Praktikum Teknologi Perekayasaan Data

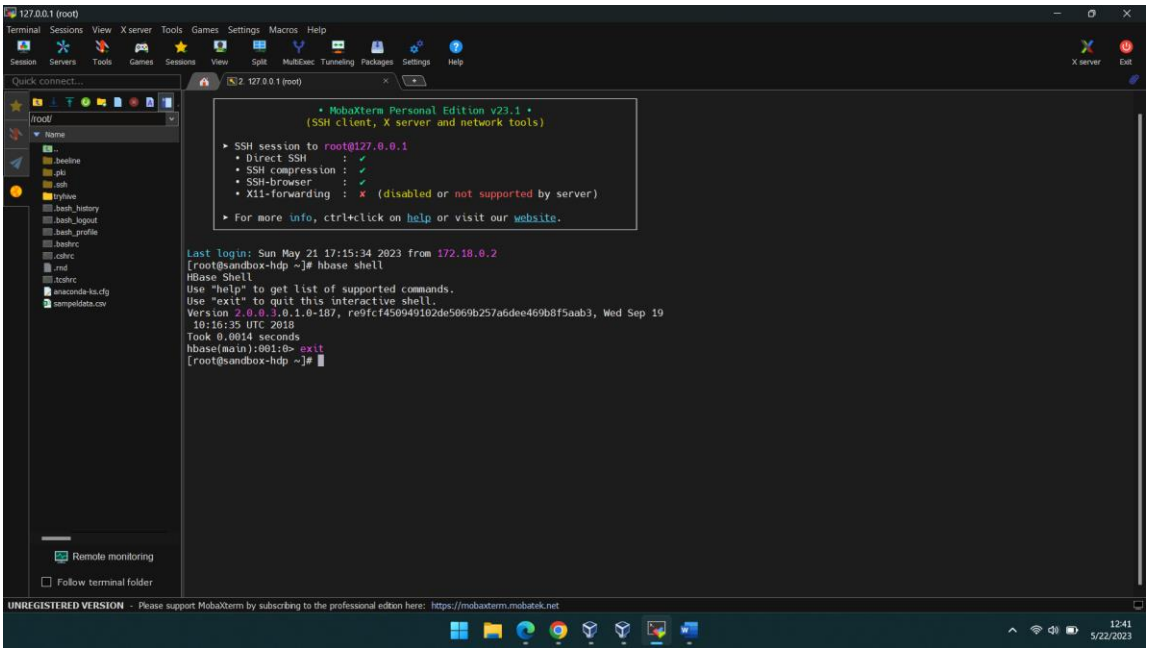
Tanggal Praktikum : Monday, May 22, 2023

