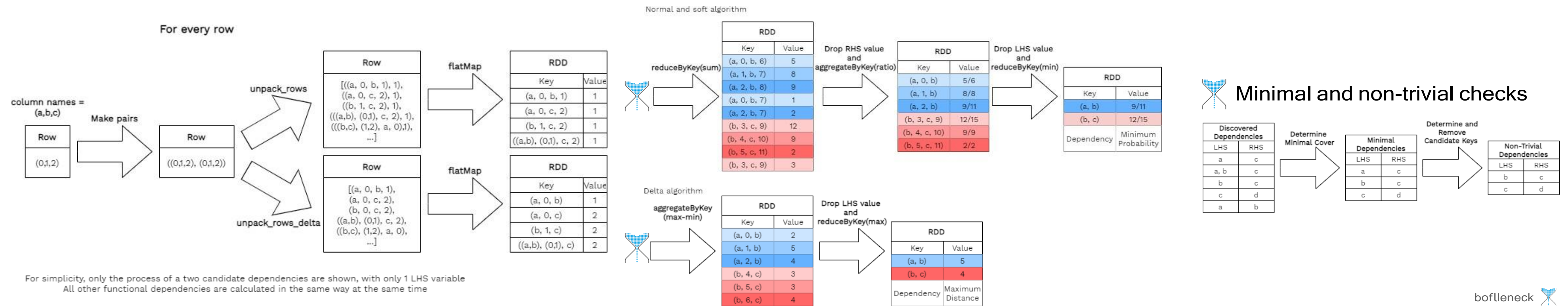
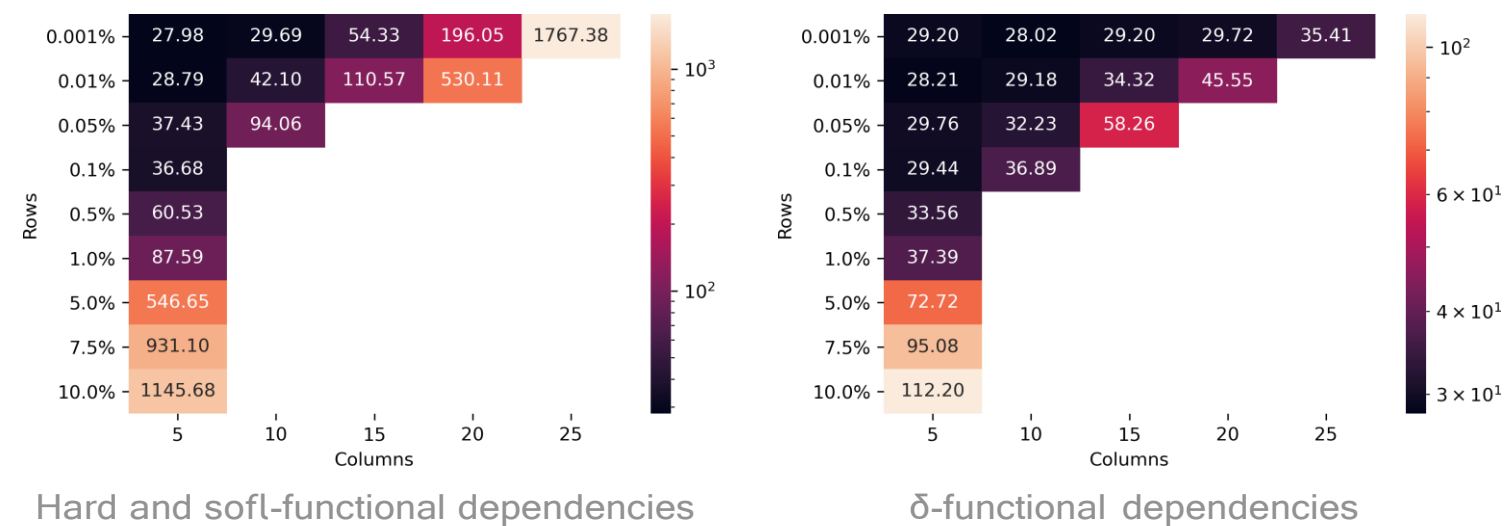


# Voters Dataset | Big Data Management | Milestone 2

## distribution of computation



## experimental performance and scalability



Execution time for different combinations of number of rows and columns. Time in seconds, blank spaces represent a failed run.

CPU: Ryzen 7 3700X 8 Cores / 16 Threads  
RAM: 32GB DDR4 3600Mhz  
Disk: SSD NVME Western Digital SN550 2400 MB/s

## detected dependencies

### hard dependencies

{house\_num, last\_name} -> {voter\_status\_reason\_desc}  
{first\_name} -> {state\_cd}  
{first\_name, street\_name} -> {mail\_state}  
{first\_name, zip\_code} -> {res\_city\_desc}  
{county\_desc} -> {county\_id}  
{first\_name, street\_name} -> {voter\_status\_desc}  
{midl\_name} -> {state\_cd}  
{county\_id, last\_name} -> {county\_desc}  
{midl\_name, voter\_status\_reason\_desc} -> {voter\_status\_desc}  
...

### soft dependencies for τ=0.8

{county\_id} -> {county\_desc}  
{voter\_status\_reason\_desc} -> {voter\_status\_desc}  
{state\_cd} -> {mail\_state}  
{voter\_status\_desc} -> {mail\_state}  
...

### delta dependencies for δ=25

{age\_group} -> {age}  
{sex} -> {salary}  
{age, voter\_reg\_num} -> {county\_id}  
{last\_name, paky\_desc, street\_name} -> {county\_id}  
{county\_desc, ncid} -> {zip\_code}  
{midl\_name, res\_city\_desc, street\_name} -> {zip\_code}  
{first\_name, house\_num, race\_desc} -> {salary}  
{house\_num, last\_name, midl\_name} -> {county\_id}  
{county\_id, house\_num, last\_name} -> {zip\_code}  
...

interesting unexpected dependencies  
expected dependencies

