

# CouchDB, PHP & PHPillow

<http://joind.in/1248>

Kore Nordmann  
<kore@php.net>  
@koredn

January 30, 2010

- ▶ Kore Nordmann, <kore@php.net>
- ▶ Long time PHP developer
- ▶ Regular speaker, author, etc.
- ▶ Studies computer science in Dortmund, currently writing thesis
- ▶ Active open source developer:
  - ▶ eZ Components (Graph, WebDav, Document), *Arbit*, PHPUnit, Torii, *PHPillow*, KaForkL, Image 3D, WCV, ...

Introduction

General

Structure

Consistency

API

Applications

QA



- ▶ Structure

- ▶ Structure
- ▶ Consistency

- ▶ Structure
- ▶ Consistency
- ▶ API

- ▶ Structure
- ▶ Consistency
- ▶ API
- ▶ Applications



Introduction

**General**

Structure

Consistency

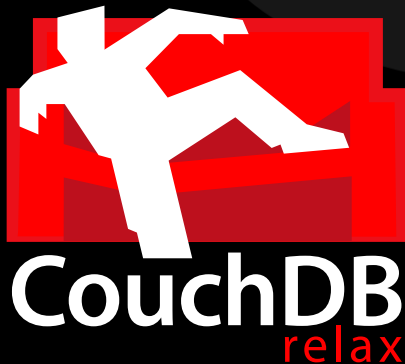
API

Applications

QA

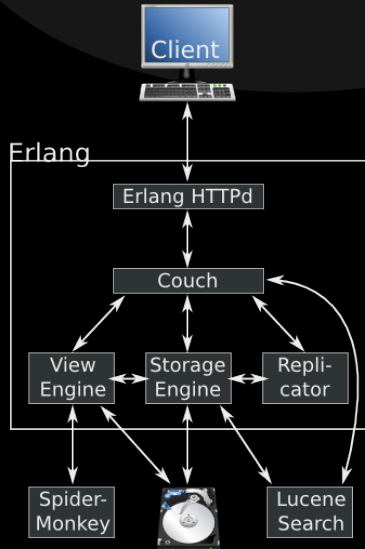


CouchDB  
relax

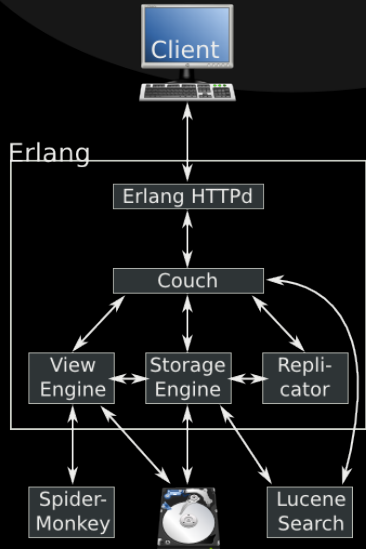


- ▶ Apache top-level project

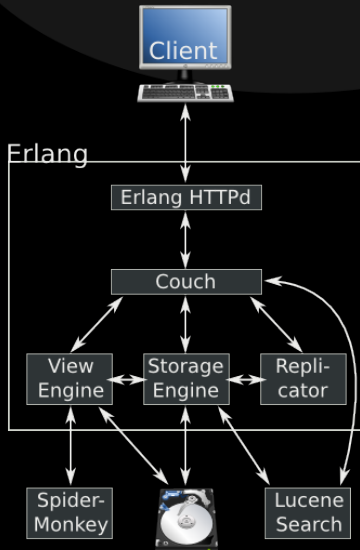
- ▶ Erlang/OTP virtual machine, developed by Ericsson



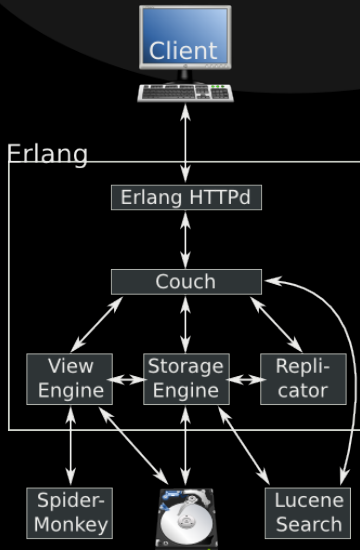
- ▶ Erlang/OTP virtual machine, developed by Ericsson
- ▶ Highly concurrent