Kelola Data dalam HBase

1. Masuk ke HBase shell

   hbase shell

   exit

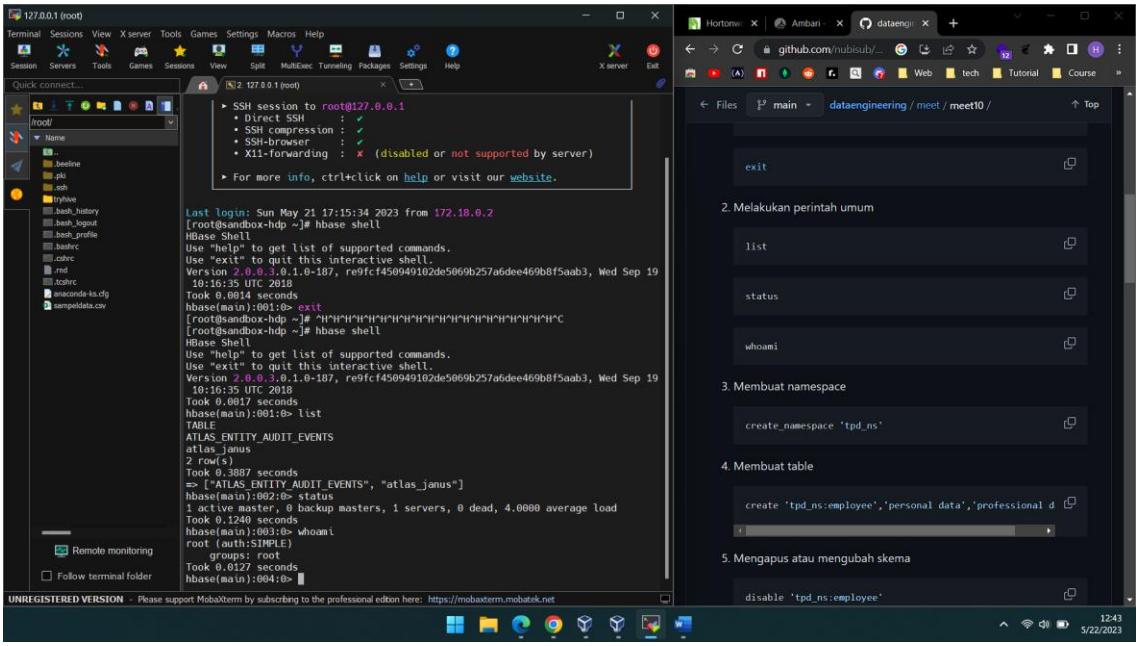


2. Melakukan perintah umum

   list

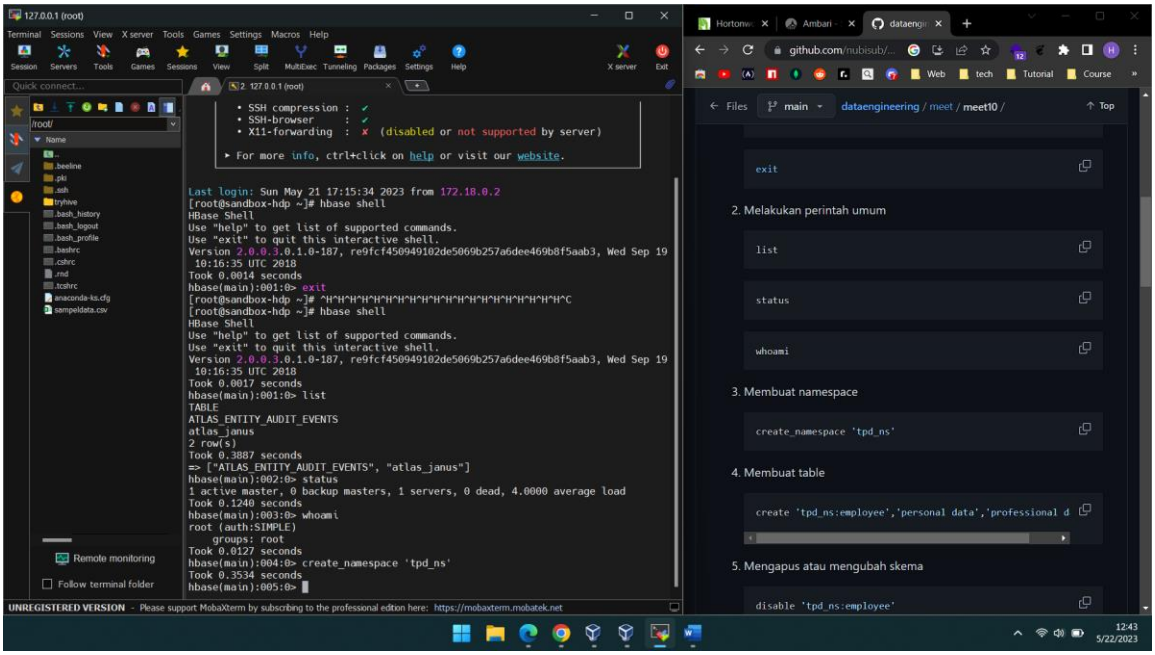
   status

   whoami



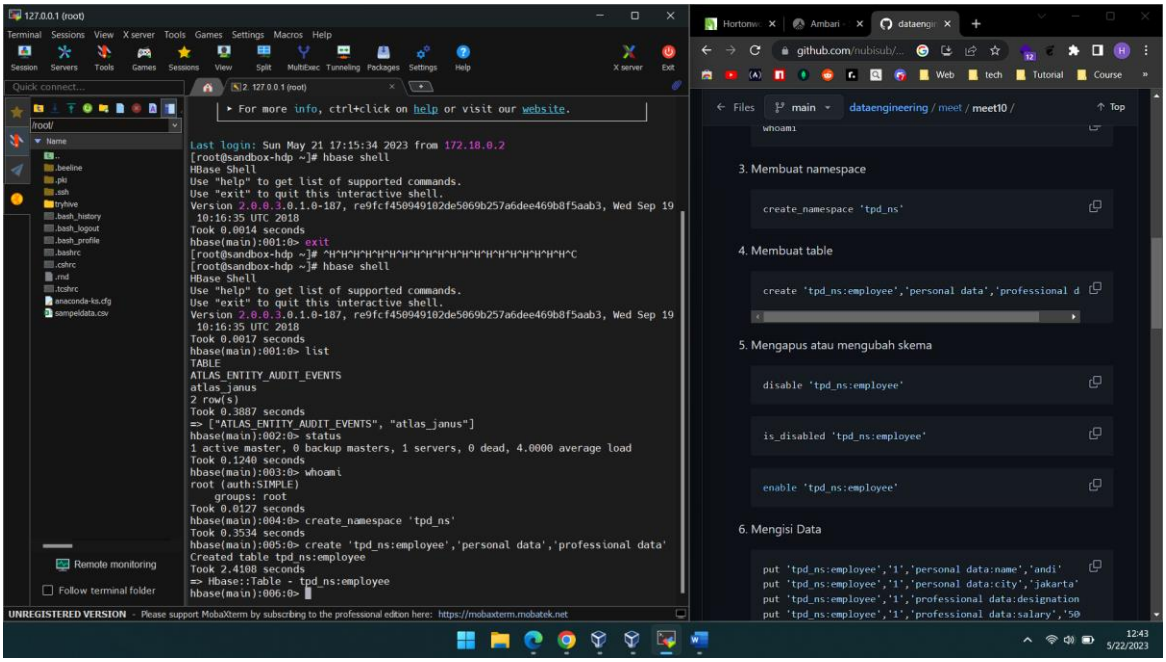
3. Membuat namespace

create_namespace 'tpd_ns'



4. Membuat table

create 'tpd_ns:employee','personal data','professional data'

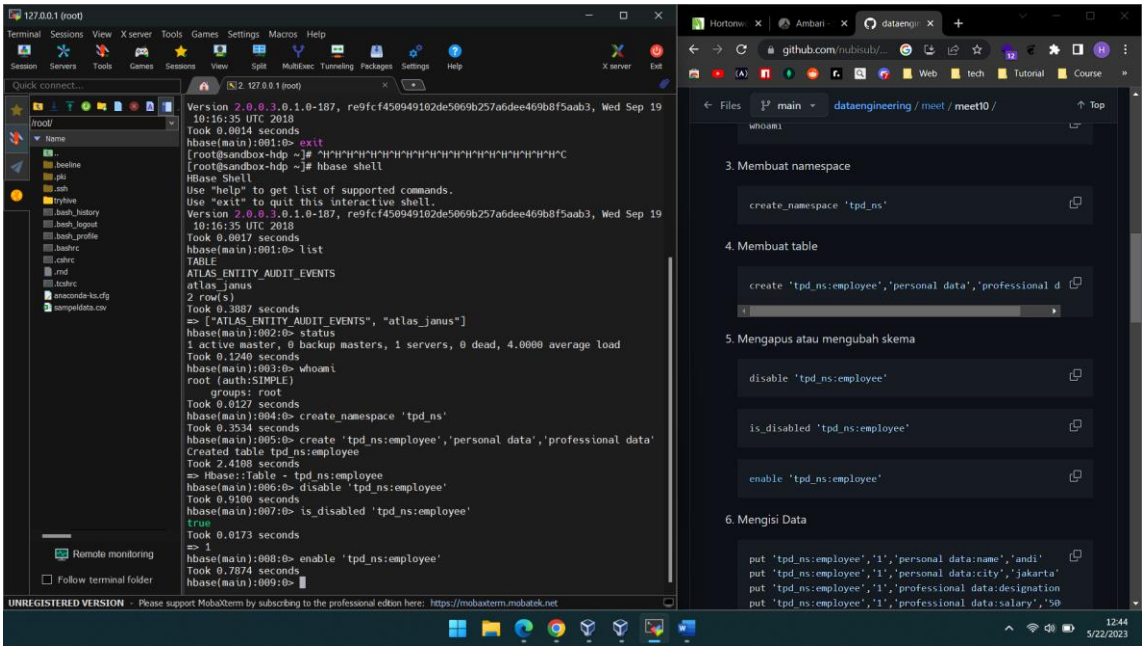


5. Mengapus atau mengubah skema

disable 'tpd_ns:employee'

is_disabled 'tpd_ns:employee'

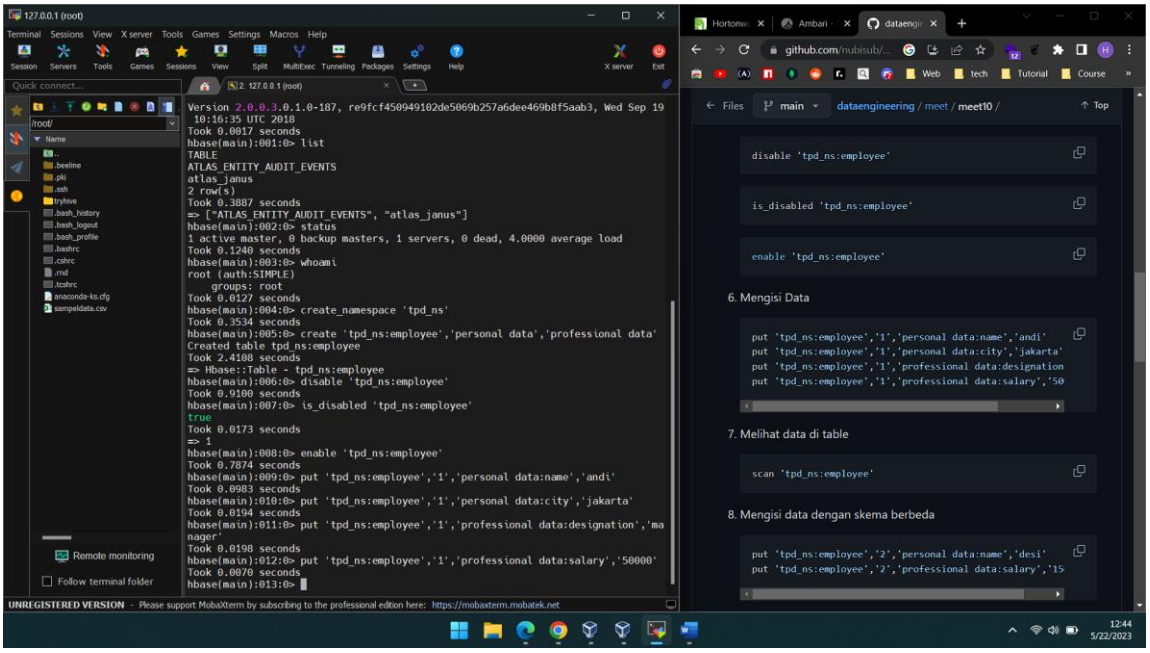
enable 'tpd_ns:employee'



6. Mengisi Data

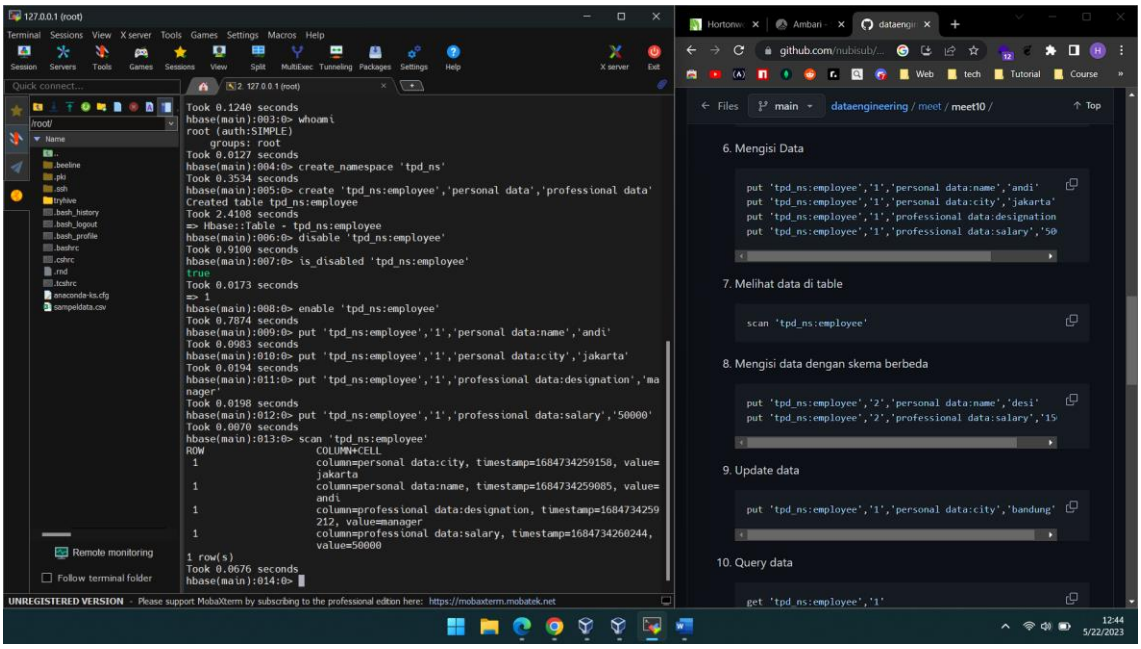
7. put 'tpd_ns:employee','1','personal data:name','andi'
8. put 'tpd_ns:employee','1','personal data:city','jakarta'

