# Experiment 1

Aim : Case Study on any application(netflix) wrt it's big data architecture.

Case Study :

Netflix utilizes extensive viewer data, including demographics, viewing habits, and preferences, gathered over its years in the streaming industry. By employing sophisticated algorithms and mechanisms, the platform anticipates user preferences, suggesting content even before completion. This data-driven approach enhances service efficiency, guiding Netflix in delivering personalized and engaging content to its audience.

Some of these tools and features are:
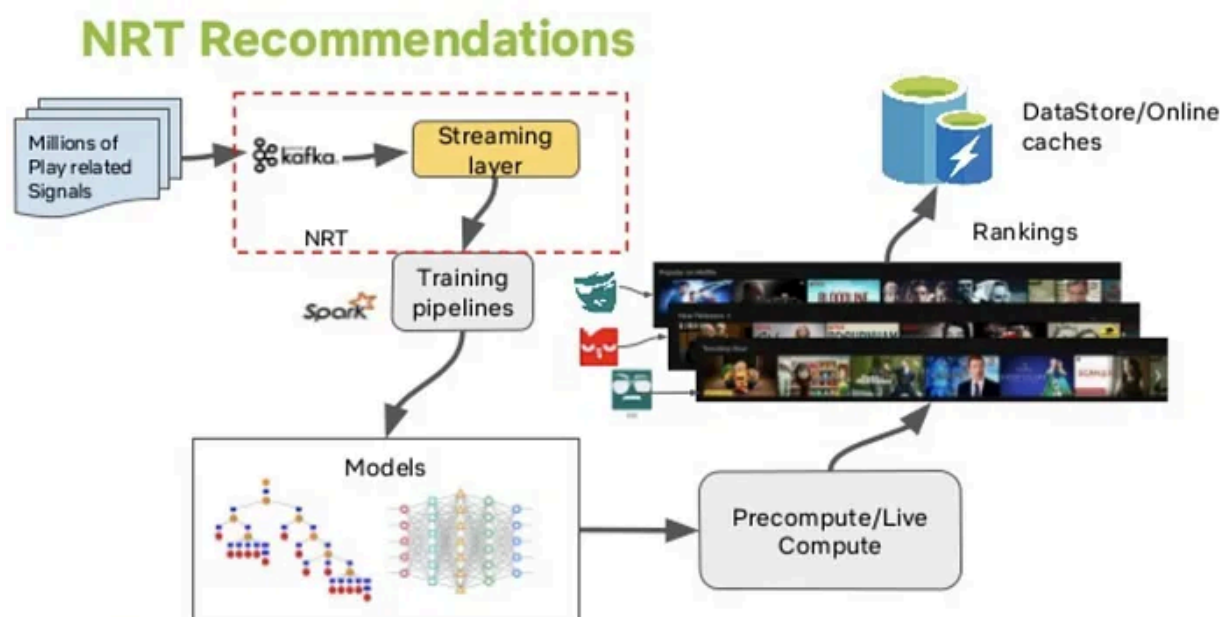
- **Real-Time Recommendation Engine**
  With a sea of users, each user generates hundreds of ratings per day based on what they watch, search and add to their watch-list, this data ultimately becomes a part of Big Data. Netflix stores all of this information and using key machine learning algorithms, it builds a pattern indicating the viewer's taste. This pattern may never match with another viewer because of how everyone's taste is unique.Based on the ratings, Netflix categorizes its media and suggests the viewer what the recommendation system thinks they might like to watch next.Netflix will know everything.

- **Artwork & Imagery Selection**
  The tool behind this is called AVA, which is essentially an algorithm that selects what artworks and images to show to whom. Short for Aesthetics Visual Analysis, AVA sifts through every video available and identifies the frames that are best suitable to be used as artworks.AVA takes a lot of metrics into consideration before finalizing on images, such as facial expressions of actors, the scene lighting, areas of interest, positioning of subjects on screen. It even categorizes and sorts artworks to show to users categorized into several taste groups.

- **Production Planning**
  Data plays an integral part when creators come up with an idea about a new show or movie. A lot of brainstorming takes place before anything gets on the paper, and that's where data comes in.With prior experience in creating new and original content and loads of data about how the viewers perceived the previous content, Big Data helps bring out the possible solutions to many of the challenges faced during the planning phase.These challenges could include identifying shoot locations, time and day of the shoot, and more. Even with simple prediction models, Netflix can save a significant amount of effort put into planning, further reducing expenses.

