

Movie ratings

PROJECT IDEA

Ru Chen

A G E N D A

- About Movie Rating
 - User types
 - Challenges
 - Solution
 - Data Assessment
 - FAIR Assessment
 - Quality control
- A presentation file as a PDF stored in Github that you will present in class. Focus on a summary of your project that touches on the main points described above for:
 - Info story or who is the user and why is it important to them
 - Existing structure and FAIR assessment,
 - How you decided to improve the structure,
 - What your new structure is, and
 - How would quality be identified and addressed

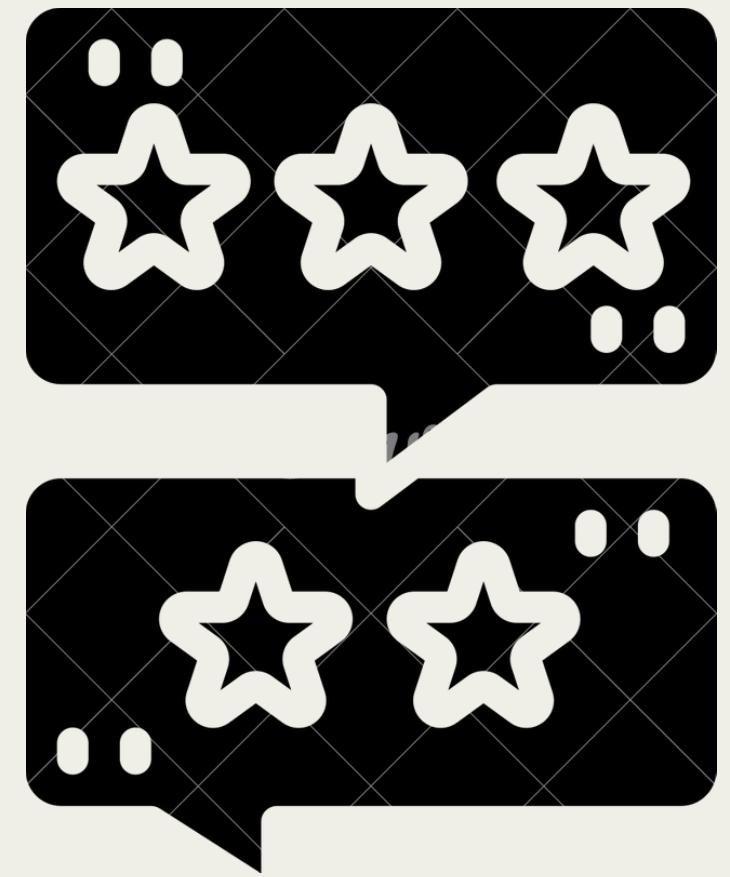
ABOUT MOVIE RATING

An application that consolidates movie ratings from three major sources:

- IMDb
- TMDb
- Metacritic.

To provide a comprehensive view of each movie's reception across these platforms.

Helps users to make informed viewing choices to choose movies based on the information.



USER TYPES

- 1. Movie Enthusiasts**
- 2. Casual Viewers**
- 3. Film Critics and blogger**



USER TYPES

1. Movie Enthusiasts

who are passionate about movies and frequently seek out new films to watch. often rely on ratings and reviews to decide which movies to see next.

