Relax with CouchDB

Mark Drew
Railo Consulting
cfObjective 2011
Minneapolis, MN,USA

Quick Info!

tw: #cfobjective & @markdrew

e: mark@getrailo.com

w: http://www.getrailo.com

v: http://bit.ly/couch_cfo

n: http://bit.ly/doc_cfo

Count Yourselves!

10:15-11:15	Don't Forget About Custom Tags - Steve Bryant	Relax with CouchDB - Mark Drew	Application Intrusion, Detection and Tracking - Dave Ferguson	Flex and ColdFusion, a Love Affair - Douglas Knudsen	ORM and ValidateThis - Jamie Krug

Who be MD?

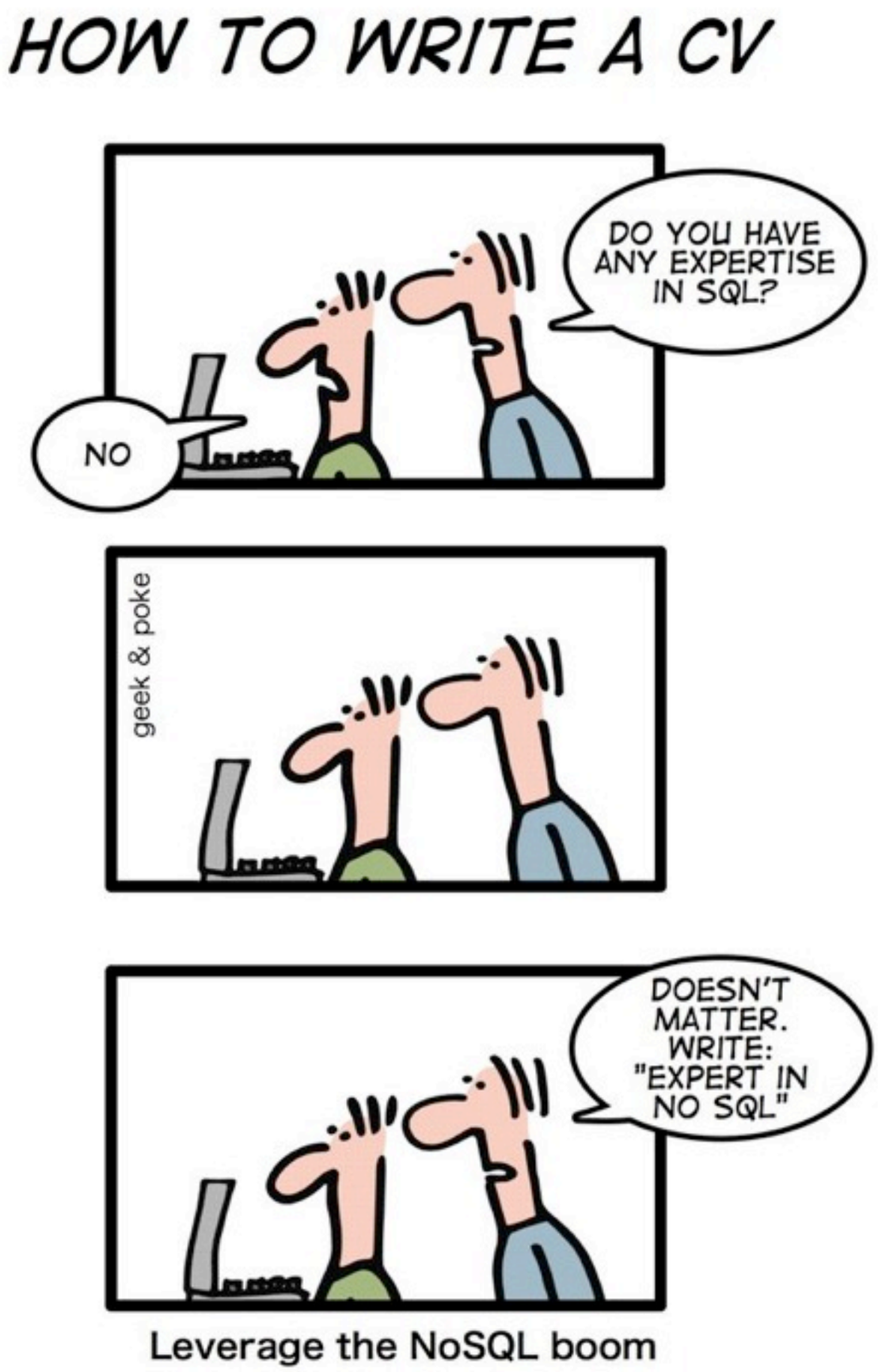
"I am just this guy... you know?"

Monday, 16 May 2011

Railo Consulting UK's CEO
- Web Development Consultancy based in London, Zurich, New York and San Francisco
- Excel at CFML application development, training, performance tuning

ColdFusion & CFML Developer since '97 and Web developer since '94

Involved in various CFML Frameworks, ORMs and IDE's
- CFEclipse's (CFML IDE) Lead Developer
- Reactor ORM Project Manager
- ColdSpring (Dependency Injection) Contributor Contributor
- Model Glue and ColdBox (MVC) Contributor
- ColdBox Contributor



“Do you have any expertise in SQL?”

“No”

“Doesn’t Matter. Write: ‘Expert in No SQL’”

How many have heard of noSQL?

Take away first?

!=

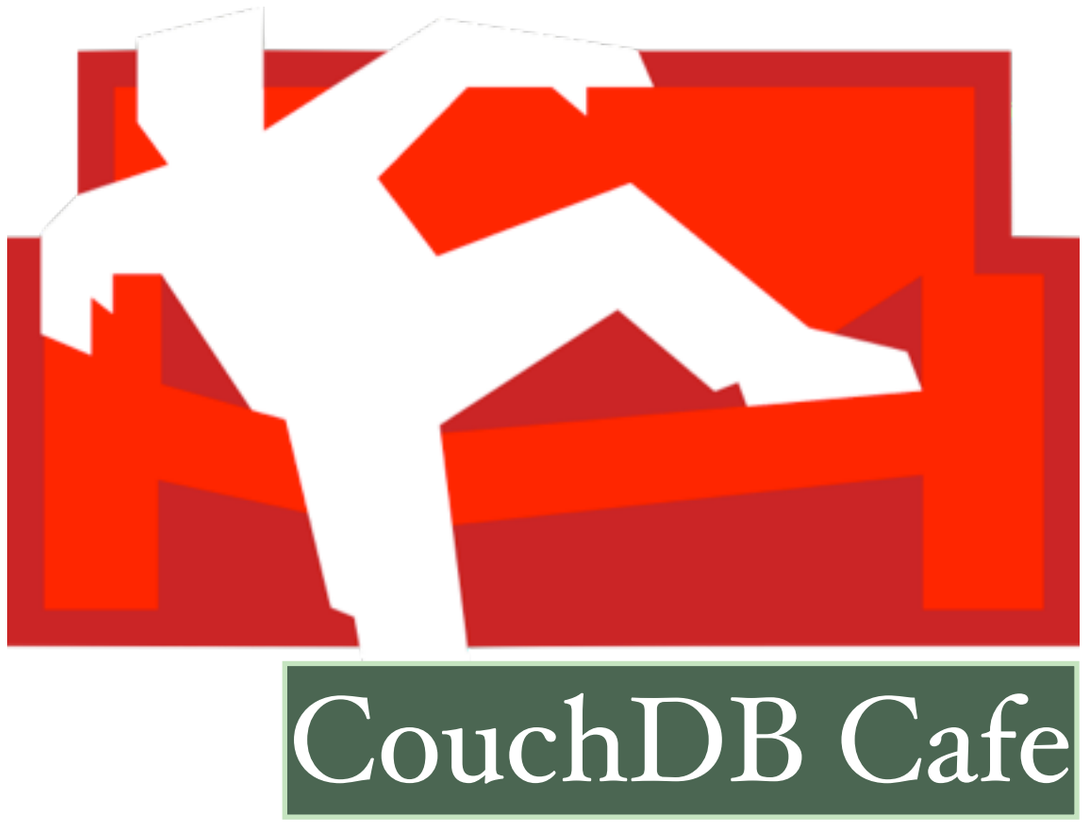
Not a direct replacement for databases.

Take away first?

!=

You can do other things

What's on the menu?



Salads and Starters

RDMBS Mezze

noSQL Salad

Mains

Couch DB Steak

REST Kebab

Desserts

CouchDB App platter

On the menu
Looking at Relational Databases and where they fail
Background of noSQL
Look into Couch DB
And How to use it
Finally check out some other things you can do with it

SQL a Go-Go!

Relational Databases are widely known and used. How many people use them here?
Why do we use them?

SQL a Go-Go!

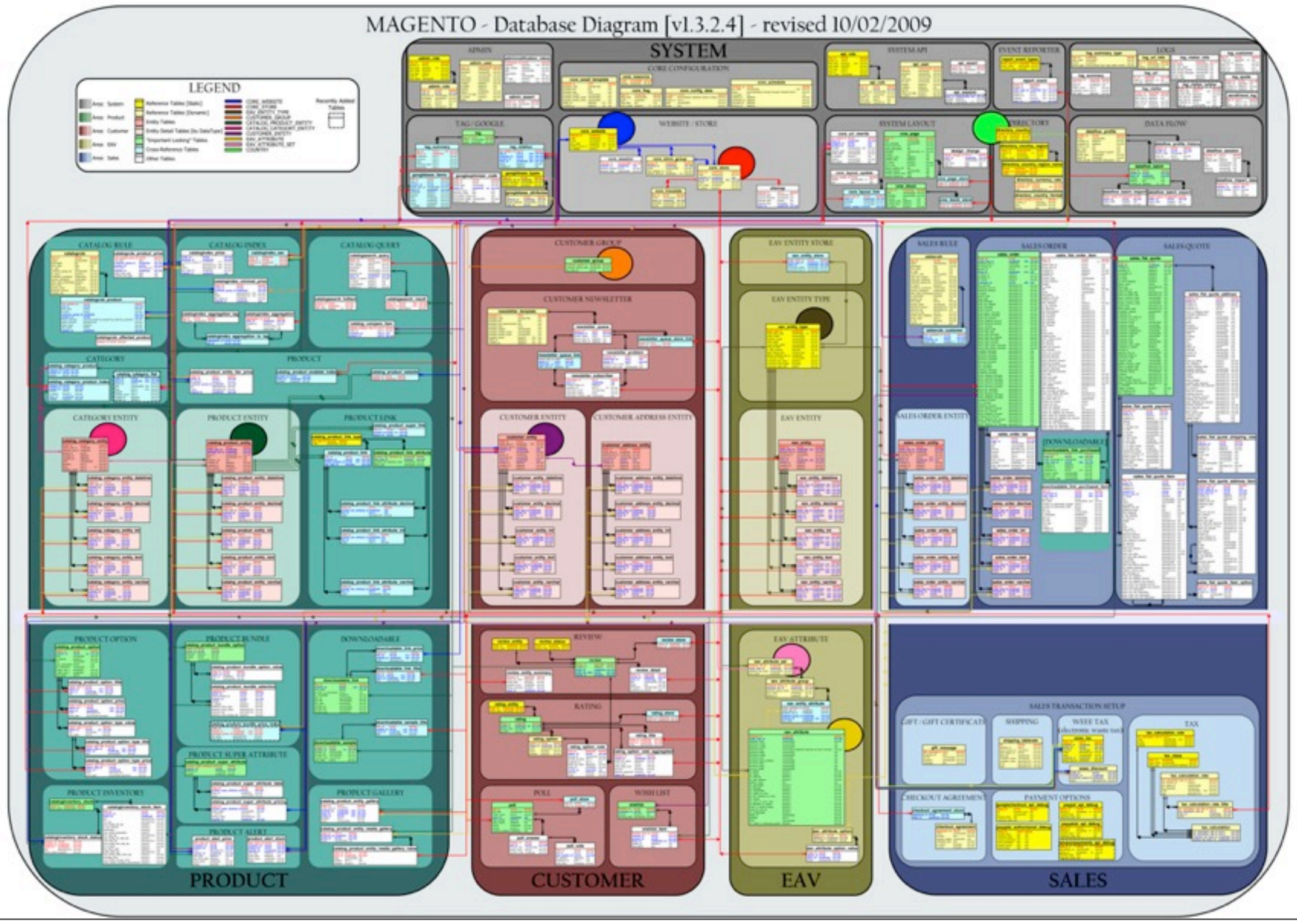
- ACID!
  - **A**tomic
  - **C**onsistent
  - **I**solated
  - **D**urable
- Reliable Proven Technologies
- Strong Ecosystem

Monday, 16 May 2011

Atomic – all or nothing
Consistent – The consistency property ensures that any transaction the database performs will take it from one consistent state to another.
Isolated – Isolation refers to the requirement that other operations cannot access data that has been modified during a transaction that has not yet completed
Durable – Durability is the ability of the DBMS to recover the committed transaction updates against any kind of system failure

Lots of Computing brain power have gone into the alogarithms within
Lots of companies have a lot of money invested.

e.g: Magento



Monday, 16 May 2011

Lots of complicated problems already solved with it
But it isn't everything

BUT!!!

