

MongoDB

une introduction

Par Jean-René Rouet / **CCIN2P3/CNRS**

MongoDB

- de "humongous" en toute modestie
- Base de données orientée document
- Open Source
- La plus utilisée des bases «NoSQL»
- Ecrit en C++
- Langage javascript

Orientée Document

- BSON format
 - Json + types natifs (date, binaryData)

```
{
  _id: 1,
  name: { first: 'John', last: 'Backus' },
  birth: new Date('Dec 03, 1924'),
  death: new Date('Mar 17, 2007'),
  contribs: [ 'Fortran', 'ALGOL', 'Backus-Naur Form', 'FP' ],
  awards: [
    {
      award: 'W.W. McDowell Award',
      year: 1967,
      by: 'IEEE Computer Society'
    },
    {
      award: 'National Medal of Science',
      year: 1975,
      by: 'National Science Foundation'
    },
    {
      award: 'Turing Award',
      year: 1977,
      by: 'ACM'
    }
  ]
}
```

Pas de schéma

On stocke ce que l'on veut où l'on veut

Une collection (unité de stockage sous la base) peut contenir des documents complètement différents

Modélisation des données

- Documents embarqués
 - relation "contient"
 - relation "one to many(peu nombreux)"
- Documents référencés
 - relation "many(nombreux) to one"
 - relation "many to many"
- pas de jointure donc 2 requêtes pour les documents référencés

Indexation

- Tous les attributs peuvent être indexés
 - simple
 - composé
 - unique
- Types d'index évolués
 - géospatial
 - text

Réplication et haute disponibilité

- Primaire et secondaires
 - si le primaire disparaît, un vote est effectué par les secondaires pour élire le nouveau primaire.
 - transparent pour le client

Répartition

Quand un noeud contient trop de données : on ajoute un nouveau noeud

On envoie une partie des données sur ce nouveau noeud, c'est automatique

Et on maîtrise les clés de répartition

Requêtes

- Puissantes
 - attributs du documents
 - attributs des sous-documents
 - tableaux
- Prédicats variés
 - = != > < ...
 - regexp

Requêtes

- BSON format
 - Json + types natifs (date, binaryData)

```
{
  _id: 1,
  name: { first: 'John', last: 'Backus' },
  birth: new Date('Dec 03, 1924'),
  death: new Date('Mar 17, 2007'),
  contribs: [ 'Fortran', 'ALGOL', 'Backus-Naur Form',
  awards: [
    {
      award: 'W.W. McDowell Award',
      year: 1967,
      by: 'IEEE Computer Society'
    },
    {
      award: 'National Medal of Science',
      year: 1975,
      by: 'National Science Foundation'
    },
    {
      award: 'Turing Award',
      year: 1977,
      by: 'ACM'
    }
  ]
}
```

```
db.bios.find(
  {
    awards: {
      $elemMatch: {
        award: 'Tur
        year: { $gt
      }
    }
  }
)
```

Mise à jour atomique

On peut modifier un document de façon atomique (find et update)

Map/Reduce

- On peut implémenter des traitements map/reduce
- Mais pour des tâches simples (aggrégations) il existe des alternatives

```
{
  "_id": "10280",
  "city": "NEW YORK",
  "state": "NY",
  "pop": 5574,
  "loc": [
    -74.016323,
    40.710537
  ]
}
```

```
db.zipcodes.aggregate( { $group :
  { _id : "$state",
    totalPop : { $sum : "$pop" } } },
  { $match : {totalPop : { $gte : 10*1000*1000 } } } )
```

```
SELECT "state", SUM("pop") AS "totalPop" FROM "zipcodes" GROUP BY "state" HAVING "totalPop">
```

Gros documents

- BSON est limité à 16 mo par document
- GridFS permet de dépasser cette limite

10gen

- Il est possible d'acheter des formations, du support et de la consultance
- Utilisateurs
 - Disney online gaming (1400 instances)
 - Cern CMS