External Sort

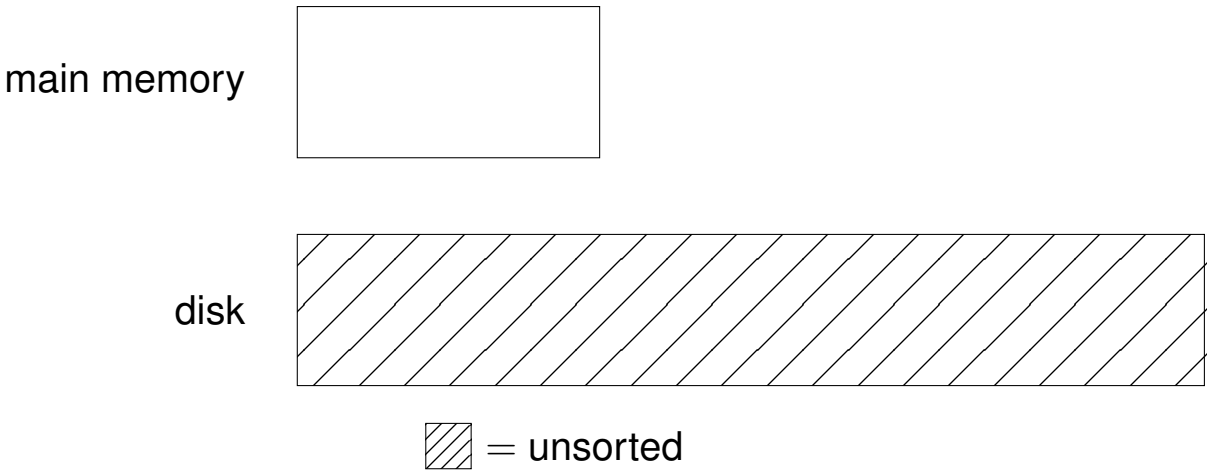
Problem:

- want to sort an arbitrary amount of data stored on disk
- accessing disk is slow so we do not want to access each value individually
- sorting in main memory is fast but main memory size is limited

Solution:

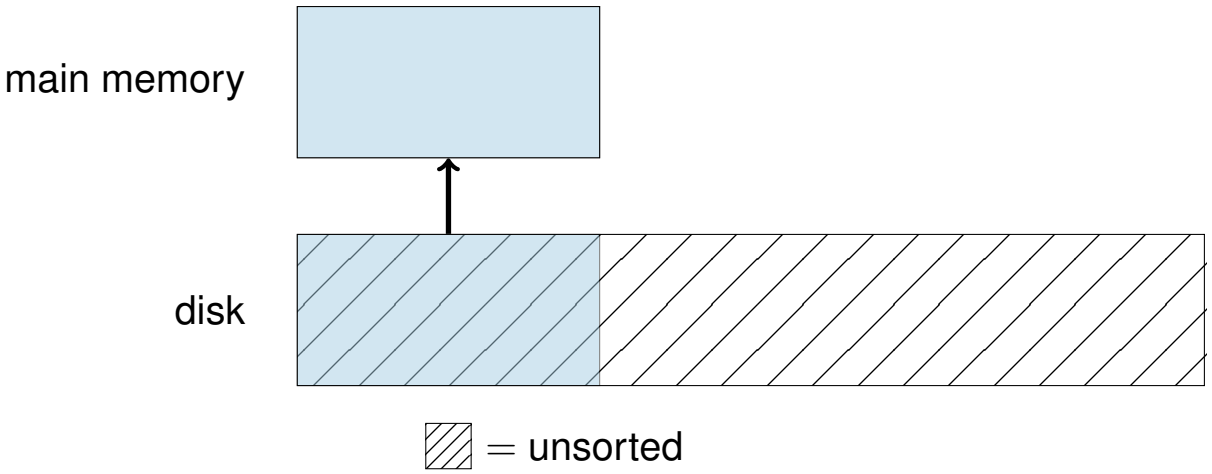
- load pieces (called “runs”) of the data into main memory and sort them
- with m values fitting into main memory and d values that should be sorted this results in $k = \lceil \frac{d}{m} \rceil$ sorted runs
- do a k-way merge of all runs

