

TP avec **Mapreduce** en MongoDB

Exmple d'un document livre de la collection biblio :

```
{
  "type": "Book",
  "title": "Modern Database Systems: The Object Model,
Interoperability, and Beyond.",
  "year": 1995,
  "publisher": "ACM Press and Addison-Wesley",
  "authors": ["Won Kim"],
  "source": "DBLP"
}
```

1. Décompresser le fichier `biblio.json.zip`
2. Dans le même répertoire, lancer l'importation du fichier, dans `mongodb` `compass` :

Quelques rappels

Pour exprimer une requête MapReduce, il faut définir une fonction de *map*, une fonction de *reduce*, un filtre (optionnel), lancer le tout avec l'opérateur `mapReduce` et consulter le résultat (optionnel!). Soit:

```
map    var mapFunction = function () {emit(this.year, 1)};
reduce var reduceFunction = function (key, values) {return
Array.sum(values)};
filtre  var queryParam = {query : {}, out : "result_set"}
requête db.publis.mapReduce(mapFunction, reduceFunction, queryParam);
résultat db.result_set.find();
```

Questions :

1. Pour chaque document de type livre (book), émettre le document avec pour clé « title »

```
[
  {
    _id: 'The knowledge acquisition and representation language KARL.',
    value: {
      articles: [
        {
          _id: 'books/daglib/0084198',
          type: 'Book',
          title: 'The knowledge acquisition and representation language KARL.',
          pages: 'I',
          publisher: 'Kluwer',
          year: 1995,
          authors: [ 'Dieter Fensel' ],
          isbn: [ '978-0-7923-9601-7' ]
        }
      ]
    }
  },
  {
    _id: 'Stability Analysis of Fuzzy-Model-Based Control Systems - Linear-Matrix-Inequality Approach',
    value: {
      articles: [
        {
          _id: 'books/sp/LamL11',
          type: 'Book',
          title: 'Stability Analysis of Fuzzy-Model-Based Control Systems - Linear-Matrix-Inequality Approach',
          volume: '264',
          year: 2010,
          pages: { start: 1, end: 215 },
          publisher: 'Springer',
          series: 'Studies in Fuzziness and Soft Computing',
          url: 'http://dx.doi.org/10.1007/978-3-642-17844-3',
          authors: [ 'Hak-Keung Lam', 'Frank Hung-Fat Leung' ],
          isbn: [ '978-3-642-17843-6' ]
        }
      ]
    }
  },
  {
    _id: 'Discrete scale-space formulation and multiscale edge extraction toward higher dimensions.',
    value: {
      articles: [
        {
          _id: 'books/daglib/0010393',
          type: 'Book',

```

2. Pour chacun de ces livres, donner le nombre de ses auteurs

```
[
  {
    series: 'Springer Briefs in Computer Science',
    url: 'http://dx.doi.org/10.1007/978-1-4614-8890-3',
    _id: 'An introduction to fuzzy logic for practical applications.',
    value: { articles: [ 1 ] }8889-7', '978-1-4614-8890-3' ]
  },
  {
    _id: 'Economics of Information Security',
    value: { articles: [ 0 ] }
  },
  {
    _id: 'Understanding CORBA - common object request broker architecture.',
    _id: 'Information Retrieval Evaluation',
    value: { articles: [ 1 ] }
  },
  {
    _id: 'books/daglib/0081916',
    _id: 'Learning JavaScript - add life to your web pages: covers Ajax and DOM.',
    value: { articles: [ 1 ] }g CORBA - common object request broker architecture.',
    publisher: 'Prentice Hall',
    year: 1995,
    _id: 'Running Linux - a Unix-compatible system for the personal computer (2. ed.',
    value: { articles: [ 2 ] }9884-0' ]
  },
  {
    _id: 'Hypertext Hands-on! - an introduction to a new way of organizing and access',
    value: { articles: [ 2 ] }
  },
  {
    e "it" for more
    _id: 'Advanced Statistical Steganalysis',
    value: { articles: [ 1 ] }
  },
  {
    _id: 'Semantics in Adaptive and Personalized Services: Methods, Tools and Applications',
    value: { articles: [ 0 ] }
  }
]
```

- 3- Pour l'éditeur « Springer », donner le nombre de publications par année

```
[
  { _id: 1993, value: 59 },
  { _id: 2012, value: 236 },
  { _id: 1982, value: 18 },
  { _id: 1979, value: 4 },
  { _id: 1978, value: 7 },
  { _id: 1976, value: 4 },
  { _id: 2001, value: 64 },
  { _id: 1995, value: 73 },
  { _id: 1997, value: 140 },
  { _id: 2009, value: 196 },
  { _id: 1990, value: 46 },
  { _id: 1969, value: 1 },
  { _id: 2014, value: 255 },
  { _id: 2003, value: 68 },
  { _id: 1999, value: 77 },
  { _id: 2007, value: 154 },
  { _id: 1998, value: 103 },
  { _id: 1987, value: 45 },
  { _id: 1975, value: 5 },
  { _id: 2002, value: 80 }
]
```

