数据集的制作

> 后续处理数据时直接使用 public dataset in BigQuery，这里仅展示数据获取方法

使用 https://github.com/blockchain-etl/ethereum-etl/ 项目

运行环境：

```
1    CPU Model          : Common KVM processor
2    CPU Cores          : 3 @ 3399.996 MHz
3    CPU Cache          : 16384 KB
4    AES-NI             : ✗ Disabled
5    VM-x/AMD-V          : ✗ Disabled
6    Total Disk         : 96.8 GB (30.3 GB Used)
7    Total Mem          : 3.8 GB (2.6 GB Used)
8    System uptime      : 10 days, 10 hour 16 min
9    Load average       : 1.84, 1.26, 0.76
10   OS                 : Ubuntu 22.04.4 LTS
11   Arch               : x86_64 (64 Bit)
12   Kernel             : 5.15.0-112-generic
13   TCP CC             : cubic
14   Virtualization     : KVM
15   IPv4/IPv6          : ✓ Online / ✓ Online
16   Organization       : AS49770 Internetport Sweden AB
17   Location           : Stockholm / SE
18   Region             : Stockholm
```

安装Ethereum-ETL

进入Python3.8虚拟环境后执行

```
1    pip install ethereum-etl
```

获取 infura API

在 https://app.infura.io/ 获得API

My First Key
Last updated on 2024-06-17

993f75d8e6034b9b848ff6b90e978a81

Active Endpoints   ⊽ 3/24 ACTIVE

HTTPS   WebSockets   ADD ENDPOINTS

◆ **Ethereum** Holesky | DOCS   ᶻᶻ NO REQUESTS IN 1 HOUR   ∧ HIDE URL

https://holesky.infura.io/v3/993f75d8e6034b9b848ff6b90e978a81

◆ **Ethereum** Mainnet | DOCS   ᶻᶻ NO REQUESTS IN 1 HOUR   ∧ HIDE URL

https://mainnet.infura.io/v3/993f75d8e6034b9b848ff6b90e978a81

导出 blocks

```
1  ethereumetl export_blocks_and_transactions --start-block 0 --end-block 1000 \
2  --blocks-output blocks.csv --transactions-output transactions.csv \
3  --provider-uri https://mainnet.infura.io/v3/993f75d8e6034b9b848ff6b90e978a81
```

```
19:59:13,796 - BatchWorkExecutor [INFO] - The batch of size 25 will be retried one item at a time.
19:59:16,326 - ProgressLogger [INFO] - 775 items processed. Progress is 77%.
19:59:16,374 - ProgressLogger [INFO] - 875 items processed. Progress is 87%.
19:59:16,452 - ProgressLogger [INFO] - 926 items processed. Progress is 92%.
19:59:16,745 - ProgressLogger [INFO] - 1001 items processed. Progress is 100%.
19:59:16,745 - ProgressLogger [INFO] - Finished work. Total items processed: 1001. Took 0:00:24.728463.
19:59:16,746 - CompositeItemExporter [INFO] - block items exported: 1001
19:59:16,746 - CompositeItemExporter [INFO] - transaction items exported: 0
```

```
1  number,hash,parent_hash,nonce,sha3_uncles,logs_bloom,transactions_root,
   state_root,receipts_root,miner,difficulty,total_difficulty,size,extra_data,
   gas_limit,gas_used,timestamp,transaction_count,base_fee_per_gas,
   withdrawals_root,withdrawals,blob_gas_used,excess_blob_gas
2  400,0x761103b53941451225dd7f7ab2c3d1eed0e1901f1475d1a753a20147ee5f9429,
   0xd93bd199c7dd249fd353d4aa58521a7e960a8b5e6c1dc10a3226ce144562cc88,
   0xb33dbe58ae31e4f3,
   0x1dcc4de8dec75d7aab85b567b6ccd41ad312451b948a7413f0a142fd40d49347,
   0x00000000000000000000000000000000000000000000000000000000000000000
   0000000000000000000000000000000000000000000000000000000000000000000
   0000000000000000000000000000000000000000000000000000000000000000000
   0000000000000000000000000000000000000000000000000000000000000000000
   0000000000000000000000000000000000000000000000000000000000000000000
   0000000000000000000000000000000000000000000000000000000000000000000
   00000000000000000000000000000000000000000000000000000000000000,
   0x56e81f171bcc55a6ff8345e692c0f86e5b48e01b996cadc001622fb5e363b421,
   0xcefeb930c0faeb5477ab34ae5333f143b91c75a0b8f5973c813546e0112435a3,
   0x56e81f171bcc55a6ff8345e692c0f86e5b48e01b996cadc001622fb5e363b421,
   0x28921e4e2c9d84f4c0f0c0ceb991f45751a0fe93,20742135970,7556615279346,539,
   0x476574682f76312e302e302f6c696e75782f676f312e342e32,5000,0,1438271032,
   0,,,,,
3  300,0xb3e37f7c14742bc54d08163792d38ada69c3951817b8dde6ef96776aa5c0f00c,
   0x989b8bf2af0be6c18c9c95bfde81492e0b47bcc1c26d555bb7cea2d09e92c6c3,
   0x424b554fa4a7a04f,
   0x1dcc4de8dec75d7aab85b567b6ccd41ad312451b948a7413f0a142fd40d49347,
   0x00000000000000000000000000000000000000000000000000000000000000000
   0000000000000000000000000000000000000000000000000000000000000000000
```

