

# Dirty NoSQL

How simple is your database?

# About



Contributor



node.js driver



Felix Geisendörfer



Co-founder



node.js driver



**Dirty**

# Dirty



**CouchDB**  
relax

JavaScript views



Dirty



In-memory

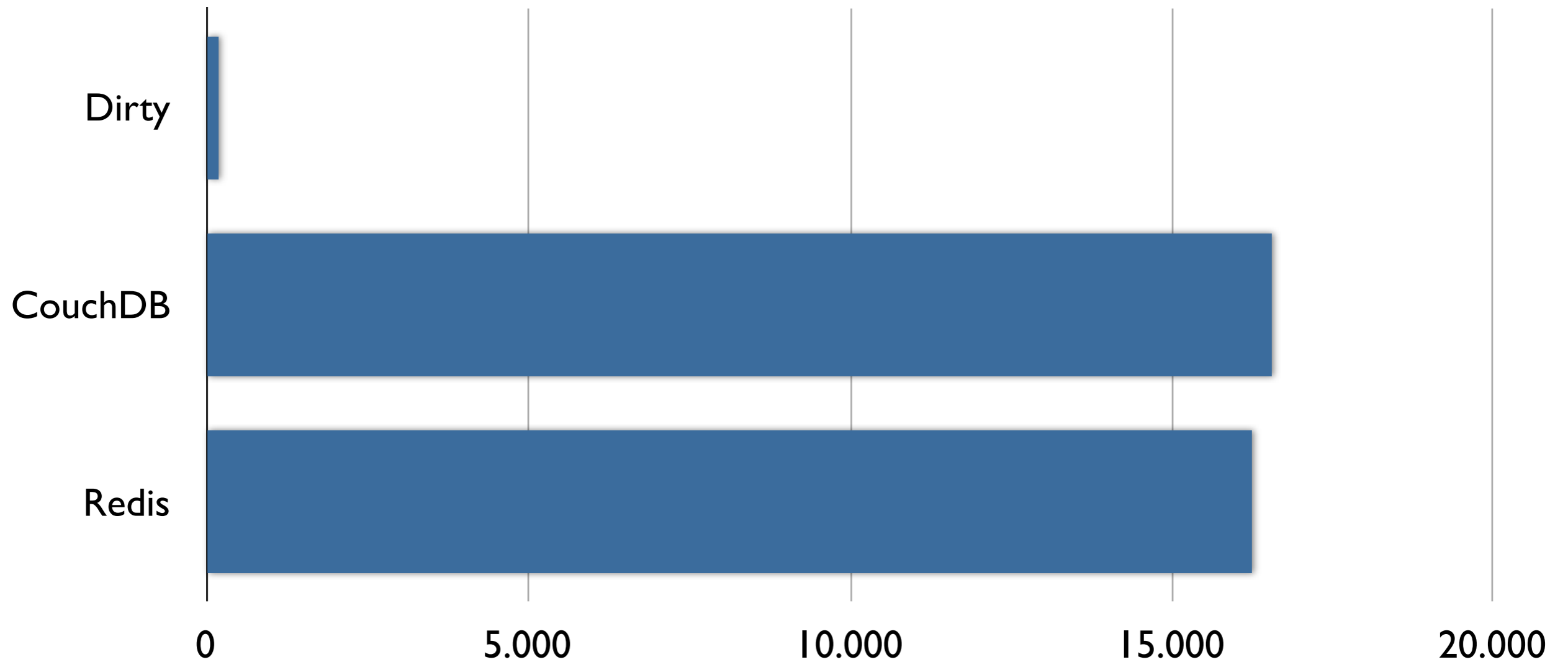
# Design choices



**Non-blocking I/O FTW**

# 150 lines of code

Lines of code





# Append-only JSON log

```
$ cat dirty.db
```

```
{"key": "first", "val": {"foo": "bar"}}  
{"key": "second", "val": "A string"}  
{"key": "a-number", "val": 23}
```

# No Networking

```
var db = require('dirty')('test.db');
```

```
db.set('foo', 'bar');
```

```
db.get('foo'); // => bar
```

# No Networking

```
var db = require('dirty')('languages.db');  
  
db.set('javascript', {age: 15});  
db.set('python', {age: 19});  
db.set('perl', {age: 23});  
  
db.forEach(function(key, doc) {  
  console.log('%s is %d years old.', key, doc.age);  
});
```

**Performance**

# Benchmarks

**Do your own!**

`dirty.get()`

**50 MHz**

(50 million / s)

v8: 160 MHz

`dirty.set()`

**5 MHz**

(5 million / s)

v8: 12 MHz

# dirty.set()

With flushing to disk

Numbers

**200 kHz**

(200-thousand / s)

256 byte string

**70 kHz**

(70-thousand / s)