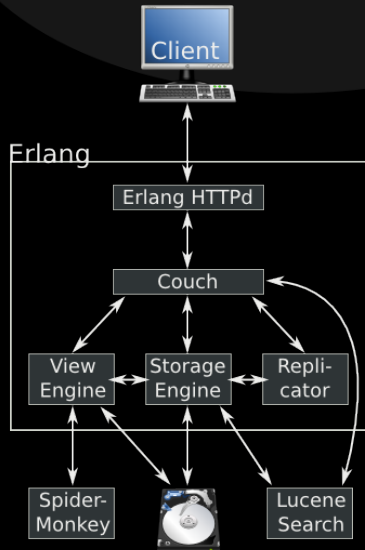
- ▶ Erlang/OTP virtual machine, developed by Ericsson
- ▶ Highly concurrent
- ▶ Scales nearly linearly with the amount of CPUs



- ▶ Erlang/OTP virtual machine, developed by Ericsson
- ▶ Highly concurrent
- ▶ Scales nearly linearly with the amount of CPUs
- ▶ High reliability (nine nines)



- ▶ Erlang/OTP virtual machine, developed by Ericsson
- ▶ Highly concurrent
- ▶ Scales nearly linearly with the amount of CPUs
- ▶ High reliability (nine nines)
- ▶ CouchDB is fast (enough)





Introduction

General

Structure

Consistency

API

Applications

QA

- ▶ No pre-defined structure (tables)

- ▶ No pre-defined structure (tables)
- ▶ Document based database

- ▶ No pre-defined structure (tables)
- ▶ Document based database
- ▶ Put in any JSON object you want

- ▶ No pre-defined structure (tables)
- ▶ Document based database
- ▶ Put in any JSON object you want
  - ▶ Even deep structures (arrays of objects)

- ▶ No pre-defined structure (tables)
- ▶ Document based database
- ▶ Put in any JSON object you want
  - ▶ Even deep structures (arrays of objects)
  - ▶ You may attach any number of files to documents

- ▶ Change document structure at any time

- ▶ Change document structure at any time
- ▶ No need for non-transaction-safe DDL



- ▶ Change document structure at any time
- ▶ No need for non-transaction-safe DDL
- ▶ Fits rapid development approaches with common customer requested changes to the data structure

- ▶ How to query such a mess?

- ▶ How to query such a mess?
  - ▶ Views are small scripts, run for all documents in a database

- ▶ How to query such a mess?
  - ▶ Views are small scripts, run for all documents in a database
  - ▶ Views are built iteratively, results stored in BTrees

- ▶ How to query such a mess?
  - ▶ Views are small scripts, run for all documents in a database
  - ▶ Views are built iteratively, results stored in BTrees
  - ▶ Mostly JavaScript, but also PHP, Ruby, Perl, Erlang, ...

- ▶ How to query such a mess?
  - ▶ Views are small scripts, run for all documents in a database
  - ▶ Views are built iteratively, results stored in BTrees
  - ▶ Mostly JavaScript, but also PHP, Ruby, Perl, Erlang, ...
  - ▶ A view may emit any number of key-value pairs for each document

- ▶ How to query such a mess?
  - ▶ Views are small scripts, run for all documents in a database
  - ▶ Views are built iteratively, results stored in BTrees
  - ▶ Mostly JavaScript, but also PHP, Ruby, Perl, Erlang, ...
  - ▶ A view may emit any number of key-value pairs for each document
  - ▶ Key and value may be any JSON structure

► Index all wiki documents by their title

```
1 function( doc )
2 {
3     if ( doc.type == "wiki_page" )
4     {
5         emit( [doc.namespace, doc.title], doc._id );
6     }
7 }
```

