

```
In [31]: # import package(s) and create connection below

from pymongo import MongoClient
from bson.json_util import dumps

client = MongoClient("[uri string]")
swoosh = client.swoosh
```

```
In [32]: # Use Python comment to answer Q1 below
# Q1A
# The data type of the name is string.

# Q1B
# The data type of isStudent is boolean.

# Q1C
# The type of start_at is string.

# Q1D
# The type of topic is an array or a string.

# Q1E
# The minimum array length is 1, the average is 2.01, and the maximum is 3.
```

```
In [33]: # Q2A below

# SELECT *
# FROM users
# WHERE isStudent = true AND user_id = 4

cursor = swoosh.users.find( {"$and": [{"user_id": "4"}, {"isstudent": True}]} , {} )
print(dumps(cursor, indent = 2))
```

```
[
  {
    "_id": {
      "$oid": "6271b52f5f2b9f0eefdf7b73"
    },
    "user_id": "4",
    "email": "ryanhoward@penn state.edu",
    "name": {
      "first_name": "Daniel",
      "last_name": "Duarte"
    },
    "occupation": "Building services engineer",
    "isinstructor": false,
    "isstudent": true
  }
]
```

```
In [34]: # Q2B below

cursor = swoosh.posts.find( {"$and": [{"topics": "homework2" },
                                     {"created_at": { "$gt": "2020-07-01 00:00" } } ] }).sort("created_at", -1 ).limit(3)
print(dumps(cursor, indent = 2))
```

```
[
  {
    "_id": {
      "$oid": "6271b5055f2b9f0eefdf761c"
    },
    "body": "Always seven dream happy first message. Instead month great consider view. Politics but beat once. Six s
ing beyond lay.",
    "created_at": "2020-12-23 9:15:40",
    "replied_to_post_id": "293",
    "topics": [
      "finalexam",
      "homework2"
    ],
    "post_id": "452",
    "user_id": "169",
    "meeting_id": "825",
    "post_type": "note"
  },
  {
    "_id": {
      "$oid": "6271b5055f2b9f0eefdf772b"
    },
    "body": "Always seven dream happy first message. Instead month great consider view. Politics but beat once. Six s
ing beyond lay.",
    "created_at": "2020-12-23 9:15:40",
    "replied_to_post_id": "293",
    "topics": [
      "finalexam",
      "homework2"
    ],
    "post_id": "452",
    "user_id": "169",
    "meeting_id": "825",
    "post_type": "note"
  }
]
```

```

    },
    "body": "Later bag attorney offer table management machine. Rate food song agreement find focus life. Majority ca
re soldier since rule it. Else effect police put.",
    "created_at": "2020-12-23 8:08:08",
    "replied_to_post_id": "465",
    "topics": [
        "module4",
        "project3",
        "homework2"
    ],
    "post_id": "723",
    "user_id": "183",
    "meeting_id": "825",
    "post_type": "question"
},
{
    "_id": {
        "$oid": "6271b5055f2b9f0eefdf7730"
    },
    "body": "Benefit group arrive. Travel power bill stop tax. Size chair respond including PM stock. Nothing discuss
ion cell share brother. Age else civil its structure. Room time himself game close scene all.",
    "created_at": "2020-12-23 5:50:21",
    "replied_to_post_id": "30",
    "topics": [
        "finalproject",
        "module2",
        "homework2"
    ],
    "post_id": "728",
    "user_id": "141",
    "meeting_id": "825",
    "post_type": "note"
}
]

```

```

In [35]: # Q2C below

swoosh.posts.update_many( {"post_type": ""}, { "$set": { "post_type": "miscellaneous" } } )
cursor = swoosh.posts.find( { "post_type": "miscellaneous" }, { "_id": 0, "post_id": 1, "post_type": 1 } )
print(dumps(cursor, indent = 2))

```

```

[
  {
    "post_id": "271",
    "post_type": "miscellaneous"
  },
  {
    "post_id": "272",
    "post_type": "miscellaneous"
  },
  {
    "post_id": "279",
    "post_type": "miscellaneous"
  }
]

```

```

In [36]: # Q2D below

cursor = swoosh.users.find( {"$and": [{ "isinstructor": True }, { "education.school": "UCI" },
                                     { "education.graduation_year": { "$gte": 2000 } } ]},
                             { "_id": 0, "user_id": 1, "first_name": "$name.first_name",
                               "last_name": "$name.last_name" } ).sort("name.last_name").limit(5)

print(dumps(cursor, indent = 2))

```

```

[
  {
    "user_id": "325",
    "first_name": "Tiffany",
    "last_name": "Bryant"
  },
  {
    "user_id": "426",
    "first_name": "Jeffrey",
    "last_name": "Burns"
  },
  {
    "user_id": "94",
    "first_name": "Brian",
    "last_name": "Conner"
  },
  {
    "user_id": "347",
    "first_name": "Lindsay",
    "last_name": "Cooke"
  },
  {
    "user_id": "180",
    "first_name": "Donna",
    "last_name": "Cox"
  }
]