9. put 'tpd_ns:employee','1','professional data:designation','manager'
   put 'tpd_ns:employee','1','professional data:salary','50000'



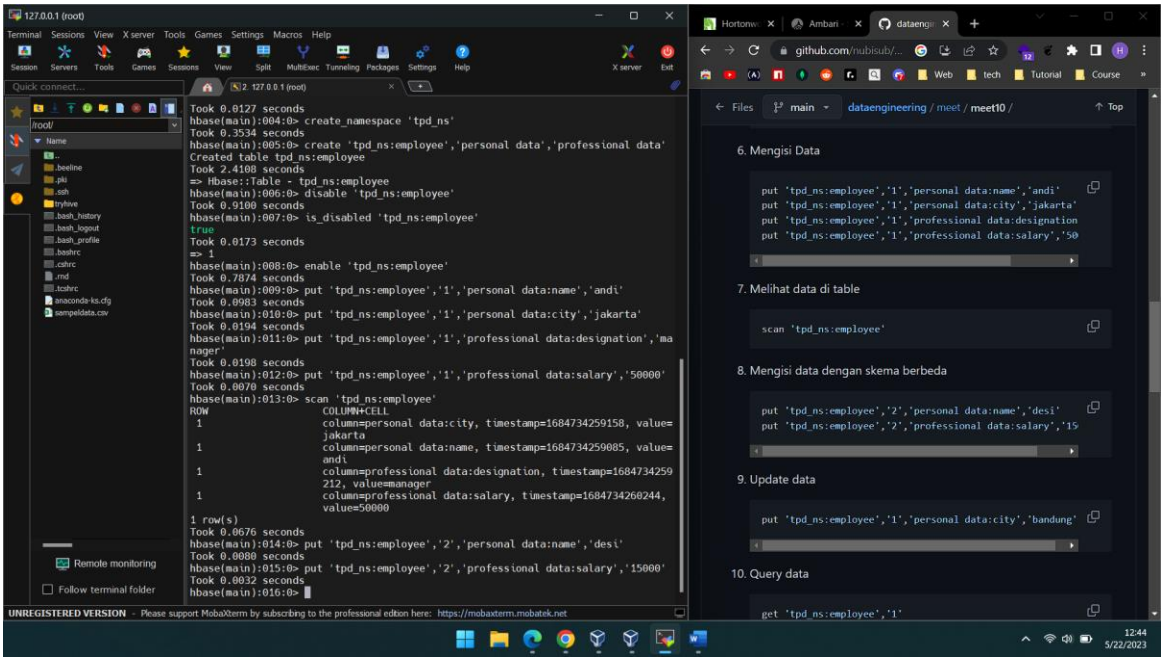
10. Melihat data di table

    scan 'tpd_ns:employee'



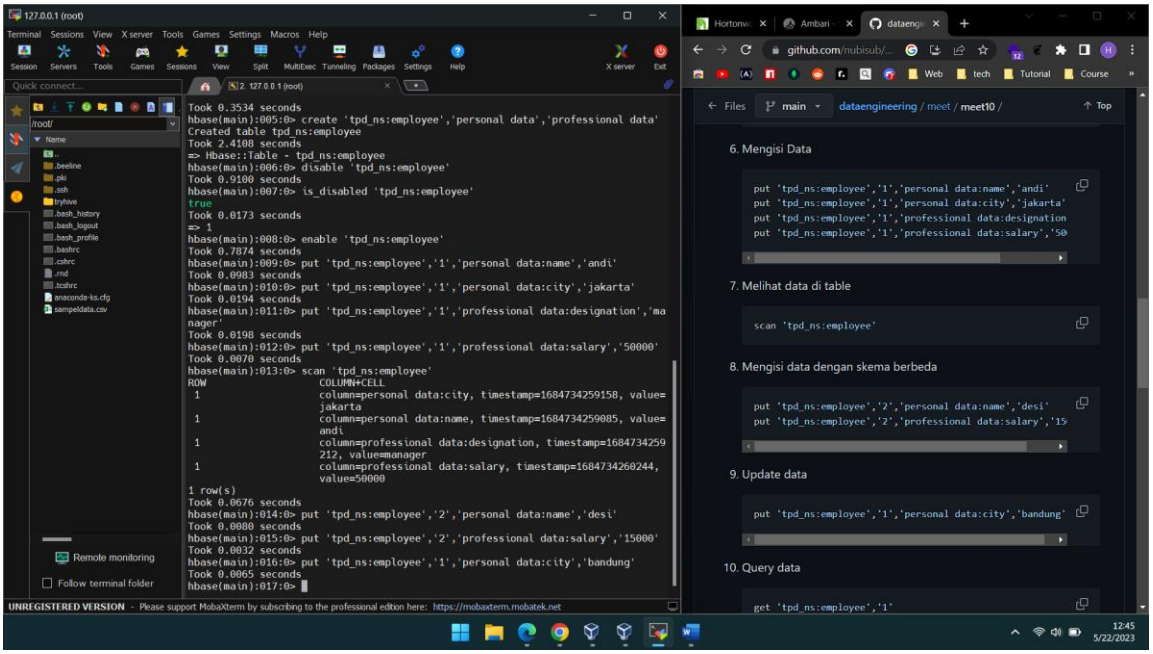
11. Mengisi data dengan skema berbeda

12. put 'tpd_ns:employee','2','personal data:name','desi'
    put 'tpd_ns:employee','2','professional data:salary','15000'



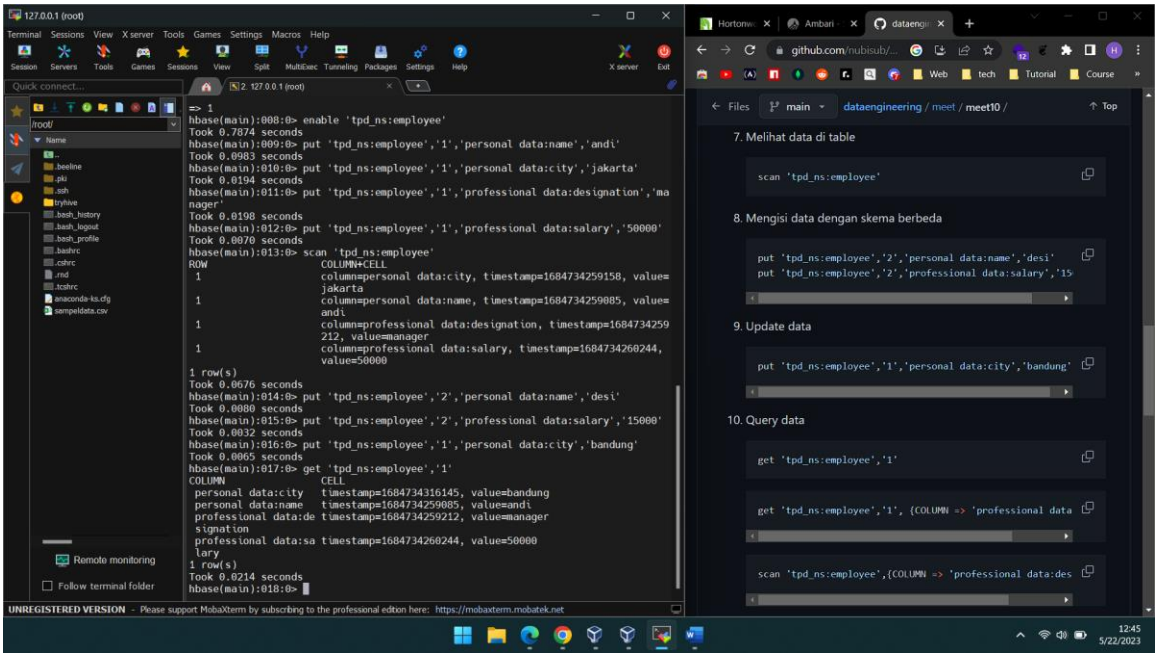
13. Update data

put 'tpd_ns:employee','1','personal data:city','bandung'