# dirty.forEach()

## **4.5 MHz**

(4.5 million / s)

v8: 33 MHz

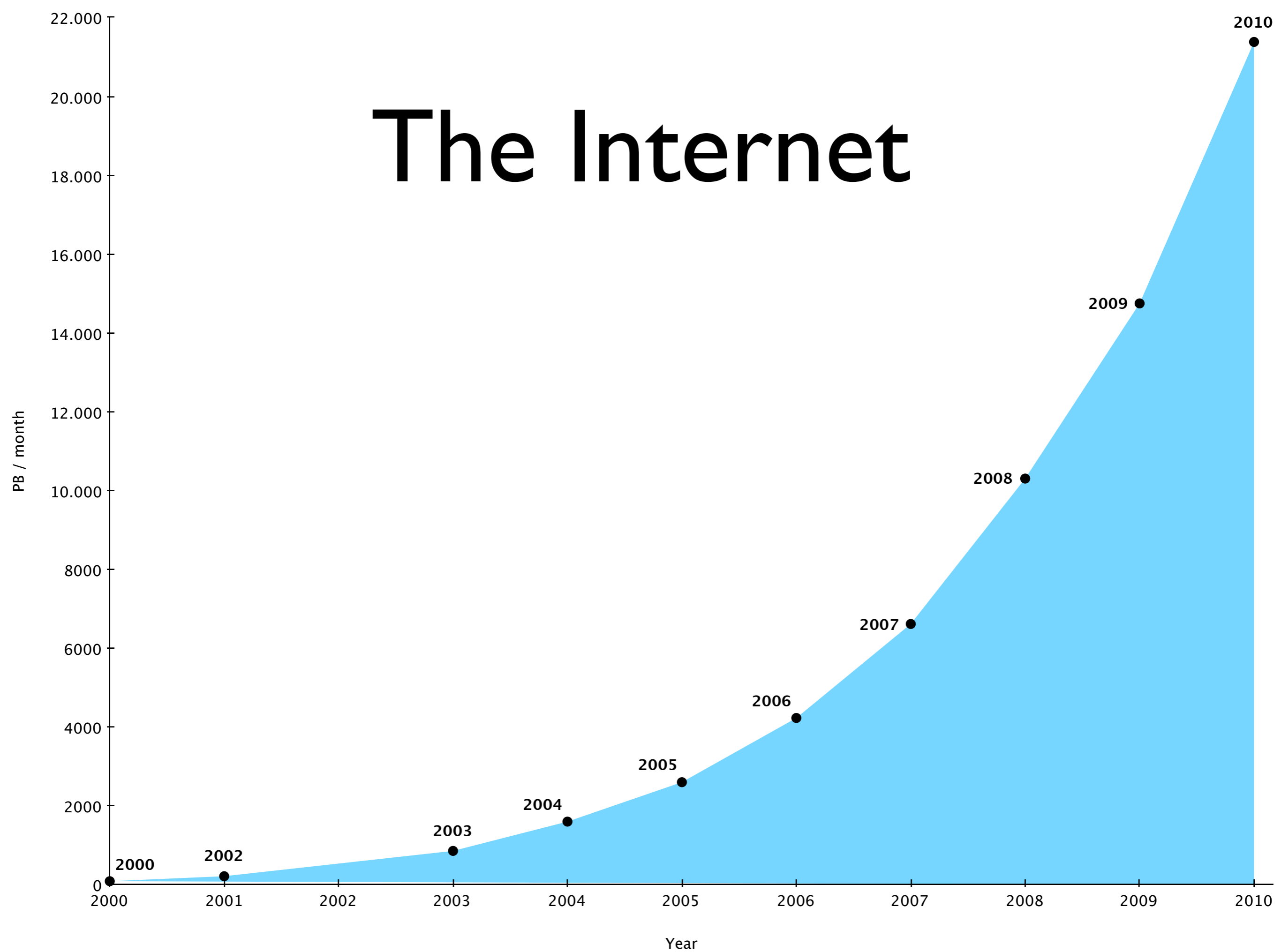
# The Wall

- Dirty is a wonderful database as long as you have  $< 1$  million records
- After that, you hit “The Wall”

# Scaling beyond

Possibilities for future node.js based databases

# The Internet



**22 exabyte / month**

**(that is 22 billion gigabytes)**

# Flexible guarantees

```
db.set('my-key', 'my-value', function(err) {  
  if (err) throw err;  
  
  console.log('Record written to disk.');});  
  
console.log('Record written to memory.');
```

# Memory / Disk Hybrids

- “Memcached” built into your database
- Better knowledge about your data than any general purpose algorithm

# Replication

- Node.js = perfect for streaming between instances
- Node could “hold” the connection if a not-yet replicated key is being requested



# Web Services

- Node.js could act as a proxy for different database backends (written in node or not)
- One could query 3rd party services for information

