Rendu n°2 : MONGO

Groupe : Adèle MONTOYA, Eleonor KIOULOU, Paul RUNAVOT
Classe : ESILV S8, DIA 4

Presentation of the dataset

We worked on the Companies (2) dataset.
This database contains:
- basic information about companies like the name, the location, the URL of the homepage, the category of companies
- specific data like the amount about investments, acquisitions, type, currency…

Insertion of the dataset

Creation of the database :

```
data> db.createCollection("mycollection")
{ ok: 1 }

data> show collections
mycollection
```

Simple queries (6)

1. Récupérer les entreprises dont le nombre d'employés est supérieur à 1000 :
db.myCollection.find({number_of_employees: {$gt: 1000}}, {_id: 0, name: 1, number_of_employees: 1})

```
data> db.myCollection.find({number_of_employees: {$gt: 1000}}, {_id: 0, name: 1, number_of_employees: 1})
[
  { name: 'Apple', number_of_employees: 80000 },
  { name: 'NetApp', number_of_employees: 8000 },
  { name: 'SuccessFactors', number_of_employees: 1200 },
  { name: 'Motorola Solutions', number_of_employees: 51000 },
  { name: 'Expedia', number_of_employees: 4400 },
  { name: 'Salesforce', number_of_employees: 3500 },
  { name: 'Experian', number_of_employees: 15500 },
  { name: 'Netflix', number_of_employees: 2000 },
  { name: 'Nykredit Realkredit', number_of_employees: 4296 },
  { name: 'Los Angeles Times Media Group', number_of_employees: 10000 },
  { name: 'StubHub', number_of_employees: 1250 },
  { name: 'BBC', number_of_employees: 25000 },
  { name: 'Intuit', number_of_employees: 8000 },
  { name: 'PCCW', number_of_employees: 18200 },
  { name: 'Mobiluck', number_of_employees: 1234 },
  { name: 'Dell', number_of_employees: 80000 },
  { name: 'Logitech', number_of_employees: 8000 },
  { name: 'LG', number_of_employees: 177000 },
  { name: 'Zoho', number_of_employees: 1600 },
  { name: 'Facebook', number_of_employees: 5299 }
]
Type "it" for more
```

2. Récupérer les entreprises dont la catégorie est "Software" :
db.myCollection.find({category_code: "software"}, {name: 1, category_code: 1, _id: 0})

```
data> db.myCollection.find({category_code: "software"}, {name: 1, category_code: 1, _id: 0})
[
  { name: 'LiveJournal', category_code: 'software' },
  { name: 'RipCode', category_code: 'software' },
  { name: 'Bunchball', category_code: 'software' },
  { name: 'NetApp', category_code: 'software' },
  { name: 'Workface', category_code: 'software' },
  { name: '4Home', category_code: 'software' },
  { name: 'SugarCRM', category_code: 'software' },
  { name: 'SurePeak', category_code: 'software' },
  { name: 'Mumboe', category_code: 'software' },
  { name: 'Roost', category_code: 'software' },
  { name: '5pm', category_code: 'software' },
  { name: 'Reshade', category_code: 'software' },
  { name: ']project-open[', category_code: 'software' },
  { name: 'LiquidPlanner', category_code: 'software' },
  { name: 'Vaakya Technologies', category_code: 'software' },
  { name: 'Sikka Software', category_code: 'software' },
  { name: 'Offermatica', category_code: 'software' },
  { name: 'CoActLive', category_code: 'software' },
  { name: 'ioko', category_code: 'software' },
  { name: 'Gliffy', category_code: 'software' }
]
Type "it" for more
```

3. Récupérer les entreprises dont la date de fondation est entre 2005 et 2007 :
db.myCollection.find(
... {
...   "founded_year": { $gte: 2005, $lte: 2007 }
... },
... {
...   "name": 1,
...   "founded_year": 1,
...   "_id": 0
... })

```
data> db.myCollection.find(
...   {
...       "founded_year": { $gte: 2005, $lte: 2007 }
...   },
...   {
...       "name": 1,
...       "founded_year": 1,
...       "_id": 0
...   }
... )
[
  { name: 'TringMe', founded_year: 2007 },
  { name: 'CrossLoop', founded_year: 2006 },
  { name: 'eventsbot', founded_year: 2007 },
  { name: 'Vidyo', founded_year: 2005 },
  { name: 'iMemories', founded_year: 2006 },
  { name: 'Criticker', founded_year: 2005 },
  { name: 'Dapper', founded_year: 2006 },
  { name: 'Twipster', founded_year: 2007 },
  { name: 'ClickTale', founded_year: 2006 },
  { name: 'TradeKing', founded_year: 2005 },
  { name: 'Twenga', founded_year: 2006 },
  { name: 'SocialPicks', founded_year: 2006 },
  { name: 'Jacked', founded_year: 2006 },
  { name: 'FFWD Wheels', founded_year: 2006 },
  { name: 'Shipwire', founded_year: 2006 },
  { name: 'CafeMom', founded_year: 2006 },
  { name: 'eBurl', founded_year: 2007 },
  { name: 'Uversity', founded_year: 2007 },
  { name: 'Vemo', founded_year: 2005 },
  { name: 'Bahu', founded_year: 2007 }
]
Type "it" for more
```