But it isn’t everything
Are we using them for the right things?
What are the problems with them?

How about changes to data?

	id	title	author_id	status	last_modified
1	0343F64E-BE56-40A8-8056AEA08B93F140	FusionDebug 3.0: now for Railo!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-04-09 10:24:19'}
2	03BE5824-C603-9309-55927EBD568ACE5F	How to be a dufus...	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-12-04 15:44:12'}
3	04821AFF-C797-ED0B-62543B59B2AD74EE	Strange expandPath bug in ColdFusion 8	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-23 12:22:03'}
4	05E78B9B-0E9F-4596-8B521FC9FD827D7B	And we're off! First day of business at Mark Drew Consulting	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-07-18 13:38:48'}
5	05FCEE52-FCEA-347F-19A033712C5902B2	CFUnited 09: Cast your vote and a little question for you...	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-12-05 02:11:48'}
6	0646C651-7C82-4294-9FF58337460E4A7E	The Railo Crew invade New York, May the CFML be with you.	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-04-28 09:17:13'}
7	0815446F-5058-40F0-BD79EDF9EFCF46C2	Presenting at cf.Objective (ANZ)	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2010-07-26 07:07:33'}
8	13163767-0D48-447A-AFBA9097188B4F53	Quick Tip on Railo Email Send Queues	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2010-07-22 09:54:20'}
9	145A6678-EB80-601D-2B5E50A48AD1AB2C	Like Model-Glue? Give the team some feedback!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-10-19 05:05:47'}
10	166C7C24-A41D-720C-C240DF59BCEE0B78	UKCFUG meetup tomorrow, Send us your questions!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-12-08 06:50:15'}

What happens when you want to change the schema?
Migrations?
Blogg.no user table

How about changes to data?

	id	title	name	author_id	status	last_modified
1	0343F64E-BE56-40A8-8056AEA08B93F140	FusionDebug 3.0: now for Railo!	fusiondebug-3-0-now-for-railo	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-04-09 10:24:19'}
2	03BE5824-C603-9309-55927EBD568ACE5F	How to be a dufus...	how-to-be-a-dufus	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-12-04 15:44:12'}
3	04821AFF-C797-ED0B-62543B59B2AD74EE	Strange expandPath bug in ColdFusion 8	strange-expandpath-bug-in-coldfusion-8	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-23 12:22:03'}
4	05E78B9B-0E9F-4596-8B521FC9FD827D7B	And we're off! First day of business at Mark Drew Consulting	and-we-re-off-first-day-of-business-at-mark-drew-consulting	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-07-18 13:38:48'}
5	05FCEE52-FCEA-347F-19A033712C5902B2	CFUnited 09: Cast your vote and a little question for you...	cfunited-09-cast-your-vote-and-a-little-question-for-you	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-12-05 02:11:48'}
6	0646C651-7C82-4294-9FF58337460E4A7E	The Railo Crew invade New York, May the CFML be with you.	the-railo-crew-invade-new-york-may-the-cfml-be-with-you	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-04-28 09:17:13'}
7	0815446F-5058-40F0-BD79EDF9EFCF46C2	Presenting at cf.Objective (ANZ)	presenting-at-cf-objective-anz	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2010-07-26 07:07:33'}
8	13163767-0D48-447A-AFBA9097188B4F53	Quick Tip on Railo Email Send Queues	quick-tip-on-railo-email-send-queues	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2010-07-22 09:54:20'}
9	145A6678-EB80-601D-2B5E50A48AD1AB2C	Like Model-Glue? Give the team some feedback!	like-model-glue-give-the-team-some-feedback	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-10-19 05:05:47'}
10	166C7C24-A41D-720C-C240DF59BCEE0B78	UKCFUG meetup tomorrow, Send us your questions!	ukcfug-meetup-tomorrow-send-us-your-questions	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2008-12-08 06:50:15'}

Other types of data that are harder to map to relational dbs?

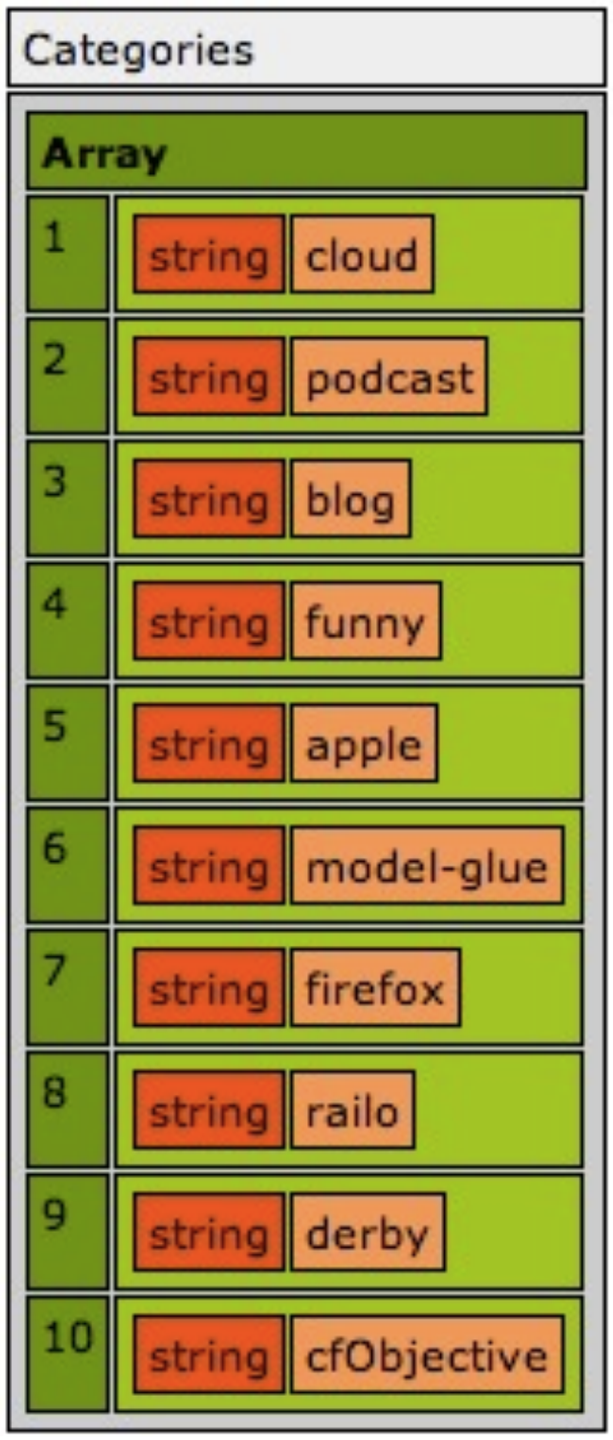
Can all data be defined?

Query
Template:/Library/WebServer/Documents/Development/couchdemo/Types.cfm
Execution Time (ms):3
Recordcount:10
Cached:No
SQL:
SELECT id, title, author_id, status, last_modified
FROM mng_entry
ORDER BY last_modified DESC
LIMIT 10

	id	title	author_id	status	last_modified
1	B3163B70-A3C9-B01B-0B978FA3911F7E3D	MDC, easy as 1-2-3. Say hello to Mark Drew Consulting	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-02-26 09:56:32'}
2	89C88345-DA01-222E-A1C2E1206DF08F23	Scotch-On-The-Rocks 2009 Tickets availale!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-02-18 09:27:13'}
3	4C1AC4DE-D7A1-BC0A-D8FCF6F5C3C8CBB9	UKCFUG Meeting: Little Fluffy Clouds!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-02-06 10:01:23'}
4	04821AFF-C797-ED0B-62543B59B2AD74EE	Strange expandPath bug in ColdFusion 8	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-23 12:22:03'}
5	FE0799BA-ECDE-382E-A660D87CA3138260	Want to win a Railo Enterprise License?	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-22 06:09:17'}
6	FA68C5E7-09B6-75A0-2DC45D71BB65FA91	Don't underestimate the power of ad placement!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-21 13:16:57'}
7	F00D61B1-0C1B-F111-162526555C4317E5	Want to know more about Elastic Computing and Virtualistaion?	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-19 13:00:55'}
8	EE83A386-E8ED-0941-E7247AFF60785648	Episode 3 of the CFDocs podcast is live!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-19 05:50:51'}
9	DFD9C2C6-03E5-A0B1-E85ABAA557EA7080	Support MG Joe (and Dale)	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-16 09:30:36'}
10	B66E0058-E624-A1A7-54F227DB4453EF98	CF Docs Podcast Episode 2 released Today!	8A974E31-C466-658A-A63060C8F841FBC4	published	{ts '2009-01-08 08:28:28'}

But is all data the same?

What about...?



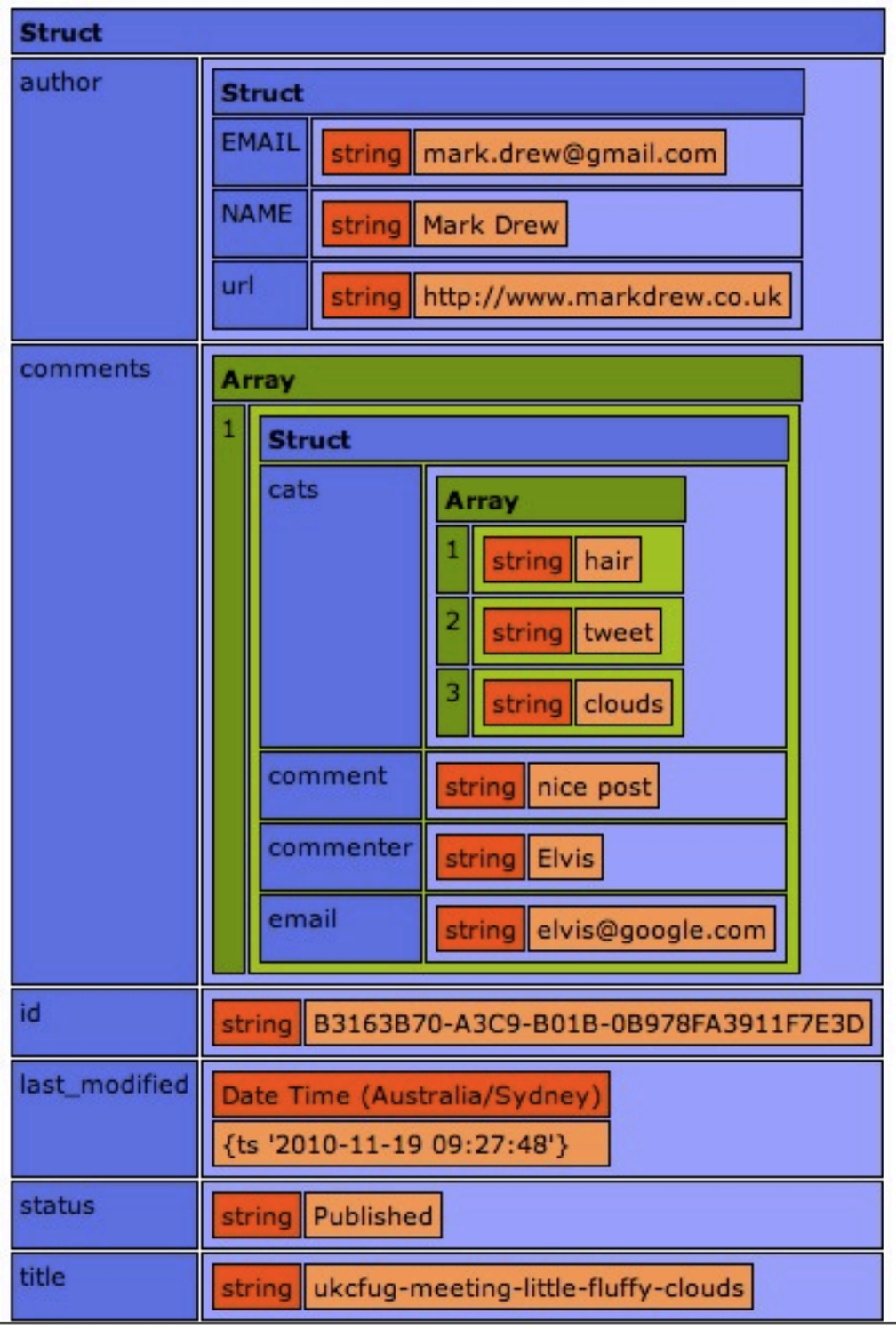
How about arrays? That's cool.

What about...?

Struct	
author	string: Mark Drew
id	string: B3163B70-A3C9-B01B-0B978FA3911F7E3D
last_modified	Date Time (Australia/Sydney) {ts '2010-11-19 09:19:57'}
status	string: Published
title	string: ukcfug-meeting-little-fluffy-clouds

One Level maps (or structures) ... cool still

But what about...?



But what about more complicated objects... such as items with arrays and structures?

But what about...?


```
- <entries>
  - <entry>
      <id>0343F64E-BE56-40A8-8056AEA08B93F140</id>
      <title>FusionDebug 3.0: now for Railo!</title>
      <author_id>8A974E31-C466-658A-A63060C8F841FBC4</author_id>
    - <comments>
      - <comment>
          <creator>Luis Majano</creator>
        - <content>
            That is amazing Mark, what a great debugger!! This just makes Railo development even better.
          </content>
        </comment>
      - <comment>
          <creator>Gianna Benson</creator>
        - <content>
            Just thought I would let you know that we have come out with yet another new release! The new FusionDebug 3.5 introduces Non-Intrusive Debugging as well as other advanced
            debugging features. You can check it out online at http://www.fusion-debug.com/fd/.
          </content>
        </comment>
      - <comment>
          <creator>Gianna Benson</creator>
        - <content>
            Hi Mark! FusionDebug 3.0 is now released! People can download a free trial edition of FusionDebug 3.0 from the FusionDebug website at http://www.fusion-reactor.com
            /fd/downloads.cfm.
          </content>
        </comment>
      </comments>
    </entry>
  - <entry>
      <id>03BE5824-C603-9309-55927EBD568ACE5F</id>
      <title>How to be a dufus...</title>
      <author_id>8A974E31-C466-658A-A63060C8F841FBC4</author_id>
    - <comments>
      - <comment>
          <creator>todd sharp</creator>
        - <content>
            Some of these might have been appropriate for your wufoo form --> http://zacster.blogspot.com/2008/12/best-out-of-office-automatic-e-mail.html
```

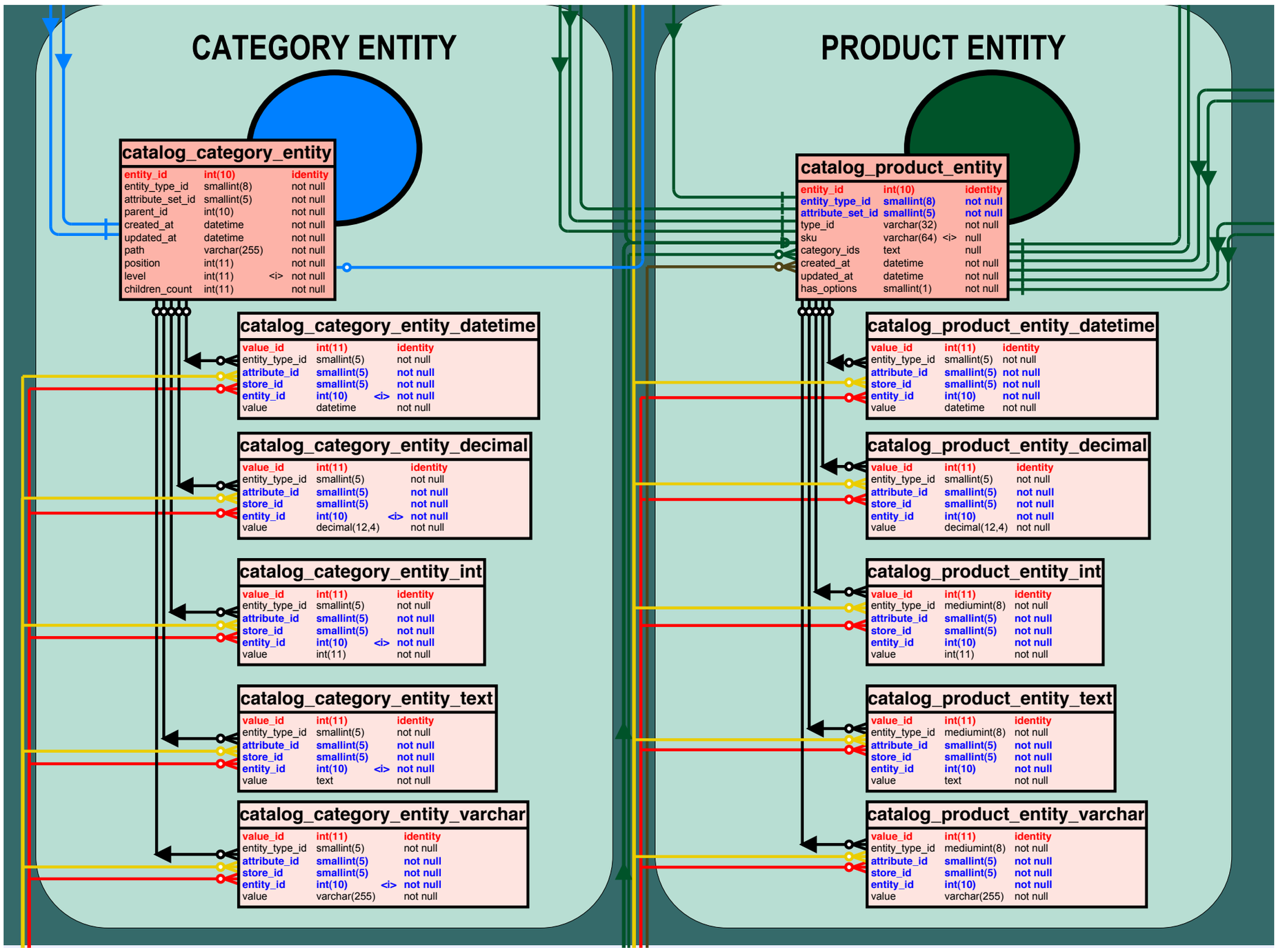

How about XML document trees?

But what about...?


```
{
   "entries": [{
      "author_id": "8A974E31-C466-658A-A63060C8F841FBC4",
      "comments": [{
         "creator": "Luis Majano",
         "content": "That is amazing Mark, what a great debugger!! This just makes Railo development even better."
      },
      {
         "creator": "Gianna Benson",
         "content": "Just thought I would let you know that we have come out with yet another new release! The new FusionDebug 3.5 introduces Non-Intrusive Debugging as well as other advanced debugging features. You can check it out online at http:\/\/www.fusion-debug.com\/fd\/."
      },
      {
         "creator": "Gianna Benson",
         "content": "Hi Mark! FusionDebug 3.0 is now released! People can download a free trial edition of FusionDebug 3.0 from the FusionDebug website at http:\/\/www.fusion-reactor.com\/fd\/downloads.cfm."
      }],
      "title": "FusionDebug 3.0: now for Railo!",
      "id": "0343F64E-BE56-40A8-8056AEA08B93F140"
   },
   {
      "author_id": "8A974E31-C466-658A-A63060C8F841FBC4",
      "comments": [{
         "creator": "todd sharp",
         "content": "Some of these might have been appropriate for your wufoo form --> http:\/\/zacster.blogspot.com\/2008\/12\/best-out-of-office-automatic-e-mail.html"
      },
      {
         "creator": "Gary Gilbert",
         "content": "Thanks Mark, that is exactly what I needed this morning, a good laugh"
      },
```

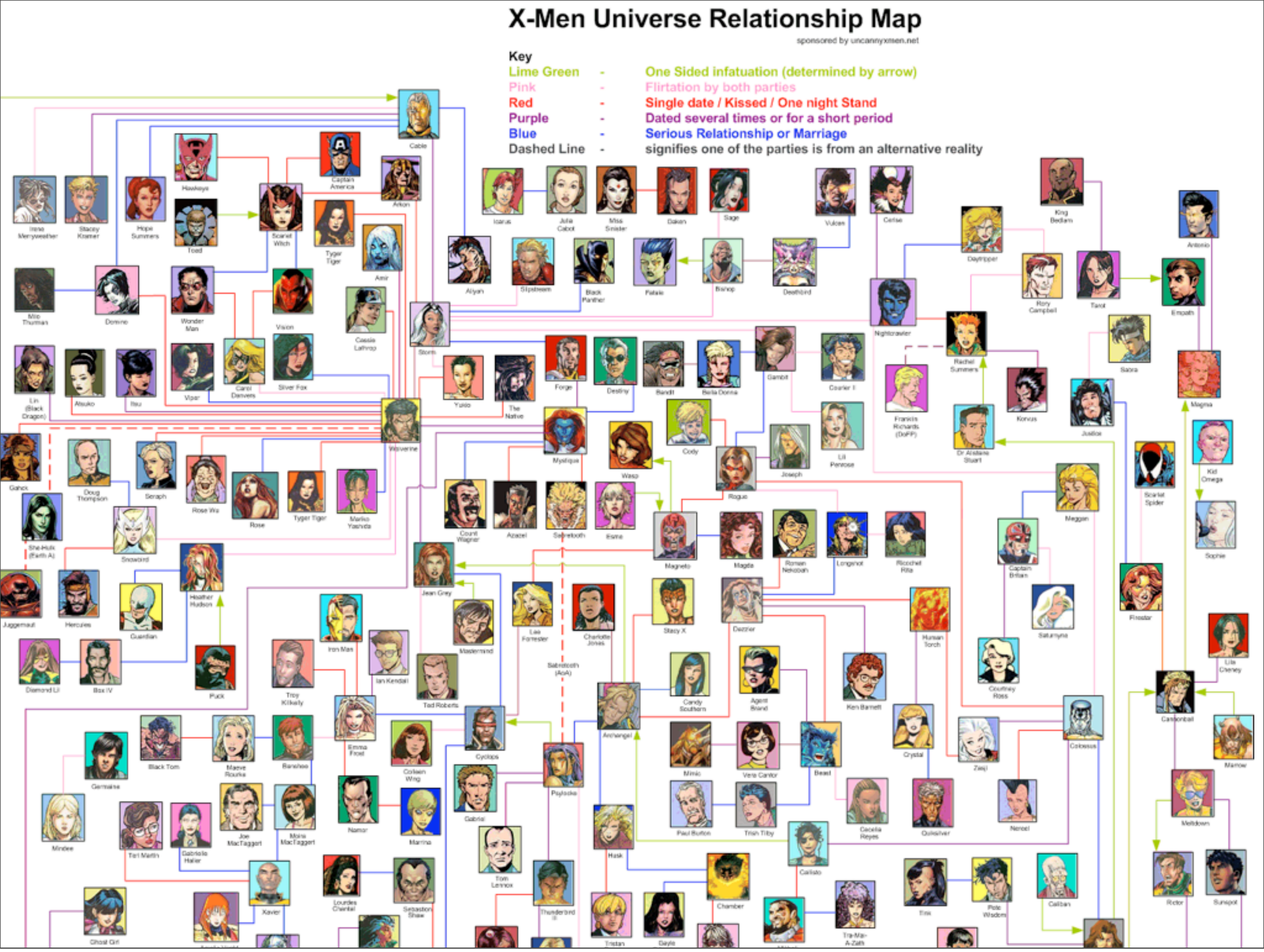

JSON trees? Who uses JQuery or other Javascript libraries?

Can be done...?