number	hash	parent_has	nonce	sha3_uncles	logs_bloom	transaction	state_root	receipts_ro	miner	difficulty	total_diffic	size	extra_data	gas_limit	gas_used	timestamp
400	0x761103b5	0xd93bd199	0xb33dbe58	0x1dcc4de8	0x00000000	0x56e81f171	0xcefeb930c	0x56e81f171	0x28921e4e;	#######	#######	539	0x47657468;	5000	0	1438271032
300	0xb3e37f7c1	0x989b8bf2;	0x424b554fc	0x1dcc4de8	0x00000000	0x56e81f171	0x34e5b524!	0x56e81f171	0xbb7b8287	#######	#######	544	0x47657468;	5000	0	1438270848
0	0xd4e56740!	0x00000000	0x00000000	0x1dcc4de8	0x00000000	0x56e81f171	0xd7f8974ft	0x56e81f171	0x00000000	1.718E+10	1.718E+10	540	0x11bbe8db	5000	0	0
100	0xdfe2e70d(	0xdb10afd3	0x37129c7f;	0x1dcc4de8	0x00000000	0x56e81f171	0x90c25f6d;	0x56e81f171	0xbb7b8287	#######	#######	542	0x47657468;	5000	0	1438270443
200	0x75208f95c	0x4dcc8333<	0xcb7bac79	0x1dcc4de8	0x00000000	0x56e81f171	0x9b6f2db3	0x56e81f171	0xbb7b8287	#######	#######	543	0x47657468;	5000	0	1438270646
401	0xcda52e43;	0x761103b5	0xec5294da	0x1dcc4de8	0x00000000	0x56e81f171	0x2970c5a3!	0x56e81f171	0x5088d623	#######	#######	546	0x47657468;	5000	0	1438271035
301	0x18e141524	0xb3e37f7c1	0x15c29cb9(	0x1dcc4de8	0x00000000	0x56e81f171	0x0a45b7b3	0x56e81f171	0x5088d623	#######	#######	546	0x47657468;	5000	0	1438270849
1	0x88e96d45;	0xd4e56740!	0x539bd497	0x1dcc4de8	0x00000000	0x56e81f171	0xd67e4d45	0x56e81f171	0x05a56e2d	#######	#######	537	0x47657468;	5000	0	1438269988
201	0xb840bb51	0x75208f95c	0xd37447f5!	0xea2138c6;	0x00000000	0x56e81f171	0xeb883899;	0x56e81f171	0xbb7b8287	#######	#######	1078	0x47657468;	5000	0	1438270647
101	0x4f66fd024	0xdfe2e70d(	0x4587d7ffc	0x1dcc4de8	0x00000000	0x56e81f171	0xc66500dc;	0x56e81f171	0xbb7b8287	#######	#######	542	0x47657468;	5000	0	1438270445
402	0xda32605b	0xcda52e43;	0x399aaa6d	0xd866254d	0x00000000	0x56e81f171	0x2608069e<	0x56e81f171	0xbb7b8287	#######	#######	1085	0x47657468;	5000	0	1438271036
2	0xb495a1d7	0x88e96d45;	0xb853fa26	0x1dcc4de8	0x00000000	0x56e81f171	0x4943d941	0x56e81f171	0xdd2f1e6e<	#######	#######	544	0x47657468;	5000	0	1438270017
302	0x2d4e5352;	0x18e141524	0x59d62198	0x15080878!	0x00000000	0x56e81f171	0x41ccd808l	0x56e81f171	0x1b7047b4	#######	#######	1076	0x42697463(	5000	0	1438270851
202	0x321aecc5!	0xb840bb51	0x8ffc76459	0xc0e09c60!	0x00000000	0x56e81f171	0x40b38881<	0x56e81f171	0x28921e4e;	#######	3.641E+12	1073	0x47657468;	5000	0	1438270648
102	0x16110f3a(	0x4f66fd024	0x8d36eecc(	0x1dcc4de8	0x00000000	0x56e81f171	0x2607447(	0x56e81f171	0x28921e4e;	#######	#######	537	0x47657468;	5000	0	1438270447
