

SPARK ASSIGNMENT

Read datasets



Create a FlatMap for each dataset containing the attribute value as key and the attribute name as value and unify datasets into one dataset to get the cells



Group by attribute value, use collect_set on the attribute names and discard the attribute value column to get the attribute sets



Remove all entries with empty inclusion lists and print the result



Group the inclusion lists by key attribute and call mapGroups to intersect the attribute sets (with reduce)



Use explode on the attribute sets to get one row for each set and every element in the set, then remove the respective element from the set of dependents to get the inclusion lists



Team Taube_Nuesschen



Team Taube_Nuesschen