Big Data

(Netflix's Success Story)

Instructor: Thanh Binh Nguyen

September 1st, 2019

S³Lab
Smart Software System Laboratory


“Big data is at the foundation of all the megatrends that are happening today, from social to mobile to cloud to gaming.”
– Chris Lynch, Vertica Systems

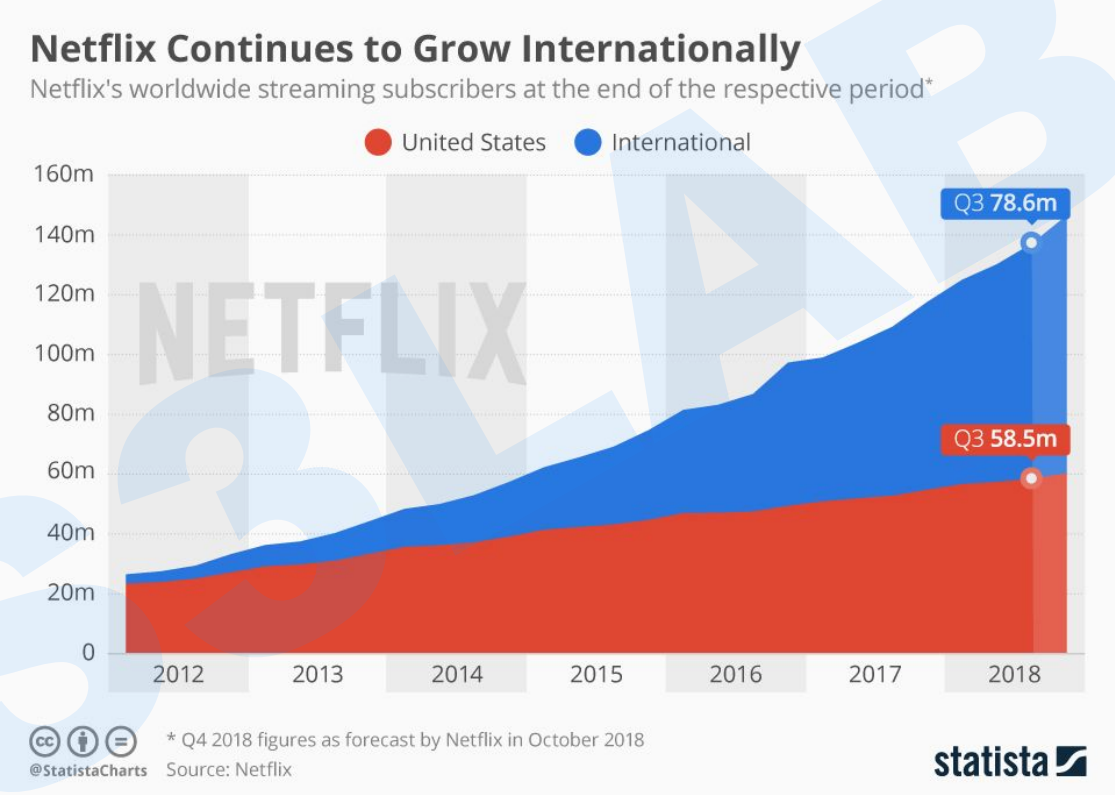
Netflix in numbers

- Netflix streaming service is said to account for ⅓ of peak time internet traffic in the US.
- Last year, Netflix announced that it signed on 100 million accounts worldwide.
- Netflix estimate that its algorithms save $1 billion a year in value from customer retention
- A typically Netflix member loses interest after perhaps 60 to 90 seconds of choosing something to watch, having reviewed 10 to 20 tiles.

Netflix in numbers

- Majority of Netflix users consider recommendations with 80% of Netflix views coming from the service's recommendations.
- Netflix has set up 1300 recommendation clusters based on users viewing preferences.
- Netflix segments its viewers into over 2K taste groups. Based on the taste group a viewer falls, it dictates the recommendations.
- With over 7K TV shows, movies in the catalogue, it is actually impossible for a viewer to find movies they like to watch on their own. Netflix's recommendation engine automates this search process for its users.

Netflix in numbers



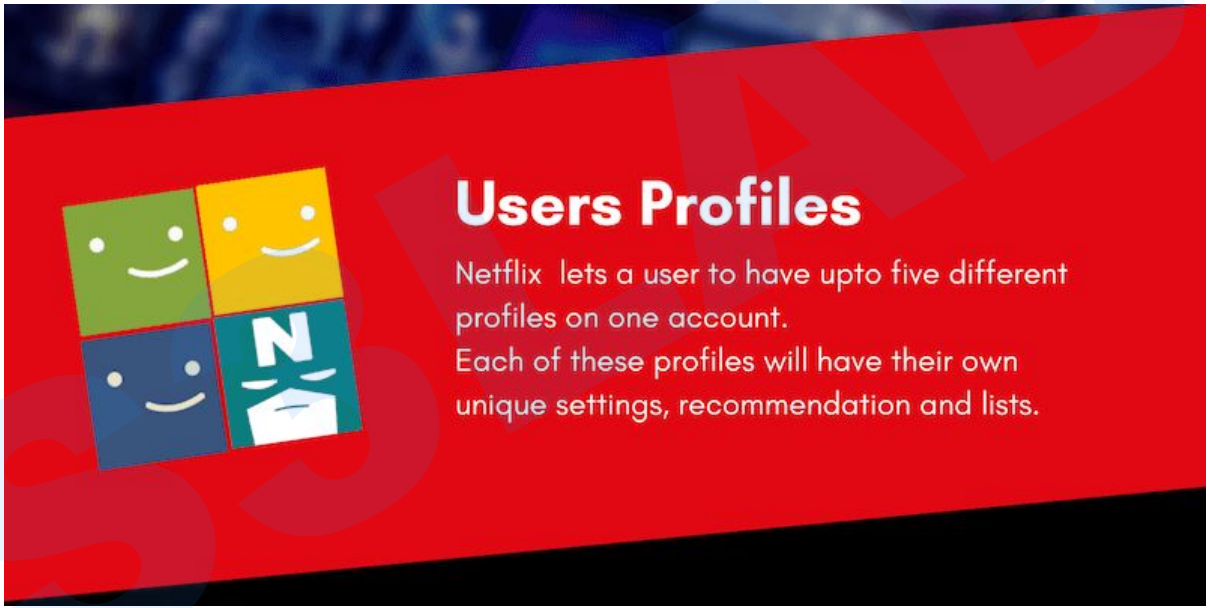
Key Factors Behind the Success

Unlimited Options

- Subscribers can find a wide range of movies, tv series, shows, performances, documentaries licensed from distribution partners with new options entering the offer on a regular basis. This makes Netflix's offer practically unlimited. I've never heard anyone say they've seen everything on Netflix.