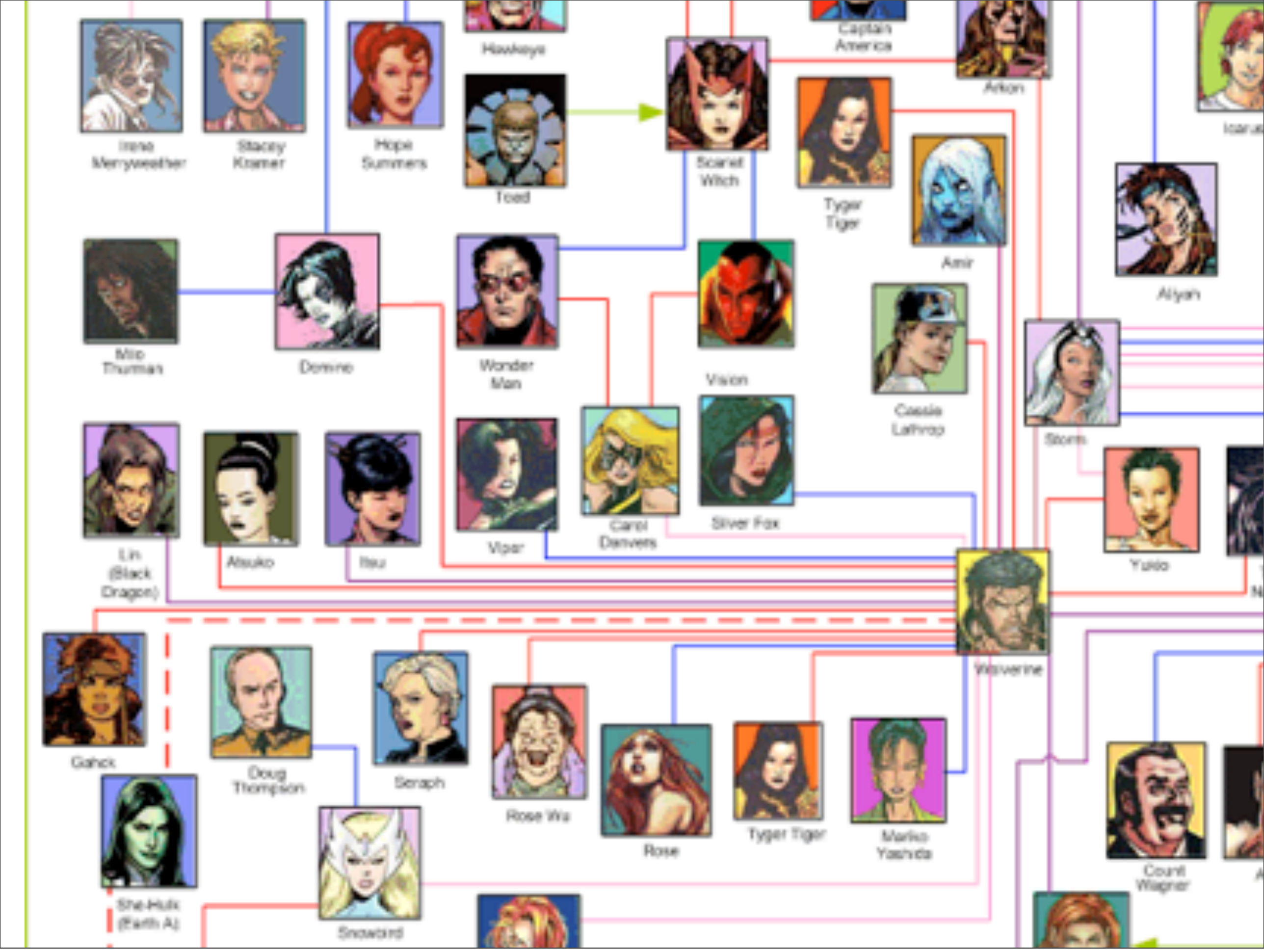
Monday, 16 May 2011

There are solutions, such as relations... as seen here in Magento



Monday, 16 May 2011

But not all relationships are easy, nor optimised



Monday, 16 May 2011

Are RDBM's the only way to solve these problems?

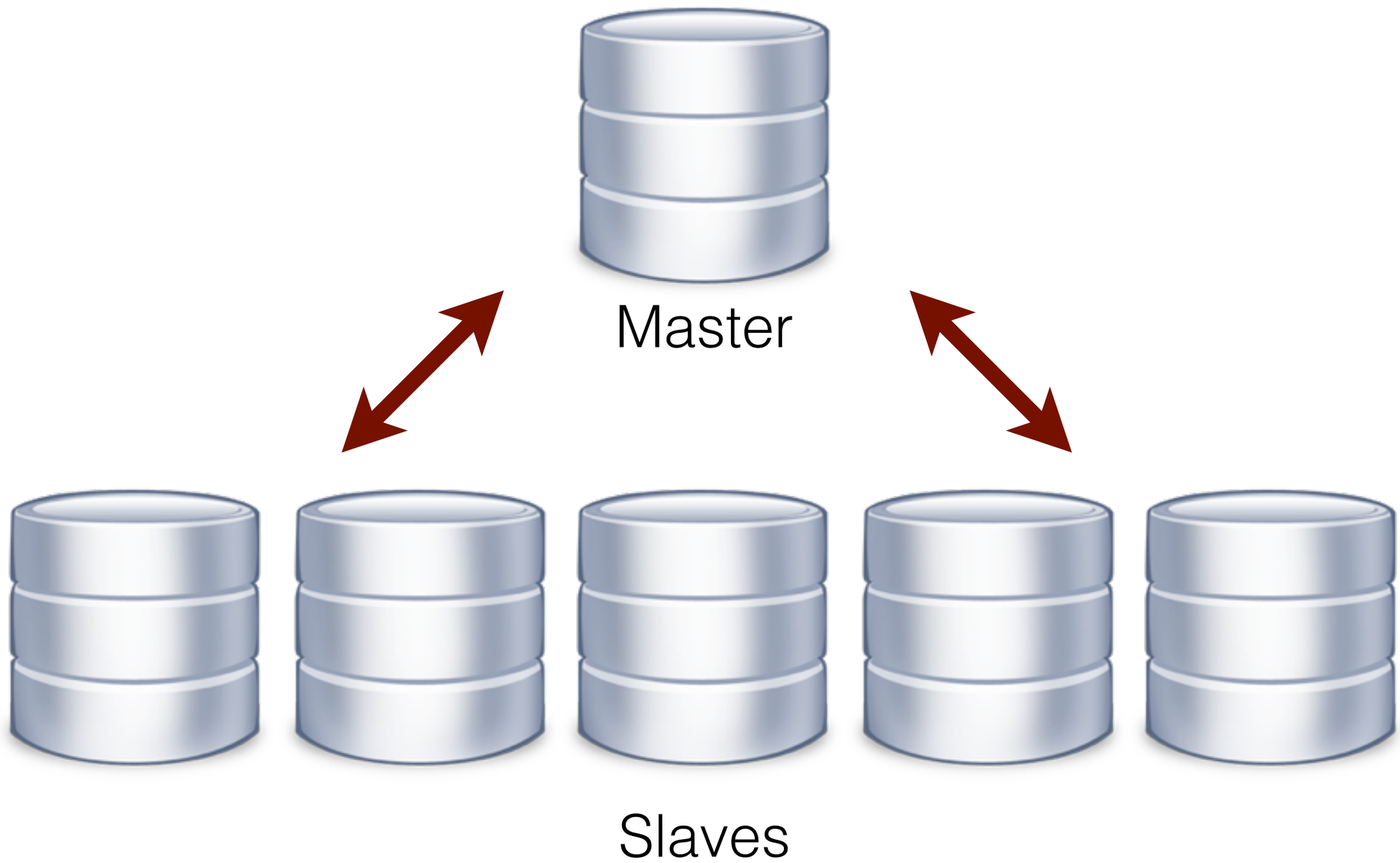
This is kinda like facebook or linked in, or the 6 degrees of separation between me and Kevin Bacon

Clustering?

Monday, 16 May 2011

Easy to set up massive scale clustering

Clustering?



Performance?

Isolated transactions means locking... Not so good when you are getting high traffic!

SQL #fail

- Flexibility (fixed schema)
- Doesn't model domain
- Horizontal Scaling
- Performance

Back on topic...

noSQL WTF?

NoSQL?

Monday, 16 May 2011

Not “NO SQL” but “Not Only SQL”
1998 By Carlo Strozzi, for a OSS Relational Database with no SQL api
Rackspace Eric Evans and Last.fm Johan Oskarsson re-used the term in 2009 for an even to discuss open source distributed databases
http://www.vineetgupta.com/2010/01/nosql-databases-part-1-landscape.html
OLTP - Online Transaction Processing
DW - Datawarehousing

