

# Distributed Systems

What Nobody Told You



greplin

**My name is Shaneal,  
and I build Distributed Systems**

My name is Shaneal,  
~~and I build Distributed Systems~~

**My name is Shaneal,  
and I'm an addict**

# Greplin?



# Launched Sept 2010

# Un-launched 3 Days Later

# Re-launched Feb 2011



- 100s of (Virtual) Servers
- Billions of Documents
- Adding 5,000+ docs/sec

# 12 Steps!

# Step 1

We admitted we were  
powerless over complexity -  
that our systems had become  
unmanageable.



# PAGERDUTY



PagerDuty (415) 349-0382 - work

5/13/11 2:06 PM 4 months ago

Inbox



PagerDuty: ALRT #5905 on Nagios: Service 'GEC stuck errors' on host 'Hydra-49' is CRITICAL.. Reply 3404:Ack, 3406:Resolv, 3408:Escal8 lvl2 (D Gross) 7:19 AM  
PagerDuty: ALRT: #5906,#5907,#5908,#5909 on Nagios. Reply 3414:Ack all, 3416:Resolv all, 3418:Escal8 all lvl2 (D Gross) 7:47 AM  
PagerDuty: ALRT: 6 open incidents on Nagios. Reply 3424:Ack all, 3426:Resolv all, 3428:Escal8 all lvl2 (D Gross) 7:50 AM  
PagerDuty: ALRT: 11 open incidents on Nagios. Reply 3434:Ack all, 3436:Resolv all, 3438:Escal8 all lvl2 (D Gross) 7:53 AM  
PagerDuty: ALRT: 7 open incidents on Nagios. Reply 3444:Ack all, 3446:Resolv all, 3448:Escal8 all lvl2 (D Gross) 7:56 AM  
PagerDuty: ALRT: #5937,#5938,#5939,#5940,#5941 on Nagios. Reply 3454:Ack all, 3456:Resolv all, 3458:Escal8 all lvl2 (D Gross) 8:01 AM  
PagerDuty: ALRT: 6 open incidents on Nagios. Reply 3464:Ack all, 3466:Resolv all, 3468:Escal8 all lvl2 (D Gross) 8:04 AM  
PagerDuty: ALRT: 8 open incidents on Nagios. Reply 3474:Ack all, 3476:Resolv all, 3478:Escal8 all lvl2 (D Gross) 8:08 AM  
PagerDuty: ALRT: 8 open incidents on Nagios. Reply 3484:Ack all, 3486:Resolv all, 3488:Escal8 all lvl2 (D Gross) 8:11 AM  
PagerDuty: ALRT: #5975,#5976 on Nagios. Reply 3494:Ack all, 3496:Resolv all, 3498:Escal8 all lvl2 (D Gross) 8:15 AM  
PagerDuty: ALRT #5977 on Nagios: Service 'CPU Stats' on host 'Hydra-5' is CRITICAL.. Reply 3504:Ack, 3506:Resolv, 3508:Escal8 lvl2 (D Gross) 8:18 AM  
PagerDuty: ALRT: 16 open incidents on Nagios. Reply 3514:Ack all, 3516:Resolv all, 3518:Escal8 all lvl2 (D Gross) 8:26 AM  
PagerDuty: ALRT: 32 open incidents on Nagios. Reply 3524:Ack all, 3526:Resolv all, 3528:Escal8 all lvl2 (D Gross) 8:28 AM  
PagerDuty: ALRT: #5937,#5941,#5986 on Nagios. Reply 3534:Ack all, 3536:Resolv all, 3538:Escal8 all lvl2 (D Gross) 8:33 AM  
PagerDuty: ALRT #5987 on Nagios: Service 'Ping' on host 'Hydra-23' is CRITICAL.. Reply 3544:Ack, 3546:Resolv, 3548:Escal8 lvl2 (D Gross) 8:38 AM  
PagerDuty: ALRT: 8 open incidents on Nagios. Reply 3554:Ack all, 3556:Resolv all, 3558:Escal8 all lvl2 (D Gross) 8:48 AM  
PagerDuty: ALRT #5996 on Nagios: Service 'RAID Health' on host 'Hydra-46' is CRITICAL.. Reply 3564:Ack, 3566:Resolv, 3568:Escal8 lvl2 (D Gross) 8:51 AM  
PagerDuty: ALRT: #5996,#5997 on Nagios. Reply 3574:Ack all, 3576:Resolv all, 3578:Escal8 all lvl2 (D Gross) 8:53 AM  
PagerDuty: ALRT #5905 on Nagios: Service 'GEC stuck errors' on host 'Hydra-49' is CRITICAL.. Reply 3584:Ack, 3586:Resolv, 3588:Escal8 lvl2 (D Gross) 9:32 AM  
PagerDuty: ALRT: #5905,#5999 on Nagios. Reply 3594:Ack all, 3596:Resolv all, 3598:Escal8 all lvl2 (D Gross) 10:04 AM  
PagerDuty: ALRT: #6001,#6002 on Nagios. Reply 3604:Ack all, 3606:Resolv all, 3608:Escal8 all lvl2 (D Gross) 10:08 AM  
PagerDuty: ALRT #5905 on Nagios: Service 'GEC stuck errors' on host 'Hydra-49' is CRITICAL.. Reply 3614:Ack, 3616:Resolv, 3618:Escal8 lvl2 (D Gross) 10:37 AM  
PagerDuty: ALRT #5905 on Nagios: Service 'GEC stuck errors' on host 'Hydra-49' is CRITICAL.. Reply 3624:Ack, 3626:Resolv, 3628:Escal8 lvl2 (D Gross) 11:10 AM  
PagerDuty: ALRT #6005 on Nagios: Service 'Current Load' on host 'localhost' is CRITICAL.. Reply 3634:Ack, 3636:Resolv, 3638:Escal8 lvl2 (D Gross) 11:45 AM  
PagerDuty: ALRT #6006 on Nagios: Service 'Current Load' on host 'localhost' is CRITICAL.. Reply 3644:Ack, 3646:Resolv, 3648:Escal8 lvl2 (D Gross) 12:05 PM  
PagerDuty: ALRT #6007 on Nagios: Service 'CPU Load' on host 'Hydra-5' is CRITICAL.. Reply 3654:Ack, 3656:Resolv, 3658:Escal8 lvl2 (D Gross) 12:16 PM  
PagerDuty: ALRT #6008 on Nagios: Service 'User Capacity' on host 'mysql' is CRITICAL.. Reply 3664:Ack, 3666:Resolv, 3668:Escal8 lvl2 (D Gross) 12:26 PM  
PagerDuty: ALRT #6012 on Nagios: Service 'Current Load' on host 'localhost' is CRITICAL.. Reply 3674:Ack, 3676:Resolv, 3678:Escal8 lvl2 (D Gross) 1:30 PM