Key Factors Behind the Success

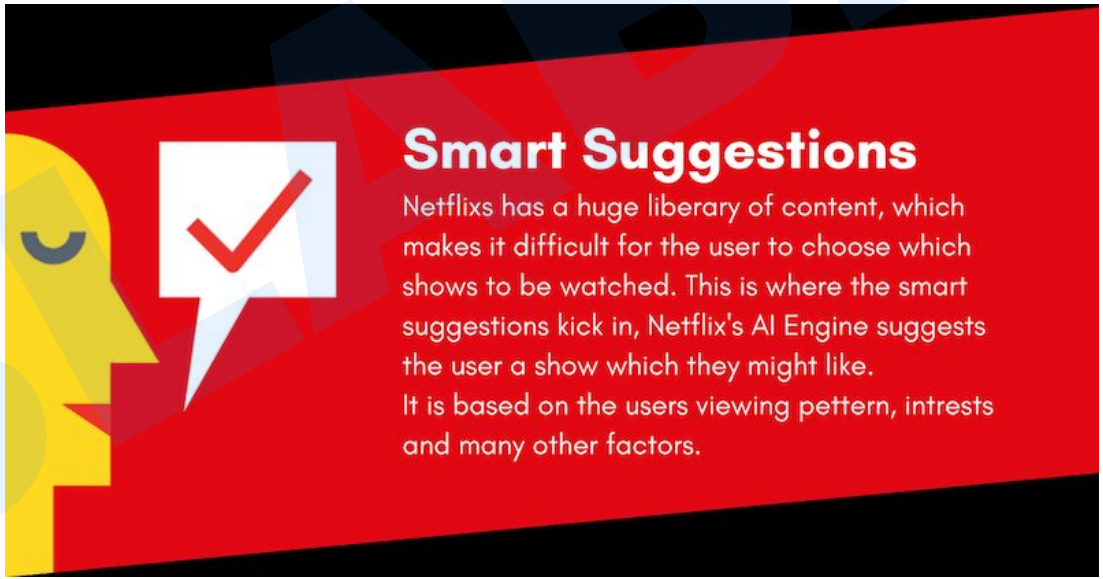
Flexible Configuration



Key Factors Behind the Success

Personalization via Recommendation engine

- Capture everything user Do on their platform.
- Support by Machine Learning and BigData



Netflix Researches: https://research.netflix.com/research-area/recommendations

Key Factors Behind the Success

AI

- Auto-Generation and Personalization of Thumbnails / Artwork.
- Location Scouting for Movie Production (Pre-Production).
- Movie Editing (Post-Production)
- Streaming Quality

Netflix Researches: https://research.netflix.com/research-area/recommendations

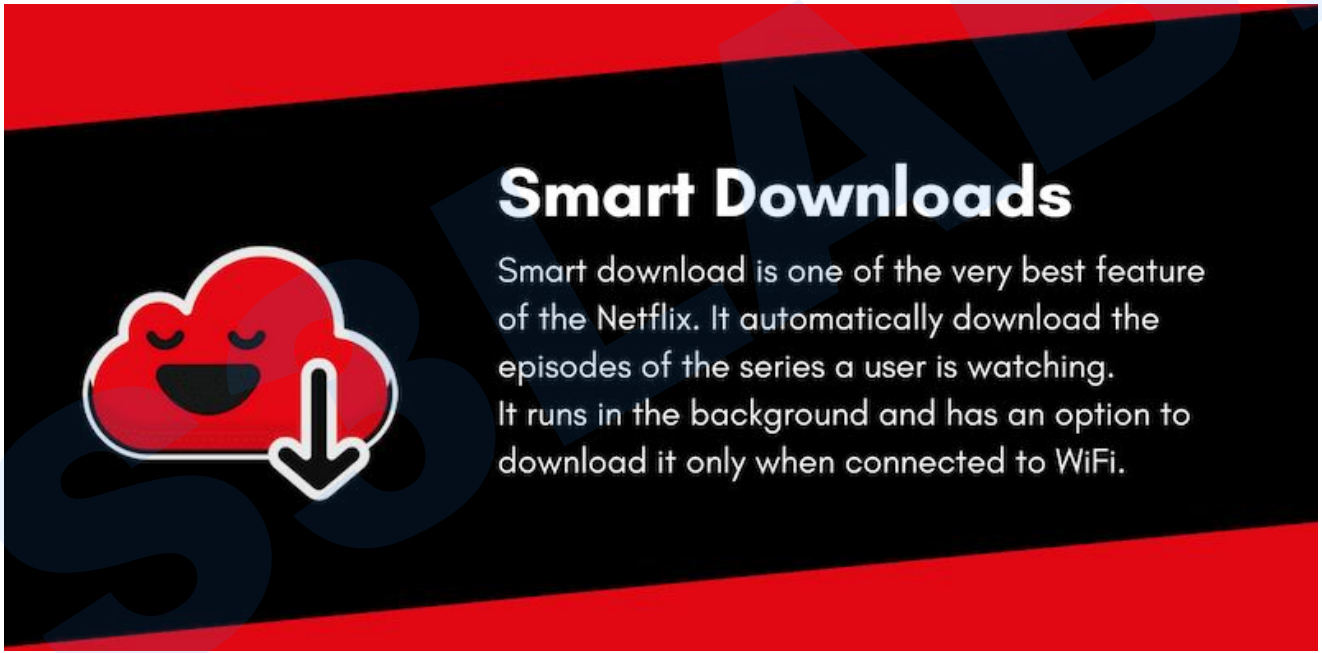
Key Factors Behind the Success

Investing in Original Content

- Spend $15 billion on content in 2019.
- About 85% of new spending is on originals which means over $12 billion