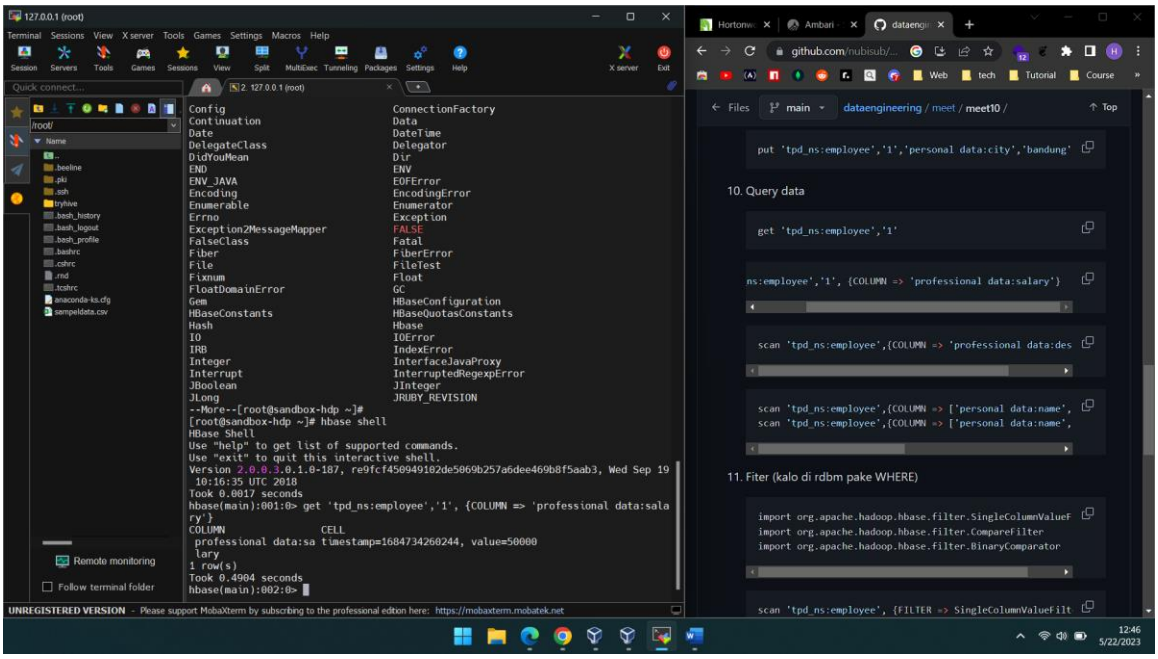


14. Query data

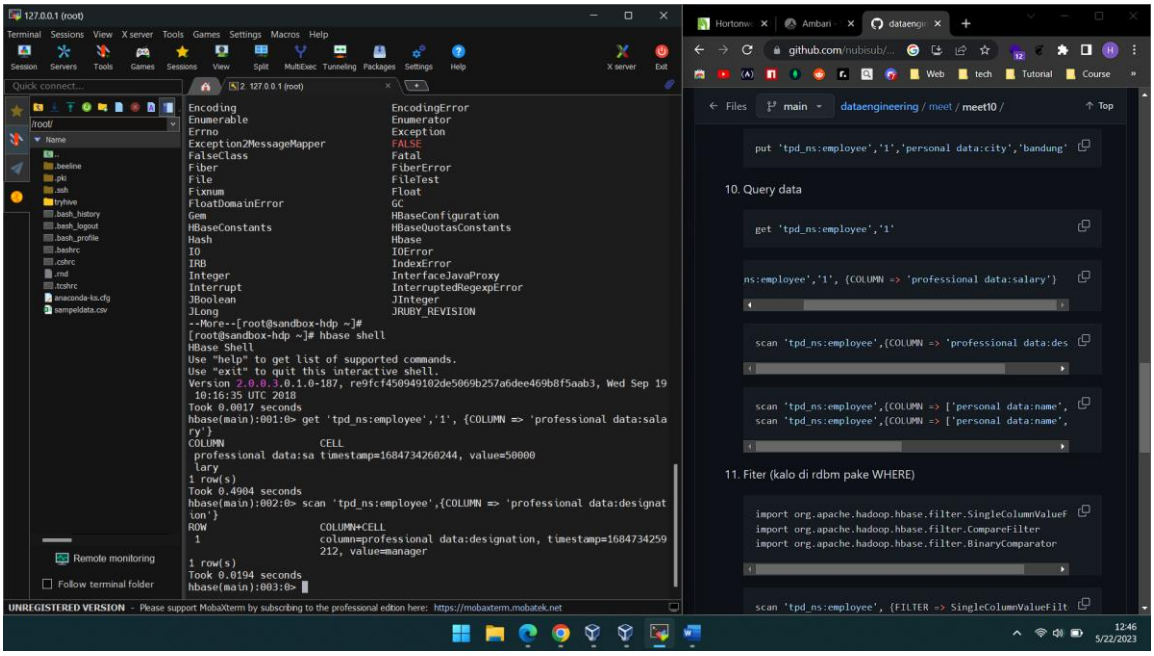
get 'tpd_ns:employee','1'



get 'tpd_ns:employee','1', {COLUMN => 'professional data:salary'}

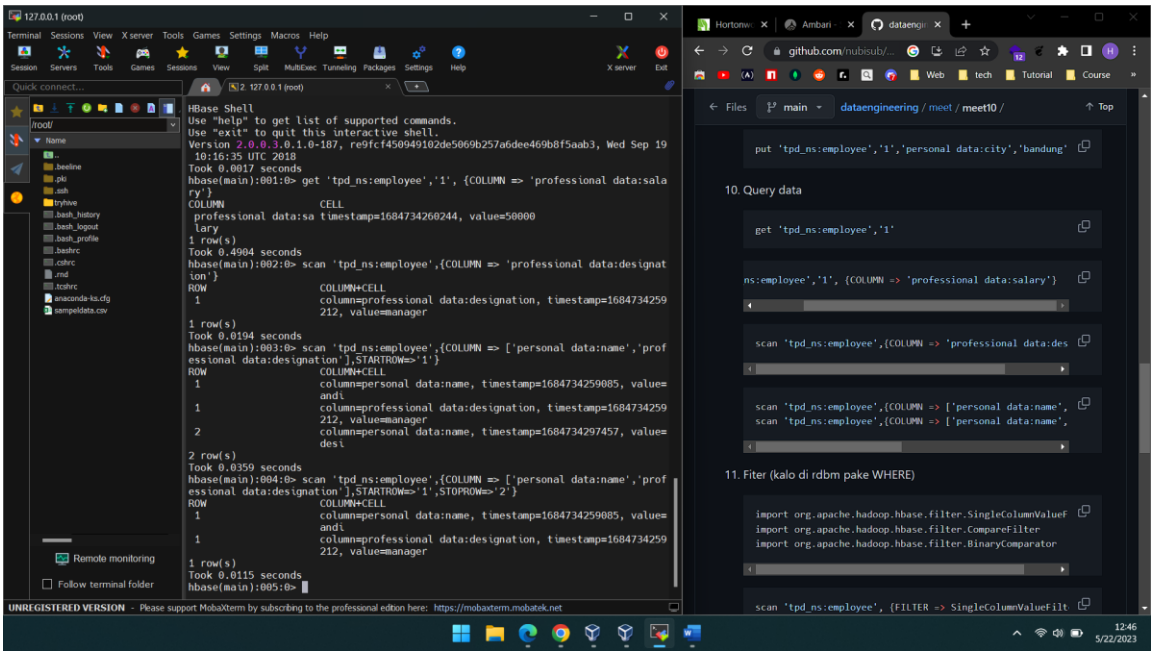


scan 'tpd_ns:employee',{COLUMN => 'professional data:designation'}



scan 'tpd_ns:employee',{COLUMN => ['personal data:name','professional data:designation'],STARTROW=>'1'}

