

Big Data

MARTIN FOWLER

ThoughtWorks®

<http://martinfowler.com>
<http://thoughtworks.com>
@martinfowler

REBECCA PARSONS CTO

ThoughtWorks®

rparsons@thoughtworks.com
<http://thoughtworks.com>

Big Data

MARTIN FOWLER

ThoughtWorks®

<http://martinfowler.com>
<http://thoughtworks.com>
@martinfowler

REBECCA PARSONS CTO

ThoughtWorks®

rparsons@thoughtworks.com
<http://thoughtworks.com>

The Evolving Panorama of Data

MARTIN FOWLER

ThoughtWorks®

<http://martinfowler.com>
<http://thoughtworks.com>
@martinfowler

REBECCA PARSONS CTO

ThoughtWorks®

rparsons@thoughtworks.com
<http://thoughtworks.com>

Changing Nature of Data

Response

How we use data now

Changing Nature of Data

Response

How we use data now

Data is:

Data is: *Growing*

Data is: *Growing*



Jeff Xiong
@gigix

Following



"big data" is when the size of the data itself becomes part of the problem

35
RETWEETS

9
FAVORITES



6:00 AM - 27 Feb 12 via web · Embed this Tweet

[Reply](#) [Retweeted](#) [Favorite](#)

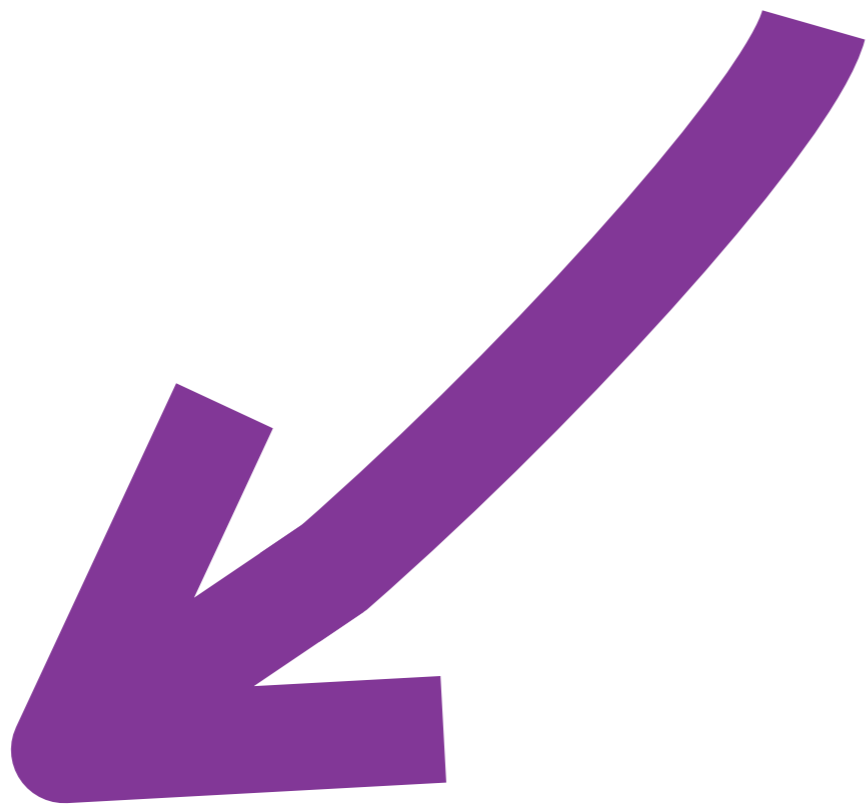
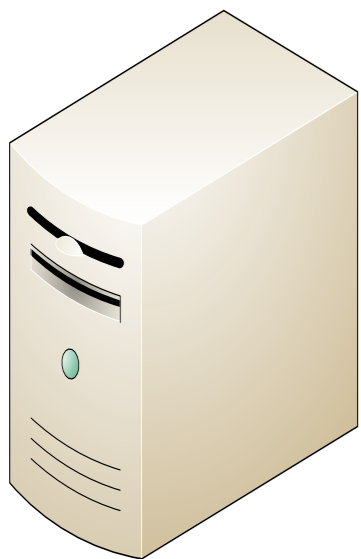
Data is: *Growing*

Walmart: 1 million transactions per hour

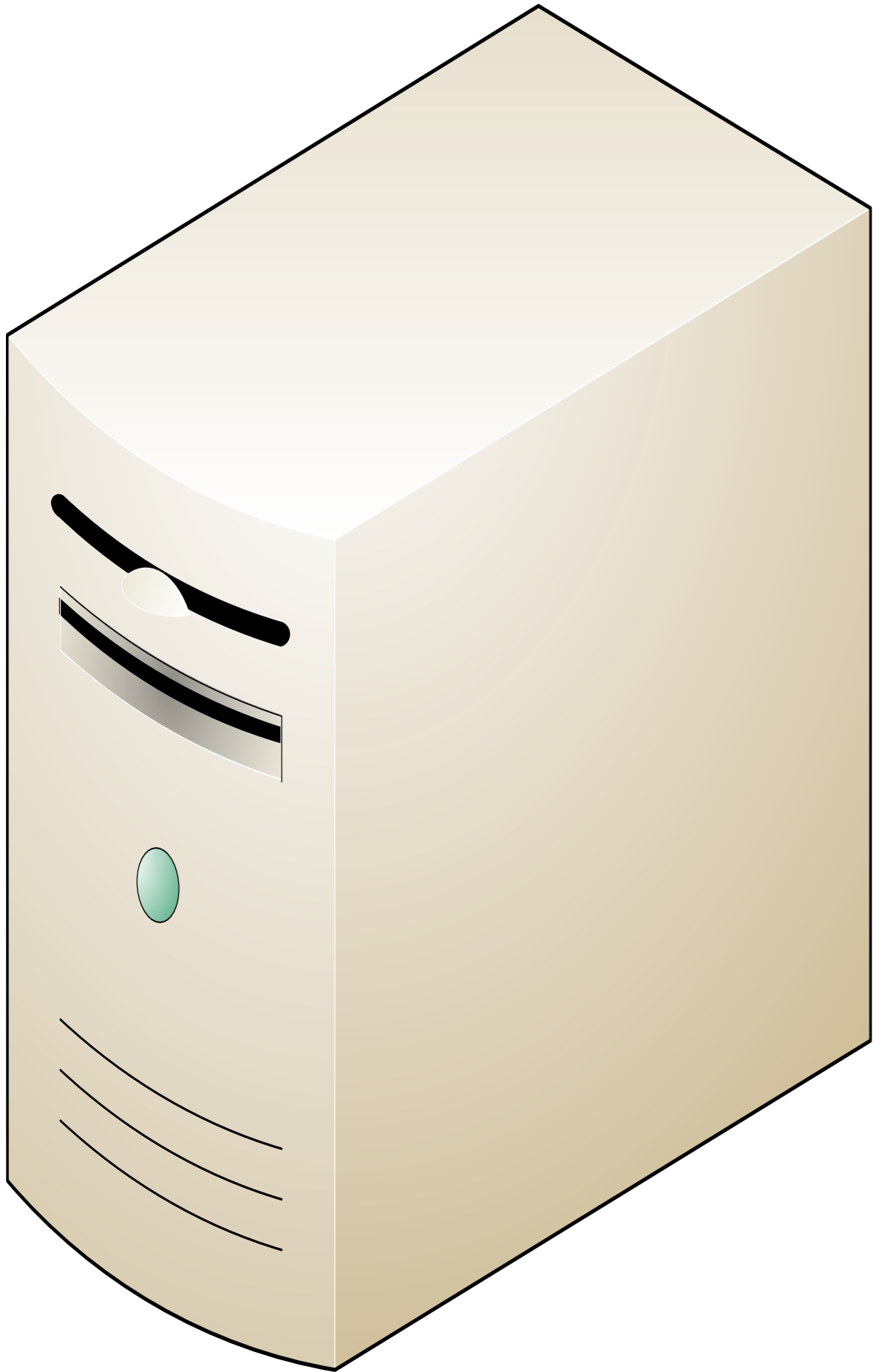
Facebook: 40 billion photos

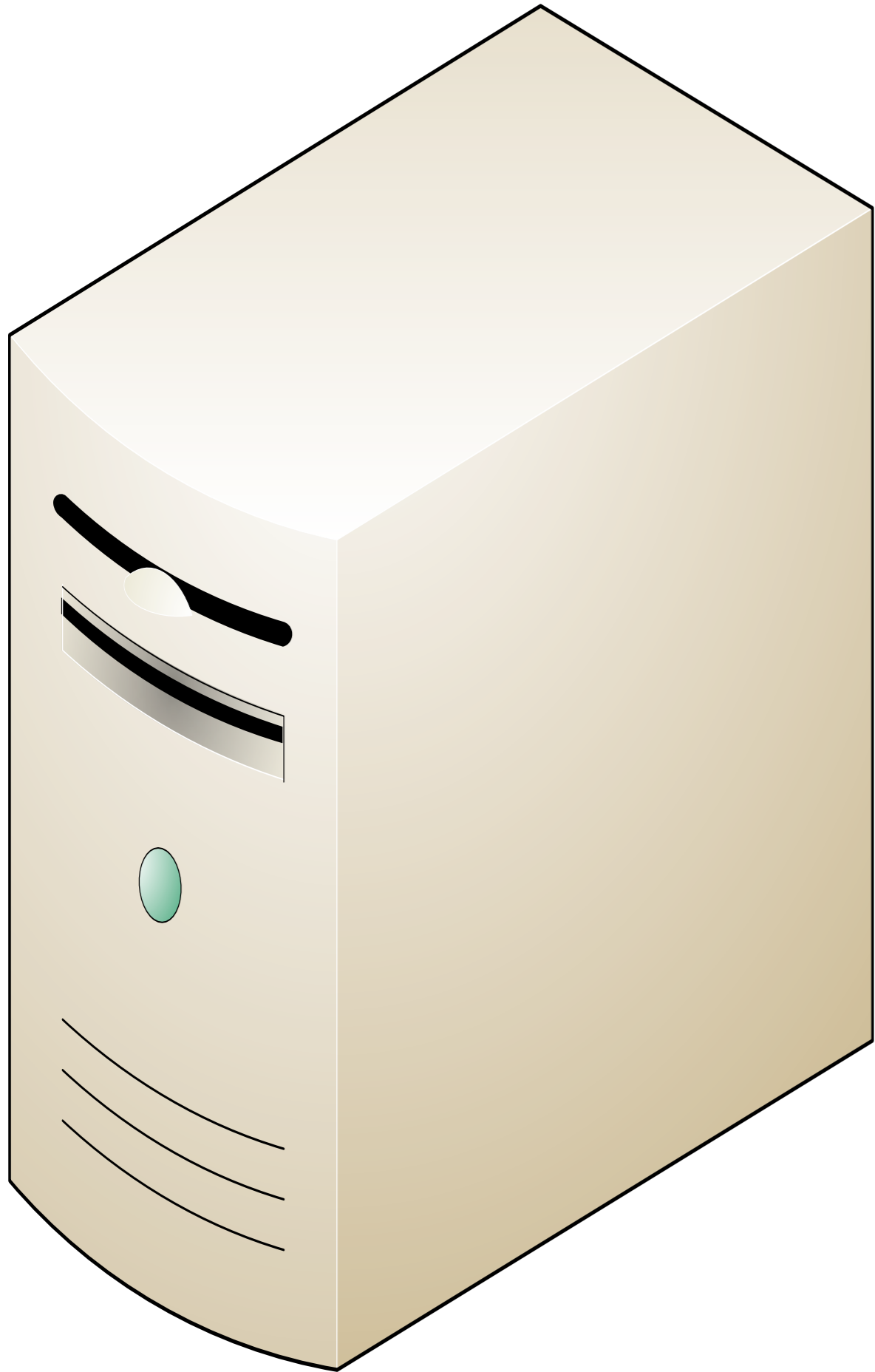
The Economist: Feb 25th 2010

640K ought to be
enough for anybody!



Lots of
Traffic

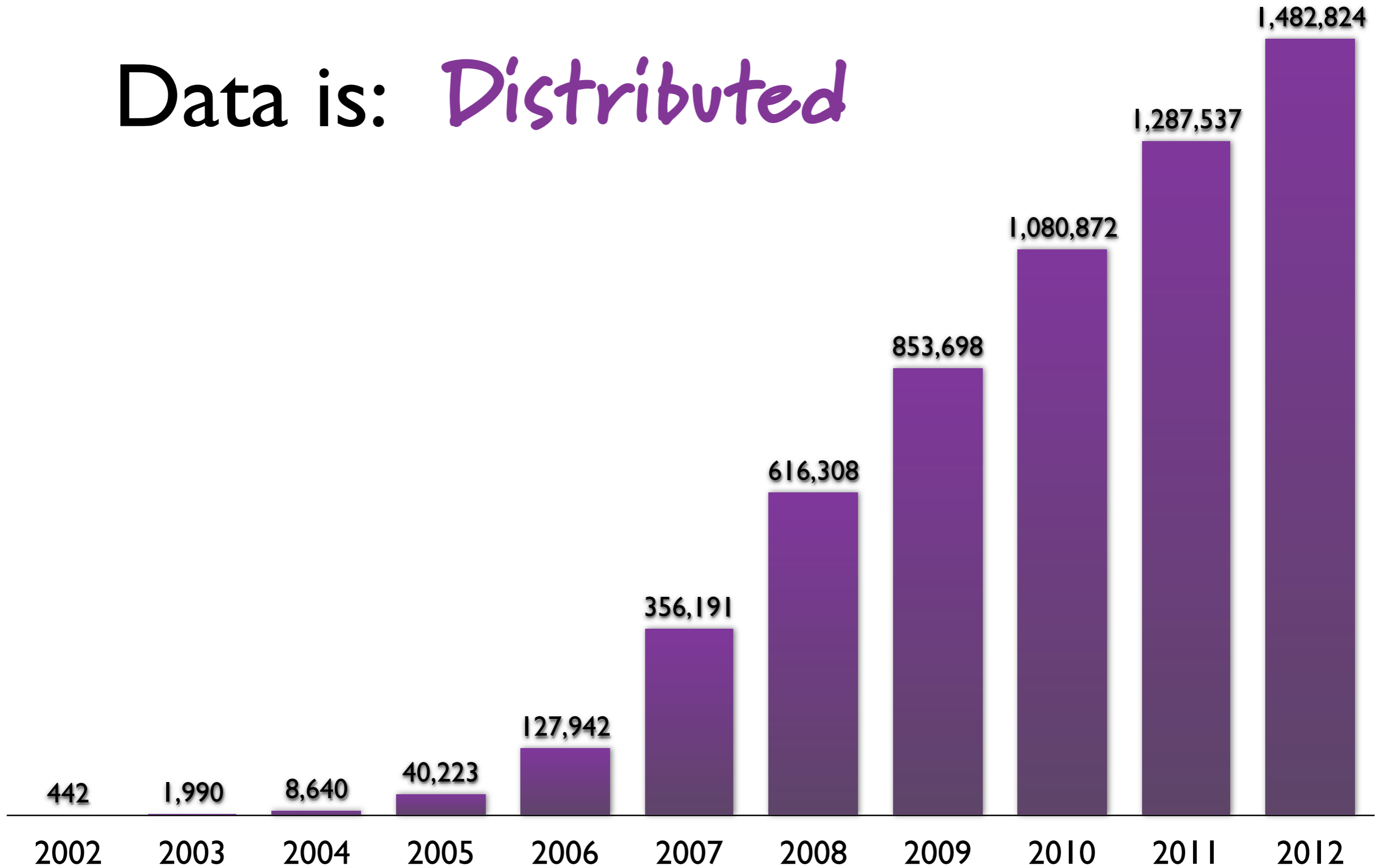




Data is:

Data is: *Distributed*

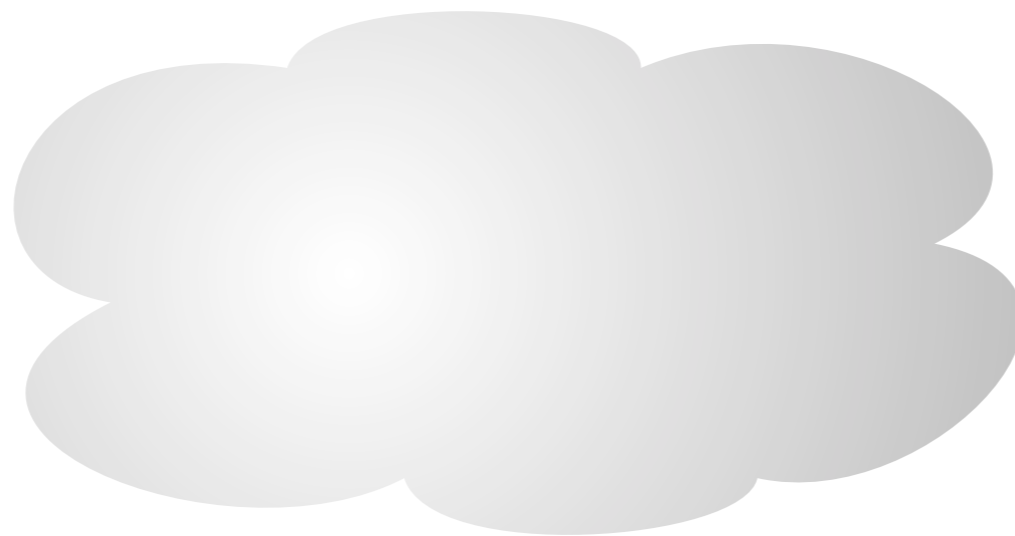
Data is: *Distributed*



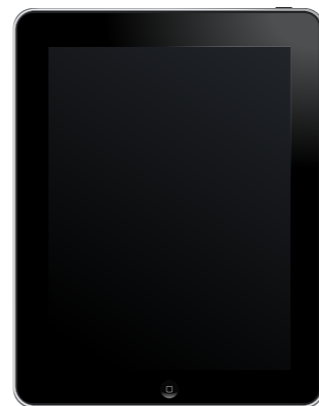
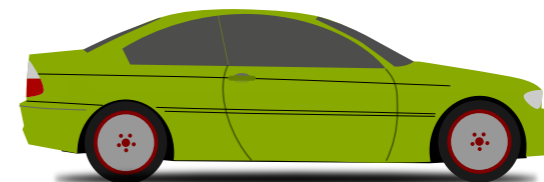
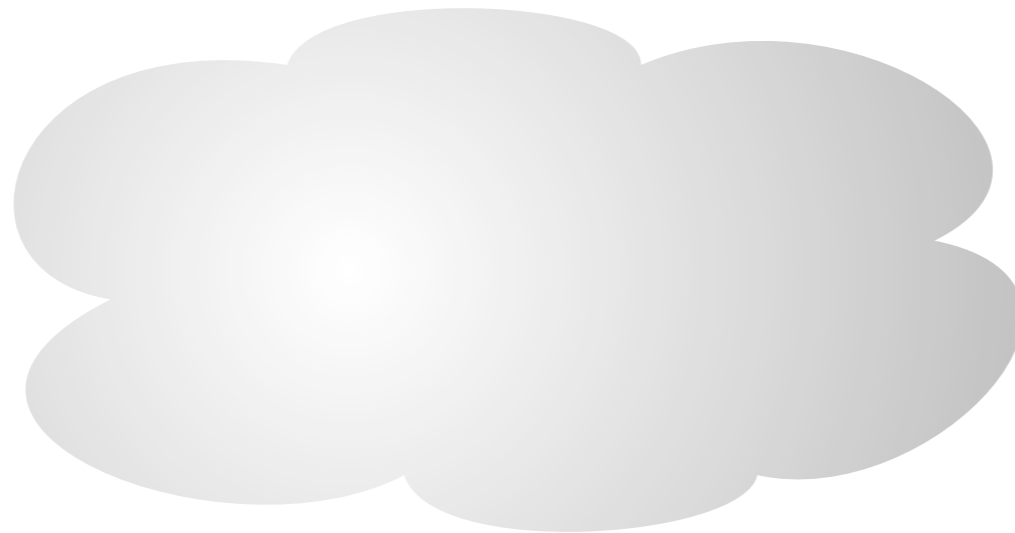
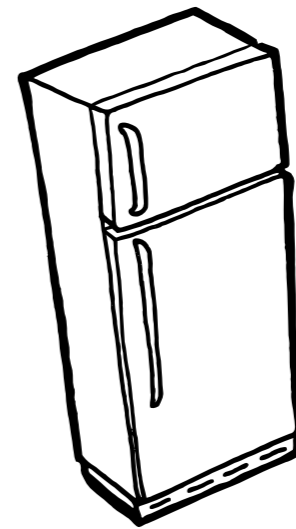
Monthly Contributors to Wikipedia

source: wikipedia

Data is: *Distributed*



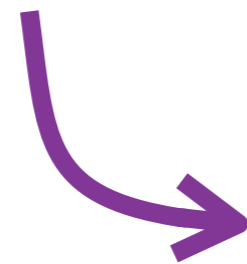
Data is: *Distributed*



Data is: *Distributed*

98% of internet access points in Africa are **mobile**

30 million networked sensor nodes



growing 30% per year

McKinsey Global Institute: Big data: The next frontier for innovation, competition, and productivity

Data is:

Data is: *Valuable*

Data is: *Valuable*

\$300 billion / year for US health care

60% increase in retail margins

McKinsey Global Institute: Big data: The next frontier
for innovation, competition, and productivity

