MongoDB @ SourceForge

Mark Ramm

We had a problem

six weeks

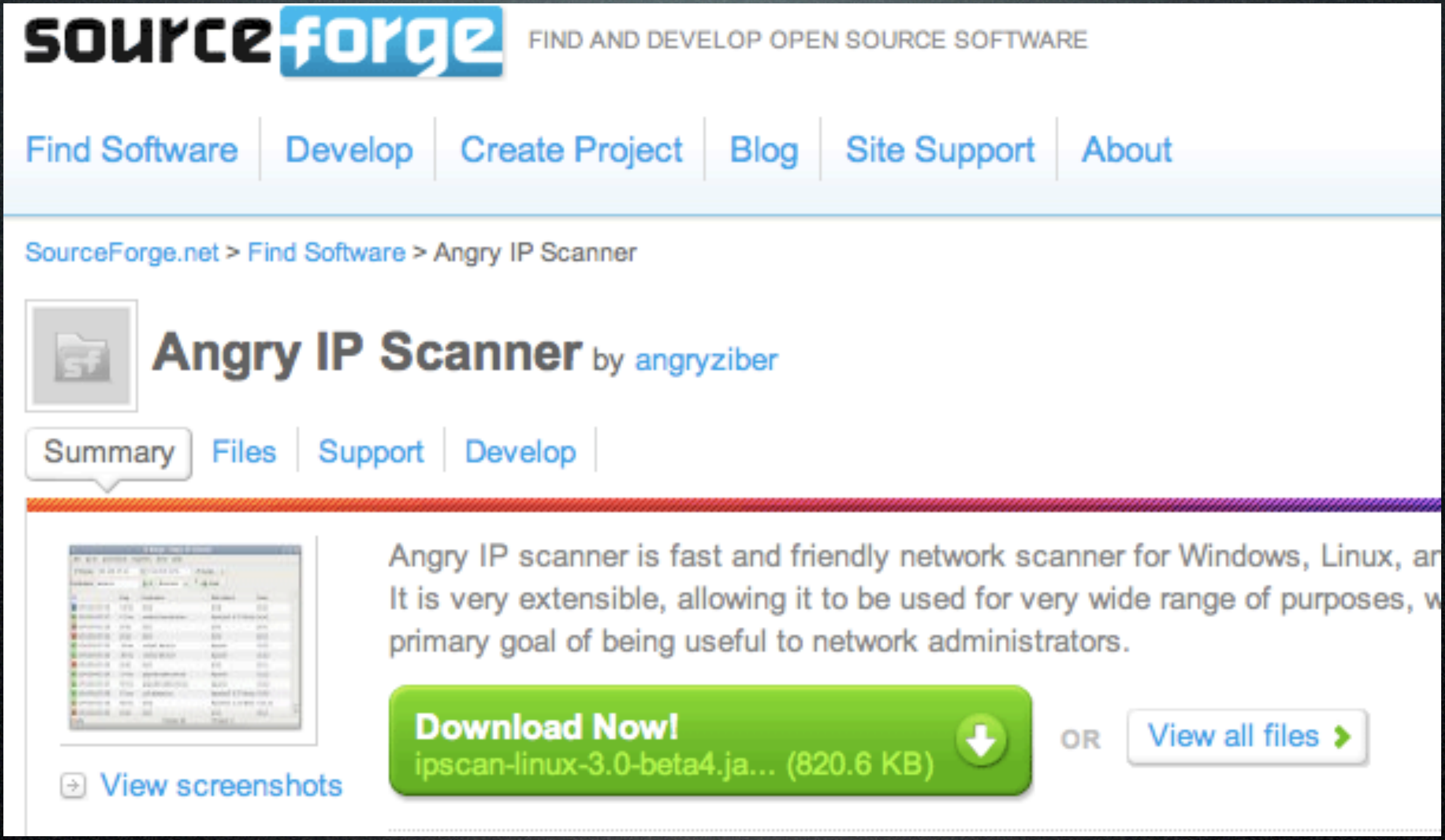
the other sourceforge

over 90% of traffic

Design goals

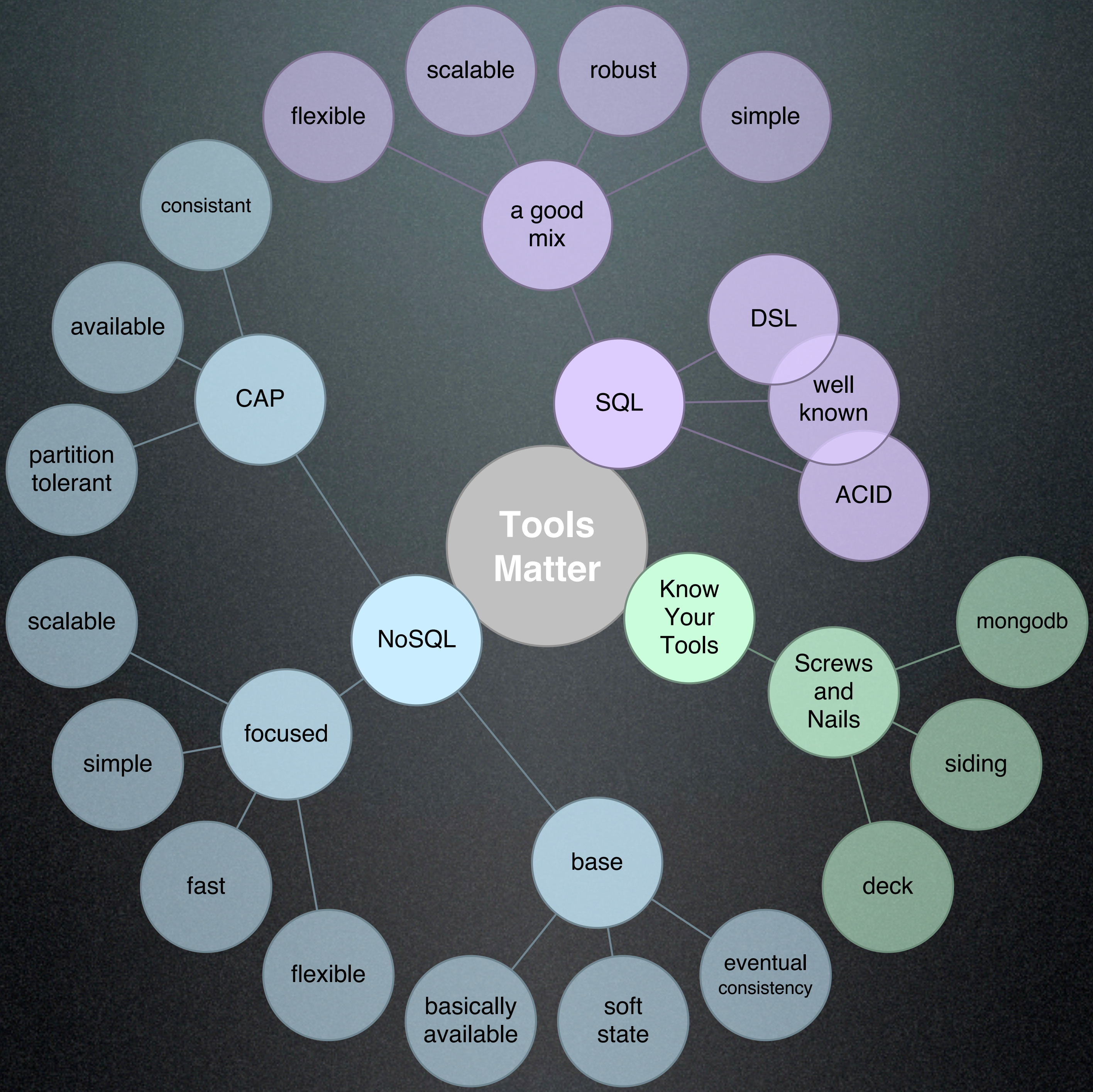
- Improve Usability
  (more data, more dynamic pages)
- Improve Performance
- Improve Reliability

Big Green Button


sourceforge
FIND AND DEVELOP OPEN SOURCE SOFTWARE
Find Software
Develop
Create Project
Blog
Site Support
About
SourceForge.net > Find Software > Angry IP Scanner

Angry IP Scanner by angryziber

Angry IP scanner is fast and friendly network scanner for Windows, Linux, ar
It is very extensible, allowing it to be used for very wide range of purposes, w
primary goal of being useful to network administrators.