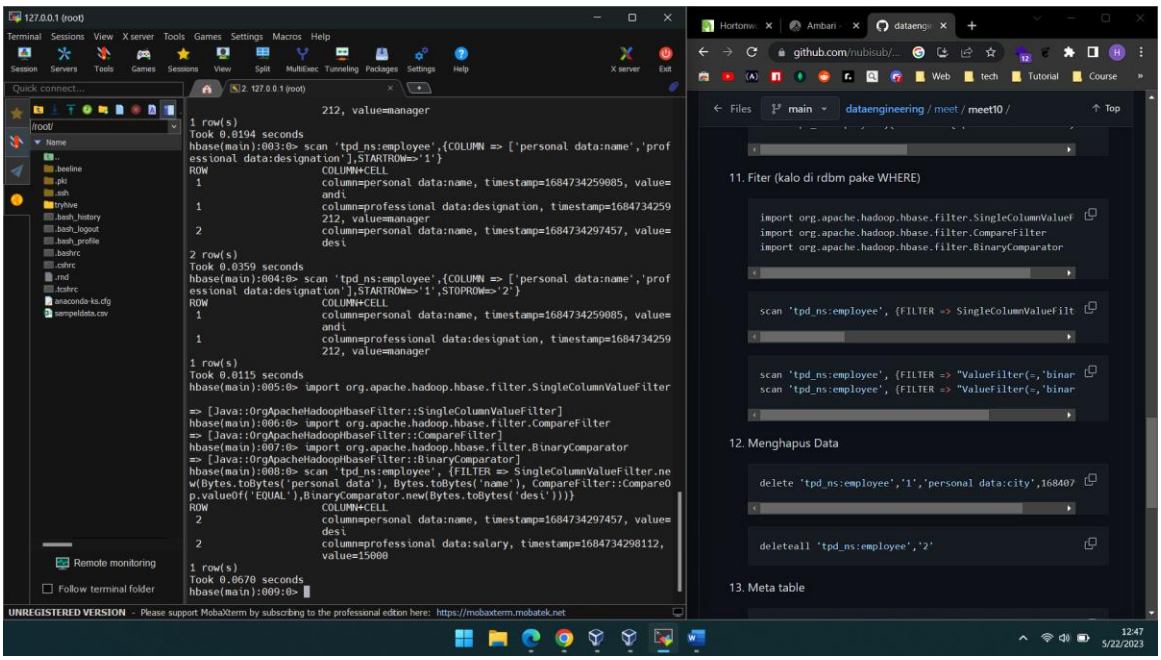
scan 'tpd_ns:employee',{COLUMN => ['personal data:name','professional data:designation'],STARTROW=>'1',STOPROW=>'2'}



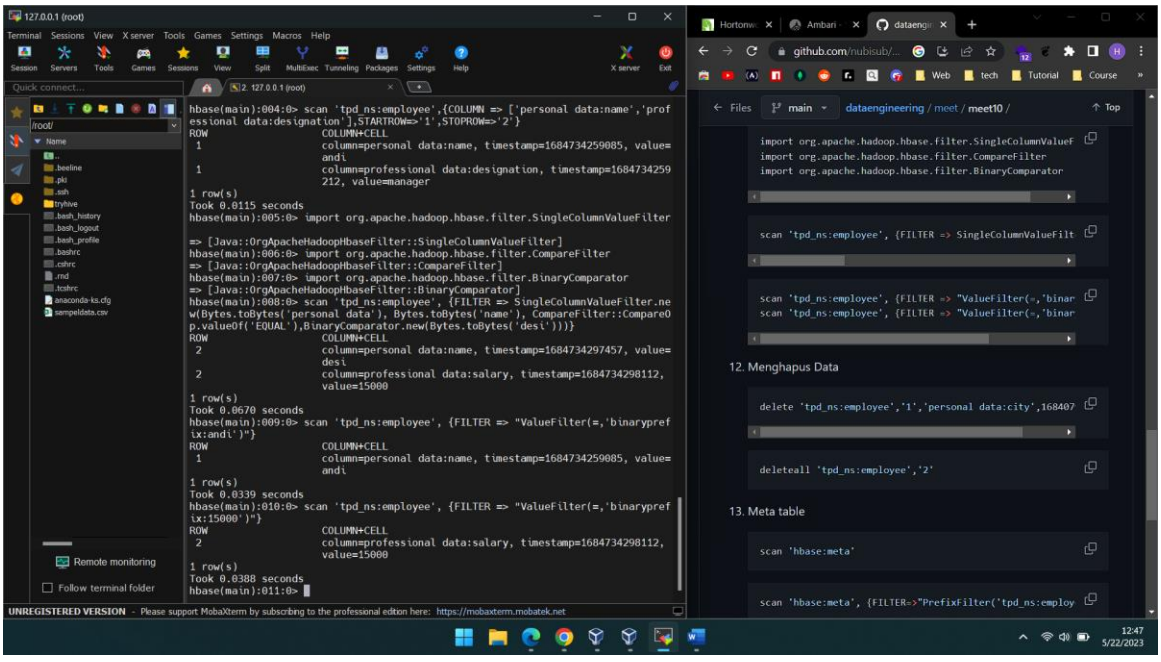
15. Fiter (kalo di rdbm pake WHERE)

16. import org.apache.hadoop.hbase.filter.SingleColumnValueFilter
17. import org.apache.hadoop.hbase.filter.CompareFilter
    import org.apache.hadoop.hbase.filter.BinaryComparator

scan 'tpd_ns:employee', {FILTER => SingleColumnValueFilter.new(Bytes.toBytes('personal data'), Bytes.toBytes('name'), CompareFilter::CompareOp.valueOf('EQUAL'),BinaryComparator.new(Bytes.toBytes('desi')))}



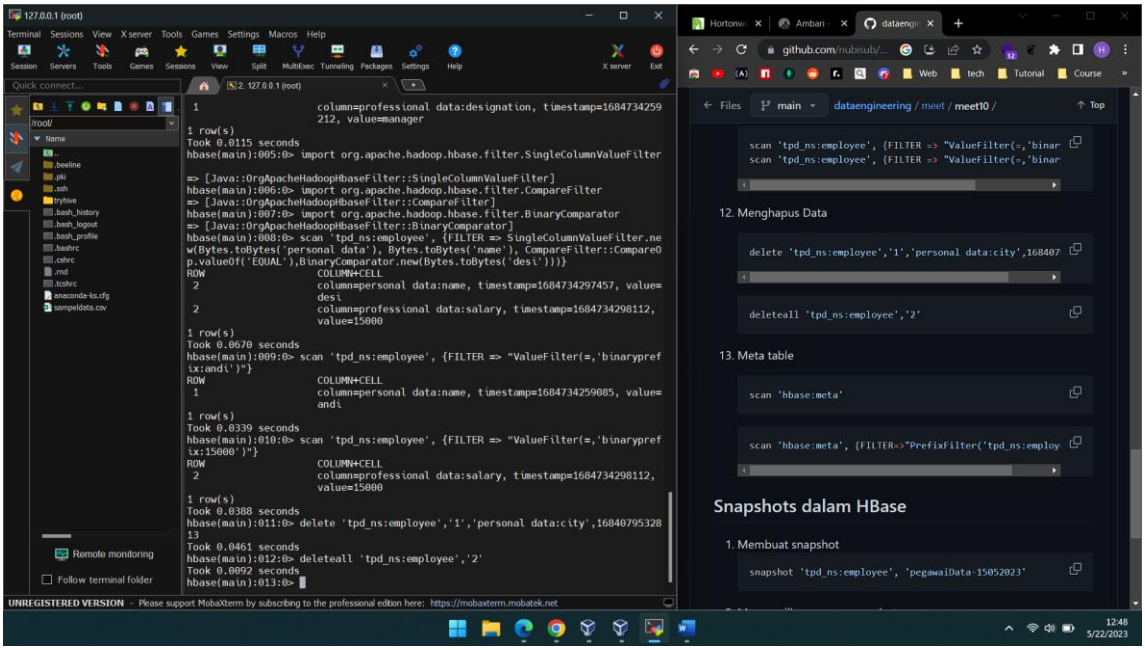
scan 'tpd_ns:employee', {FILTER => "ValueFilter(=,'binaryprefix:andi')"}
scan 'tpd_ns:employee', {FILTER => "ValueFilter(=,'binaryprefix:15000')"}