```

```
}  
]
```

```
In [37]: # Q2E below  
  
cursor = swoosh.users.count_documents( { "$or": [ { "$and": [ { "isstudent": True },  
    { "education.graduation_year": { "$gt": 2010 } } ] },  
    { "$and": [ { "isinstructor": True },  
    { "education.graduation_year": { "$lt": 2000 } } ] } ] } )  
  
print(cursor)
```

87

```
In [38]: # Q2F below  
  
match_stage = { "$match": { "user_id": "355" } }  
count_stage = { "$count": "courses_count" }  
cursor = swoosh.enrolledin.aggregate( [ match_stage, count_stage ] )  
print(dumps(cursor, indent=2))
```

```
[  
  {  
    "courses_count": 6  
  }  
]
```

```
In [39]: # Q2G below  
  
group_stage = { "$group": { "_id": "user_id", "course_count": { "$sum": 1 } } }  
rank_stage = { "$setWindowFields": { "partitionBy": "user_id", "sortBy": { "course_count": -1 },  
    "output": { "rank": { "$denseRank": {} } } } }  
match_stage = { "$match": { "rank": { "$lte": 5 } } }  
project_stage = { "$project": { "_id": 0, "instructor_id": "$_id", "course_count": "$course_count" } }  
  
cursor = swoosh.teaches.aggregate( [ group_stage, rank_stage, match_stage, project_stage ] )  
print(dumps(cursor, indent=2))
```

```
[  
  {  
    "instructor_id": "486",  
    "course_count": 10  
  },  
  {  
    "instructor_id": "366",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "223",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "347",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "100",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "300",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "474",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "492",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "426",  
    "course_count": 6  
  },  
  {  
    "instructor_id": "195",  
    "course_count": 5  
  },  
  {  
    "instructor_id": "469",  
    "course_count": 5  
  },  
  {  
    "instructor_id": "440",  
    "course_count": 5  
  },  
  {  
    "instructor_id": "10",  
    "course_count": 5  
  },  
]
```

```
{
  "instructor_id": "175",
  "course_count": 5
},
{
  "instructor_id": "8",
  "course_count": 5
},
{
  "instructor_id": "40",
  "course_count": 5
},
{
  "instructor_id": "22",
  "course_count": 5
},
{
  "instructor_id": "160",
  "course_count": 5
},
{
  "instructor_id": "94",
  "course_count": 5
},
{
  "instructor_id": "37",
  "course_count": 5
},
{
  "instructor_id": "117",
  "course_count": 4
},
{
  "instructor_id": "142",
  "course_count": 4
},
{
  "instructor_id": "59",
  "course_count": 4
},
{
  "instructor_id": "187",
  "course_count": 4
},
{
  "instructor_id": "159",
  "course_count": 4
},
{
  "instructor_id": "267",
  "course_count": 4
},
{
  "instructor_id": "171",
  "course_count": 4
},
{
  "instructor_id": "408",
  "course_count": 4
},
{
  "instructor_id": "410",
  "course_count": 4
},
{
  "instructor_id": "358",
  "course_count": 3
},
{
  "instructor_id": "178",
  "course_count": 3
},
{
  "instructor_id": "242",
  "course_count": 3
},
{
  "instructor_id": "176",
  "course_count": 3
},
{
  "instructor_id": "212",
  "course_count": 3
},
{
  "instructor_id": "12",
  "course_count": 3
},
{
  "instructor_id": "228",
  "course_count": 3
}
```

```

    },
    {
      "instructor_id": "110",
      "course_count": 3
    },
    {
      "instructor_id": "464",
      "course_count": 3
    },
    {
      "instructor_id": "109",
      "course_count": 3
    },
    {
      "instructor_id": "82",
      "course_count": 3
    },
    {
      "instructor_id": "334",
      "course_count": 3
    },
    {
      "instructor_id": "182",
      "course_count": 3
    },
    {
      "instructor_id": "73",
      "course_count": 3
    },
    {
      "instructor_id": "335",
      "course_count": 3
    }
  ]

```

In [40]: *# Q2H below*

```

match_stage = {"$match": {"course_name": "French II"}}
lookup_stage = {"$lookup": {"from": "meetings", "localField": "course_id", "foreignField": "course_id",
                             "as": "meeting_list"}}
sort_stage = {"$sort": {"meeting_list.duration": 1}}
project_stage = {"$project": {"_id": 0, "meeting_list.meeting_id": 1, "meeting_list.meeting_name": 1,
                              "meeting_list.instructor_id": 1, "meeting_list.duration": 1}}
cursor = swoosh.courses.aggregate( [ match_stage, lookup_stage, sort_stage, project_stage ] )
print(dumps(cursor, indent=2))