Not Only SQL => NOSQL

Web Workloads != Online Transaction Processing or Data Warehousing

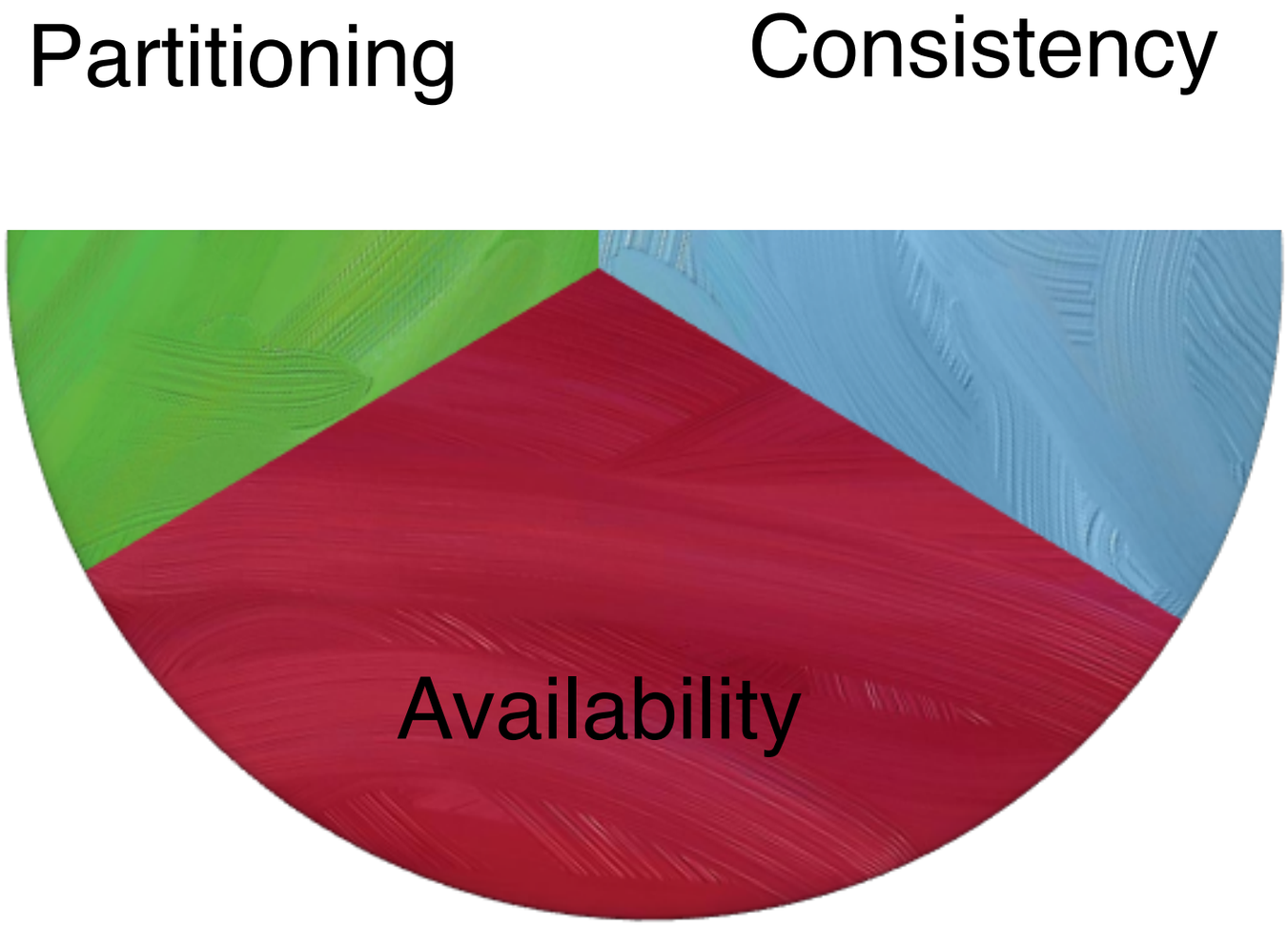
Ecosystem that addresses #fail of sql

Three Keys to NoSQL

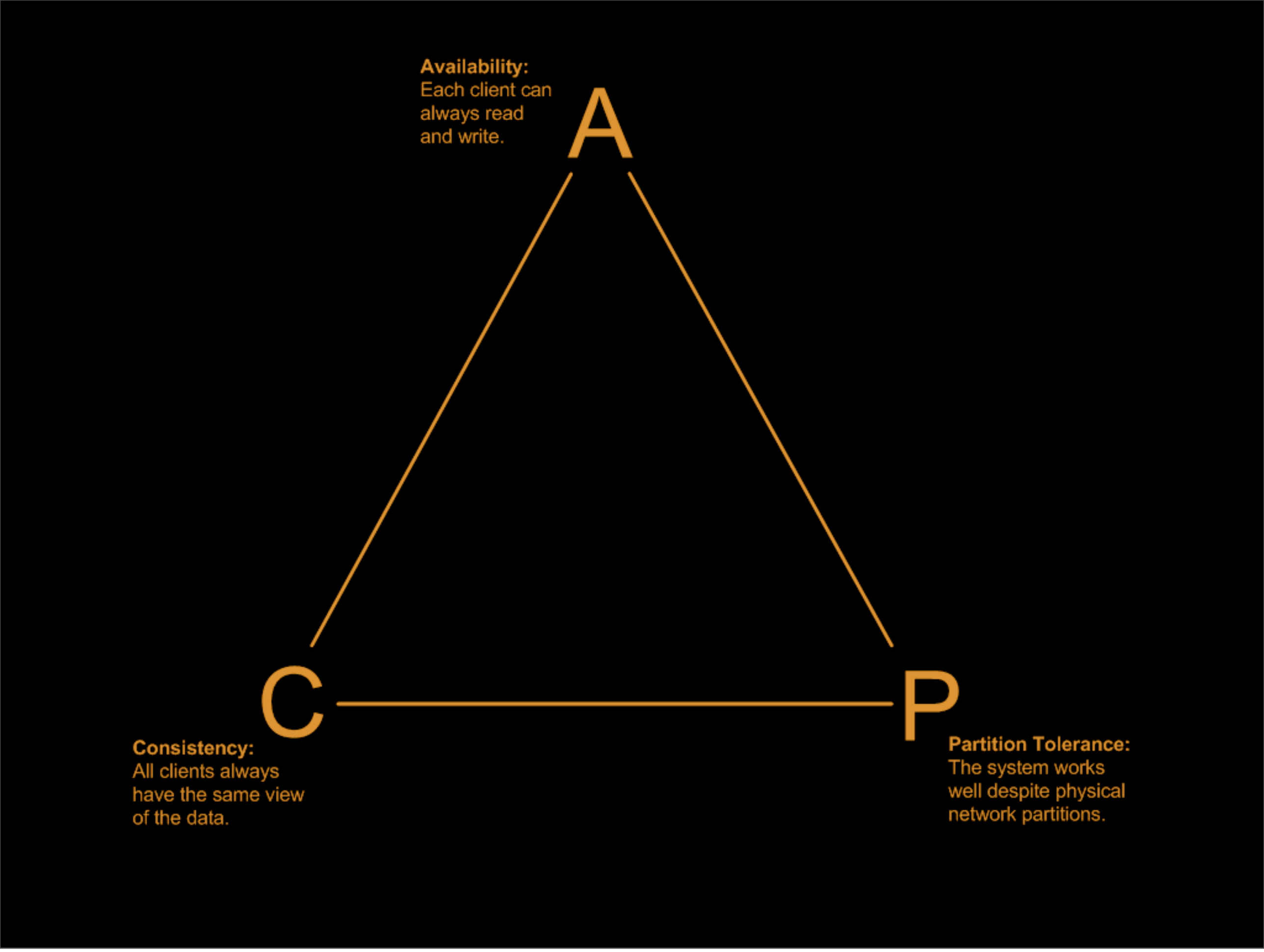
- CAP Theorem
- Dynamo implementation
- Map/Reduce algorithm

http://camelcase.blogspot.com/2007/08/cap-theorem.html
http://www.allthingsdistributed.com/2007/10/amazons_dynamo.html
http://labs.google.com/papers/mapreduce.html

CAP Theorem

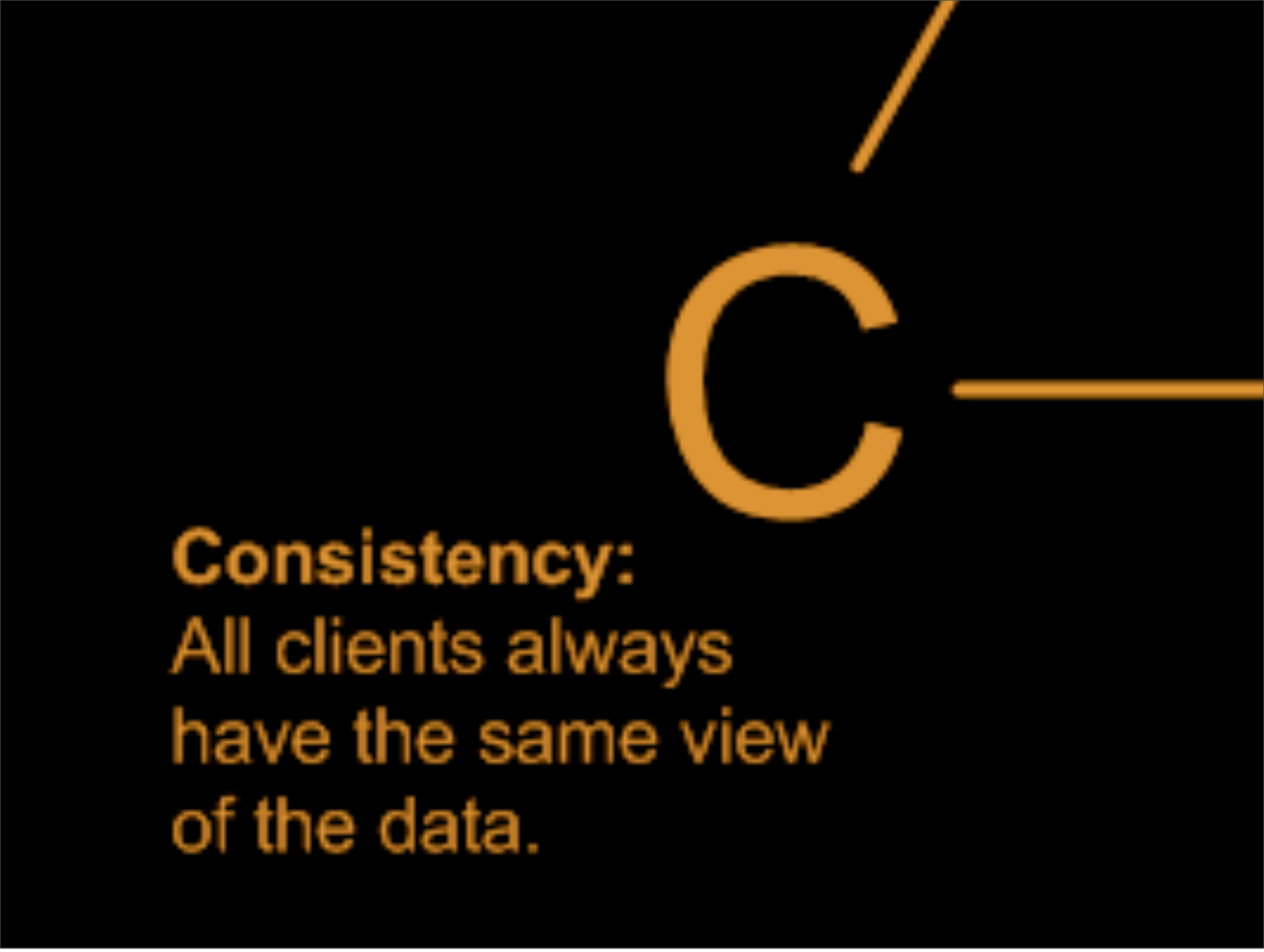


http://www.julianbrowne.com/article/viewer/brewers-cap-theorem
Eric Brewer at the ACM Symposium on the Principles of Distributed Computing (PODC). 2004
http://blog.nahurst.com/visual-guide-to-nosql-systems
Consistency means that each client always has the same view of the data.
Availability means that all clients can always read and write.
Partition tolerance means that the system works well across physical network partitions.



http://blog.nahurst.com/visual-guide-to-nosql-systems

- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.



Monday, 16 May 2011

http://blog.nahurst.com/visual-guide-to-nosql-systems

- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.

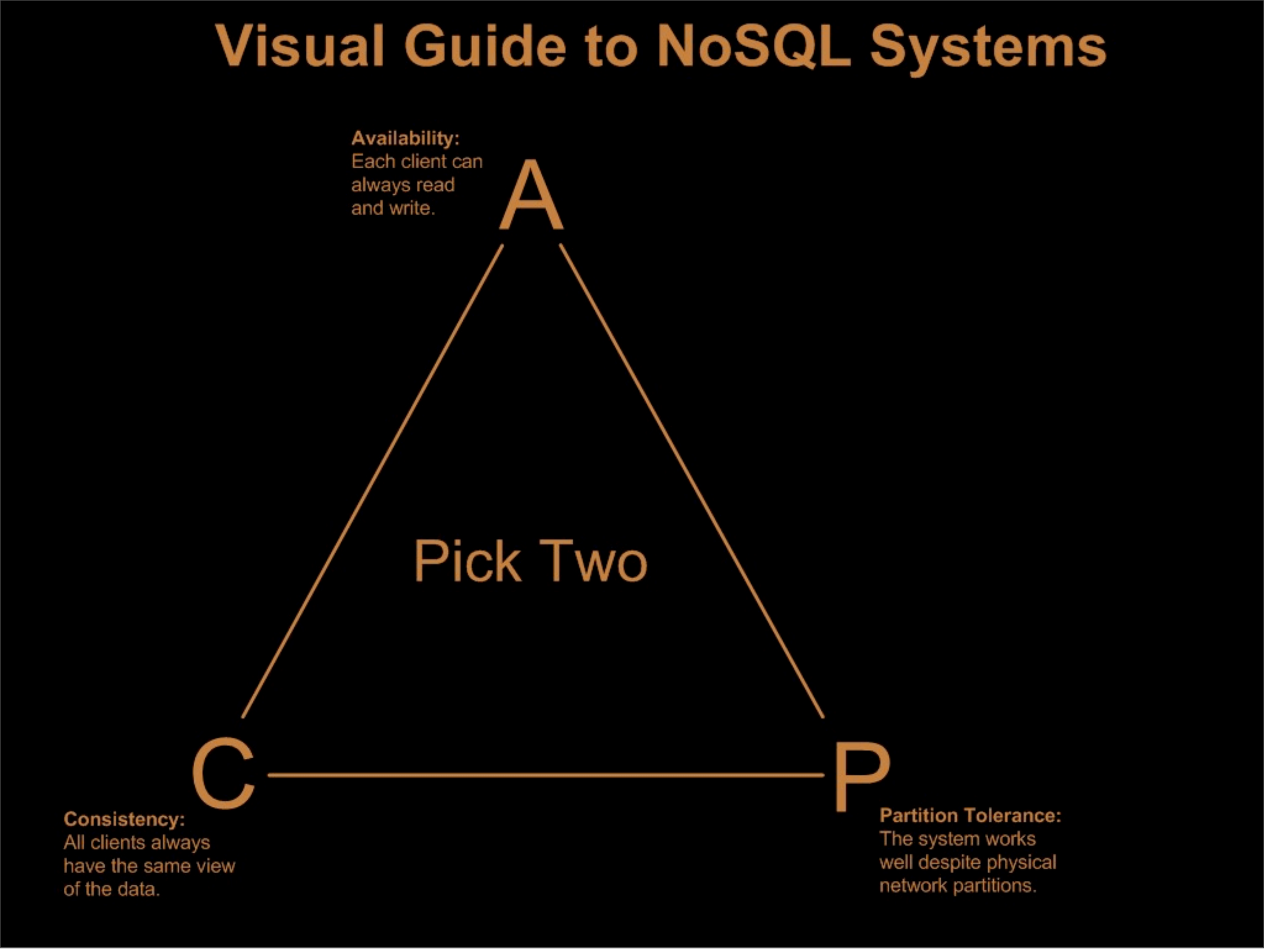
http://blog.nahurst.com/visual-guide-to-nosql-systems

- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.

Partition Tolerance:
The system works well despite physical network partitions.

http://blog.nahurst.com/visual-guide-to-nosql-systems

- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.



Monday, 16 May 2011

http://blog.nahurst.com/visual-guide-to-nosql-systems

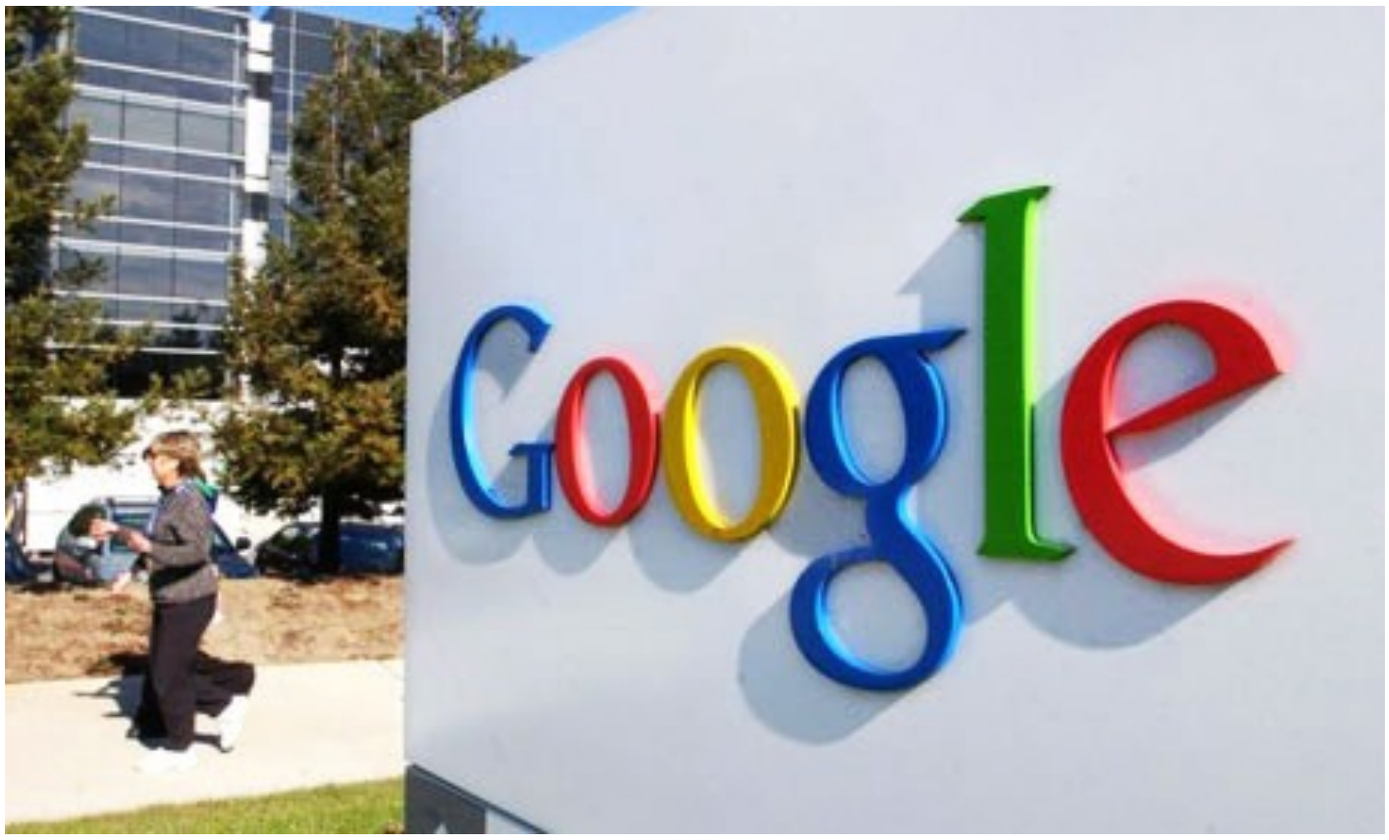
- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.