- Tools Matter
  - SQL
    - a good mix
      - flexible
      - scalable
      - robust
      - simple
    - DSL
    - well known
    - ACID
  - NoSQL
    - CAP
      - consistant
      - available
      - partition tolerant
    - focused
      - scalable
      - simple
      - fast
      - flexible
    - base
      - basically available
      - soft state
      - eventual consistency
  - Know Your Tools
    - Screws and Nails
      - mongodb
      - siding
      - deck

ACID

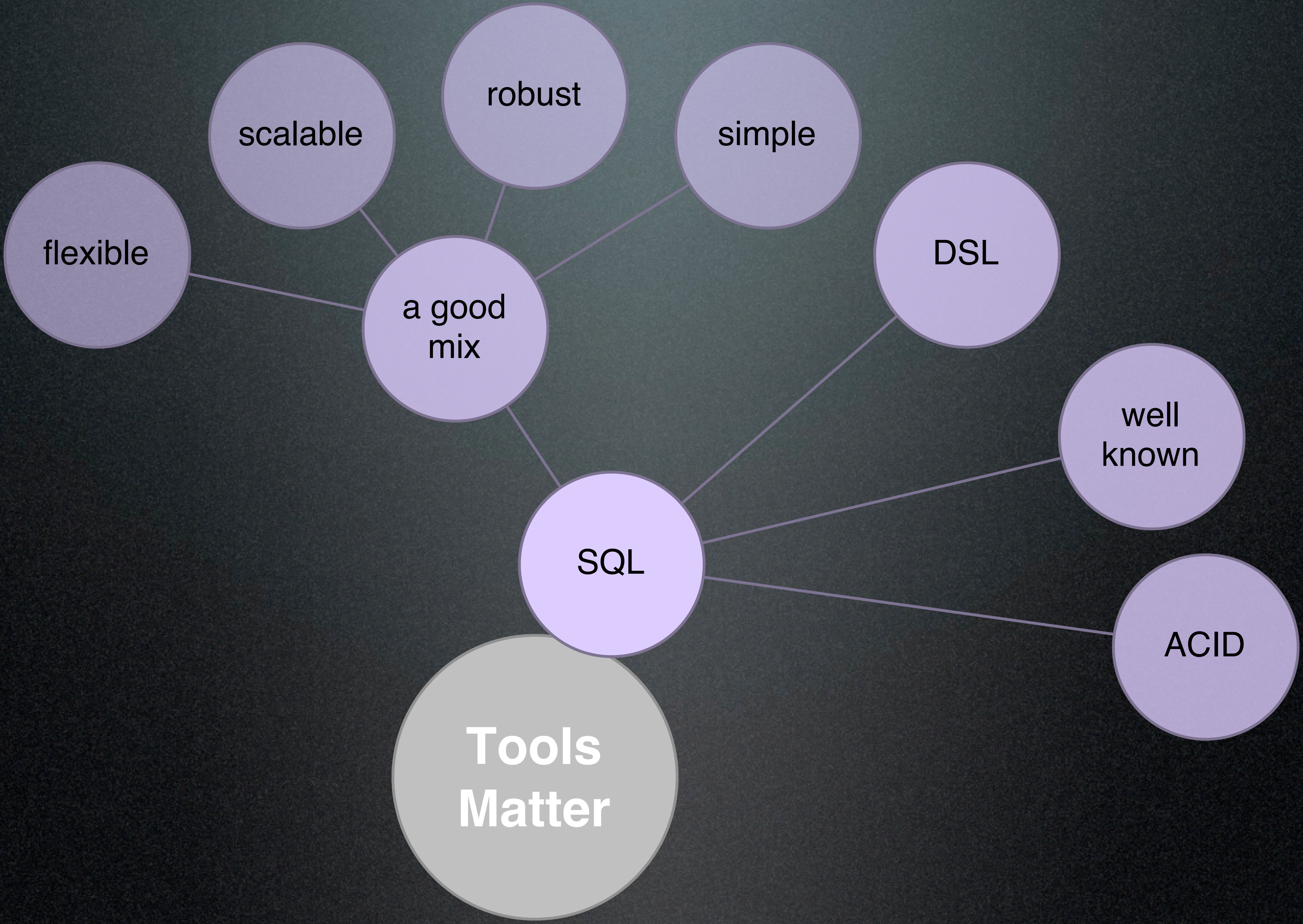
http://www.flickr.com/photos/di4b0liko/2292648884/

Why I NEED Relational

- I have to have ACID because...
  - It's financial data (need consistency)
  - My data is relational

BULLSHIT

But...