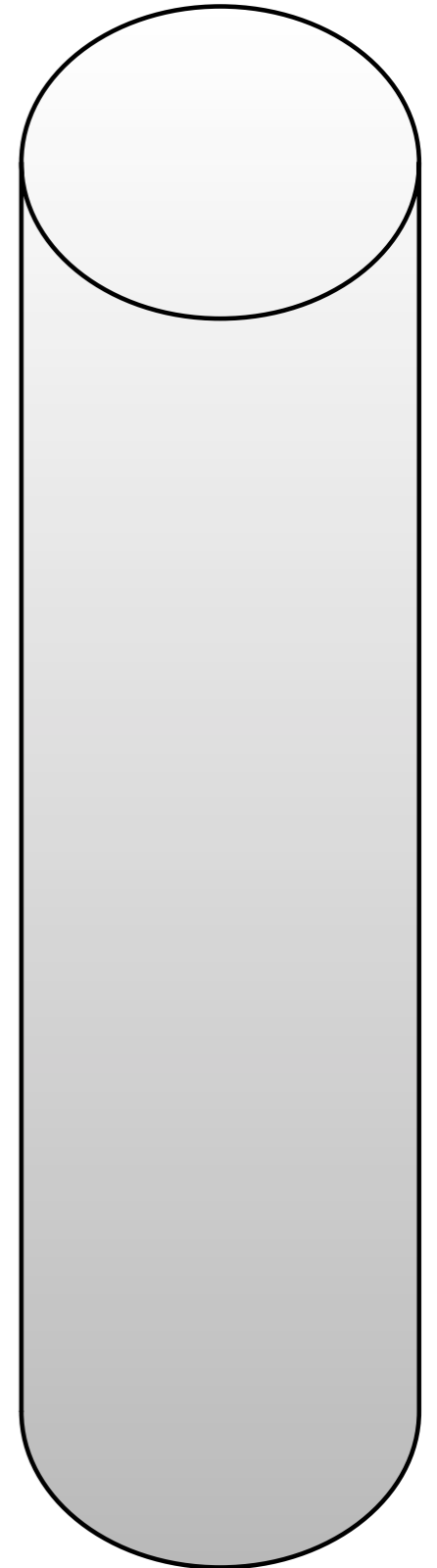
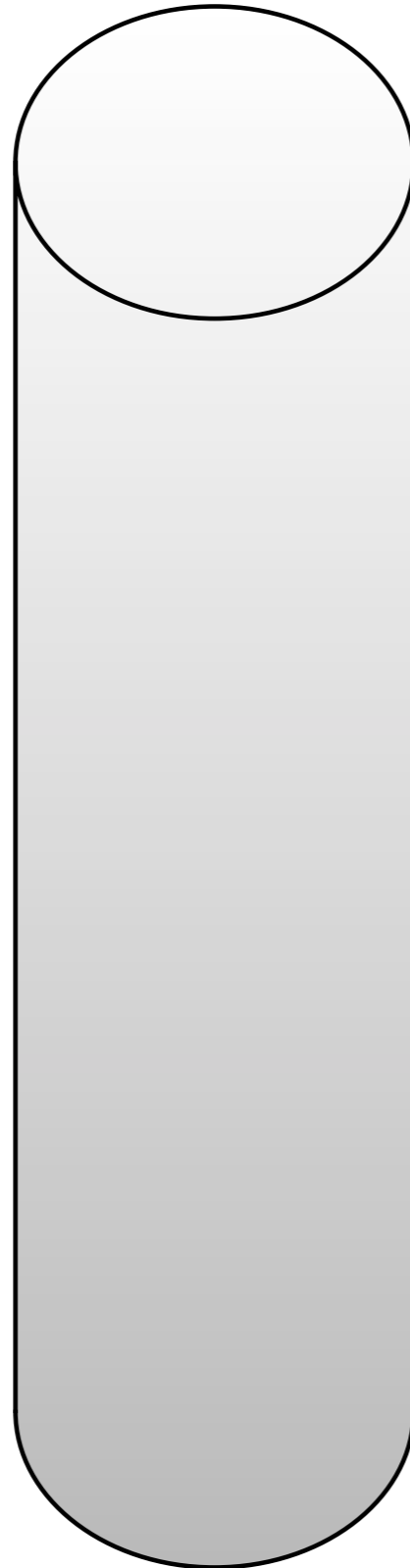
Data is:

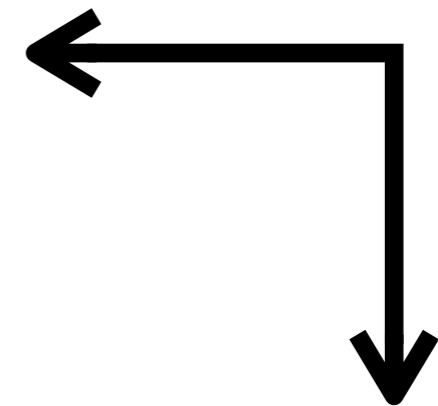
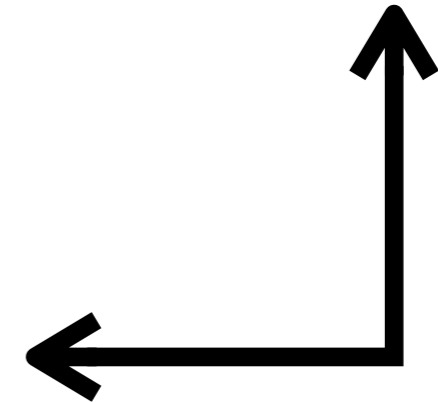
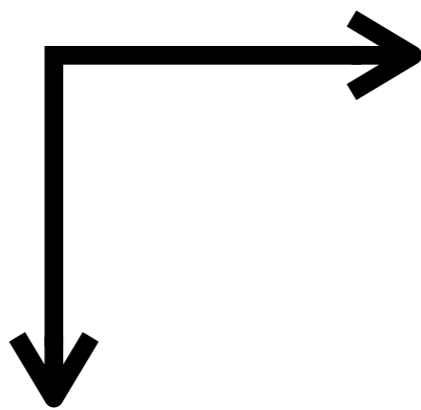
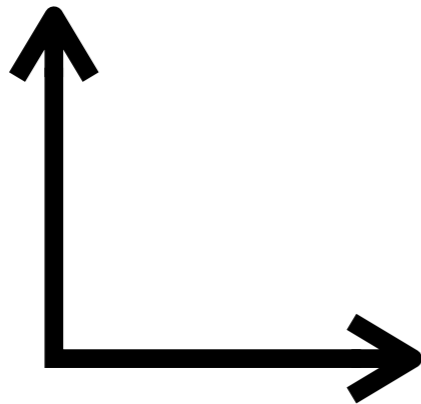
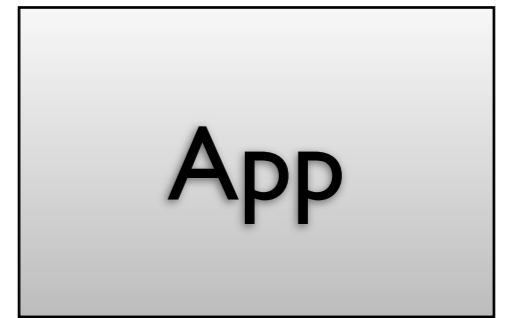
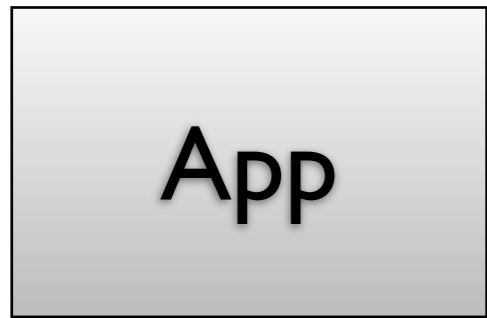
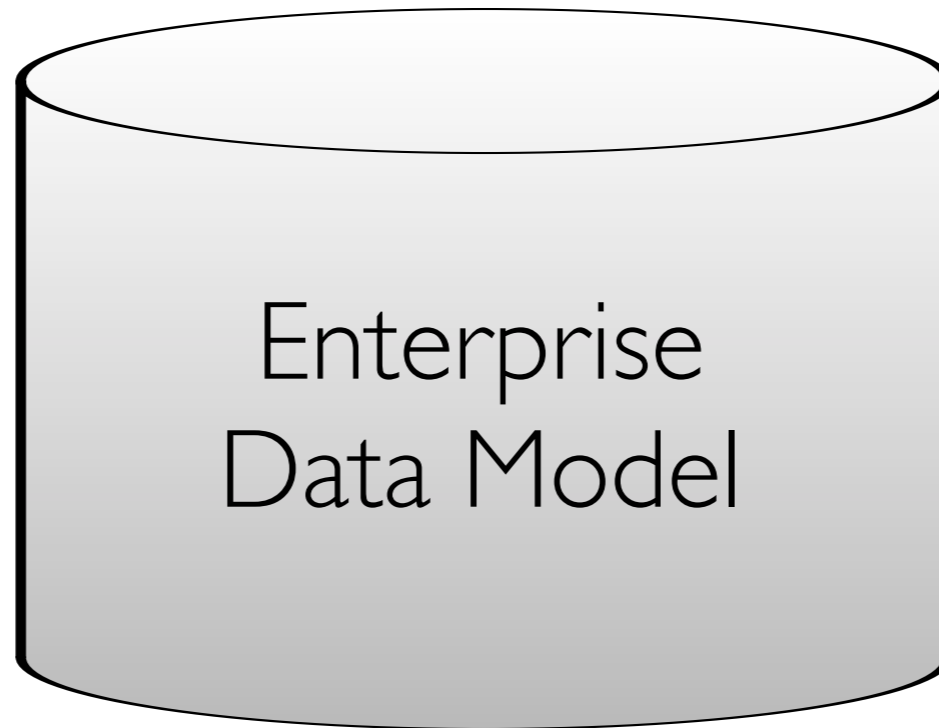
Data is: *Urgent*

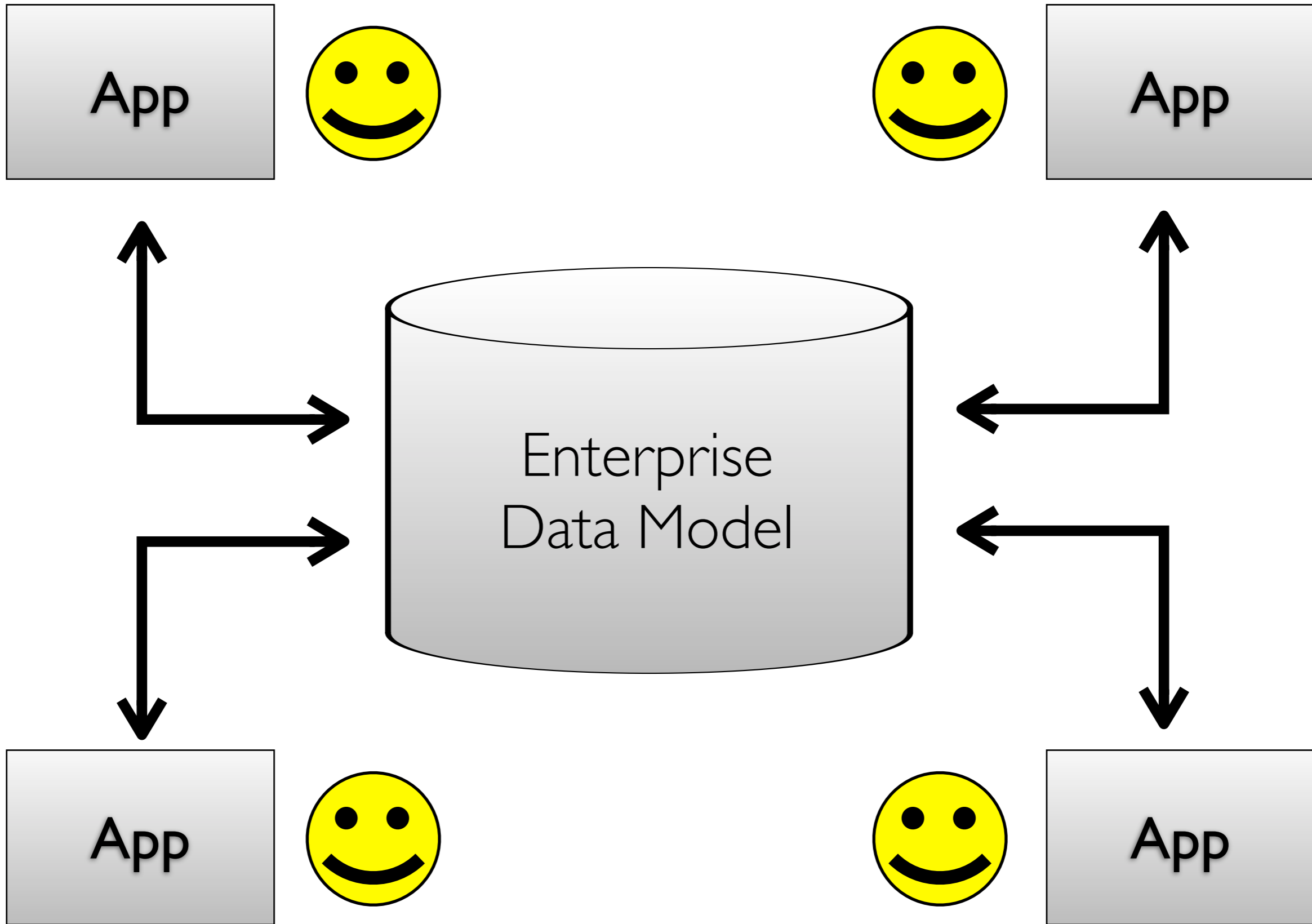
Data is:

Data is: *Connected*

Data Silos







Changing Nature of Data

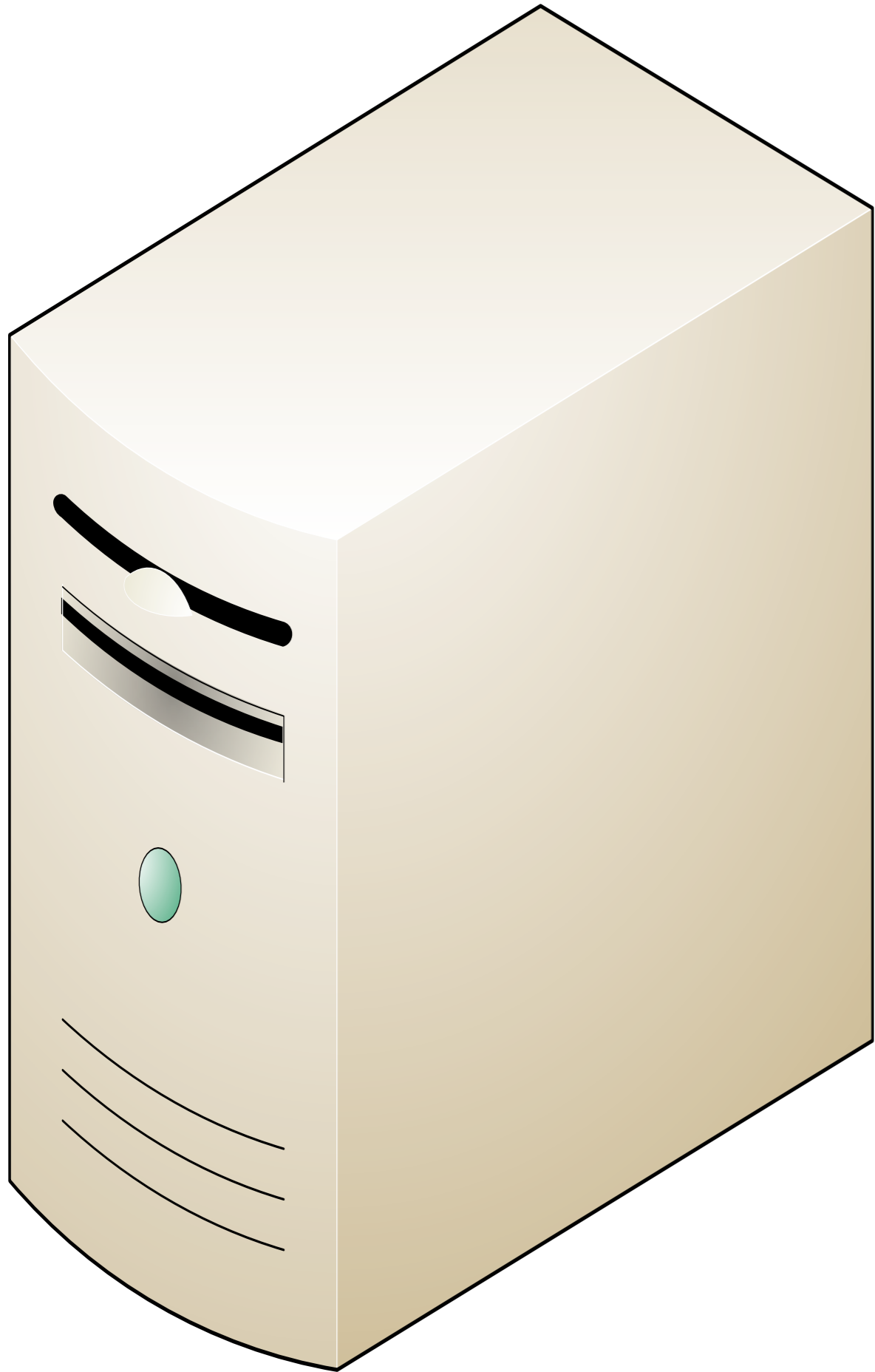
Response

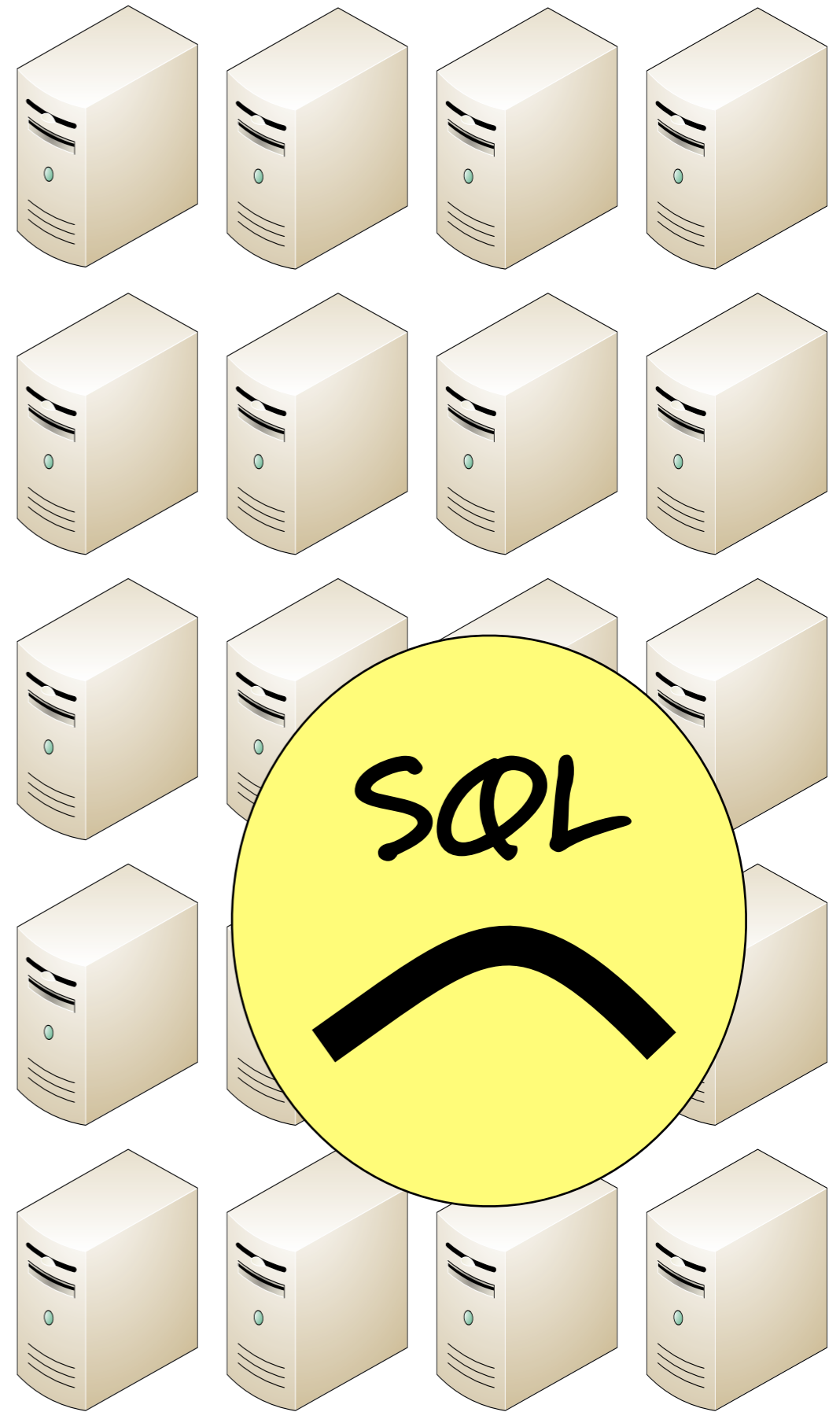
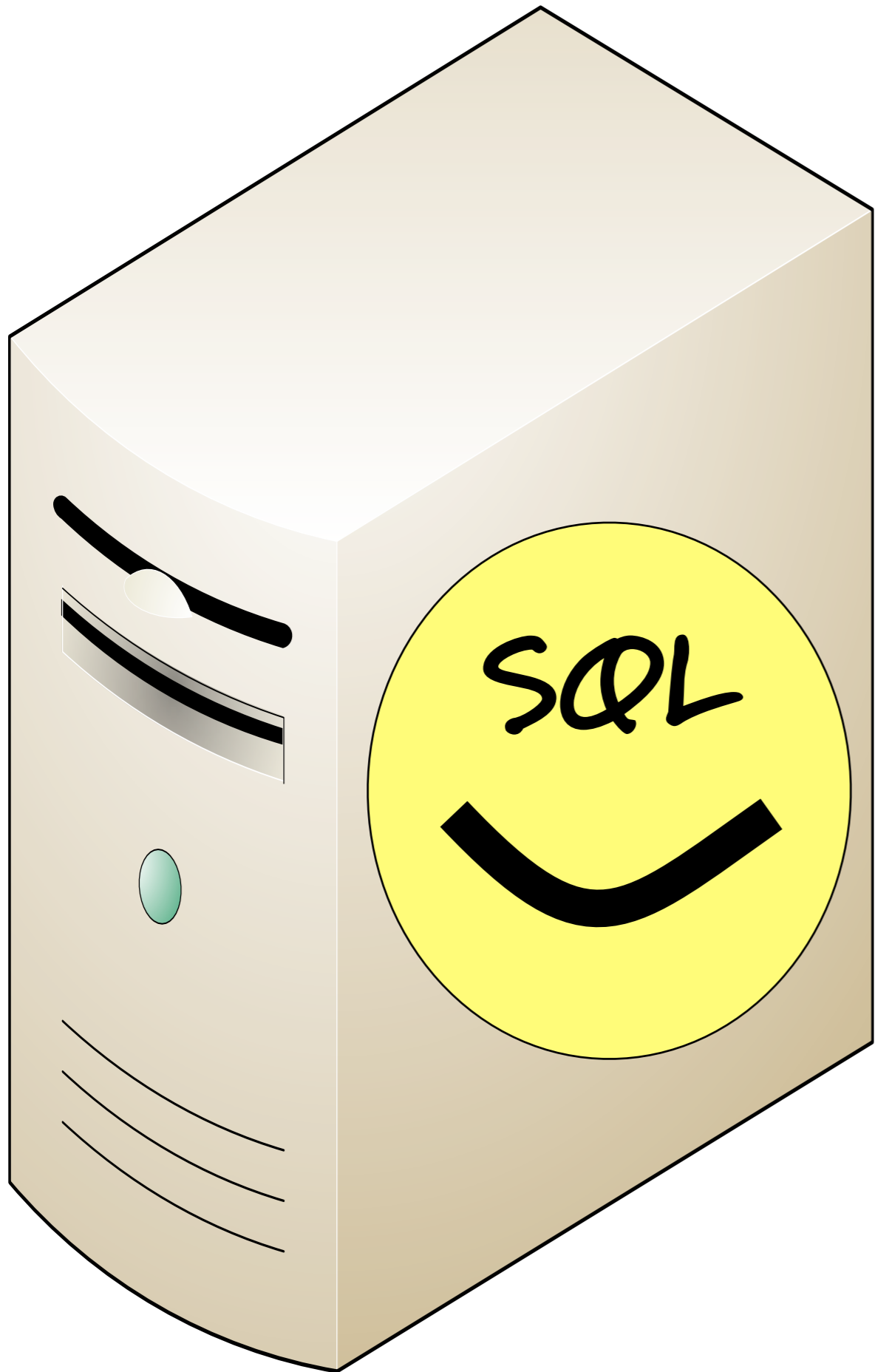
How we use data now

Changing Nature of Data

Response

How we use data now

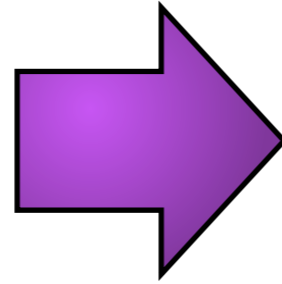




Google™

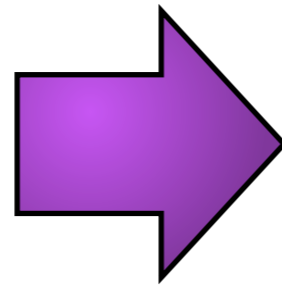
amazon.com®

Google™



Bigtable

amazon.com®



Dynamo

"NoSQL"

Definition of NoSQL

Characteristics of NoSQL

Characteristics of NoSQL

non-relational

open-source

cluster-friendly

21st Century Web

schema-less

Document



Column-family



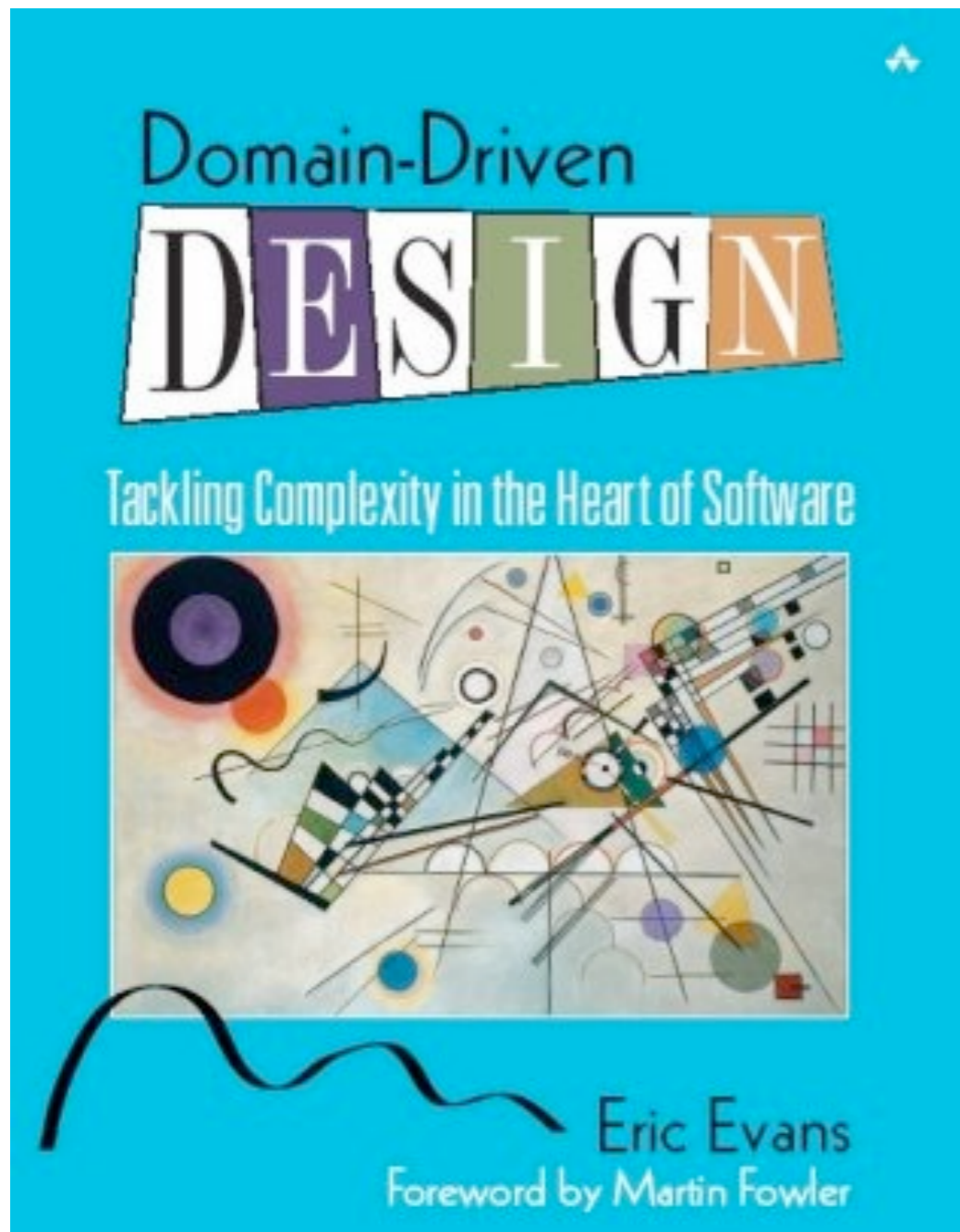
Graph



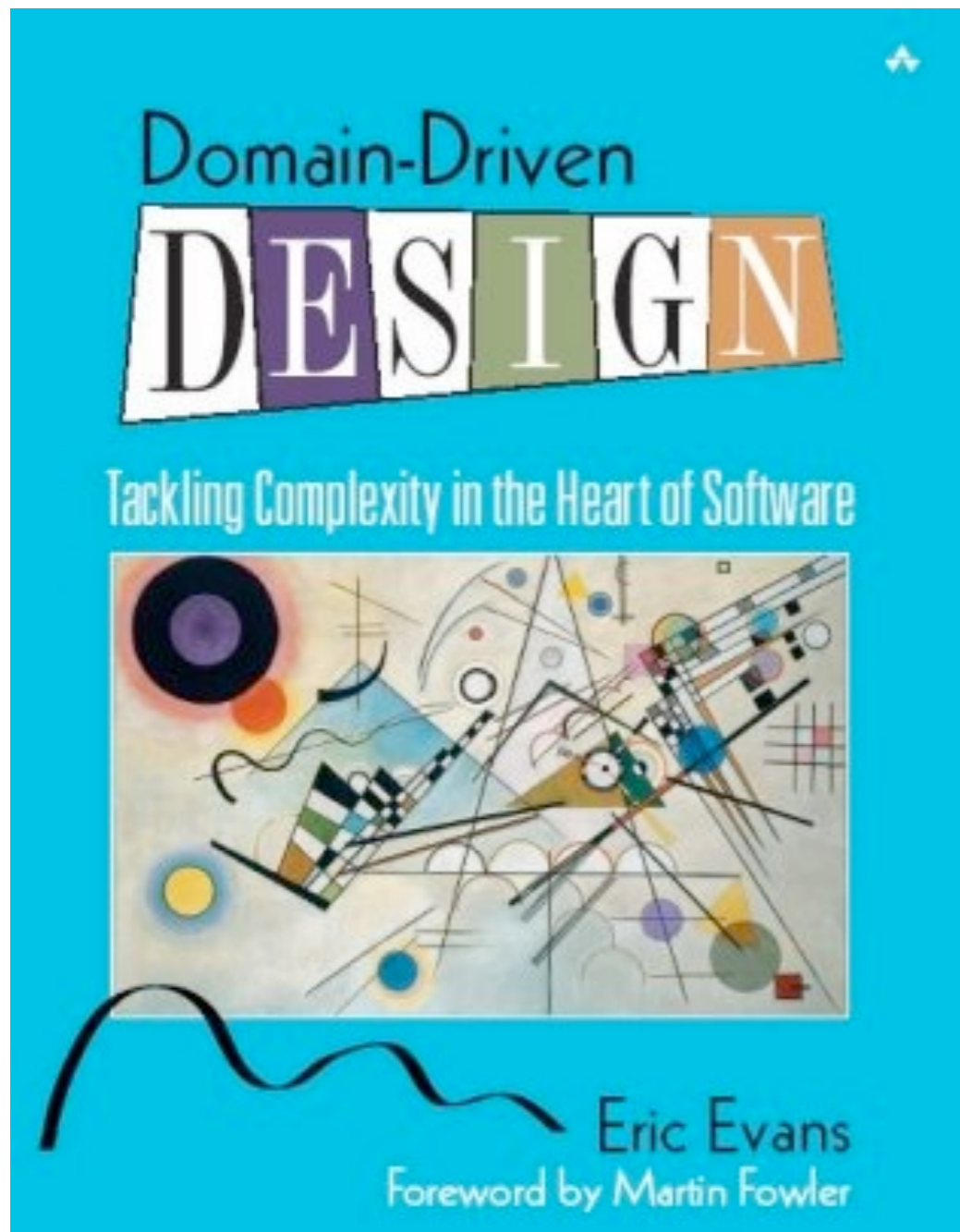
Key-value



Aggregate



Aggregate



Eric Evans

ID: 1001

customer: Ann

line items:

0321293533	2	\$48	\$96
0321601912	1	\$39	\$39
0131495054	1	\$51	\$51

payment details:

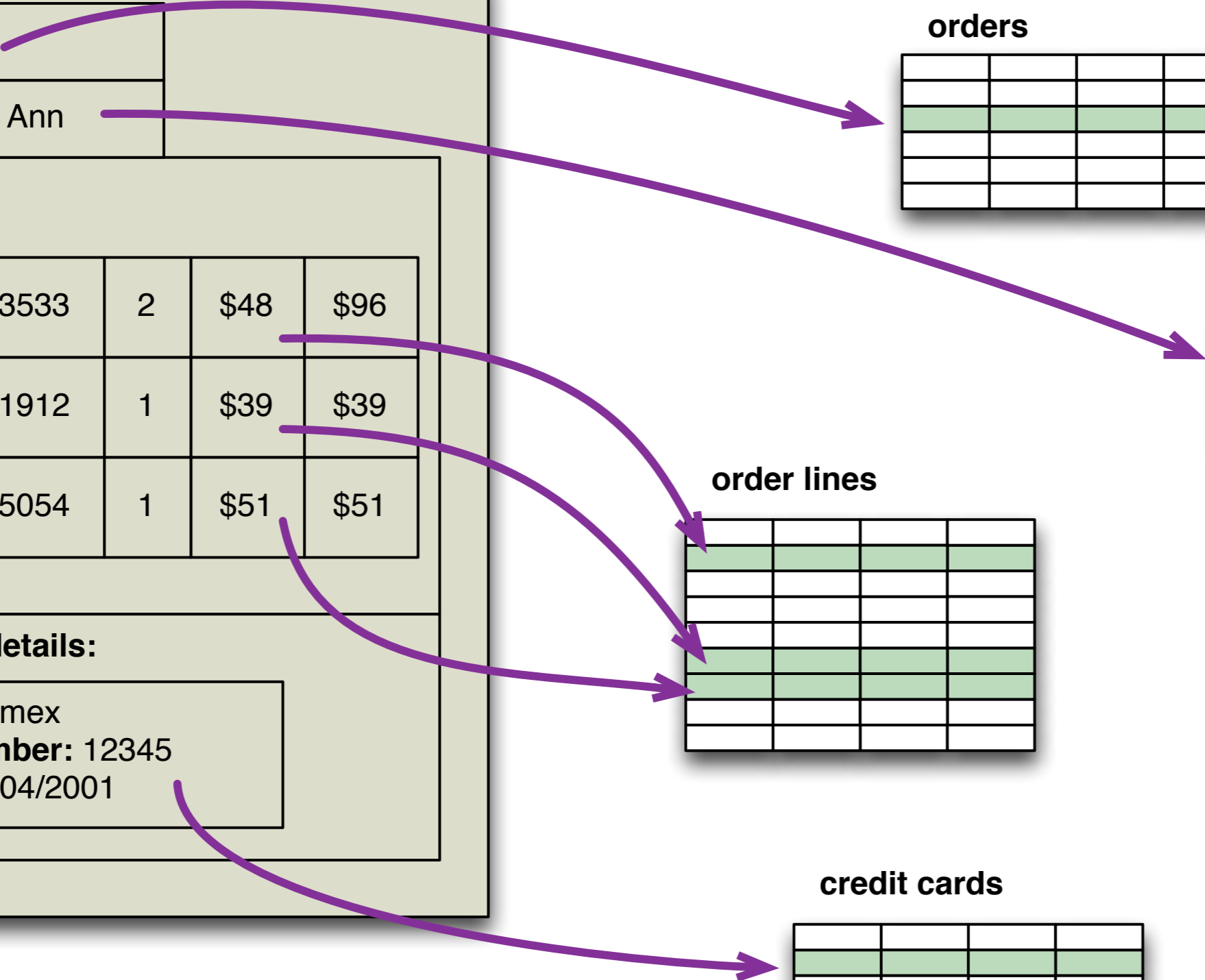
Card: Amex
CC Number: 12345
expiry: 04/2001

orders

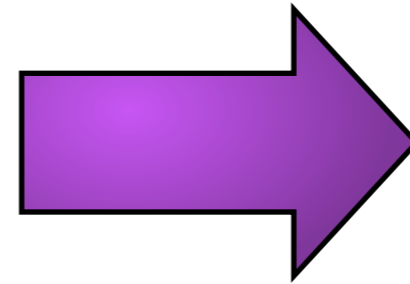
customers

order lines

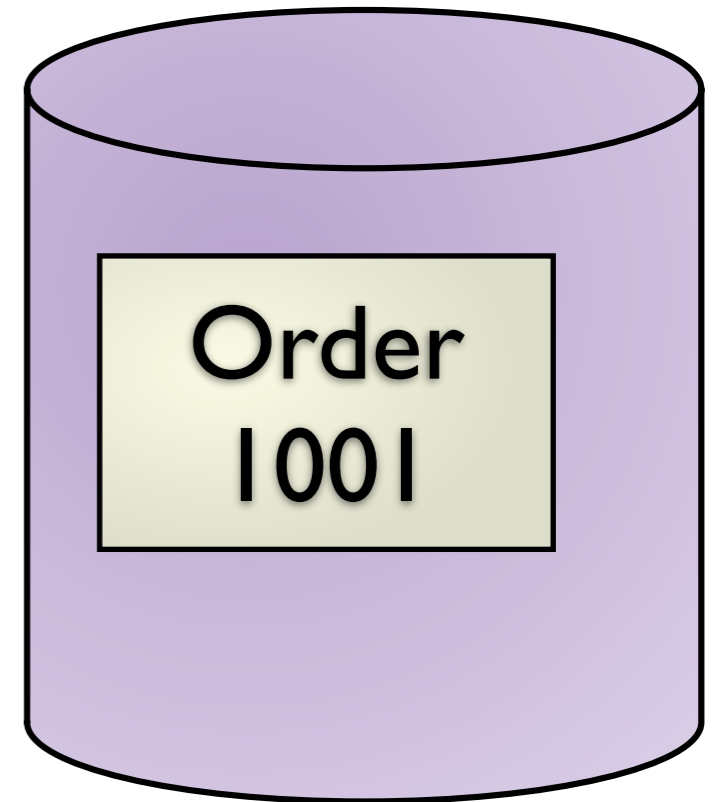
credit cards



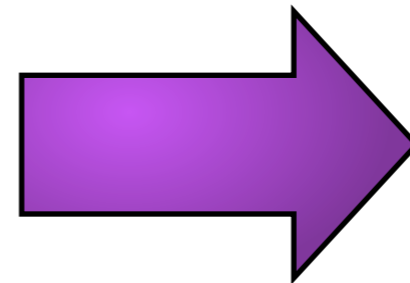
ID: 1001			
customer: Ann			
line items:			
0321293533	2	\$48	\$96
0321601912	1	\$39	\$39
0131495054	1	\$51	\$51
payment details:			
Card: Amex CC Number: 12345 expiry: 04/2001			



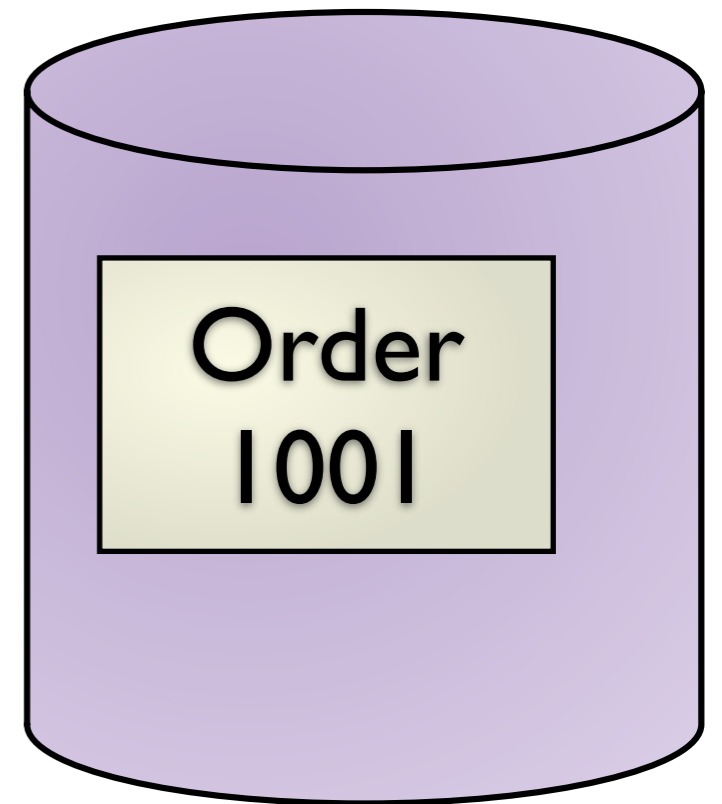
aggregate



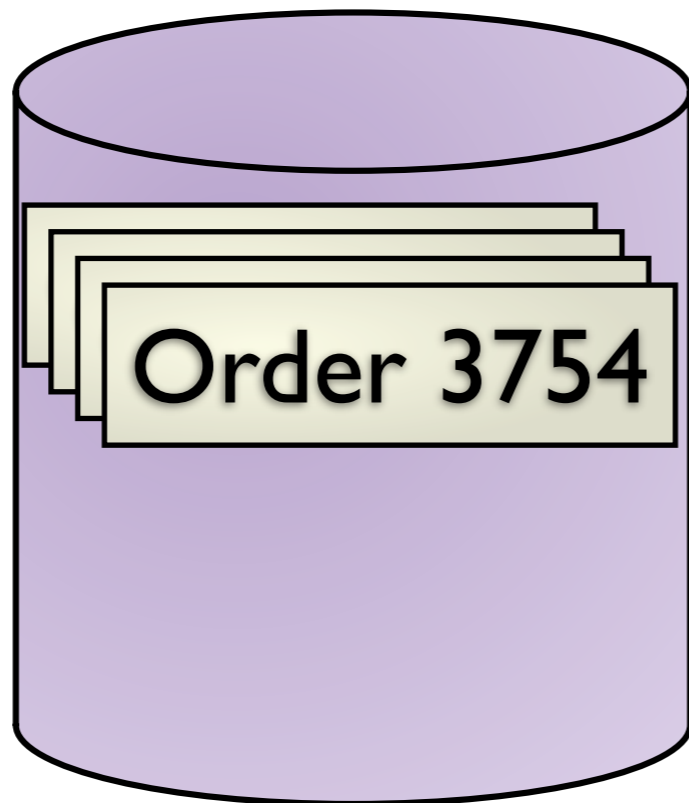
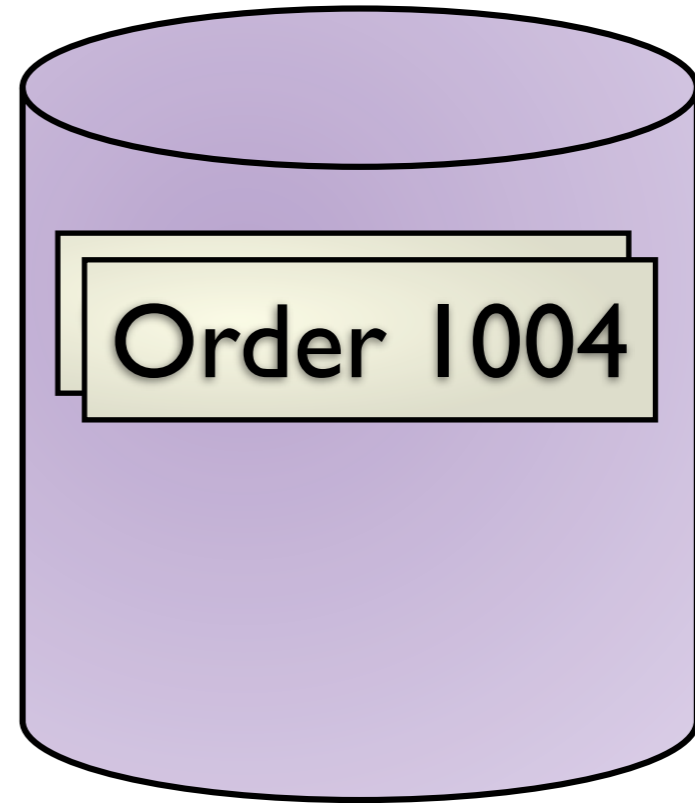
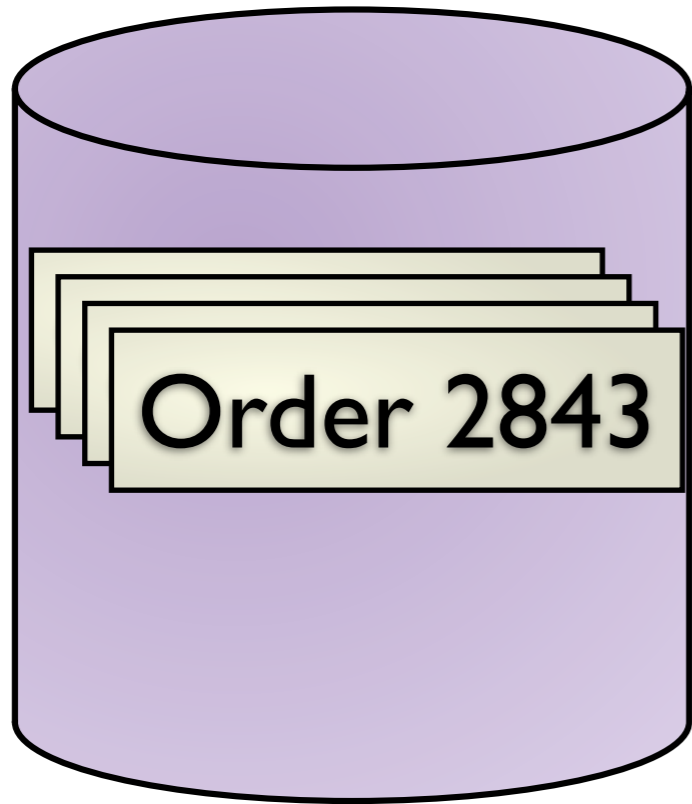
ID: 1001			
customer: Ann			
line items:			
0321293533	2	\$48	\$96
0321601912	1	\$39	\$39
0131495054	1	\$51	\$51
payment details:			
Card: Amex CC Number: 12345 expiry: 04/2001			



