



input tuples

-> read in each input, union together

cells

-> map rows into tuples, flatmap into dataset

attribute lists

-> key by value, aggregate group to attribute list for each value and remove values

inclusion lists

-> flatMap to inclusion lists



INDs

-> group by dependant, calculate intersection between all inclusion lists for dependent

