

Guide Complet : MapReduce avec CouchDB pour Débutants

Introduction

MapReduce est un paradigme de programmation qui permet de traiter de grandes quantités de données de manière distribuée. Dans ce tutoriel, nous allons découvrir comment utiliser MapReduce dans un contexte centralisé avec CouchDB, une base de données NoSQL orientée document.

Pour rendre ce concept plus accessible, imaginons que vous devez trier une énorme bibliothèque de livres. Au lieu de tout trier vous-même, vous pourriez :

1. Distribuer les livres à plusieurs personnes (Map)
2. Leur demander de les organiser selon certains critères
3. Rassembler et combiner leurs résultats (Reduce)

Installation de CouchDB

Sous Windows

1. Téléchargez l'installateur depuis le site officiel de CouchDB :

```
https://couchdb.apache.org/#download
```

2. Exécutez l'installateur et suivez les instructions
3. Acceptez l'installation du service Windows
4. Gardez les paramètres par défaut pour le mode "single node"
5. Définissez un nom d'utilisateur admin et un mot de passe

Sous Linux (Ubuntu/Debian)

1. Ajoutez la clé du repository :

```
curl -L https://couchdb.apache.org/repo/keys.asc | gpg --dearmor | sudo tee /usr/share/keyrings/couchdb-archive-keyring.gpg >/dev/null
```

2. Ajoutez le repository CouchDB :

```
echo "deb [signed-by=/usr/share/keyrings/couchdb-archive-keyring.gpg] https://apache.jfrog.io/artifactory/couchdb-deb/ $(lsb_release -cs)" | sudo tee /etc/apt/sources.list.d/couchdb.list
```

3. Mettez à jour et installez :

```
sudo apt update
sudo apt install couchdb
```

4. Pendant l'installation, choisissez :
 - "standalone" pour le mode
 - "127.0.0.1" pour l'adresse d'écoute
 - Définissez un mot de passe admin

Sous MacOS

Avec Homebrew :

```
brew install couchdb
brew services start couchdb
```

Vérification de l'Installation

1. Ouvrez votre navigateur et accédez à :

```
http://127.0.0.1:5984/_utils
```

2. Connectez-vous avec vos identifiants admin
3. Vous devriez voir l'interface Fauxton de CouchDB

Comprendre MapReduce dans CouchDB

Les Bases

Dans CouchDB, MapReduce est utilisé pour créer des "vues" qui permettent d'indexer et d'interroger vos données efficacement. Une vue est composée de deux parties :

1. Une fonction **map** qui sélectionne et transforme les données
2. Une fonction **reduce** (optionnelle) qui agrège les résultats

Création d'une Base de Données de Test

Dans l'interface Fauxton :

1. Cliquez sur "Create Database"
2. Nommez-la "bibliotheque"
3. Ajoutez quelques documents de test :

```
{
  "_id": "livre1",
  "titre": "Le Petit Prince",
  "auteur": "Saint-Exupéry",
  "annee": 1943,
  "genre": "Roman"
}

{
  "_id": "livre2",
  "titre": "1984",
  "auteur": "Orwell",
  "annee": 1949,
  "genre": "Science-fiction"
}

{
  "_id": "livre3",
  "titre": "Dune",
  "auteur": "Herbert",
  "annee": 1965,
  "genre": "Science-fiction"
}
```

Création d'une Vue Simple

Créons une vue qui compte les livres par genre :

1. Dans Fauxton, sélectionnez votre base "bibliotheque"
2. Cliquez sur "Design Documents" puis "New View"
3. Créez un nouveau design document "_design/bibliotheque"
4. Nommez votre vue "livres_par_genre"

Fonction Map :

```
function (doc) {
  if (doc.genre && doc.titre) { // Vérifie que le document a un genre et un titre
    emit(doc.genre, 1);        // Émet le genre comme clé et 1 comme valeur
  }
}
```

Fonction Reduce :

```
function (keys, values, rereduce) {
  return sum(values); // Additionne simplement les valeurs
}
```

Comprendre le Résultat

Cette vue va produire un résultat comme :

```
{
  "rows": [
    {
      "key": "Roman",
      "value": 1
    },
    {
      "key": "Science-fiction",
      "value": 2
    }
  ]
}
```

Exemple Plus Avancé : Analyse des Publications par Décennie

Créons une vue plus complexe qui analyse les publications par décennie :

```
// Fonction Map
function (doc) {
  if (doc.annee) {
    // Calcule la décennie (1943 -> 1940)
    var decennie = Math.floor(doc.annee / 10) * 10;
    emit(decennie, {
      titre: doc.titre,
      auteur: doc.auteur,
      annee: doc.annee
    });
  }
}

// Fonction Reduce
function (keys, values, rereduce) {
  var result = {
    count: 0,
    livres: []
  };

  if (rereduce) {
    // Combine les résultats précédents
    values.forEach(function(value) {
      result.count += value.count;
      result.livres = result.livres.concat(value.livres);
    });
  } else {
    // Premier passage
    result.count = values.length;
    result.livres = values;
  }

  return result;
}
```

Bonnes Pratiques

1. Indexation Efficace

- Émettez uniquement les données nécessaires
- Utilisez des clés pertinentes pour votre cas d'usage

2. Performance

- Évitez les fonctions reduce trop complexes
- Limitez la taille des données émises

3. Validation

- Vérifiez toujours l'existence des propriétés avant de les utiliser

- Gérez les cas particuliers

Interrogation des Vues

Pour interroger une vue via HTTP :

```
http://127.0.0.1:5984/bibliotheque/_design/bibliotheque/_view/livres_par_genre
```

Avec des paramètres :

```
http://127.0.0.1:5984/bibliotheque/_design/bibliotheque/_view/livres_par_genre?group=true
```

Via curl :

```
curl -X GET http://admin:password@127.0.0.1:5984/bibliotheque/_design/bibliotheque/_view/livres_par_genre?group=true
```

Debuggage et Outils

Console JavaScript dans Fauxton

Fauxton intègre une console JavaScript pour tester vos fonctions map/reduce :

1. Allez dans l'onglet "Verify Installation"
2. Utilisez la console pour tester votre code

Logs CouchDB

Les logs se trouvent généralement :

- Windows : C:\CouchDB\log
- Linux : /var/log/couchdb/
- MacOS : /usr/local/var/log/couchdb/

Exercices Pratiques

1. Créez une vue qui compte les livres par auteur
2. Modifiez la vue des décennies pour inclure le nombre moyen de pages par décennie
3. Créez une vue qui liste les livres les plus récents en premier

Conclusion

MapReduce avec CouchDB offre une approche puissante pour l'analyse de données. Bien que nous ayons travaillé en mode centralisé, les concepts appris ici s'appliquent également aux environnements distribués.

Les prochaines étapes pourraient inclure :

- L'exploration des vues temporaires
- L'utilisation de vues pour des cas réels
- L'optimisation des performances
- L'intégration avec des applications web

Ressources Additionnelles

- Documentation officielle de CouchDB : <https://docs.couchdb.org/>
- Guide MapReduce : <https://guide.couchdb.org/draft/views.html>
- Communauté CouchDB : <https://blog.couchdb.org/>