<https://blog.pingpong.us/reformer-review/>

Reformer: The Efficient Transformer(ICLR 2020)

데이터 토큰이 조금만 커져도 모델 크기가 엄청나게 커지는 기존 Transformer

이는 Attention 모델이 데이터 열에서 모든 데이터 포인트 쌍 간의 가중치를 나타내기 때문(??)

기존 Transformer는 번역을 위해, 아니면 언어 이해를 위한 구조로 입출력이 그렇게 크지 않아서 모델이 커지지 않았다. 따라서 문제가 없었다.(512 토큰 정도)

하지만 여러 Task에 적용되고 있고, 입출력이 문서 단위, 이미지 단위로 들어간다면 문제가 된다.