THE ROAD TO 2 MILLION WEB SOCKET CONNECTIONS WITH PHOENIX

BLOG POST

http://www.phoenixframework.org/blog/the-road-to-2-million-websocket-connections



GARY RENNIE Gazler @TheGazler

TALK FORMAT

- Why?
- How? (Beware: contains XML)
- Results
- Future



- WhatsApp used as a reference for all Phoenix presentations
- Lots of Phoenix HTTP benchmarks, 0 for persistent connections

<u>https://blog.whatsapp.com/196/1-million-is-so-2011</u> <u>https://github.com/mroth/phoenix-showdown</u> <u>http://www.littlelines.com/blog/2014/07/08/elixir-vs-ruby-showdown-phoenix-vs-rails/</u>

WHAT MAKES BENCHMARKING WEB SOCKETS DIFFERENT?

- Uses Web Socket protocol
- Connections have to stay open
- Can't recycle connections
- Increased memory usage

• Hard limit on connections per machine (~64k however adding additional ip addresses/using different ports can fix this)

CHAT APPLICATION

Phoenix Chat	source	Learn more about the Phoenix Framework
[Gazler] Hello		
[Chris] Hi		
@ Gazler		

https://github.com/chrismccord/phoenix chat example

diff --git a/web/channels/room_channel.ex b/web/channels/room_channel.ex

index bc92759..66ead5c 100644

--- a/web/channels/room_channel.ex

+++ b/web/channels/room_channel.ex

@@ -14,8 +14,6 @@ defmodule Chat.RoomChannel do

def join("rooms:lobby", message, socket) do
 Process.flag(:trap_exit, true)

- :timer.send_interval(5000, :ping)

- send(self, {:after_join, message})

{:ok, socket}
end

[418 entered] [418] foo				
[420 entered] [420] foo				
[419 entered] [419] foo				
[421 entered] [421] foo				
[422 entered] [422] foo				
[423 entered]				
[423] foo				
[424 entered]				
[424] foo				
[425 entered]				
[425] foo				
[426 entered] [426] foo				
[
@ useri				

TSUNG

- An open-source multi-protocol distributed load testing tool
- Written in Erlang
- Started in 2001 as Idx-Tsunami
- Supports Web Sockets (As of version 1.5)
- Supports multiple machines
- Produces pretty charts (web dashboard as of 1.6)

https://github.com/processone/tsung

TSUNG CONFIG

<?xml version="1.0"?> <!DOCTYPE tsung SYSTEM "/user/share/tsung/tsung-1.0.dtd"> <tsung loglevel="debug" version="1.0"> <clients></clients> <servers></servers> <load></load> <sessions></sessions> </tsung>

- Machine running with the config is the controller
- Other machines are clients (the controller can also be a client)

TSUNG CLIENTS

<clients>

<client host="phoenix1" weight="1" cpu="4" use_controller_vm="false" maxusers="60000" />
<client host="phoenix2" weight="2" cpu="4" use_controller_vm="false" maxusers="60000">
<ip value="10.9.195.12"></ip>
<ip value="10.9.195.13"></ip>
</client>
<client host="phoenix3" weight="1" cpu="4" use_controller_vm="false" maxusers="60000" />
/clients>

- One client per machine
- Keep maxusers below machine limit
- Can set a weight on the machines (load ratio)
- Can set multiple virtual IP addresses (we didn't use this)
- use_controller_vm="false" don't share Erlang VM

TSUNG SERVERS

<servers>

<server host="server1" port="4000" type="tcp" weight="4"></server>
<server host="server2" port="4000" type="tcp" weight="1"></server>
</servers>

- Can set multiple servers (we only used 1)
- Can set a weight on the machines (load ratio)
- Can set a connection type (tcp, ssl, udp, websocket)
- We actually used the tcp type instead of websocket as websocket didn't stay open

TSUNG LOAD

<lead duration="1" unit="hour"> <arrivalphase phase="1" duration="10000" unit="second"> <users maxnumber="100000" arrivalrate="1000" unit="second" /> </arrivalphase> </load>

