

ArtistsAndBands Writeup

Phase 0:

```
shimm@TABLET-7SLA7USR MINGW64 ~/coding/Greengart_Shimon_2018998974/ModernDataManagement/assignments/ArtistsAndBands (main)
$ time python main.py
So far: processed 0 lines
So far: processed 10000 lines
So far: processed 20000 lines
So far: processed 30000 lines
So far: processed 40000 lines
So far: processed 50000 lines
So far: processed 60000 lines
So far: processed 70000 lines
So far: processed 80000 lines
So far: processed 90000 lines
So far: processed 100000 lines
Total lines processed: 106270
Useful lines processed: 106089

real    0m3.800s
user    0m0.000s
sys      0m0.000s

shimm@TABLET-7SLA7USR MINGW64 ~/coding/Greengart_Shimon_2018998974/ModernDataManagement/assignments/ArtistsAndBands (main)
$
```

Phase 1: Schema

- 'bands_of:<artist_name>' connecting to the set of bands
- 'artists_in:<band_name>' connecting to the set of artists
- '<artist_name>:roles_in:<band_name>' connecting to a set of the artist's roles in the band

Query 1. Which artists were members of “The Beatles”?

smembers 'artists_in:The Beatles'

```
shimm@TABLET-7SLA7USR: /mnt/c/Users/shimm/coding/Greengart_Shimon_2018998974/ModernDataManagement/assignments/ArtistsAndBands$ redis-cli
127.0.0.1:6379> select 4
OK
127.0.0.1:6379[4]> smembers 'artists_in:The Beatles'
1) "Paul McCartney"
2) "George Harrison"
3) "Stuart Sutcliffe"
4) "Ringo Starr"
5) "John Lennon"
6) "Pete Best"
127.0.0.1:6379[4]> |
```

Query 2. Which bands was Paul McCartney a member of?

smembers 'bands_of:Paul McCartney'

```

6) "Pete Best"
127.0.0.1:6379[4]> smembers 'bands_of:Paul McCartney'
1) "Paul McCartney & Stevie Wonder"
2) "Band Aid 20"
3) "Paul McCartney & Michael Jackson"
4) "The Smokin' Mojo Filters"
5) "Paul & Linda McCartney"
6) "Band Aid"
7) "Paul McCartney & Eric Clapton"
8) "The Quarrymen"
9) "The Fireman"
10) "Wings"
11) "John Lennon & Paul McCartney"
12) "Suzy and the Red Stripes"
13) "Johnny Cash & Paul McCartney"
14) "The Christians, Holly Johnson, Paul McCartney, Gerry Marsden & Stock Aitken Waterman"
15) "The Beatles"
127.0.0.1:6379[4]> |

```

Query 3. Which roles did “Cale Harper” play in the band “Horses About Men”?

```
smembers 'Cale Harper:roles_in:Horses About Men'
```

```

15) "The Beatles"
127.0.0.1:6379[4]> smembers 'Cale Harper:roles_in:Horses About Men'
1) "Lyricist"
2) "Acoustic guitar"
3) "Songwriter"
4) "Harmonica"
5) "Rhythm guitar"
6) "Lead vocalist"
127.0.0.1:6379[4]> |

```

Phase 2: Schema

```

DB artists_and_bands COLLECTION band_collection: {
  name: String
  artists: [{
    name: String
    roles: [String] // empty array if missing
  }]
}

```

You know how I previously said that Mongo was much better than you made it seem? I have changed my mind. There are some things that should be straightforward to do – like take the union of arrays during a group operation – that are instead very difficult. I’m starting to understand the importance of a system with an underlying mathematical foundation. You should Shep Nachas.

Query 1. Display a sample band document

```
db.band_collection.aggregate([
  {$sample: {size: 1}}
])
```

```
]
artists_and_bands> db.band_collection.aggregate([
...   {$sample: {size: 1}}
... ])
[
  {
    _id: ObjectId('661328d2aa466b767aeaed48'),
    name: 'Jukkapoika ja Jenkkarekka',
    artists: [ { name: 'Jukka Poika', roles: [] } ]
  }
]
artists_and_bands> |
```

Query 2: How many distinct bands in the data set?

```
db.band_collection.aggregate([
  {$group: {_id: "$name"}},
  {$count: "name"}
])
```

```
MongoServerError[Location:40324]: unrecognized pipeline stage name: 'group'
artists_and_bands> db.band_collection.aggregate([{$group: {_id: "$name"}}, {$count: "name"}])
[ { name: 38583 } ]
artists_and_bands> |
```

