Data on the internet grows by 50 percent annually. More than 90% of the data has been generated in recent years. This is the time for big data. How can we effectively transfer this huge amount of data?

The researchers will investigate caching techniques used by online video platforms. Among them, YouTube is a leading online video provider worldwide. Before 2012, video streaming in YouTube was done using Real Time Messaging Protocol (RTMP)-based servers. This requires a streaming server and a near-continuous connection between the server and user. Requiring a streaming server can increase implementation cost, while RTMP-based packets can be blocked by firewalls. In 2012, this was replaced by HTTP (Hypertext Transfer Protocol) based servers known as MPEG DASH (Dynamic Adaptive Streaming over HTTP). By using this technology, the servers were able to use HTTP-Caching. This latter capability decreased total bandwidth costs associated with delivering the video, since more data could be served from web-based caches rather than the origin server, and improved quality of service, since cached data is generally closer to the viewer and more easily retrievable.

The essay will discuss different kinds of caching techniques, optimizations, data analysis, prediction techniques used by YouTube, including their advantages/disadvantages and potential social impacts.