

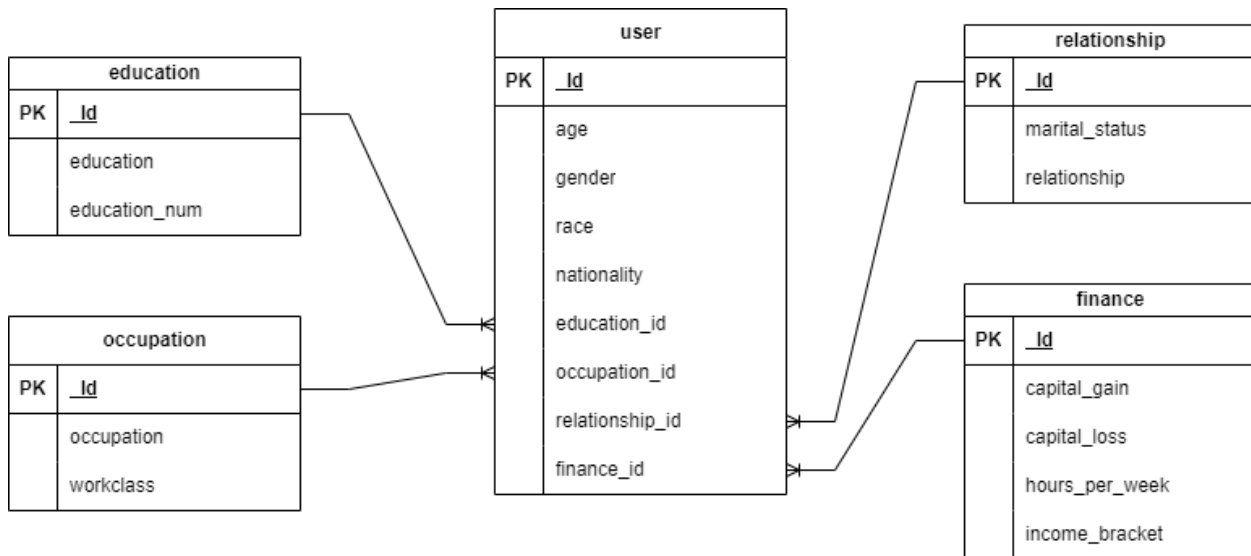
BÁO CÁO

Assignment 02 - DEP302

Họ và tên: Nguyễn Hải Dương

Mã SV: duongnhfx16133@funix.edu.vn

I. ERD của Database



II. Các mối quan hệ của các collections:

Giải thích mối quan hệ giữa các Collection và cấu trúc của các Document trong đó:

Mỗi bảng được kết nối với bảng **user** thông qua khóa ngoại, cho thấy rằng thông tin về giáo dục, nghề nghiệp, tài chính, và quan hệ đều được liên kết với một người dùng cụ thể.

Bảng **user** có các trường cơ bản như age, gender, race, nationality cùng với các khóa ngoại liên kết tới các bảng khác.

Bảng **education** có thể chứa thông tin về tên và mức độ giáo dục.

Bảng **occupation** chứa thông tin về nghề nghiệp và lớp công việc.

Bảng **relationship** chứa thông tin về tình trạng hôn nhân và mối quan hệ.

Bảng **finance** quản lý các thông tin về thu nhập và giờ làm việc.

III. Liệt kê các Business Query:

1. Có bao nhiêu người là Nữ và làm việc nhiều hơn 30 tiếng / tuần ?

```
use('DEP302_ASM2')
db.user.aggregate([
  {
    $match: {gender:"Female"}
  },
  {
    $lookup:{
      from: "finance",
      localField: "finance_id",
      foreignField: "_id",
      as: "finance"
    }
  },
  {
    $match:{"finance.hours_per_week":{$gt:30}}
  },
  {
    $count:"total"
  }
])
```

2. Có bao nhiêu người ở Mỹ có mức thu nhập > 50K

```
use('DEP302_ASM2')
db.user.aggregate([
  {
    $match: {
      native_country:"United-States"
    }
  },
  {
    $lookup:{
      from: "finance",
      localField: "finance_id",
      foreignField: "_id",
      as: "finance"
    }
  },
  {
    $match:{"finance.income_bracket": ">50K"}
  },
  {
    $count:"total"
  }
])
```

3. Tính tổng số dư tài khoản của những người đang ở Mỹ.

```
use('DEP302_ASM2')
db.user.aggregate([
  {
    $match: {
      native_country: "United-States"
    }
  },
  {
    $lookup: {
      from: "finance",
      localField: "finance_id",
      foreignField: "_id",
      as: "finance"
    }
  },
  {
    $unwind: "$finance"
  },
  {
    $group: {
      _id: null,
      total_balance: { $sum: "$finance.total" }
    }
  }
])
```

4. Tính tổng số giờ làm việc một tuần của những người có mức thu nhập <= 50K

//Cách 1: Tổng hợp từ bảng user

```
use('DEP302_ASM2')
db.user.aggregate([
  {
    $lookup:{
      from: "finance",
      localField: "finance_id",
      foreignField: "_id",
      as: "finance"
    }
  },
  {
    $unwind:"$finance"
  },
  {
    $match:{"finance.income_bracket":"<=50K"}
  },
  {
    $group:{
      _id:null,
      total_hours:{$sum:"$finance.hours_per_week"}
    }
  }
])
```

//Cách 2: Tổng hợp từ bảng finance

```
use('DEP302_ASM2')
db.user.createIndex({finance_id:1});
db.user.createIndex({income_bracket:1});
db.finance.aggregate([
  {
    $match: {income_bracket: "<=50K"}
  },
  {
    $lookup: {
      from: "user",
      localField: "_id",
      foreignField: "finance_id",
      as: "finance"
    }
  },
  {
    $unwind: "$finance"
  },
  {
    $group: {
      _id: null,
      total_hours: { $sum: "$hours_per_week" }
    }
  }
])
```

5. Tìm những người có tổng số tiền trong tài khoản > 1000000 và có số giờ làm việc hàng tuần < 55

```
use('DEP302_ASM2')
db.user.aggregate([
  {
    $lookup:{
      from: "finance",
      localField: "finance_id",
      foreignField: "_id",
      as: "finance"
    }
  },
  {
    $unwind:"$finance"
  },
  {
    $match:{"finance.total":{$gt:1000000},"finance.hours_per_week":{$lt:55}}
  },
  {
    $count:"count"
  }
])
```

IV. Thiết lập các index

Trường native_country trong collection user được sử dụng nhiều lần để lọc user theo quốc gia.

```
db.user.createIndex({native_country:1});
// Trường native_country được sử dụng trong phép so sánh nhiều lần
```

Trường finance_id và income_bracket được sử dụng nhiều trong collection user để phân loại user theo các mức thu nhập

```
db.user.createIndex({finance_id:1});
db.user.createIndex({income_bracket:1});
```