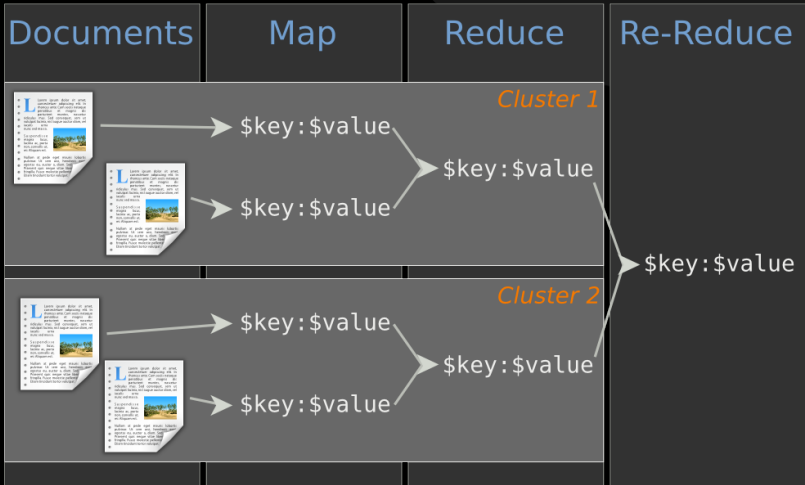


► Index all documents by their title

```
1 ["development_wiki", "BuildModuleDesign"] => "
  wiki_page-development_wiki_buildmoduledesign"
2 ["development_wiki", "CodingGuidelines"] => "
  wiki_page-development_wiki_codingguidelines"
3 ["development_wiki", "DiscussionProtocols"] => "
  wiki_page-development_wiki_discussionprotocols"
4 ["development_wiki", "ModuleDesign"] => "
  wiki_page-development_wiki_moduledesign"
5 ["development_wiki", "Protocol_08_02_07"] => "
  wiki_page-development_wiki_protocol_08_02_07"
6 ["development_wiki", "VCSModuleDesign"] => "
  wiki_page-development_wiki_vcsmoduledesign"
7 ...
```



- ▶ “MapReduce is a software framework introduced by Google to support distributed computing on large data sets on clusters of computers.” [Wik09]
- ▶ Used by CouchDB to implement views



- ▶ Map and reduce functions are custom

- ▶ Map and reduce functions are custom
- ▶ Reduce is optional, plain view serves as a document index

- ▶ Map and reduce functions are custom
- ▶ Reduce is optional, plain view serves as a document index
- ▶ Reduce may be applied to subsets of the documents

- ▶ Map and reduce functions are custom
- ▶ Reduce is optional, plain view serves as a document index
- ▶ Reduce may be applied to subsets of the documents
- ▶ Reduce may be grouped



Introduction

General

Structure

Consistency

API

Applications

QA

- ▶ Multi-Version Concurrency Control

- ▶ Multi-Version Concurrency Control
- ▶ All documents in the database are versioned

- ▶ Multi-Version Concurrency Control
- ▶ All documents in the database are versioned
  - ▶ Don't use it for application level document versioning

- ▶ Multi-Version Concurrency Control
- ▶ All documents in the database are versioned
  - ▶ Don't use it for application level document versioning
- ▶ Updates and deletes need to specify the revision ID

- ▶ Multi-Version Concurrency Control
- ▶ All documents in the database are versioned
  - ▶ Don't use it for application level document versioning
- ▶ Updates and deletes need to specify the revision ID
- ▶ Changing outdated documents result in conflicts

- ▶ There is no ensured inter document consistency in CouchDB

ents:

- ▶ There is no ensured inter document consistency in CouchDB
- ▶ Different possibilities of relating documents:



- ▶ There is no ensured inter document consistency in CouchDB
- ▶ Different possibilities of relating documents:
  - ▶ List IDs of related documents in document (n:m)

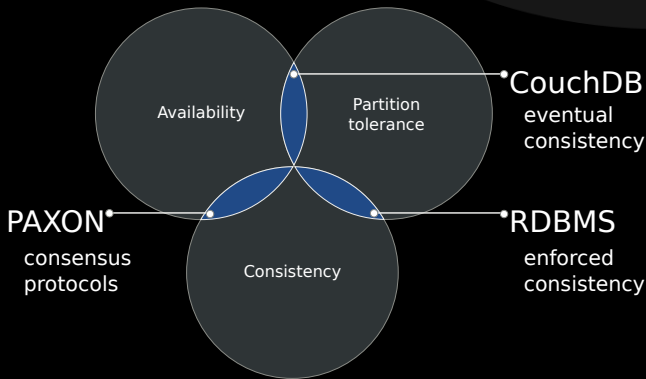
- ▶ There is no ensured inter document consistency in CouchDB
- ▶ Different possibilities of relating documents:
  - ▶ List IDs of related documents in document (n:m)
  - ▶ ... both directions are feasible

- ▶ There is no ensured inter document consistency in CouchDB
- ▶ Different possibilities of relating documents:
  - ▶ List IDs of related documents in document (n:m)
  - ▶ ... both directions are feasible
  - ▶ Embed the whole related document (1:n)

