**FASTPLEX MOVIES HANDLING SYSTEM**

**WELCOME TO THE WORLD OF MOVIES**

**DATASET:**

**Contains over 200k movies watched by 7k+ unique users along with ratings and genre**

**SIZE: (2GB)**

**Group Members:**

* **Anas Ali 20k-0181**
* **Muhammad Umer 20k-0225**
* **Zeeshan Aijaz 20k-0361**

**INTRODUCTION:**

Our project was to handle a complex data set and perform operations on it such as cleaning, handling, managing, searching, et cetera.

We chose a dataset from Kaggle(An online website commonly used for acquiring datasets ) as it was more than enough and was challenging enough to handle the data and display everything related.

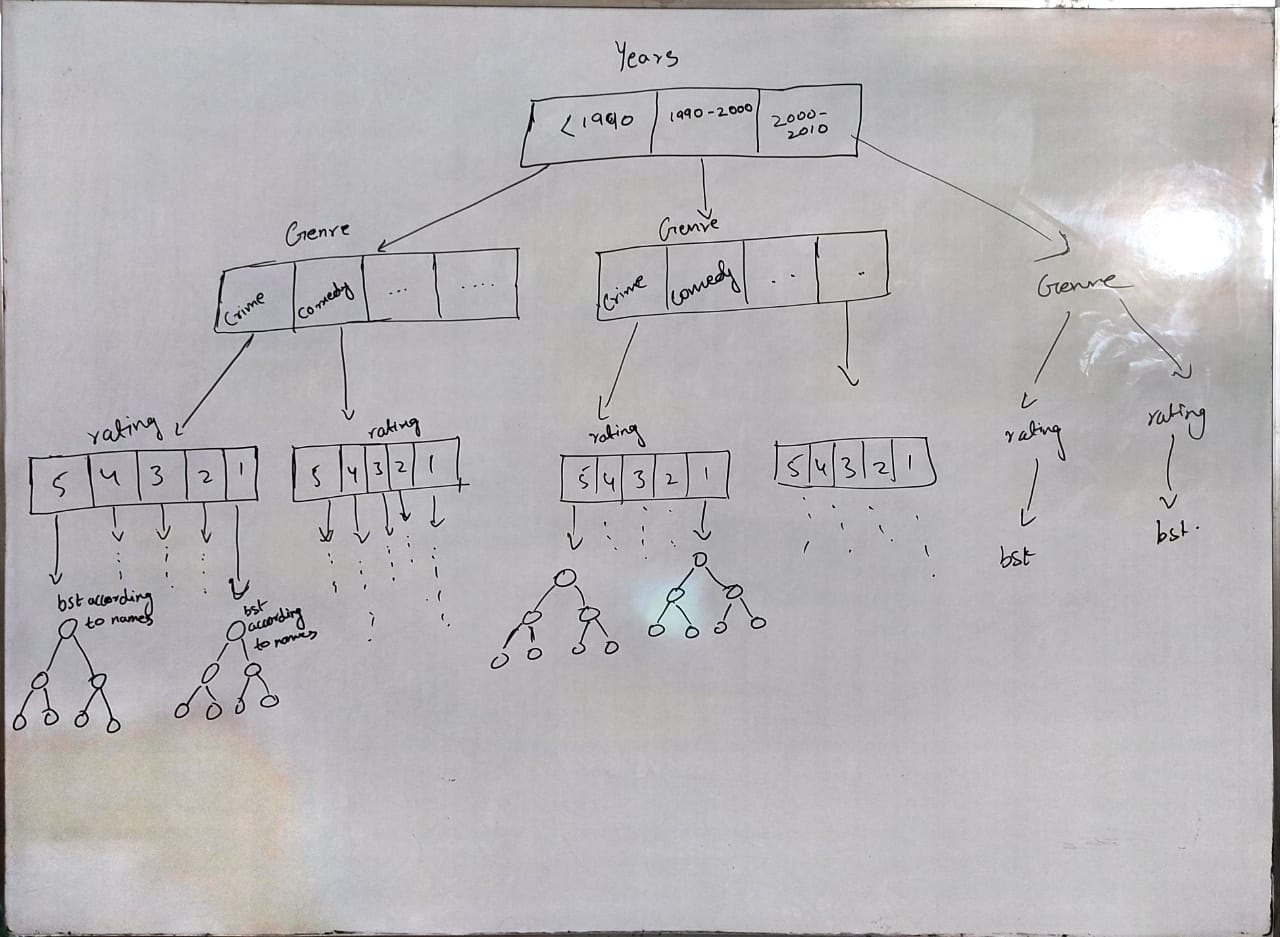
**PROLEM DEFINATION:**

The problems that were faced during the process of the project were

* Which data structure to use?
* How to load .csv file to .cpp program?
* How to clean such large file.
* How to access data so that least searching time is consumed?
* What type of operations should be performed?

**METHODOLOGY:**

* After discussing we decided to use a unique combination of B-tree and binary search tree for loading the .csv file, as it was the most efficient approach and traversal time was efficient O(1) is best cases.
* Data-structure used:



* For loading the file to our data-structure, we used filing and its commands.
* For accessing and searching the movies, genres ratings and years, we used integers and floats and for strings, we used functions from the string library, both of which return the properties that have the highest similarity as of entered by the user.
* Built-in libraries were used for abstraction and for increasing readability and optimization.
* Queues were also used for BFS traversal in BST, which matched with the requirements of the user and displayed the movies accordingly.

**FEATURES:**

The salient features of our project were:

* Efficient searching in big data
* Encapsulation
* Suggestions of movies for user
* Movie name searching
* Searching through genres
* Searching through time period

**IMPLEMENTATION, TESTING & DESIGN:**

We created our project in CPP. The IDE we used was visual studio code, as it is the best platform present for a project. Our basic motive was to create a program which helps the user to find their favorite of potential favorite movie(s), from a data of millions!

The work flow was as follows:

* Decision on using a combination of b-trees and BST for implementation
* Loading the tree from .csv file
* Creating searching functions/algorithms that gave the best results

**CONCLUSION:**

The two important things that were kept in mind while creating the project were best time complexity and searching with maximum productivity in this voluminous data set.

The project was created with the help of all the team members, which share equal participation and contribution.