Exercise Sheet 8: Query Optimization 3

Exercise 1- Join Trees and Join Cost Estimation

Assume we are given the SQL query from last week's exercise sheet 7 and the information from the table below.

SELECT a.aliasName, pt.label
FROM persons p, aliases a, superPowers sp, powerType pt, powerOrigin po
WHERE pt.id=sp.powerType_id **AND** p.id=sp.persons_id **AND** po.id=p.powerOrigin_id **AND** p.id=a.persons_id **AND** po.label='Mutation' **AND** sp.powerStrength=>7
AND NOT a.aliasName='Magneto'

a. How many different Join trees can be built for this query? How many different Deep Left Join trees exist?

b. Perform an intermediate result estimation for the different JOINS required to answer the query and explain them.

c. In which cases, data statistics can help to improve intermediate result estimations? Give 2 examples how histograms could be used to improve the estimations.

d. Find a left join tree based on your intermediate result estimations as costs using the greedy procedure presented in the lecture. What are the overall costs?

Table	Column	Unique Values
Persons	*id*	45823
	powerOrigin	198
	firstName	10034
	lastName	23923
Aliases	*Id*	55021
	Persons_id	45711
	aliasName	55009
superPowers	*id*	74124
	powerType_id	13
	Persons_id	44998
	powerStrength	35
powerType	*Id*	23
	Label	23
	Description	23
PowerOrigin	*Id*	199
	Label	189
	Description	189