## Scalable, Good, Cheap

a tale of sexiness, puppets, shell scripts, and python

From this...



...to this!



# Get your infrastructure started right!

(not just preparing for incident and rapid event response)

## Who we are?

Avleen Vig (@avleen)

- Senior Systems Engineer at Etsy
- Good at: Scaling frontends, python
- Previous companies: WooMe, Google, Earthlink

Marc Cluet (@lynxman)

- Senior Systems Engineer at WooMe
- Good at: Backend scaling, bash/python, languages
- Previous companies: RTFX, Tiscali, World Online

## Overview

- Workflow
- Why planning for scaling is important
- How do you choose your software
- Setting up your infrastructure
- Managing your infrastructure

## The background

- Larger startup, \$32m in funding
- 6 million+ active users
- Dozens of developers
- 6 systems administrators
- 4 DBAs
- 10+ code releases every day
- Geographically distributed employees
  - $\circ$  Brooklyn HQ
  - Satellites in Berlin, San Francisco
  - Small number of remote employees



## The background

- Small, funded start up
- 6 python developers
- 2 front end developers
- 3 systems administrators
- 1 DBA (moustache included)
- Multiple code releases every day
- Geographically distributed employees
  - Berlin, Copenhagen, Leeds, London, Los Angeles, Oakland, Paris, Portland, Zagreb



introducing the world

## Workflow

- Ticket systems

   Ticket, or it didn't happen!
- Documentation

   Wikis are good
- Don't Repeat Yourself
   If you keep doing the same thing manually, automate
- Version control everything

   All of your scripts
   All of your configurations

## Workflow

- Everything will change
- Technical debt vs Premature optimisation
   If you try to be too accurate too early, you'll fail



## **Team integration**

- Be sure to hire the right people
   Beer recruitment interview
- Encourage speed

   Release soon and release often
- Embrace mistakes as part of your day to day
   Learn to work with it
- Ask for peer reviews for important components
   O Helps sanity checking your logic
- Developers, Sysadmins, DBAs, one team

## **Team communication**

- Team communication is the most critical factor
- Make sure everyone is in the loop
- Useful applications
  - $\circ$  IRC
  - Skype
  - o email
  - o shout!
- Don't be afraid to use the phone to avoid miscommunication

## Layering! Not just for haircuts.

Separate your systems

- Front end
- Application
- Database
- Caching

## Choosing your software

- What does your software need to do?

   FastCGI / HTTP proxy? Use nginx
   PHP processing? Use apache
- What expertise do you already have?
   Stick to what you're 100% good at
- Don't rewrite everything
   If it does 70% of what you need it's good for you

## Release management

- Fast and furious
- Automate, automate, automate
- Script your deploys and rollbacks
- Continuous deployment
- MTTR vs MTBF

## MTTR vs MTBF

## Maintainability





## **MTTR Optimized**

## **MTBF** Optimized

More info here: http://ti.arc.nasa.gov/projects/ishem/Papers/ONeill\_Maintainability.doc

## Logging

- Centralize your logging
  - $\circ$  syslog-ng
- Parsing web logs the secret troubleshooting weapon

 $\circ$  SQL

 $\circ$  Splunk

## Web logs in a database!

CREATE TABLE access ( ip inet, hostname text, username text, date timestamp without time zone, method text, path text, protocol text, status integer, size integer, referrer text, useragent text, clienttime double precision, backendtime double precision, backendip inet, backendport integer, backendstatus integer, ssl\_cipher text, ssl\_protocol text, scheme text );

## Web logs in a database!





• Alerting vs Trend analysis

Last Update: Thu Novi 11 16:42 2 PST 2010   Name   Borne    Borne   Borne  <		Tactical Monitoring	Overview		N	Ionitoring Perfo	ormance		
General       Nagoo 82.11 - www.nagoc.co         Wome       Documentation         Home       0.00 / 0.38 / 0.14 sec         Home       0.00 / 0.00 / 0.00 / 0.00 / 0.00 sec         # Active Overview       8 ervice Checks: 8 / 1 / 524         # Service Detail       Network Outages         Host Detail       Network Outages         Ø Outages       0 Outages         # Service Problems       0 Down         # Service Problems       0 Down         Ø Down       0 Unreachable       81 Up         Ø Downine       Services Problems         Ø Down       0 Unreachable       81 Up         Ø Downine       Services         Ø Downine       Services         Ø Downine       Service Problems         Ø Downine       Service Problems         Ø Downine       Services         Ø Downine       Service Problems         Ø Downine       Service Problems         Ø Downine       Service Problems         Ø Downine       Service Problems	Naglos	Last Updated: Thu Nov 1 Updated every 90 second	1 16:44:22 PST 2010 Is		9	Service Check E	xecution Time:	0.01/1	0.12 / 0.416 sec
Nome       Nome       Host       0.01 / 5.04 / 0.247 sec         Monicoring       0.00 / 0.00 / 0.00 / 0.000 sec       # Active Check Execution Time:       0.01 / 5.04 / 0.247 sec         Host Check Latency:       0.00 / 0.00 / 0.000 sec       # Active Host / Service Checks:       0.1 / 5.04 / 0.247 sec         Factical Overview       Service Detail       Image: Check Execution Time:       0.01 / 5.04 / 0.247 sec         Monicoring       0       Outros       # Active Host / Service Checks:       0 / 0.00 / 0.000 sec         Services Detail       Network Outages       Network Health       Network Health:       Image: Check Execution Time:       0 / 0.00 / 0.000 sec         Services Poblems       0       Outages       Network Outages       Network Health:       Image: Check Execution Time:       Service Checks:       0 / 0.00         Services Poblems       0       Outros       0       Unreachable       81 Up       0       Pending         Show Host:       Services       0       Down       0       Unreachable       81 Up       0       Pending         Process Info       6       Gritical 6       Warning       0       Unknown       1512 Ok       0       Pending         Process Info       Services       Service Feabled       Monitoring Features       <	General	Nagios® 2.11 - www.nagi	os.org		9	Service Check L	atency:	0.00 / 0	.36 / 0.144 sec
• Documentation             Monitoring             • Tractical Overview             Service Detail             Host proup Overview             Bervice Detail             Host proup Overview             Bostricegroup Summary             Service Problems             Show Host:             Comments             Porcess Info         Porcestretide         Porcest Info         Porcess Info	• Home	Logged in de drieen			H	lost Check Exe	cution Time:	0.01 / 5	.04 / 0.247 sec
Monitoring       # Active Host / Service Checks: 81/1524         * Tactical Overview       Service Datail         * Host Overview       Service Datail         * Host Overview       Network Outages         * Servicegroup Summary       Network Outages         * Servicegroup Summary       Network Health         * Servicegroup Summary       Network Health         * Servicegroup Summary       Hosts         * Servicegroup Summary       Hosts         * Servicegroup Summary       Network Outages         * Service Problems       0 Down         * Network Outages       0 Down         * Network Outages       0 Down         * Services       0 Down         * Opending       Network Outages         * Services       0 Down         * Opending       Network Outages         * Show Host:       Services         * Opending       Intervices         * Opending       Outnowdo Problems         * Altor Histogram       Notifications         * Performance Info       Scheduling Queue         * Altor Histogram       Notifications         * Availability       Alt Hosts Enabled       M Services Enabled         * Alt Histogram       Alt Hosts Enabled       M Hosts	Documentation				H	lost Check Late	ncy:	0.00 / 0	.00 / 0.000 sec
Tactical Overview       # Passive Host / Service Checks: 0 / 0         Service Datail       Network Outages         Mestgroup Overview       Network Outages         Servicegroup Summary       0 Outages         Servicegroup Summary       Service Health:         Servicegroup Grid       Service Problems         Mestgroup Grid       0 Outages         Service Problems       0 Down         Mestgroup Summary       Network Outages         Service Problems       0 Down         Mestgroup Summary       Services         Services       0 Down         Oburneachable       81 Up         Opending       Services         Status Map       0 Unreachable         Network Outages       Services         Show Host:       Services         Services       0 Unnandog Problems         Show Host:       Services         Services       Services         Monitoring Features       Services Enabled         Process Info       Notifications         Scheduling Queue       Notifications         Maint Histogram       All Services Enabled         All Hosts Enabled       All Services Enabled         All Hosts Enabled       All Hosts Enabled	Monitoring				#	# Active Host / S	ervice Checks:	81 / 152	24