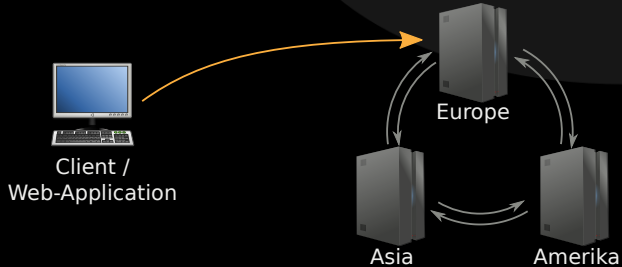
- ▶ There is no ensured inter document consistency in CouchDB
- ▶ Different possibilities of relating documents:
  - ▶ List IDs of related documents in document (n:m)
  - ▶ ... both directions are feasible
  - ▶ Embed the whole related document (1:n)
- ▶ Solution depends on update-ratio

```
1  { "type": "discussion",
2    "wiki": "wiki-Start",
3    "text": "...",
4    "comments": [
5      { "comment": "..."}],
6  ],
7  "creator": "user-8247",
8  }
```

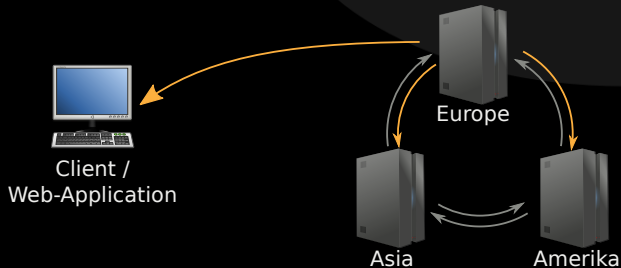
- ▶ The CAP theorem, read more in “CouchDB: The Definitive Guide” [JCA09]



- ▶ CouchDB employs “Eventual Consistency”

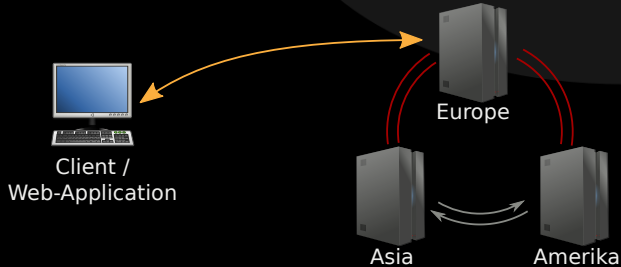


► Structure your



- ▶ Delayed, triggered synchronization (push, pull)
  - ▶ Deterministic (manual) conflict resolution on replication on all nodes

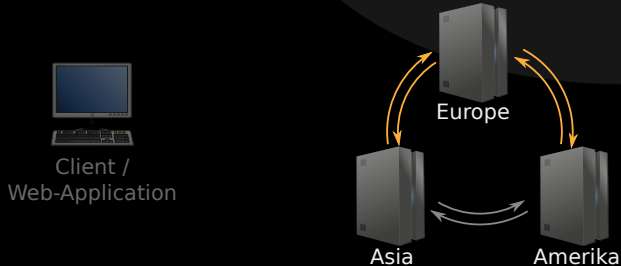
▶ Structure your



- ▶ Delayed, triggered synchronization (push, pull)
  - ▶ Deterministic (manual) conflict resolution on replication on all nodes

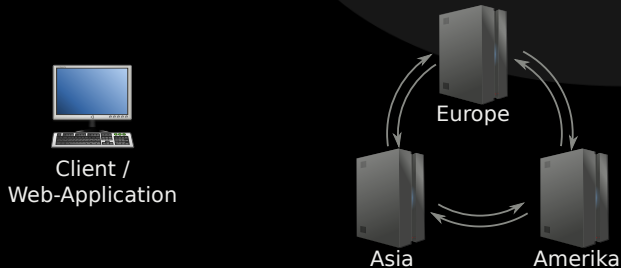
▶ Structure your...





- ▶ Delayed, triggered synchronization (push, pull)
  - ▶ Deterministic (manual) conflict resolution on replication on all nodes
- ▶ Scales well for seldom concurrent writes

▶ Structure your...



