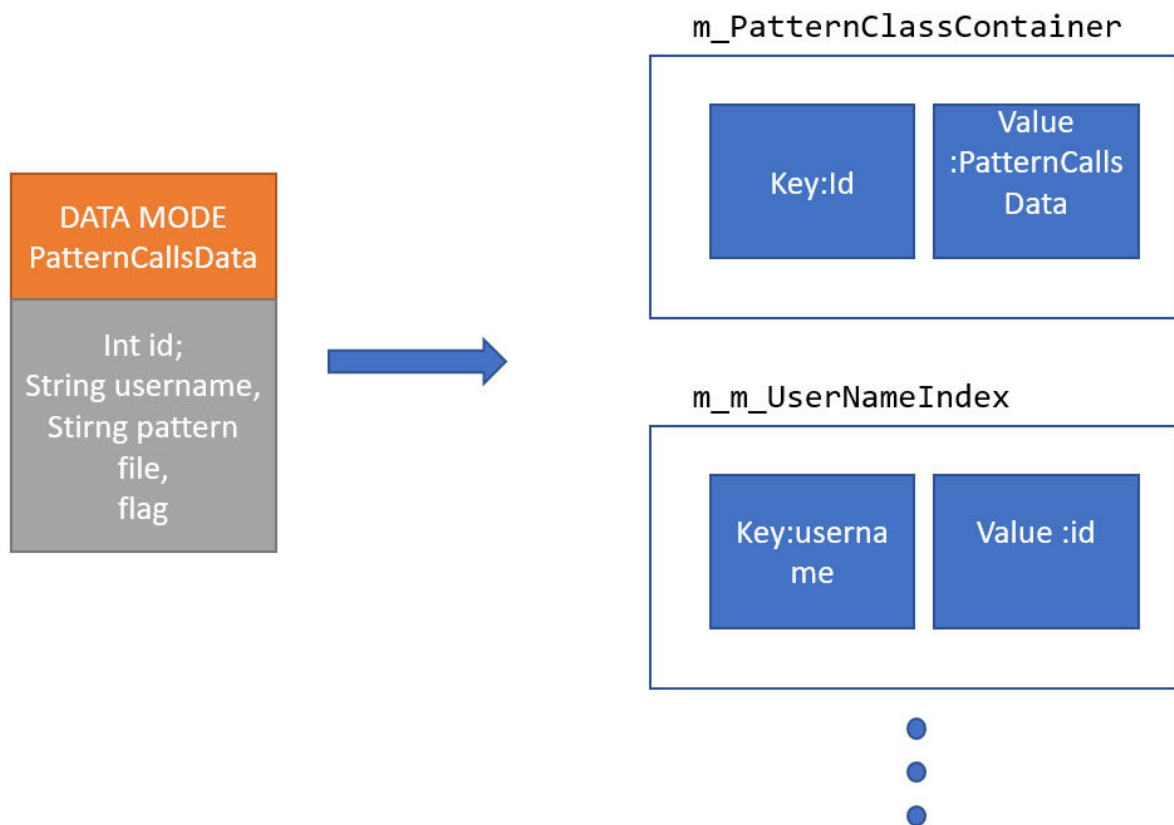
The following use-cases need in this project:

- 1- the tuples should hold and store in the memory. In this project, the tuples were not modified and just adding to the storage.
- 2- Reporting from the in-memory data storage with considering the complexity order.
- 3- the program should be production-ready such as maintainable and documented.
- 4- The program should contain the test unit. In the instruction, the program needs a production test to test the basic behavior of the program. Such as accuracy, reliability, time, etc....