Network Outages       Network Outages         Hostgroup Summary       0 Outages         Servicegroup Overview       0 Outages         Servicegroup Grid       0 Outages         Servicegroup Grid       0 Outages         Servicegroup Grid       0 Down         Servicegroup Grid       0 Down         Servicegroup Grid       0 Down         Services       0 Down         O Down       0 Unreachable         81 Up       0 Pending         Network Outages       0 Down         Show Host:       Services         Comments       0 Critical         Obowntime       6 Critical         Services       6 Critical         Scheduling Queue       Monitoring Features         Reporting       Flap Detection         Notifications       Event Handlers         Alter Histogram       Altors Enabled         Alter Histogram       Altors Enabled         Alter Summary       NIA         Outages       Altors Enabled         Configuration       Configuration	Tactical Overview     Service Detail				?	Passive Host /	Service Checks	:0/0	
Network Outages       Network Outages         I Hostsproup Grid       0 Outages         Servicegroup Grid       0 Outages         Servicegroup Grid       0 Outages         Servicegroup Grid       Service Problems         I Obwn       0 Unreachable         8 Host Problems       0 Down         I Network Outages       0 Down         Show Host:       Services         Somments       0 Continue         Process Info       Performance Info         Performance Info       Services Frabeled         Protest Info       Notifications         Ethory       NA Hosts Enabled         Monitoring Features       Notifications         Flap Detection       Notifications         Services Enabled       Al Hosts Enabled         NA Hosts Enabled       Al Hosts Enabled         Alt History       Al Hosts Enabled       Al Hosts Enabled         Notifications       Event Handlers       Active Checks       Passive Checks         Pasit History       NA Hosts Enabled       Bit History       Al Hosts Enabled       Bit History         Notifications       Event Handlers       Active Checks       Passive Checks         Process Info       Performance Info       Al Services Enab	Hostgroup Overview	Natural Outside					Matural	Useláb	
Servicesproup Overview       Bervicesproup Overview       Host Health:         Servicesproup Grid       Status Map         3-D Status Map       Bervices Problems         0 Down       0 Unreachable       81 Up       0 Pending         Service Problems       0 Down       0 Unreachable       81 Up       0 Pending         Shatus Map       0 Down       0 Unreachable       81 Up       0 Pending         Show Host:       Services       Services       Services         Comments       6 Critical       6 Warning       0 Unknown       1512 Ok       0 Pending         Process Info       Scknowloged       Scknowloged       Services Filap Detection       Notifications       Event Handlers       Active Checks       Passive Checks         Process Info       Scknowloged       Notifications       Event Handlers       Active Checks       Passive Checks         Process Info       Notifications       Event Handlers       Active Checks       Passive Checks         Process Info       Not Mitoring Features       Monitoring Features       Not Mitoring Features       Not Hosts Enabled       Not Services Enabled         All Hosts Enabled       NA Services Enabled       Nat Hosts Enabled       Nat Hosts Enabled       Nat Hosts Enabled       Nat Hosts Enabled	Hostgroup Summary Hostgroup Grid	Network Outages					Network	Health	
Servicegroup Summary Servicegroup Side Status Map Service Problems Network Outages Network Outages Show Host: Comments Downtime Process Info Scheduling Queue Scheduling Queue NA Difference Disabled All Hosts Enabled All Hosts Enabled NA Services Enabled Scheduling Queue NA Services Enabled All Hosts Enabled NA Services Enabled All Hosts Enabled Services Enabled All Hosts Enabled Services Enabled All Hosts Enabled Services Enabl	Servicegroup Overview	0 Outages					Host He	alth:	
<ul> <li>Status Map</li> <li>Service Problems</li> <li>Network Outages</li> <li>Down</li> <li>Unreachable</li> <li>Up</li> <li>Down</li> <li>Unreachable</li> <li>Up</li> <li>Unreachable</li> <li>Up</li> <li>Unknown</li> <li>Unknow</li></ul>	Servicegroup Summary Servicegroup Grid						Service	Health:	
