

small mind,
big data

michael.glaesemann@myyearbook.com
senior data architect

Emerging Technologies for the Enterprise
Philadelphia, PA 2011-04-28

**we're
hiring!**

<http://www.myyearbook.com/apps/careers>

Who?

Luke

normalization

orthogonality

degrees of
freedom

measure

Einstein

日本

teach

fax

UTF-8

PostgreSQL

community

Codd

Darwven

Date

relational model values & operators

normalization

orthogonality

first-order predicate logic

PHP

Ruby, Rails

object-relational
impedance mismatch

imperative
versus
declarative

logical
versus
physical

thought

word

deed

functional

NoSQL

myYearbook.com

casual social network

founded in 2006

Google Analytics

by gender

52% female

48% male

by age

37% 13–17

27% 18–24

13% 25–34

12% 35–44

11% 45+

myYearbook Teens and Twitter Survey (August 2009, ~%)

myYearbook

70% meet people

50% play games

40% keep in touch with friends

35% flirt/date

Facebook

80% keep in touch with friends

35% meet people

30% share photos

25% play games

comScore Teens Category (April 2009)

| rank | site | visits (K) | uniques (K) | minutes (M) | page views (M) |
|----------|---------------------|---------------|--------------|-------------|----------------|
| 1 | myYearbook | 55,808 | 4,604 | 851 | 1,630 |
| 2 | MEEZ | 8,629 | 1,407 | 226 | 469 |
| 3 | Zwinky | 11,558 | 3,691 | 153 | 108 |
| 4 | Hearst Teen Network | 6,940 | 2,314 | 53 | 117 |
| 5 | Quizilla | 7,924 | 2,058 | 75 | 75 |

comScore page views
July 2009

| rank | site | views (M) |
|-----------|-------------------|------------|
| 20 | GaiaOnline | 1,105 |
| 21 | Chase | 1,056 |
| 22 | ESPN | 984 |
| 23 | myYearbook | 953 |
| 24 | Wikipedia | 903 |
| 25 | Onemanga | 840 |
| 26 | Mapquest | 833 |
| 27 | Foxsports | 815 |

comScore time spent
July 2009

| rank | site | minutes (M) |
|-----------|-------------------|-------------|
| 20 | Amazon | 806 |
| 21 | CNN | 744 |
| 22 | GaiaOnline | 709 |
| 23 | Bing | 707 |
| 24 | MSNBC | 691 |
| 25 | myYearbook | 678 |
| 26 | Iwin | 670 |
| 27 | NickJr | 665 |

2007

100M HTTP requests/month

1 PostgreSQL server

1 Apache/PHP server

static content servers

1 outage per night

RT @DrErnie:

Scalability is the ability to use abstraction;

Efficiency is the ability to avoid abstraction.

Performance comes from knowing which to use.

resources

1

use

abstraction

function

API

schema

avoid

abstraction

(1)

measure

pgfovine

process

1

avoid

abstraction

(2)

Relative data access latencies

| | |
|---------------|-------------------|
| 1 | CPU Register |
| 1–2 | L1 cache |
| 6–10 | L2 cache |
| 25–50 | main memory (RAM) |
| 10,000,000 | hard disk |
| 100,000,000 | LAN |
| 1,000,000,000 | WAN |