aggregate



*document
key-value
column family*



Document



Column-family



Graph



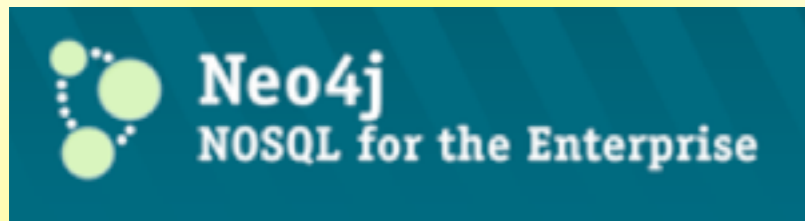
Key-value



Document



Column-family



Graph



Key-value











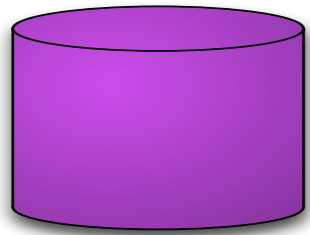
Polyglot



Persistence



Billing

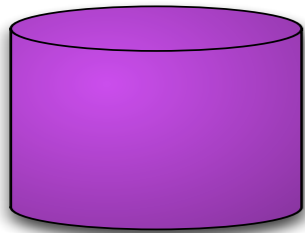


Inventory

Billing



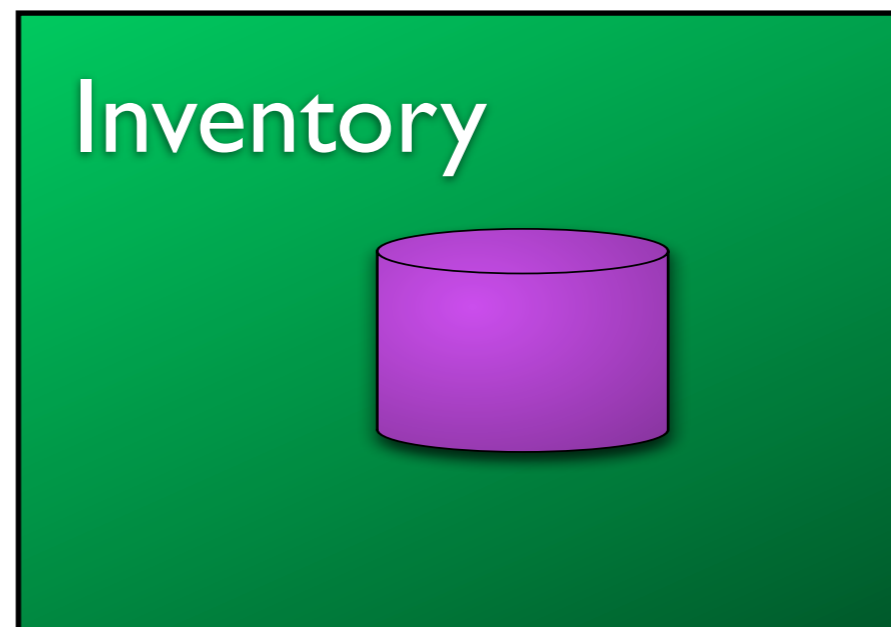
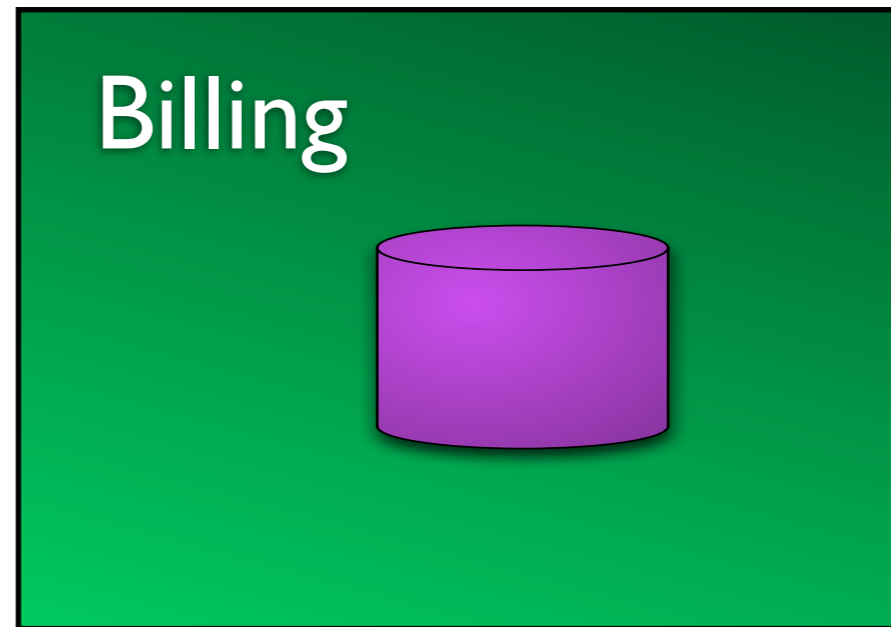
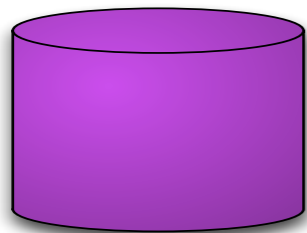
Integration Database



Inventory

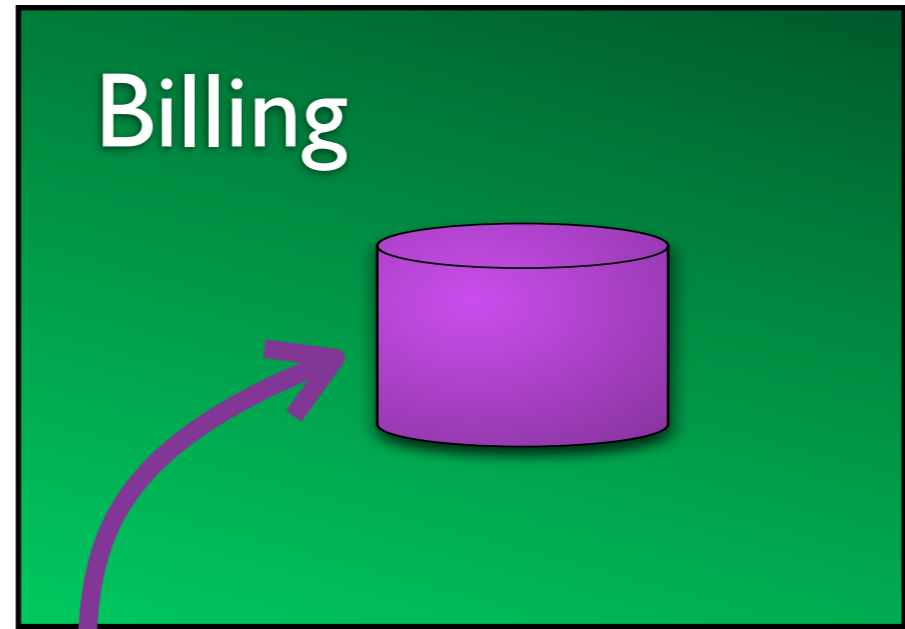
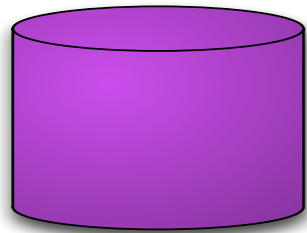


*Integration
Database*

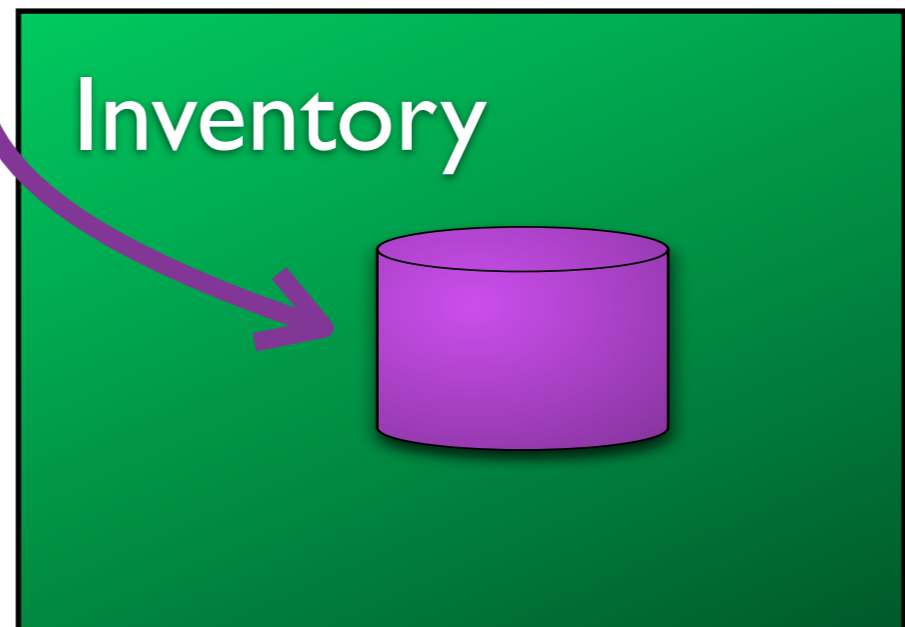




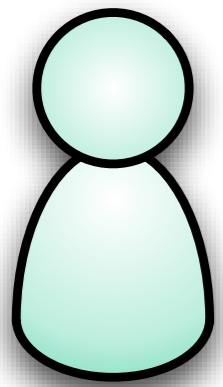
Integration
Database



Application
Database



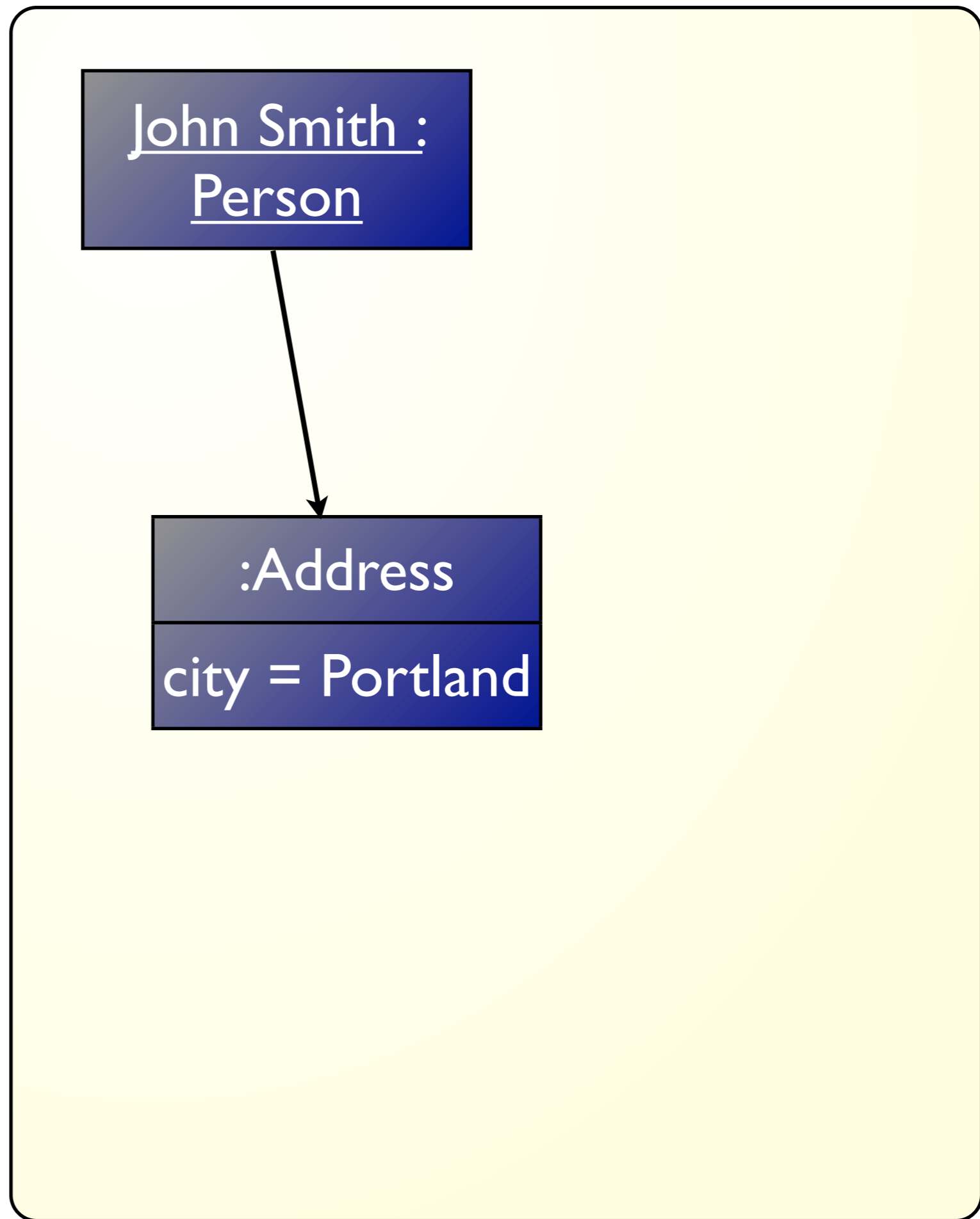
Event Sourcing

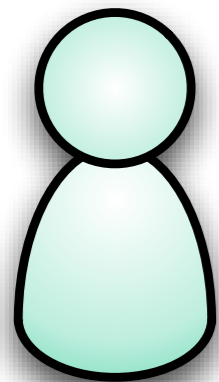


John Smith :
Person



:Address
city = Portland

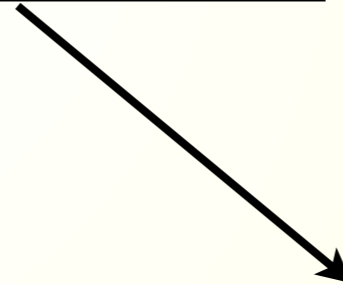


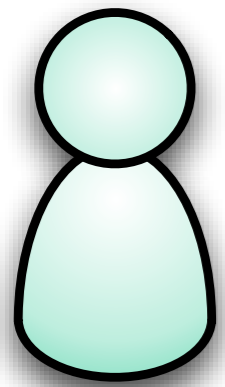


change my
address

John Smith :
Person

:Address
city = Boston



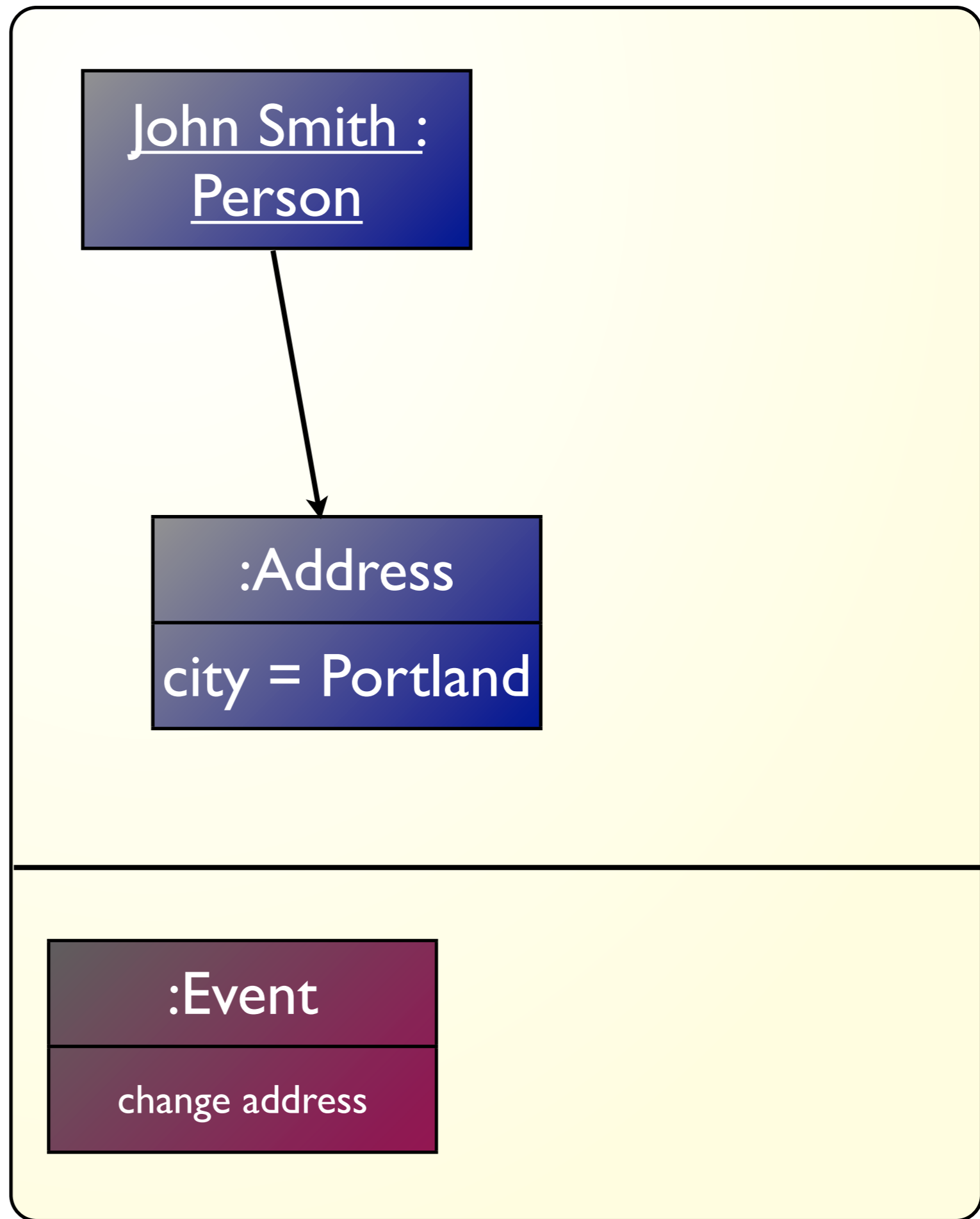


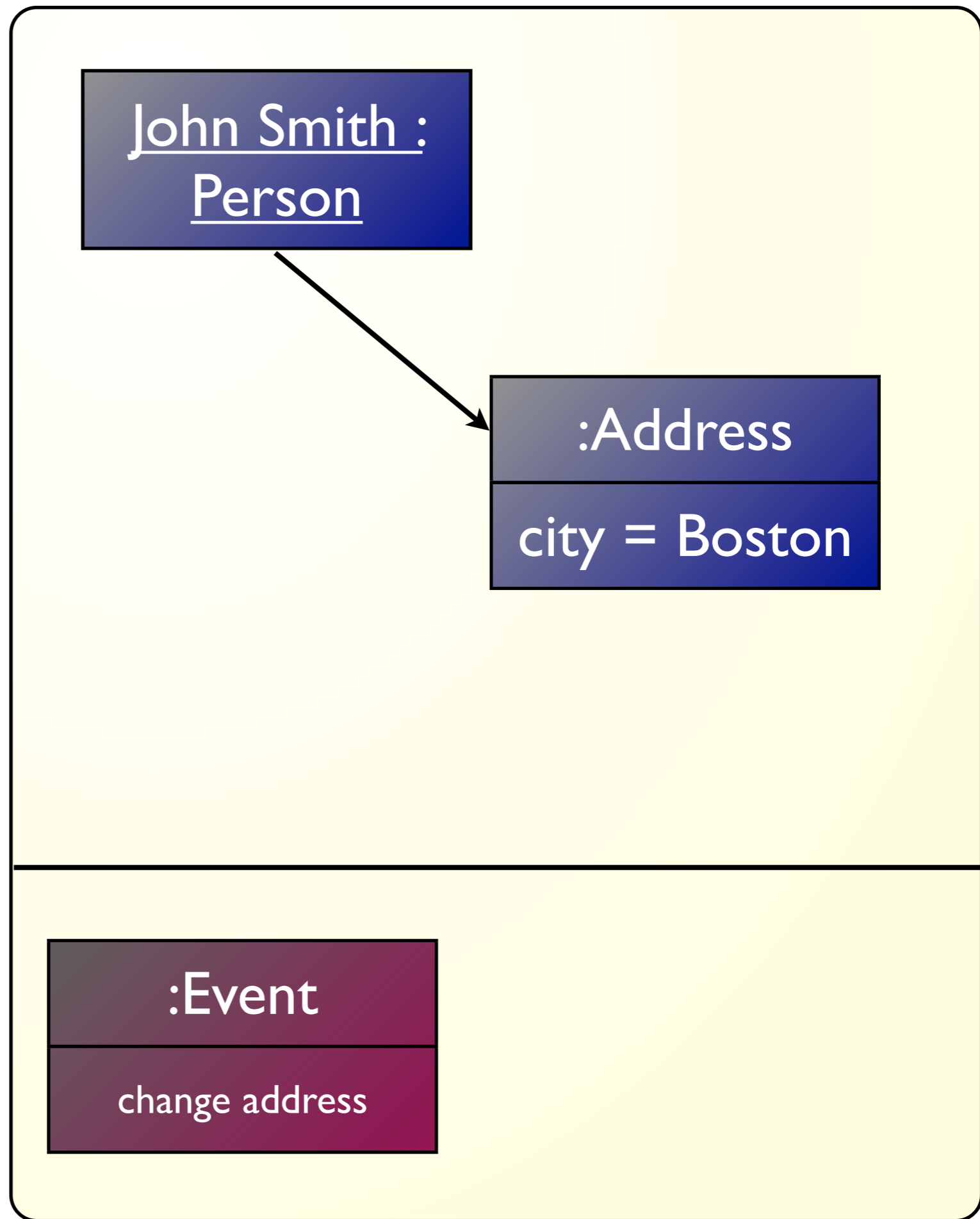
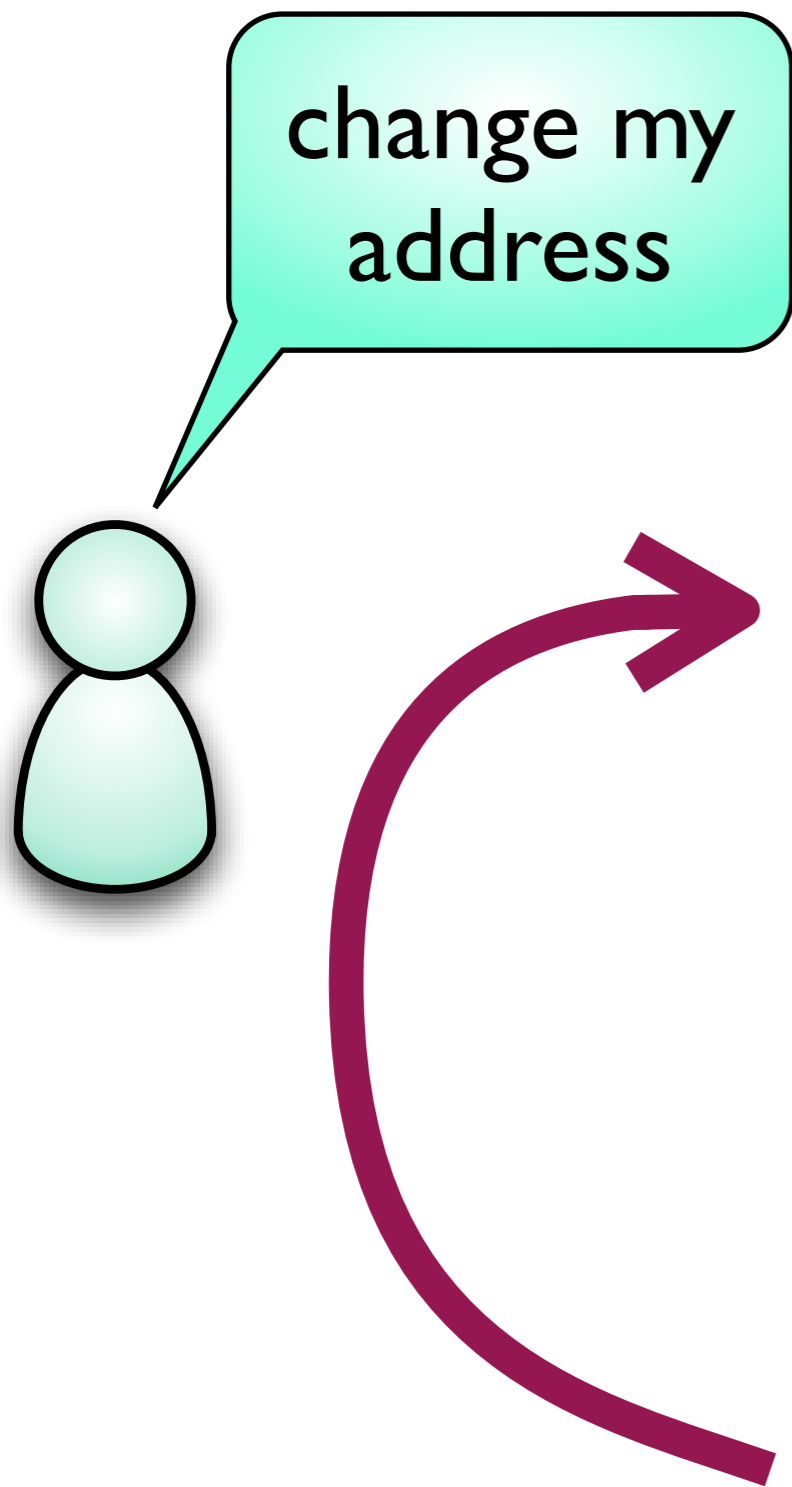
change my
address

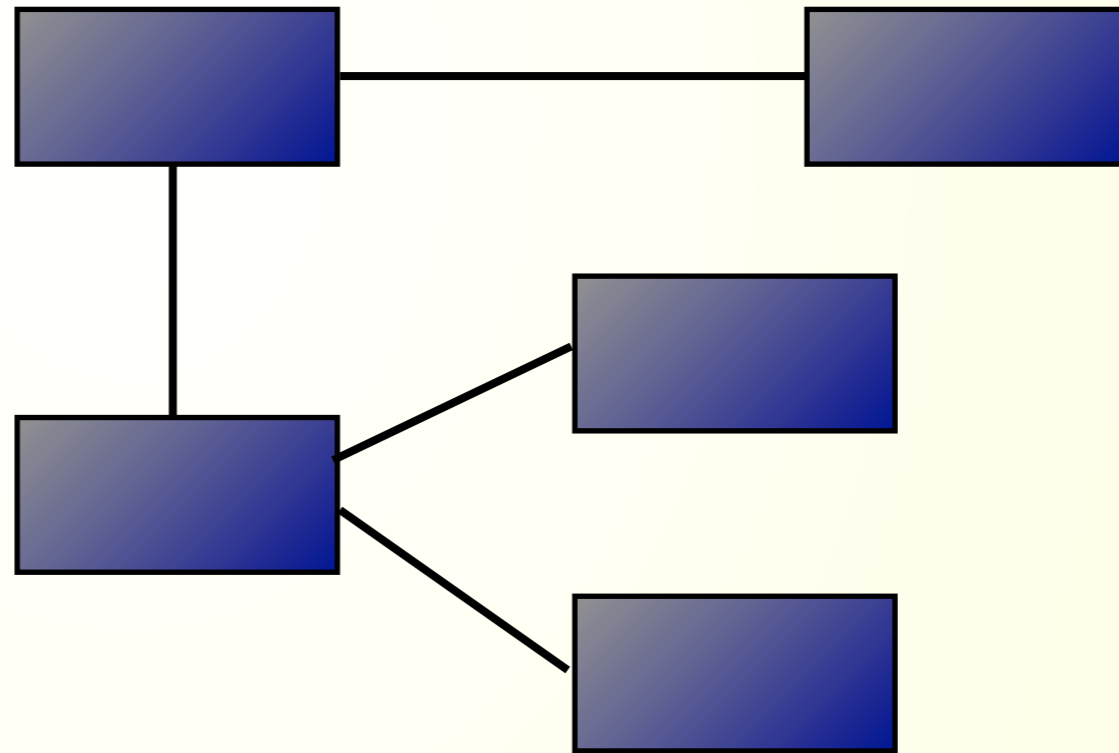
John Smith :
Person

:Address

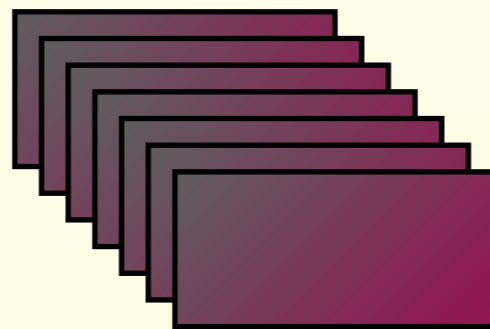
city = Portland







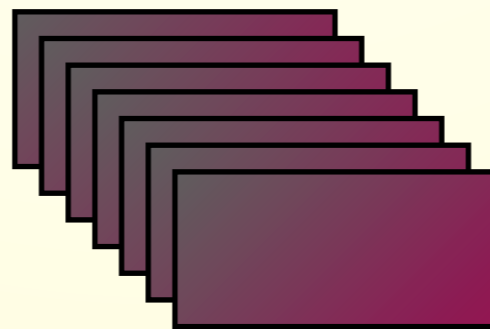
Application State

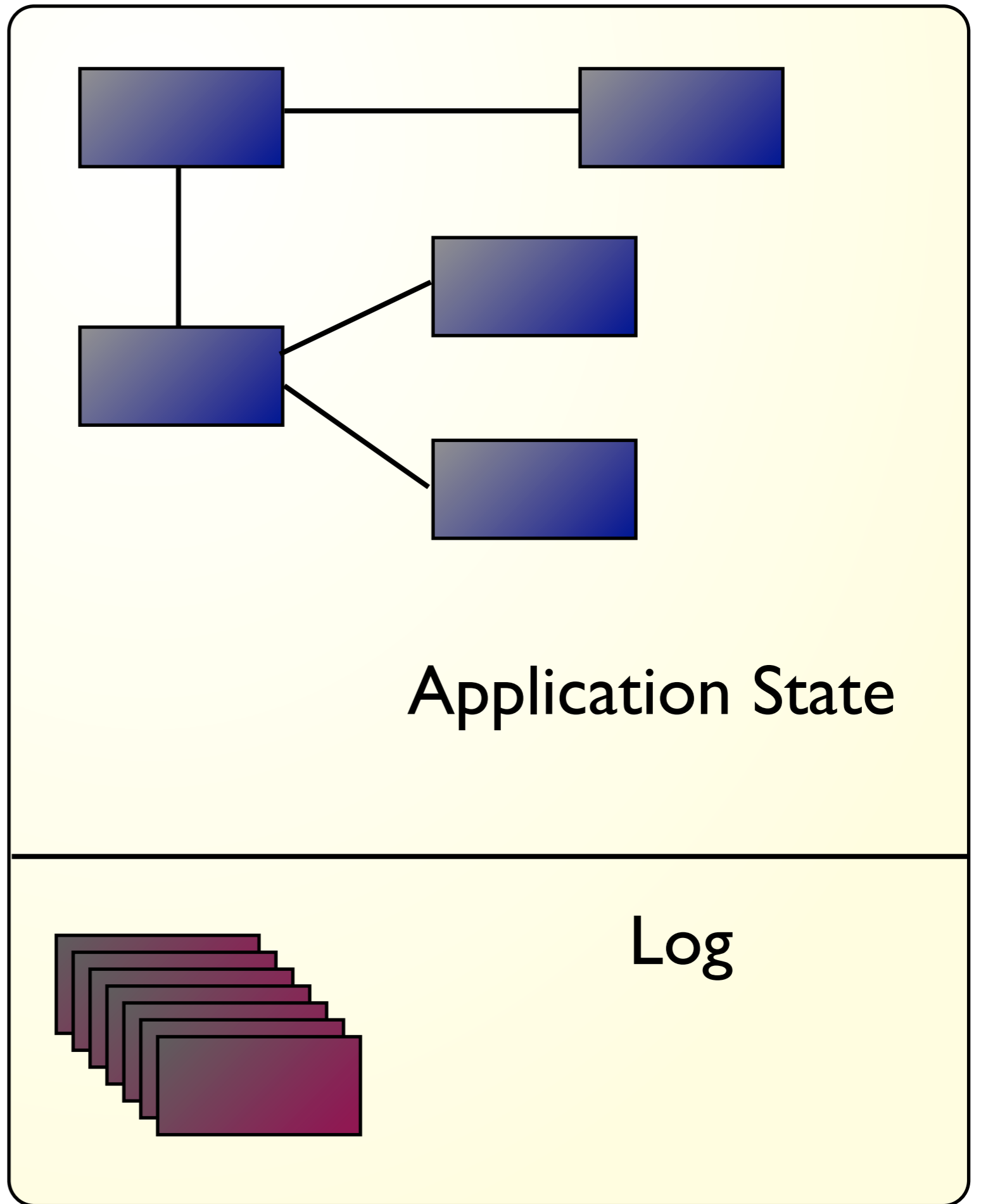


Log

Application State

Log





on-demand
self-service

rapid elasticity

broad network
access

measured
service

resource pooling

NIST

special publication 800-145

Data Sources

Data Sources

were

Data Sources

were

will be

text

image

video

connections

Analytics

Analytics

were

roll-ups

trends

variance

Analytics

were

will be

roll-ups

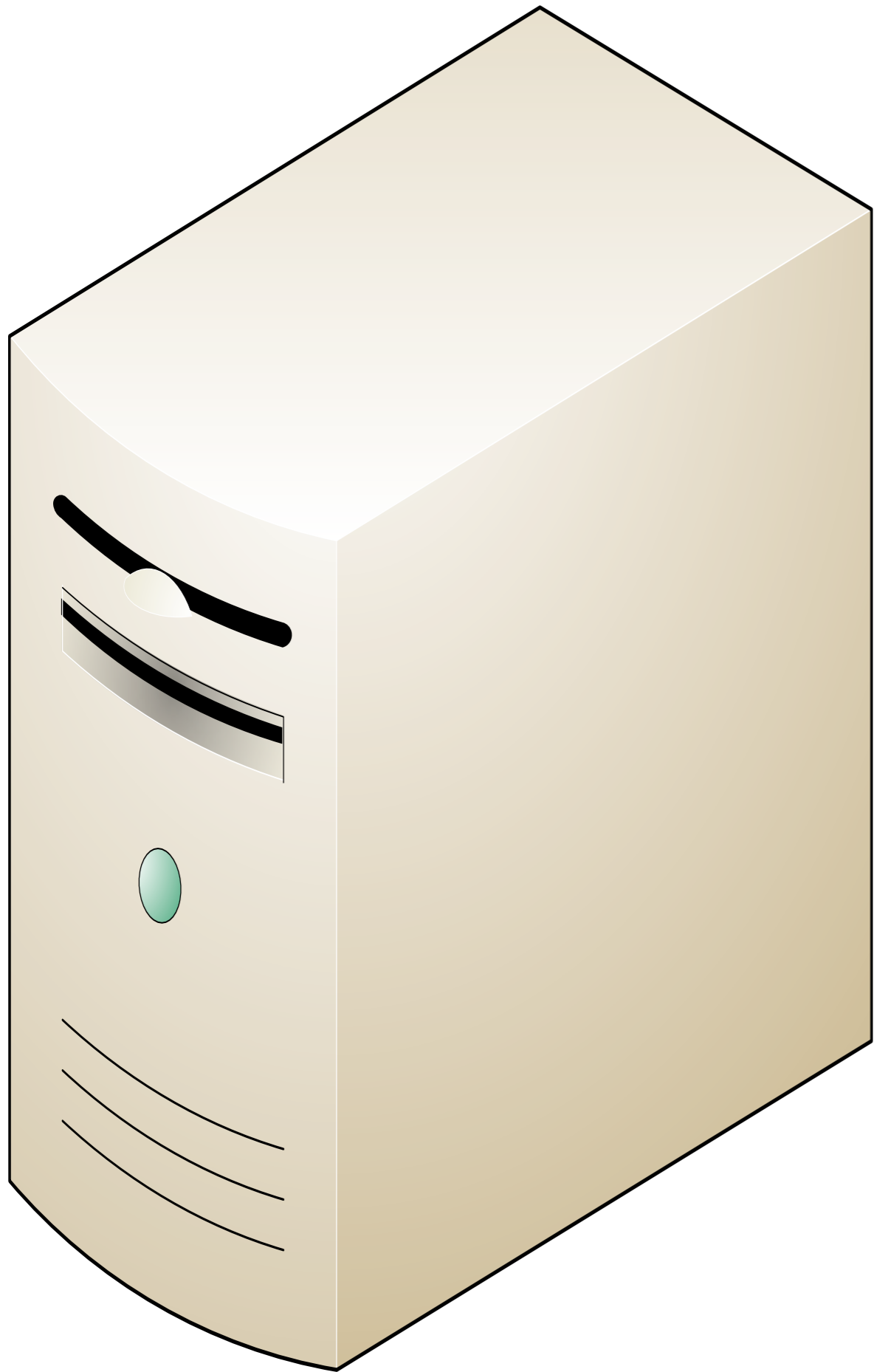
pattern recognition

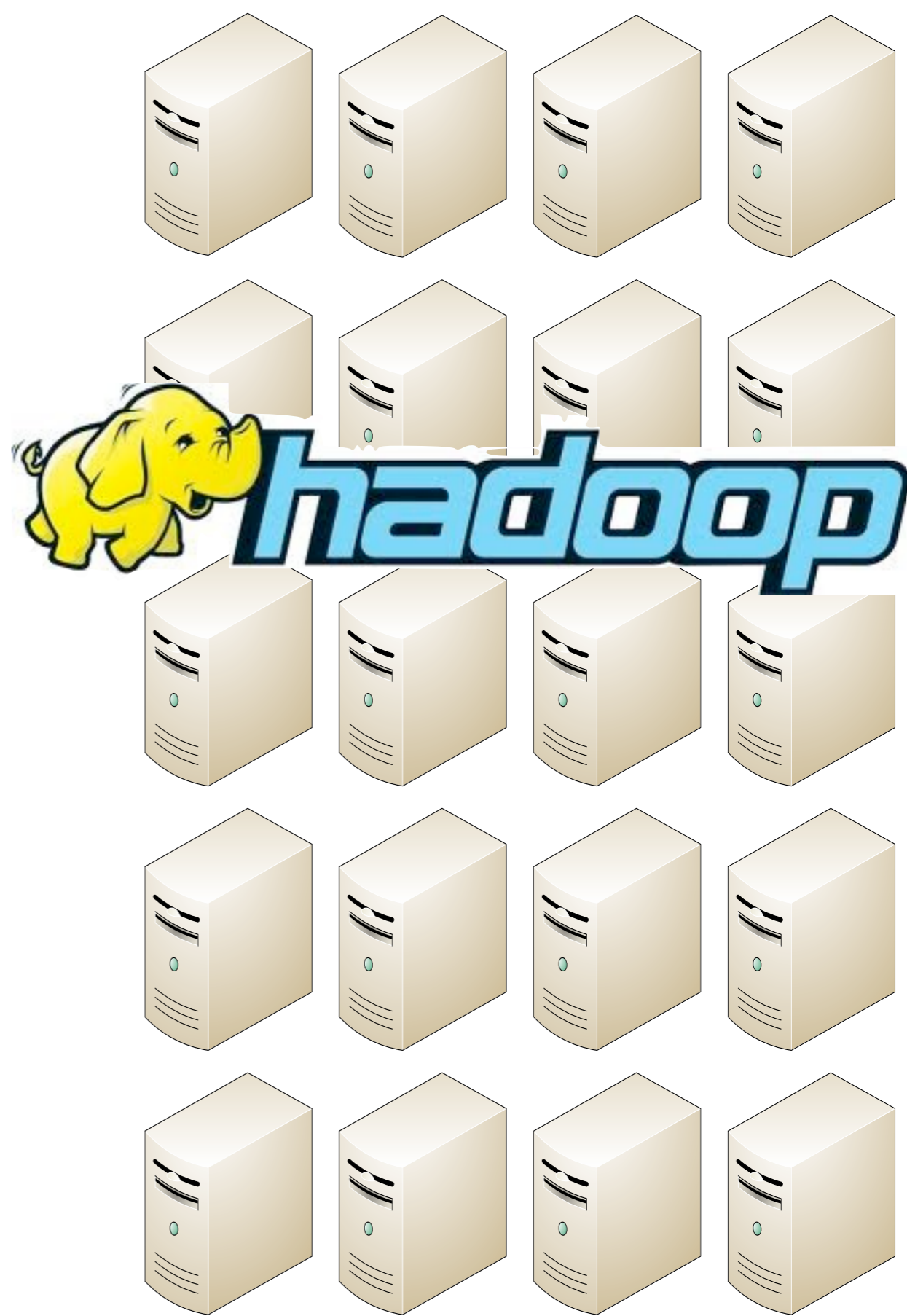
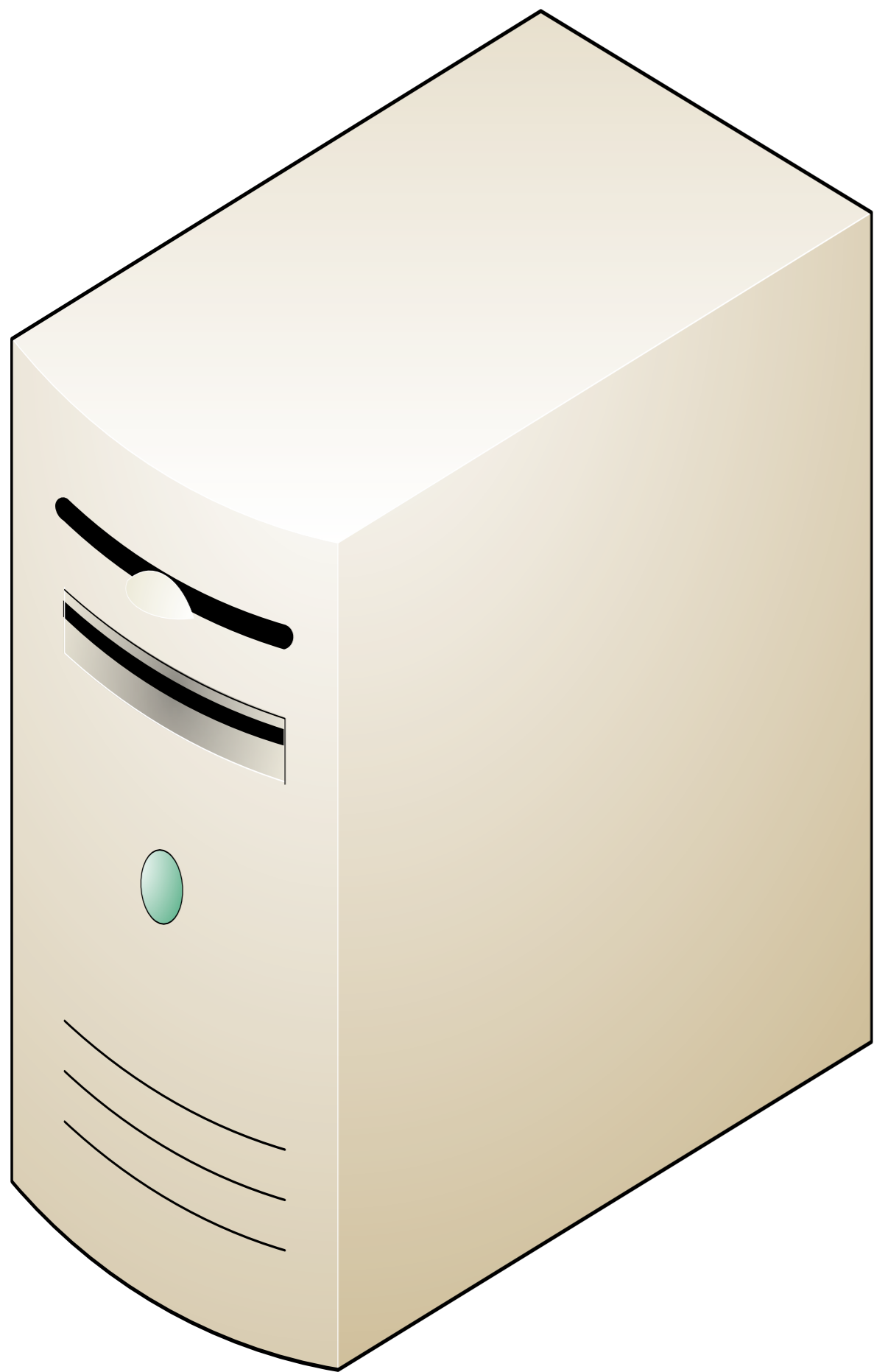
trends

data mining

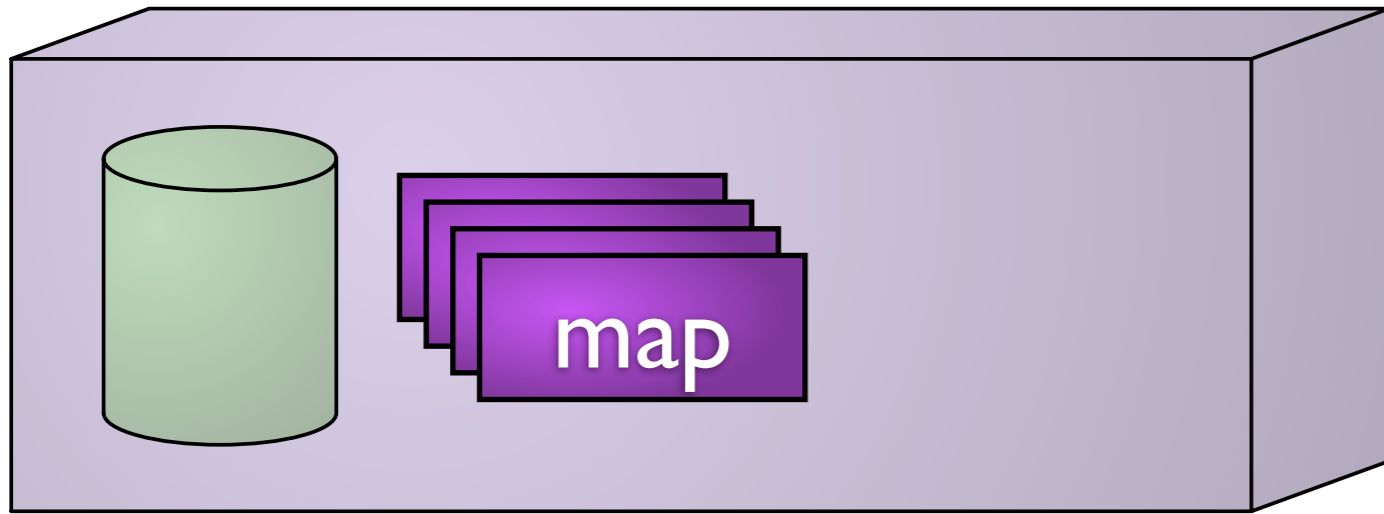
variance

chasing connections

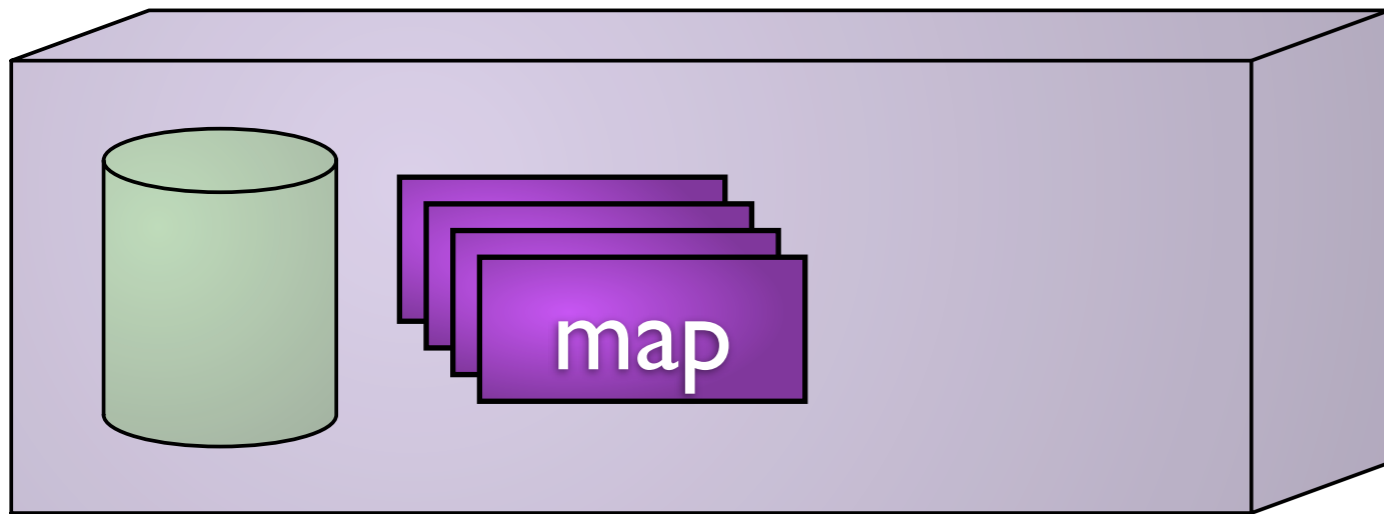
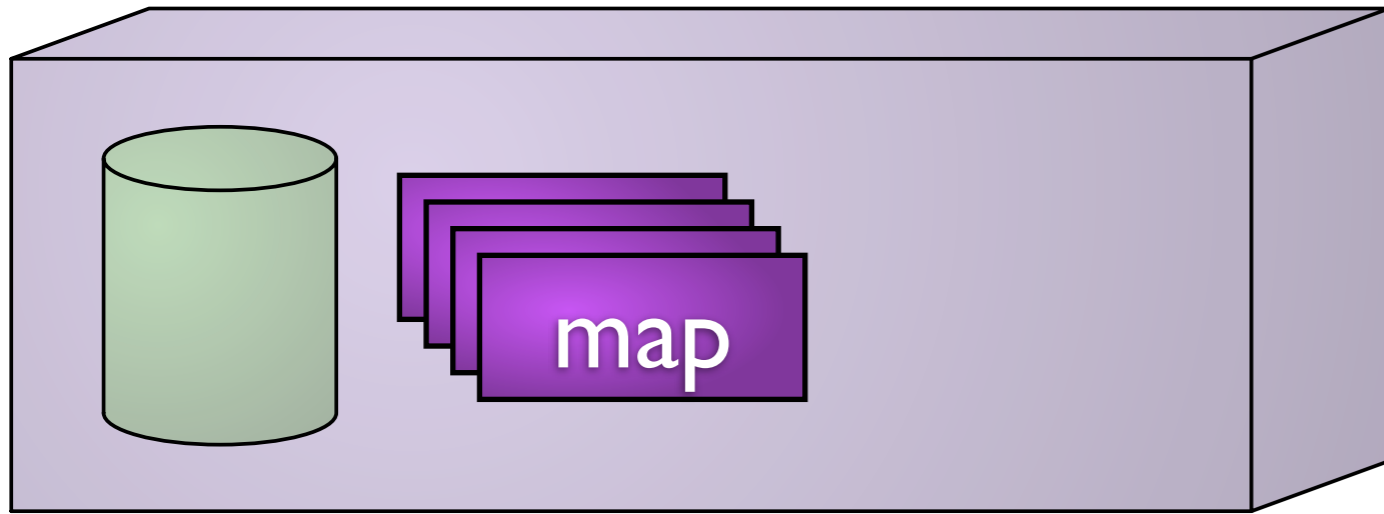


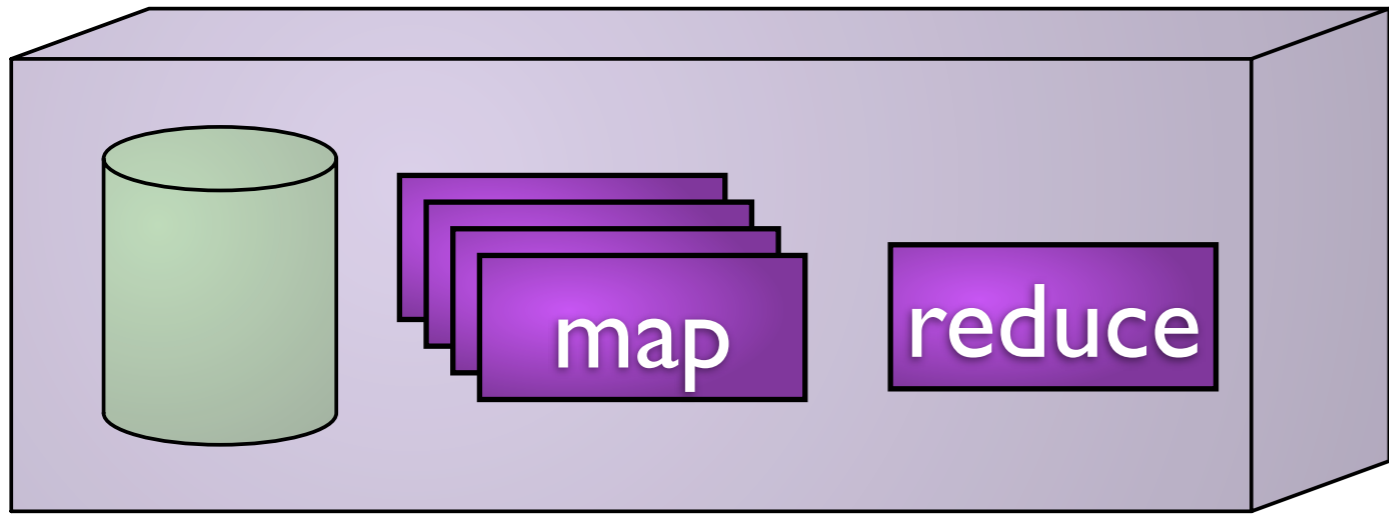




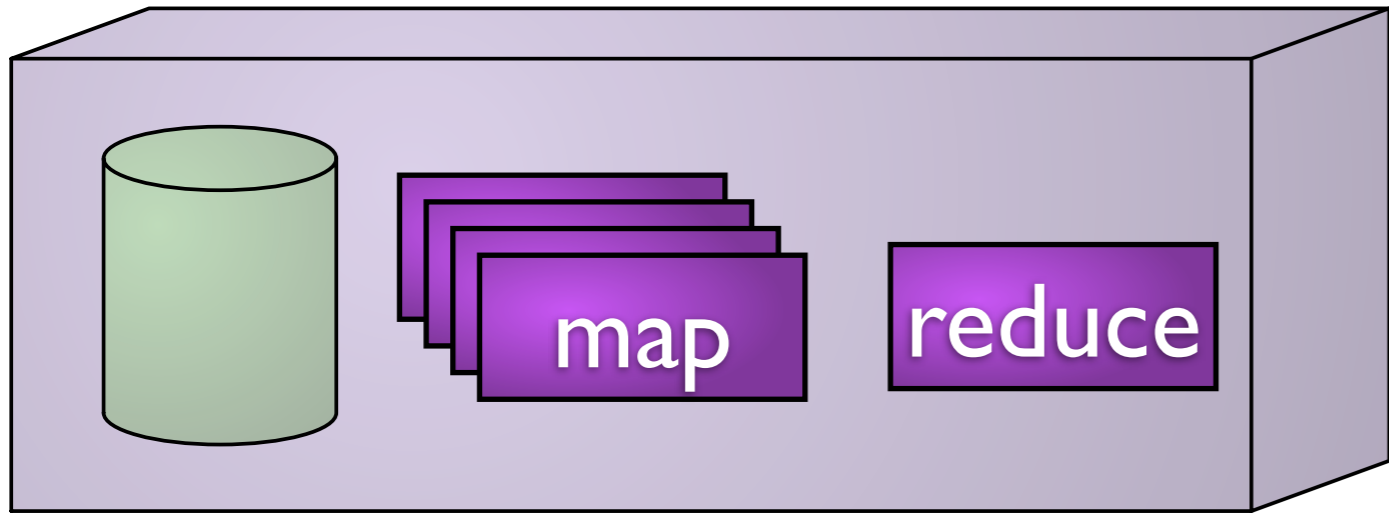
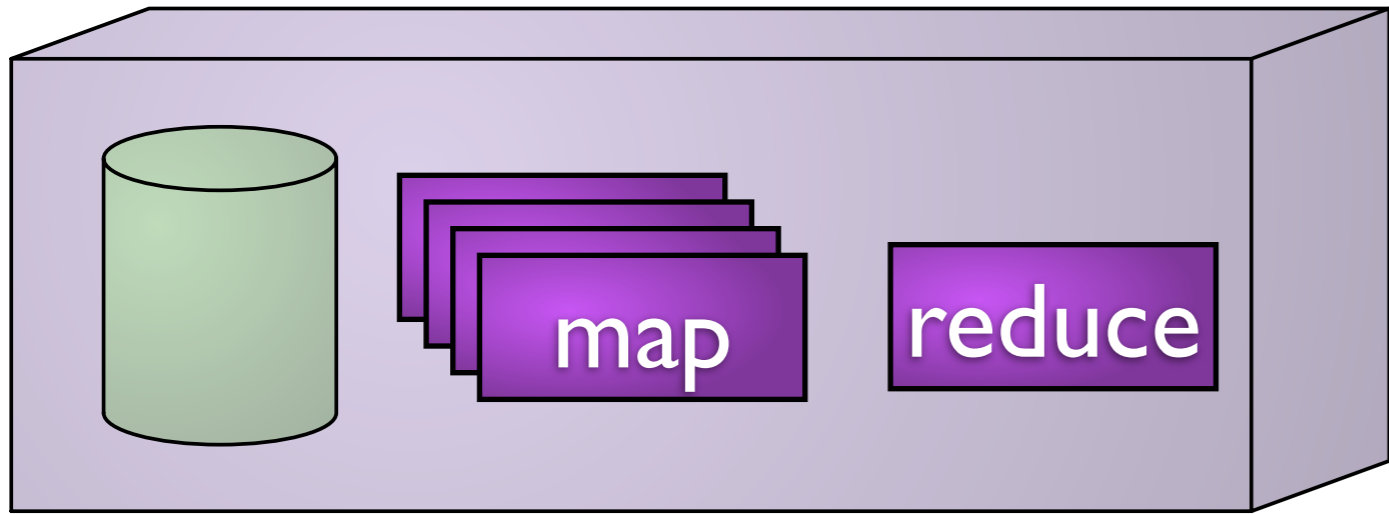


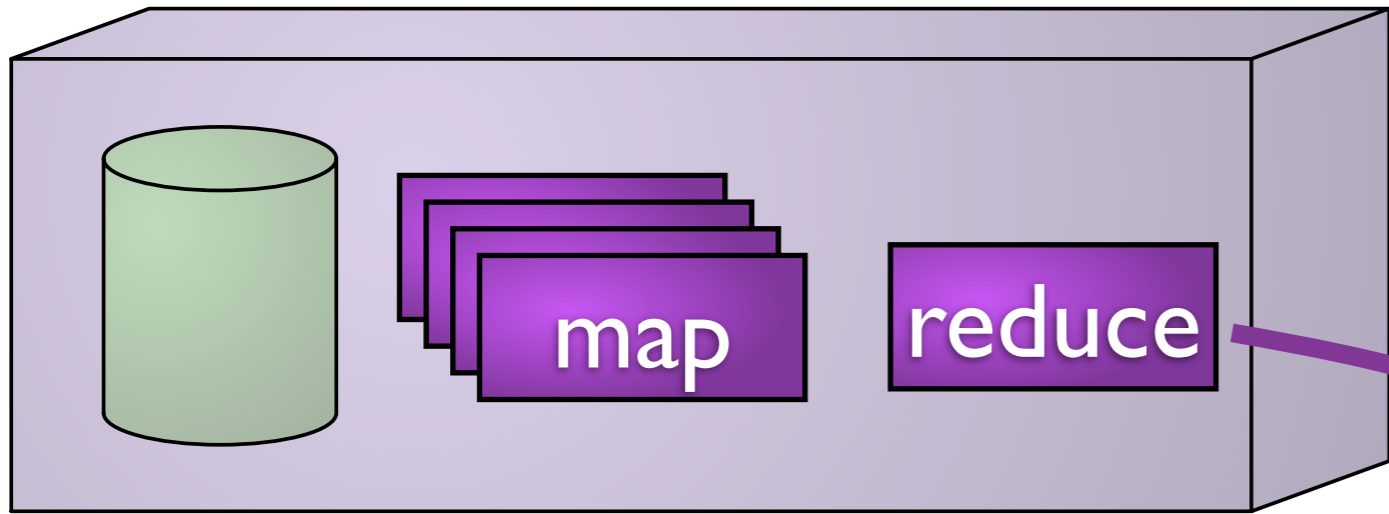
per order



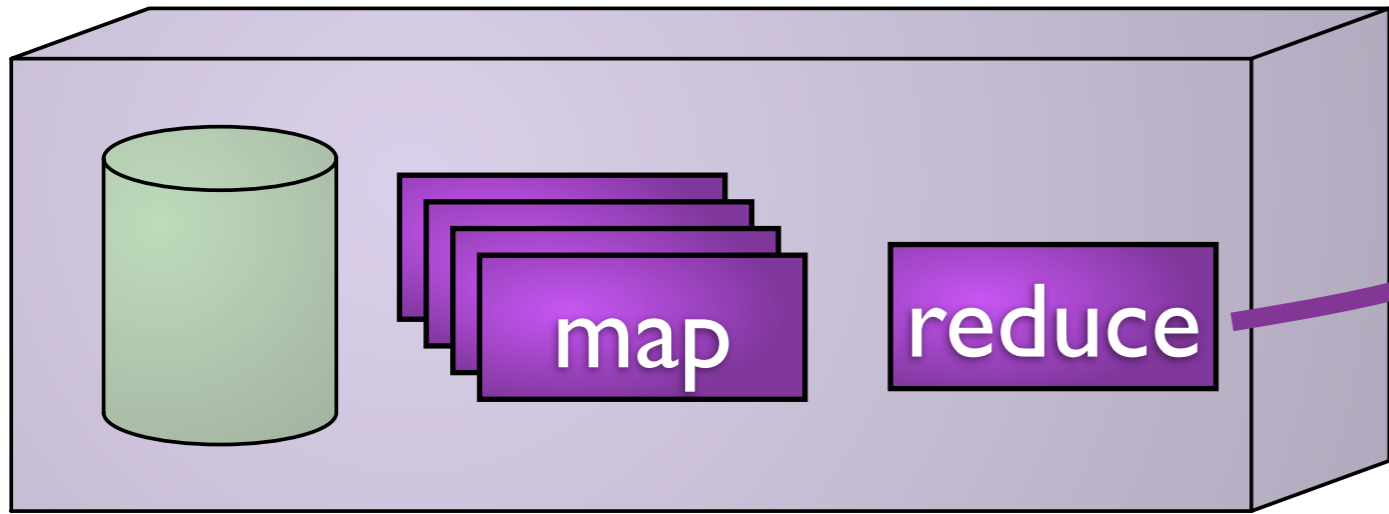
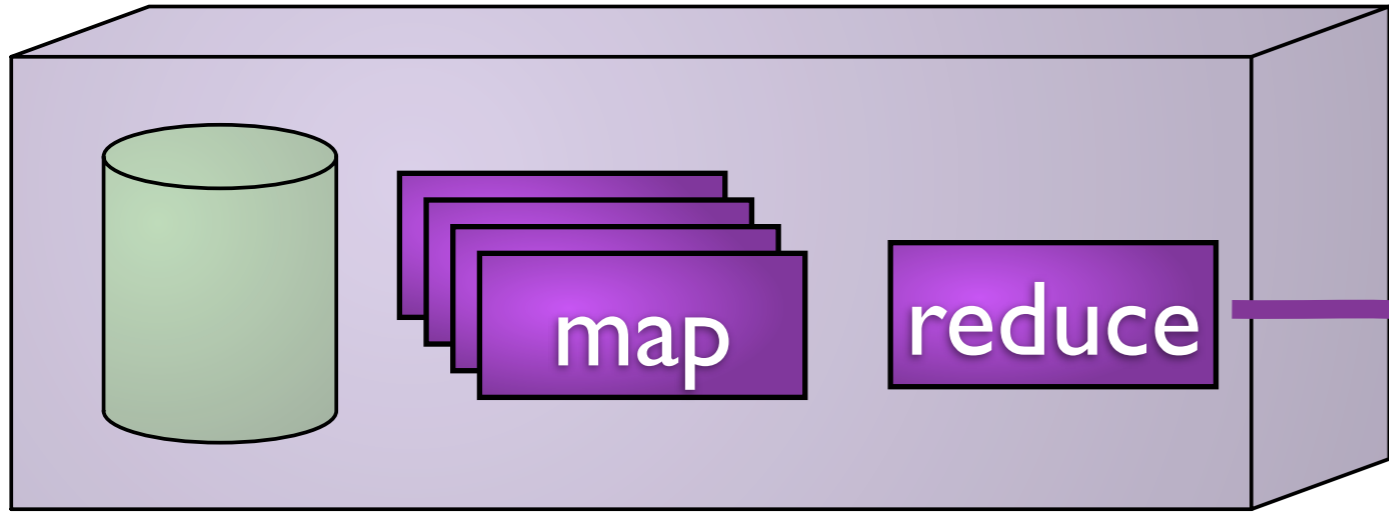