Service Problems       Hosts         I Most Problems       0 Down       0 Unreachable       81 Up       0 Pending         I Most Problems       0 Down       0 Unreachable       81 Up       0 Pending         Show Host:       Services       6 Critical       6 Warning       0 Unknown       1512 Ok       0 Pending         I Unrandled       6 Untandled Problems       Services       6 Critical       6 Warning       0 Unknown       1512 Ok       0 Pending         Process Info       Performance Info       Scheduling Queue       Monitoring Features       Flap Detection       Notifications       Event Handlers       Active Checks       Passive Checks         Part Histogram       All Hosts Enabled       NIA       Total Services Enabled       Total Services Enabled       Total Services Enabled       All Hosts Enabled       Total Services Enabled       T	Status Map								
Service Problems Network Outages Show Host: Show Host: Comments Down 0 Unreachable 81 Up 0 Pending Services 6 Critical 6 Warning 0 Unknown 1512 Ok 0 Pending 6 Critical 6 Warning 0 Unknown 1512 Ok 0 Pending 8 Unhandied Problems 5 Acknowledged Process Info 9 Process Info 9 Anitoring Features 1 Unhandied Problems 9 All Services Enabled All Hosts Enabled All Hosts Enabled All Hosts Enabled All Hosts Enabled 1 Hosts	Operatus Map	Hosts							
Network Outages Show Host: Show Host: Comments Comments Downtime Porcess Info Performance Info Scheduling Queue Monitoring Features Reporting Trends Availability Alert History Alert History Alert History Alert History Alert History Alert History Configuration Configuration Configuration	Service Problems Host Problems	0 Down	0 Unreachable	81 Up	0 Pending				
Show Host:              • Comments         • Downtime         • Porcess Info         • Performance Info         • Scheduling Queue         • Scheduling Queue         • Trands         • Availability         • Alert Histogram         • Alert Histogram         • Alert Histogram         • Alert Histogram         • Notifications         • Event Log         • Configuration	Network Outages								
<ul> <li>Services</li> <li>Comments</li> <li>Downtime</li> <li>Process Info</li> <li>Performance Info</li> <li>Scheduling Queue</li> <li>Monitoring Features</li> <li>Flap Detection</li> <li>Notifications</li> <li>Alert Histogram</li> <li>Configuration</li> </ul>	Show Host:								
• Comments         • Downtime           • Comments         • Downtime           • Comments         • Downtime           • Comments         • Downtime           • Comments           • O Pending             • Process Info         • Performance Info         • Scheduling Queue           • Comments           • Co		Services							
Comments   Downtime   Process Info   Performance Info   Scheduling Queue   Monitoring Features   Flap Detection   Notifications   Event Handlers   Active Checks   Passive Checks   All Services Enabled   All Services Enabled   All Hosts Enabled   All Hosts Enabled   NA   Performance     Image: Comments     Solution:     Configuration     Configuration	<b>.</b> .	6 Critical	6 Warning	0 Unknown	1512 Ok	0 Pe	nding		
<ul> <li>Process Info</li> <li>Performance Info</li> <li>Scheduling Queue</li> <li>Acknowledged</li> </ul> Monitoring Features           Reporting         Monitoring Features           Flap Detection         Notifications         Event Handlers         Active Checks         Passive Checks           Alert History         Alert Summary         N/A         Tendle         All Services Enabled         All Hosts Enabled <t< td=""><td>Comments Downtime</td><td>1 Unhandled</td><td>6 Unhandled Problems</td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	Comments Downtime	1 Unhandled	6 Unhandled Problems						
Performance Info     Scheduling Queue      Monitoring Features      Iservice Disabled     Availability     Alert History     Alert History     Alert Summary     NA     Deg      Iservice Disabled     All Hosts Enabled     All Hosts Enable	Process Info	5 Acknowledged							
Scheduling Queue       Monitoring Features         Reporting       Monitoring Features         Trends       Flap Detection       Notifications       Event Handlers       Active Checks       Passive Checks         Availability       All Service Disabled       All Services Enabled       All Hosts Enabled         Alert History       Alert Summary       N/A       Base of the second	Performance Info								
Reporting       Flap Detection       Notifications       Event Handlers       Active Checks       Passive Checks         Availability       Availability       All Service Disabled       All Services Enabled       All Services Enabled       All Hosts	Scheduling Queue	<b>Monitoring Features</b>							
<ul> <li>Trends</li> <li>Availability</li> <li>Alert Histogram</li> <li>Alert History</li> <li>Alert Summary</li> <li>Notifications</li> <li>Event Log</li> </ul>	Reporting	Flap Detection	Notifications	Event Handlers	Active	Checks	Passive Checks		
© Event Log	<ul> <li>Trends</li> <li>Availability</li> <li>Alert Histogram</li> <li>Alert History</li> <li>Alert Summary</li> </ul>	Disabled V/A	All Hosts Enabled	All Services Enabled All Hosts Enabled	All Serv All Host	ices Enabled s Enabled	All Services Enable All Hosts Enabled	d	
	© Event Log								