18. Menghapus Data

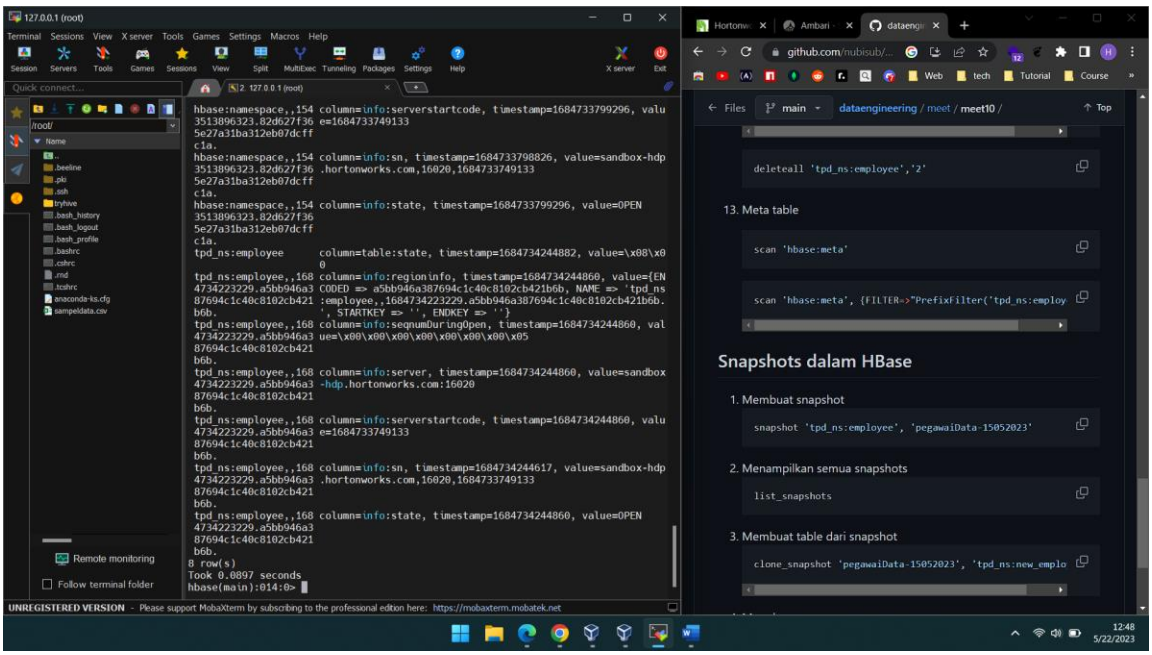
delete 'tpd_ns:employee','1','personal data:city',1684079532813

deleteall 'tpd_ns:employee','2'

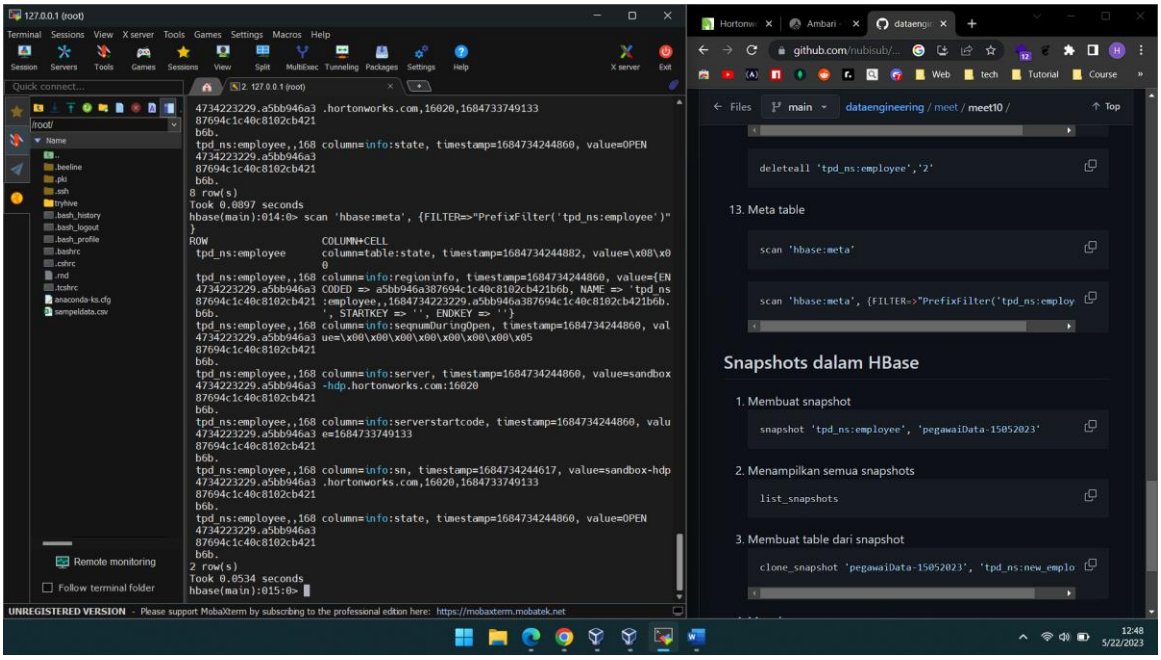


19. Meta table

scan 'hbase:meta'



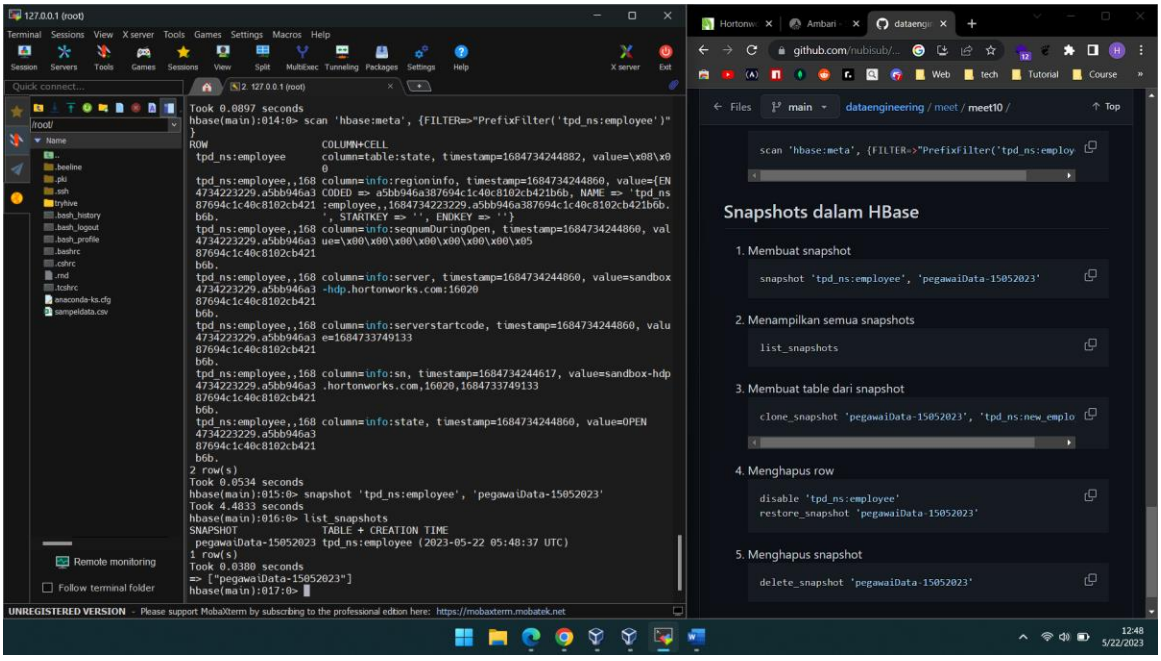
scan 'hbase:meta', {FILTER=>"PrefixFilter('tpd_ns:employee')"}



Snapshots dalam HBase

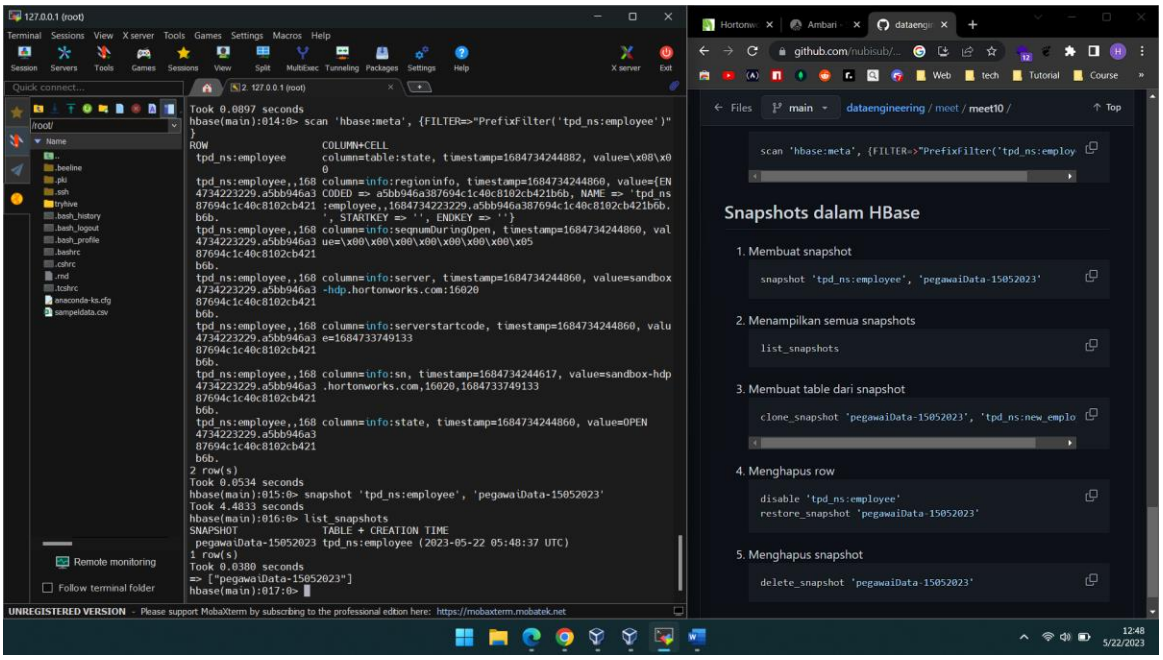
1. Membuat snapshot

   snapshot 'tpd_ns:employee', 'pegawaiData-15052023'



