

×

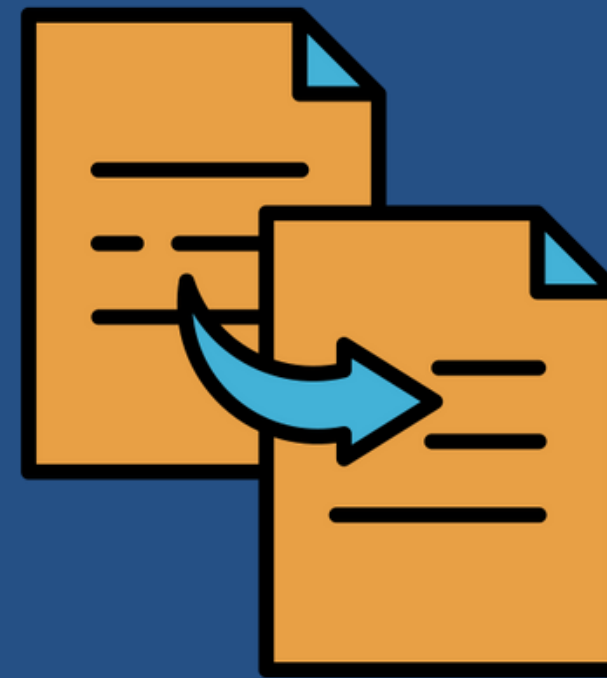
KEY

**VOCABULARY FOR
THE READING
ACTIVITY #2**



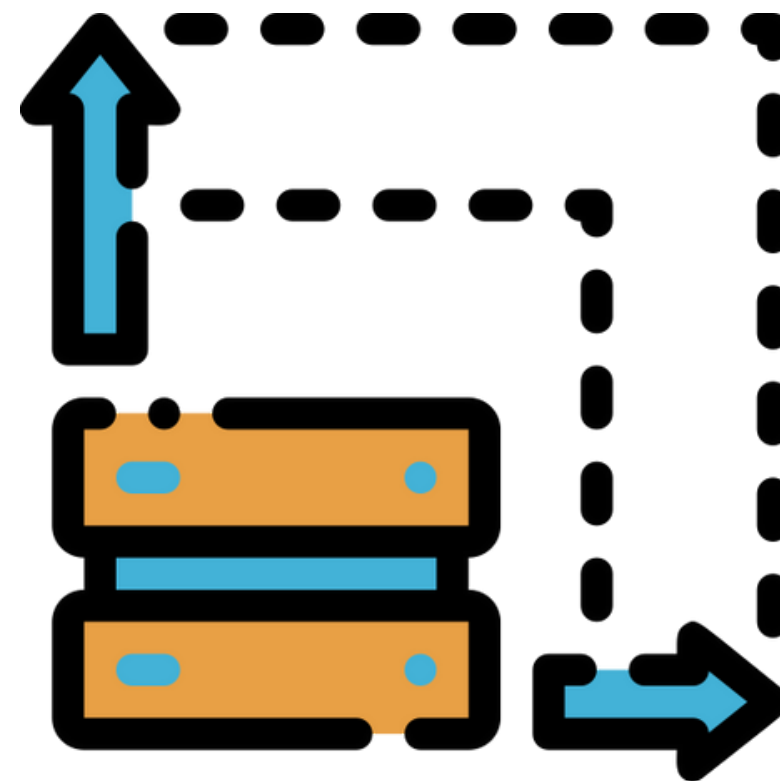
DATA PARTITIONING:

A TECHNIQUE INVOLVING THE DIVISION OF A LARGE DATASET INTO SMALLER, MORE MANAGEABLE SUBSETS OR PARTITIONS BASED ON SPECIFIC CRITERIA SUCH AS VALUES, GEOGRAPHICAL LOCATION, OR OTHER RELEVANT ATTRIBUTES.