Data Modeling

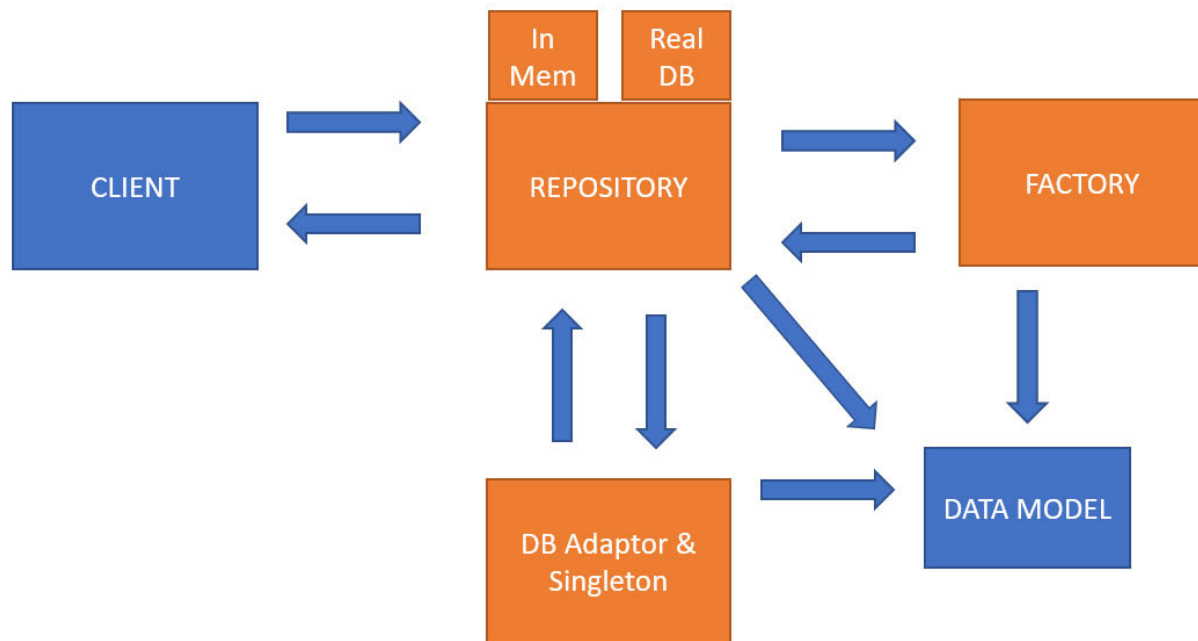
The Id of the tuples are unique. At the first glance, using a map is good. The map key is the Id of tuples and the objects that store is a tuple. For modeling the data, I constructed the object name as *PatternCallsData* containing the Id, username string, pattern file string, Boolean flag. It may use the *std::tuple* for this purpose. because I want to have control over the object, I constructed it as a class. In the first run, this is enough for storing the tuples on the map. But the second requirement of the project is reporting with considering the complexity order. If we want to find the specific Id, the complexity order is $O(1)$. But for finding other information such as username, pattern files, flag that is modeling as an object, we should use a search algorithm over the map to find the specific data. In this way, the complexity order is $O(n)$. We can use a different type of searching to decrease the complexity order. The best complexity order for this work is $O(1)$. If we want to achieve this, we should index all other databases when inserting it into a map. For example, after adding a tuple to map, we should add its username, pattern file, flag status to separate multimap. In this way, we first find the Id of the specific data and retrieve the data from the

original map. The complexity order will be $O(1)$. But the drawback of this work is needing an amount of RAM more than twice. That is a trade-off between the resource and complexity level. It is correct that we can improve some index technique to approach the lower amount of RAM. But in this project, I consider a simple index. The structure of the program that will be demonstrated later show that importing the new index and in-memory data storage can be easily imported.



Software architecture

The software should be ready for production. There are many aspects that should be taken into consideration that the program is ready for production. The first one, the reliability, and accuracy of the program should be guaranteed. The program should work in a good manner based on time, resources, and size. Second, the code should be readable and maintainable by other engineers. Third, the code should be extended at a low cost. If there is some modification that should be occurred in the projects that are common especially in the agile project systems, this modification should be done easily. At least, the design has this capability to support this issue in a good way. In this project, I consider the following architecture.



The client or program should work with a different part of the system by repositories. I use a repository design pattern to handle a unit of work and can extend the gateway design pattern to support different types of proxies. For example, I have a repository base class to retrieve the information to the real database, in-memory database, a different type of storage, and etc. by the means of their implemented repositories. On the other hand, I use the factory base design pattern for creating objects in the system. For example, there is a factory that retrieves the data from the string and creates the object. In this way, other types of creating objects from different resources can be handle. If creating the object should be done by parsing a JSON or XML, it can be done easily.

There is a database adaptor in the application that it needs to read the database or in-memory database gateway. In this project, in-memory data storage should be implemented in the program. I use the singleton design pattern to model the database. The in-memory data storage is communicating to its repository by adaptor design pattern.

Testing Stage

in this the project, writing a production test is required. I have written a test program to evaluate the performance, and accuracy of the program. However, due to the structure of the program that everything is encapsulated, applying a Unit test for testing each part of the system can be handled. googletest and googlemock can be used to test each repository, factories, DB adaptor, etc. separately. For example, to

test a real database, we can implement another repository that is producing mock data for googlemock. In this way, we can test all parts of the systems.

Questions

What are the benefits of your design?

The design may not be perfect. But I think It has this capability to improve along with low cost in implementation and designing. I tried to encapsulate everything that I can handle them easily.

Do you see improvement potential?

There are many available potential improvements. The data model section can be improved to use a much less amount of RAM. In the class architecture, I have used a copy constructor and copy assignment. It is much better to use also move constructor and move assignment to avoid deep copying. Because working with templates, inserting and work on them, they occupy and free the objects in repeat. The architecture of the program can be changed to more reliable. Consider the in-memory data storage to support more different types. The test unit can be added by googletest and googlemack for example to test a specific part of classes. I think the error handling and exception handling should be designed to support all possible failure.

What assumptions did you make and what trade-offs did you consider?

I assume that there is enough RAM space to store data. Because I need twice the amount of RAM in normal without indexing. In this way, I have a trade-off between the amount of RAM and complexity level. The complexity level is $O(1)$ that is considered as good in comparison to $O(n)$.

What is the complexity (Big-O notation) of the queries you provide?

The complexity level is $O(1)$.

Which part of your solution took the most time (e.g. design, coding, documentation) and why?

In my case the design and coding are equal. At first, I try to design a system architecture and then implement it. About the documentation, I think, it takes some time to perform documentation. But we consider the documentation as a way to clear all the things obvious, it will convert to an enjoyable task even though that it is time-consuming.

Eventually, I did my best to do the project in the time that I have. I know that there are some disadvantages and limitations. But I always eager to learn more about programming.

Best Regards,

Mostafa