Topic: Projection Operator Implementation

Assigned reading: Section 14.3 from Raghu Ramakrishnan's book

Projection operator selects a set of attributes from the given relation. Note that projection operator can create duplicates. By default, SQL implementations retain duplicates. However, they also provide constructs to remove duplicates.

Without duplicate elimination, projection is easy to implement. Single scan of the datafile is enough to retrieve all records with required attributes.

For duplicate elimination, there are two possible approaches: Sorting and Hashing. Sorting puts duplicates one after the other. With a single scan of the qualifying records, duplicates can be removed after sorting. Hashing on the other hand only partitions records. For each record, its duplicates are located in its bucket only. Each bucket can be sorted independently to eliminate duplicates. If the size of a bucket is too large, then we can go for more rounds of hashing with new hashing function to split the bucket into smaller groups. However, we still need sorting per bucket at the end.