Dynamo

Dynamo: Amazon's Proprietary Highly Available Key-value Store
Inspires: Apache Hadoop project develops open-source software for reliable, scalable, distributed computing

Map/Reduce

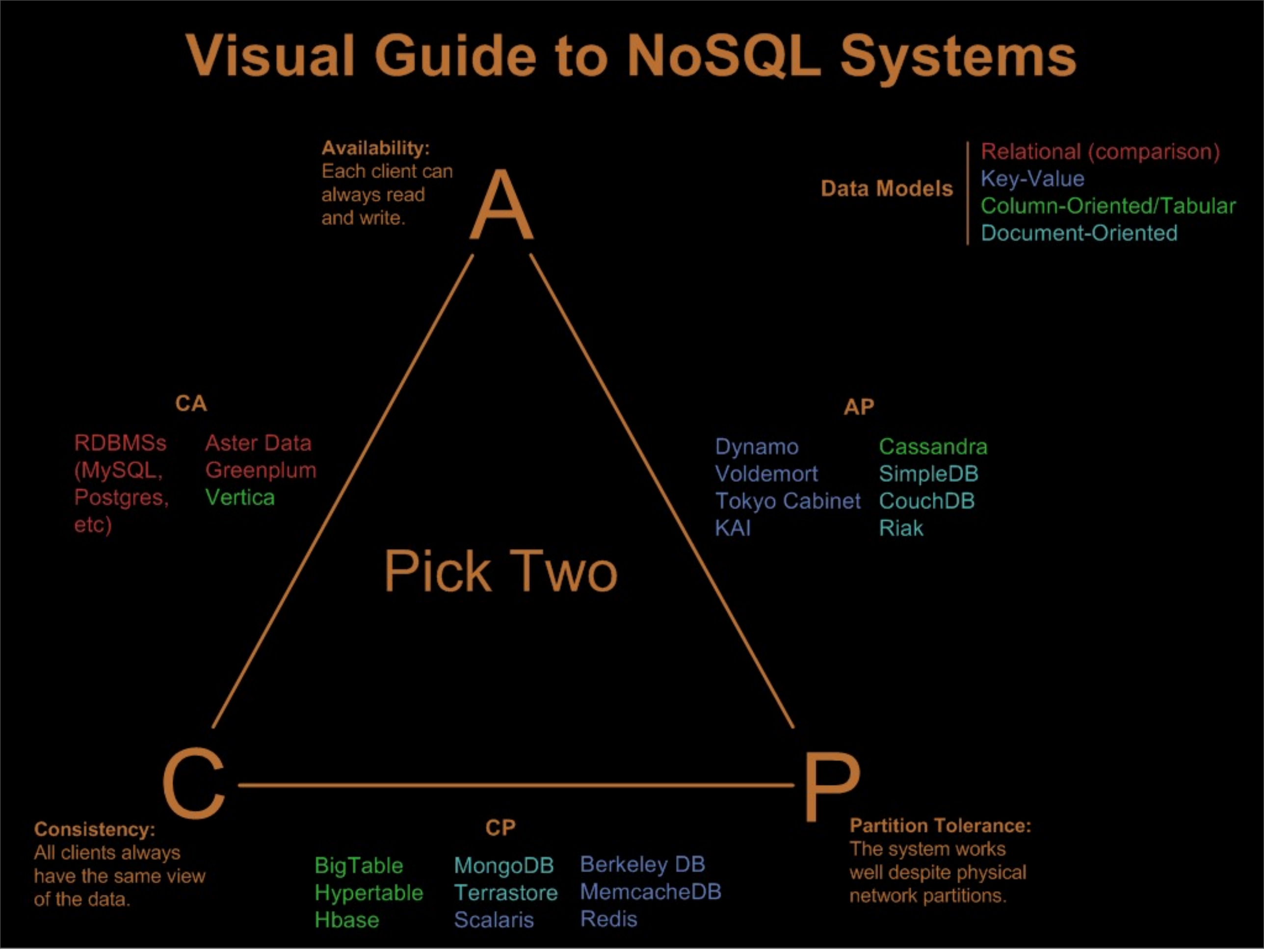
- Patented Google Algorithm
- Massively Parallel
- Indexes added per update
- Reduce contents of B-Tree



http://labs.google.com/papers/bigtable.html
http://arin.me/blog/wtf-is-a-supercolumn-cassandra-data-model

Google is canonical example
Digg with Cassandra
Twitter for analytics

NoSQL Landscape



http://blog.nahurst.com/visual-guide-to-nosql-systems

- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.

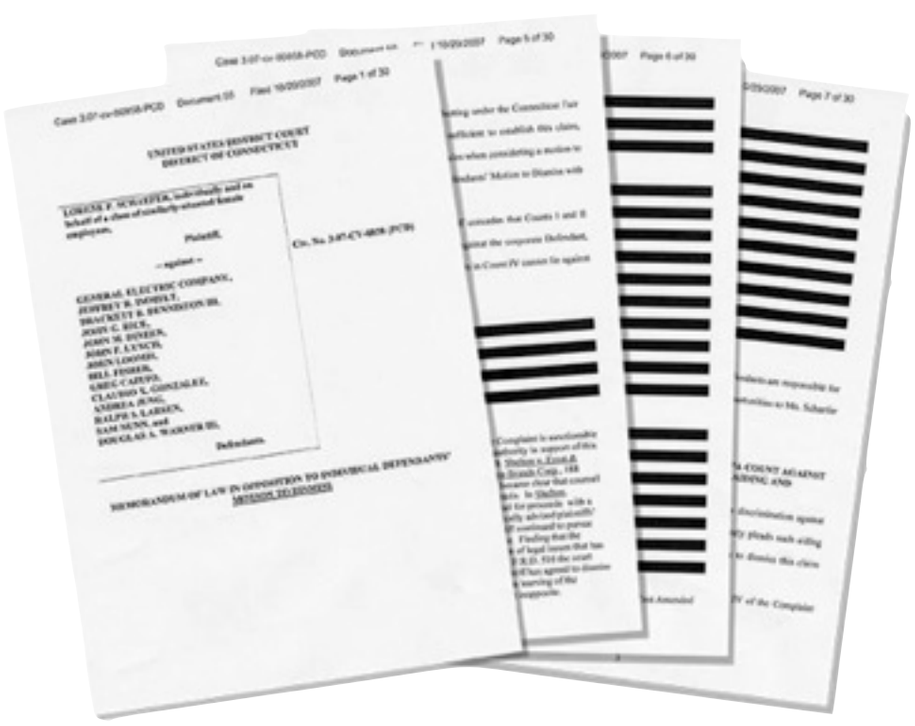
Data Models

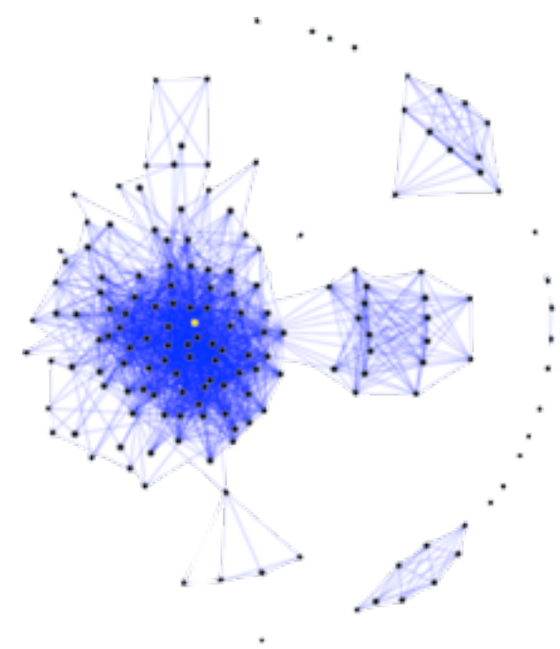
Relational (comparison)
Key-Value
Column-Oriented/Tabular
Document-Oriented

http://blog.nahurst.com/visual-guide-to-nosql-systems

- Consistency means that each client always has the same view of the data.
- Availability means that all clients can always read and write.
- Partition tolerance means that the system works well across physical network partitions.

Types of noSQL DB's





Monday, 16 May 2011

Key/Value
Column
Document
Graph
(probably there will be more things coming too!)

Key-Value Store

scaling, scaling, scaling
big hash table -- keys and values
transparent partition and replication
lower emphasis on transactions
“eventual consistency”
optimised for low-latency, high-throughput
A/P
Dynamo
Voldermort
Tokyo Cabinet
KAI
C/P
Berkey DB
MemcahedDB
Redis

Column Stores

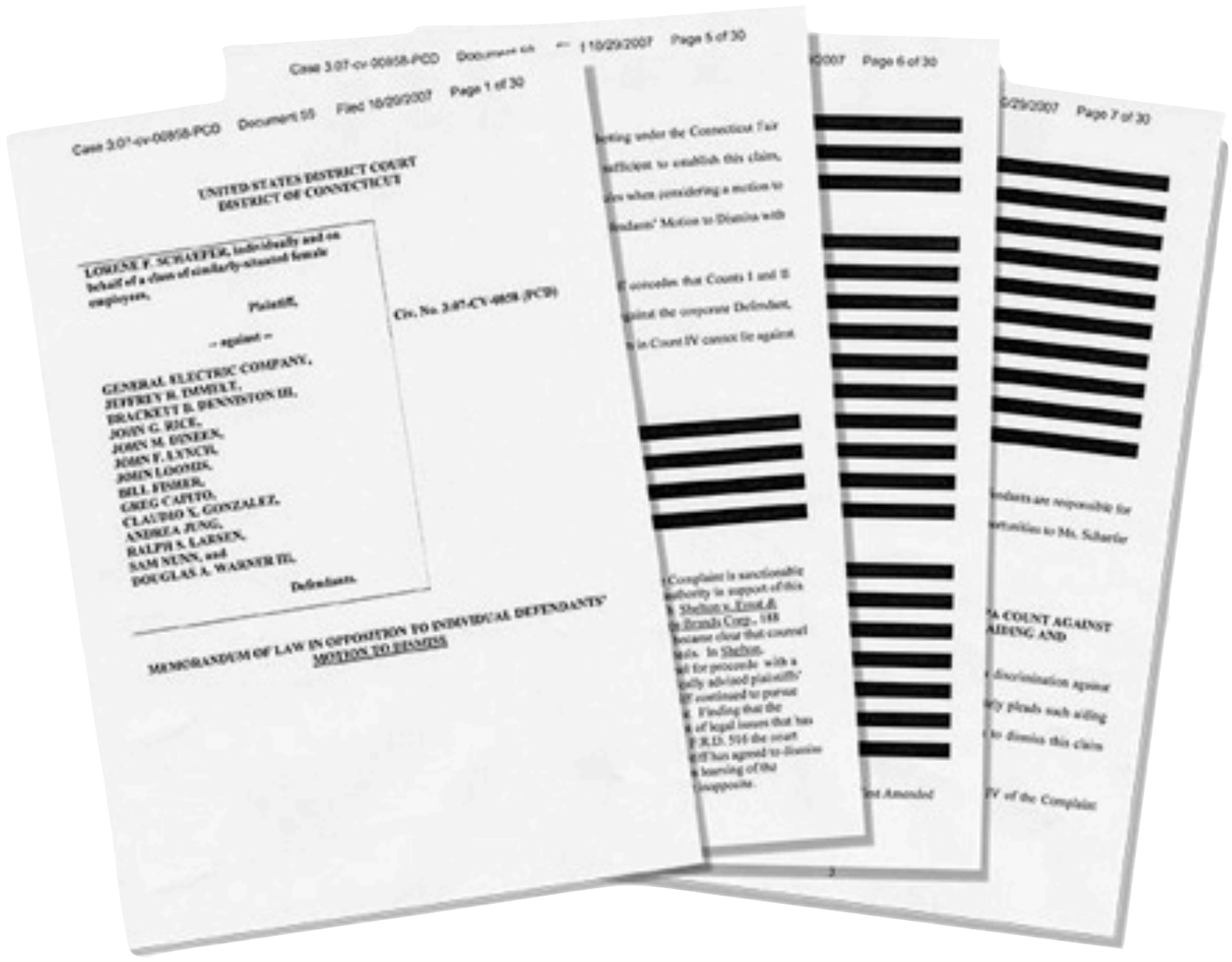
Monday, 16 May 2011

http://labs.google.com/papers/bigtable.html
http://arin.me/blog/wtf-is-a-supercolumn-cassandra-data-model

column is tuple of name, value, timestamp
stored on (custom) filesystem
no joins, no sorting after insert
very large (TB => PB) data set
distributed (local => global) data set
need for MapReduce-style processing

