**Apache Spark Interview Preparation**

To be read :

<https://luminousmen.com/post/dive-into-spark-memory>

<https://www.linkedin.com/pulse/apache-spark-memory-management-deep-dive-deepak-rajak/>

1. Apache Spark Memory Management
2. ResourceManager handles memory requests and allocates executor container up to maximum allocation size settled by yarn.scheduler.maximum-allocation-mb configuration. Memory requests higher than the specified value will not take effect.
3. When Does Spark gets Out of memory issues

* Driver OOM :

1. Collect Operation
2. Broadcast Join:

While broadcasting , the data which is getting broadcasted is collected from data nodes to driver memory. Then only it will be broadcasted to all the worker nodes.

This can be controlled by spark.sql.autoBroadcastJoinThreshold

* Executor OOM: