Title Page

Name: Seainin Keenan

Student No: D00190985

Course: Computing in Cloud Computing

Date: 17/01/2022

**Submitted in accordance with the requirements for the degree of B.Sc. (Hons) in Cloud Computing**

Copyright © 2022, Seainin Keenan

# Acknowledgements

Thank you

# Declaration

|  |
| --- |
| I hereby declare that the work described in this project is, except where otherwise stated, entirely my own work and has not been submitted as part of any degree at this or any other Institute/University.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Seainin Keenan, 2022 |

# Abstract

The main objective of this project was to create an application/website which will track new releases of a different variety of entertainment media, i.e. movies, books, games, and shows Users will be able to add different entertainment media to their in-app calendar; The user receives a notification when their media is released via phone or email

Table of Contents

[Acknowledgements ii](#_Toc102669852)

[Declaration iii](#_Toc102669853)

[Abstract iv](#_Toc102669854)

[Table of Figures ix](#_Toc102669855)

[Introduction 10](#_Toc102669856)

[Aims 10](#_Toc102669857)

[Objectives 10](#_Toc102669858)

[Literature Review-why use this 11](#_Toc102669859)

[Existing Applications 11](#_Toc102669860)

[Hosting Platforms 11](#_Toc102669861)

[Amazon Web Services 11](#_Toc102669862)

[Google 11](#_Toc102669863)

[Azure 11](#_Toc102669864)

[Technologies 11](#_Toc102669865)

[What kind of app will be developed 12](#_Toc102669866)

[Integrated Development Environment 12](#_Toc102669867)

[Programming Languages 12](#_Toc102669868)

[Web Scraper 13](#_Toc102669869)

[Design-what is used, how I would used it 14](#_Toc102669870)

[Brief Overview of Project 14](#_Toc102669871)

[APP 14](#_Toc102669872)

[API 14](#_Toc102669873)

[Webscaper 14](#_Toc102669874)

[Implementation-how I used it physically in the code 15](#_Toc102669875)

[API Development 15](#_Toc102669876)

[Timeline 15](#_Toc102669877)

[What I did 16](#_Toc102669878)

[Hiccups 16](#_Toc102669879)

[Deployment 16](#_Toc102669880)

[Future Work 16](#_Toc102669881)

[Observations 16](#_Toc102669882)

[APP Development 16](#_Toc102669883)

[Timeline 16](#_Toc102669884)

[What I did 17](#_Toc102669885)

[Storage 17](#_Toc102669886)

[Hiccups 17](#_Toc102669887)

[Future Work 17](#_Toc102669888)

[Observations 17](#_Toc102669889)

[Web Scraper Development 17](#_Toc102669890)

[Timeline 17](#_Toc102669891)

[What I did 17](#_Toc102669892)

[Hiccups 17](#_Toc102669893)

[Deployment 17](#_Toc102669894)

[Future Work 17](#_Toc102669895)

[Observations 17](#_Toc102669896)

[Conclusions 19](#_Toc102669897)

[Source Code 20](#_Toc102669898)

# Table of Figures

[Figure 1 the numpty 7](#_Toc95301044)

# Introduction

This project looks a designing and implementing an application that can be used as a centralised point to find release dates for moves, games and music.

## Aims

1. Design a mobile application that will allow users to access entertainment release dates.
2. Implement the application backend on a cloud platform.
3. Design a database to maintain user and media information
4. Use a web scraper to find release dates automatically on the web.
5. Implement push notifications.

## Objectives

1. Understand the tools and applications used to develop mobile applications
2. Research cloud platforms
3. Research database options such as sql and noSQL
4. Research the operation of web scrapers to find information on the Internet.

# Literature Review-why use this

## Existing Applications

There id not any application to keep track of different media releases that I could find, however there were multiply websites that tell you when specific medias release i.e. movies releases or game releases. They had the information but no straight forward method to keep track of them. You could keep track of themon your on calendear and add them manullly.

There are multiple websites to find information on media releases including list. All have their own way of showing the data, as you might see games and books have a higher quantity of releases.