Google is canonical example
Digg with Cassandra
Twitter for analytics
Like a huge excel spreadsheet

Document Store

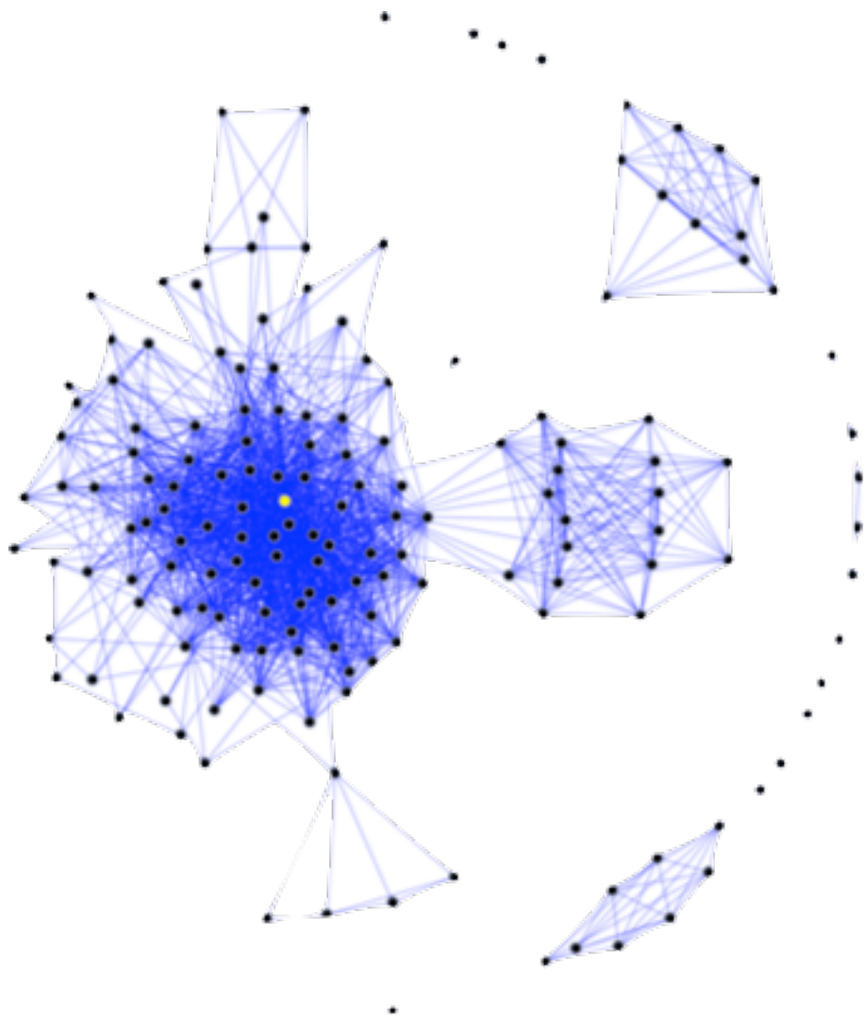


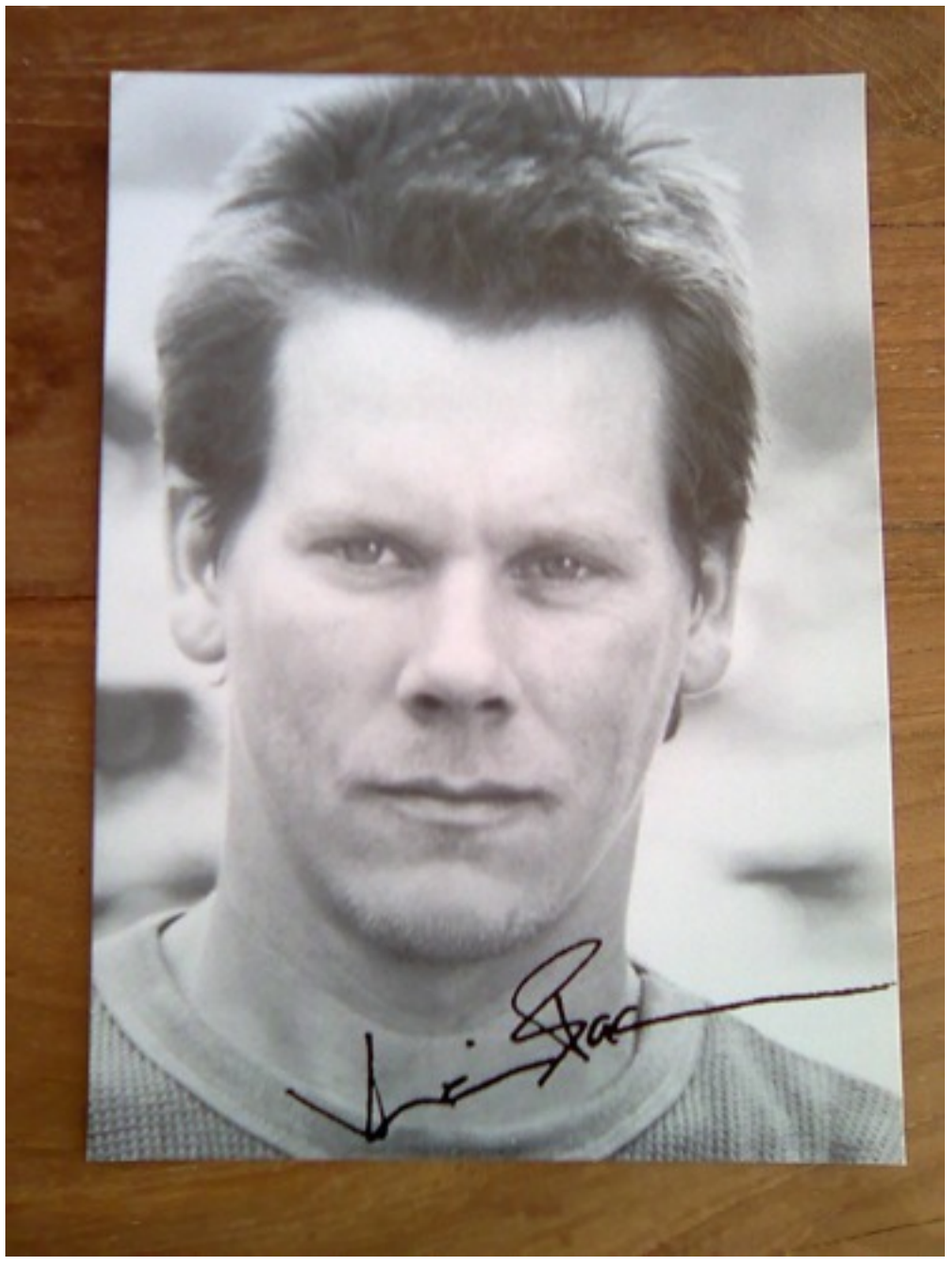
Monday, 16 May 2011

schema-less (!!!!!)
tree of objects with attribute values and lists
similar to XML or JSON storage
direct access to structure

A/P
Simple DB
Couch DB
Riak
C/P
MongoDB
TerraStore

Graph Databases





not really NoSQL in sense of other tools
stores data in graphs (networks)
nodes, relationships, properties
traverse the graph to find the data
great for social networks, collaborative filtering

noSQL #win

Thought Experiment: eCommerce
Baskets
Orders
Related to Products?
Related to Customer?
Duplication? How many tables make an order?
Messaging systems?
Distributed caches

CouchDB

- Document Store
- JSON
- REST
- MVCC

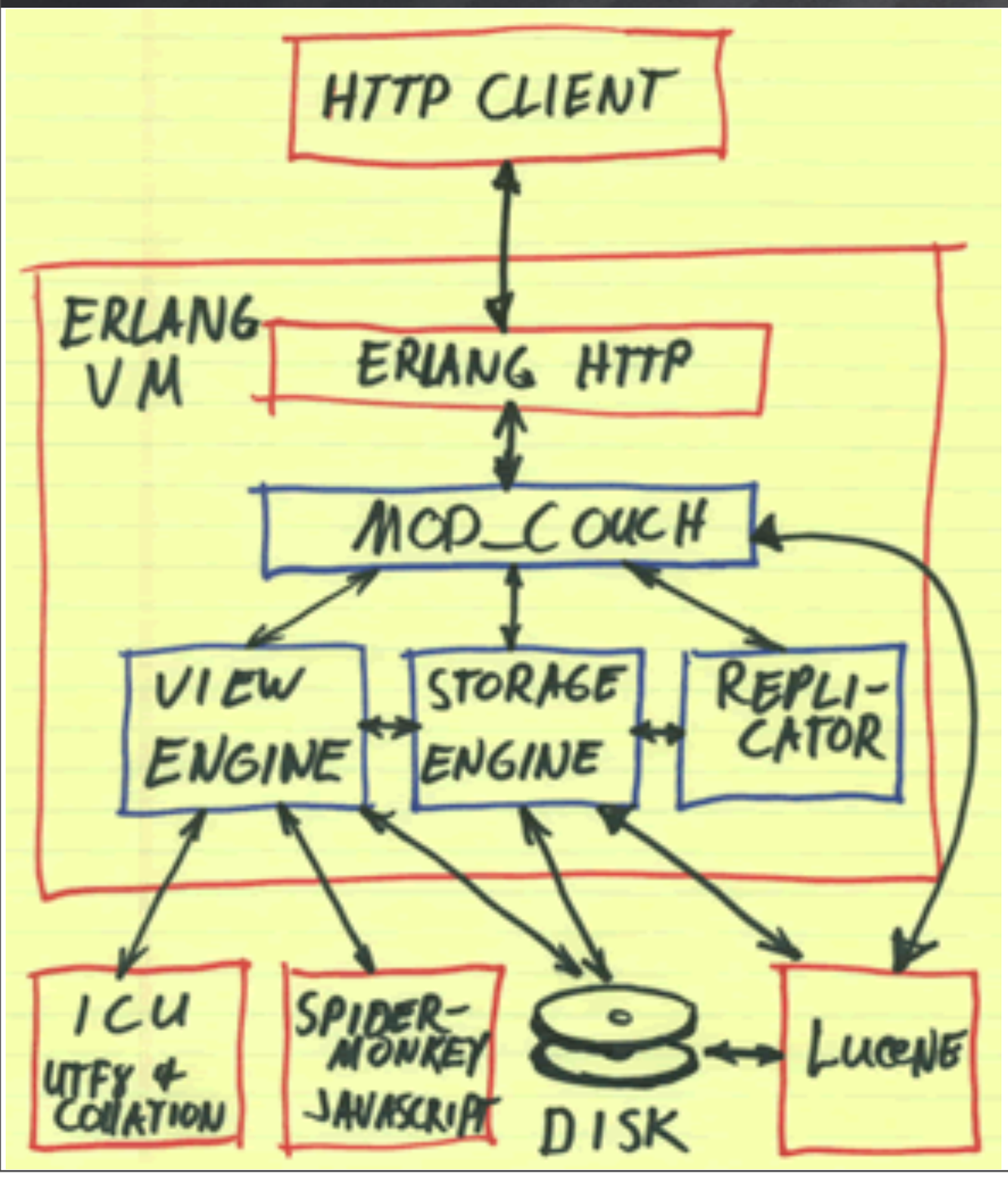
Couchbase

http://www.couchbase.com
http://couchdb.apache.org/

It's by Apache you know?

lockless optimistic updates
multiview concurrency control of documents means it's ACID

Why CouchBD?