4. Récupérer les entreprises dont l'URL Crunchbase est renseignée :
db.myCollection.find({ crunchbase_url: { $ne: null } }, { name: 1, crunchbase_url: 1, _id: 0 })

```
data> db.myCollection.find({ crunchbase_url: { $ne: null } }, { name: 1, crunchbase_url: 1, _id: 0 })
[
  {
    name: 'Fraud Sciences',
    crunchbase_url: 'http://www.crunchbase.com/company/fraud-sciences'
  },
  {
    name: 'WeGame',
    crunchbase_url: 'http://www.crunchbase.com/company/wegame'
  },
  {
    name: 'IRSeeK',
    crunchbase_url: 'http://www.crunchbase.com/company/irseek'
  },
  {
    name: 'Intronis',
    crunchbase_url: 'http://www.crunchbase.com/company/intronis'
  },
  {
    name: 'TringMe',
    crunchbase_url: 'http://www.crunchbase.com/company/tringme'
  },
  {
    name: 'CrossLoop',
    crunchbase_url: 'http://www.crunchbase.com/company/crossloop'
  },
  {
    name: 'eventsbot',
    crunchbase_url: 'http://www.crunchbase.com/company/eventsbot'
  },
  {
    name: 'Vidyo',
    crunchbase_url: 'http://www.crunchbase.com/company/vidyo'
  },
  {
    name: 'iMemories',
    crunchbase_url: 'http://www.crunchbase.com/company/imemories'
  },
  {
    name: 'Criticker',
    crunchbase_url: 'http://www.crunchbase.com/company/criticker'
  },
  {
    name: 'Trezr',
    crunchbase_url: 'http://www.crunchbase.com/company/trezr'
  },
  {
    name: 'expensr',
    crunchbase_url: 'http://www.crunchbase.com/company/expensr'
  },
  {
    name: 'Dapper',
    crunchbase_url: 'http://www.crunchbase.com/company/dapper'
  },
  {
    name: 'Twipster',
    crunchbase_url: 'http://www.crunchbase.com/company/twipster'
  },
  {
    name: 'ClickTale',
    crunchbase_url: 'http://www.crunchbase.com/company/clicktale'
  },
  {
    name: 'TradeKing',
    crunchbase_url: 'http://www.crunchbase.com/company/tradeking'
  },
  {
    name: 'Twenga',
    crunchbase_url: 'http://www.crunchbase.com/company/twenga'
  },
  {
    name: 'SocialPicks',
    crunchbase_url: 'http://www.crunchbase.com/company/socialpicks'
  },
  {
    name: 'Viadeo',
    crunchbase_url: 'http://www.crunchbase.com/company/viadeo'
  },
  {
    name: 'Jacked',
    crunchbase_url: 'http://www.crunchbase.com/company/jacked'
  }
]
Type "it" for more
data>
```

5. Récupérer le montant total des fonds levés par l'entreprise :
db.myCollection.find({}, {"name": 1, "total_money_raised": 1, "_id": 0})

```
data> db.myCollection.find({}, {"name": 1, "total_money_raised": 1, "_id": 0})
[
  { name: 'Fraud Sciences', total_money_raised: '$11M' },
  { name: 'WeGame', total_money_raised: '$3.5M' },
  { name: 'IRSeeK', total_money_raised: '$0' },
  { name: 'Intronis', total_money_raised: '$25.9M' },
  { name: 'TringMe', total_money_raised: '$0' },
  { name: 'CrossLoop', total_money_raised: '$10.5M' },
  { name: 'eventsbot', total_money_raised: '$0' },
  { name: 'Vidyo', total_money_raised: '$141M' },
  { name: 'iMemories', total_money_raised: '$16M' },
  { name: 'Criticker', total_money_raised: '$0' },
  { name: 'Trezr', total_money_raised: '$0' },
  { name: 'expensr', total_money_raised: '$0' },
  { name: 'Dapper', total_money_raised: '$3M' },
  { name: 'Twipster', total_money_raised: '$0' },
  { name: 'ClickTale', total_money_raised: '$23M' },
  { name: 'TradeKing', total_money_raised: '$6M' },
  { name: 'Twenga', total_money_raised: '€2.6M' },
  { name: 'SocialPicks', total_money_raised: '$500k' },
  { name: 'Viadeo', total_money_raised: '$50.2M' },
  { name: 'Jacked', total_money_raised: '$9M' }
]
Type "it" for more
```