Tools Matter

SQL

a good mix

flexible

scalable

robust

simple

DSL

well known

ACID

NoSQL

hadoop

HYPERTABLE

HBASE

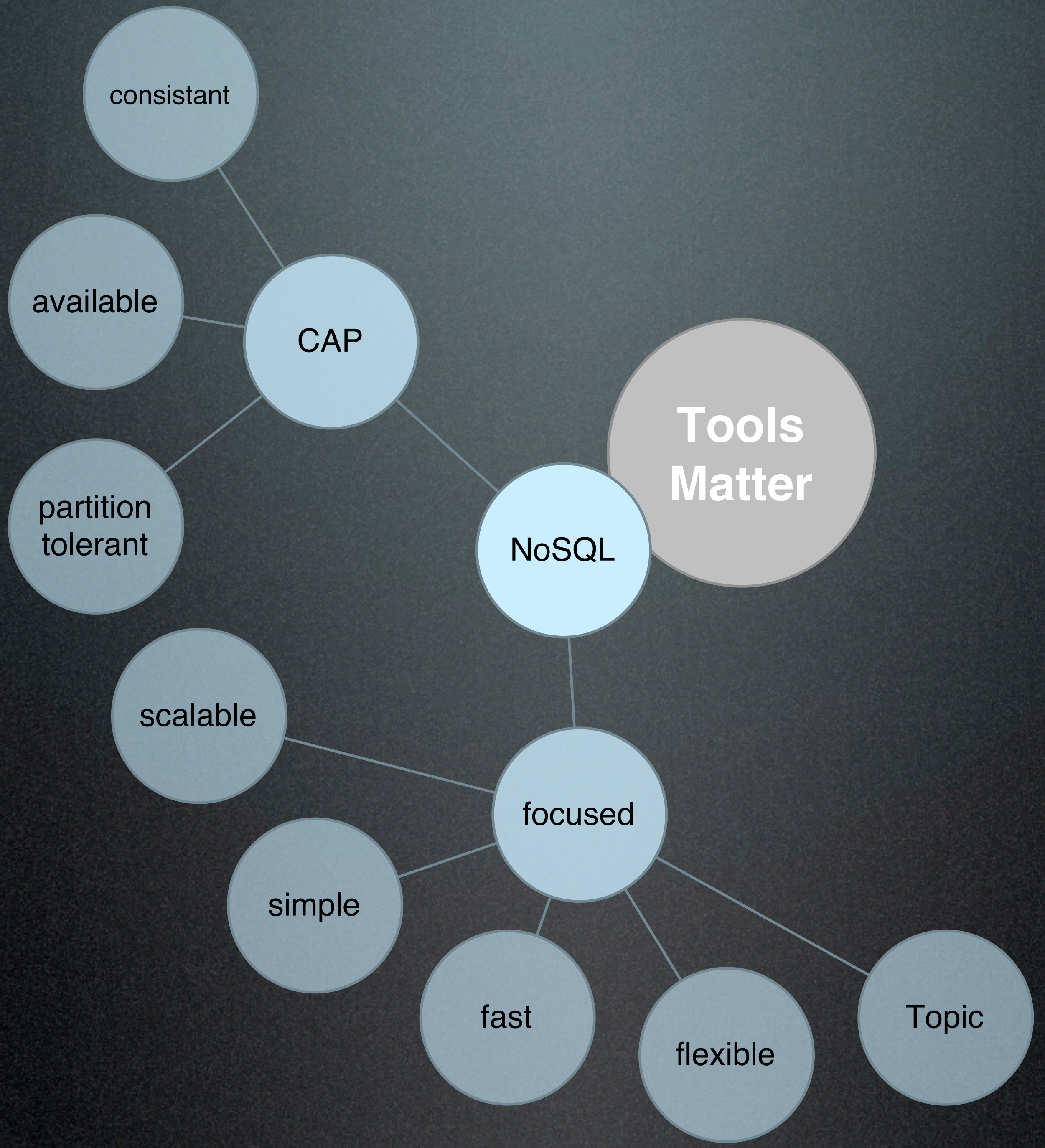
CouchDB
relax

Redis

mongoDB

Cassandra

Tokyo
Cabinet 8192PiB


consistant
available
CAP
partition tolerant
Tools Matter
NoSQL
scalable
focused
simple
fast
flexible
Topic

CAP

http://www.cs.berkeley.edu/~brewer/cs262b-2004/PODC-keynote.pdf

http://www.flickr.com/photos/beigephotos/900974545/

blah, blah,
blah, blah,
blah

typology of NoSQL

- key/value store
- distributed key/value stores
- column oriented stores
- map-reduce store/system
- document oriented store
- graph oriented stores

Enough theory

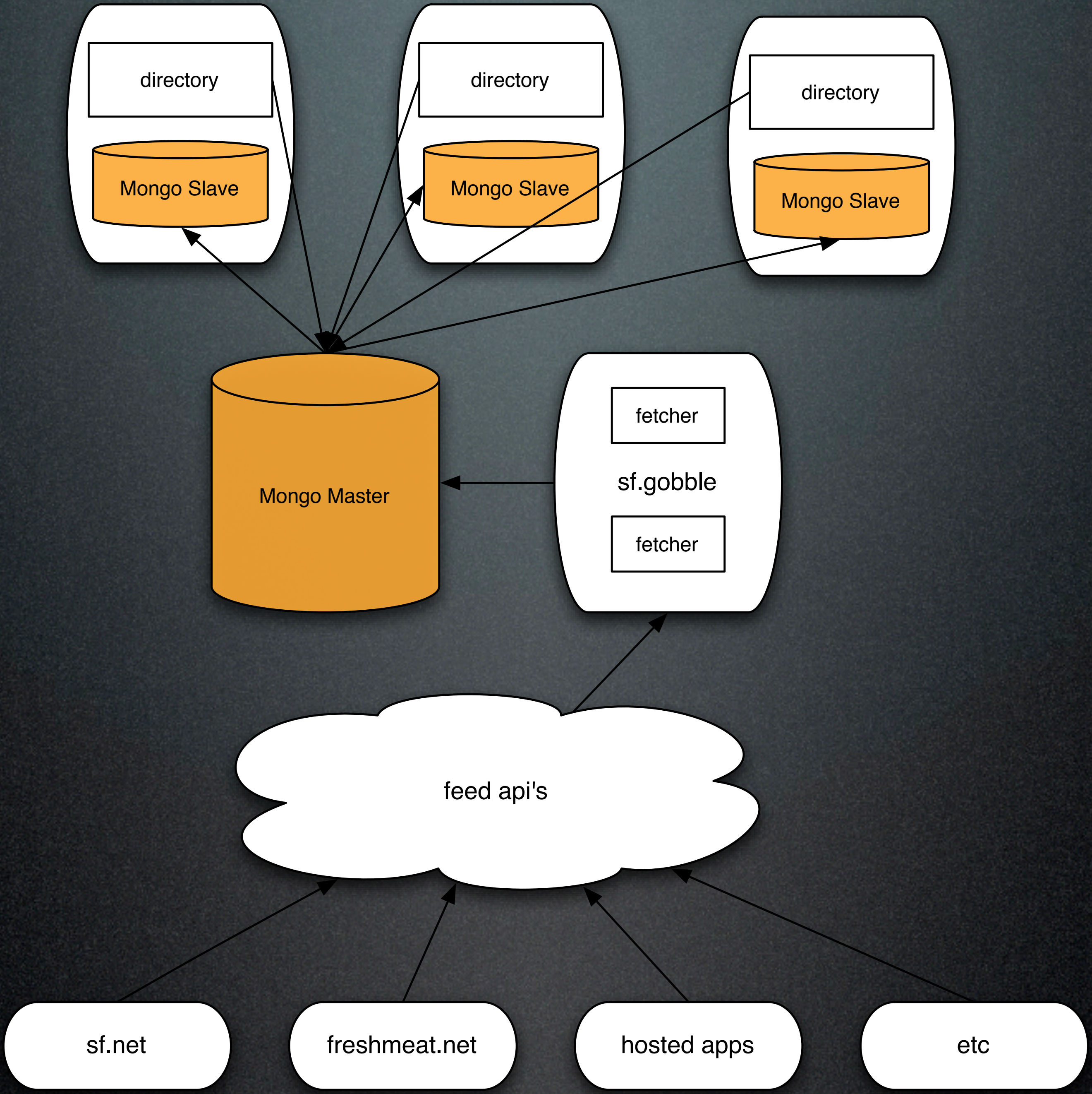
We had documents

```
{        'source': 'sf.net',
   'shortname': 'azureus',
   'related': [
               'shortname': 'foo', 'description':'bar', 'screenshots':[...],
   'project_url': 'http://asdf', 'name'; 'Azureus',]
            'sf_id': 5383,
            'sf_piwik_siteid': '2',

            'name': 'Azureus',
            'doap': 'http://sourceforge.net/api/project/name/azureus/doap',
            'created': datetime.datetime(2003, 6, 24, 0, 0),
            'homepage': 'http://azureus.sourceforge.net',
            'project_url': 'http://sourceforge.net/projects/combined-for-all-data',
            'resources': {
                'news': [{'feed': 'http://sourceforge.net/api/news/index/...,
                          'name': 'News',
                          'url': 'http://sourceforge.net/news/?group_id=84122'}],
                'forums': [{'feed': 'http://sourceforge.net/api/post/index/.../rss',
                            'name': 'Help',
                            'url': 'http://sourceforge.net/forum/forum.php...',
                            'item_count': 1,},
                           {'feed': 'http://sourceforge.net/api/post/index.../rss',
                            'name': 'Discussion',
                            'url': 'http://sourceforge.net/forum/forum.php/...',
                            'item_count': 28216,}],
```

we did get some “lucky breaks”

- consistency not critical
- scale reads, not writes

We wanted replication


directory
Mongo Slave
directory
Mongo Slave
directory
Mongo Slave
Mongo Master
fetcher
sf.gobble
fetcher
feed api's
sf.net
freshmeat.net
hosted apps
etc

We didn’t have
a lot of time

MongoDB has a query language

```
select * from document where x=3
and y="foo"
```

```
db.things.find(
        { x : 3, y : "foo" }
        );
```

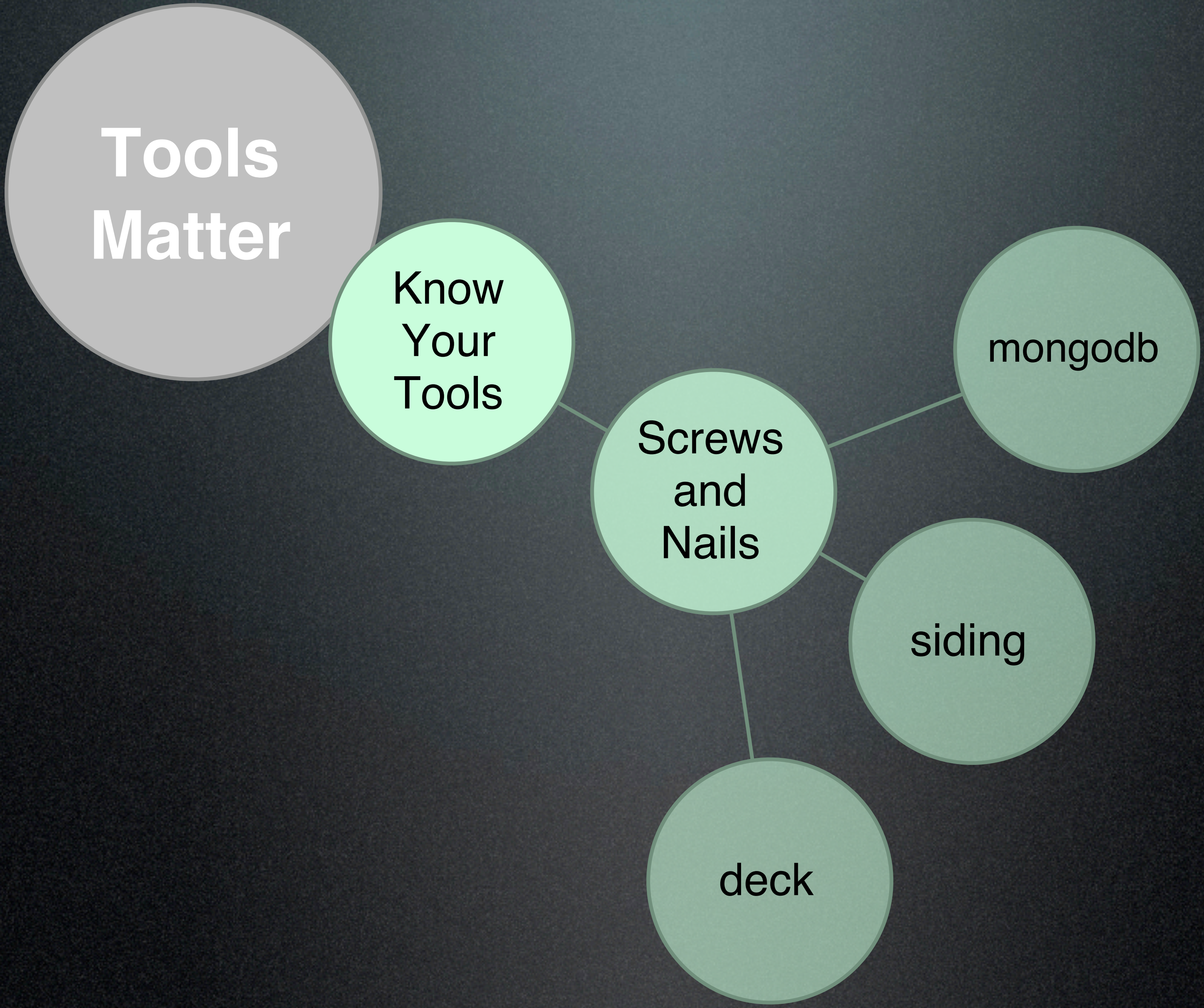
partial updates

- $inc
- $set
- $unset
- $push
- $pushAll
- $addToSet
- $pop
- $pull
- $pullAll

```
{ $inc : { field : value } }
```