fsync

memory

2008

2.5G page views

30 PostgreSQL servers

120 Apache/PHP servers

8 ActiveMQ message brokers

99.84% uptime

2010

2.5G HTTP requests/month

Database servers

connection poolers, routers

Web servers

Apache/PHP, Tornado/Python, lighttpd

Message brokers

ActiveMQ, RabbitMQ

Monitoring, R&D, other servers

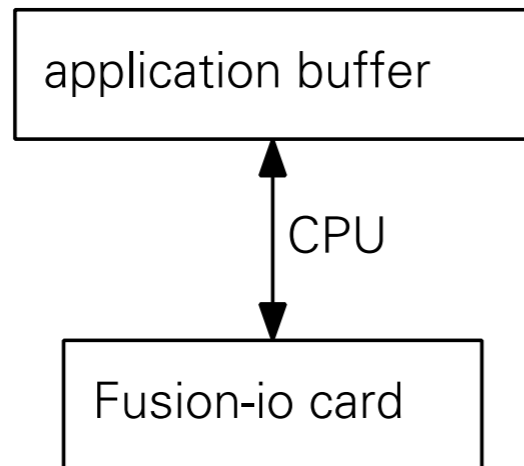
avoid

abstraction

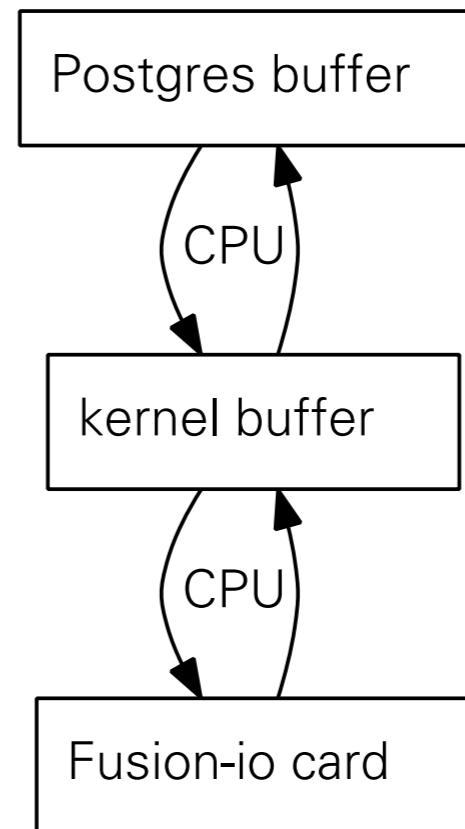
(3)

Fusion-io

direct IO



buffered IO



logical
versus
physical

mind

body

diet
exercise

rapid
growth

bottleneck

time

thinking

represent

高橋 喬

たかかはし

夕カハハシ

Takahashi

高橋
たかはし
Takahashi

Glaeseemann

グレスズマン

gurezman

gure**z**uman

concepts

7 ± 2

4

1, 2, many

counting
is an
algorithm

biggger than
your head

levels

names

colors

symbols

math

the territory
is not
the map

visualization

but!

more activity?

views/month 2007 100M → 2009 1.5G

more TPS

more servers

more connections

more configuration

more pain!