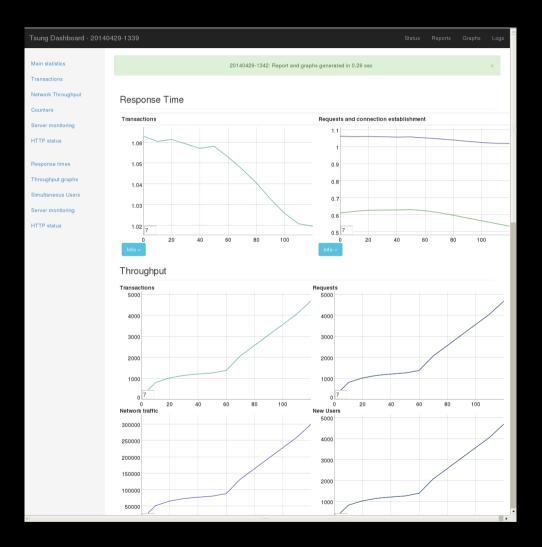
- Can set multiple phases (we used 1)
- Phases have a duration
- Total load can have a duration (max just under 50 days!)
- Phases can be looped
- We set duration high and manually terminated
- Arrival rate is the number of connections arriving

TSUNG SESSIONS

<sessions>

- How users interact with the application (we connect and wait)
- Can use Erlang terms (%%ts_user_server:get_unique_id%%)
- Can use different request types (websocket, http)
- Different probabilities per session

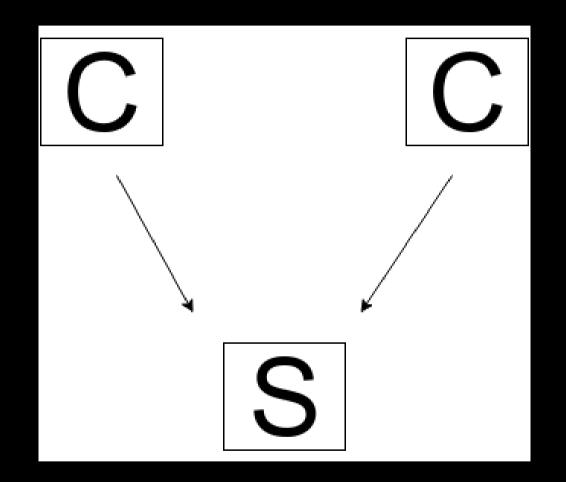
Tsung Dashboard - 20160417-1240	
Status	
Running users	
Connected users	
4452	
Request rate: 1626.45 req/sec	
Active nodes:	
1	
Current phase (total is 1)	
Controller CPU usage	
Tauna varian 1.6.0	
Tsung version 1.6.0 Contact: tsung-users@process-one.net	

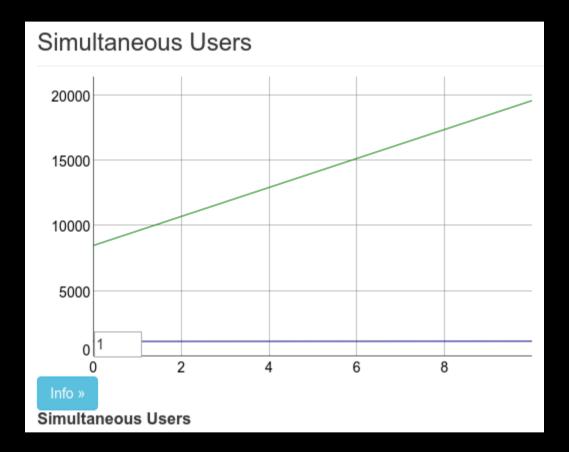


Simultaneous Users

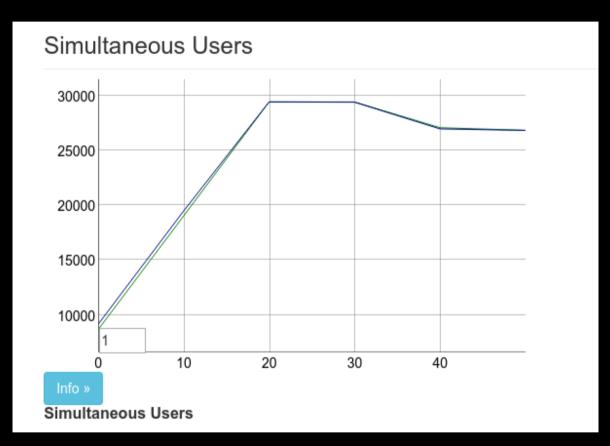
THE FIRST RUN OF TSUNG
tsung -f config.xml start

- 1 server Rackspace I/O v1
- 15GB RAM, 4 cores
- 2 clients as above





SERVER CONFIG ulimit -n 2000000 Make sure it is set after restarts!

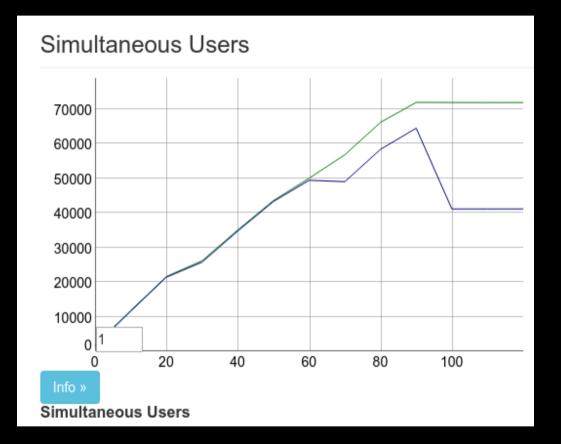


FIRST OPTIMIZATION

061c69b Only rely on ETS tables inside local (5 months ago, José Valim)

lib/phoenix/pubsub/local.ex | 83 +++++++++++++++ 1 file changed, 14 insertions(+), 69 deletions(-)

- Remove local HashDict of PIDs
- Use topics ETS tables to find subscribers (PIDs)