- ▶ Delayed, triggered synchronization (push, pull)
  - ▶ Deterministic (manual) conflict resolution on replication on all nodes
- ▶ Scales well for seldom concurrent writes
  - ▶ Structure your documents accordingly

Introduction

General

Structure

Consistency

**API**

Applications

QA

- ▶ RESTful HTTP access

- ▶ RESTful HTTP access
- ▶ HTTP is available on “all” platforms natively
  - ▶ No PHP extension required
  - ▶ Just use PHPs HTTP stream wrapper, pecl/http or curl

- ▶ RESTful HTTP access
- ▶ HTTP is available on “all” platforms natively
  - ▶ No PHP extension required
  - ▶ Just use PHPs HTTP stream wrapper, pecl/http or curl
- ▶ You can use all your known HTTP middleware
  - ▶ Reverse proxies for scaling reads (Varnish, Squid)
  - ▶ Simple custom proxy configuration for direct “AJAX” access

```
1 $ curl -i -X PUT http://localhost:5984/phpbnl_wiki
2
3 HTTP/1.1 201 Created
4 Server: CouchDB/0.10.0 (Erlang OTP/R13B)
5 Location: http://localhost:5984/phpbnl_wiki
6 Date: Fri, 13 Jan 2009 14:07:57 GMT
7 Content-Type: text/plain; charset=utf-8
8 Content-Length: 12
9 Cache-Control: must-revalidate
10
11 {"ok": true}
```

```
1 $ curl -i -X PUT http://localhost:5984/phpbnl_wiki/  
   Start --data '{"name": "Start", "text": "Hello  
   World!"}'  
2  
3 HTTP/1.1 201 Created  
4 Server: CouchDB/0.10.0 (Erlang OTP/R13B)  
5 Location: http://localhost:5984/phpbnl_wiki/Start  
6 Etag: "1-6bfd4885b6c62bb5169a19d5a81927e3"  
7 Date: Fri, 13 Jan 2009 14:14:55 GMT  
8 Content-Type: text/plain; charset=utf-8  
9 Content-Length: 68  
10 Cache-Control: must-revalidate  
11  
12 {"ok":true,"id":"Start","rev":"1-6  
   bfd4885b6c62bb5169a19d5a81927e3"}
```



```
1 $ curl -i -X GET http://localhost:5984/phpbnl_wiki/  
   Start  
2  
3 HTTP/1.1 200 OK  
4 Server: CouchDB/0.10.0 (Erlang OTP/R13B)  
5 Etag: "1-6bfd4885b6c62bb5169a19d5a81927e3"  
6 Date: Fri, 13 Jan 2009 14:15:48 GMT  
7 Content-Type: text/plain; charset=utf-8  
8 Content-Length: 97  
9 Cache-Control: must-revalidate  
10  
11 {"_id": "Start", "_rev": "1-6  
    bfd4885b6c62bb5169a19d5a81927e3", "name": "Start",  
    "text": "Hello World!"}
```

- ▶ Simple database based access restrictions

- ▶ Simple database based access restrictions
- ▶ Using HTTP plain auth

- ▶ Simple database based access restrictions
- ▶ Using HTTP plain auth
- ▶ More fine grained access control is under discussion

- ▶ Object-oriented client for CouchDB
- ▶ PHP  $\geq$  5.2 since last release (5.3 only before)
- ▶  $>96\%$  test coverage

- ▶ Object-oriented client for CouchDB
- ▶ PHP  $\geq$  5.2 since last release (5.3 only before)
- ▶  $>96\%$  test coverage
- ▶ Still in alpha state

- ▶ Object-oriented client for CouchDB
- ▶ PHP  $\geq$  5.2 since last release (5.3 only before)
- ▶  $>96\%$  test coverage
- ▶ Still in alpha state
  - ▶ Since CouchDB just got “beta” recently, and no new release was required.

## ► Document creation example

```
1 // Create a document
2 $doc = new phpillowUserDocument();
3 $doc->login = 'kore';
4 $doc->name = 'Kore_Nordmann';
5 $doc->data = array(
6     'mail' => "kore@php.net",
7     // ...
8 );
9 $id = $doc->save();
10
11 // Fetch a document by ID
12 $doc = new phpillowUserDocument( $id );
```



Introduction

General

Structure

Consistency

API

**Applications**

QA

- ▶ CouchDB allows you to attach files to documents

- ▶ CouchDB allows you to attach files to documents
- ▶ Files are replicated

- ▶ CouchDB allows you to attach files to documents
- ▶ Files are replicated
- ▶ You can serve full Web-Applications from a CouchDB
- ▶ Deploy using PUSH-replication

- ▶ Mirror database into userspace

- ▶ Mirror database into userspace
- ▶ Offline usage and synchronization of Browser applications