External Sort



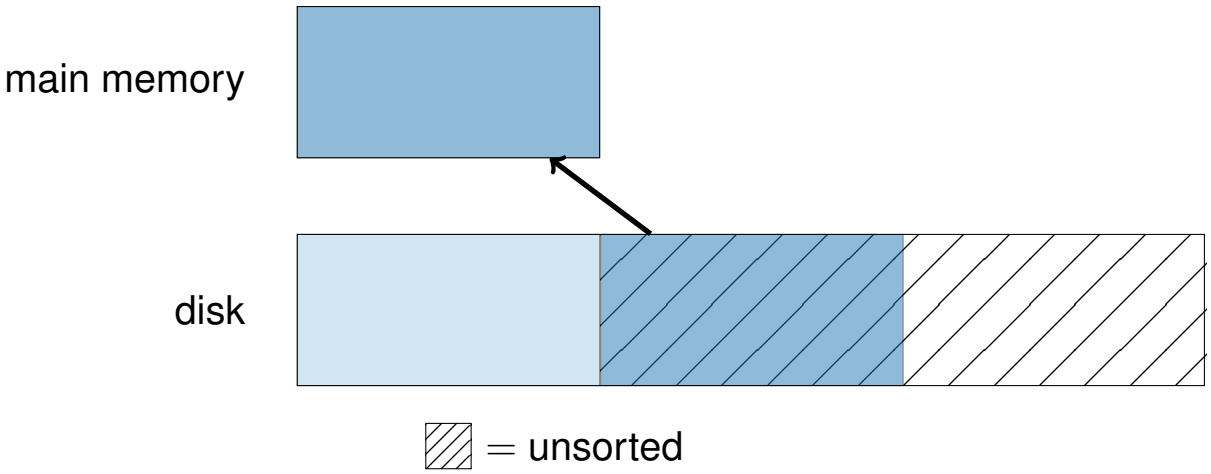
Step: sort k runs

External Sort



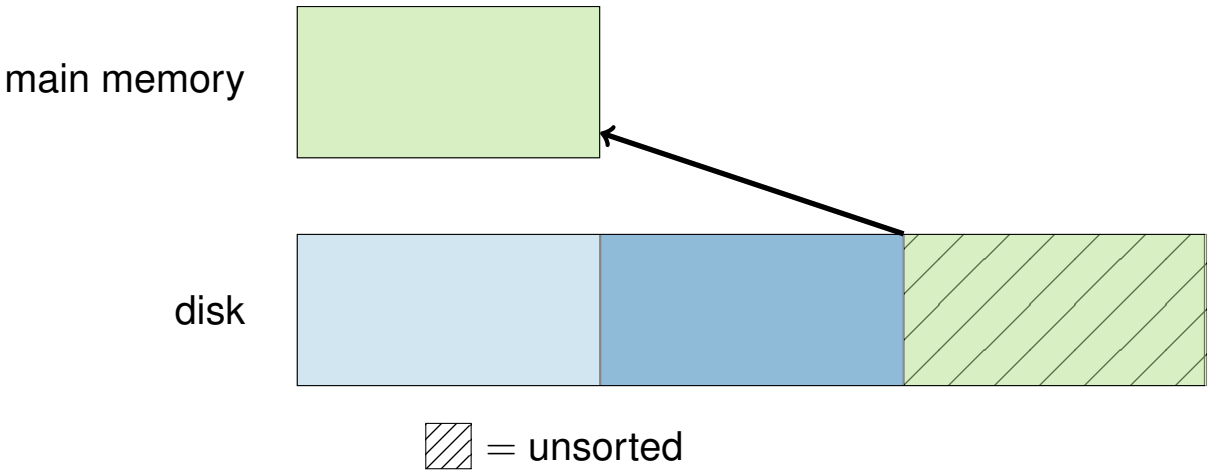
Step: sort k runs

External Sort