per order *per month*





per order *per month*



A PERIODIC TABLE OF VISUALIZATION METHODS

C continuum																G graphic facilitation						
Tb table	Ca cartesian coordinates																Me meeting trace	Mm metro map	Tm temple	St story template	Tr tree	Ct cartoon
Pi pie chart	L line chart																Co communication diagram	Fp flight plan	Cs concept skeleton	Br bridge	Fu funnel	Ri rich picture
B bar chart	Ac area chart	R radar chart cobweb	Pa parallel coordinates	Hy hyperbolic tree	Cy cycle diagram	T timeline	Ve venn diagram	Mi mindmap	Sq square of oppositions	Cc concentric circles	Ar argument slide	Sw swim lane diagram	Gc gant chart	Pm perspectives diagram	D dilemma diagram	Pr parameter ruler	Kn knowledge map					
Hi histogram	Sc scatterplot	Sa sankey diagram	In information lense	E entity relationship diagram	Pt petri net	Fl flow chart	Cl clustering	Le layer chart	Py minto pyramid technique	Ce cause-effect chains	Tl toulmin map	Dt decision tree	Cp cpm critical path method	Cf concept fan	Co concept map	Ic iceberg	Lm learning map					
Tk tukey box plot	Sp spectrogram	Da data map	Tp treemap	Cn cone tree	Sy system dyn./ simulation	Df data flow diagram	Se semantic network	So soft system modeling	Sn synergy map	Fo force field diagram	Ib ibis argumentation map	Pr process event chains	Pe pert chart	Ev evocative knowledge map	V Vee diagram	Hh heaven 'n' hell chart	I informal					

Data Visualization
Visual representations of quantitative data in schematic form (either with or without axes)

Strategy Visualization
The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.

Information Visualization
The use of interactive visual representations of data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space. The image can be changed by users as they proceed working with it

Metaphor Visualization
Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed

Concept Visualization
Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses.

Compound Visualization
The complementary use of different graphic representation formats in one single schema or frame

- Cy** **Process Visualization**
- Hy** **Structure Visualization**
- Overview**
- Detail**
- Detail AND Overview**
- Divergent thinking**
- Convergent thinking**

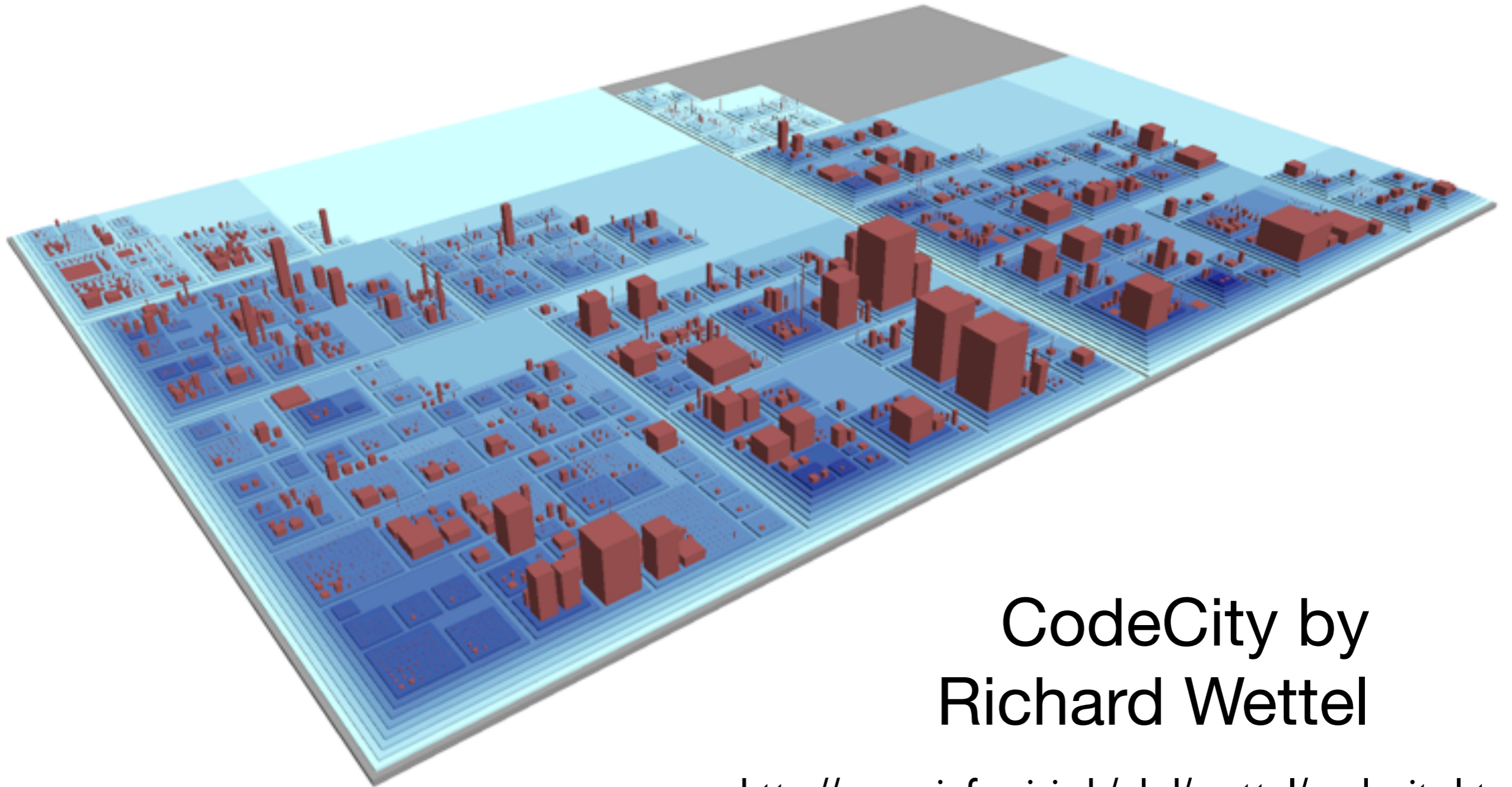
Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

version 1.5

© Ralph Lengler & Martin J. Eppler, www.visual-literacy.org

Su supply demand curve	Pc performance charting	St strategy map	Oc organisation chart	Ho house of quality	Fd feedback diagram	Ft failure tree	Mq magic quadrant	Ld life-cycle diagram	Po porter's five forces	S s-cycle	Sm stakeholder map	Is ishikawa diagram	Tc technology roadmap
Ed edgeworth box	Pf portfolio diagram	Sg strategic game board	Mz mintzberg's organigraph	Z zwick's morphological box	Ad affinity diagram	De decision discovery diagram	Bm bcg matrix	Stc strategy canvas	Vc value chain	Hy hype-cycle	Sr stakeholder rating map	Ta taps	Sd spray diagram

10,000 ft view (literally)



CodeCity by
Richard Wetzel

<http://www.inf.unisi.ch/phd/wetzel/codecity.html>

All Spending

Types of Spending

Changes

Department Totals

Mandatory \$2.5 trillion

About 70 percent of budgetary spending is controlled by existing laws, including entitlements like Medicare, Medicaid and Social Security.



EDUCATION

Office of Innovation and Improvement
Discretionary spending

\$4.3 billion **+183.5%**

Discretionary \$1.1 trillion

Only about 30 percent of the budget is controlled by the annual budget process. Last August, the White House and Congress agreed to a cap on this spending.

The New York Times

Changing Nature of Data

Response

How we use data now

Changing Nature of Data

Response

How we use data now

Data Scientist

Data

Data Journalist

Google

Google Search

I'm Feeling Lucky

Google

Google Search

I'm Feeling Lucky



+Martin Search Images Videos Maps News Shopping Gmail More - martinfowler.com@gmail.com - ⚙

iGoogle

[Change theme](#)

Add gadgets

Home

- Google Calendar
- Weather
- Gmail (1)
- DaylightMap
- Google Reader (286)

[Enable chat](#)

Google Reader (286)

All items (286) refresh mark all as read

- 5 Things You Need To Know Today—Tuesday, from [Melrose Patch](#)
- Time to doff the maroon cap to a king of the from [Sport, Cricket | guardian.co.uk](#)
- Andy Marks: Where university stops... and from [www.thoughtworks.com aggregator](#)
- Temporary Post Used For Theme Detection from [Tim Anderson's ITWriting](#)
- Update On Kasich And Federal Tornado Aid, from [Stan Collender's Capital Gains and](#)

Weather

Melrose, MA

20°F Current: Partly Cloudy
Wind: NW at 20 mph
Humidity: 46%

Tue	Wed	Thu	Fri
39° 28°	55° 39°	61° 43°	52° 30°

Gmail (1)

Inbox (1)

- Erik .. Ajeay, Sam, Nick (10) - Top Tech, Github and Coderwall - Hi
- John J Ritsko - Review of "Assessing technical debt by
- Floyd Marinescu - small favour for a worthy cause: vote on our first
- Neal Ford - Presentation Patterns book Friends & Family review -
- Cindy Chabot - receipts - no receipt: Amex 2/17/12 Qlosk.com.NY \$5.00

Google Calendar

March 2012

M	T	W	T	F	S	S
27	28	29	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25

DaylightMap

NETFLIX

Watch Instantly

Just for Kids

Browse DVDs

Your Queue

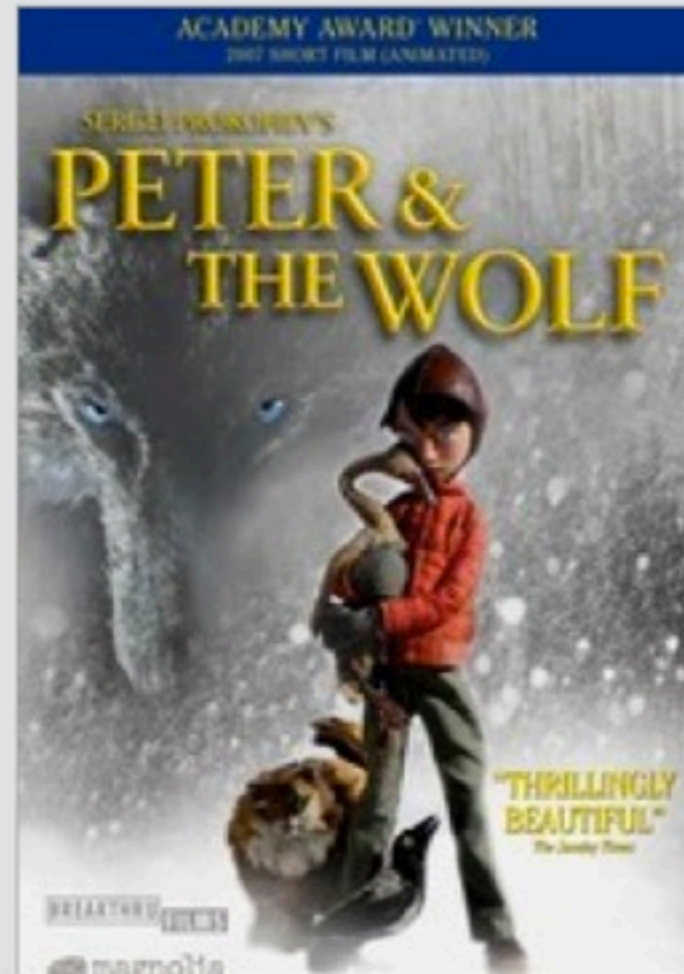
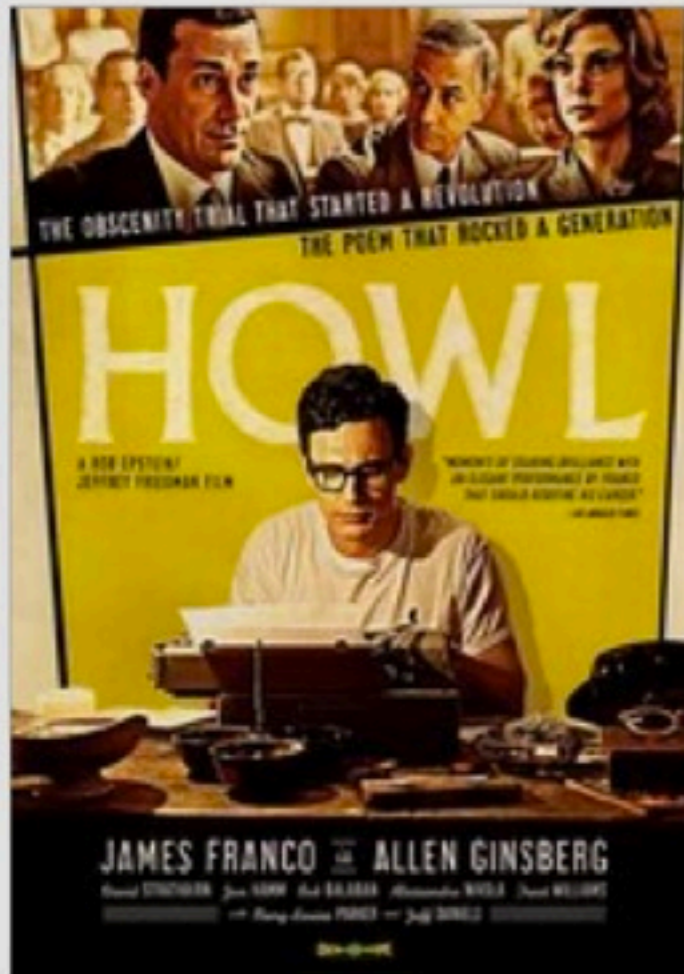
★ Suggestions for You

Genres ▼

New Arrivals

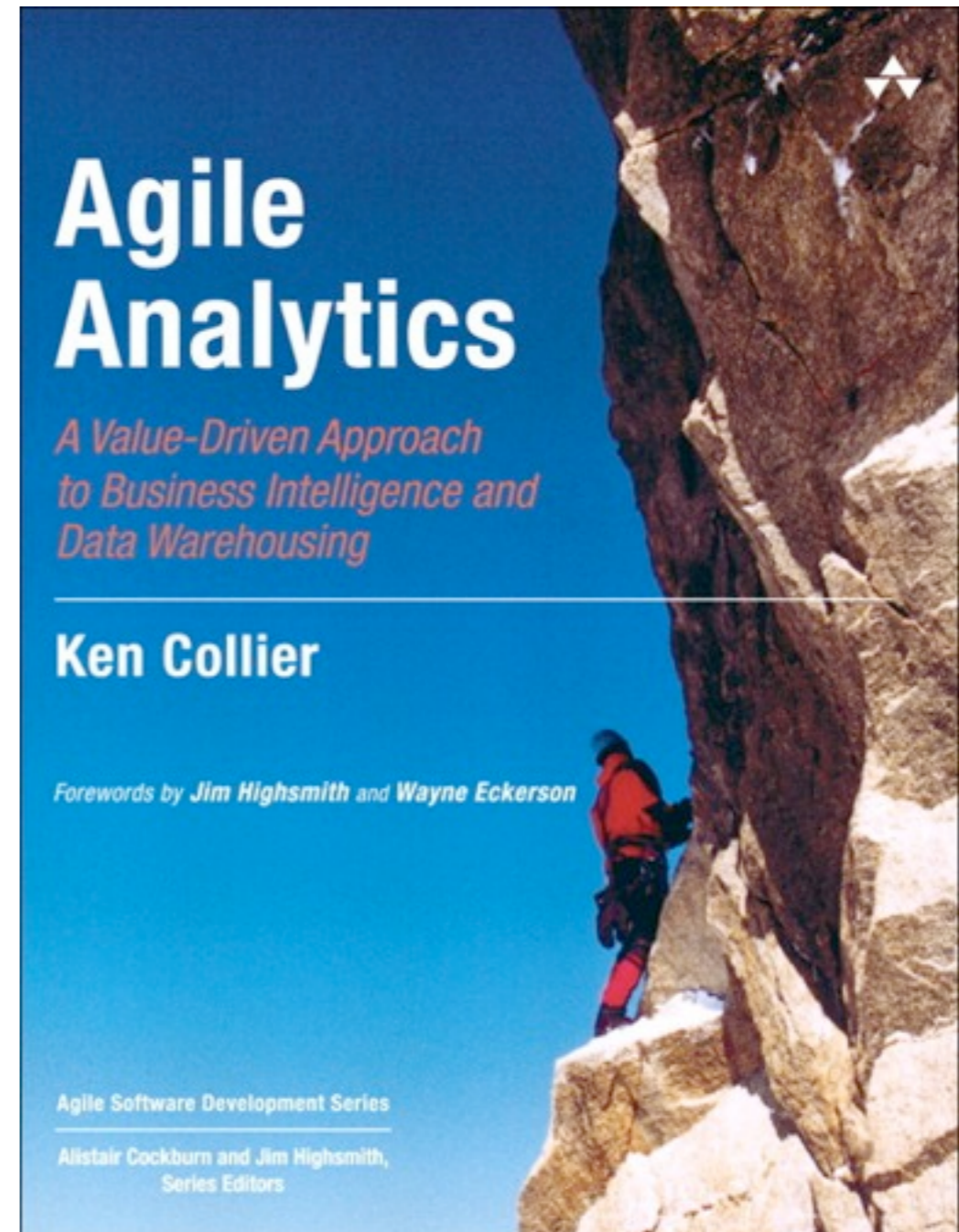
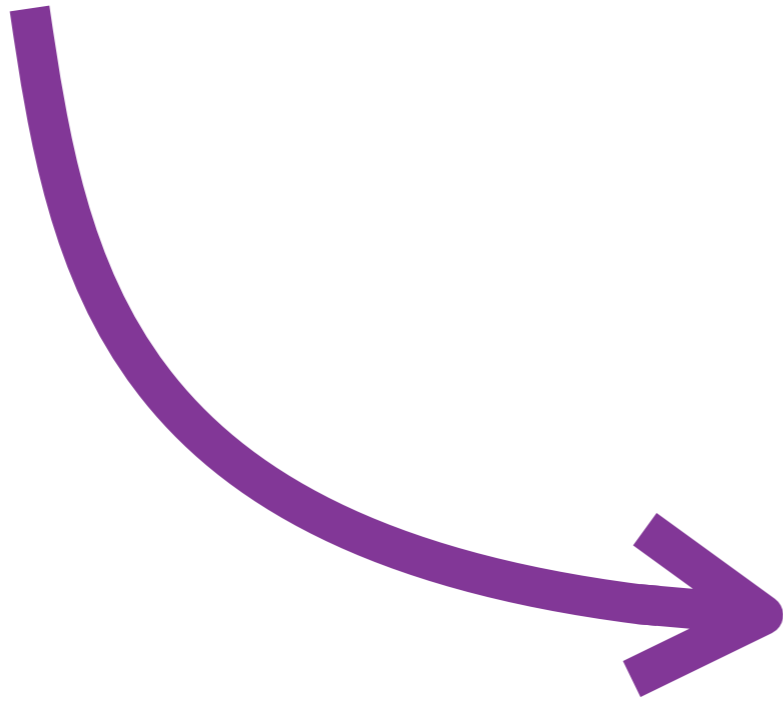
Instantly to your TV

Top 10 for Martin



Data Warehousing

Data Warehousing



Shop All Stores ▾

Search Keywords, Model # or Item #

Marketplace ▾

Narrow Results

Guided **Advanced Power** >

Useful Links

- Free Shipping (122)
- Discount Item (99)
- ShopRunner Eligible (64)
- Mail-in Rebate (21)
- Top Sellers (11)
- New Product (4)
- First From Asia (3)

Manufacturer

- OCZ Technology (30)
- Kingston Technology Corp. (18)
- Crucial (13)
- Intel (12)
- Corsair (11)
- SAMSUNG (9)

More

Capacity

- 40GB (1)
- 48GB (1)

Home > Computer Hardware > Hard Drives > SSD > Internal SSD (x) > Price : \$100 - \$200 (x) (1-100 of 137 Results)

Search Results

Search Within

Sort by : View:

Select Items

Showing 1-100 of 137 Products



(563)

Compare

Crucial M4 CT064M4SSD2 2.5" 64GB SATA III MLC Internal Solid State Drive (SSD)

- Sustained Sequential Read: Up to 500 MB/s (SATA 6Gb/s)
- Sustained Sequential Write: Up to 95 MB/s (SATA 6Gb/s)
- 4KB Random Read: Up to 45,000 IOPS
- 4KB Random Write: Up to 20,000 IOPS
- Model #: CT064M4SSD2
- Item #: N82E16820148441
- Return Policy: Limited Replacement Only Return Policy

was: \$139.99

\$104.99

Free Shipping



Crucial M4 CT128M4SSD2 2.5" 128GB SATA III MLC Internal Solid State Drive (SSD)

- Sustained Sequential Read: Up to 500 MB/s (SATA 6Gb/s)
- Sustained Sequential Write: Up to 175 MB/s (SATA 6Gb/s)

was: \$199.99

\$174.99

Free Shipping

Home > PCs & Laptops > Laptop Accessories > Internal SSD > Crucial > Item#: N82E16820148442



Crucial M4 CT128M4SSD2 2.5" 128GB SATA III MLC Internal Solid State Drive (SSD)

In stock.

was: \$199.99

\$174.99

save: \$25.00

QTY.