View Config

Alerting vs Trend analysis

 Nagios is great for raising alerts on problems



	Time and String Metrics
Last Boot Time	Tue, 31 Aug 2010 16:36:59 -0700
Gexec Status	OFF
Gmond Started	Mon, 06 Sep 2010 12:16:42 -0700
Last Reported	0 days, 0:00:03
Machine Type	x86
mysql_version	5.1.47-rel11.2-log
Operating System	Linux
Operating System Release	2.6.26-2-xen-686
Uptime	72 days, 1:08:55

	Constant Metrics
CPU Count	2 CPUs
CPU Speed	2795 MHz
Memory Total	917700 KB
Swap Space Total	0 KB



- Alerting vs Trend analysis
  - $\circ$  Nagios is great for raising alerts on problems
  - Ganglia is great at long term trend analysis
  - $\circ$  Know when something is out of the "ordinary"

- Alerting vs Trend analysis
  - Nagios is great for raising alerts on problems
  - Ganglia is great at long term trend analysis
  - $\circ$  Know when something is out of the "ordinary"
- What should you monitor?
  - $\circ$  Anything which breaks once
  - $\circ$  Customer facing services

- Alerting vs Trend analysis
  - Nagios is great for raising alerts on problems
  - Ganglia is great at long term trend analysis

 $\circ$  Know when something is out of the "ordinary"

- What should you graph?
  - Everything! If it moves, graph it.

 $\circ$  Customer facing rates and statistics

Get statistics from your logs:

- PostgreSQL: pgfouine
- MySQL: mk-query-digest
- Web servers: webalizer, awstats, urchin
- Custom applications: Do it yourself! Integrate with Ganglia





• Caches are disposable





- Caches are disposable
- But what about the thundering herd?



August 2003 Northeastern US and Canada blackout

 $\circ$  Caused by poor process execution

 $\circ$  Lack of good monitoring

 $\circ$  Poor scaling



- Massive destruction avoided!
  - $\circ$  256 power stations automatically shut down
  - $\odot$  85% after disconnecting from the grid
  - Power lost but plants saved!

## Caching

- Caches are disposable
- But what about the thundering herd?

○ Increase backend capacity along with cache capacity

- $\circ$  Plan for cache failure
- $\circ$  Reduce demand when cache fails

## Caching

- Find out how your caching software works
  - Memcache + peep!
  - $\circ$  Is it better with lots of keys and small objects?
  - $\odot$  Or fewer keys and large objects?
  - How is memory allocated?

## Caching

- Caches are disposable
   Solved!
- But what about the thundering herd?
   Solved!
- Now we get into database scaling!
   Over to Marc...