4. Pour chaque couple « publisher & année » (il faut que publisher soit présent), donner le nombre de publications.

Important

la clé du `emit()` doit être un document.

```
[
  { _id: { publisher: 'Universiteit Twente', year: 2002 }, value: 1 },
  {
    _id: { publisher: 'Cambridge University Press', year: 2011 },
    value: 14
  },
  {
    _id: { publisher: 'Information Science Reference', year: 2009 },
    value: 1
  },
  { _id: { publisher: 'Birkh?user', year: 2006 }, value: 2 },
  { _id: { publisher: 'Macmillan', year: 1986 }, value: 1 },
  { _id: { publisher: 'Addison-Wesley', year: 2003 }, value: 7 },
  { _id: { publisher: 'Prentice-Hall', year: 1981 }, value: 2 },
  { _id: { publisher: 'University Amsterdam', year: 2011 }, value: 1 },
  { _id: { publisher: 'McGraw-Hill', year: 1958 }, value: 1 },
  { _id: { publisher: 'VDI-Verlag', year: 1993 }, value: 1 },
  { _id: { publisher: 'Logos', year: 2002 }, value: 1 },
  { _id: { publisher: 'Morgan Kaufmann', year: 1994 }, value: 10 },
  {
    _id: { publisher: 'Manning Publications Company', year: 2004 },
    value: 1
  }
]
```

5. Pour l'auteur « Toru Ishida », donner le nombre de publications par année

```
base202> db.result_set06.find();
[
  { _id: 1958, value: 2 },
  { _id: 2006, value: 6431 },
  { _id: 2011, value: 8005 },
  { _id: 1988, value: 569 },
  { _id: 1980, value: 426 },
  { _id: 1955, value: 2 },
  { _id: 1991, value: 1151 },
  { _id: 1966, value: 17 },
  { _id: 1989, value: 857 },
  { _id: 1974, value: 56 },
  { _id: 1992, value: 1093 },
  { _id: 1971, value: 67 },
  { _id: 2008, value: 8760 },
  { _id: 1961, value: 1 },
  { _id: 1956, value: 3 },
  { _id: 2002, value: 3387 },
  { _id: 2014, value: 4553 },
  { _id: 1950, value: 2 },
  { _id: 2007, value: 6948 },
  { _id: 1965, value: 20 }
]
```

6. Pour chaque auteur, lister les titres de ses publications

Attention

la sortie du *map* et du *reduce* doit être un document (pas un tableau)

```
[
  {
    _id: 'Oliver Oechslein',
    value: {
      titles: [
        'The Value of a Recommendation: The Role of Social Ties in Social Recomm',
        'Examining Trust within the Team in IT Startup Companies-An Empirical St',
        'An Application of UTAUT2 on Social Recommender Systems: Incorporating S
ancy.'
      ]
    }
  },
  {
    _id: 'Alice Iosefina Renner',
    value: { titles: [ 'Anisotropic diffusion in Riemannian colour space.' ] }
  },
  {
    _id: 'Giovanni Russello',
    value: { titles: [ 'Trust Transfer in Distributed Systems.' ] }
  },
  {
    _id: 'Hans-Peter Bernhard',
    value: {
      titles: [
        'Nonlinear long-term prediction of speech signals.',
        'Performance analysis of the mutual information function for nonlinear a
      ]
    }
  },
  {
    id: 'Eugenio de Oliveira Simonetto',

```

7. Pour chaque auteur, lister le nombre de publications associé à chaque année

```
base202> db.result_set09.find();
[
  { _id: { author: 'Masataka Goto', year: 2007 }, value: 2 },
  { _id: { author: 'Kentaro Ogawa', year: 2011 }, value: 1 },
  { _id: { author: 'Bing-Yu Hsieh', year: 2003 }, value: 1 },
  { _id: { author: 'Griff L. Bilbro', year: 1983 }, value: 1 },
  { _id: { author: 'Eleni Tsalapati', year: 2009 }, value: 1 },
  { _id: { author: 'Hanku Lee', year: 2013 }, value: 4 },
  { _id: { author: 'Fan Zhai', year: 2007 }, value: 2 },
  { _id: { author: 'I. Lu', year: 2008 }, value: 1 },
  { _id: { author: 'Danny Little', year: 2005 }, value: 1 },
  { _id: { author: 'Zhongfeng Wang', year: 2004 }, value: 1 },
  { _id: { author: 'Fangwei Hu', year: 2013 }, value: 1 },
  { _id: { author: 'John F. Siebenand', year: 1979 }, value: 1 },
  { _id: { author: 'J. D. Brand', year: 2002 }, value: 1 },
  { _id: { author: 'Danilo Ardagna', year: 2006 }, value: 2 },
  { _id: { author: 'Yu Gong', year: 2011 }, value: 1 },
  { _id: { author: 'Issam Aib', year: 2008 }, value: 2 },
  { _id: { author: 'Celina Alexandre', year: 2014 }, value: 1 },
  { _id: { author: 'James R. McDonald', year: 2004 }, value: 1 },
  { _id: { author: 'Simone Paolo Ponzetto', year: 2010 }, value: 1 },
  { _id: { author: 'Mark Hartong', year: 2012 }, value: 1 }
]
```

8. Compter les publications de plus de 3 auteurs

```
[ { _id: null, value: 25724 } ]
```