Lecture 31: The IO Model 2 Repacking

Professor Xiannong Meng Spring 2018 Lecture and activity contents are based on what Prof Chris Ré of Stanford used in his CS 145 in the fall 2016 term with permission Repacking

Repacking for even longer initial runs

- With B+1 buffer pages, we can now start with *B***+1-length initial runs** (and use *B***-way merges) to get 2N(\left[\log_B \frac{N}{B+1}\right] + 1) IO cost...**
- Can we reduce this cost more by getting even longer initial runs?
- Use <u>repacking</u>- produce longer initial runs by "merging" in buffer as we sort at initial stage























Summary

- Basics of IO and buffer management. • See notebook for more fun! (Learn about *sequential flooding*)
- We introduced the IO cost model using sorting.
 Saw how to do merges with few IOs,
 Works better than main-memory sort algorithms.
- Described a few optimizations for sorting