Transformer

Before Transformer, RNN (Recurrent Neural Network) was mainly used to process text sequence data. When RNN takes the sequence as input, it processes the sentence word by word. It adopts a sequential processing and cannot be executed in parallel. Additionally, when such sequences are too long, models tend to forget content at distant locations in the sequence or mix it with content at subsequent locations.

Here's a simple ASCII representation:

No.	Fruit	Price ($)
1	Apple	1.5
2	Banana	3.5
3	Pear	2.7
4	Orange	3

Transformer proposes a new approach. It proposes to encode each position and apply an attention mechanism to associate two distant words, which can then be parallelized, thus speeding up training.