Protect Your Investment!

1-year extended warranty **\$16.99.** (more options)

FREE SHIPPING (restrictions apply)

FREE 2-Day Shipping [learn more](#) | [sign in](#)

Average Rating



(329 reviews)

Write a Review

Add to Wishlist

Price Alert

Print This Page

Share



Like 370

4

Overview

Details

Feedback

Learn more about the Crucial CT128M4SSD2

General

Brand	Crucial
Series	M4
Model	CT128M4SSD2
Device Type	Internal Solid State Drive (SSD)
Used For	Consumer

Quick Info

Warranty

Limited Warranty period (parts): 3 years

Limited Warranty period (labor): 3 years

Manufacturer Contact Info

Manufacturer Product Page

Website:

<http://www.crucial.com/index.aspx>

Support Phone: 1-800-336-8915

- ») HOME
- ») POLL RESULTS
- ») ABOUT UREPORT
- ») HOW TO JOIN

Ureport is a free SMS-based system that allows young Ugandans to speak out on what's happening in communities across the country, and work together with other community leaders for positive change.

Join Ureport 
Text "JOIN" to 8500
IT'S FREE!

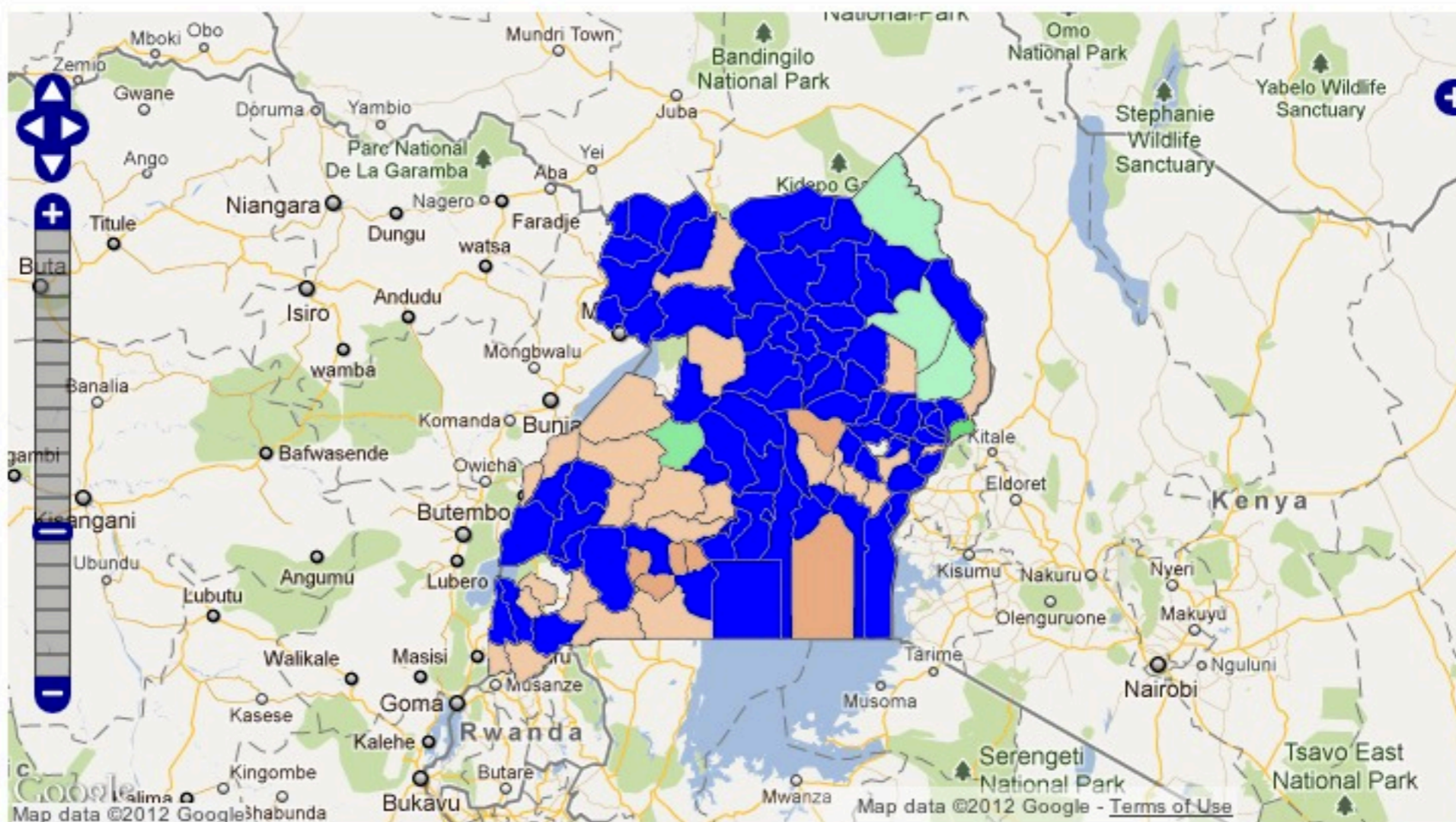
YOUR SAY: Where do women in your community go to get antenatal care?

"In our community we do not have health centre, but we have one in Alebtong which is about 25 km from our community."

Geoffrey, Lira UReporter

CURRENT POLL - 29/Feb/2012 :

 Hi U-reporters. Do you think the Youth Fund will help the youth in your community? Answer YES or NO, and tell us why.



<http://ureport.ug/>



<http://ushahidi.com/>

Libya Crisis Map

SEARCH



HOME

BIG MAP

REPORTS

GET ALERTS

DOWNLOAD REPORTS

⚠ The LibyaCrisisMap platform was activated by the request of the Office for the Coordination of Humanitarian Affairs (OCHA) to the Standby Task Force (SBTF). The platform was fully handed over to OCHA supervision in April 2011, and continued to be supported by a team of volunteers until June 4th of 2011.

The platform is not currently accepting new reports, but the data will remain available for a couple of months.

Click on map icons to see local reports

↓ CATEGORY FILTER [HIDE]

- All Categories
- Internal :: Sources
- Events :: Armed Confrontations
- Events :: Attacks on Protected Persons
- Events :: Attacks on Protected Objects
- Events :: Explosive Remnants of War
- Needs
- Response
- Events :: Mass Displacements
- Internal :: Verification



<http://libyacrisismap.net/>

Open Government Partnership

Initiative: The bigger picture

Kenya has stated its intention to participate in the global Open Government Partnership, a global movement of Countries that seek to empower their citizens to be involved in active governance, engaging Kenyans in the development of their Communities. This is just a taste of what's to come. Our information is a national asset, and it's time it was shared: this data is key to improving transparency; unlocking social and economic value; and building Government 2.0 in Kenya.

Open Government Partnership

Latest Datasets and Views

Poverty Rate, by District 

Kenya Primary Schools, 2007 

County Expenditures by Administration 2002-3 to 2008-9 

Health Facilities 

County Urbanization-2009 



Kenya County Fact Sheets Dec 2011

Immediately after the next General Elections, forty-seven County Governments will replace the Provincial and Local Government administration systems, which were created at independence.



Budget 2010 to 2011 recurrent plus development

Chart shows total expenditure by administrative vote. Size of box shows how large the vote is in the budget. Color shows changes from last year.



Quarterly GDP 2000 - 2011 Chart

KNBS: Quarterly Gross Domestic Product by Economic Sector, 2000 - 2011, KSh Million at Constant 2001 prices



Counties

Information on the 47 Counties can be found here. Easily find data sets relating to a specific County.



Developer? Click here!

Find out about our powerful developer APIs and tools.



USING DATA
IN THE SERVICE
OF HUMANITY



[HOME](#)

[WHO WE ARE](#)

[GET INVOLVED](#)

[EVENTS](#)

[BLOG](#)

NEW YORK CITY DATADIVE

On the weekend of October 14th, 2011, over 50 engaged data scientists, developers, and digital do-gooders got together to work on using data for social change at our New York Datadive. Teams worked with the [UN Global Pulse](#), [The Microfinance Information Exchange Market \(MIX Market\)](#), and the [New York Civil Liberties Union](#) to better understand their data and build visualizations, analyses, and applications around them. You can view the results for each group below and see [pictures of the event on our Flickr account](#).

// MIX MARKET



HOW DO YOU SCRAPE 20 DIFFERENT SITES INTO ONE DATABASE?

The Microfinance Information Exchange Market was interested in gathering more data about microfinance loans taking place in Africa. In order to do this, they needed to scrape information from over 20 banking institutions in different formats and from different websites. See how Data Without Borders worked with MIX to deliver a year's worth of data scraping work to them in just over 24 hours.

[Overview of Results](#) | [Wiki Page](#)

// THE NEW YORK CIVIL LIBERTIES UNION

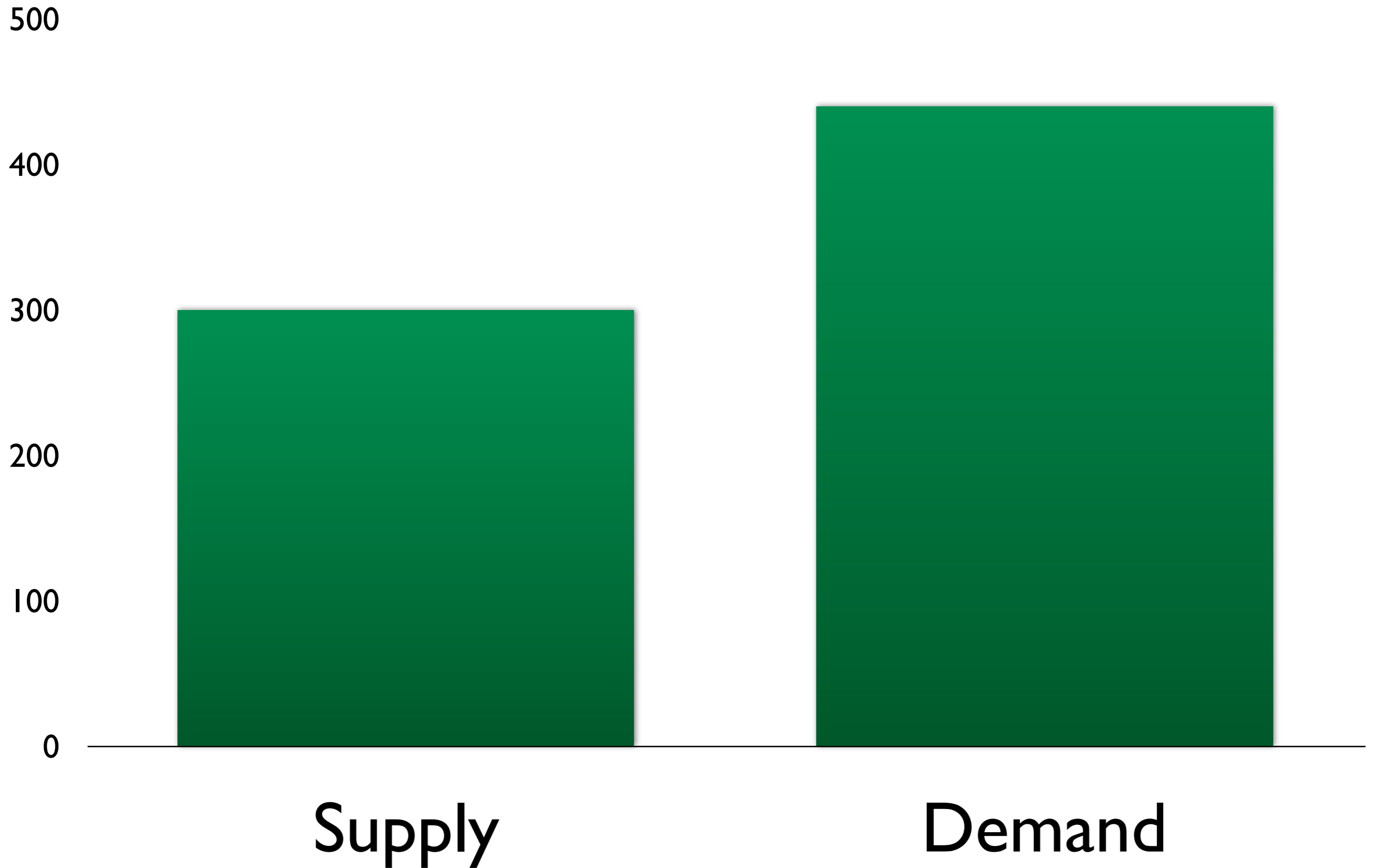


IS THE NYPD USING RACIAL DISCRIMINATION WHEN STOPPING AND FRISKING?

The NYCLU came to our Datadive wondering if they could better understand whether the NYPD was using racial discrimination or not. By looking at publicly available data of stop-and-frisks going back to 2006, the team was able to better understand this thorny issue and created maps and visualizations that breathed life into the data. See how Data Without Borders worked with NYCLU to create tools to better understand the NYPD's data.

<http://datawithoutborders.cc/>

Deep Analytical Talent in 2018



Deep Analytical Talent in 2018

500

400

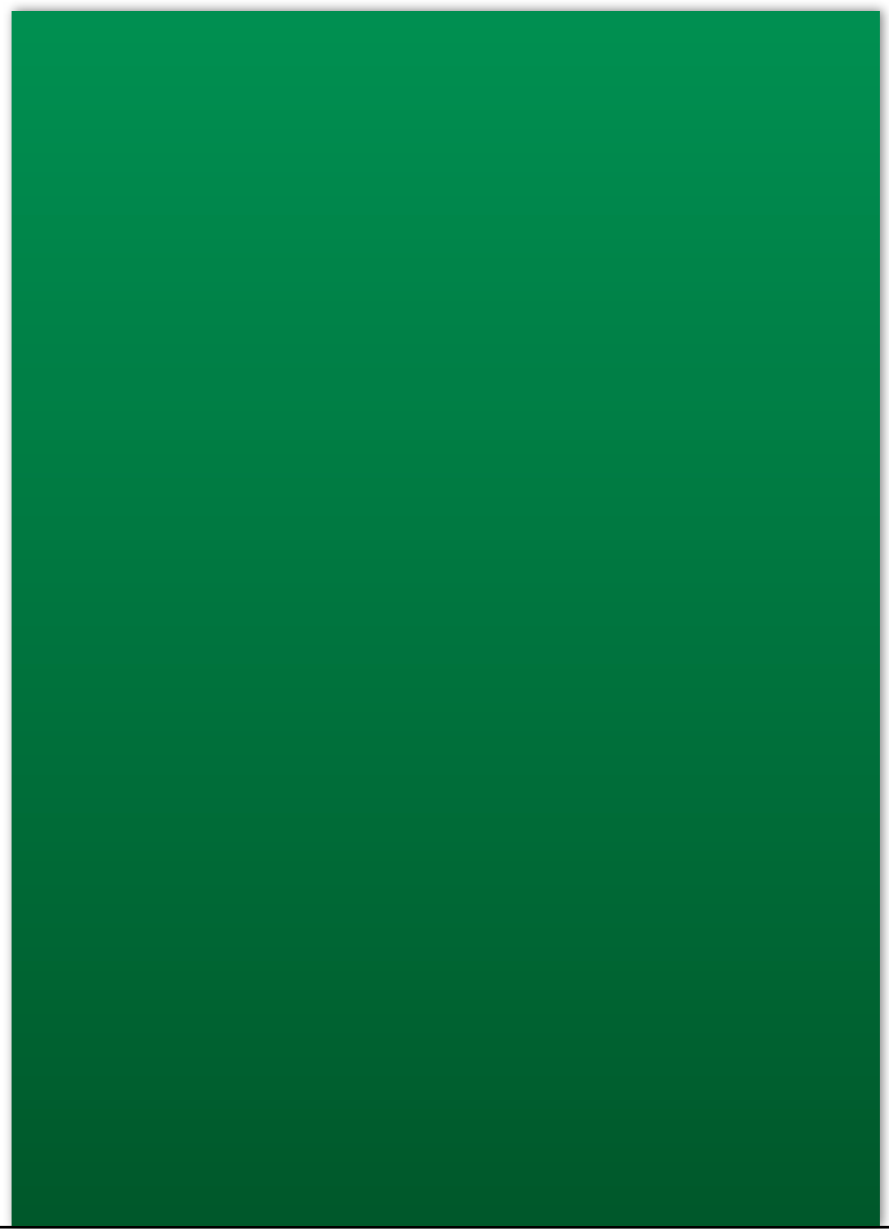
300

200

100

0

50% of supply



Supply

Demand

THINKING,
FAST AND SLOW



DANIEL
KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

Changing Nature of Data

Response

How we use data now

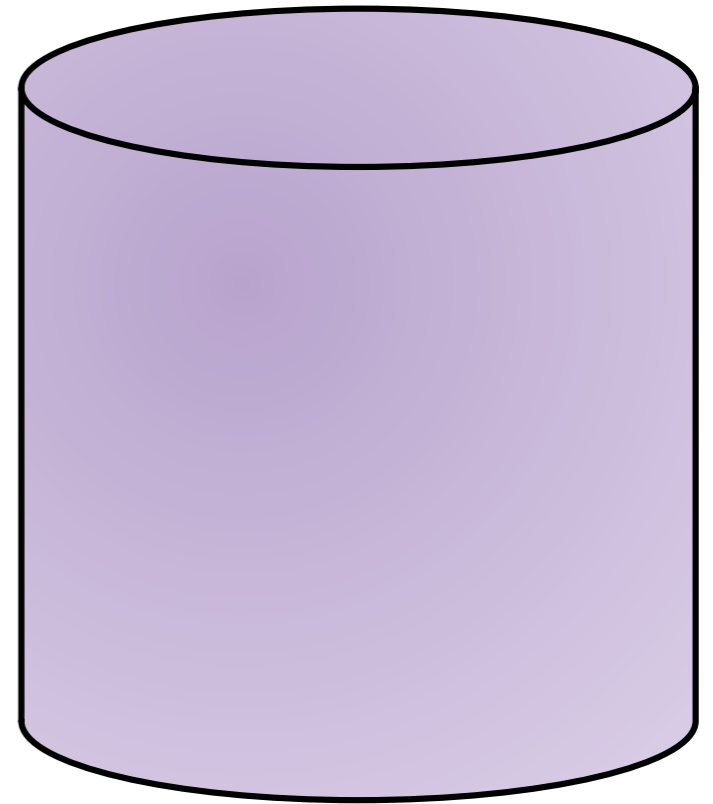
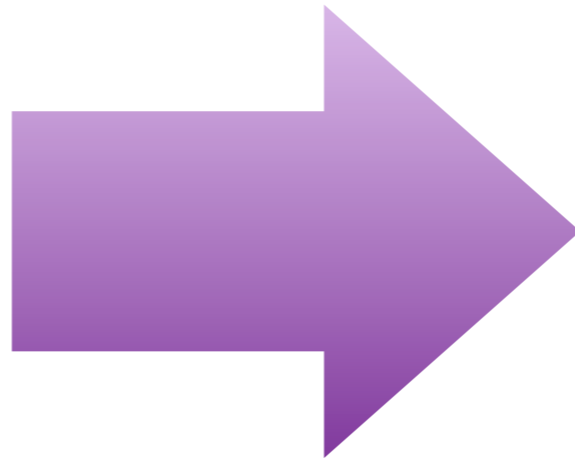
Changing Nature of Data

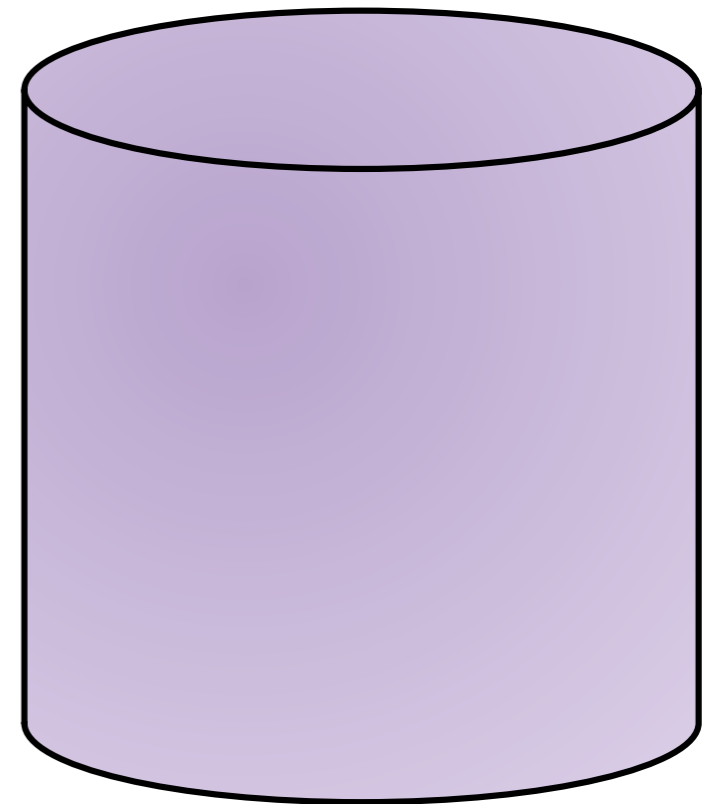
Response

How we use data now

What about

vs?







MARTIN FOWLER

ThoughtWorks®

<http://martinfowler.com>
<http://thoughtworks.com>
[@martinfowler](#)

REBECCA PARSONS CTO

ThoughtWorks®

rparsons@thoughtworks.com
<http://thoughtworks.com>

clip art from <http://openclipart.org>