Goals

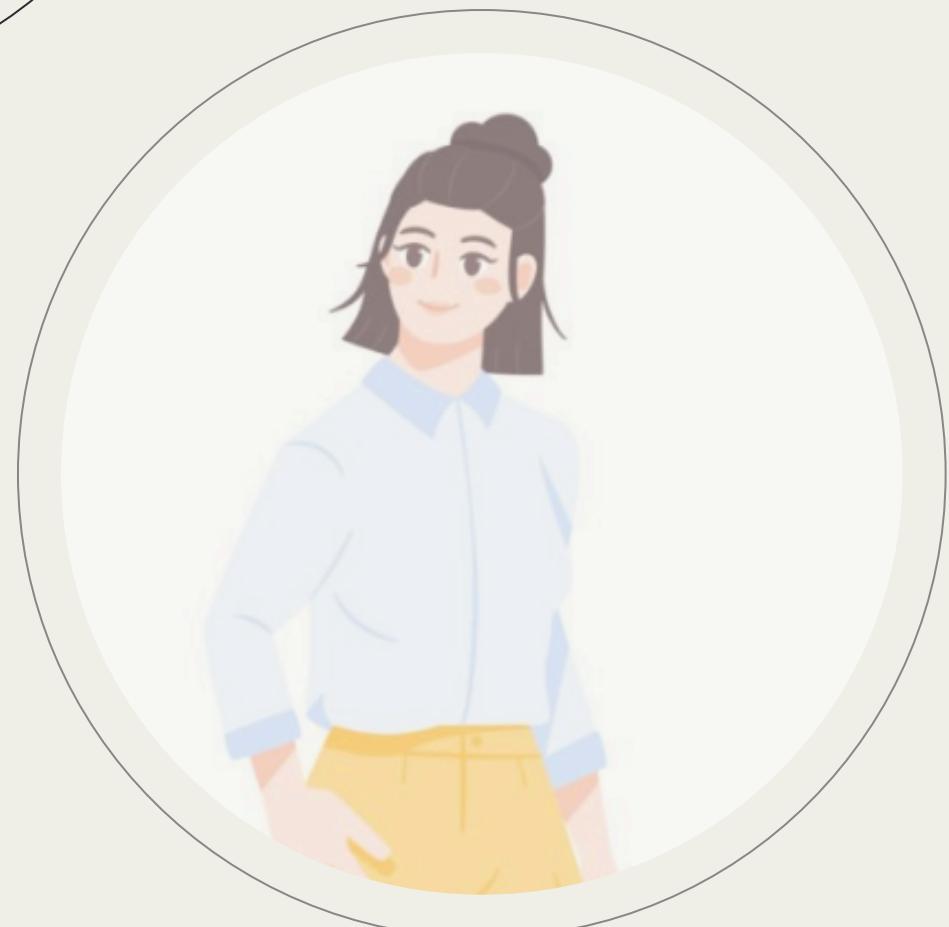
- Discover new movies based on ratings
- compare ratings across different platforms
- make informed decisions about what to watch

Challenges

- finding consolidated and reliable ratings
- navigating multiple websites to gather movie information

Usage

- check average ratings to quickly decide on movie
- view aggregated ratings from IMDb, TMDb and Metacritic.



USER TYPES

2. Casual Viewers

who watch movies occasionally and prefer to watch highly-rated or popular movies.

Goals

- Quickly find highly-rated movies
- get recommendations for movies based on ratings

Challenges

- lack of time to research and compare movie

Usage

- Use the application to quickly access top-rated movies



CHALLENGES

Search options:

1. Search movie titles on Google
2. direct search from other movie reviews

A screenshot of a Google search results page for the query "avatar 2". The top result is for "Avatar: The Way of Water" (PG-13, 2022). The page includes a large image of the movie, a trailer thumbnail, and sections for Overview, Cast, Watch movie, Reviews, Trailers & clips, and Behind the scenes. A red box highlights the "Where to watch" section, which shows Disney+ Subscription and other streaming options. Below this, a red box highlights the ratings section, which includes IMDb (7.6/10), Rotten Tomatoes (76%), and Metacritic (67%).

Avatar: The Way of Water | Official Trailer

2:29

76% Rotten Tomatoes | 7.6/10 IMDb | 4/5 Common Sens...

88% liked this movie Google users

Jake Sully and Ney'tiri have formed a family and are doing everything to stay together. However, they must leave their home and explore the regions of Pandora. When an ancient threat resurfaces, Jake must fight a difficult war against the humans.

Release date: December 16, 2022 (USA)

Director: James Cameron

Budget: 250 million USD

Distributed by: 20th Century Studios

Box office: \$2.320 billion

Edited by: Stephen Rivkin; David Brenner; John Refoua; James Cameron

Feedback

CHALLENGES

Search options:

1. Search movie titles on Google
2. Direct search from other movie reviews

The collage consists of five separate sections:

- IMDb Charts - Most Popular Movies:** Shows the top 100 movies on IMDb, including Civil War, Dune: Part Two, Joker: Folie à Deux, and Monkey Man.
- Sponsored Content - ABIGAIL:** A movie poster for "ABIGAIL IN THEATERS FRIDAY GET TICKETS".
- You have rated:** A section showing 0/100 (0%) rated movies, with an option to hide them.
- More to explore:** A section titled "Charts" with a link to "Top Box Office (US) > From the past weekend".
- Popular Streaming Movies:** A list of movies like Fandango at Home, Netflix, Prime Video, Max, etc., with their popularity percentages.
- Most Popular TV on RT:** A list of TV shows with their popularity percentages.
- KINGDOM of PLANET OF THE APES:** A movie poster for "KINGDOM of PLANET OF THE APES" in IMAX, released on MAY 10, with a "GET TICKETS" button.
- Common Sense Media - AVATAR: THE WAY OF WATER:** A review page for the movie, including a Common Sense Selection badge, parents' reviews, and a video review.
- metacritic - Critic Reviews:** A screenshot of the metacritic website showing the Metascore (67), user reviews, and cast & crew information for "Avatar: The Way of Water".

ABOUT MOVIE RATING

IMDb Charts
Most Popular Movies
As determined by IMDb users

Sort by: Ranking

Rank	Title	Rating	Year	Length	Genre
1 (+ 8)	Civil War	7.6 (15K)	2024	1h 49m	R
2 (+ 1)	Dune: Part Two	8.7 (40K)	2024	2h 46m	PG-13
3 (+ 30)	Joker: Follie à Deux	8.0	2024	R	
4 (+ 1)	Monkey Man	7.3 (14K)	2024	2h 1m	R

You have rated
0/100 (0%)

More to explore

Top Box Office (US) > From the past weekend



POPULAR STREAMING MOVIES [VIEW ALL](#)

- Fandango at Home | Netflix | Prime Video | Max | More...
- Argyle** ★ 33%
- Baby Reindeer** ★ 100%
- One Life** ★ 90%
- The Zone of Interest** ★ 93%
- Scoop** ★ 78%
- Kung Fu Panda 4** ★ 72%
- Poor Things** ★ 92%
- The Bricklayer** ★ 50%
- North of Normal** ★ 100%
- Anatomy of a Fall** ★ 96%
- The Sympathizer** ★ 89%

MOST POPULAR TV ON RT [VIEW ALL](#)

- Fallout** ★ 93%
- Baby Reindeer** ★ 100%
- Ripley** ★ 86%
- 3 Body Problem** ★ 80%
- The Sympathizer** ★ 89%
- Parasite: The Grey** ★ 100%
- Sugar** ★ 81%
- Shogun** ★ 99%
- LaRoy, Texas** ★ 100%
- Franklin** ★ 67%
- A Gentleman in Moscow** ★ 90%

NEW TV THIS WEEK

WHAT'S ON TONIGHT

KINGDOM OF THE PLANET OF THE APES [IMAX IN THEATERS MAY 10 GET TICKETS](#)



metacritic Games Movies TV Shows Music News Search Register

Critic Reviews User Reviews Cast & Crew Details

67 METASCORE Generally Favorable

Positive	45 (66%)
Neutral	21 (31%)
Negative	2 (3%)

Showing 68 Critic Reviews [All Reviews](#) [Metascore](#)

100 Empire DEC 13, 2022

James Cameron has surfaced with a cosmic marine epic that only he could make: eccentric, soulful, joyous, dark and very, very blue. Yes, he's still leagues ahead of the pack.

This website uses cookies to enhance user experience and to analyze performance and traffic on our website. We also share information about your use of our site with our social media, advertising and analytics partners. [Privacy Policy](#)

Do Not Sell My Personal Information [Accept Cookies](#)



age 13+ ★★★★½

Long but dazzling return to Pandora has sci-fi violence.

Movie | PG-13 | 2022 | 192 minutes

[Rate movie](#)

Parents Say: age 13+ 39 reviews

Any Iffy Content? Read more

Watch Our Video Review Watch now

Where to Watch Disney+ Amazon MAX Show more ▾

Want more movie suggestions? Add your streaming services

Our Editors Recommend

A Lot or a Little? What you will—and won't—find in this movie.

Positive Messages ★★★½ Messages about acceptance, unity, ...

Positive Role Models ★★★½ The women leaders of the clan ar ...

Diverse Representations ★★★½ The Na'vi species is divided into ...

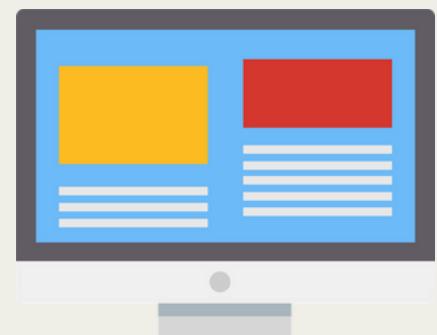
Violence & Scariness ★★★½ Sci-fi action violence. Supporting ...