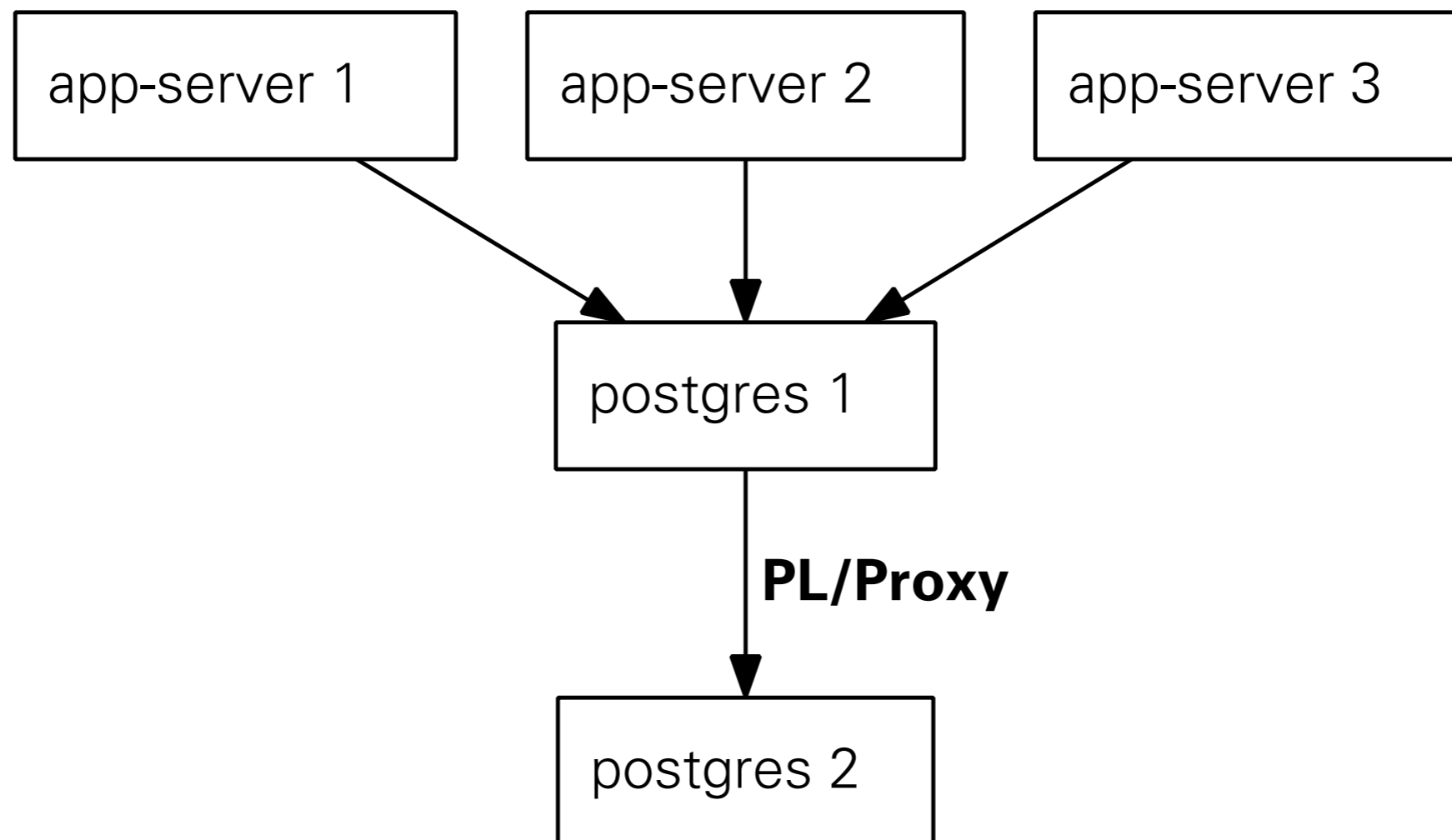
Skype

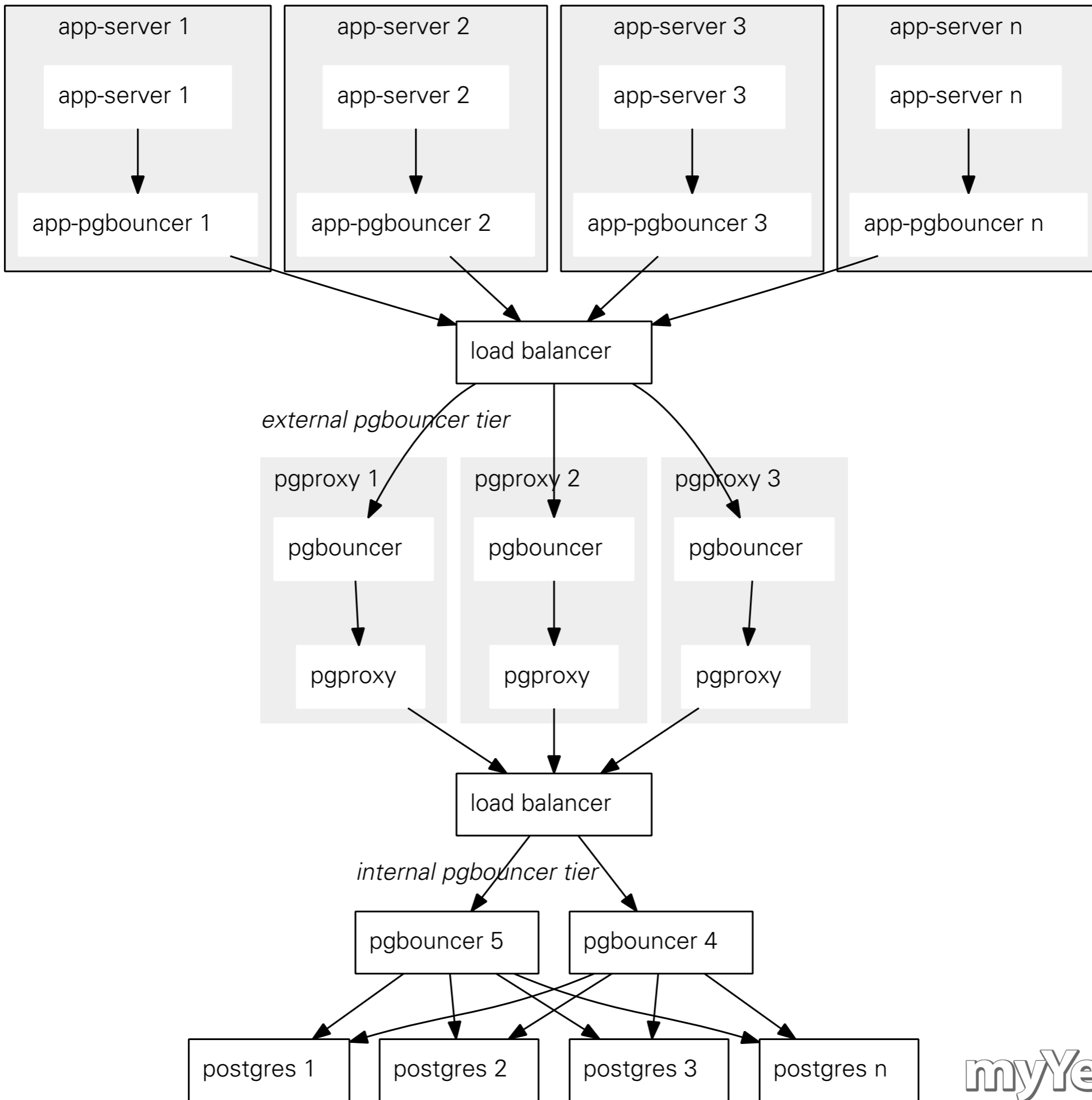
PL/Proxy

pgBouncer

less configuration

fewer connections





1

DRY

truth

everything in
its right place

SOA

REST

DB API

time

generate

tools

measure

belief

worldview

reason

versus

rationalization

smarter

narrow

science

learn

blub

improve

automate

test

queuing theory

under-constrained

small mind,
big data

michael.glaesemann@myyearbook.com
senior data architect

We're hiring!

<http://www.myyearbook.com/apps/careers>

Selected references

Brooks, Frederick P., Jr.; (1995) The Mythical Man-Month.

Codd, E. F.; (1970) "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM 13 (6): 377–387.

Cowan, N.; (2001) "The magical number 4 in short-term memory: A reconsideration of mental storage capacity." Behavioral and Brain Sciences, 24, 87-185.

Darwen, Hugh; Date, Christopher J; (March 1995) "The Third Manifesto", ACM SIGMOD Record (New York, NY, USA: ACM Press) 24 (1): 39–49.

Date, C. J.; (2009) SQL and Relational Theory: How to Write Accurate SQL Code.

Eames, Ray; Eames, Charles; (1968) "Powers of Ten".