6. Récupérer la liste des compétiteurs de l'entreprise:
db.myCollection.find({}, {name: 1, competitions: 1, _id:0})

```
data> db.myCollection.find({}, {name: 1, competitions: 1, _id:0})
```
```
{ name: 'Fraud Sciences', competitions: [] },
{
  name: 'WeGame',
  competitions: [
    { competitor: { name: 'UGAME', permalink: 'ugame' } },
    { competitor: { name: 'Wipido', permalink: 'wipido' } },
    { competitor: { name: 'GameVee', permalink: 'gamevee' } }
  ]
},
{ name: 'IRSeeK', competitions: [] },
{
  name: 'Intronis',
  competitions: [
    { competitor: { name: 'Mozy', permalink: 'mozy' } },
    { competitor: { name: 'Carbonite', permalink: 'carbonite' } },
    { competitor: { name: 'filobite', permalink: 'filobite' } }
  ]
},
{ name: 'TringMe', competitions: [] },
{ name: 'CrossLoop', competitions: [] },
{ name: 'eventsbot', competitions: [] },
{
  name: 'Vidyo',
  competitions: [
    { competitor: { name: 'Dimdim', permalink: 'dimdim' } },
    { competitor: { name: 'Polycom', permalink: 'polycom-2' } },
    { competitor: { name: 'ooVoo', permalink: 'oovoo' } }
  ]
},
{ name: 'iMemories', competitions: [] },
{ name: 'Criticker', competitions: [] },
{ name: 'Trezr', competitions: [] },
{
  name: 'expensr',
  competitions: [
    { competitor: { name: 'Mint.com', permalink: 'mint' } },
    { competitor: { name: 'Wesabe', permalink: 'wesabe' } },
    { competitor: { name: 'sMoneyBox', permalink: 'smoneybox' } },
    { competitor: { name: 'Thrive', permalink: 'thrive' } },
    { competitor: { name: 'Jwaala', permalink: 'jwaala' } },
    { competitor: { name: 'PearBudget', permalink: 'pearbudget' } },
    {
      competitor: { name: 'GreenSherpa', permalink: 'green-sherpa' }
    }
  ]
},
  name: 'Dapper',
  competitions: [ { competitor: { name: 'Feedity', permalink: 'feedity' } } ]
},
{ name: 'Twipster', competitions: [] },
{
  name: 'ClickTale',
  competitions: [
    {
      competitor: { name: 'thatsmymouse', permalink: 'thatsmymouse' }
    },
    {
      competitor: {
        name: 'www.SeeMyVisits.com',
        permalink: 'www-seemyvisits-com'
      }
    },
    {
      competitor: { name: 'www.sessioncam.com', permalink: 'www-sessioncam-com' }
    }
  ]
},
{ name: 'TradeKing', competitions: [] },
{
  name: 'Twenga',
  competitions: [
    { competitor: { name: 'Bountii', permalink: 'bountii' } },
    {
      competitor: { name: 'Crowdstorm USA', permalink: 'crowdstorm' }
    },
    { competitor: { name: 'TheFind', permalink: 'thefind' } },
    { competitor: { name: 'ShopWiki', permalink: 'shopwiki' } }
  ]
},
{
  name: 'SocialPicks',
  competitions: [
    { competitor: { name: 'StockTwits', permalink: 'stocktwits' } },
    { competitor: { name: 'Stocial', permalink: 'stocial' } }
  ]
},
{ name: 'Viadeo', competitions: [] },
{ name: 'Jacked', competitions: [] }
]
Type "it" for more
```

Complex queries (2)

