The purpose of this document is to provide the developer and end users with information on using the Treasure Data Query Execution utility.

This utility is written in Python and enables the user to execute a query on a data set available in Treasure Data.

--------------------------------------------------------

The utility can be executed with Python in the below format:

<User Python executable directory>python TDQueryExecute.py

--db\_name <Treasure Data DB Name>

--table\_name <Treasure Data Table Name>

--col\_list <Comma separated list of columns. Is optional and will return all columns if omitted.>

--min\_time <Unix/Epoch timestamp or YYYY-MM-DD. Is optional.>

--max\_time < Unix/Epoch timestamp or YYYY-MM-DD. Is optional>

--query\_engine <Can be specified as hive or presto. Is optional. Defaults to Hive.>

--output\_format <Can be specified as csv or tabular. Is optional. Defaults to tabular>

--query\_limit <An integer limit for displaying the number of rows in the result. Is optional>

--------------------------------------------------------

For the developer/end user, to be able to run the utility locally, three key steps need to be performed:

1. Provide the API Key within the python program code where the comment “# Define API Key” is specified.
2. Specify the path and name for the csv output file where the comment “# Set CSV file path and location for csv output” is specified.
3. Ensure to install additional libraries used by the program by executing commands as follows:
4. ! pip install tdclient
5. ! pip install argparse
6. ! pip install csv
7. ! pip install urllib3
8. ! pip install prettytable

----------------------------------------------------------

A few samples of the query execution and output are attached below:

**With column list provided:**

C:\ProgramData\Anaconda3>python C:\Users\spanicke\TreasureData\TDQueryExecute.py --db\_name sid\_td\_test --table\_name party --col\_list partykey,lastupdatedate --min\_time NULL --max\_time NULL --query\_engine hive --output\_format tabular --query\_limit 100

The query you submitted is:

SELECT partykey,lastupdatedate FROM party LIMIT 100

+-----------+----------------+

| partykey | lastupdatedate |

+-----------+----------------+

| 196077719 | 4/19/2013 |

| 755374620 | 4/19/2013 |

| 755374622 | 4/19/2013 |

| 296768785 | 4/19/2013 |

| 296768786 | 4/19/2013 |

| 296768787 | 4/19/2013 |

| 197220118 | 4/19/2013 |

| 196588353 | 4/19/2013 |

| 318210263 | 4/19/2013 |

| 195301988 | 4/19/2013 |

| 197220117 | 4/19/2013 |

| 742966796 | 4/19/2013 |

| 195827349 | 4/19/2013 |

| 742966793 | 4/19/2013 |

| 196787254 | 4/19/2013 |

| 195854423 | 4/19/2013 |

| 195854422 | 4/19/2013 |

| 296842538 | 4/19/2013 |

+-----------+----------------+

**With no column list provided:**

C:\ProgramData\Anaconda3>python C:\Users\spanicke\TreasureData\TDQueryExecute.py --db\_name sid\_td\_test --table\_name party --min\_time NULL --max\_time NULL --query\_engine hive --output\_format tabular --query\_limit 100

The query you submitted is:

SELECT \* FROM party LIMIT 100

+-----------+----------------+-----------+--------------------------+---------------+------------+

| partykey | lastupdatedate | govtid | fullname | partytypecode | time |

+-----------+----------------+-----------+--------------------------+---------------+------------+

| 196077719 | 4/19/2013 | 553644627 | Goodman,Karen L Stratman | 1 | 1510539438 |

| 755374620 | 4/19/2013 | 326701852 | Jaye,Hollie | 1 | 1510539438 |

| 755374622 | 4/19/2013 | 326701852 | Jaye,Hollie | 1 | 1510539438 |

| 296768785 | 4/19/2013 | 170309216 | Kenny,Doris Marlene | 1 | 1510539438 |

| 296768786 | 4/19/2013 | 522047991 | Kenny,Kay Lynn | 1 | 1510539438 |

| 296768787 | 4/19/2013 | 462980198 | Kenny,Linda | 1 | 1510539438 |

| 197220118 | 4/19/2013 | None | Good,Jackie L. | 1 | 1510539438 |

| 196588353 | 4/19/2013 | 503962577 | Good,Jacob M | 1 | 1510539438 |

| 318210263 | 4/19/2013 | None | Good,Jacob M | 1 | 1510539438 |

| 195301988 | 4/19/2013 | 151724790 | Good,James | 1 | 1510539438 |

| 197220117 | 4/19/2013 | None | Good,Jannie R. | 1 | 1510539438 |

| 742966796 | 4/19/2013 | 147581133 | Goodwyn,Percell,Jr | 1 | 1510539438 |

| 195827349 | 4/19/2013 | 229981401 | Goodwyn,Reanita P | 1 | 1510539438 |

| 742966793 | 4/19/2013 | 150546766 | Goodwyn,Tanya S | 1 | 1510539438 |

| 196787254 | 4/19/2013 | None | Searsy,Mary L | 1 | 1510539438 |

| 195854423 | 4/19/2013 | 498846967 | Searwar,Ivan | 1 | 1510539438 |

| 195854422 | 4/19/2013 | 499842100 | Searwar,Olivia | 1 | 1510539438 |

| 296842538 | 4/19/2013 | 505560706 | Setlak,John Joseph | 1 | 1510539438 |

+-----------+----------------+-----------+--------------------------+---------------+------------+

**With non-existent table name:**

C:\ProgramData\Anaconda3>python C:\Users\spanicke\TreasureData\TDQueryExecute.py --db\_name sid\_td\_test --table\_name party1 --min\_time NULL --max\_time NULL --query\_engine hive --output\_format tabular --query\_limit 100

The query you submitted is:

SELECT \* FROM party1 LIMIT 100

Table does not exist. Check table name.

Query did not return any data.

**With non-existent database name:**

C:\ProgramData\Anaconda3>python C:\Users\spanicke\TreasureData\TDQueryExecute.py --db\_name sid\_td\_test1 --table\_name party1 --min\_time NULL --max\_time NULL --query\_engine hive --output\_format tabular --query\_limit 100

The query you submitted is:

SELECT \* FROM party1 LIMIT 100

Database does not exist. Check database name.

Query did not return any data.

**With CSV output format(CSV file attached within document below):**

C:\ProgramData\Anaconda3>python C:\Users\spanicke\TreasureData\TDQueryExecute.py --db\_name sid\_td\_test --table\_name party --min\_time NULL --max\_time NULL --query\_engine hive --output\_format csv --query\_limit 100

The query you submitted is:

SELECT \* FROM party LIMIT 100

CSV output format selected

Writing output to C:\ProgramData\Anaconda3\sampleoutput.csv