# Questions?



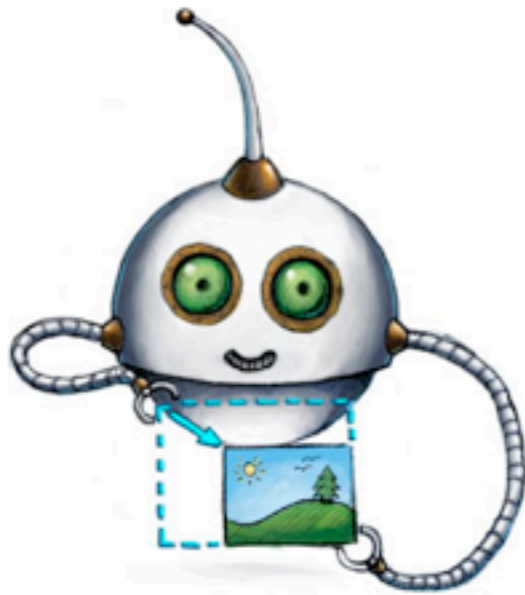
@felixge / [felix@debuggable.com](mailto:felix@debuggable.com)

# Get it

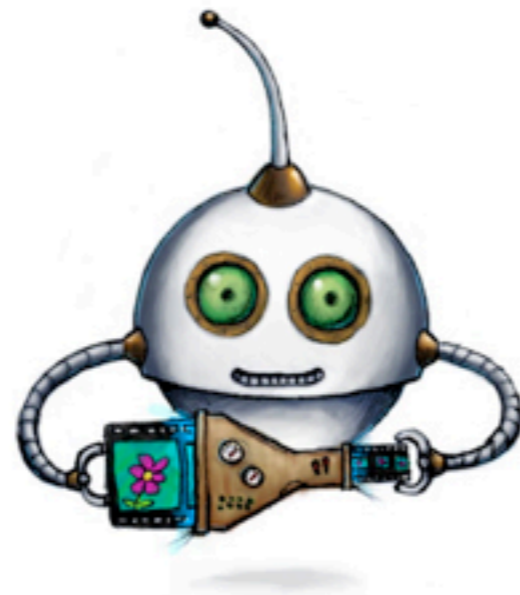
```
$ npm install dirty
```

or

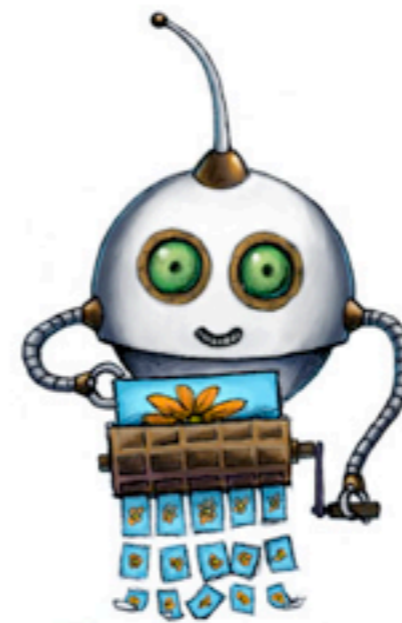
<http://github.com/felixge/node-dirty>



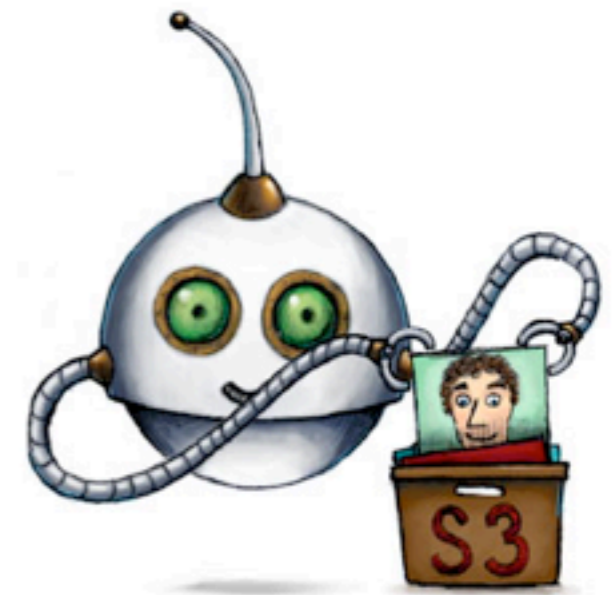
Resize images



Encode videos



Extract thumbnails



Store in S3

# Don't hit the wall



# Questions?



@felixge / [felix@debuggable.com](mailto:felix@debuggable.com)

# Sources

- Internet traffic numbers: [http://en.wikipedia.org/wiki/Internet\\_traffic](http://en.wikipedia.org/wiki/Internet_traffic)
- Wonderful dino-rider vs. volcano image: Jesse Star / [http://www.geekologie.com/2009/10/dinorider\\_geekologie\\_writer\\_vs.php](http://www.geekologie.com/2009/10/dinorider_geekologie_writer_vs.php)

# Memory overhead

(for setting numeric key / values)

- 20mb overhead / 1 million records
- 20 bytes / key