Spark Operations:

Transformations & Actions

Transformations:

1. Narrow: Operations: map(), filter(), union()
2. Partitioning : 1:1 input/output partitions
3. Performance: Efficient

* Wide: Operations: groupBy(), join(), distinct()
* Partitioning: Inputs span multiple output partitions
* Cost: 10-100x slower due to disk