# Chương 1: Design

## 1.1. Nguồn dữ liệu

### 1.1.1. API từ Website RoPhim

Các API công khai của RoPhim được phát hiện và truy xuất thông qua công cụ Developer Tools (F12) trên trình duyệt. Danh sách các API được sử dụng:

* list-movie: danh sách phim theo từng trang

https://api.rophim.me/v1/movie/filterV2?type=[1,2]&exclude\_status=Upcoming&sort=release\_date&page={page}

* cast-of-movie: dàn diễn viên chính của phim

https://api.rophim.me/v1/movie/casts/{movie\_id}

* list-comment-or-review: danh sách bình luận hoặc đánh giá

https://api.rophim.me/v1/comment/list?movie\_id={movie\_id}&is\_reviews=[0,1]

* list-reply-comment: danh sách các bình luận phản hồi

https://api.rophim.me/v1/comment/replyList?parent\_id={parent\_id}

### 1.1.2. Web Scraping

Đối với các thông tin không có sẵn qua API, chúng tôi sử dụng kỹ thuật Web Scraping để trích xuất dữ liệu trực tiếp từ giao diện web. Các trường dữ liệu thu thập theo cách này bao gồm:

* production-and-distribution: nền tảng chiếu phim, nhà sản xuất và đạo diễn

## **1.2. Dữ liệu trả về từ API là gì ?**

### 1.2.1. Danh sách phim theo trang

|  |  |  |
| --- | --- | --- |
| **Trường dữ liệu** | **Kiểu dữ liệu** | **Mô tả** |
| \_id | String | ID của phim |
| original\_title | String | Tên gốc của phim |
| english\_title | String | Tên tiếng Anh của phim |
| title | String | Tên tiếng Việt của phim |
| slug | String | SEO URL slug |
| overview | String | Mô tả nội dung phim |
| release\_date | String | Ngày phát hành (yyyy-MM-dd) |
| quality | String | Chất lượng phim |
| rating | String | Giới hạn độ tuổi |
| runtime | Integer | Thời lượng phim (phút) |
| type | Integer | Loại phim (1: phim lẻ, 2: phim bộ) |
| origin\_country | Array[String] | Quốc gia sản xuất phim |
| status | String | Tình trạng phim |
| latest\_season | Integer | Mùa phim mới nhất |
| imdb\_rating.$IntegerDecimal | String | Điểm đánh giá IMDb |
| latest\_episode | Object | Thông tin tập mới nhất theo mùa |
| year | String | Năm phát hành |
| genres | Array[Object] | Danh sách thể loại ( \_id, name, slug) |
| images.posters | Array[Object] | Danh sách ảnh poster chính |
| images.horizontal\_posters | Array[Object] | Danh sách ảnh poster ngang |
| images.backdrops | Array[Object] | Danh sách ảnh nền phía sau |
| images.titles | Array[Object] | Danh sách ảnh tiêu đề phim |

### 1.2.2. Dàn diễn viên của phim

|  |  |  |
| --- | --- | --- |
| **Trường dữ liệu** | **Kiểu dữ liệu** | **Mô tả** |
| \_id | String | ID của record cast trong phim |
| movie\_id | String | ID của phim |
| cast\_id | String | ID của diễn viên |
| character | String | Tên nhân vật do diễn viên thủ vai |
| order | Integer | Vị trí thứ tự hiển thị |
| created\_by | Integer | ID người tạo record này |
| updated\_by | Integer | ID người cập nhật record |
| created\_at | Integer | Thời gian tạo record (UNIX timestamp) |
| updated\_at | Integer | Thời gian cập nhật record |
| \_\_v | Integer | Phiên bản document trong MongoDB |
| cast.\_id | String | ID của diễn viên |
| cast.name | String | Tên diễn viên |
| cast.slug | String | Đường dẫn thân thiện với SEO của diễn viên |
| cast.profile\_path | String | Đường dẫn ảnh đại diện của diễn viên |
| cast.publish | Boolean | Diễn viên này có được công khai hay không |