```

```

[
  {
    "meeting_list": [
      {
        "meeting_id": "1108",
        "meeting_name": "French II1",
        "duration": 120,
        "instructor_id": "243"
      },
      {
        "meeting_id": "1109",
        "meeting_name": "French II2",
        "duration": 120,
        "instructor_id": "8"
      },
      {
        "meeting_id": "1110",
        "meeting_name": "French II3",
        "duration": 120,
        "instructor_id": "8"
      },
      {
        "meeting_id": "1111",
        "meeting_name": "French II4",
        "duration": 120,
        "instructor_id": "234"
      },
      {
        "meeting_id": "1112",
        "meeting_name": "French II5",
        "duration": 120,
        "instructor_id": "234"
      },
      {
        "meeting_id": "1113",
        "meeting_name": "French II6",
        "duration": 120,
        "instructor_id": "243"
      },
      {
        "meeting_id": "1114",
        "meeting_name": "French II7",
        "duration": 120,
        "instructor_id": "365"
      }
    ]
  }
]

```

```

    {
      "meeting_id": "1115",
      "meeting_name": "French II8",
      "duration": 120,
      "instructor_id": "8"
    },
    {
      "meeting_id": "1116",
      "meeting_name": "French II9",
      "duration": 120,
      "instructor_id": "234"
    },
    {
      "meeting_id": "1117",
      "meeting_name": "French II10",
      "duration": 120,
      "instructor_id": "8"
    },
    {
      "meeting_id": "1118",
      "meeting_name": "French II11",
      "duration": 120,
      "instructor_id": "234"
    }
  ]
}
]

```

```

In [41]: # Q2I below

lookup_stage = {"$lookup": {"from": "courses", "localField": "course_id", "foreignField": "course_id",
                             "as": "enrolled_course"}}

unwind_stage = {"$unwind": "$enrolled_course"}

project_stage = {"$project": {"_id": 0, "course_name": "$enrolled_course.course_name", "course_id": 1}}

group_stage = {"$group": {"_id": {"course_id": "$course_id", "course_name": "$course_name"},
                           "enrolled_count": {"$sum": 1}}}

match_stage = {"$match": {"enrolled_count": {"$gt": 5}}}

project_stage2 = {"$project": {"_id": 0, "course_name": "$_id.course_name", "course_id": "$_id.course_id",
                               "enrolled_count": "$enrolled_count"}}

sample_stage = {"$sample": {"size": 5}}

cursor = swoosh.enrolledin.aggregate( [ lookup_stage, unwind_stage, project_stage, group_stage, match_stage,
                                         project_stage2, sample_stage ] )

print(dumps(cursor, indent=2))

```

```

[
  {
    "course_name": "Neural Networks",
    "course_id": "73",
    "enrolled_count": 16
  },
  {
    "course_name": "Engineering 102",
    "course_id": "2",
    "enrolled_count": 13
  },
  {
    "course_name": "Literature 102",
    "course_id": "49",
    "enrolled_count": 19
  },
  {
    "course_name": "Communications 102",
    "course_id": "46",
    "enrolled_count": 17
  },
  {
    "course_name": "Calculus IV",
    "course_id": "89",
    "enrolled_count": 11
  }
]

```

```

In [42]: # Q2J below

...
SELECT p.user_id, COUNT(p.user_id)
FROM swoosh.post p JOIN swoosh.Thumbsup tu
ON p.user_id = tu.user_id
WHERE p.post_type = 'note'
AND p.created_at > '2020-06-21 00:00:00'
GROUP BY p.user_id
HAVING COUNT(p.user_id) > 5    <----- ???????
ORDER BY COUNT(p.user_id) DESC

```

```

LIMIT 5
'''

lookup_stage = {"$lookup": {"from": "thumbsup", "localField": "user_id", "foreignField": "user_id", "as": "result"}}
match_stage = {"$match": {"$and": [{"post_type": "note"}, {"created_at": {"$gt": '2020-06-21 00:00:00'}}]}}
group_stage = {"$group": {"_id": None, "user_count": {"$sum": 1}}}
match_stage2 = {"$match": {"user_count": {"$gt": 5}}}
sort_stage = {"$sort": {"user_count": 1}}
limit_stage = {"$limit": 5}
project_stage = {"$project": {"_id": 0, "user_id": "$_id", "user_count": "$user_count"}}

cursor = swoosh.posts.aggregate( [ lookup_stage, match_stage, group_stage, match_stage2, sort_stage, limit_stage,
                                   project_stage] )
print(dumps(cursor, indent=2))

[
  {
    "user_id": null,
    "user_count": 302
  }
]

```

In []:

In []: