F*#$! your formats, just gimme u're data... (reads: 'Collective Intelligence')

claimid.com/wesyoung

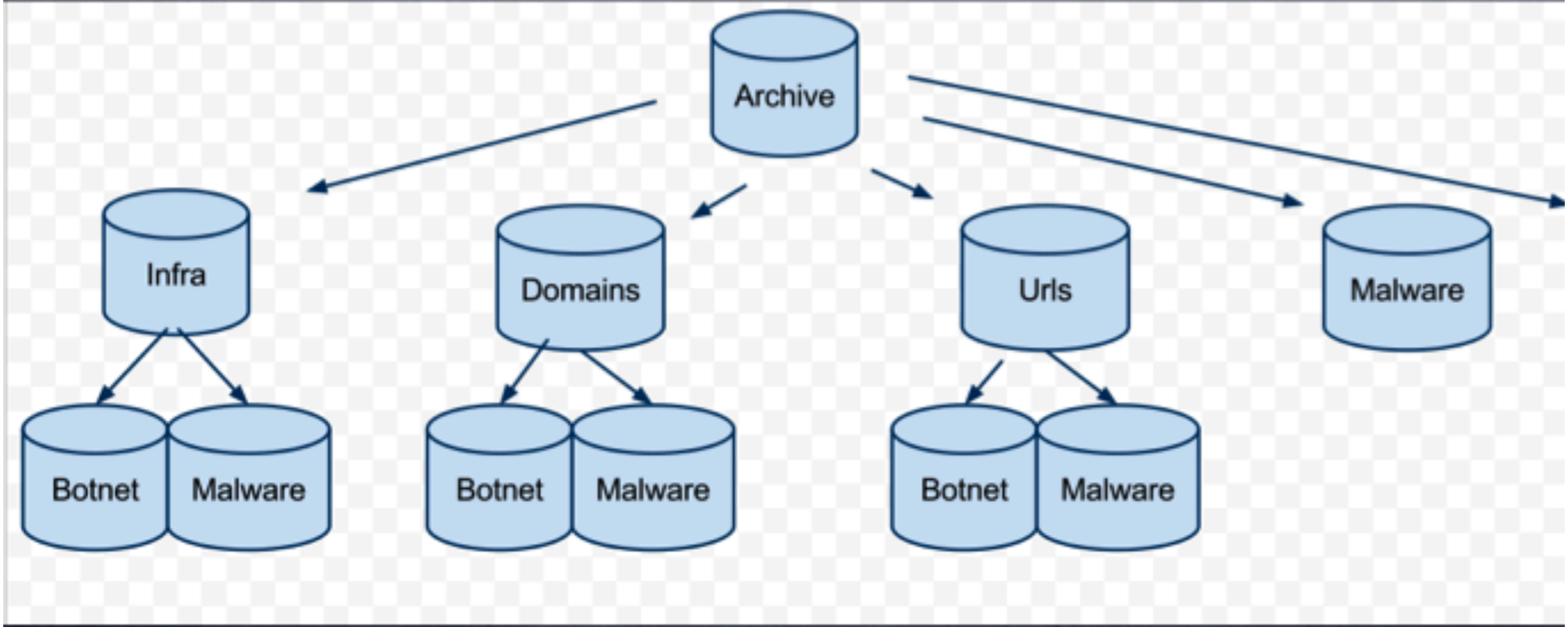
Collective Intelligence

- at it's core; it's a simple re-implementation of how social networks appear to store data. However, this framework takes in both structured (xml) and unstructured messages; indexes them on the fly, then allows you to "do stuff" to them via a set of index tables (search the index tables; relate back to the original data messages).
- This framework pulls in various data-records from any source (irregardless of ‘format’); normalizes it using whatever ‘format’ you desire; creates a series of messages "over time" (eg: observations generating reputation), then when you "do stuff" (correlation, etc), you look at a series of messages chronologically and make decisions much as you would look at an email thread around an address/hash/malware/md5/etc...
- allows you to cast your own “severity” and “weight” (reads: bias) to the multitude of data intel in cyberspace that meets your own business requirements.
- http://bret.appspot.com/entry/how-friendfeed-uses-mysql

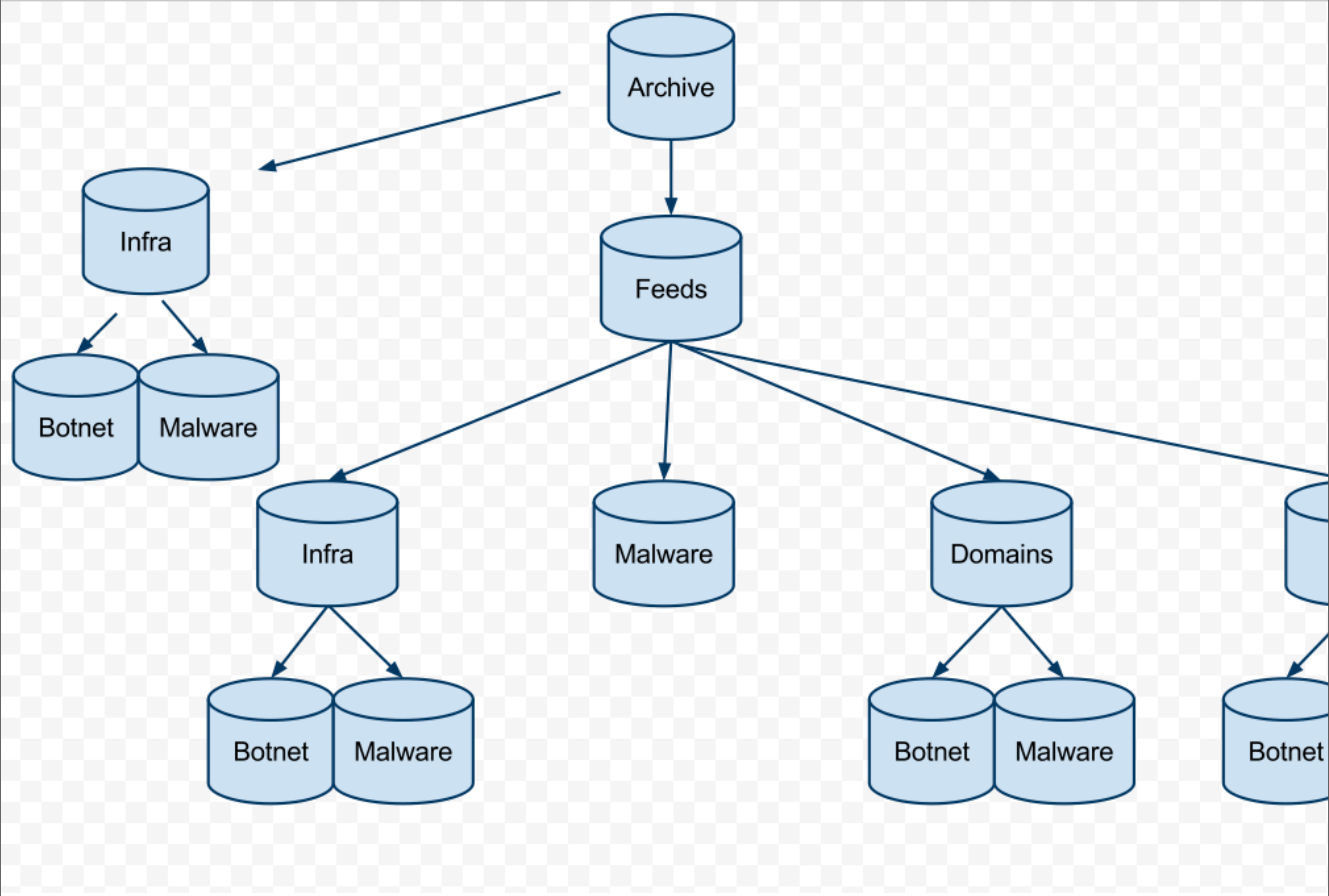
Collective Intelligence

- Allows you to....
  - evolve “new tables” on the fly from existing data without breaking existing applications
  - Tables act as a linear search index to the original data
  - actually USE the data that’s avail in the public domain with simple and open technologies (reads: the sweet spot between high end architecture and munging text-files)

searching the archive



the point, is to get back to the original data blob as quickly as possible. You can use the data in the index, but you don't have to.


Archive
Infra
Feeds
Botnet
Malware
Infra
Malware
Domains
Botnet
Malware
Botnet
Malware
Botnet

Collective Intelligence

- Everything is a plugin.... from storage to the REST API
- wanna store your data as straight up json, done by default.
- wanna store your data as IODEF, install the IODEF storage plugin (eg: a simple driver model, the plugin does all the work, the framework just passes along the data)
- no more "ip=xxx.xxx.xx..." or "domain=example.com" or "md5=xxxxx" garbage. The API figures out (regex) what you're trying to send and searches it's appropriate index

Collective Intelligence

- www.ren-isac.net/ses
- http://code.google.com/p/collective-intelligence-framework/