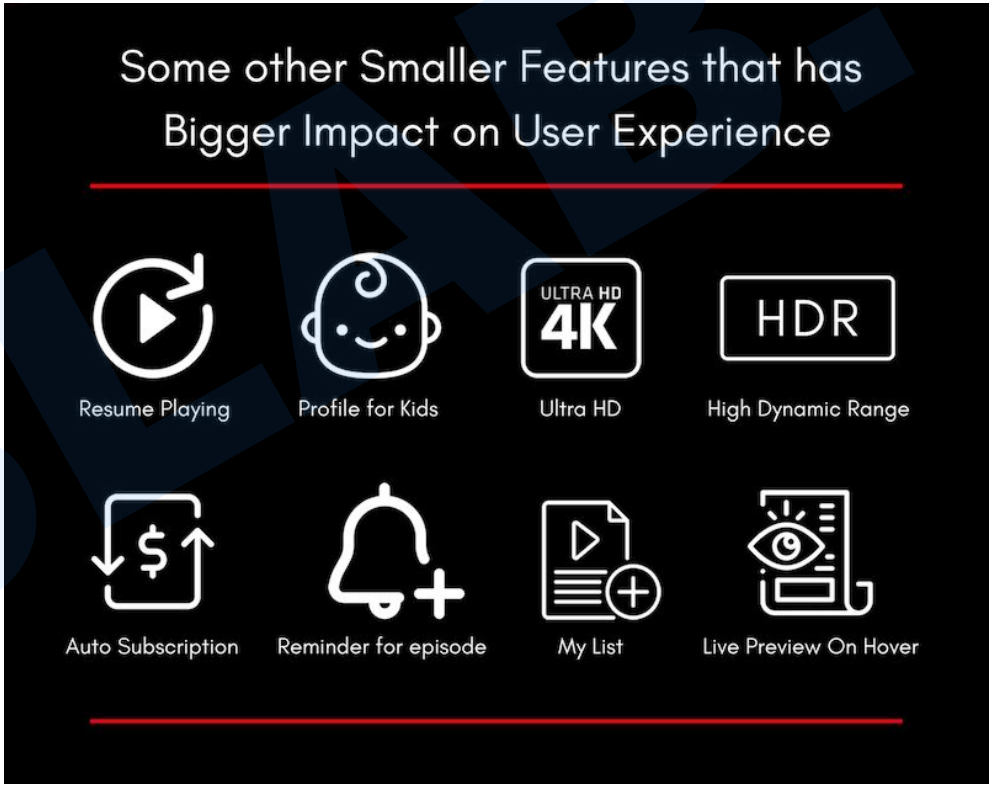
Key Factors Behind the Success

Excellent User Experience



Key Factors Behind the Success

Excellent User Experience



Key Factors Behind the Success

No AD

Key Factors Behind the Success

Binge-Watching(2016)

- Watching a TV series in like a week with multiple seasons, and getting little to no sleep.

Smart Suggestion | Recommendation System

Back in 1998

- Netflix was an online DVD rental, a DVD delivery service

Smart Suggestion | Recommendation System

The Netflix Prize

- Central to the effectiveness of its DVD delivery service was its recommendation engine
- Netflix offered $1 million to the person that could come up with the best algorithm for predicting how customers would rate a movie based on their previous ratings.
- See http://www.netflixprize.com/

Smart Suggestion | Recommendation System

With four data points

- At the time, Netflix was mainly a DVD delivery service so programmers only had access to four data points:
  - Customer ID
  - Movie ID
  - Rating
  - The date the movie was watched

Smart Suggestion | Recommendation System

Netflix predicts movie preferences

- 2009, the sophisticated winning approaches (see http://www.netflixprize.com/leaderboard.html) did not made it to production (100 million movie ratings for test - more than 5 billion ratings in real. The algorithms were **static**, meaning that they only deal with **historical data** and did not take into account the **dynamicity of users adding reviews in real-time**)
- A simple scalable (based on Big Data technology) approach made it

Smart Suggestion | Recommendation System

Nowadays, with more data points

- As streaming video became the primary focus, many more data points became available, giving insights into the customers
- Netflix specialists wade through and capture important information from a variety of different analytics streams:
  - Personalization analytics
  - Messaging analytics
  - Content delivery analytics
  - Device analytics
  - ...

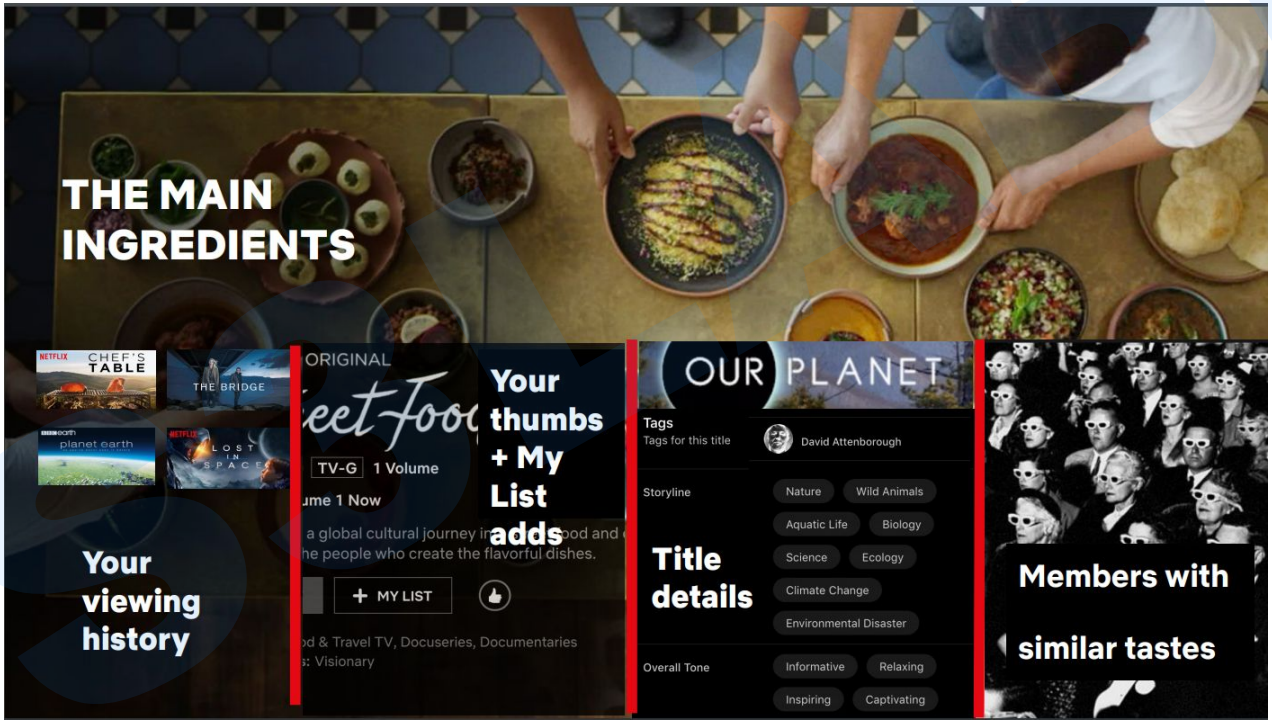
Smart Suggestion | Recommendation System

Nowadays, with more data points

- These data points included:
  - your interactions with our service (such as your viewing history and rating)
  - Information about the categories, year of release, title, genres, and more.
  - Other viewers with similar watching preferences and tastes.
  - Time duration of a viewer watching a show
  - The device on which a viewer is watching.
  - The time of the day a viewer watches -This is because Netflix has the data that there is different viewing behaviour based on the time of the day, the day of the week, the location, and the device on which a show or movie is viewed.
  - Time spent selecting movies
  - How often a movie or program was paused
  - Searches
  - ...