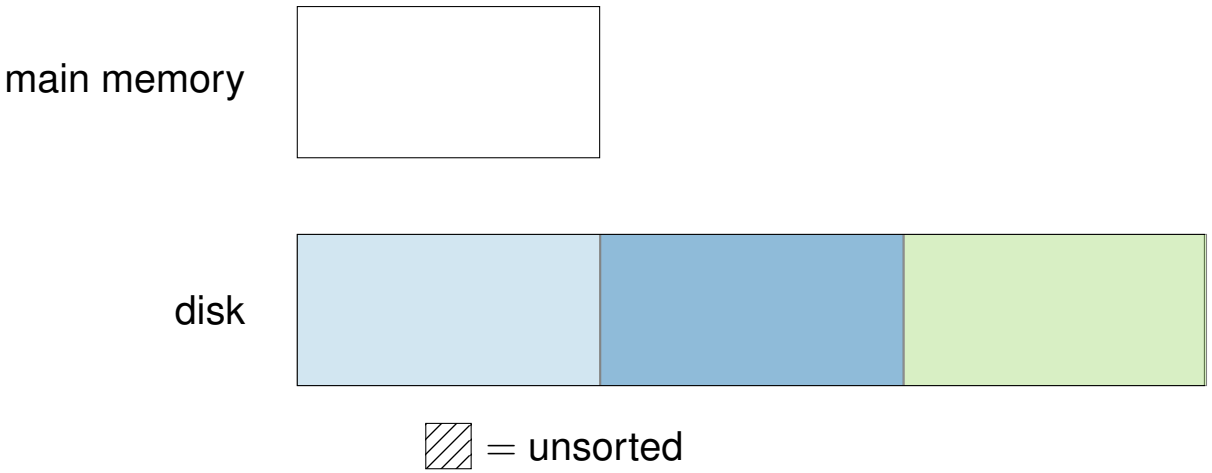
Step: sort k runs

External Sort



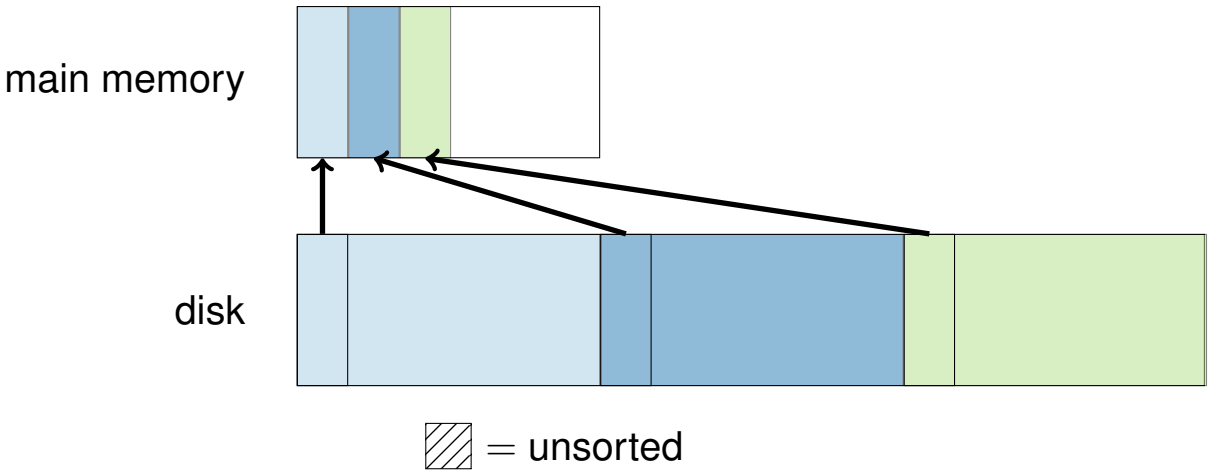
Step: sort k runs

External Sort



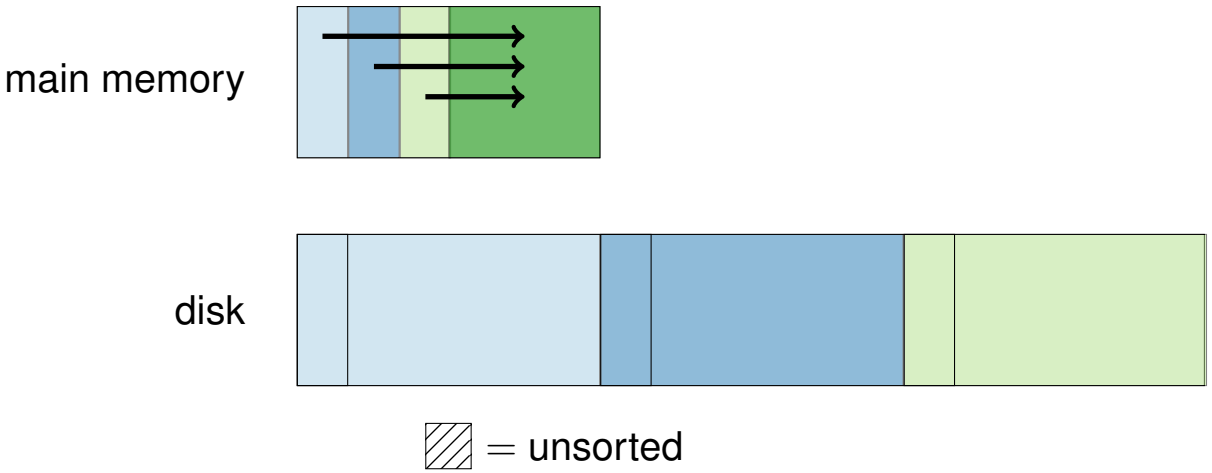
Step: *k*-way merge