# Step II

We came to believe that a  
Power greater than ourselves  
could restore us to sanity.

# Power?



- Zookeeper



- **Nagios®**

- Graphite

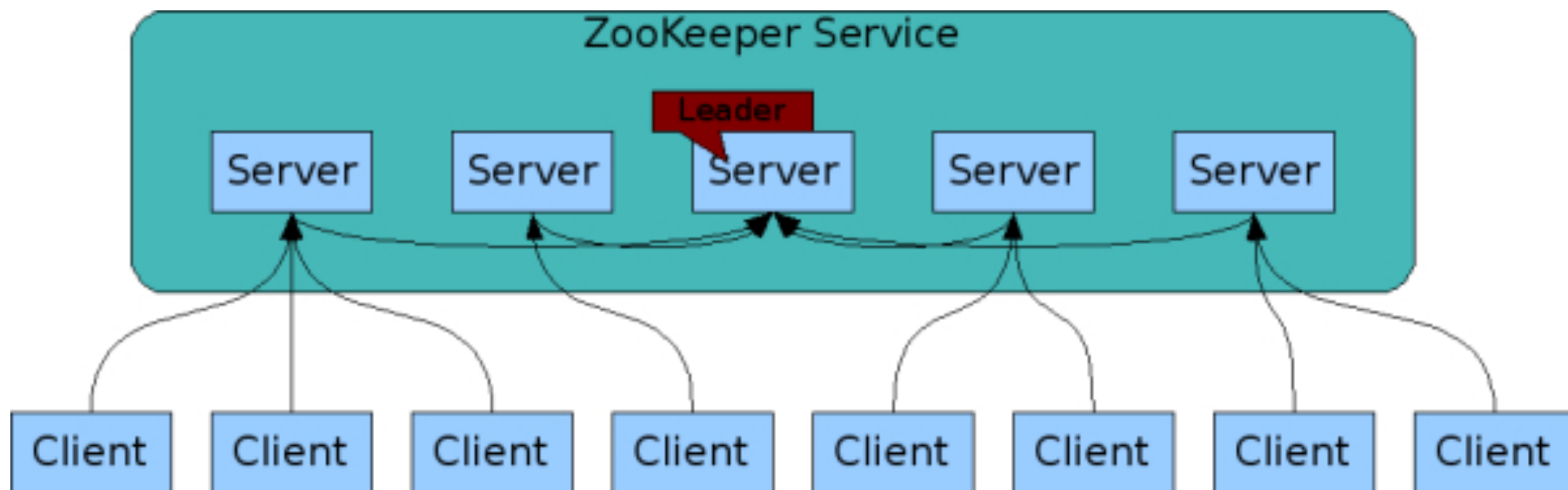
- Yammer Metrics

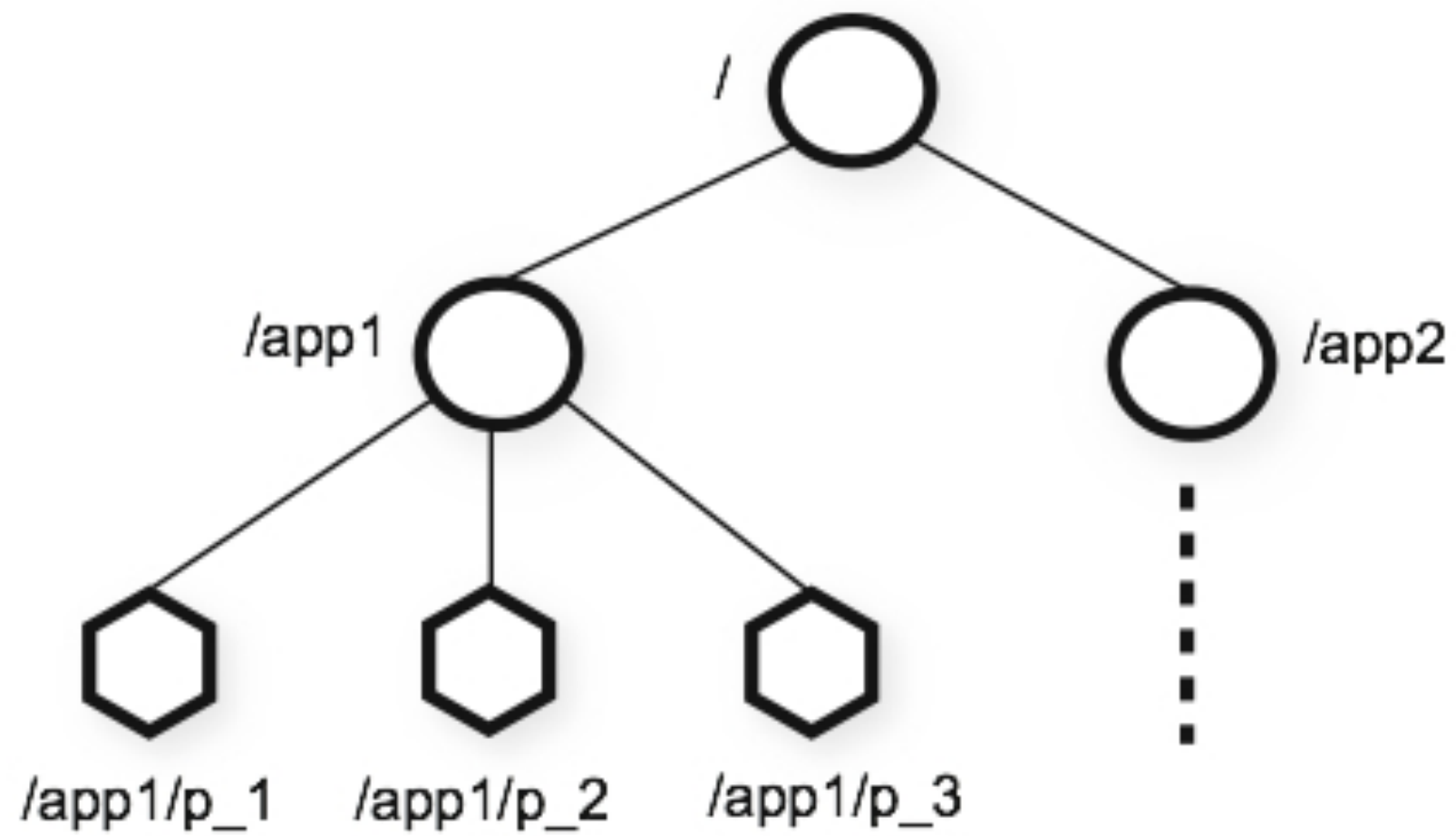
-  **riak**

-  **redis**

# Step III

**We made a decision to turn  
our systems over to the care  
of those Powers**





“A Simple Totally Ordered Broadcast Protocol”