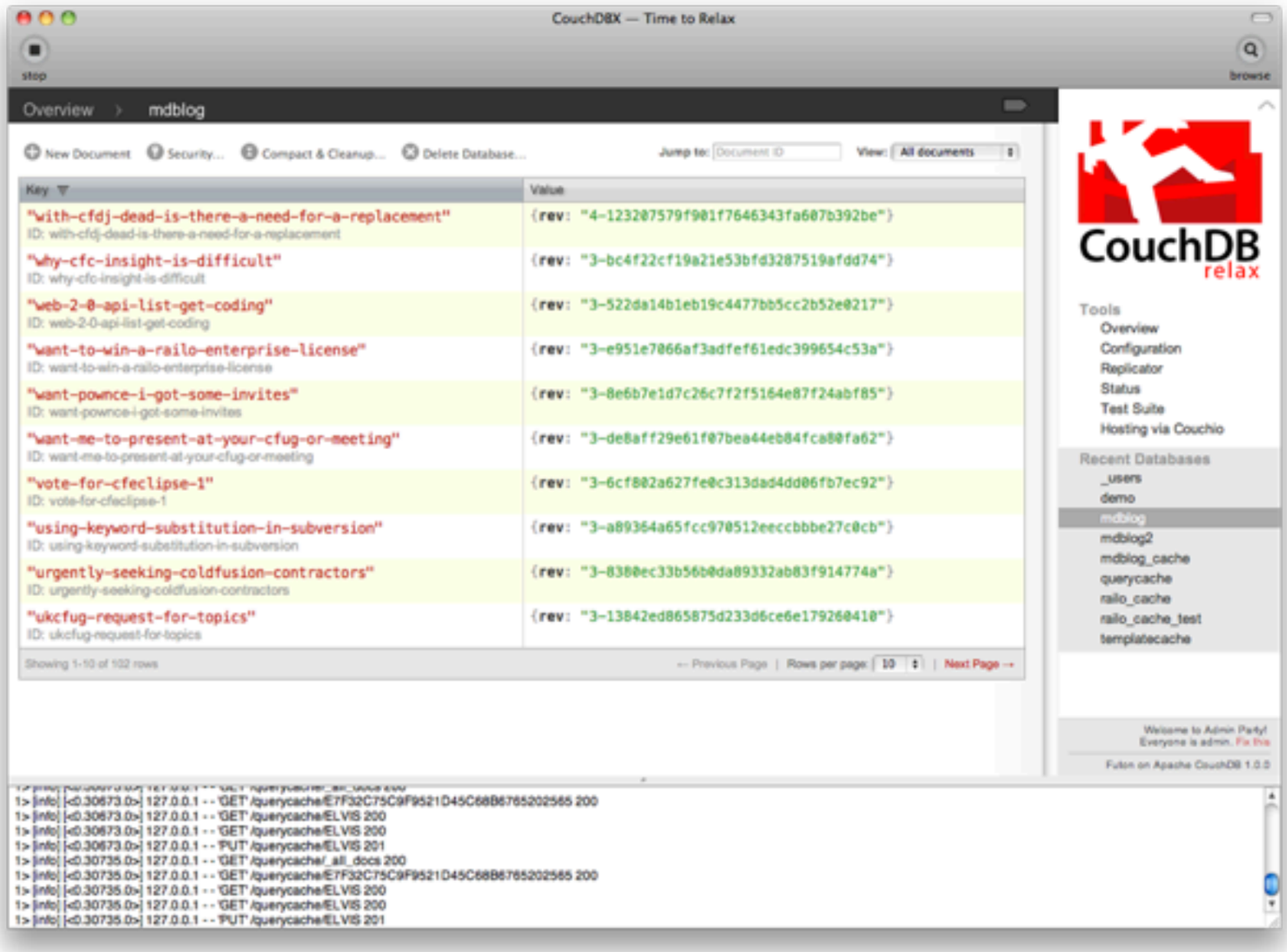


- Unstructured data
- HTTP environment
- Replication for SCALING!
- Mobile/Cloud/All OS
- Whole App Dev if you want!

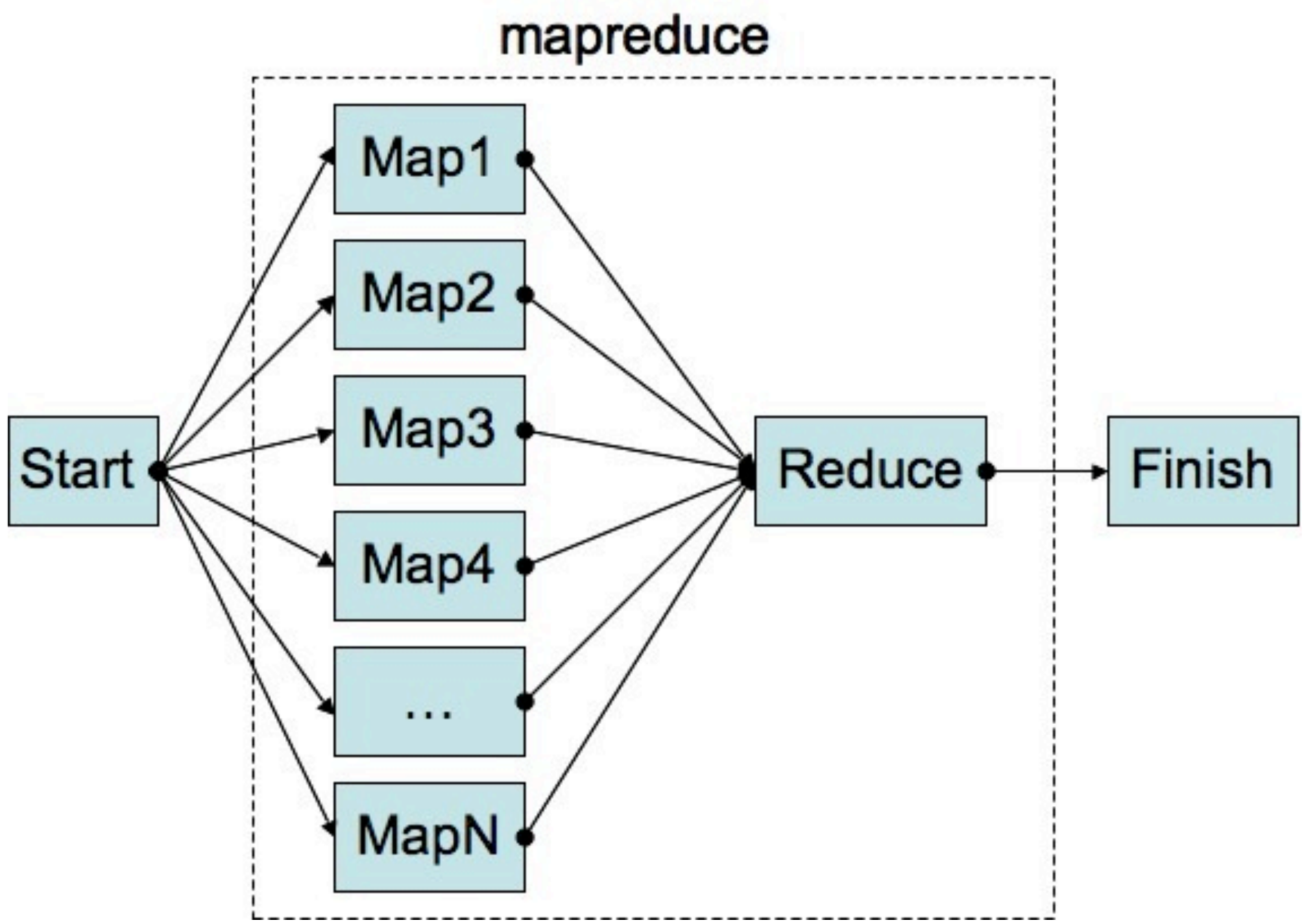
Built on Erlang
Concurrent functional language
Designed by Ericsson to support distributed, fault-tolerant, soft-real-time, non-stop applications

CouchDB Demos

- Futon
- HTTP/REST/JSON
- Replication



CouchDb "Queries"



No such thing as a query
Called Views, key value pairs, made up of a map & a reduce
Views defined by Map functions(in JavaScript)
Accessed via a URL (getting the hint?)

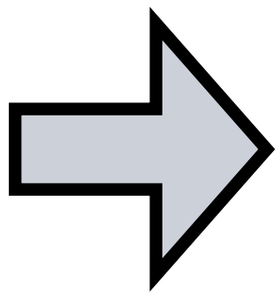
Couch DB Example Map:

```
function(doc) {
   if(doc.tags){
      for(var t in doc.tags){
         emit(doc._id, doc);
      }
   }
}
```

CouchApps?

(or A.N. Other Web Tech)

CouchApps?



http://couchapp.org

CouchApps!

> couchapp generate blog

> cd blog

> couchapp push http://localhost:5984/blog/

> http://localhost:5984/blog/_design/blog/index.html

http://couchapp.org

More on CouchDB

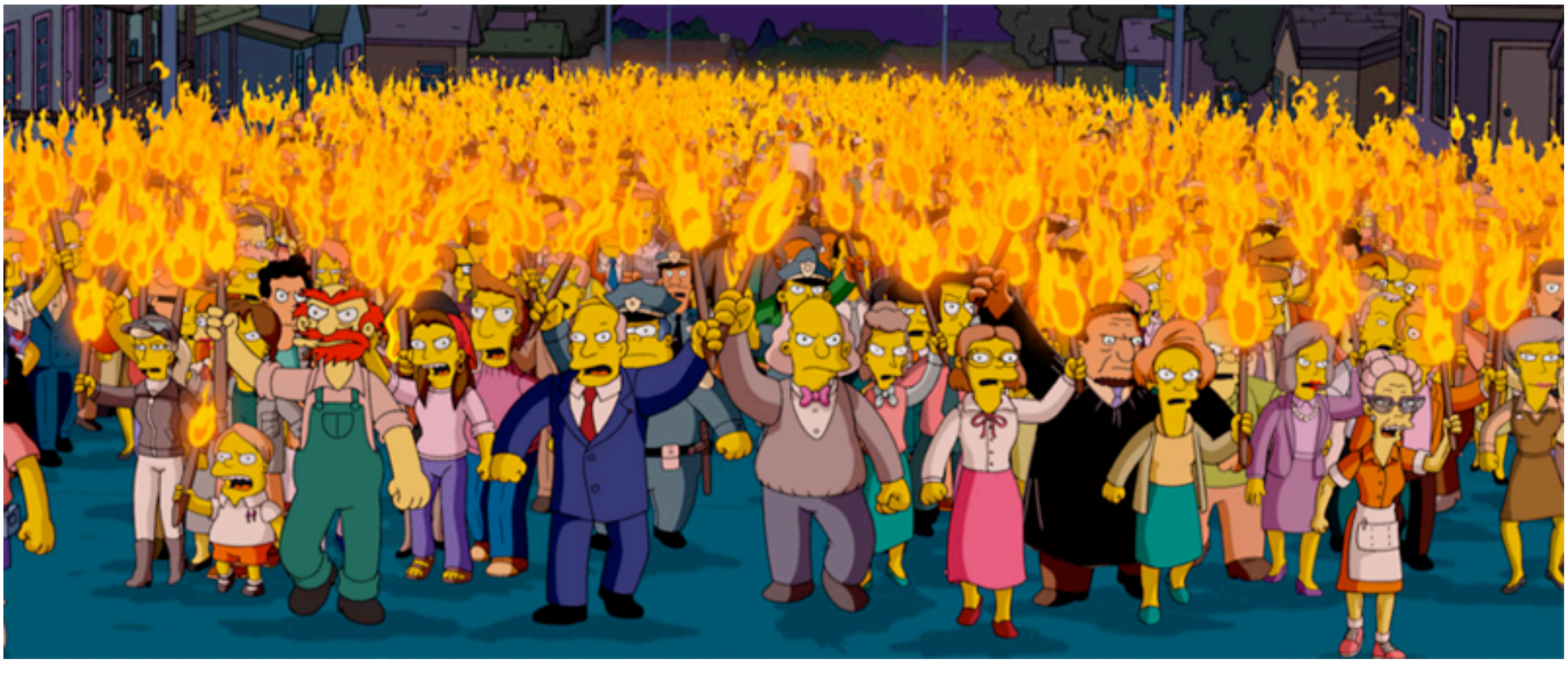
- Check out http://www.couchbase.com/
- Download/Run/Play
- Free Book http://guide.couchdb.org/
- Apps! http://couchapp.org

Takeaways

- CAP Theorem
- New tools for your toolbox
- Not everything has to be a DB

>

Thoughts?



Finally

tw: #cfobjective & @markdrew

e: mark@getrailo.com

w: http://www.getrailo.com

v: http://bit.ly/couch_cfo

n: http://bit.ly/doc_cfo

How many people were there?