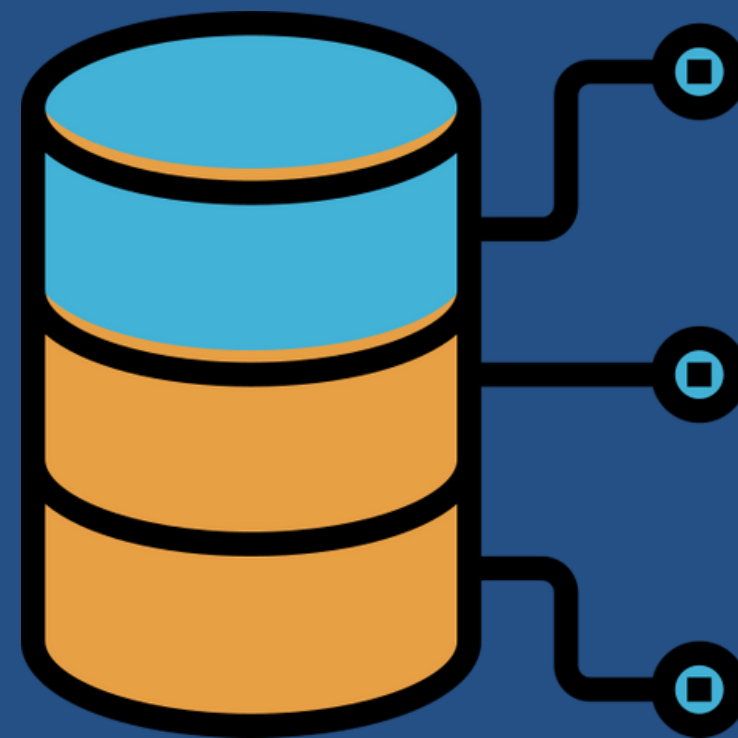
SCALABILITY:

THE ABILITY OF A SYSTEM OR PROCESS TO HANDLE A GROWING AMOUNT OF DATA, WORK, OR USERS, ENSURING THAT PERFORMANCE IS MAINTAINED OR IMPROVED AS RESOURCES INCREASE.



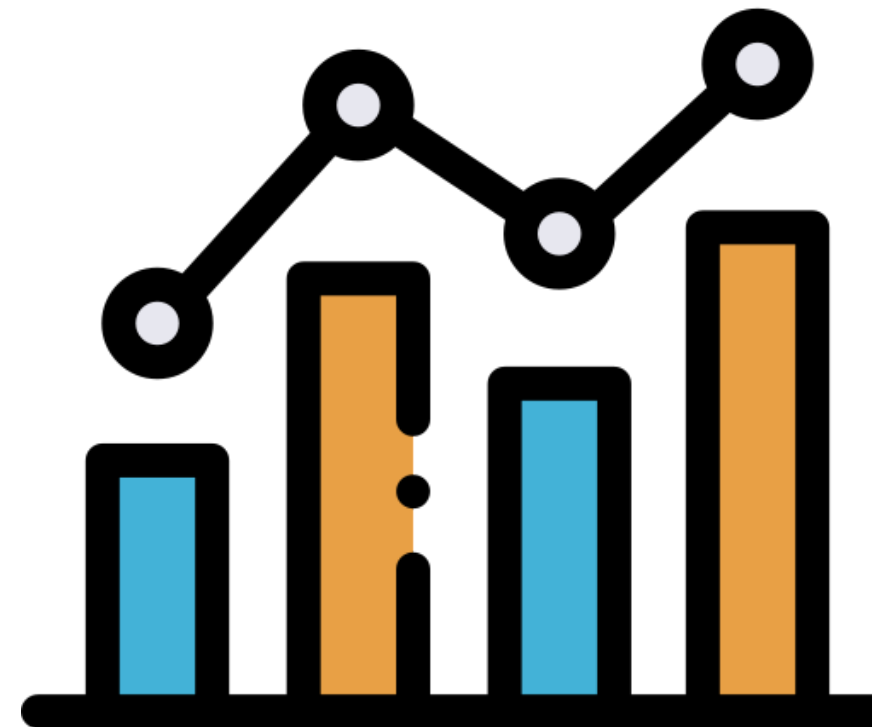
QUERY PERFORMANCE:

THE EFFICIENCY WITH WHICH A DATABASE OR SYSTEM RETRIEVES AND PROCESSES DATA IN RESPONSE TO USER QUERIES, OFTEN INFLUENCED BY FACTORS LIKE INDEXING, PARTITIONING, AND DATA DISTRIBUTION.



DATA SKEW:

UNEVEN DISTRIBUTION OF DATA ACROSS PARTITIONS, LEADING TO IMBALANCES IN WORKLOAD AND POTENTIAL PERFORMANCE ISSUES, WHICH NEEDS TO BE ADDRESSED FOR OPTIMAL SYSTEM FUNCTIONING.





RESILIENT DISTRIBUTED DATASETS (RDD):

**A FUNDAMENTAL DATA STRUCTURE IN
APACHE SPARK, PROVIDING FAULT-
TOLERANT, PARALLEL PROCESSING OF DATA
ACROSS A DISTRIBUTED COMPUTING
CLUSTER, ENABLING EFFICIENT HANDLING OF
PARTITIONED DATA.**