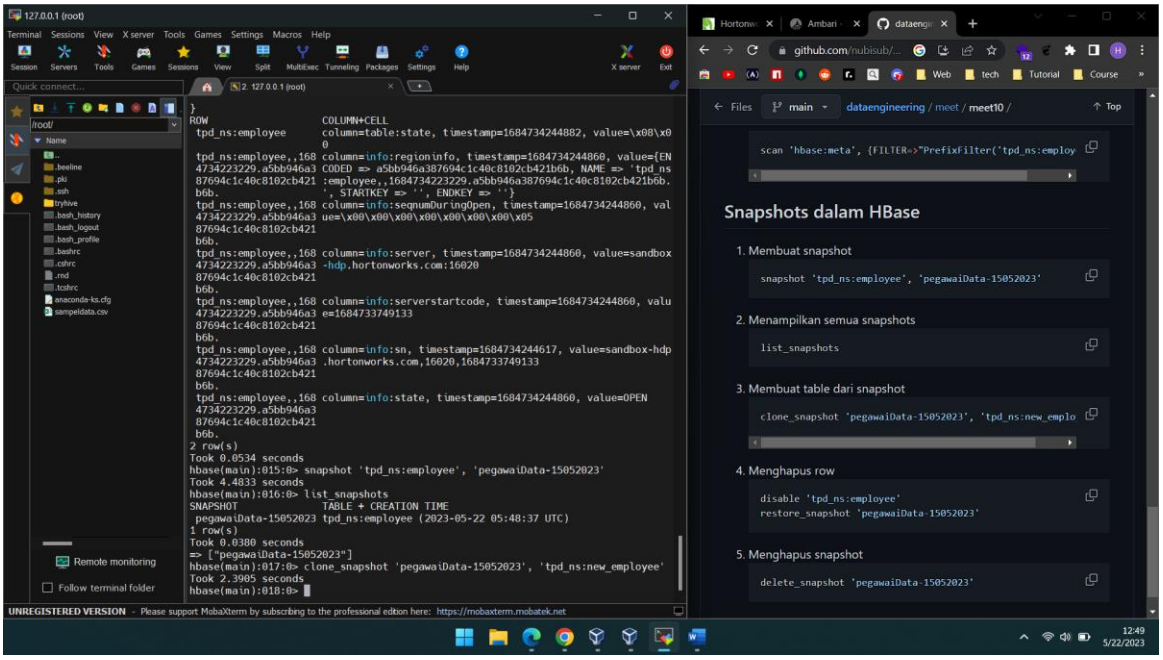
2. Menampilkan semua snapshots

   list_snapshots

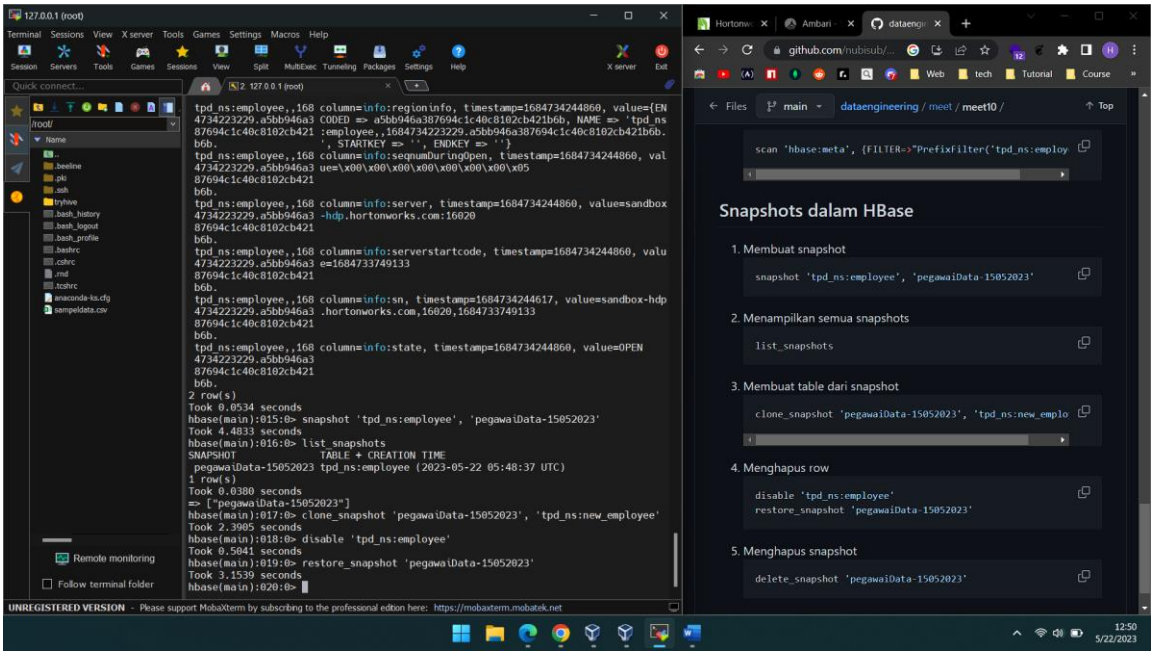


3. Membuat table dari snapshot

clone_snapshot 'pegawaiData-15052023', 'tpd_ns:new_employee'

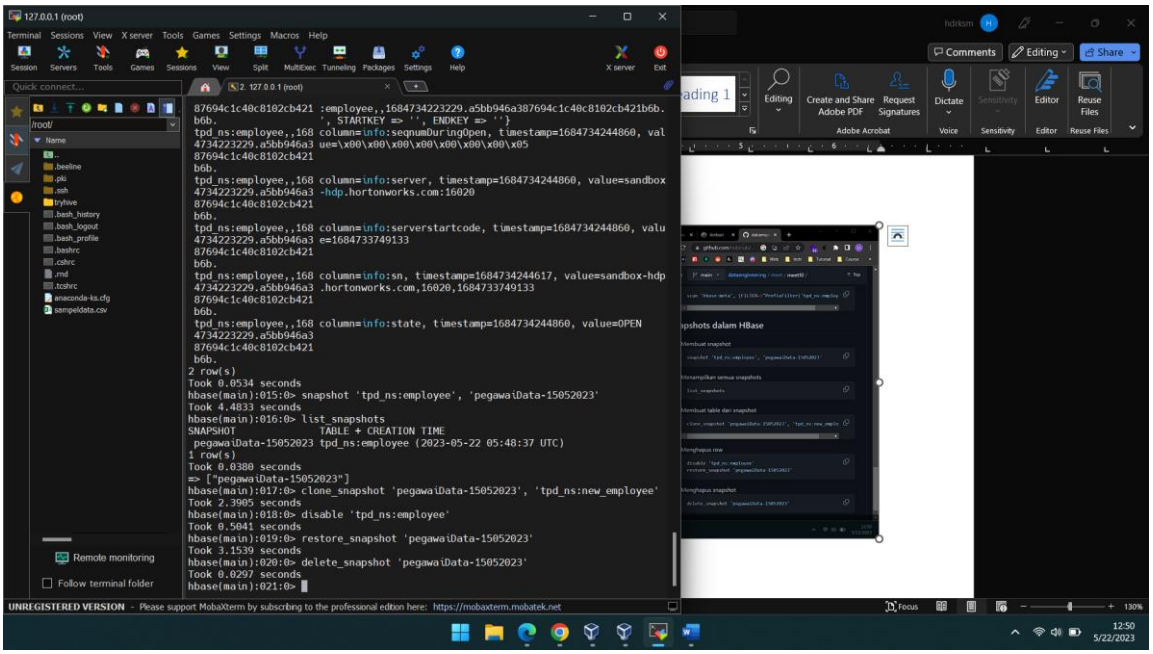


4. Menghapus row
5. disable 'tpd_ns:employee'
   restore_snapshot 'pegawaiData-15052023'



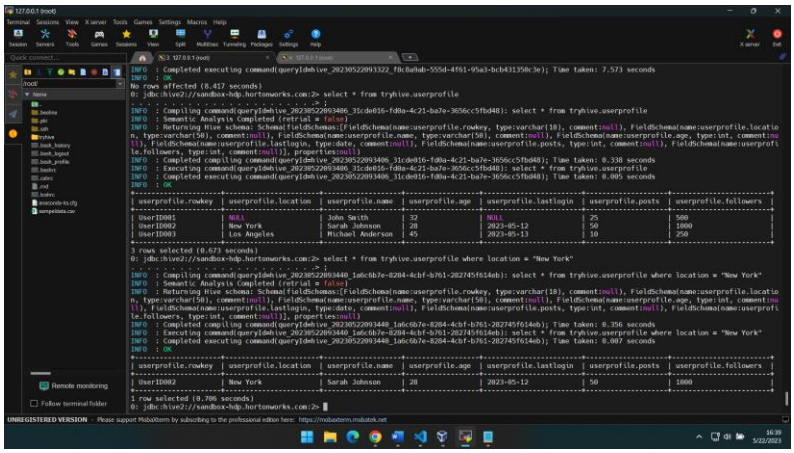
6. Menghapus snapshot

delete_snapshot 'pegawaiData-15052023'