Sex, Romance & Nudity ★★★½ Brief scene of nonsexual nudity (...

Language ★★★½ Scattered strong language includ ...



Overall movie rates



WIREFRAME



WIREFRAME

<Links to the movie rating dataset>

<input type="text" value="Movie Title"/>	<input type="text" value="Year"/>	<input type="text" value="Genres"/>	<input type="text" value="Score"/>	
IMDb	TMDb	Metacritic	Average	
8.1	8.1	8.5	8.2	
<input type="text" value="Movie Title"/>	<input type="text" value="Year"/>	<input type="text" value="Genres"/>	<input type="text" value="Score"/>	
IMDb	TMDb	Metacritic	Average	
7.1	6.7	8.8	7.6	
<input type="text" value="Movie Title"/>	<input type="text" value="Year"/>	<input type="text" value="Genres"/>	<input type="text" value="Score"/>	
IMDb	TMDb	Metacritic	Average	
7.3	7.4	6.1	6.9	
<input type="text" value="Movie Title"/>	<input type="text" value="Year"/>	<input type="text" value="Genres"/>	<input type="text" value="Score"/>	
IMDb	TMDb	Metacritic	Average	

API link

movie_title	year	genres	directors	IMDb_score	TMDb_score	Metascore	Average_scores
-------------	------	--------	-----------	------------	------------	-----------	----------------

DATA STRUCTURE



API link

Challenges:

Each data structure from different sources, like imdb, tmdb, and metascore has different unique id.

EXISTING DATA STRUCTURE - IMDB DATASET

Title.basixcs.tsv.gz

- tconst (string) - alphanumeric unique identifier of the title
- titleType (string) – the type/format of the title (e.g. movie, short, tvseries, tvepisode, video, etc)
- primaryTitle (string) – the more popular title / the title used by the filmmakers on promotional materials at the point of release
- originalTitle (string) - original title, in the original language
- isAdult (boolean) - 0: non-adult title; 1: adult title
- startYear (YYYY) – represents the release year of a title. In the case of TV Series, it is the series start year
- endYear (YYYY) – TV Series end year. '\N' for all other title types
- runtimeMinutes – primary runtime of the title, in minutes
- genres (string array) – includes up to three genres associated with the title

title.ratings.tsv.gz

- tconst (string) - alphanumeric unique identifier of the title
- averageRating – weighted average of all the individual user ratings
- numVotes - number of votes the title has received
- name.basics.tsv.gz
- nconst (string) - alphanumeric unique identifier of the name/person
- primaryName (string)– name by which the person is most often credited
- birthYear – in YYYY format
- deathYear – in YYYY format if applicable, else '\N'
- primaryProfession (array of strings)– the top-3 professions of the person
- knownForTitles (array of tconsts) – titles the person is known for

[Link to the dataset](#)



EXISTING DATA STRUCTURE - TMDB DATASET

Need to create a user account and request to access the API

Collect the data from Kaggle, who recently analyzed the data that is collected from TMDB API

- Id (int): unique identifier for each movie.
- Title: (string): title of the movie
- Vote_average (float): average vote or rating given by viewers
- Vote_count (int): total count of votes received for the movie
- Status (string): the status of the movie (e.g., released, rumored, post production, etc.)
- Release_date (string): date when the movie was released
- Revenue (int): total revenue generated by the movie
- Runtime (int): duration of the movie in minutes
- Adult (bool): indicates if the movie is suitable only for adult audiences
- Backdrop_path (string): URL of the backdrop image for the movie

EXISTING DATA STRUCTURE - METACRITIC DATASET

Needs to do web crawling

Collect the data from Kaggle, who web crawled the dataset for analyze

- Movie_title (unique values)
- Release_data
- Genre
- Meta_mixed
- Meta_negative
- Meta_positice
- Metascore
- User_mixed
- User_negative
- User_positive

DATA TRANSFORMATION - IMDB DATASET

IMDb Non-Commercial Datasets:

- 4 tsv.gz files to integrate to analyze the information

imdb id	Movie title	Year	genres	IMDb score	director1	director2
tt0013274	Istoriya grazhdanskoy voyny	2021	Documentary	6.8	Nikolai Izvolov	Dziga Vertov
tt0015414	La tierra de los toros	2000	\N	5.4	Musidora	
tt0035423	Kate & Leopold	2001	Comedy,Fantasy,Romance	6.4	James Mangold	
tt0062336	The Tango of the Widower and Its Distorting Mirror	2020	Drama	6.4	Raúl Ruiz	Valeria Sarmiento
tt0069049	The Other Side of the Wind	2018	Drama	6.7	Orson Welles	
tt0070596	Socialist Realism	2023	Drama	7.5	Raúl Ruiz	Valeria Sarmiento
tt0082328	Embodiment of Evil	2008	Horror	5.9	José Mojica Marins	
tt0088751	The Naked Monster	2005	Comedy,Horror,Sci-Fi	5.2	Wayne Berwick	Ted Newsom
tt0096056	Crime and Punishment	2002	Drama	5.5	Menahem Golan	
tt0096235	Taxi Killer	2022	Action,Crime,Drama	5.6	Stelvio Massi	
tt0097304	Everything's for You	2009	Documentary	7	Abraham Ravett	
tt0097767	Loading Ludwig	2022	\N	7	Mara Mattuschka	
tt0100275	The Wandering Soap Opera	2017	Comedy,Drama,Fantasy	6.5	Raúl Ruiz	Valeria Sarmiento
tt0102362	Istota	2000	Drama,Romance	6.3	Andrzej Czarnecki	
tt0104988	Neues in Wittstock	2021	Documentary	7.8	Volker Koepp	
tt0108116	Shepherd on the Rock	2014	Drama	5.8	Bob Keen	
tt0108549	West from North Goes South	2004	Comedy,Mystery	7.8		

DATA TRANSFORMATION - TMDB DATASET

TMDB API access
non-commercial purposes

imdb id	movie_title	Year	genres	director1	director2	IMDb score	tmdb score
tt0013274	Istoriya grazhdanskoy voyny	2021	Documentary	Nikolai Izvolc	Dziga Vertov	6.8	6
tt0015414	La tierra de los toros	2000	IN	Musidora		5.4	0
tt0035423	Kate & Leopold	2001	Comedy,Fan	James Mangold		6.4	6.328
tt0062336	The Tango of the Widower and Its Distorting Mirror	2020	Drama	Ra,à,ò,à 'l Ru	Valeria Sarmi	6.4	6.167
tt0069049	The Other Side of the Wind	2018	Drama	Orson Welles		6.7	6.698
tt0070596	Socialist Realism	2023	Drama	Ra,à,ò,à 'l Ru	Valeria Sarmi	7.5	0
tt0088751	The Naked Monster	2005	Comedy,Hor	Wayne Berwi	Ted Newsom	5.2	4
tt0096056	Crime and Punishment	2002	Drama	Menahem Golan		5.5	5.333
tt0096235	Taxi Killer	2022	Action,Crime	Stelvio Massi		5.6	3
tt0097304	Everything's for You	2009	Documentary	Abraham Ravett		7	0
tt0097767	Loading Ludwig	2022	IN	Mara Mattuschka		7	0
tt0100275	The Wandering Soap Opera	2017	Comedy,Dra	Ra,à,ò,à 'l Ru	Valeria Sarmi	6.5	6.6
tt0104988	Neues in Wittstock	2021	Documentary	Volker Koepp		7.8	0
tt0108116	Shepherd on the Rock	2014	Drama	Bob Keen		5.8	5.5
tt0109173	Auf allen Meeren	2002	Documentary	Johannes Holzhausen		7.8	6

DATA TRANSFORMATION - METACRITIC DATASET

It requires data scrapping to gather the data.
Used existing Metacritic datasets from Kaggle.

