





















































































## Other Interesting Parallel Join Implementation

Skew:

- Some partitions get more input tuples than others Reasons:
  - Range-partition instead of hash
  - Some values are very popular:
  - Heavy hitters values; e.g. 'Justin Bieber'
  - Selection before join with different selectivities
- Some partitions generate more output tuples than others

51

CSE 444 - Winter 2018



If using range partition:

- Ensure each range gets same number of tuples
- E.g.: {1, 1, 1, 2, 3, 4, 5, 6 } → [1,2] and [3,6]
- Eq-depth v.s. eq-width histograms

CSE 444 - Winter 2018

52

















