

Assignment 4 – Inclusion Discovery with Spark

Salih Mezraoui and Tobias Niedling
Group Herbstlaub

Solution

In essence, we followed the advice given in the slides:

- 1) Map each cell to (`<value>`, `<column>`) and make the result distinct
- 2) Group by `<value>` and collect the set of `<column>`
- 3) Map `<column>` sets to possible inclusions \rightarrow (`<col>`, [`<in1>`, `<in2>`, ...])
e.g. (a,b,c) \rightarrow (a,[b,c]),(b,[a,c]),(c,[a,b])
- 4) Group by possibly included column `<col>`
- 5) Intersect all collected lists, which leads to the result

Assignment 4
Salih Mezraoui,
Tobias Niedling

Chart 2