imdb id	movie_title	Year	genres	director1	director2	IMDb score	tmdb score	metascore
tt0013274	Istoriya grazh	2021	Documentar	Nikolai Izvolc	Dziga Vertov	6.8	6	0
tt0015414	La tierra de lo	2000	\N	Musidora		5.4	0	0
tt0035423	Kate & Leopo	2001	Comedy,Fam	James Mangold		6.4	6.328	44
tt0062336	The Tango of I	2020	Drama	Ra,àö,à 'l Ru	Valeria Sarm	6.4	6.167	0
tt0069049	The Other Sic	2018	Drama	Orson Welles		6.7	6.698	78
tt0070596	Socialist Rea	2023	Drama	Ra,àö,à 'l Ru	Valeria Sarm	7.5	0	0
tt0088751	The Naked M	2005	Comedy,Hor	Wayne Berwi	Ted Newsom	5.2	4	0
tt0096056	Crime and Pu	2002	Drama	Menahem Golan		5.5	5.333	0
tt0096235	Taxi Killer	2022	Action,Crime	Stelvio Massi		5.6	3	0
tt0097304	Everything's f	2009	Documentar	Abraham Ravett		7	0	0
tt0097767	Loading Ludw	2022	\N	Mara Mattuschka		7	0	0

DATA STRUCTURE - MOVIE RATING DATASET

A	B	C	D	E	F	G	H	I	J	K
imdb_id	movie_title	Year	genres	director1	director2	IMDb score	tmdb score	metascore	metascore_n	average_score
tt0013274	Istoriya grazh	2021	Documentar	Nikolai Izvolc	Dziga Vertov	6.8	6	0	0	4.3
tt0015414	La tierra de los	2000	\N	Musidora		5.4	0	0	0	1.8
tt0035423	Kate & Leopo	2001	Comedy,Fan	James Mangold		6.4	6.328	44	4.4	5.7
tt0062336	The Tango of t	2020	Drama	Ra,àò,à 'l Rui	Valeria Sarm	6.4	6.167	0	0	4.2
tt0069049	The Other Sic	2018	Drama	Orson Welles		6.7	6.698	78	7.8	7.1
tt0070596	Socialist Rea	2023	Drama	Ra,àò,à 'l Rui	Valeria Sarm	7.5	0	0	0	2.5
tt0088751	The Naked M	2005	Comedy,Hor	Wayne Berwi	Ted Newsom	5.2	4	0	0	3.1
tt0096056	Crime and Pu	2002	Drama	Menahem Golan		5.5	5.333	0	0	3.6
tt0096235	Taxi Killer	2022	Action,Crime	Stelvio Massi		5.6	3	0	0	2.9
tt0097304	Everything's f	2009	Documentar	Abraham Ravett		7	0	0	0	2.3
tt0097767	Loading Ludw	2022	\N	Mara Mattuschka		7	0	0	0	2.3

IMPROVED DATA STRUCTURE- MOVIE RATING DATASET

- imdb id (string) - alphanumeric unique identifier of the title
- movie_title(string) - the movie publish name
- Year (YYYY) – represents the movie release year
- genres (string array) – includes up to three genres associated with the title
- director1 (string) - director 1(s) of the given title
- director2 (string) - director 2(s) of the given title
- IMDb score (float) - the average score in IMDb
- tmdb score (float) - average score in tmdb
- metascore - average score in metacritic
- metascore_normalized - normalized the metascore range to 1-10
- average_score - the total average score of the movie

A	B	C	D	E	F	G	H	I	J	K
imdb id	movie_title	Year	genres	director1	director2	IMDb score	tmdb score	metascore	metascore_normalized	average_score
tt0013274	Istoriya grazh	2021	Documentar	Nikolai Izvolc	Dziga Vertov	6.8	6	0	0	4.3
tt0015414	La tierra de lo	2000	\N	Musidora		5.4	0	0	0	1.8
tt0035423	Kate & Leopo	2001	Comedy,Fan	James Mangold		6.4	6.328	44	4.4	5.7
tt0062336	The Tango of i	2020	Drama	Ra,àò,à 'l Ru	Valeria Sarm	6.4	6.167	0	0	4.2
tt0299242	The Other Side	2012	Drama	George Miller		6.7	6.222	72	7.2	7.1

FAIR ACCESSMENT

Findability

- each movie dataset is assigned a unique identifier to locate the data
- datasets are described with rich metadata

Accessibility

- Data is retrievable using standardized protocols, API
- The protocols used are open, free, and universally implementable.

Interoperability

- Data uses formal accessible, shared and broadly used application, csv

Reusability

- Data is described with accurate and relevant attributes

QUALITY AND PERFORMANCE

Test plan:

- Functional tests
 - integration test: ensure all components work well
 - end-to-end test: simulate the scenario to make sure the overall workflow function correctly
- Performance test
 - load test: test the limit of accessing the system

Thank you!

ENJOY THE SHOW.