### 1.2.3. Danh sách bình luận hoặc đánh giá

|  |  |  |
| --- | --- | --- |
| **Trường dữ liệu** | **Kiểu dữ liệu** | **Mô tả** |
| \_id | String | ID của bình luận |
| parent\_id | String / null | ID của bình luận cha (nếu là bình luận reply) |
| user\_id | String | ID của người bình luận |
| movie\_id | String | ID phim mà bình luận thuộc về |
| episode\_Integer | Integer | Số tập (nếu bình luận cho tập cụ thể), 0 nếu là bình luận chung |
| season\_Integer | Integer | Số mùa (nếu có), 0 nếu là bình luận chung |
| is\_spoil | Boolean | Có chứa spoil không (true/false) |
| mention\_id | String / null | ID người dùng được đề cập nếu có |
| total\_children | Integer | Số lượng phản hồi (reply) cho bình luận này |
| total\_like | Integer | Số lượt thích |
| total\_dislike | Integer | Số lượt không thích |
| content | String | Nội dung bình luận |
| is\_pinned | Boolean | Có được ghim không |
| is\_reviews | Boolean | Có phải là đánh giá (review) không (true/false) |
| reviews\_id | String / null | ID bài đánh giá nếu đây là comment review |
| created\_at | Integer | Thời gian tạo bình luận (UNIX timestamp) |
| author.\_id | String | ID người dùng |
| author.name | String | Tên người dùng |
| author.role | String | Vai trò người dùng |
| author.avatar\_id | String | ID avatar của người dùng |
| author.gender | Integer | Giới tính người dùng |
| author.avatar.\_id | String | ID của hình đại diện |
| author.avatar.path | String | Đường dẫn hình ảnh đại diện |
| mention\_user | Object / null | Thông tin người dùng được đề cập |
| reviews.\_id | String | ID đánh giá |
| reviews.point | Integer | Số điểm đánh giá |

### 1.2.4. Danh sách bình luận phản hồi

|  |  |  |
| --- | --- | --- |
| **Trường dữ liệu** | **Kiểu dữ liệu** | **Mô tả** |
| \_id | String | ID duy nhất của bình luận. |
| parent\_id | String | ID của bình luận gốc mà bình luận này trả lời. |
| user\_id | String | ID của người dùng viết bình luận. |
| movie\_id | String | ID bộ phim mà bình luận liên quan đến. |
| episode\_number | Integer | Số tập phim (0 nếu không áp dụng). |
| season\_number | Integer | Số mùa phim (0 nếu không áp dụng). |
| is\_spoil | Boolean | Cho biết bình luận có spoil nội dung phim hay không. |
| mention\_id | String | ID của người dùng được nhắc đến trong bình luận. |
| total\_children | Integer | Số lượng phản hồi (reply) của bình luận này. |
| total\_like | Integer | Số lượt thích bình luận. |
| total\_dislike | Integer | Số lượt không thích bình luận. |
| content | String | Nội dung văn bản của bình luận. |
| is\_pinned | Boolean | Cho biết bình luận có được ghim hay không. |
| is\_reviews | Boolean | Cho biết bình luận có phải là đánh giá (review) hay không. |
| reviews\_id | Null / String | ID của bài đánh giá liên kết, nếu có (null nếu không phải đánh giá). |
| created\_at | Integer | Thời gian tạo bình luận (Unix timestamp). |
| author.\_id | String | ID người dùng đã viết bình luận. |
| author.name | String | Tên hiển thị của người dùng viết bình luận. |
| author.role | String | Vai trò của người dùng (ví dụ: "member"). |
| author.avatar\_id | String | ID ảnh đại diện của người dùng. |
| author.gender | Integer | Giới tính người dùng (1 = nam, 2 = nữ, 3 = khác / không rõ). |
| author.avatar.\_id | String | ID của ảnh đại diện. |
| author.avatar.path | String | Đường dẫn đến ảnh đại diện của người dùng. |
| mention\_user.\_id | String | ID của người dùng bị nhắc đến (mention) trong bình luận. |
| mention\_user.name | String | Tên của người dùng bị nhắc đến. |
| reviews | Null | Đối tượng review liên kết |

## 1.3. Khoảng thời gian truy vấn API

## 1.4. Khối lượng dữ liệu cần xử lý

## 1.5. Thiết kế cơ sở dữ liệu

## 1.6. Xác định bài toán và yêu cầu phân tích dữ liệu

Business Requirement #1: Phân tích xu hướng phim