

SQL Queries

The `pandas.io.sql` module provides a collection of query wrappers to both facilitate data retrieval and to reduce dependency on DB-specific API. Database abstraction is provided by SQLAlchemy if installed. In addition you will need a driver library for your database. Examples of such drivers are [psycopg2](#) for PostgreSQL or [pymysql](#) for MySQL. For [SQLite](#) this is included in Python's standard library by default. You can find an overview of supported drivers for each SQL dialect in the [SQLAlchemy docs](#).

If SQLAlchemy is not installed, a fallback is only provided for sqlite (and for mysql for backwards compatibility, but this is deprecated and will be removed in a future version). This mode requires a Python database adapter which respect the [Python DB-API](#).

See also some [cookbook examples](#) for some advanced strategies.

The key functions are:

<code>read_sql_table(table_name, con[, schema, ...])</code>	Read SQL database table into a DataFrame.
<code>read_sql_query(sql, con[, index_col, ...])</code>	Read SQL query into a DataFrame.
<code>read_sql(sql, con[, index_col, ...])</code>	Read SQL query or database table into a DataFrame.
<code>DataFrame.to_sql(name, con[, flavor, ...])</code>	Write records stored in a DataFrame to a SQL database.

Note The function `read_sql()` is a convenience wrapper around `read_sql_table()` and `read_sql_query()` (and for backward compatibility) and will delegate to specific function depending on the provided input (database table name or sql query). Table names do not need to be quoted if they have special characters.

In the following example, we use the [SQLite](#) SQL database engine. You can use a temporary SQLite database where data are stored in “memory”.

To connect with SQLAlchemy you use the `create_engine()` function to create an engine object from database URI. You only need to create the engine once per database you are connecting to. For more information on `create_engine()` and the URI formatting, see the examples below and the SQLAlchemy [documentation](#)

```
In [516]: from sqlalchemy import create_engine

# Create your engine.
In [517]: engine = create_engine('sqlite:///memory:')
```

If you want to manage your own connections you can pass one of those instead:

```
with engine.connect() as conn, conn.begin():
    data = pd.read_sql_table('data', conn)
```

Writing DataFrames

Assuming the following data is in a DataFrame `data`, we can insert it into the database using `to_sql()`.

id	Date	Col_1	Col_2	Col_3
26	2012-10-18	X	25.7	True
42	2012-10-19	Y	-12.4	False
63	2012-10-20	Z	5.73	True

```
In [518]: data.to_sql('data', engine)
```

With some databases, writing large DataFrames can result in errors due to packet size limitations being exceeded. This can be avoided by setting the `chunksize` parameter when calling `to_sql`. For example, the following writes data to the database in batches of 1000 rows at a time:

```
In [519]: data.to_sql('data_chunked', engine, chunksize=1000)
```

SQL data types

`to_sql()` will try to map your data to an appropriate SQL data type based on the dtype of the data. When you have columns of dtype `object`, pandas will try to infer the data type.

You can always override the default type by specifying the desired SQL type of any of the columns by using the `dtype` argument. This argument needs a dictionary mapping column names to SQLAlchemy types (or strings for the sqlite3 fallback mode). For example, specifying to use the sqlalchemy `string` type instead of the default `Text` type for string columns:

```
In [520]: from sqlalchemy.types import String
```

```
In [521]: data.to_sql('data_dtype', engine, dtype={'Col_1': String})
```

Note Due to the limited support for `timedelta`'s in the different database flavors, columns with type `timedelta64` will be written as integer values as nanoseconds to the database and a warning will be raised.

Note Columns of `category` dtype will be converted to the dense representation as you would get with `np.asarray(categorical)` (e.g. for string categories this gives an array of strings). Because of this, reading the database table back in does **not** generate a categorical.

Reading Tables

`read_sql_table()` will read a database table given the table name and optionally a subset of columns to read.

