

Document Database Coursework

Task Description

In this course work you will write several MongoDB queries for an example dataset. You can test your queries on a test server available at cloud-vm-41-04.doc.ic.ac.uk with a database called ocean (user name *h412* and password *h412*). There are multiple ways to test the queries. You can install MongoDB local and use the mongo console, you can use it programmatically with whatever language you prefer and, finally, you can also use a dedicated client application (Robomongo - <https://robomongo.org/> - is recommended and available for most platforms – or TablePlus - <https://www.tableplus.io>).

An SQL/Postgres database with the same data but shredded to fit a relational database is loaded on the same server. The database name is ocean and user name is *h412* and the password is *412h*.

You should turn in your solution electronically (text document, PDF or similar) via CATe.

Data Description

The dataset used is based on real oceanography data from NOAA with two basic collections, *stations* and *oceandata*. The former contains all station information and the latter all measurements from stations. Stations can be from different vendors and the structure of the documents can consequently differ. See the examples below for the two document types.

```
{
  "_id" : ObjectId("53e4fcc42239c23dce3cb7bc"),
  "station_id" : 8461490,
  "loc" : {
    "type" : "Point",
    "coordinates" : [
      -72.09,
      41.3614
    ]
  },
  "name" : "New London",
  "products" : [
    {
      "v" : 69.4,
      "t" : ISODate("2014-08-08T16:24:00Z"),
      "name" : "water_temperature",
      "f" : "0,0,0"
    },
    {
      "v" : 77,
      "t" : ISODate("2014-08-08T16:24:00Z"),
      "name" : "air_temperature",
      "f" : "0,0,0"
    },
    {
      "d" : "360.00",
      "g" : "8.75",
      "f" : "0,0,0",
      "s" : "4.08",
      "t" : ISODate("2014-08-08T16:24:00Z"),
      "dr" : "N",
      "name" : "wind"
    },
    {
      "v" : 1015.8,
      "t" : ISODate("2014-08-08T16:24:00Z"),
      "name" : "air_pressure",
      "f" : "0,0,0"
    }
  ],
  "fetch_date" : ISODate("2014-08-08T16:37:22.640Z"),
  "id" : 8461490
}
```

Figure 1 Example oceandata document, i.e., one measurement.

```

{
  "_id" : ObjectId("53744abad3a11771ea04bac4"),
  "-name" : "Nawiliwili",
  "-ID" : 1611400,
  "metadata" : {
    "location" : {
      "lat" : 21.9544,
      "long" : -159.3561,
      "state" : "HI"
    },
    "date_established" : ISODate("1954-11-31T09:20:00Z")
  },
  "parameter" : [
    {
      "-name" : "Water Level",
      "-sensorID" : "A1",
      "-DCP" : "1",
      "-status" : "1"
    },
    {
      "-name" : "Water Temp",
      "-sensorID" : "E1",
      "-DCP" : "1",
      "-status" : "1"
    }
  ]
}

```

Figure 2 Example station document.

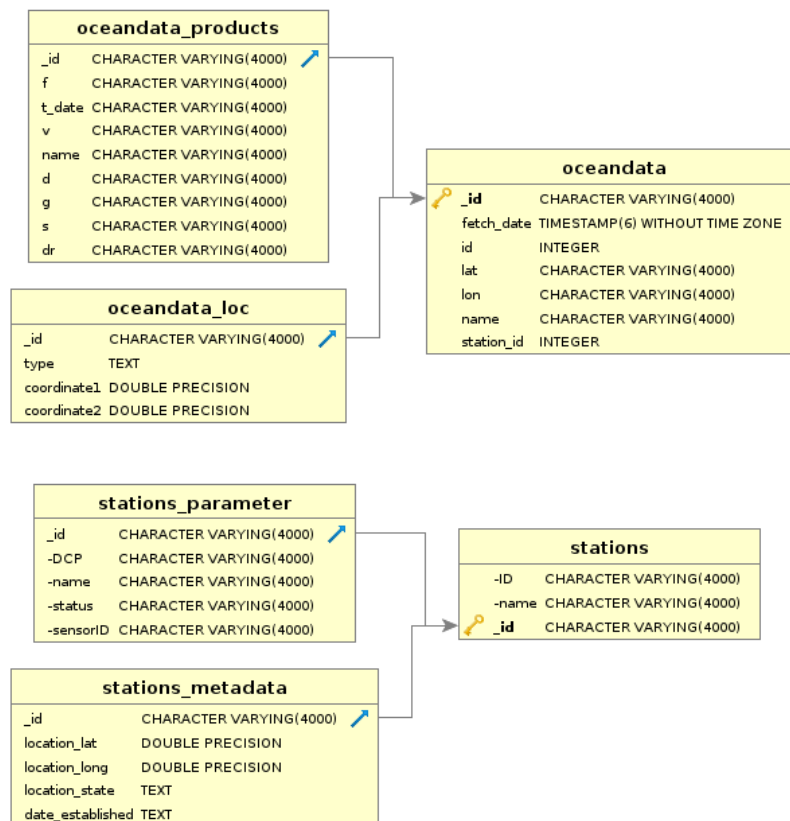


Figure 3 Relational schema

Queries

Write MongoDB queries for the questions below. Also write queries 1, 5, 6, 8, 10 in SQL given the relational schema defined before.

- 1) Find all stations with the name "Kawaihae".
- 2) Find all stations that measure air temperature and water temperature.
- 3) Find all stations that do not measure the air temperature.
- 4) Find all measurement of the air temperature reported not using a double as data type.

- 5) Find all stations and list only their name and state. Sort by name ascending.
- 6) Count the stations by state.
- 7) Get average values of measurements grouped by year where the product name is "water_temperature".
- 8) Find the nearest station to the point -117.1572600, 32.7153300.
- 9) Calculate the average water temperature of the five nearest stations to point -117.1572600, 32.7153300.
- 10) Find all measurements of stations in the great state of Hawaii.