## Hosting Platforms

Intro to cloud services, and why you would use them in a project like this

### Amazon Web Services

Amazon has a lot of simple services it provides including sns,sqs,s3, and ec2 instances. All these services can be used indepentadly from each other. They do require some knowledge to set up aand can get confusing. Go into detail about sns and lambda

### Google

Google also provide cloud services, similar to AWS. Firebase is a platform to help create and deveolpe mobile and web applications. They offer realtime and firestore databases details… the firebase sdk/sopftware development kit can be use with multiply langages like c++, java and javascript. Firebase is widely used with javascript frames like angular and react

### Azure

Azure web services are provided from Microsoft. Like the other services azure provides sql databases both on a server and serverless. Microsoft also developed visual studio, so applications created in visual studios can be easily itigrated into azure web services

## Technologies

Ionic is a open source sdk used to develop hybrid apps, i.e. web applications that are put into a native app shell

Angular is a framework used to create web applications. It is used a front end framework. Helps create web pages with functionality. It is written mostly in typescript

Bootstrap is used to create <better looking> websites. Although it can cause trouble with ionic. It uses html and css to do this

Node js is used to execute javascript code on the server side. But also runds on the front end of the application

.net core is used to build web apps and services in c# and other languages. .net core is the latest version of the .net framework. Which is a software development frame work to build and run applications on windows.it was developed by Microsoft

Selenium is a tool used to preform automated tasks on browser. This includes testing and many webscrapers using webdrivers<explain a bit more>.

Git is a version control system. It helps keep track of files & folders and stores them in the cloud<more detail>

## What kind of app will be developed

Ionic provides the capablility to create apps in both IOS, android and web, each platform required small changes to be compatible with ionic. This app will be developed for android

## Integrated Development Environment

Visual Studio, Visual Studio Code and Replit are all envrments that are used to develop code. They are used for a wide range of projects and langauages.

Visual Studio is used to develop websites webapps apis and more<details>.

Visual Studio Code is code editor. It provides an area where you can create and edit code, it supports multiply different coding languages

Replit is a online ide based in san Francisco. Supports multiply languages<details>

## Programming Languages

JavaScript used to ctreate front end websites. High level

C# can be used with .Net core to create web applications and APIs. High level

typescript is used to help improve javascript. Some javascript frameworks are written in typescript including angular and vue while react supports it. Typescript simplifies javascript code so it can be read and modified more easily

Html is a language used to design displays for web sites In a browser. Css assists in formatting html pages. They take care of the design of a webpage

Json is a frormat for data to be stored in. it helps store objects in a key value pair and lists of different objects. I can be easily read and written by humans. JSON stands for javascript object notation. It can be extremely useful for sending data across http

Python is a high level, general purpose programming language. Can be lightweight but not used for mobile or game dev becasudee slow

SQL stands from Structured Query Language. It is used to communicate with databases to preform crud operations on those databases. It is capable of other processes such as >>>>

<https://www.geeksforgeeks.org/top-programming-languages-for-android-app-development/>

## Web Scraper

Webscrapers are used to get(scrape) data from the world wide web, getting the data from html or web browsers.

Two ways of creating a webscraper are c++ with gumbo Or using python and selenium to interpret dynamiv web pages

# Design-what is used, how I would used it

### Brief Overview of Project

The client will have a mobile application which they will be able to log in and create an account. The main page will be a feed which will display a list of different types of newly released or upcoming media including movies, games, books and shows. The user will then ne able to add the media to their in-app calendar. The user will also be able to message other users and review medias. The mobile application will also store its users in a firestore database, along with other information that is specific to the user, i.e. the calendar entries. User will be notified when a upcoming media they follow is released.

The mobile application will get the media from an api. This api will communicate with a sql database. The api will be capable of preforming crud operations on the sql database. The api will have multiple endpoints, and conform to the REST API criteria.<details>. The api will also be designed using the model view controller pattern. Splitting all the data in models and used by their own controller<details>

The sql database will be filled with data via a webscraper. The webscraper will use python to scrape specific websites and store the data in the sql database

## APP

All the pages. Calendar,forums,login,media,messages,profile, services details on all, what they used

## API

Models- all plus notations used. Data interface and sql. Controllers. Microsoft.EFcore./tools/sqlserver/design, system.data.sqlClient,Newton.Json. migrations, adding imgs

## Webscaper

Scrape websites for newly released medias

Pandas,selenium,webdriver,date,pyodbc,uuid, stop duplicates

# Implementation-how I used it physically in the code

Step by step, even or especially when things don’t work, it demonstrates how you spent your time.

## API Development

Intro

### Timeline-remember why or what made me do it this way

Create model/models, form on what information I wanted in my api

Create data interfaces

Create controller for separate model,added crud to media controller

Create dummy data to get from in a class.

Had a problem with my personal computer, wasn’t able to use it for a 2 weeks. This happened near the start of development so wasn’t too big of a problem

Start again with better results-changed what I thought might need to be changed, because I had just written it before was easy enough to remember plus I hadn’t done too much up to that point. Added addbdset

Start using git and continue to used for version control, what I learned here

Computer fixed, so started using it again, same SSD so no real trouble

Compatibility issues between computer and laptop net5.0 on laptop net6.0 on computer. Not sure how this happened. I think I tried to update my old project using git with my new project . when I got pack to my laptop it wouldn’t work,told me I needed a newer version orto change the project version.Updated laptop

Create sql database in azure, set up database was simple enough as I had done it before. Just bit of setup

Implement Entity frameworks and classes, downloaded Microsoft entity frameworks. Use this to create table in database. Had to add notations so models translated to database tables [key],[required]

Problem connecting to database, wasn’t sure what format the connection string had to be but then found azure provided the string. Also had to add dbcontext in startup.cs and add scope for each interface and sql class pair for controllers to work

When I tried to connect was rejected because my ip adrees wasn’t accept. Figure out how to add ip address to security firewall

Adding migrations, problem with models as all medias were children of media class. Wouldn’t work with ef.i tried looking up a solution but couldn’t find one. Endned up seperateing each child of class media into their own separate class, didn’t like this but I wanted to make it work. Later on in development found out you could add children classes to database

Added migration and updated database

Use postman to add and get models from api. Tested out the api which help create and redesign the controllers and data interfaces and data sql classes. Used methods provide by data context instead of raw sql queries.

Used entity framework to add not mapped lists, which would later be used to fill the list in the api. These lists stored lists of objects. I had an object called series which would have a list of medias. I wasn’t able to stoe the list in the sql database so I found you could use [NotMapped] to be ignore when adding migrations. Iused it with medias as well as each media has a list of themes and genres and actors associated with them.

Reworked season models from series-season-episode to series-shows-season-epsodes. Now had movies,books,shows and games where all apart of series. Is a bit canfusing as I changed the meaning of series in the context of the api.

Used post man to add dummy data to api so there would be data to test mobile app. So all the data in media models where filled. Which turned out different when using the webscraper.

I had a problem with the mobile app and api communicating due to cors. After looking for a solution and finding that it could be solved by adding app.useCors() to the startup.cs file.

Change api each controller to accept files/wwwroot folder Using fromform. Watched tutorial on adding images. Using IWebHostEnviroment, which save the images in a folder in the project. I was going to used this to add information from the webscraper but decided to add straight to the sql instead. Will be used in the future to edit and add from by admins? Mods?

Added one to many tables to connect medias with people, themes and genres in sql database, although only themes is used as not al lot of data was gotten from the websites scraped(should have built sraper first, why did I make the api first. Because I knew how to)

Update models to store date as date in models and database instead of as string. Used [Column(TypeName = “Date”)], kept reading that I should never store dates as strings in a sql database. Also thought I would have a problem in the future if I tried to sort/order by date.

Updated models to store extra data that could be gotten from the webscraper.deleted dummy data/decided not to use api to add to sql database and add straight to database. Would have required a lot of repetitive classes. Better to try another methods to show I can do more? Not sure what I would be marked on so I tried use as much technologies that I could, which has negatively affected the project. No part is fully developed and only does the bare minimum.

Added feed api to get multiple medias with one api call. Merge all medias(because the database already order then in there individual tables by date) ordered by date. Thought this methods was better than sorting them while in an api response request. Cut down time

Deployed to azure using visual studios

### Future Work

#### Biggest issues

Another problem was that the mobile app has to wait around 10 seconds for the data to be sent back. Will have to fix this in the future. Which is kind of ridiculous

#### Other work

Plan to add api keys, shouldn’t be too hard and how I plan to implement them, use user id generated by firebase? No because then other people wont be able to use the api. How about send api request with your email, use sns to automatically send you an api key associate with your email address, if you confirm with that email. An api key would be good if I wanted to monitor how individually user(anoumously, don’t need to know who they are) use the api(data which could be used to improve the api). Could also use api keys to restrict amount of requests so I amnit charged a furtune if somebody makes thousands of requests.

Refactor api to be similier to <https://the-one-api.dev/v2>. Really liked the design of the api and how one request can be formatted different ways, using different search perametters.

### Observations

Why create api before webscraper? Different data found than system was created for had to refactor. Bit off more than I could chew. Added functionality that I never used because it would require a lot of time to used right, more complicated webscraper to get all the data I wanted, not impossible just more time than I had to complete this project again bit off more than I could chew. Although it will be easier to scale the project as some aspects are already in place, but are they in the right place?

## APP Development

Intro

### Timeline

Generated application. Generated using ionic, I tried using different templates including sidebar and tab. Deleted and restarted as I kept running into problems when I change some aspect of the routing. Just deleted them instead of figuring out the problem

Generated tabs. Decide to used this template. Have a tab for each media. This was later change to a feed that would get all medias with a parameter to search specific media

Added pages for the different medias. At the start this was all I thought I needed. And a calendar. Didn’t do much after that

Had a problem with my personal computer, wasn’t able to use it for a 2 weeks. This happened near the start of development so wasn’t too big of a problem. Was able to use a laptop in the mean time

Computer fixed.

Compatibility issue between laptop and computer. Laptop had up to date ionic while computer had older version(from december). Used version of computer on both. So downgraded on laptop to ionic 5.

Redone app to use feed instead of a page for each media

Added calendar,details-media,forums,login,sighup,media/feed,messages,profile,sidenav. Added a place holder for each of these pages. Calendar would have a calendar using a module specifically for android. Details-media was to bring up more detail on a media when clicked on. As of writing this thought of a solution? Feed api gets less information so it is faster? Just title, image, release date and media id. Then when loading details-media use a more specific api call to get the media data. Would cut down wait time.

Created a login and signup page utilizing the firestore database. I added the firebase api information into environment.ts and added the AngularFiremodule. Fireservice as a method to add a user with an email and password. Then the information email,password and generated uid is added to a firestore database. Then a user could sign up and login successfully. forums and messages were place holders which were going to be used to message other users on the app and create posts on different forums

Problems with sidenav. New releases of ionic quickly, hard to find solution on ionic version I am using. Different methods to navigate in ionic between ionic 4 5 and 6. So it was hard to find a tutorial for ionic 5. decided to move on to more important parts. Still having trouble with it. Tried using mwthods from ionic 4 and 6. And watch many tutorials and I could only get it t work once and only once. Think it is something to do with the folder placement or the sidenav only loading once as all pages are loaded inside sidenav. Can figure out solution.

I tried adding the media api I create as a service in ionic but I coulnt get it working so I moved the request to the feed page. The problem wasn’t in the service it was with how is used the observable.

When I got the api connected I then ran into another problem. Which was cors. Cross-origin resource sharing is a mechanism that restricts http requests from scripts. From what I understand it send a request to an api, the api returns the response and the domain is not the same as the request the response is rejected

I tried setting up requests that would get rid of the error but they didn’t seem to work. I Ended finding a solution to fix this cors problem on the api instead.

Added functionality on media.page.html to show data received from the api. Also added the ability to get more details by saving the tapped media in a state and bringing that state to another page where more information was show as well as more medias in the same series as the current media

I tried using different if statements in the html to sort the list of data into different media types. As each media had different bits of information, this didn’t matter too much as the information wont show up if its bot there on the webpage.The problem came when I tried to dislay the iamages. The page failed to load as the image usrls where store using just their name i.e. picture.jpeg. I added the url in the ionic app. Each image for each image was stored in a different folder in the wwroot file so I had to check what type the media was to figure out the folder its image was stored in. The user interface didn’t look too good after I spent a good amount of time developing it and didn’t look to pretty in the code either so I deleted it and decide to use bootstrap.

Later found out that ionic and bootstrap don’t work well together as both of them using the DOM in different ways.

Add calendar functionality to keep track of media added to calendar. Just so the user could see them

Figure out how to use sns with specific users. Went through few ideas. Decide to use python again, as I had created the webscraper by this point. Add table to database which stores mobile number and media id. Run python script daily if media date in today create topic and get all mobile numbers using media id. Send text to all with details without subscriping first as text messages can be sent without subscribing

### Future Work

Add messages,forum,user reviews api keys per account, better css as app doesn’t look pleasant at all.