Conditional Updates

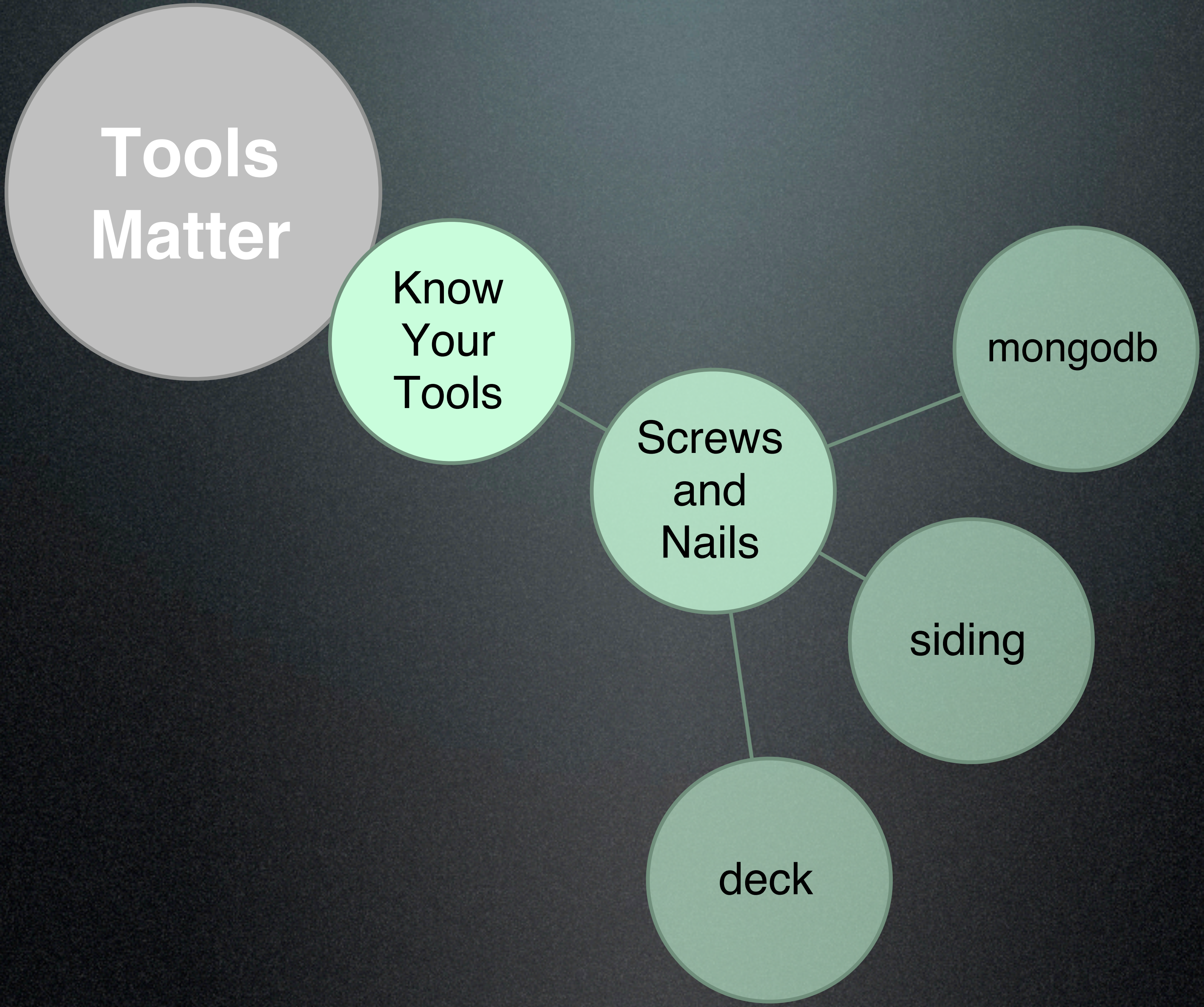
```
db.people.update( { name:"Joe" },
{ $inc: { x:1, y:1 } }, true
{        'source': 'sf.net',
         'shortname': 'azureus',
         'related': [
               'shortname': 'foo', 'description':'bar', 'screenshots':[...],
            'project_url': 'http://asdf', 'name'; 'Azureus',]
            'sf_id': 5383,
            'sf_piwik_siteid': '2',

            'name': 'Azureus',
            'doap': 'http://sourceforge.net/api/project/name/azureus/doap',
            'created': datetime.datetime(2003, 6, 24, 0, 0),
            'homepage': 'http://azureus.sourceforge.net',
            'project_url': 'http://sourceforge.net/projects/combined-for-all-data',
            'resources': {
                'news': [{'feed': 'http://sourceforge.net/api/news/index/...,
                          'name': 'News',
                          'url': 'http://sourceforge.net/news/?group_id=84122'}],
                'forums': [{'feed': 'http://sourceforge.net/api/post/index/.../rss',
                            'name': 'Help',
                            'url': 'http://sourceforge.net/forum/forum.php...',
                            'item_count': 1,},
                           {'feed': 'http://sourceforge.net/api/post/index.../rss',
                            'name': 'Discussion',
                            'url': 'http://sourceforge.net/forum/forum.php/...',
                            'item_count': 28216,}],
    ],
},
'screenshot_page': 'http://sourceforge.net/project/screenshots.php?group_id=93438',
'screenshots': [{'url': 'http://sourceforge.net/dbimage.php?id=208967',
                 'thumb': 'http://sourceforge.net/dbimage.php?id=208966',
                 "name" : "Table structure view"},
                {'url': 'http://sourceforge.net/dbimage.php?id=99723',
                 'thumb': 'http://sourceforge.net/dbimage.php?id=99722'}],

# ID,shortname,description only present for SF projects.
'categories': {'Development Status': [{'description': '4 - Beta',
                                        'id': 10,
                                        'name': '4 - Beta',
                                        'shortname': 'beta'}],
                'Intended Audience': [{'description': 'Developers',
                                       'id': 3,
                                       'name': 'Developers',
                                       'shortname': 'developers'},
                                      {'description': 'End Users/Desktop',