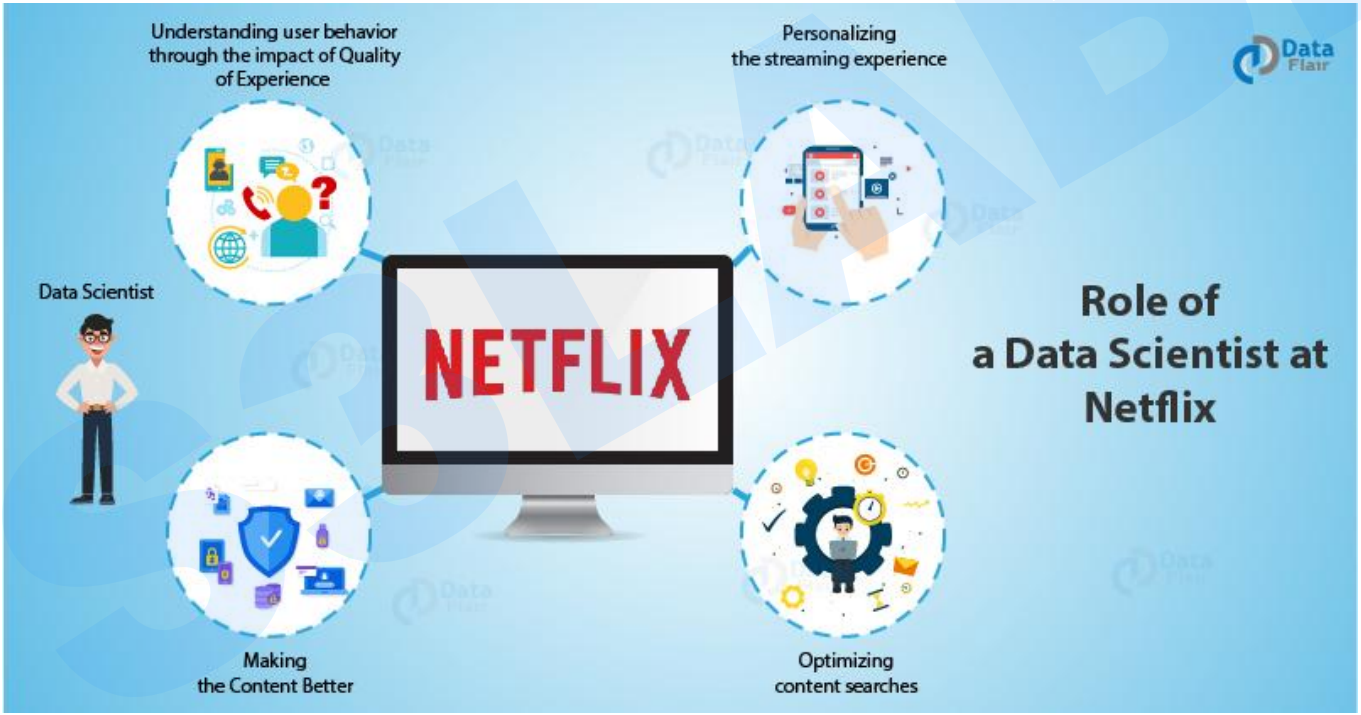
Smart Suggestion | Recommendation System

Nowadays, with more data points



Smart Suggestion | Recommendation System

Data Scientist at Netflix



Smart Suggestion | Recommendation System

Netflix predicts “perfect situation”

- With no ads, Netflix’s business model relies on customers who subscribe to their service for the long run. The happier the customers are, the longer they will stay subscribed to the service.
- Using these data points, Netflix developers built models to predict the “perfect situation” in which customers continuously receive programs they enjoy.
- Netflix uses **Ranking Algorithms** to provide a ranked list of movies and TV shows that appeal the most to its users. They implemented the **interleaving technique** to utility the various of ranking algorithms together.

Smart Suggestion | Recommendation System

Utility the various of ranking algorithms together

- Orders the entire Netflix collection for each member profile in a personalised way. The same genre row for each member has an entirely different selection of videos (**Personalised Video Ranker**).
- Netflix can now analyze how these factors impacted viewers' enjoyment (based on **ratings given to movies**).

Smart Suggestion | Recommendation System

Utility the various of ranking algorithms together

- Picks out the top personalized recommendations from the entire catalogue, focusing only on the titles that are top of ranking. (**Top N Video Ranker**)
- Videos trending with a mix of personalised for members based on viewing history and beacons (the capture of all events and user activities on Netflix) (**Trending now**)

Smart Suggestion | Recommendation System

Utility the various of ranking algorithms together

- Sorts your recently viewed titles and estimates whether the member will **continue watching** or **re-watch**, or whether perhaps they **stopped watching** something because it was less interesting than they envisaged (**Continue Watching**).
- Because you watched one video you may also like a similar video. Even though the similarity ranking is not personalised, it provides a good estimate of what a member might like based on what they previously watched (**Video-Video Similarity algorithm**)

Smart Suggestion | Recommendation System

Specific category suggestions

- Another method of data analytics is **movie tagging**. When a viewer rates a movie highly, the Netflix algorithm suggests other movies they may enjoy based on **matching tags**
- The service has created **76,897 unique ways to describe types of movies**.
- These are called “alt-genres,” which is what leads to Netflix’s scarily specific category suggestions (e.g.,“Comedies Featuring a Strong Female Lead”).

Smart Suggestion | Recommendation System

A/B Tests

- Netflix run 250 A/B tests each year. Around 100.000 people are randomly selected for each test, with another 100.000 used as a control group. Using two slightly different experiences (player look, mechanisms by which people find shows) to see how user respond.

Smart Suggestion | Recommendation System

Landing Cards

- Netflix creates multiple different landing cards - the images that are shown as people scroll through shows - for each of its titles. The idea is to find the most popular options.