Note In order to use `read_sql_table()`, you **must** have the SQLAlchemy optional dependency installed.

```
In [522]: pd.read_sql_table('data', engine)
Out[522]:
```

	index	id	Date	Col_1	Col_2	Col_3
0	0	26	2010-10-18	X	27.50	True
1	1	42	2010-10-19	Y	-12.50	False
2	2	63	2010-10-20	Z	5.73	True

You can also specify the name of the column as the DataFrame index, and specify a subset of columns to be read.

```
In [523]: pd.read_sql_table('data', engine, index_col='id')
Out[523]:
```

	index	Date	Col_1	Col_2	Col_3
id					
26	0	2010-10-18	X	27.50	True
42	1	2010-10-19	Y	-12.50	False
63	2	2010-10-20	Z	5.73	True

```
In [524]: pd.read_sql_table('data', engine, columns=['Col_1', 'Col_2'])
Out[524]:
```

	Col_1	Col_2
0	X	27.50
1	Y	-12.50
2	Z	5.73

And you can explicitly force columns to be parsed as dates:

```
In [525]: pd.read_sql_table('data', engine, parse_dates=['Date'])
Out[525]:
```

	index	id	Date	Col_1	Col_2	Col_3
0	0	26	2010-10-18	X	27.50	True
1	1	42	2010-10-19	Y	-12.50	False
2	2	63	2010-10-20	Z	5.73	True

If needed you can explicitly specify a format string, or a dict of arguments to pass to `pandas.to_datetime()`:

```
pd.read_sql_table('data', engine, parse_dates={'Date': '%Y-%m-%d'})
pd.read_sql_table('data', engine, parse_dates={'Date': {'format': '%Y-%m-%d %H:%M:%S'}}
```

You can check if a table exists using `has_table()`

Schema support

Reading from and writing to different schema's is supported through the `schema` keyword in the `read_sql_table()` and `to_sql()` functions. Note however that this depends on the database flavor (sqlite does not have schema's). For example:

```
df.to_sql('table', engine, schema='other_schema')
pd.read_sql_table('table', engine, schema='other_schema')
```

Querying

You can query using raw SQL in the `read_sql_query()` function. In this case you must use the SQL variant appropriate for your database. When using SQLAlchemy, you can also pass SQLAlchemy Expression language constructs, which are database-agnostic.

```
In [526]: pd.read_sql_query('SELECT * FROM data', engine)
Out[526]:
```

	index	id	Date	Col_1	Col_2	Col_3
0	0	26	2010-10-18 00:00:00.000000	X	27.50	1
1	1	42	2010-10-19 00:00:00.000000	Y	-12.50	0
2	2	63	2010-10-20 00:00:00.000000	Z	5.73	1

Of course, you can specify a more “complex” query.

```
In [527]: pd.read_sql_query("SELECT id, Col_1, Col_2 FROM data WHERE id = 42;", eng
Out[527]:
```

	id	Col_1	Col_2
0	42	Y	-12.5

The `read_sql_query()` function supports a `chunksize` argument. Specifying this will return an iterator through chunks of the query result:

```
In [528]: df = pd.DataFrame(np.random.randn(20, 3), columns=list('abc'))
In [529]: df.to_sql('data_chunks', engine, index=False)
```

```
In [530]: for chunk in pd.read_sql_query("SELECT * FROM data_chunks", engine, chunk
.....:     print(chunk)
.....:
```

	a	b	c
0	0.280665	-0.073113	1.160339
1	0.369493	1.904659	1.111057
2	0.659050	-1.627438	0.602319

```

3  0.420282  0.810952  1.044442
4 -0.400878  0.824006 -0.562305
      a      b      c
0  1.954878 -1.331952 -1.760689
1 -1.650721 -0.890556 -1.119115
2  1.956079 -0.326499 -1.342676
3  1.114383 -0.586524 -1.236853
4  0.875839  0.623362 -0.434957
      a      b      c
0  1.407540  0.129102  1.616950
1  0.502741  1.558806  0.109403
2 -1.219744  2.449369 -0.545774
3 -0.198838 -0.700399 -0.203394
4  0.242669  0.201830  0.661020
      a      b      c
0  1.792158 -0.120465 -1.233121
1 -1.182318 -0.665755 -1.674196
2  0.825030 -0.498214 -0.310985
3 -0.001891 -1.396620 -0.861316
4  0.674712  0.618539 -0.443172

