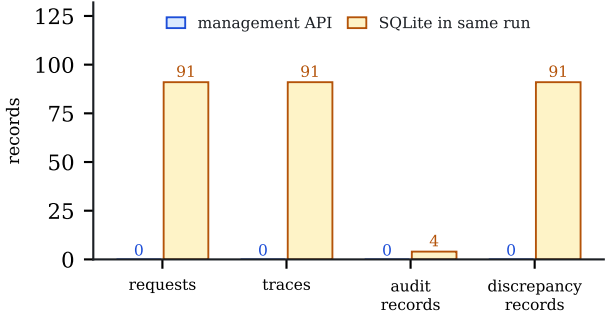
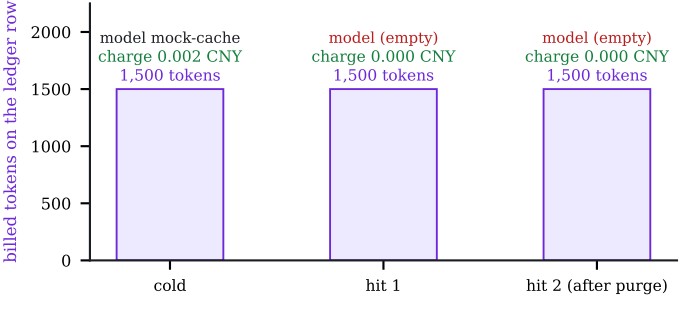

the control plane reports nothing the data plane wrote
management API
SQLite in same run
records
125
100
75
50
25
0
0
91
0
91
0
4
0
91
requests
traces
audit records
discrepancy records


cache replays keep the token count and drop the model (run total CNY 0.162)
billed tokens on the ledger row
2000
1500
1000
500
0
model mock-cache
charge 0.002 CNY
1,500 tokens
model (empty)
charge 0.000 CNY
1,500 tokens
model (empty)
charge 0.000 CNY
1,500 tokens
cold
hit 1
hit 2 (after purge)