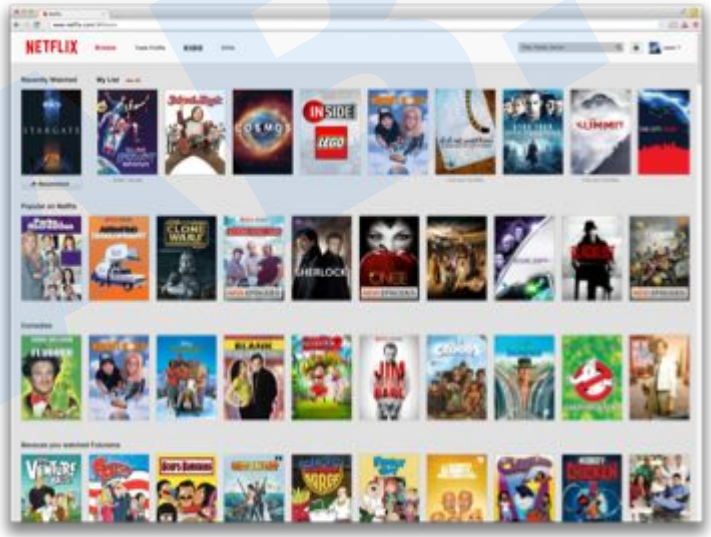
Smart Suggestion | Recommendation System

Timing

- Netflix may show you shorter programmes, or ones you're halfway through, when you login late at night and may not be looking to watch an entire show from scratch.

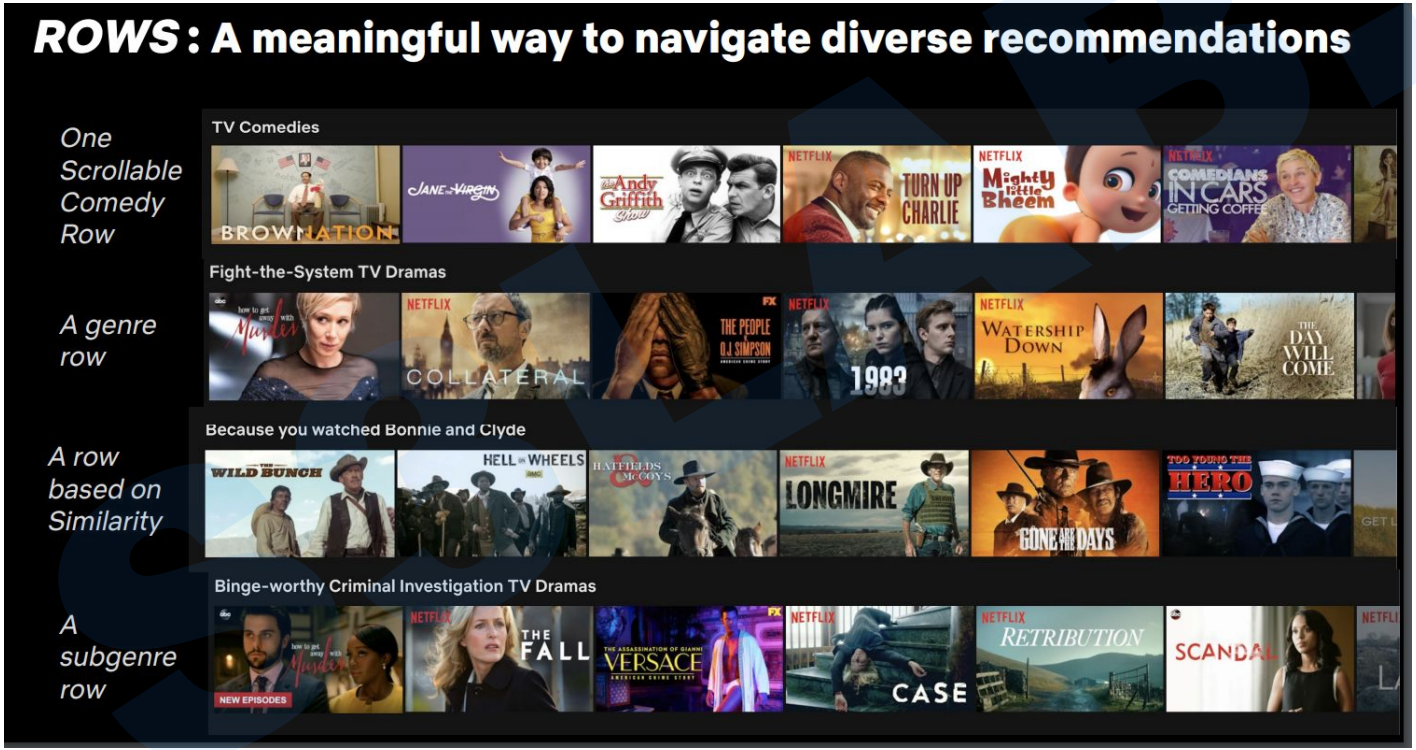
Personalized Homepage

- How to best tailor each member's homepage to make it relevant, cover their interests and intents, and still allow for exploration of our catalog.
- Two dimensional layout, member can scroll either horizontally on a row to see more videos in that row or vertically to see other rows.