- **Metaflow**
  Netflix has open-sourced Metaflow, their cloud native, human-centric framework aimed at boosting data scientist productivityThe idea behind Metaflow was to shift the focus of data scientists from worrying about the infrastructure of models to solving problems. Metaflow allowed them the freedom to experiment with their ideas by offering a set of fine-tuned features that almost makes Metaflow feel like a plug-and-play framework.

- **Polynote**
  Developed and open-sourced by Netflix, Polynote is a polyglot notebook with support for Scala and various other features. Polynote allows smooth integration of JVM based machine learning platform with Python to data scientists and machine learning

researchers.
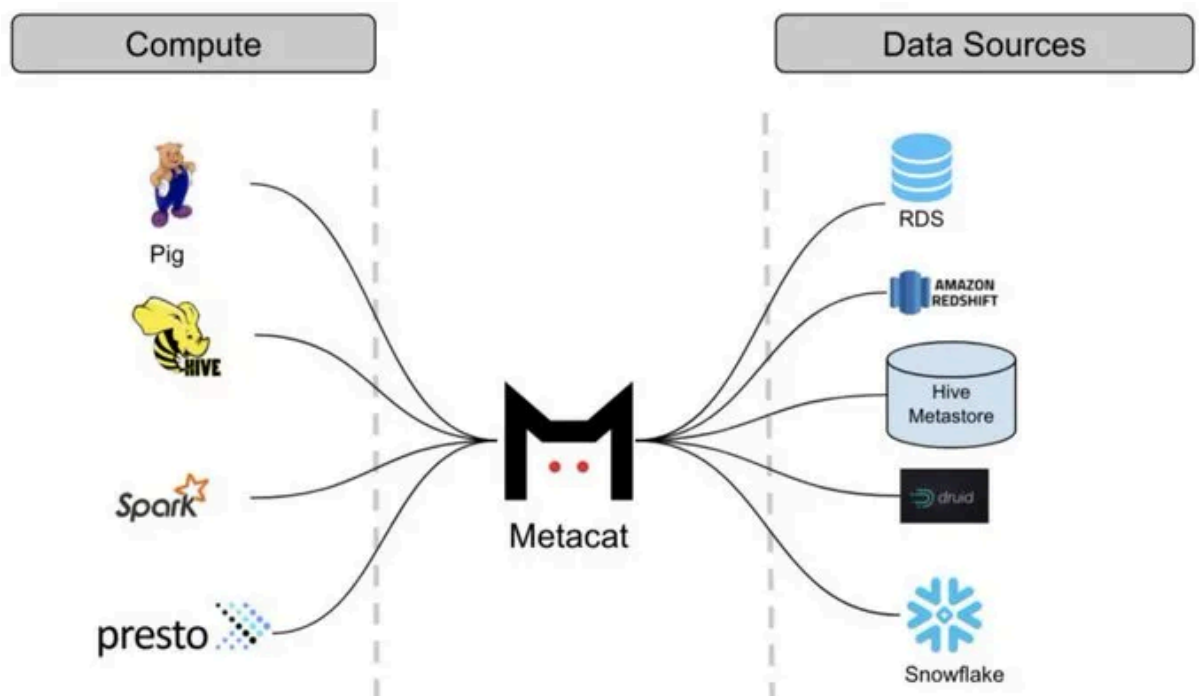
- **Metacat**
  The vast pool of data that Netflix operates on is spread across multiple platforms such as Amazon S3, Druid, Redshift and MySQL, to name a few. To maintain seamless interoperability among these data stores, Netflix needed a service.This need for simplicity gave birth to Metacat, whose sole purpose was to provide centralized metadata access for all data stores

- **Druid**
  Netflix uses Apache Druid for ensuring that its users get a high-quality user experience every time. Delivering a top-notch user experience every time is not a simple feat. It requires constant analysis of several events, gathering the necessary data and analyzing it. This data could be anything from the playback information, to device information, to measuring platform performance and several others. All these event metrics make raw data complicated, and that's where Druid comes into play.Druid's task is to provide real-time analytics on databases where queries execute regularly and at uncertain time-periods. It is highly scalable and offers excellent performance for any given workload.

Conclusion : Using big data, Netflix saves $1 billion per year on customer retention. Big Data plays a critical role in not just deciding the functioning of Netflix but also presents them with newer opportunities to grow. New technologies often bring their fair share of issues with them, but at Netflix, they have been tackling those issues head-on, consistently by taking community inputs. By open-sourcing several of the libraries and frameworks to the community, Netflix aims to improve not just itself, but other companies as well.