1. Bandingkan performa read/write dengan Hive di praktikum modul sebelumnya, manakah yang relative lebih cepat?

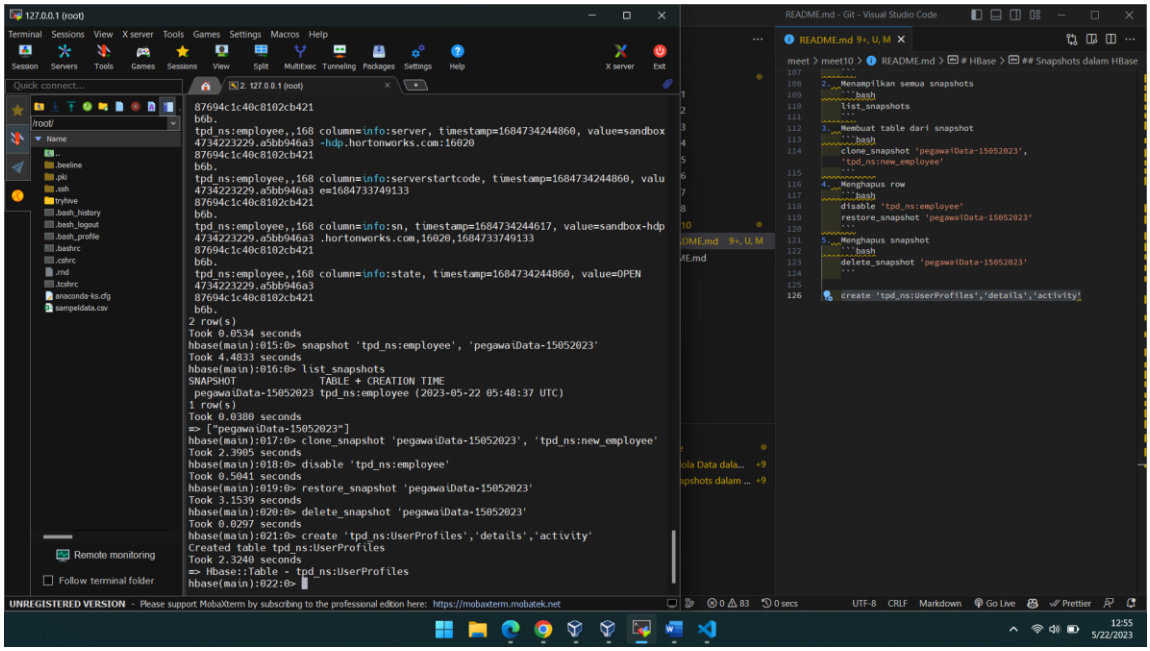
   Dilakukan pengujian menggunakan data user profile dengan 3 row baris saja

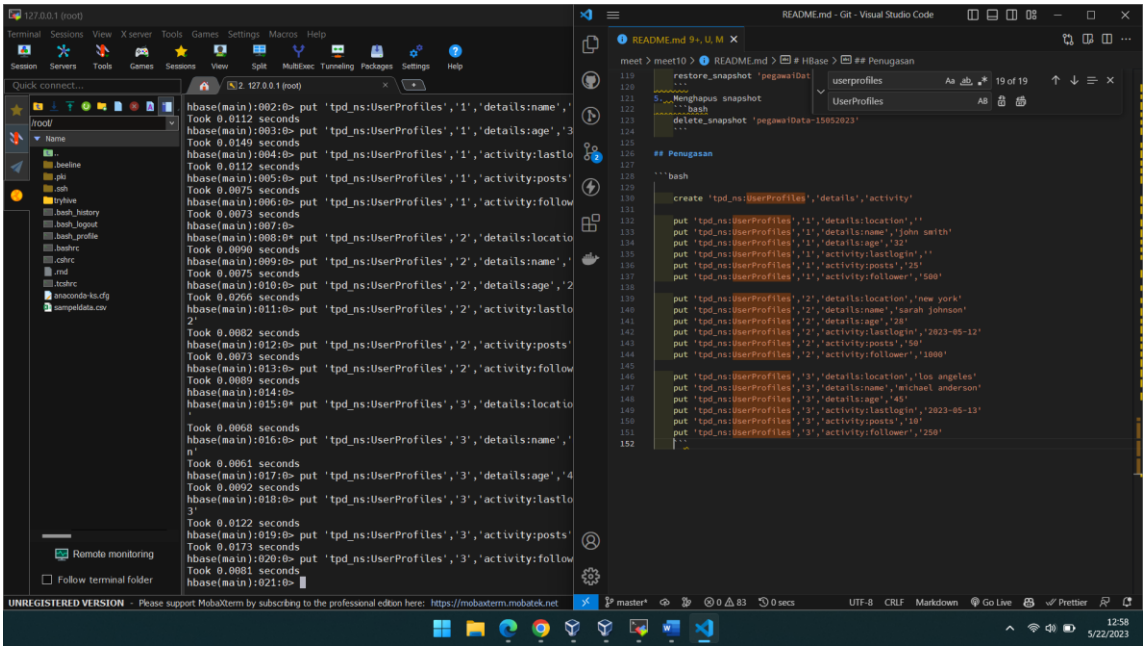


	Hive (seconds)	**Hbase** (seconds)
Write	8.417	0.192
Read	1.846	0.706

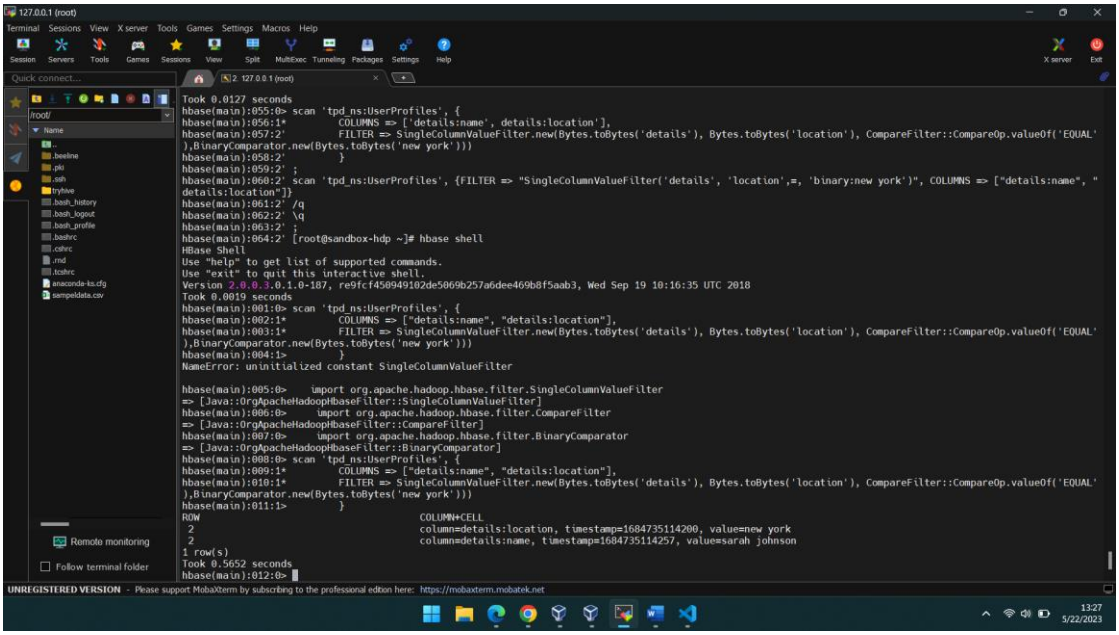
Dari pengujian di atas dapat disimpulkan bahwa performa read dan write pada Hbase lebih baik dari Hive

2. Buat table UserProfiles





a. Query untuk Nama user yang berlokasi di New York



b. Query untuk Nama dan last login user yang melakukan posting lebih dari 20 kali