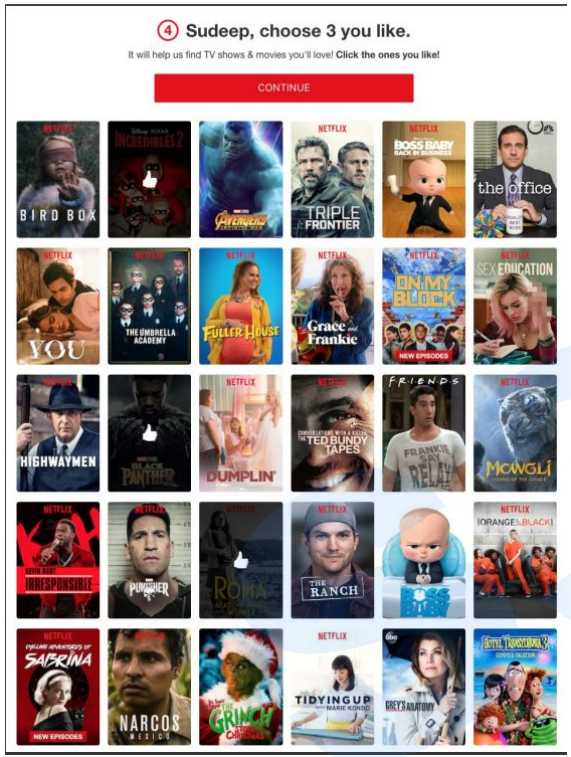


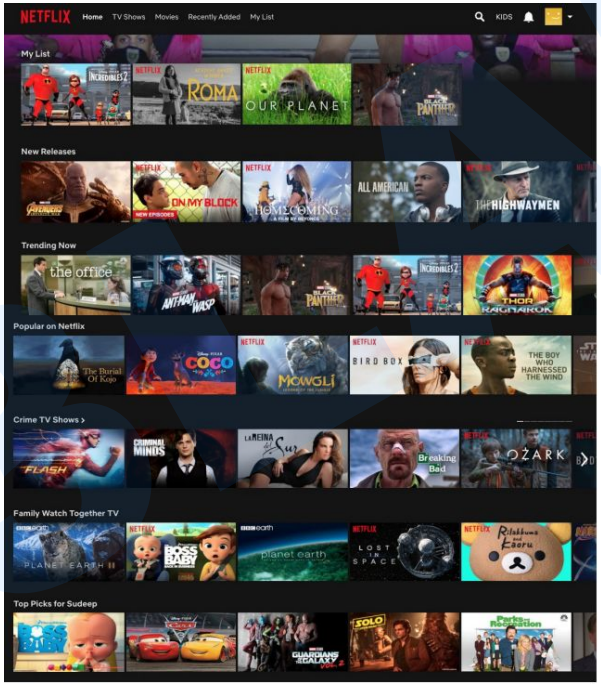
Personalized Homepage

Personalized Homepage



Personalized Homepage

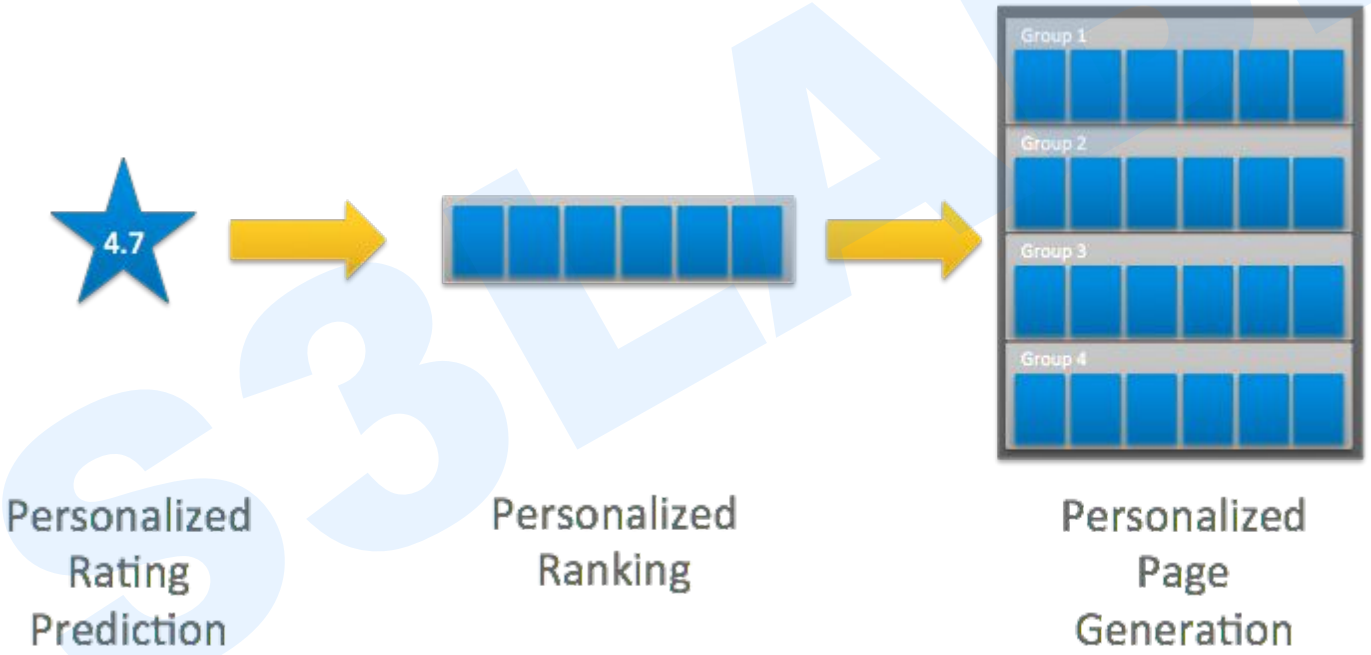




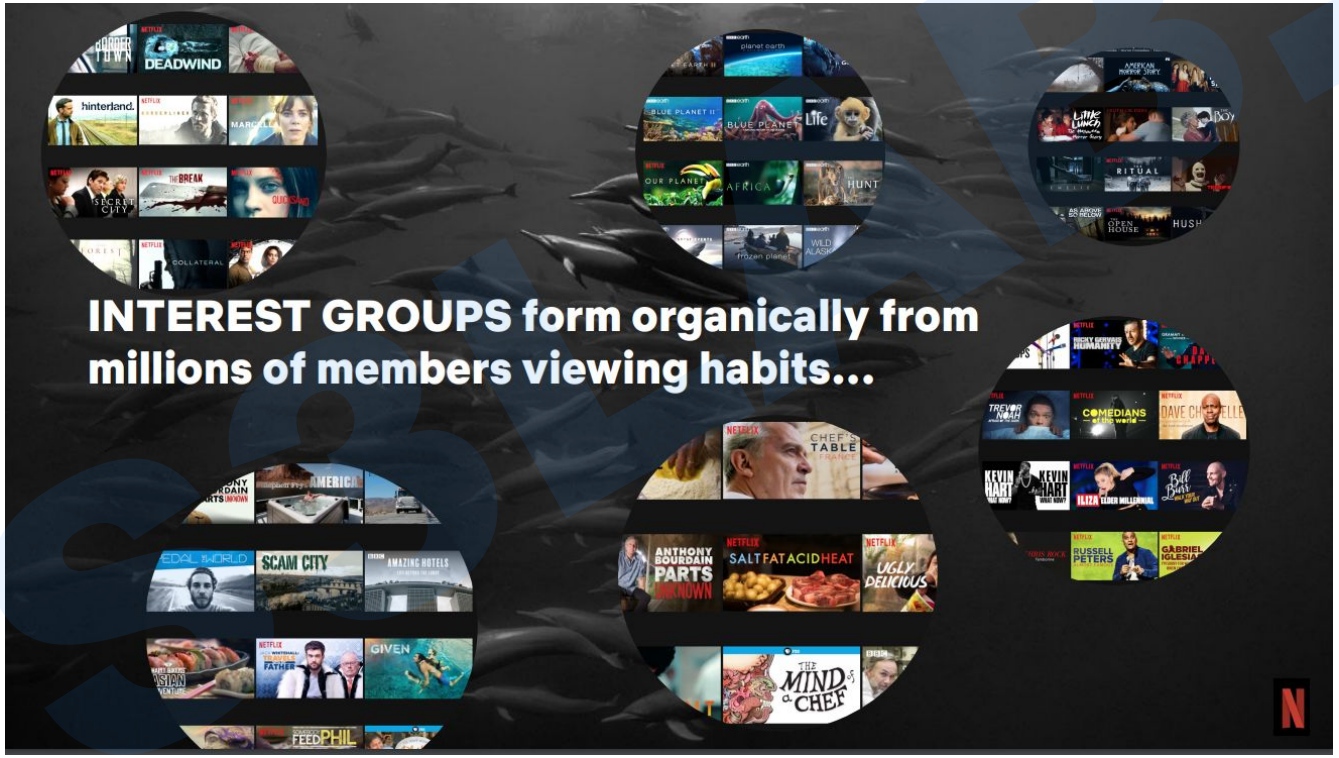
The algorithm's first guesses are mostly based on POPULARITY but gently nudged toward *YOUR* INITIAL CHOICES