Foer, Joshua; (2011-02-15) "Secrets of a Mind-Gamer: How I trained my brain and became a world-class memory athlete." <http://www.nytimes.com/interactive/2011/02/20/magazine/mind-secrets.html> Retrieved 2011-04-27.

Graham, Paul; (2001-04) "Beating the Averages" <http://www.paulgraham.com/avg.html> Retrieved 2011-04-28.

Hickey, Rich; (2010) "Hammock-Driven Development", Clojure Conj 2011 keynote address.

McConnell, Steve; (2004) Code Complete: A Practical Handbook of Software Construction, 2d Edition.

McDougall, Christopher; (2009). Born to Run.

McNally, Joe; (2009). The Hot Shoe Diaries: Big Light from Small Flashes.

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review 63 (2): 81–97.

Mooney, Chris; (2011-04-18) "The Science of Why We Don't Believe Science: How our brains fool us on climate, creationism, and the vaccine-autism link."; <http://m.motherjones.com/politics/2011/03/denial-science-chris-mooney> Retrieved 2011-04-27.

Pinker, Steven; (2007) The Stuff of Thought: Language as a Window into Human Nature.

Reynolds, Garr; (2005-09-13) "Living large: 'Takahashi Method' uses king-sized text as a visual" http://presentationzen.blogspot.com/presentationzen/2005/09/living_large_ta.html Retrived 2011-04-27

Seibel, Peter; (2005) Practical Common Lisp.

Taubes, Gary; (2007) Good Calories, Bad Calories.