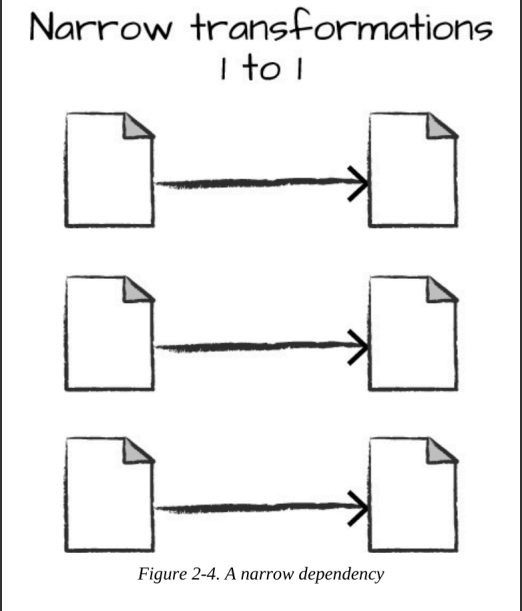


Spark will not act on transformations until we call an action

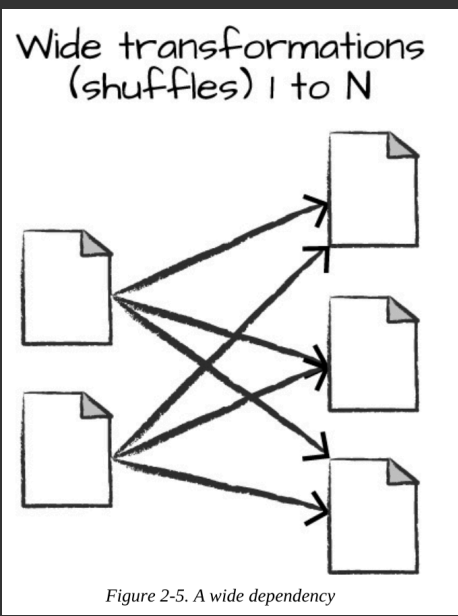
### Narrow dependencies |

- Each input partition will contribute to only one output partition
- Spark will automatically perform pipelining, meaning that if we specify multiple filters on DataFrames, they'll all be performed in-memory.

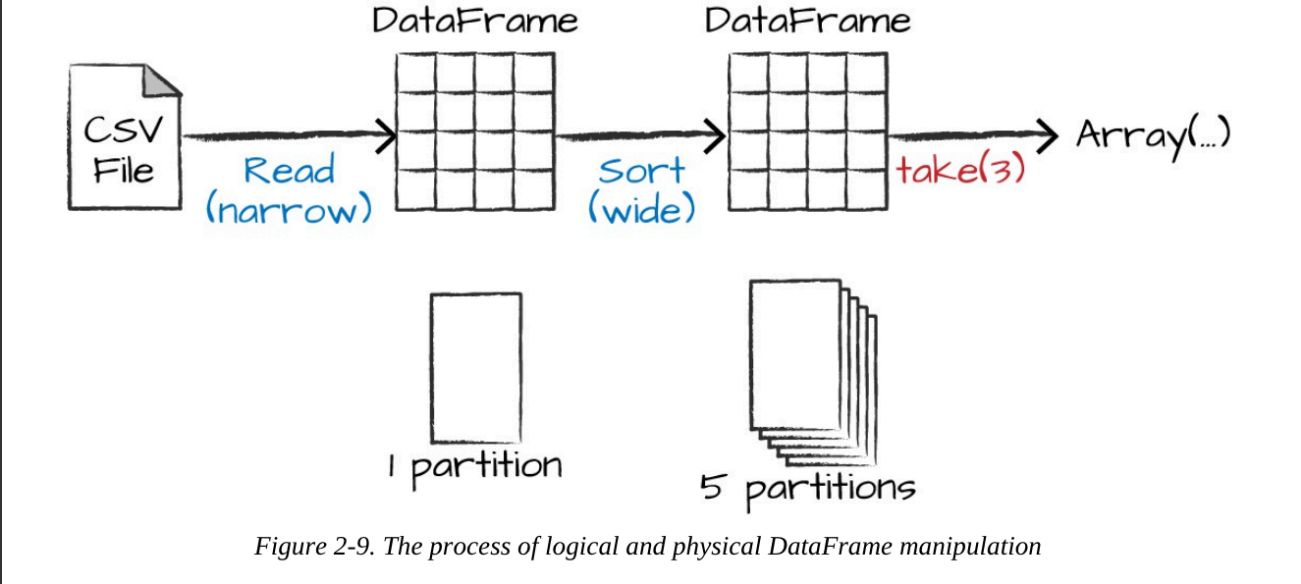


### Wide dependencies

- Input partitions contributing to many output partitions.
- Referred to as a *shuffle* - Spark will exchange partitions across the cluster
- When we perform a shuffle, Spark writes the results to disk



Read is narrow transformation, sort is wide transformation:



*Lazy Execution* - Spark prepares a plan of execution before executing (via Directed Acyclic Graph)