<http://research.yahoo.com/files/ladis08.pdf>

```
final int port = 2182;

final EmbeddedZookeeperServer server =
    EmbeddedZookeeperServer.builder().clientPort(port).build();
final RobustZooKeeper client =
    new RobustZooKeeper("localhost:" + port);

client.withLock("emailerLock",
    new Runnable() {
        @Override
        public void run() {
            System.out.println("Got the distributed lock!");
        }
    });

client.shutdown();
server.shutdown();
```

<https://github.com/Greplin/greplin-zookeeper-utils>

```
from greplin.nagios import parseArgs, parseJsonFile, statValue,
ResponseBuilder
```

```
def check(argv):
```

```
    """Runs the check."""
```

```
    args = parseArgs('check_sample.py',
                      ('HOST', str),
                      ('TIMEOUT', int),
                      argv=argv);
```

```
    data = parseJson(wgetWithTimeout(
                      args['HOST'],
                      8081,
                      '/metricsUrl',
                      args['TIMEOUT']));
```

```
    ResponseBuilder()
```

```
        .addValue('SomeStatToTrack',
                   statValue(data, 'myStat', default = 0))
```

```
        .addRule('OutOfMemoryErrors', Maximum(0, 5),
                  statValue(data, 'jsonKeyOfErrors'))
```

```
        .finish()
```

<https://github.com/Greplin/greplin-nagios-utils>



# Step IV

Made a searching and fearless  
inventory of our systems.

# Metrics!

- Yammer Metrics for JVM
- Greplin Scales for Python
- Greplin Nagios-Utils

```
class Commands {  
  @Timed  
  def search(q: String): Seq[Result] = {  
    // ...  
  }  
  
  @Metered  
  def putDocuments(docs: DocSchema) {  
    // ...  
  }  
}
```

<https://github.com/codahale/metrics/>

```
from greplin import scales
from greplin import graphite
import greplin.scales.flaskhandler as statserver

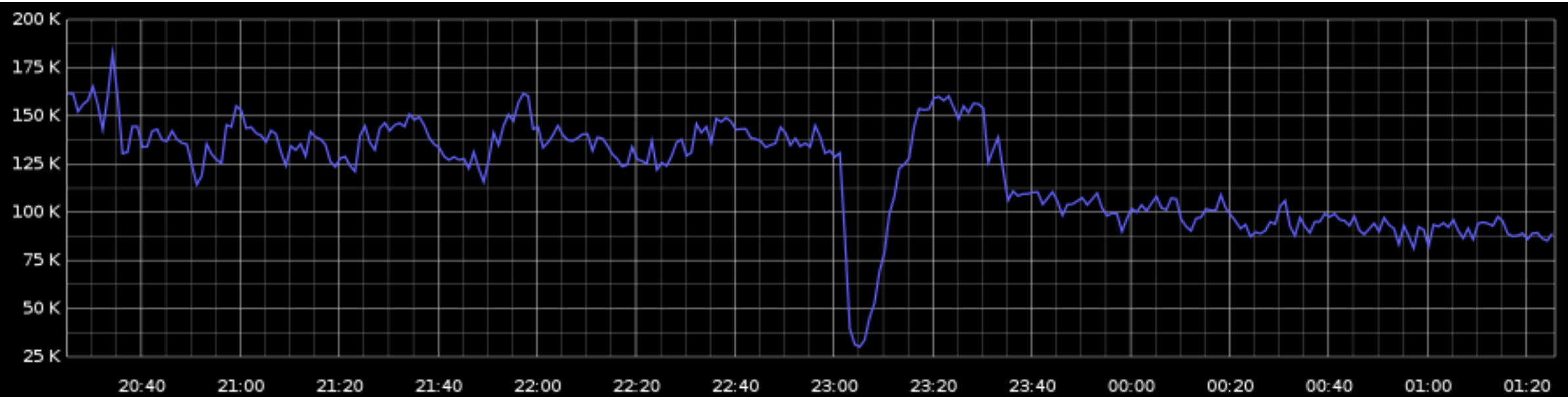
STATS = scales.collection('/web', scales.PmfStat('latency'))

# In a request handler
with STATS.latency.time():
    do_something_expensive()

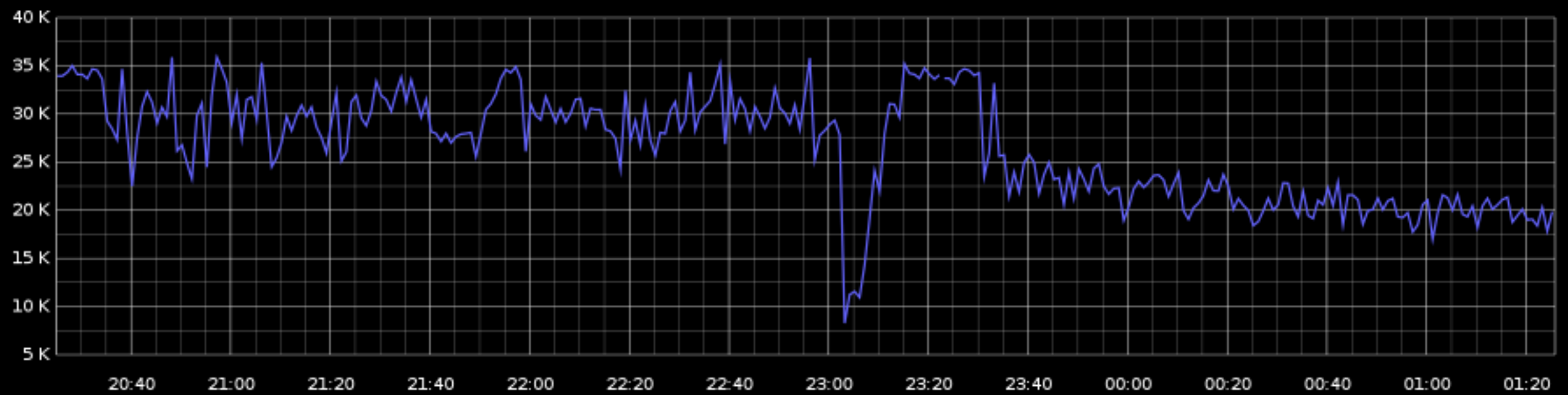
statserver.serveInBackground(8765)
graphite.GraphitePeriodicPusher('graphite-collector', 2003,
'my.server.prefix.').start()
```

<https://github.com/Greplin/scales>

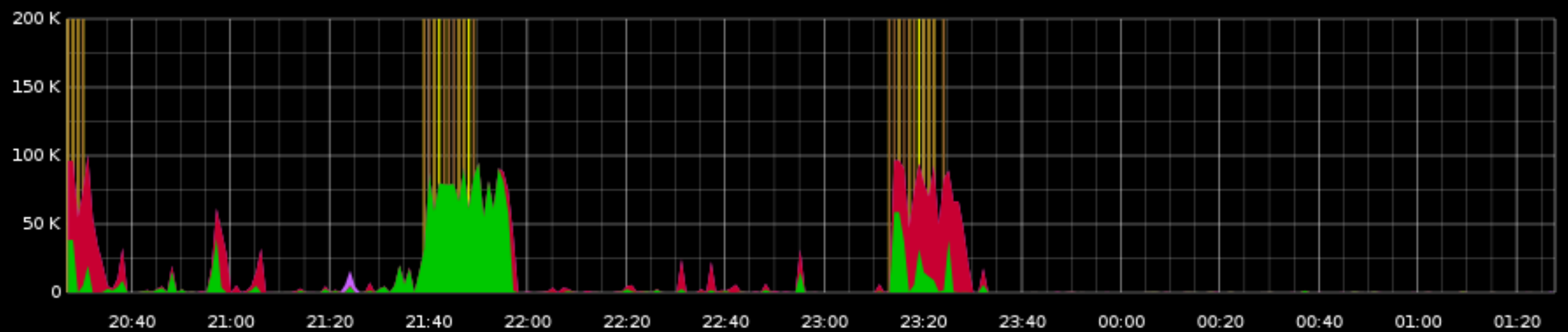
```
"putDocuments": {
  "type": "meter",
  "event_type": "puts",
  "unit": "minutes",
  "count": 4414354,
  "mean": 2316.209153753676,
  "m1": 1330.3353098477928,
  "m5": 1553.2901540398534,
  "m15": 1631.4387098821426
},
"search": {
  "type": "timer",
  "duration": {
    "unit": "milliseconds",
    "min": 34,
    "max": 862,
    "mean": 436.2857142857143,
    "std_dev": 373.0637632209425,
    "median": 443.5,
    "p75": 559,
    "p95": 615,
    "p98": 704,
    "p99": 736,
    "p999": 862
  },
  "rate": {
    "unit": "minutes",
    "count": 700,
    "mean": 27.36730649480839943,
    "m1": 32.133478797150441,
    "m5": 24.688311197707771,
    "m15": 26.76132761570310987
  }
}
```