### Observations

I found ionic to be useful in that it has a lot to modules that can be used to enchance to application but I have always found css,html and a lot of ionic to be tedious. Typescript can be hard to understand by just looking at it. Although I can seem why it is well used as it can be extremely powerful when used properly.

## Web Scraper Development

### Timeline

Watch a tutorial about using a webscraper with python and selenium

Use replit to create webscraper using python. As I didn’t want to install all the python,pi and dependencies locally especially when the final webscraper would be running on a aws lambda.

Imported selenium with webdrivers, these are automatically installed with replit so I didn’t have to worry about setting up the selenium webdriver.

Tested getting a page, and printed the results, after getting the webpages information, I then started looking for websites that had the realease dates of movies,books,games and shows, I also wanted them to be reoccurring as if they weren’t it would kind of defeat the reason of this webscraper. So if they were reoccurring I could scrape then once a day and have up too date data in my sql database. I got two or three websites from each media.

I tried out all on the movie website metacritc first before implementing it with other media i.e. repeat with books, shows, games. After setting up the driver with chrome using selenium I was able to get the page. Then used the find\_elements(By.TAG\_NAME,’tr’). This would get every tag,and everything in that tag and return a list. The tutorial I was following was using a find\_elements method that was deprecated so I had to search up the current way to usethe method, luckily it hadn’t changed much. I have never formolly learned python only in small past project so I wasn’t familiar with all the syntax. I did watch a few basic tutorials so I would understand what I was doing a bit better. Someof the syntax used I the webscraper tutorial I didn’t understand at all, but by the time I finished the webscraper I understood exactly what I was doing. Although I do think there are glaring gaps in my usage of python.

To parse the medias I created a function for each media. Parse\_movie\_metacritic would take in one moviei.e. element from the webpage and return a python object I could use.the webscraper tutorial returned a data structure I didn’t understand and python threw no exceptions. Weird that python didn’t need me to manually initiate object as return method otr oject so culdnt find out what it was. Found out it was called a dictionary store string key object value pair. Parse\_movie\_metacritic threw exception because each second tr was empty with other empy one randomly. Had problems to figure out how to handle exceptions try except

Had 2 loops one went through list of trs and removed empty tags. Then other loop parse the data into a list of dictionarys. Refactor using if is not None(learning new things) split time in half especially when all web pages where queried

Turned list of dictioary into csv file using panda dataframe.

Some movies didn’t have release dates, so I didn’t want to use them, added another if statement to handle if the dictionary[date] had None

I ran into this problem with both shows and games websites

Then I had to researched how to add the data to an sql database.

Replit doesn’t support pyodbc as it needs a sql driver

Set up python and pip,on my computer and laptop,Download odbc driver for sql server. Now have to configure selenium on pc. Install webdriver manager to install driver during runtime

Little problem writing add\_to\_db methods. With sanitization and dictionary(which is what I was adding to the data base) being null

Created method to change string date to type date so it would be stored properly in the database. Methods can change multiple date string formats into suitable date format

Problem with null fields and unique ids for mediaId. Imported uuid<> which is the same as guid<>. Learned what these both do<detail>.

The method is inside a loop so for each element add to database. A lot of sql queries should try do less. Could have let the add\_to\_db method take in a list and all the data to be inserted together and let the sql database handle it as one sql statement. However I didn’t want to set up a second loop to handles the now full set(no null columns) data. Would have increased the runtime by a good amount according to the big O(n) this shouldn’t matter. Lambda has a 15 minute limit so wanted to keep the time down especially if I want to add more to this webscrpare, although I could just create another separate webscraper.

Created stored procedure to delete duplicates if title release date and description are the same. Little bit complicated. What if media changes date? Or two different media by the same name ae coming out??

Created triggers for stored procedure, to trigger when insert into media/movies/shows/games/books. (if so then stored procedure is triggered every time there is a request, will need to refactor if I want to be more effecient)

Add another stored procedure to order all medias by release date. Faster than order every request.

Deploy webscrpaer on aws lambda(steps to take)

### Future Work

Code doesn’t look neat and hard to read/ refator or redesign

Add profanity filter

Make another wenscraper to search for trailers for medias in sql and add them to sql database

### Observations

Learned a lot about python and enjoyed it

# Conclusions

The application …

# Source Code

|  |  |
| --- | --- |
|  |  |