External Sort



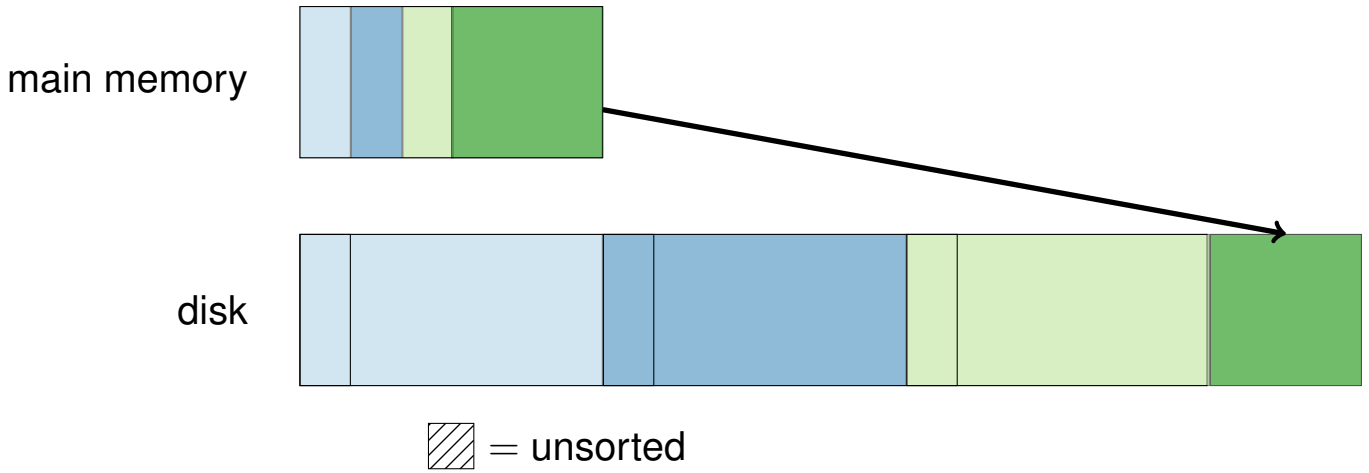
Step: *k*-way merge

External Sort



Step: k-way merge

External Sort



Step: *k*-way merge

External Sort

Step: done