                                       'id': 2,
                                       'name': 'End Users/Desktop',
                                       'shortname': 'endusers'},
                                      {'description': 'System Administrators',
                                       'id': 4,
                                       'name': 'System Administrators',
                                       'shortname': 'sysadmins'}],
                'License': [{'description': 'Apache License V2.0',
                             'id': 401,
                             'name': 'Apache License V2.0',
                             'shortname': 'apache2'},
                            {'description': 'GNU Library or Lesser General Public License (LGPL)',
                             'id': 16,
                             'name': 'GNU Library or Lesser General Public License (LGPL)',
                             'shortname': 'lgpl'}],
                'Operating System': [{'description': 'All POSIX (Linux/BSD/UNIX-like OSes)',
                                      'id': 200,
                                      'name': 'All POSIX (Linux/BSD/UNIX-like OSes)',
                                      'shortname': 'posix'},
                                     {'description': 'OS Independent (Written in an interpreted language)',
                                      'id': 235,
                                      'name': 'OS Independent (Written in an interpreted language)',
                                      'shortname': 'independent'}],
                'Programming Language': [{'description': 'Python',
                                          'id': 178,
                                          'name': 'Python',
                                          'shortname': 'python'}]
```


Tools Matter
Know Your Tools
Screws and Nails
mongodb
siding
deck

AKA

Learning by doing

AKA Horror stories



Tools Matter

Know Your Tools

Screws and Nails

mongodb

siding

deck

merciless.sourceforge.net

- Figure out what YOUR app needs
- Don’t obsess about SCALE you’ll never achieve
- Use the right tool for the job

Lessons learned

- a tool is only right when you know how to use it
- DomainModel style setup is critical if you use more than one persistance type

Mongo Lessons learned

- you will have to repeat yourself
- autosharding (still) not ready
- local mongo on the web server is *really* fast
- be carefull if the index does not fit in memory