USING OBSERVER

observer

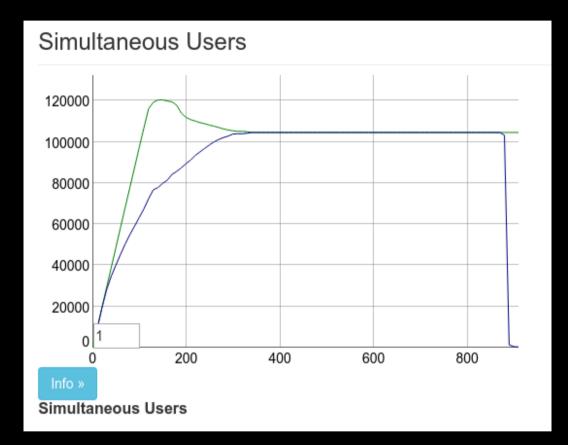
•••				nonode@nohos	st		
	System	Load Charts	Memory Allocators	Applications	Processes	T ole Vi	ewer Trace Overview
Pi	d	Name or Initial Func		Reds	Memory	MsgQ Curre	nt Function
<0.61	.0> t	imer_server		1202342	24776	120 gen_se	erver:loop/6
<0.10	5.0> g	gen:init_it/6		0	8920	0 wx_obj	ject:loop/6
<0.97	'.0> e	erlang:apply/2		0	2752	0 observ	/er_backend:flag_holder_proc/1
<0.95	.0> e	erlang:apply/2		21645	142808	0 observ	ver_pro_wx:table_holder/1
<0.94	.0> (gen:init_it/6		889	24808	0 wx_obj	ject:loop/6
<0.93	.0> g	gen:init_it/6		0	8920	0 wx_obj	ject:loop/6
<0.92	.0> g	gen:init_it/6		0	7048	0 wx_obj	ject:loop/6
<0.91	.0> 0	gen:init_it/6		0	8920	0 wx_obj	ject:loop/6
<0.90	.0> (gen:init_it/6		0	88728	0 wx_obj	ject:loop/6
<0.89	.0> e	erlang:apply/2		572	8744	0 timer:s	leep/1
<0.88	.0> \	wxe_master		0	21680	0 gen_se	erver:loop/6
<0.87	.0> \	wxe_server:init/1		72924	27664	0 gen_se	erver:loop/6
<0.55	.0> E	Elixir.Logger.Watcher	:init/1	0	2864	0 gen_se	erver:loop/6
<0.54	.0> E	Elixir.Logger.Watcher	:init/1	0	2968	0 gen_se	erver:loop/6
<0.53	.0> E	Elixir.Logger.Watcher		0	7016	0 gen_se	erver:loop/6
< 0.52	.0> E	Elixir.Logger.Watcher	:init/1	0	2968	0 gen_se	erver:loop/6
<0.51	.0> 8	Elixir.Logger		0	7264	0 Elixir.G	enEvent:fetch_msg/5
<0.50).0> E	Elixir.Logger.Supervis	or	0	10880	0 gen_se	erver:loop/6
<0.49	.0> a	application_master:st	art_it/4	0	6912	0 applica	ation_master:loop_it/4
<0.48	.0> a	application_master:in	it/4	0	2864	0 applica	ation_master:main_loop/2
<0.45	.0> E	Elixir.IEx.Config		0	2824	0 gen_se	erver:loop/6
<0.44	.0> E	Elixir.IEx.Supervisor		0	5872	0 gen_se	erver:loop/6
<0.43	.0> a	application_master:st	art_it/4	0	2760	0 applica	ation_master:loop_it/4

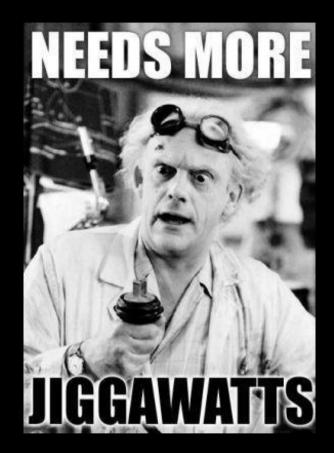
SECOND OPTIMIZATION

7b252f4 Remove uneeded heartbeat since cowboy handles timeouts (5 months ago, Chris McCord)

lib/phoenix/transports/websocket.ex	30 ++
<pre>test/phoenix/integration/websocket_test.exs</pre>	; 11 ++
<pre>test/phoenix/socket_test.exs</pre>	2 +-
3 files changed, 5 insertions(+), 38 deleti	.ons(-)

- :timer.send_after is expensive
- Remove 30s heartbeat
- Cowboy handles heartbeat

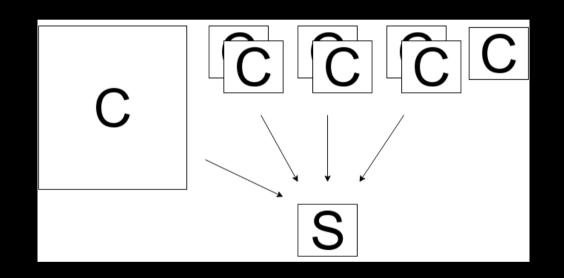


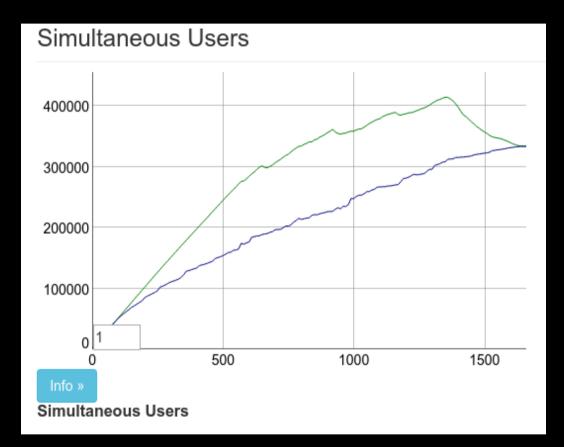


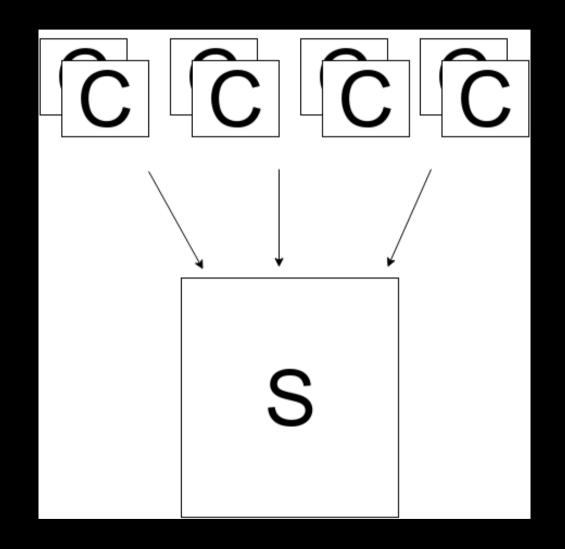
MACHINE SPECS

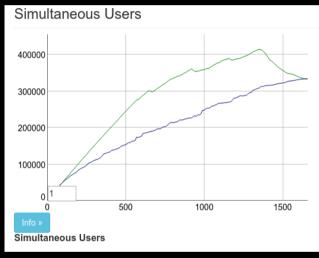
- 1 server Rackspace I/O v1 15GB RAM, 4 cores
- 2 clients as above
- 1 client Rackspace OnMetal 128GB RAM, 40 cores
- 5 client Rackspace general purpose 4GB RAM, 4 cores

1 [0.0%	11 [0.0%	21 [0.0%	31 [0.0%
2 [0.0%	12 [0.0%	22 [0.0%	32 [0.0%
3 [0.0%	13 [0.0%	23 [0.0%	33 [0.0%
4 [0.0%	14 [0.0%	24 [0.0%	34 [0.0%
5 [0.0%	15 [0.0%	25 [0.0%	35 [0.0%
6 [0.0%	16 [0.0%	26 [0.0%	36 [0.0%
7 [0.0%	17 [0.0%	27 [0.0%	37 [0.0%
8 [0.0%	18 [0.0%	28 [0.0%	38 [0.0%
9 [0.0%	19 [0.0%	29 [0.0%	39 [0.0%
10 [0.0%	20 [0.0%	30 [0.5%	40 [0.0%
Mem[618,	/128906MB]	Tasks: 18	3, 6 thr; 1	running	
Swp[0/0MB]	Load aver	rage: 0.08 0	0.06 0.04	
				Uptime: 0	00:06:35		

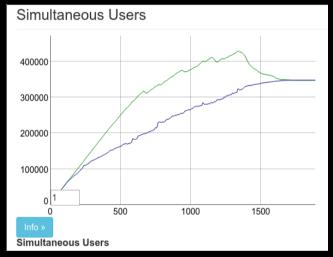








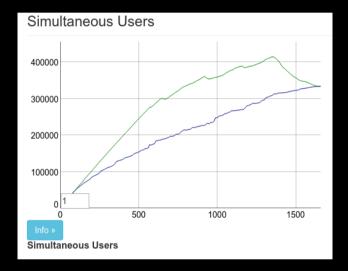
15GB - 4 cores 128GB - 40 cores

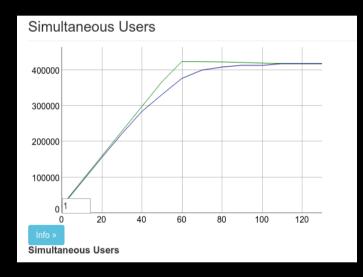


THIRD OPTIMIZATION

bad6b11 Change PG2 ETS from bag to duplicate_bag to improve insert performance for elements with same key (5 months ago, Gabi Zuniga)

lib/phoenix/pubsub/local.ex | 2 +1 file changed, 1 insertion(+), 1 deletion(-)

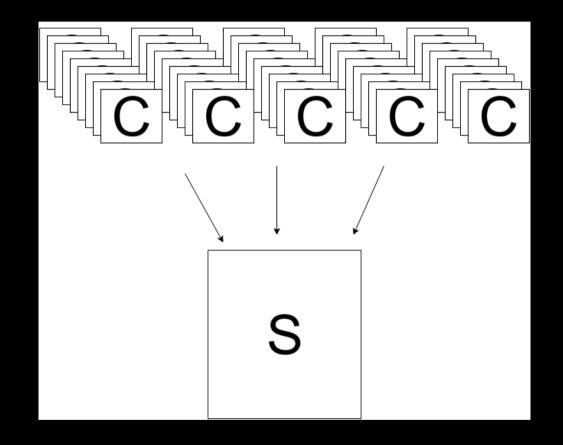




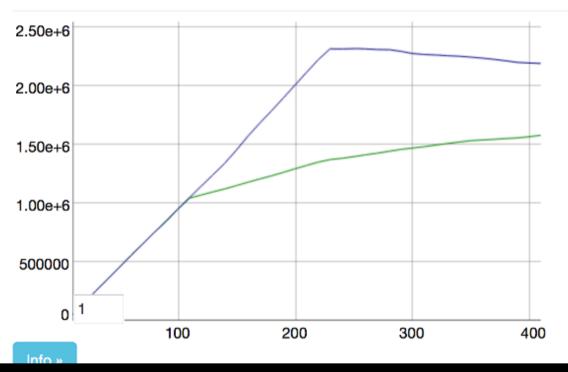
THIRD OPTIMIZATION

- ^local = :ets.new(local, [:bag, :named_table, :public, + ^local = :ets.new(local, [:duplicate_bag, :named_table, :publ

- Know your ETS types!
- Duplicates don't matter since each subscriber unique
- 10x arrival rate increase



Simultaneous Users



PROBLEMS

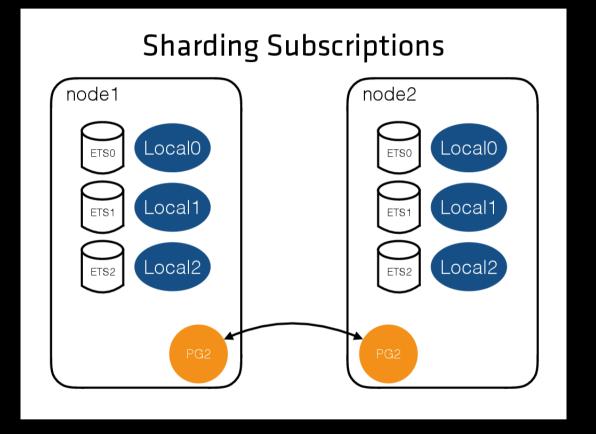
- All low hanging fruit gone (as far as we know!)
- Client crash was resulting in 60k down messages
- Server was timing out on connections
- Broadcasts taking 5 seconds to send to all users
- We tried parallelizing broadcasts but still got timeouts

FOURTH OPTIMIZATION

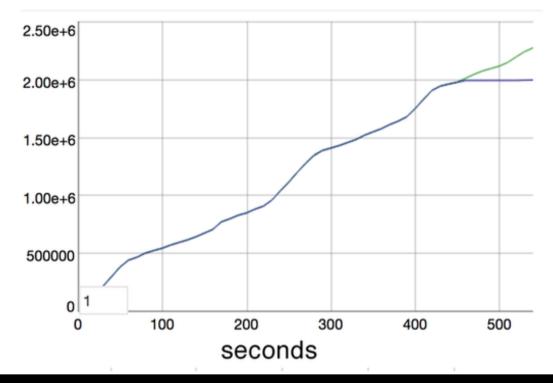
8eb9dfa Add random local pool and sharded subscribers (5 months ago, Chris McCord)

- First time we added code!
- Sharding with a pool of servers and ETS tables
- Shard based on PID
- Use :erlang.phash2(pid, shard_size)
- Configure with `pubsub: [pool_size: 40] `
- Able to maintain 1-2s broadcasts

Sharding Subscriptionsnode1Pub.Loca1



Simultaneous Users



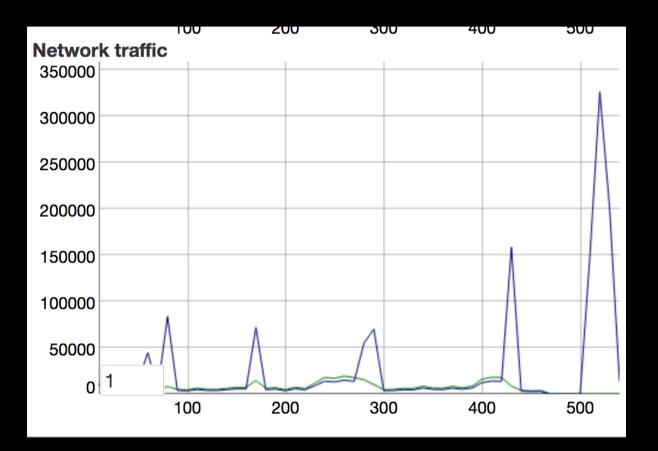
1] 19	7000 7630 9999 9999	630 975	subso	ribers					
i									
	1	[0.0%]	11 []	0.5%	21 [0.0%	31 [0.0%
	2	Ĩ.	0.0%	12	0.5%	22 [0.0%	32	0.0%
	3	Ĩ.	0.0%	13	0.0%	23 [0.0%	33	0.0%
	4	Ĩ.	1.0%	14 [0.0%	24 [0.5%	34 [0.0%
	5	[0.5%]	15 [0.0%	25 [0.0%	35 [0.0%
	6	[0.5%	16 [0.0%	26 [0.0%	36 [0.0%
	7	[0.0%	17 [0.0%	27 [0.0%	37 [0.0%
	8	[]	1.0%	18 [0.0%	28 [0.5%	38 [0.0%
	9	[0.0%	19 [0.0%	29 [0.0%	39	0.0%
	10	[0.0%	20 [0.0%	30 [0.0%	40	0.0%
	Mer	n[83765/12	8906MB]	Tasks:	22, 150 t	hr; 2 rur	nning
	Sw) [0/0MB	Load av	erage: 5.	98 5.45	3.98
						Uptime:	5 days,	11:17:13	

HOW TO OPTIMIZE

- Be José Valim
- Use the tools available to you (observer)
- Isolate the bottlenecks
- Know your data types (ETS)
- Use a pool if a process is bottlenecked

SO DO THE CONNECTIONS DO ANYTHING?

• We have a chat room with 2 million people

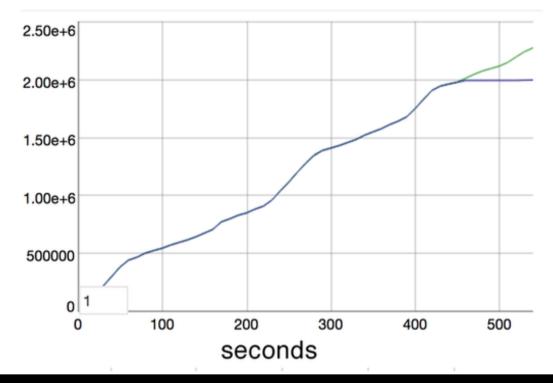




Ci cu o	Not logged in Talk Contributions Create account Log in					
	Article Talk Read Edit View history Search Q					
WIKIPEDIA The Free Encyclopedia	Elixir From Wikipedia, the free encyclopedia					
Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store	For other uses, see Elixir (disambiguation). This article does not cite any sources. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (December 2011) An elixir (from Arabic: الركسيد) - al-ïksīr) is a clear, sweet-flavored liquid used for medicinal purposes, to be taken orally					
Interaction Help About Wikipedia Community portal Recent changes Contact page	and intended to cure one's ills. When used as a pharmaceutical preparation, an elixir contains at least one active ingredient designed to be taken orally. Contents [hide] 1 Types 1.1 Non-medicated elixirs 1.2 Medicated elixirs					
What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page	1.3 East Asian vitamin drinks 2 Composition 3 Storage 4 See also 5 References					
Print/export Create a book Download as PDF	Types [edit] Non-medicated elixirs [edit]					
Printable version In other projects Wikimedia Commons	They are used as solvents or vehicles for the preparation of medicated elixirs: aromatic elixirs (USP), isoalcoholic elixirs (NF), or compound benzaldehyde elixirs (NF). Active ingredient dissolved in a solution that contains 15 to 50% by volume of ethyl alcohol.					
Languages 🔅 العربية	Medicated elixirs [edit]					

	Article Talk	Read	Let view source		Falk Contributions Co	Reate account Log in			
WIKIPEDIA The Free Encyclopedia	Troll From Wikipedia, the free encyclopedi	ia							
Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store	This article is about beings from Scandinavian folklore and mythology. For the internet term, see Internet t								
teraction Help About Wikipedia Community portal Recent changes Contact page	and are rarely helpful to human be Later, in Scandinavian folklore, tro where they live far from human ha considered dangerous to human b which accounts of trolls stem, thei	eings. Ills became beings in the bitation, are not Christia beings. Depending on the	eir own right, nized, and ar e region from	A THE					
ools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page	be ugly and slow-witted, or look ar with no particularly grotesque chan Trolls are sometimes associated w times may be explained as formed are depicted in a variety of media Contents [hide] 1 Norse mythology	nd behave exactly like h racteristic about them. vith particular landmarks I from a troll exposed to	uman beings, , which at sunlight. Trol	sons! You won't find	ou won't find more be	other said. Look at my ⁶³ more beautiful troils on 6. (1915) by John Bauer			
rint/export Create a book Download as PDF Printable version	2 Scandinavian folklore 3 See also 4 Notes 5 References								
other projects Wikimedia Commons Wikiquote	6 External links Norse mythology								
anguages 🌼	In Nerse mythelegy, troll like thur	a is a form applied to its	nor and are	montioned the	roughout the Old N	!			

Simultaneous Users



TSUNG COMMON ISSUES

- Everything requires a host name (clients and controller)
- If /etc/hosts names don't match then you will get an error
- Needs to be correct for every client
- SSH keys need to be set up from the controller to each client
- ulimit needs to be set

PHOENIX CONFIGURATION

- Use production environment
- Disable logging

Parameters: %{"user" => "1569"} Parameters: %{"user" => "1570"} Parameters: %{"user" => "1572"} Parameters: %{"user" => "1578"} Parameters: %{"user" => "1579"} Parameters: %{"user" => "1580"} 16:35:15.082 [info] Replied rooms:lobby :ok

EASIEST WAY TO DO THIS AT SCALE

- Create a server, install erlang, tsung, etc.
- Create SSH key and add it to ~.ssh/.authorized_keys
- Set up own hostname (tsung-controller or something)
- Set ulimit in `/etc/security/limits`
- Create an image
- Spawn new servers from that image

WHAT NEXT?

- Benchmarking is expensive!
- multi-node
- Chat room with messages being sent periodically
- Benchmarking more "chat rooms" on a single server
- Use IP aliasing to make testing require fewer machines
- Automate the tests for each release (both websocket and HTTP)

THANKS FOR LISTENING