Personalized Homepage

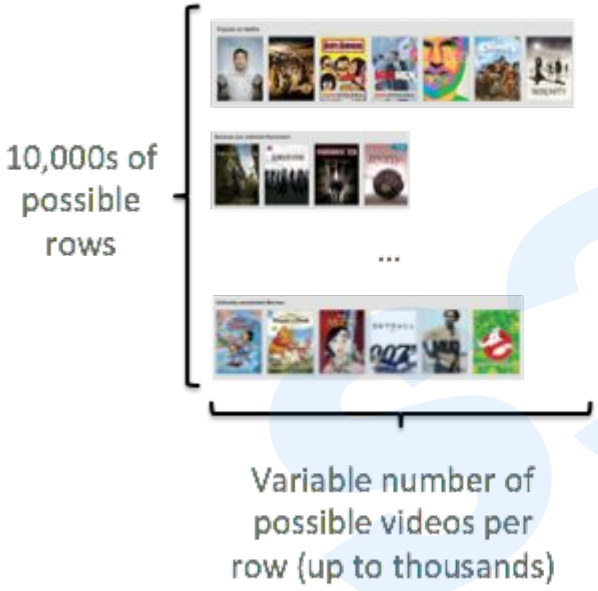
Evolution of Netflix's Personalization approach

