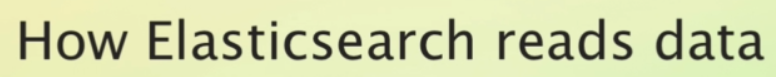
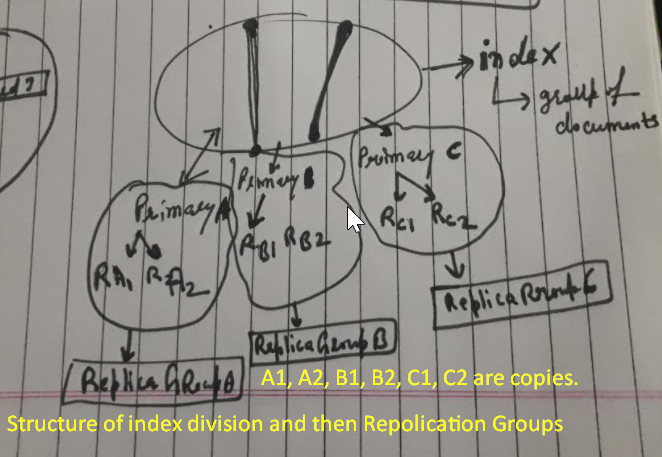
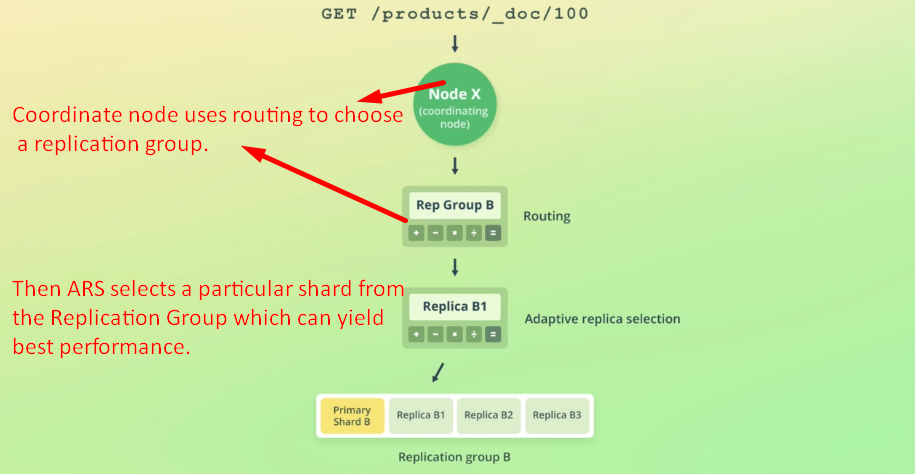
1.   
   First understand the following diagram then read further to understand properly.  
   
2. We saw what routing is.
3. Let’s talk about how Elasticsearch reads data. Routing plays a key role in it but the there is more to it than that.
4. **Agenda**: I want to mention that what we will cover in this lecture, is related to reading a single document and not how search queries work; we will cover that later in this course when the time is right.
5. How reading works?
   1. The very first thing that happens is that a given node (coordinate) receives the read request.
   2. This node is responsible for coordinating the request, and is therefore referred to as **coordinating node.  
      So what does this coordination involve?**
   3. The first step is to figure out where the document we’re looking for is stored and we know that is done with **routing**.  
      And we know that routing resolves to a shard that stores the document but more specifically it resolves to the primary shard (or replication group).  
      That’s of course assuming that the shard has been replicated, but that is almost always the case, for reasons that we discussed earlier.
   4. If Elasticsearch just retrieves the document directly from the primary shard, then all retrievals would end up on the same shard, which of course doesn’t scale well.  
      
   5. Instead, a shard is chosen from the replication group. Elasticsearch uses a technique called **Adaptive Replica Selection** (ARS) for this purpose.
   6. What **ARS** essentially does, is to select the shard copy that is deemed to be the best. This is evaluated by a formula that takes a number of factors into account, but I will not get into that right now. We will talk about ARS in more detail a bit later in the course.  
      For now just know that Elasticsearch tries to select the shard copy that it believes can yield the best performance. Once a shard has been selected, the coordinating node sends the read request to that shard. When the shard responds, the coordinating node collects the response and sends it to the client.
   7. The client will typically be an app using one of the Elasticsearch SDKs, but it could also be Kibana or your development machine if you send requests from the command line.  
      
6. Now that you know how reading data works at a high level, let’s move on to talking about how writing data works.