```

You can also run a plain query without creating a dataframe with `execute()`. This is useful for queries that don't return values, such as INSERT. This is functionally equivalent to calling `execute` on the SQLAlchemy engine or db connection object. Again, you must use the SQL syntax variant appropriate for your database.

```

from pandas.io import sql
sql.execute('SELECT * FROM table_name', engine)
sql.execute('INSERT INTO table_name VALUES(?, ?, ?)', engine, params=[('id', 1, 12.3456789)])

```

Engine connection examples

To connect with SQLAlchemy you use the `create_engine()` function to create an engine object from database URI. You only need to create the engine once per database you are connecting to.

```

from sqlalchemy import create_engine

engine = create_engine('postgresql://scott:tiger@localhost:5432/mydatabase')

engine = create_engine('mysql+mysqldb://scott:tiger@localhost/foo')

engine = create_engine('oracle://scott:tiger@127.0.0.1:1521/sidname')

engine = create_engine('mssql+pyodbc://mydsn')

# sqlite://<nohostname>/<path>
# where <path> is relative:
engine = create_engine('sqlite:///foo.db')

# or absolute, starting with a slash:
engine = create_engine('sqlite:///absolute/path/to/foo.db')

```

For more information see the examples the SQLAlchemy [documentation](#)

Advanced SQLAlchemy queries

You can use SQLAlchemy constructs to describe your query.

Use `sqlalchemy.text()` to specify query parameters in a backend-neutral way

```
In [531]: import sqlalchemy as sa

In [532]: pd.read_sql(sa.text('SELECT * FROM data where Col_1=:col1'), engine, para
Out[532]:
```

	index	id	Date	Col_1	Col_2	Col_3
0	0	26	2010-10-18 00:00:00.000000	X	27.5	1

If you have an SQLAlchemy description of your database you can express where conditions using SQLAlchemy expressions

```
In [533]: metadata = sa.MetaData()

In [534]: data_table = sa.Table('data', metadata,
.....:     sa.Column('index', sa.Integer),
.....:     sa.Column('Date', sa.DateTime),
.....:     sa.Column('Col_1', sa.String),
.....:     sa.Column('Col_2', sa.Float),
.....:     sa.Column('Col_3', sa.Boolean),
.....: )
.....:

In [535]: pd.read_sql(sa.select([data_table]).where(data_table.c.Col_3 == True), en
Out[535]:
```

	index	Date	Col_1	Col_2	Col_3
0	0	2010-10-18	X	27.50	True
1	2	2010-10-20	Z	5.73	True

You can combine SQLAlchemy expressions with parameters passed

to `read_sql()` using `sqlalchemy.bindparam()`

```
In [536]: import datetime as dt

In [537]: expr = sa.select([data_table]).where(data_table.c.Date > sa.bindparam('da

In [538]: pd.read_sql(expr, engine, params={'date': dt.datetime(2010, 10, 18)})
Out[538]:
```

	index	Date	Col_1	Col_2	Col_3
0	1	2010-10-19	Y	-12.50	False
1	2	2010-10-20	Z	5.73	True

Sqlite fallback

The use of sqlite is supported without using SQLAlchemy. This mode requires a Python database adapter which respect the [Python DB-API](#).

You can create connections like so:

```
import sqlite3
con = sqlite3.connect(':memory:')
```

And then issue the following queries:

```
data.to_sql('data', cnx)
pd.read_sql_query("SELECT * FROM data", con)
```