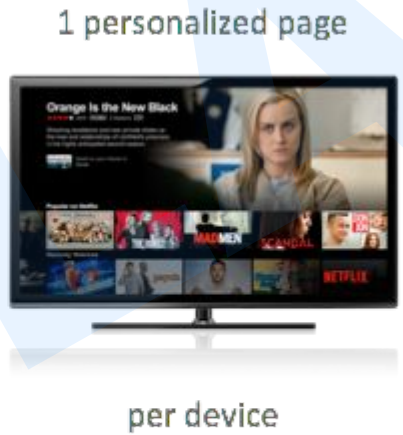


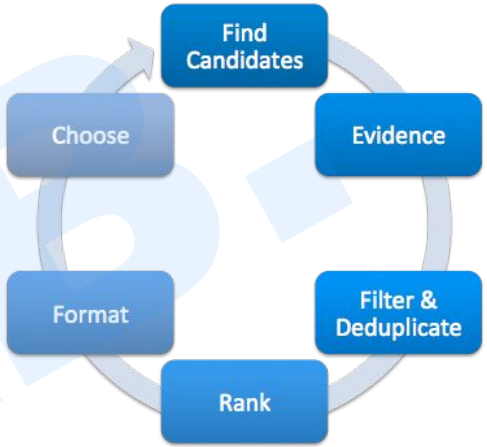
Personalized Homepage



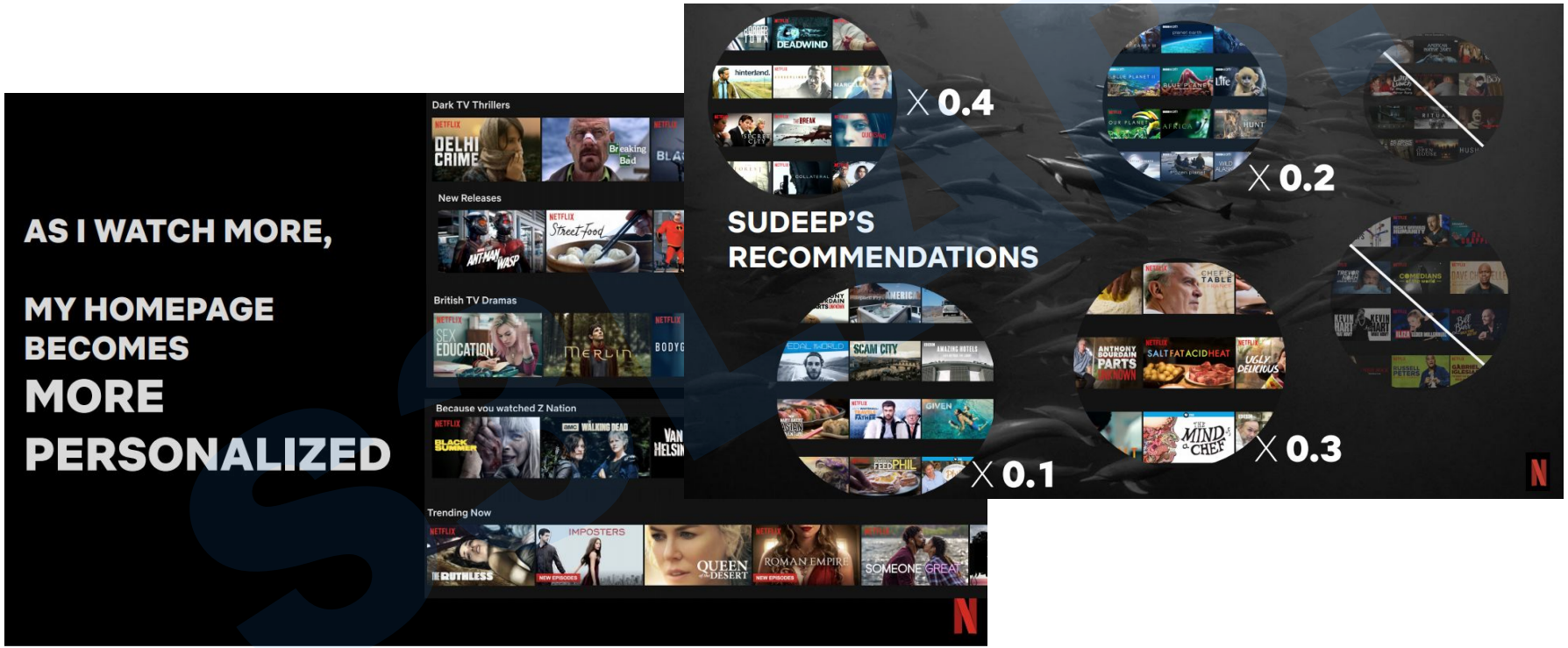
Personalized Homepage



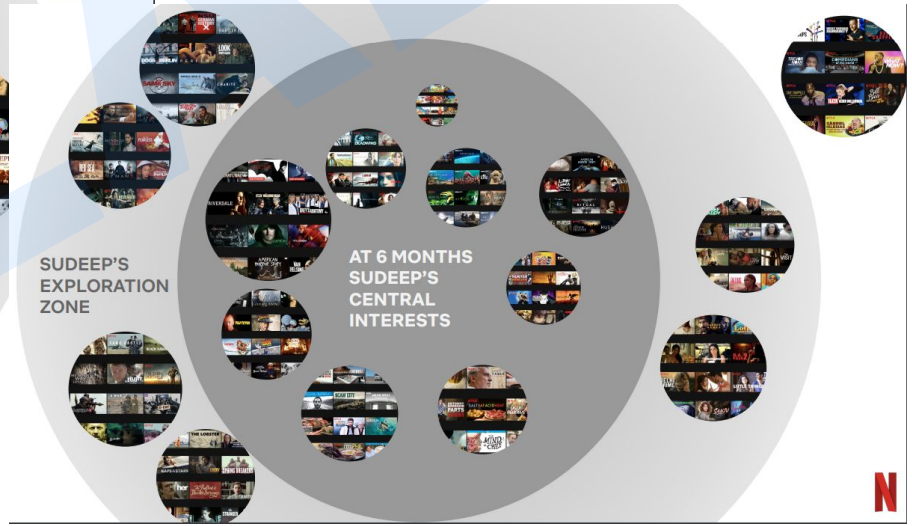




Personalized Homepage



Personalized Homepage



Bids on original programming

Using Big Data to finding the next Smash-hit series

- When the series **House of Cards** began shopping around for a home, **Netflix** aggressively jumped on it, outbidding major cable networks with a massive **two-season order**.
- Netflix spent $100 million on 26 episodes of House of Cards, as they were confident the show could be marketed successfully to their audience.

Bids on original programming

Why the enthusiasm

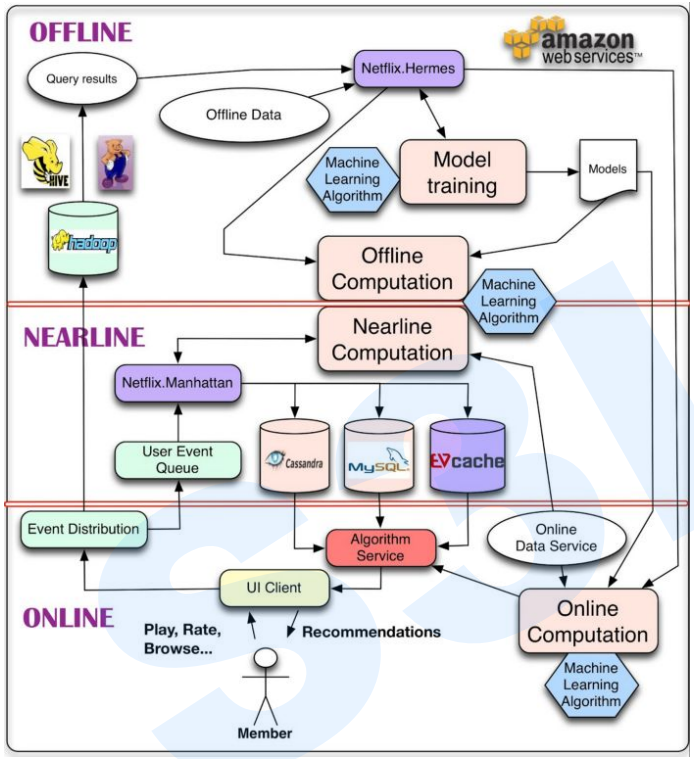
- **By analyzing its data**, Netflix saw that a large majority of its viewers enjoy programs :
  - Directed by David Fincher (who directed Se7en, Fight Club and The Social Network)
  - Starring Kevin Spacey.
- **House of Cards** was exactly that!

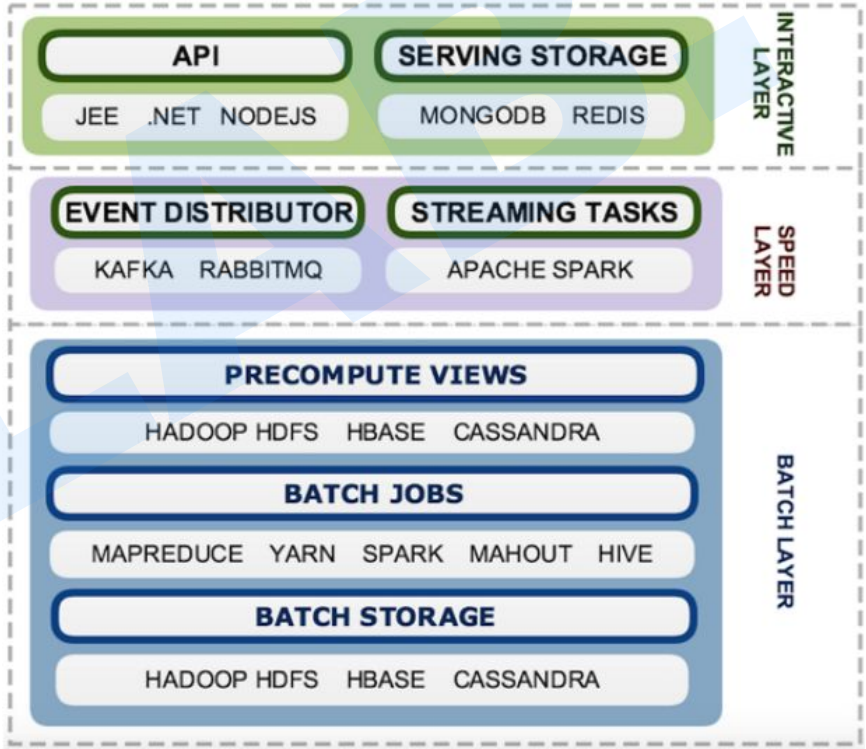
Bids on original programming

Promote programs

- Netflix took a data-driven approach to promote **House of Cards**.
- *For example*, the company modelled the show's cover image on the colours and styles for successful, similarly tagged programs, to help draw new viewers in

Netflix's System Architecture





Netflix's System Architecture

Data Pipeline Evolution

- ~500 billion events and ~1.3 PB per day
- ~8 million events and ~24 GB per second during peak hours
- There are several hundred event streams flowing through the pipeline. For example:
  - Video viewing activities
  - UI activities
  - Error logs
  - Performance events
  - Troubleshooting & diagnostic events

Conclusion

- Netflix disrupted the TV industry using data analytics to provide viewers with exactly the content they want.

Q & A

Cảm ơn đã theo dõi

Chúng tôi hy vọng cùng nhau đi đến thành công.