Beatrix Haddock

12/16/2022

# Profiling: Synthetic Population Scale-Up

Notes from trying to scale up different simulation parameters. I think most of the time the increase in memory is due to the synthetic\_pii data. Never tried running with name\_ids AND address\_ids (instead of names and addresses)

There are a few more profiling notes in some of my jira tickets related to profiling (attached as screenshots, etc); I’m sorry I am making this too late and don’t have access to those, now. However, this should contain the overarching info (most of those were re-runs or less granular runs asking the same questions answered here)

## TESTING ADDRESS IDS INSTEAD OF ADDRESSES

Profiling with:

* 10.75 million employers
* running for 10 years starting in April 2020
* address\_ids instead of addresses for both households and businesses
* max time 10 hours
* 300 GB

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 20,000 sims | 200,000 sims | 2,000,000 sims | 20,000,000 sims |
| Setup | 1.566 min | 1.127 min | ~1 min | ~ 1 min |
| Init | 1.683 min | 4.636 min | 53 min |  |
| Mean timestep | 8.104 seconds | 41 seconds | ~15 min |  |
| Main loop | 18.369 min | 1.6 hours |  |  |
| Failure notes |  |  | Ran out of time in 2023. | OOMed during initialization. Probably due to name sampling. |

## TESTING SCALING UP POP SIZE

* Small number of employers (exact number given by using standard employer size generation parameters:
  + EXPECTED\_EMPLOYEES\_PER\_BUSINESS = 90.105203

np.random.lognormal(

4, 1, size=int(n\_need\_employers // data\_values.EXPECTED\_EMPLOYEES\_PER\_BUSINESS)

).round()

* Using all synthetic\_pii data (names, home addresses, business addresses)

|  |  |  |
| --- | --- | --- |
| Pop size | Runtime | memory |
| 60,000 | 2h3 | 9361 |
| 80,000 | 4h6 | 9365MB |
| 100,000 | 1h40m | 10222MB |
| 200,000 | 2h7 | 10398MB |
| 400,000 | 8h0m | 14190MB |
| 600,000 | 15h22m | 26947MB |
| 800,000 | 19h33m | 44738MB |
| 1,000,000 | 22h40m | 71515 |

## TESTING SCALING UP EMPLOYER TABLE

* 2,000 simulants
* Using addresses, names