Query 3. How many artists in the band “Men at Work”?

```
db.band_collection.aggregate([
  {$unwind: "$artists"},
  {$match: {name: "Men at Work"}},
  {$count: "name"}
])
```

```
[ { artist_name: 22 } ]
artists_and_bands> db.band_collection.aggregate([
...     {$unwind: "$artists"},
...     {$match: {name: "Men at Work"}},
...     {$count: "name"}
... ])
[ { name: 22 } ]
artists_and_bands> |
```

4. How many bands did Paul McCartney play in?

```
db.band_collection.aggregate([
    {$unwind: "$artists"},
    {$match: {"artists.name": "Paul McCartney"}},
    {$count: "name"}
])
```

```
[ { name: 15 } ]
artists_and_bands> db.band_collection.aggregate([
...     {$unwind: "$artists"},
...     {$match: {"artists.name": "Paul McCartney"}},
...     {$count: "name"}
... ])
[ { name: 15 } ]
artists_and_bands> |
```

5. Find all information about artist John Cooper in the band Skillet. Note: your query can only return information about this artist: i.e., may not return information about any other band member.

```
db.band_collection.aggregate([
    {$unwind: "$artists"},
    {$match: {"artists.name": "John Cooper", name: "Skillet"}},
    {$project: {artist: "$artists.name", band: "$name", roles:
"$artists.roles", _id: 0}}
])
```

```

]
artists_and_bands> db.band_collection.aggregate([
...     {$unwind: "$artists"},
...     {$match: {"artists.name": "John Cooper", name: "Skillet"}},
...     {$project: {"artist": "$artists.name", band: "$name", roles: "$artists.roles", _id: 0}}
... ])
[
  {
    artist: 'John Cooper',
    band: 'Skillet',
    roles: [ 'Acoustic guitar', 'Bass', 'Lead vocalist' ]
  }
]
artists_and_bands> |

```

6. List the names of every band that is associated with more than 50 artists, along with the number of artists in that band.

```

db.band_collection.aggregate([
    {$project: {_id: 0, name: 1, num_artists: {$size: "$artists"}}},
    {$match: {"num_artists": {$gt: 50}}}
])

```

```

artists_and_bands> db.band_collection.aggregate([
...     {$project: {_id: 0, name: 1, num_artists: {$size: "$artists"}}},
...     {$match: {"num_artists": {$gt: 50}}}
... ])
[
  { name: 'Israel Philharmonic Orchestra', num_artists: 114 },
  { name: 'The Drifters', num_artists: 60 },
  { name: 'Santana', num_artists: 69 },
  { name: 'ハロー!プロジェクト', num_artists: 83 },
  { name: 'Pigface', num_artists: 121 },
  { name: 'Duke Ellington', num_artists: 62 },
  { name: 'Les Enfoirés', num_artists: 134 },
  { name: 'Koor Fontys Conservatorium', num_artists: 55 }
]
artists_and_bands> |

```

7. List the names of artists with more than ten roles, along with the number of roles, sorted by ascending number of roles. I had originally thought that repeat roles from multiple bands should be ignored, but you said otherwise on Piazza.

```

db.band_collection.aggregate([
    {$unwind: "$artists"},
    {$unwind: "$artists.roles"},
    {$group: {_id: "$artists.name", roles: {$push: "$artists.roles"}}},
    {$project: {name: "$_id", _id: 0, role_size: {$size: "$roles"}}},

```

```

    {$match: {role_size: {$gt: 10}}},
    {$sort: {role_size: 1}}
  ])
}

```

```

]
artists_and_bands> db.band_collection.aggregate([
...   {$unwind: "$artists"},
...   {$unwind: "$artists.roles"},
...   {$group: {_id: "$artists.name", roles: {$push: "$artists.roles"}}},
...   {$project: {name: "$_id", _id: 0, role_size: {$size: "$roles"}}},
...   {$match: {role_size: {$gt: 10}}},
...   {$sort: {role_size: 1}}
... ])
[
  { name: 'Dom Totino', role_size: 11 },
  { name: 'Adam R. Young', role_size: 11 },
  { name: 'Dave Hallowell', role_size: 12 },
  { name: 'Chris Frangou', role_size: 13 },
  { name: 'Chris Squire', role_size: 14 },
  { name: 'Jon Anderson', role_size: 15 },
  { name: 'Steve Howe', role_size: 18 },
  { name: 'Richard Libeton', role_size: 18 },
  { name: 'Rick Wakeman', role_size: 19 },
  { name: 'Francois Breytenbach Blom', role_size: 20 },
  { name: 'Les Fradkin', role_size: 36 }
]
artists_and_bands> |

```