403	0x149e2a1d	0xda32605b	0x7dd24b3d	0xcf09a642<	0x00000000	0x56e81f171	0xf88e157d!	0x56e81f171	0x5088d623	#######	#######	1083	0x47657468;	5000	0	1438271039
3	0x3d612266(	0xb495a1d7	0x2e9344e0(	0x6b17b938	0x00000000	0x56e81f171	0x76ab0b89	0x56e81f171	0x5088d623	#######	#######	1079	0x47657468;	5000	0	1438270048
303	0x746a9defl	0x2d4e5352;	0xfa150394!	0x1dcc4de8	0x00000000	0x56e81f171	0x6e597b3a	0x56e81f171	0x1b7047b4	#######	5.591E+12	540	0x42697463(	5000	0	1438270854
203	0x0f72e36b1(	0x321aecc5!	0x08a646b0	0x1dcc4de8	0x00000000	0x56e81f171	0xc5b9bb07	0x56e81f171	0xbb7b8287	1.884E+10	#######	543	0x47657468;	5000	0	1438270649
103	0x39bef3da;	0x16110f3a(	0x1f7dfbfc9	0x1dcc4de8	0x00000000	0x56e81f171	0xf7c9ae63;	0x56e81f171	0x28921e4e;	#######	#######	537	0x47657468;	5000	0	1438270451
4	0x23adf5a3(	0x3d612266(	0xc2535b5e!	0x83a8da89	0x00000000	0x56e81f171	0xe6d9f6e95	0x56e81f171	0xc8ebccc5l	#######	#######	1079	0x59617465;	5000	0	1438270077
404	0xb9a105b6	0x149e2a1d	0x3a0758fe1	0x1dcc4de8	0x00000000	0x56e81f171	0xb6bdc14c	0x56e81f171	0x5088d623	#######	#######	546	0x47657468;	5000	0	1438271040
304	0x8a486e1d(	0x746a9defl	0x6091ac62!	0x1dcc4de8	0x00000000	0x56e81f171	0x1906cd49:	0x56e81f171	0x28921e4e;	#######	#######	539	0x47657468;	5000	0	1438270857
104	0x7faae5e9(	0x39bef3da;	0xdda64b66	0x1dcc4de8	0x00000000	0x56e81f171	0x4a38f3e8(	0x56e81f171	0xbb7b8287	#######	#######	542	0x47657468;	5000	0	1438270452
204	0x25bcb630(	0x0f72e36b1(	0x63f9b283<	0x1dcc4de8	0x00000000	0x56e81f171	0x77375b19	0x56e81f171	0xbb7b8287	1.885E+10	#######	543	0x47657468;	5000	0	1438270650
5	0xf37c632d;	0x23adf5a3(	0xfba9d0cff	0x1dcc4de8	0x00000000	0x56e81f171	0x4470f3dc1	0x56e81f171	0x05a56e2d	#######	#######	537	0x47657468;	5000	0	1438270083
405	0x9b125905(	0xb9a105b6	0x8df50c82(	0x1dcc4de8	0x00000000	0x56e81f171	0xca5910cc!	0x56e81f171	0x126b8ba0	#######	#######	546	0x47657468;	5000	0	1438271041
105	0x07efc5598	0x7faae5e9(	0x363fd3fbc	0x8aaaafa1	0x00000000	0x56e81f171	0xefadfb4c<	0x56e81f171	0x28921e4e;	1.796E+10	#######	1078	0x47657468;	5000	0	1438270453
305	0x5abb817d	0x8a486e1d(	0xb9107e0f!	0x1dcc4de8	0x00000000	0x56e81f171	0x8ca30980;	0x56e81f171	0x6923f88fc	#######	#######	540	0x59617465;	5000	0	1438270858
205	0x9e9be721<	0x25bcb630(	0x3e0632acl	0x1dcc4de8	0x00000000	0x56e81f171	0x62343949!	0x56e81f171	0x1b7047b4	#######	#######	539	0x42697463(	5000	0	1438270651
6	0x1f1aed8e;	0xf37c632d;	0x7bb9369d	0x1dcc4de8	0x00000000	0x56e81f171	0xc211464e(	0x56e81f171	0x0193d941	#######	#######	537	0x47657468;	5000	0	1438270107

访问BigQuery API

设置认证环境变量

在GCP的 IAM和管理 -> 服务账号 中新增账号并创建秘钥

设置环境变量

```
1  export GOOGLE_APPLICATION_CREDENTIALS="/Users/nanmener/Downloads/proven-mercury-382209-
   efdbd404ef1a.json"
```

安装 google-cloud-bigquery 库

```
1  pip install google-cloud-bigquery
```

查询2022年9月15日的前200条交易数据

```python
1  from google.cloud import bigquery
2  import pandas as pd
3
4  client = bigquery.Client()
5
6  query = """
7      SELECT *
8      FROM `bigquery-public-data.crypto_ethereum.transactions`
9      WHERE DATE(block_timestamp) = '2022-09-15'
10     LIMIT 200
11 """
12
13 query_job = client.query(query)
14 results = query_job.result()
15 df = pd.DataFrame([dict(row) for row in results])
16 df.to_csv("ethereum_transactions_2022_09_15.csv", index=False)
17 print("Saved to ethereum_transactions_2022_09_15.csv")
```

```
1  hash,nonce,transaction_index,from_address,to_address,value,gas,gas_price,input,
   receipt_cumulative_gas_used,receipt_gas_used,receipt_contract_address,receipt_root,receipt_status,
   block_timestamp,block_number,block_hash,max_fee_per_gas,max_priority_fee_per_gas,transaction_type,
   receipt_effective_gas_price,max_fee_per_blob_gas,blob_versioned_hashes,receipt_blob_gas_price,
   receipt_blob_gas_used
2  0x7de22d3acf345e8770806b1be128e286ac3d6f758d69116d97c29580d178513f,51,13,
   0xf9b11f0b190c30119a2b9a44757b72eb2069e03f,0xfa52274dd61e1643d2205169732f29114bc240b3,
   69820605014025365,38000,11435376408,0xf7654176,2686133,34573,,,1,2022-09-15 00:20:11+00:00,
   15535872,0x1fedccb9c460a5ea9e4ecd93fd6d6fc69dad60fe7f675369fc705c744ae78f96,12234008305.0,
   5000000000.0,2,11435376408,,[],,
3  0xe7ff37d421d29797786eae30f13701d463c57ca32199393173b249840248cb4e,1255,43,
   0x6a5a13d97212474af3e20d5be4dafbc5a84dabd9,0x32176455bea5fd4fd305efbd55a45bf2f8e92380,0,685802,
   8695376408,0xee28b4a00000000000000000000000000000000000000000000000000000000000000003,10554974,
   529835,,,1,2022-09-15 00:20:11+00:00,15535872,
   0x1fedccb9c460a5ea9e4ecd93fd6d6fc69dad60fe7f675369fc705c744ae78f96,18720000000.0,2260000000.0,2,
   8695376408,,[],,
4  0xfa6229d4f91f5dcd8347f942fe1d4ea334d3355e65abebd64a180c335c0f571b,154,34,
   0x1398f87f6cc2c64ffdcff962cbfdf75bf99dd13f,0x32176455bea5fd4fd305efbd55a45bf2f8e92380,0,635802,
```

hash	nonce	transaction	from_addre	to_address	value	gas	gas_price	input	receipt_cum	receipt_gas	receipt_con	receipt_roo	receipt_stat	block_times	block_numl	block_hash
0x7de22d3a	51	13	0xf9b11f0b1	0xfa52274d	#######	38000	#######	0xf7654176	2686133	34573			1	2022-09-15 (	15535872	0x1fedccb9c
0xe7ff37d42	1255	43	0x6a5a13d9	0x32176455l	0	685802	8695376408	0xee28b4a0(	10554974	529835			1	2022-09-15 (	15535872	0x1fedccb9c
0xfa6229d4f	154	34	0x1398f87f6	0x32176455l	0	635802	9435376408	0xee28b4a0(	8449196	529835			1	2022-09-15 (	15535872	0x1fedccb9c
0x099290f63	868	42	0xb31fa805	0x32176455l	0	685802	8695376408	0xee28b4a0(	10025139	529835			1	2022-09-15 (	15535872	0x1fedccb9c
0xfbfb5968f	61	67	0x2bb98d23	0x32176455l	0	635802	8535376408	0xee28b4a0(	13755180	529835			1	2022-09-15 (	15535872	0x1fedccb9c
0xb3f1eb6e8	186	65	0x74dc8d62	0x32176455l	0	635802	8535376408	0xee28b4a0(	12695510	529835			1	2022-09-15 (	15535872	0x1fedccb9c
0xf1c1d99b(	51	39	0xa666817a	0xb4afbd84	0	240000	8935376408	0x328beaff(	9378879	41418			0	2022-09-15 (	15535872	0x1fedccb9c
0x6c7def98f	51	14	0x8b52539f	0xd5c7c800	8.5E+17	150927	1.086E+10	0x9fb17e34(	1625598	100618			1	2022-09-15 (	15536128	0x4eb913e9
0x9e174fa4f	975	25	0xa6a97c1e	0xdef1c0dec	0	147255	1.564E+10	0xaa77476c	1555057	132447			1	2022-09-15 (	15536384	0xd525e734
0x101c8f218	918	28	0x2043c220	0xb5c9985d	1	255764	#######	0x25046071(	1913913	219890			1	2022-09-15 (	15536384	0xd525e734
0xf421fc1d1	259	49	0xc2b6a4b6	0x8d96b4ab	0	77460	#######	0xa9059cbb	3454353	51640			1	2022-09-15 (	15536384	0xd525e734
0x80e952086	691	22	0x4d470e0a	0x003f1083c	0	939730	#######	0x51e3fc174	1250634	555818			1	2022-09-15 (	15536384	0xd525e734
0x070f895b9	1953	30	0x27c5932e	0x00000000	0	47029	#######	0x5b34b966	2030425	47029			1	2022-09-15 (	15536384	0xd525e734
0xa88391feb	131	45	0xc8d6753e	0xa605ee5d	0	46567	#######	0xa22cb465	3262674	46567			1	2022-09-15 (	15536384	0xd525e734
0x690cad71	7	14	0x674babb8	0x11111112	#######	156832	#######	0x2e95b6c8(	1508683	96162			1	2022-09-15 (	15536640	0x8a381a0c
0xb4b17aaf	294	30	0x1750faa0	0xf4dd946d	0	100328	1.61E+10	0x1249c58b	4177161	69483			1	2022-09-15 (	15536640	0x8a381a0c
0x82351e2b9	9534	15	0x9b1a1d93	0x11111112!	5E+16	289357	#######	0x7c025200(	1582862	109287			0	2022-09-15 (	15536896	0x2e40be299
0x0b478c07	26	13	0x54ef2505	0x03f772418	1.1E+17	179609	#######	0xb6f9de95(	1421957	151619			0	2022-09-15 (	15536896	0x2e40be299
0xbcbea245	352	216	0x7182aae3	0x283af0b2	0	46267	#######	0xf14fcbc82	24191439	46267			1	2022-09-15 (	15536896	0x2e40be299
0x414087ea	3296	178	0x33569c10	0x874dd3f2	0	127449	#######	0x42842e0e(	21109750	117849			1	2022-09-15 (	15536896	0x2e40be299
0xe8ae8860(	44299	2	0xd7e1236c	0x5f62593c7	0	259834	#######	0x0000ed13(	336803	159834			1	2022-09-15 (	15536896	0x2e40be299
0x40827a77	1	98	0xe5c1b687	0x881d4023	4E+18	312073	#######	0x5f575529(	13356190	195469			1	2022-09-15 (	15536896	0x2e40be299
0x48773eff8	5207	1	0x479bc006	0x57c1e0c2	0	296345	8532026095	0x000600ed	176969	155969			1	2022-09-15 (	15536896	0x2e40be299
0x3b839f257	24	119	0x39d46c9e	0x00000000	1E+16	182870	#######	0xfb0f3ee10	15018002	175099			1	2022-09-15 (	15536896	0x2e40be299
0x8b840365	195	48	0x2fc7bd88	0x5400e430(	0	69871	#######	0x095ea7b3	3117796	46581			1	2022-09-15 (	15536896	0x2e40be299
0xb672daef	281	89	0x4b4f2c82	0xc02aaa39	0	39643	#######	0x2e1a7d4d	12500848	30428			1	2022-09-15 (	15536896	0x2e40be299
0xd8ae102c	195	289	0xb9278388	0x68b34658	#######	169560	#######	0x5ae401dc	29574051	117336			1	2022-09-15 (	15536896	0x2e40be299
0x9b9f58507	1867	29	0x0039b625	0x57c1e0c2	0	200764	#######	0x000200ed	2597773	105687			1	2022-09-15 (	15536896	0x2e40be299
0x4cd908ae:	2063	92	0x64df7af69	0xc02aaa39	0	46000	#######	0x2e1a7d4d	12728028	31096			1	2022-09-15 (	15536896	0x2e40be299
0x4d7bfb5a	32	181	0x5ab75956	0xc3a923b8	0	318781	#######	0x0fbf0a93(	21639544	309305			1	2022-09-15 (	15536896	0x2e40be299
0xa9981244	316	265	0x6843d905	0xa0b86991	0	73224	#######	0xa9059cbb	27962712	43749			1	2022-09-15 (	15536896	0x2e40be299

以上仅展示BigQuery API访问方法，可将数据下载到本地用于模型训练。由于数据规模较大，实际进行数据的分析处理时在云上进行（BigQuery Studio），相关代码已在论文中呈现。