Genoom Simpsons Family Tree

# Preface

This exercise consists on creating:

* Data structure to store the Simpsons family tree
* Web API to access the tree.

Goals:

* Good analysis of the requisites of the project.
* Have a good starting point for the project, that means:
  + Documentation
  + Tests
* Demonstrate the software architecture skills creating a good project structure and allow evolving the project in the future.
* Demonstrate the developer skills using the latest technologies that will be required in the position:
  + .net Core, Web Api, Azure, MongoDb …

# Database design

This is the **key topic** to analyze for this exercise. The schema of the data seems simple at first sight but digging into it, we face the following problems:

* Each family member (node in a graph) has 3 levels for navigation:
  + Up 🡪 Parents
  + Down 🡪 Children
  + Side 🡪
    - Partner
    - Sibling
* Each relation between family members, is **symmetric**, that means that automatically in a standard RDBMs would **duplicate the amount of entries needed**.
* There is the transitive property between the family members. But in the exercise, should be taken in account for getting the tree.
  + We should use **recursive calls** to navigate through the parents tree, while keeping the code needed low.

## SQL Server (or any other standard RDBMS)

* Pros:
  + Pretty standard, very easy to work with because is ubiquitous.
  + Well supported in Azure.
  + Good and stable tools and documentation.
  + ACID compliant.
* Cons:
  + We generate a lot of entries in the table pretty quick due to the symmetric of nature of the data.
  + We need a good set of indexes to not income with serious performance problems soon.
  + We need to set data partitions to manage the large amount of data.
  + We should avoid large transactions for updating the data at all (should not be a big problem because the Create/Update operation is adding a new children member to an existing one directly without needing to traverse large trees).

## MongoDb (NoSQL document based Dbs)

* Pros:
  + The cost of accessing a family member and its direct tree levels is 1. As the document should contain this info.
  + The returning data format is JSON (even better BSON), we could use it in combination with **node.js server so no extra serialization is needed and would be very performant**.
* Cons:
  + Not as good support in Azure as the SQL Server.
  + The documents could become big (be aware of 10MB limit per doc).
  + Costly Add/Update operations because the large documents. (More than the SQL Server).

Should not be a big problem because most of the operations are for reading the tree not updating it.

* + Increased complexity, we should be aware of the **CAP theorem (**Consistency, Availability, Partition tolerance.)

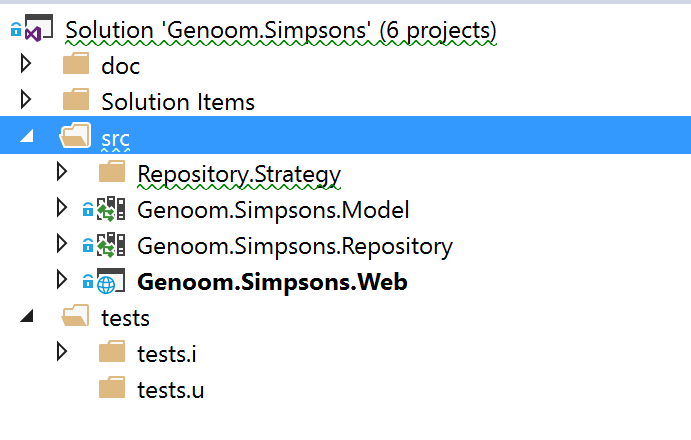
# Project architecture proposal:

## Logic and Code strategy

Before analyzing the architecture let’s set some points:

* There are 4 responsibilities for the project:
  + **Web stuff**: Web Api, Controllers, data serialization to JSON, etc.
  + **Business logic**: like checking if the person has a partner when adding a child.
  + **Data access repository**: to access the data here due to the topics discussed above, we could use a strategy like pattern so we can access the data on SQL Server and MongoDb and we could compare which option is better, also would allow creating other repositories for other technologies.
  + **Model**: simple C# POCOs representing the family members.

## Solution Structure Diagram



For organizing the code I’ve decided the following structure:

* **doc**: contains the documentation for the project, currently this document.
* **src**: contains the source code projects.
* **tests**:
  + **tests.i**: contains the integration tests. Actually, contains the **tests for the repositories** of Sql and MongoDb, because makes sense to test against the real database.
  + **tests.u**: contains the unit tests, they should not have any dependency to a real database so we need to mock those dependencies.

Actually, contains the **unit tests for the web controllers** with mocked in memory data structures.

* + For integration tests of the controllers, due to time constriction, I’ve decided to **rely on swagger** to call the API methods.

## Project Architecture Diagram



## Genom.Simpsons.Web:

* This project contains all the code related to the Web like the Web Api code, and Controllers.
* This will be the entry point for the program.
  + **It also contains all the initialization code for wiring up. Like the DI container**
* **As the business logic is quite simple, for now it’s included here in the controller**.

It should be evaluated if in the future gets more complex to move it to another project: Genoom.Simpsons.Business

## Genom.Simpsons.Repository:

* Contains only the interface for the People Repository, it’s done this way to avoid adding extra references and avoid circular references to the other projects.

## Genom.Simpsons.Repository.Sql:

Contains the implementation of the People Repository in SQL Server, but it could be used another RBDMS instead.

I’ve decided to use **Dapper** as **light ORM**. Because it’s integrated well with .net core and specially because the **performance is a lot better than Entity Framework**.

## Genom.Simpsons.Repository.MongoDb:

Contains the implementation of the People Repository in MongoDb.

**Currently, This is a work in progress and not implemented yet**

## Genom.Simpsons.Model:

Contains the Model for the application as POCOs.