

1. Prepare the table with the statistics (total articles, GBs) and datasets before and after ***deduplication***. [10 pts]

Before De-duplication		Before De-duplication	
Total articles		Total articles	
Volume in GBs		Volume in GBs	