

rdbms

March 28, 2021

```
[1]: from pathlib import Path
import os
import sqlite3

import s3fs
import pandas as pd

current_dir = Path(os.getcwd()).absolute()
results_dir = current_dir.joinpath('results')
kv_data_dir = results_dir.joinpath('kvdb')
kv_data_dir.mkdir(parents=True, exist_ok=True)

def read_cluster_csv(file_path, endpoint_url='https://storage.budsc.
↳midwest-datascience.com'):
    s3 = s3fs.S3FileSystem(
        anon=True,
        client_kwargs={
            'endpoint_url': endpoint_url
        }
    )
    return pd.read_csv(s3.open(file_path, mode='rb'))
```

```
[2]: def create_measurements_table(conn):
    sql = """
    CREATE TABLE IF NOT EXISTS measurements (
        visit_id integer NOT NULL,
        person_id text NOT NULL,
        quantity text,
        reading real,
        FOREIGN KEY (visit_id) REFERENCES visits (visit_id),
        FOREIGN KEY (person_id) REFERENCES people (people_id)
    );
    """

    c = conn.cursor()
    c.execute(sql)
```

```
def load_measurements_table(conn):
    create_measurements_table(conn)
    df = read_cluster_csv('data/external/tidynomicon/measurements.csv')
    measurements = df.values
    c = conn.cursor()
    c.execute('DELETE FROM measurements;') # Delete data if exists
    c.executemany('INSERT INTO measurements VALUES (?, ?, ?, ?)', measurements)
```

```
[3]: def create_people_table(conn):
    sql = """
    CREATE TABLE IF NOT EXISTS people (
        person_id integer NOT NULL,
        personal_name text NOT NULL,
        family_name text,
        FOREIGN KEY (person_id) REFERENCES people (people_id)
    );
    """

    c = conn.cursor()
    c.execute(sql)

    def load_people_table(conn):
        create_people_table(conn)
        df = read_cluster_csv('data/external/tidynomicon/person.csv')
        people = df.values
        c = conn.cursor()
        c.execute('DELETE FROM people;') # Delete data if exists
        c.executemany('INSERT INTO people VALUES (?, ?, ?)', people)
```

```
[4]: def create_sites_table(conn):
    sql = """
    CREATE TABLE IF NOT EXISTS sites (
        site_id text PRIMARY KEY,
        latitude double NOT NULL,
        longitude double NOT NULL
    );
    """

    c = conn.cursor()
    c.execute(sql)

    def load_sites_table(conn):
        create_sites_table(conn)
        df = read_cluster_csv('data/external/tidynomicon/site.csv')
        sites = df.values
```

```

c = conn.cursor()
c.execute('DELETE FROM sites;') # Delete data if exists
c.executemany('INSERT INTO sites VALUES (?,?),' ,sites)

```

```

[5]: def create_visits_table(conn):
    sql = """
    CREATE TABLE IF NOT EXISTS visits (
        visit_id integer PRIMARY KEY,
        site_id text NOT NULL,
        visit_date text,
        FOREIGN KEY (site_id) REFERENCES sites (site_id)
    );
    """

    c = conn.cursor()
    c.execute(sql)

    def load_visits_table(conn):
        create_visits_table(conn)
        df = read_cluster_csv('data/external/tidynomicon/visited.csv')
        visits = df.values
        c = conn.cursor()
        c.execute('DELETE FROM visits;') # Delete data if exists
        c.executemany('INSERT INTO visits VALUES (?,?),' ,visits)

```

```

[9]: db_path = results_dir.joinpath('patient-info.db')
conn = sqlite3.connect(str(db_path))
# TODO: Uncomment once functions completed
load_people_table(conn)
load_sites_table(conn)
load_visits_table(conn)
load_measurements_table(conn)

conn.commit()

df1 = pd.read_sql_query("SELECT * from people", conn)
print(df1.head())

df2 = pd.read_sql_query("SELECT * from sites", conn)
print(df2.head())

df3 = pd.read_sql_query("SELECT * from visits", conn)
print(df3.head())

df4 = pd.read_sql_query("SELECT * from measurements ", conn)
print(df4.head())

```

```
conn.close()
```

	person_id	personal_name	family_name
0	dyer	William	Dyer
1	pb	Frank	Pabodie
2	lake	Anderson	Lake
3	roe	Valentina	Roerich
4	danforth	Frank	Danforth

	site_id	latitude	longitude
0	DR-1	-49.85	-128.57
1	DR-3	-47.15	-126.72
2	MSK-4	-48.87	-123.40

	visit_id	site_id	visit_date
0	619	DR-1	1927-02-08
1	622	DR-1	1927-02-10
2	734	DR-3	1930-01-07
3	735	DR-3	1930-01-12
4	751	DR-3	1930-02-26

	visit_id	person_id	quantity	reading
0	619	dyer	rad	9.82
1	619	dyer	sal	0.13
2	622	dyer	rad	7.80
3	622	dyer	sal	0.09
4	734	pb	rad	8.41