1. Trier les entreprises par ordre décroissant du nombre d'employés et n'afficher que les 10 premières :

```
data> db.myCollection.aggregate([ { $sort: { number_of_employees: -1 } }, { $limit: 10 }, { $project: { name: 1, number_of_employees: 1 , _id:0} }])
[
  { name: 'Siemens', number_of_employees: 405000 },
  { name: 'IBM', number_of_employees: 388000 },
  { name: 'Toyota', number_of_employees: 320000 },
  { name: 'PayPal', number_of_employees: 300000 },
  {
    name: 'Nippon Telegraph and Telephone Corporation',
    number_of_employees: 227000
  },
  { name: 'Samsung Electronics', number_of_employees: 221726 },
  { name: 'Accenture', number_of_employees: 205000 },
  { name: 'Tata Consultancy Services', number_of_employees: 200300 },
  { name: 'Flextronics International', number_of_employees: 200000 },
  { name: 'Safeway', number_of_employees: 186000 }
]
```

2. Récupérer les entreprises ayant levé plus de 50 millions de dollars, triées par ordre décroissant du nombre de rounds de financement :

```
data> db.myCollection.aggregate([ { $match: { total_money_raised: { $gte: "50M" } } }, { $project: { _id: 1, name: 1, num_rounds: { $size: "$funding_rounds" }, _id:0 } }, { $sort: { num_rounds: -1 } }])
[
  { name: 'Clear2Pay', num_rounds: 7 },
  { name: 'Shutl', num_rounds: 6 },
  { name: 'Symphogen', num_rounds: 6 },
  { name: 'EnOcean', num_rounds: 5 },
  { name: 'EnOcean', num_rounds: 5 },
  { name: 'ARKeX', num_rounds: 5 },
  { name: 'Imagination Technologies', num_rounds: 5 },
  { name: 'Proximic', num_rounds: 4 },
  { name: 'Qype', num_rounds: 4 },
  { name: 'Fon', num_rounds: 4 },
  { name: 'Blyk', num_rounds: 4 },
  { name: 'Akvo', num_rounds: 4 },
  { name: 'Tupalo', num_rounds: 4 },
  { name: 'Image Metrics', num_rounds: 4 },
  { name: 'CamSemi', num_rounds: 4 },
  { name: 'Powervation', num_rounds: 4 },
  { name: 'Specle', num_rounds: 4 },
  { name: 'Taptu', num_rounds: 3 },
  { name: 'Seedcamp', num_rounds: 3 },
  { name: 'CloudMade', num_rounds: 3 }
]
Type "it" for more
```

Hard queries (1)

mapReduce : Récupérer le nombre total d'employés pour chaque catégorie d'entreprise ayant plus de 100 employés

```
data> db.myCollection.mapReduce(
...      function() {
...          emit(this.category_code, this.number_of_employees);
...      },
...      function(key, values) {
...          return Array.sum(values);
...      },
...      {
...        query: { number_of_employees: { $gt: 100 } },
...        out: { inline: 1 }
...      }
... )
```

results: [
  { _id: 'security', value: 22648 },
  { _id: 'social', value: 12941 },
  { _id: 'finance', value: 300990 },
  { _id: 'other', value: 664956 },
  { _id: 'fashion', value: 34900 },
  { _id: 'enterprise', value: 532926 },
  { _id: 'news', value: 25520 },
  { _id: 'music', value: 5572 },
  { _id: 'analytics', value: 1815 },
  { _id: 'messaging', value: 300 },
  { _id: 'ecommerce', value: 22505 },
  { _id: 'hardware', value: 1563817 },
  { _id: null, value: 758871 },
  { _id: 'real_estate', value: 600 },
  { _id: 'advertising', value: 31443 },
  { _id: 'web', value: 120588 },
  { _id: 'search', value: 43115 },
  { _id: 'travel', value: 6000 },
  { _id: 'consulting', value: 790403 },
  { _id: 'hospitality', value: 250 },
  { _id: 'software', value: 737187 },
  { _id: 'mobile', value: 332536 },
  { _id: 'health', value: 930 },
  { _id: 'games_video', value: 148379 },
  { _id: 'biotech', value: 82080 },
  { _id: 'education', value: 350 },
  { _id: 'nonprofit', value: 143 },
  { _id: 'transportation', value: 93696 },
  { _id: 'automotive', value: 171500 },
  { _id: 'semiconductor', value: 51959 },
  { _id: 'cleantech', value: 17296 },
  { _id: 'network_hosting', value: 89605 },
  { _id: 'public_relations', value: 276325 }