- ▶ Mirror database into userspace
- ▶ Offline usage and synchronization of Browser applications
- ▶ Mozilla develops a JavaScript implementation of the CouchDB API [Moz09]

- ▶ Ubuntu One uses CouchDB
- ▶ Synchronize contacts & date between nodes, or to a server



- ▶ Ubuntu One uses CouchDB
- ▶ Synchronize contacts & date between nodes, or to a server
- ▶ Yes, all Ubuntu Karmic users already have a CouchDB running

- ▶ Arbit uses CouchDB for issue tracking, wiki, FAQ and more
- ▶ Other applications: `http://wiki.apache.org/couchdb/CouchDB_in_the_wild`

- ▶ CouchDB is fast (enough)
- ▶ Document oriented approach allows new application development approaches
- ▶ CouchDB scales really well, horizontally and vertically
- ▶ CouchDB fits web applications really well

Introduction

General

Structure

Consistency

API

Applications

QA

- ▶ Apache CouchDB: <http://couchdb.org/>
- ▶ Free CouchDB book: <http://books.couchdb.org/relax/>
- ▶ PHPillow: <http://arbitracker.org/phpillow.html>
- ▶ Arbit: <http://arbitracker.org/>

- ▶ Open questions?
- ▶ Further remarks?
- ▶ Contact
  - ▶ Mail: <kore@php.net>
  - ▶ Web: <http://kore-nordmann.de/> (Slides will be available here soonish)
  - ▶ Twitter: <http://twitter.com/koredn>
  - ▶ Comment: <http://joind.in/1248>

- [JCA09] Noah Slater J. Chris Anderson, Jan Lehnardt, *Couchdb: The definitive guide*, O'Reilly Media, Inc., 2009.
- [Moz09] Mozilla, *Browsercouch documentation*, November 2009.
- [Wik09] Wikipedia, *Mapreduce — wikipedia, the free encyclopedia*, 2009, [Online; accessed 27-August-2009].

- ▶ Index all documents by all their words

```
1 function( doc ) {
2   if ( doc.type == "tracker_issue" ) {
3     // Simple word indexing, does not respect overall
4     // occurrences of words,
5     // stopwords, different word separation characters,
6     // or word variations.
7     var text = doc.title.replace( /[\s:.,!?-]+/g, " " )
8     +
9     doc.text.replace( /[\s:.,!?-]+/g, " " );
10    var words = text.split( " " );
11    for ( var i = 0; i < words.length; ++i ) {
12      value = {};
13      value[doc._id] = 1;
14      emit( words[i].toLowerCase(), value );
15    }
16  }
17 }
```



► Index all documents by all their words

```
1  ...
2  "a"           => {tracker_issue -8: 1}
3  "a"           => {tracker_issue -8: 1}
4  "a"           => {tracker_issue -8: 1}
5  "a"           => {tracker_issue -8: 1}
6  "a"           => {tracker_issue -81: 1}
7  "a"           => {tracker_issue -83: 1}
8  "a"           => {tracker_issue -83: 1}
9  "able"        => {tracker_issue -39: 1}
10 "able"         => {tracker_issue -56: 1}
11 "able"         => {tracker_issue -73: 1}
12 "able"         => {tracker_issue -80: 1}
13 "about"        => {tracker_issue -24: 1}
14 "about"        => {tracker_issue -43: 1}
15 "about"        => {tracker_issue -85: 1}
16  ...
```

## ► Reduce by word count

```
1 function( keys , values ) {
2     var count = {};
3     for ( var i in values ) {
4         for ( var id in values[i] ) {
5             if ( count[id] ) {
6                 count[id] = values[i][id] + count[id];
7             } else {
8                 count[id] = values[i][id];
9             }
10        }
11    }
12    return count;
13 }
```

► Index all documents by all their words

```
1  ...
2  "a"           => {
3                      tracker_issue -68: 6,
4                      tracker_issue -66: 6,
5                      tracker_issue -22: 4,
6                      tracker_issue -63: 3,
7                      tracker_issue -60: 2,
8                      tracker_issue -35: 2,
9                      tracker_issue -34: 1,
10                     tracker_issue -31: 1,
11                     ...
12                     }
13  "able"       => { tracker_issue -86: 1, tracker_issue -80:
14                     1, tracker_issue -73: 1, tracker_issue -56: 1,
15                     tracker_issue -39: 1 }
16  "about"      => { tracker_issue -85: 1, tracker_issue -43:
17                     1, tracker_issue -24: 1 }
18  ...
```