Why is Kafka fast?  
.  
.  
Kafka achieves low latency message delivery through Sequential I/O and Zero Copy Principle. The same techniques are commonly used in many other messaging/streaming platforms.  
  
The diagram below illustrates how the data is transmitted between producer and consumer, and what zero-copy means.  
  
🔹Step 1.1 - 1.3: Producer writes data to the disk  
  
🔹Step 2: Consumer reads data without zero-copy  
2.1: The data is loaded from disk to OS cache  
2.2 The data is copied from OS cache to Kafka application  
2.3 Kafka application copies the data into the socket buffer   
2.4 The data is copied from socket buffer to network card  
2.5 The network card sends data out to the consumer  
  
🔹Step 3: Consumer reads data with zero-copy  
3.1: The data is loaded from disk to OS cache  
3.2 OS cache directly copies the data to the network card via sendfile() command  
3.3 The network card sends data out to the consumer  
  
–  
Subscribe to our weekly newsletter to get a Free System Design PDF (158 pages): <https://bit.ly/42Ex9oZ>  
  
[#systemdesign](https://www.linkedin.com/feed/hashtag/?keywords=systemdesign&highlightedUpdateUrns=urn%3Ali%3Aactivity%3A7080426219938164736) [#coding](https://www.linkedin.com/feed/hashtag/?keywords=coding&highlightedUpdateUrns=urn%3Ali%3Aactivity%3A7080426219938164736) [#interviewtips](https://www.linkedin.com/feed/hashtag/?keywords=interviewtips&highlightedUpdateUrns=urn%3Ali%3Aactivity%3A7080426219938164736)  
.

Activate to view larger image,