Databases...

or how to live and die dangerously



#### SQL or NoSQL?



#### • SQL

- $\circ$  Gives you transactional consistency
- $\circ$  Good known system
- $\circ$  Hard to scale

#### NoSQL

- Transactionally consistent "eventually"
- $\circ$  New cool system
- $\circ$  Easy to scale

#### • SQL

- Gives you transactional consistency
- $\circ$  Good known system
- $\circ$  Hard to scale

#### NoSQL

- Transactionally consistent "eventually"
- $\circ$  New cool system
- $\circ$  Easy to scale

#### You may end up using BOTH!

• Be smart about your table design

Be smart about your table design

 Keep it simple but modular to avoid surprises

#### You need to design your database right!



- Be smart about your table design
  - $\circ$  Keep it simple but modular to avoid surprises
  - Don't abuse many-to-many tables, they will just give you hell



- Be smart about your table design
  - $\circ$  Keep it simple but modular to avoid surprises
  - Don't abuse many-to-many tables, they will just give you hell

#### • YOU WILL GET IT WRONG

You'll need to redesign parts of your DB semi-regularly
 Be prepared for the unexpected

#### The read dilemma

- As the tables grow so do read times and memory. Several options:
  - Check your slow query log, tune indexes
  - Partition to read smaller numbers of rows
  - Master / Slave, but this adds replication lag!

#### The read dilemma

- As the tables grow so do read times and memory. Several options:
  - $\circ$  Check your slow query log, tune indexes
    - Single most common problem with slow queries and capacity
    - Be careful about foreign keys

#### The read dilemma

- As the tables grow so do read times and memory. Several options:
  - Check your slow query log, tune indexes
  - $\circ$  Partition to read smaller numbers of rows
    - By range (date, id)
    - By hash (usernames)
    - By anything you can imagine!

#### The write conundrum

- As the database grows so do writes
- Writes are bound by disk I/O

   RAID1+0 helps
- Don't shoot yourself in the foot!

   Don't try to solve this early
   Have monitoring ready to foresee this issue
   Bring pizza

Divide writes!

• Remember about modular? This is it





#### How to give a consistent view to the servers?

Use a query director!

- pgbouncer on Postgres
- gizzard on MySQL

## Web frontend

- Hardware load balancers Good but expensive!
- Software load balancers Good and cheap! (more pizza)

Web server frontends
 nginx, lighttpd, apache

Reverse proxies
 varnish, squid

○ Kernel stuff■ Linux ipvs

## Web frontend

Which way should I go?

Web servers as load balancers

 Gives you nice add on features
 You can offload some process in the frontend
 Buffering problems

Reverse proxies

- Caching stuff is good
- $\circ$  Fast reaction time
- $\circ$  No buffering problems

## Web frontend

Divide your web clusters!

- You can send different requests to different clusters
- You can use an API call to connect between them

## **Configuration management**

- Be ready to mass scale
   Keep all your machines in line
- Automated server installs

   Use it to install new software
   Also to rapidly deploy new versions

## Writing tools

- If you do something more than 2 times it's worth scripting
- Write small tools when you need them
- Stick to one or two languages
   And be good at them

## Writing tools

- Even better
- Have your scripts repo in a cvs and push it everywhere

changeset:	3646:455b84f75c21
user:	mcluet
date:	Mon Oct 25 10:14:56 2010 -0700
summary:	Marc hates typos

• It's important to have backups

- It's important to have backups
- It's even more important to exercise them!
   Having backups without testing recovery is like having no backups

- It's important to have backups
- It's even more important to exercise them!
   Having backups without testing recovery is like having no backups
- How can we exercise backups for cheap?

- It's important to have backups
- It's even more important to exercise them!

   Having backups without testing recovery is like having no backups
- How can we exercise backups for cheap?
   O Cloud computing!

## **Cloud computing**

- Cloud computing help us recreate our platform on the cloud
- Giving us a more than credible recovery scenario
- Also very useful to spawn more instances if we run into problems

## Interesting things to read

Wikipedia

http://en.wikipedia.org/wiki/DevOps

Web Operations and Capacity Planning

http://kitchensoap.com

High scalability (if you get there)

http://highscalability.com/

If you really fancy databases, explain extended

http://explainextended.com/

## **Questions?**

#### Work at Etsy! http://etsy.com/jobs

#### Work at WooMe! http://bit.ly/work4woome





introducing the world

@lynxman