hydra.all.rates.putDocuments.m1



redis-master.4.systemStats.RedisOps.result



lamp-post.systemStats.PipelineQueueLengths.perfdata.LinusInLength  
 lamp-post.systemStats.PipelineQueueLengths.perfdata.GlassOutLength  
 lamp-post.systemStats.PipelineQueueLengths.perfdata.WidmoreInLength  
 lamp-post.systemStats.PipelineQueueLengths.perfdata.GlassInLength  
 drawAsInfinite(events.pipeline.barriers.globalBackoff.widmore.up)  
 drawAsInfinite(events.pipeline.barriers.globalBackoff.widmore.down)

# Step V



**Admitted to ourselves the  
exact nature of our wrongs.**

# Greplin Exception Catcher

Filters:

Project	Level	Error type	Error message	Count	When	Actions
Hydra	None	java.io.IOException	<b>read past EOF</b> prod Hydra 23	330	4 minutes ago	<a href="#">resolve</a>
Hydra	None	java.io.IOException	<b>read past EOF</b> prod Hydra 23	330	4 minutes ago	<a href="#">resolve</a>
Hydra	None	java.io.IOException	<b>read past EOF</b> prod Hydra 23	149	5 minutes ago	<a href="#">resolve</a>
Hydra	None	java.io.IOException	<b>read past EOF</b> prod Hydra 23	149	5 minutes ago	<a href="#">resolve</a>
Hydra	None	N/A	<b>Unable to acquire flushing lock after 120 seconds (0/3</b> prod Hydra 48, Hydra 49, Hydra 40, Hydra 41, Hydra 42, Hydra 43, Hydra 44, Hydra 45, Hydra 46, Hydra 47, Hydra 19, Hydra 54, Hydra 17, Hydra 16, Hydra 59, Hydra 14, Hydra 13, Hydra 12, Hydra 11, Hydra 53, Hydra 52, Hydra 51, Hydra 50, Hydra 57, Hydra 56, Hydra 55, Hydra 18, Hydra 3, Hydra 2, Hydra 1, Hydra 7, Hydra 6, Hydra 5, Hydra 9,	21346	8 minutes ago	<a href="#">resolve</a>

<https://github.com/Greplin/greplin-exception-catcher>



Host	Service	Status	Last Check	Duration	Attempt	Status Information
<a href="#">Hydra-1</a>	<a href="#">Flush Status</a>	WARNING	2011-09-29 22:11:50	0d 0h 1m 28s	2/4	WARN: FlushActiveTime: 960797ms > 900000ms
<a href="#">Hydra-19</a>	<a href="#">Flush Status</a>	WARNING	2011-09-29 22:11:12	0d 0h 2m 6s	2/4	WARN: FlushActiveTime: 1034166ms > 900000ms
<a href="#">Hydra-22</a>	<a href="#">System Health</a>	WARNING	2011-09-29 22:11:53	0d 0h 0m 25s	1/4	WARN: CpuStats.perfdata.CpuloWait: 84.99 > 80
<a href="#">Hydra-25</a>	<a href="#">Flush Status</a>	WARNING	2011-09-29 22:10:41	0d 0h 4m 37s	4/4	WARN: FlushActiveTime: 1092563ms > 900000ms
<a href="#">Hydra-4</a>	<a href="#">Flush Status</a>	WARNING	2011-09-29 22:11:21	0d 0h 1m 58s	2/4	WARN: FlushActiveTime: 1034618ms > 900000ms
<a href="#">Linus-Worker-1</a>	<a href="#">System Health</a>	CRITICAL	2011-09-29 22:11:38	0d 0h 19m 52s	4/4	CRIT: TotalProcesses.message: 271 > 200
<a href="#">Linus-Worker-2</a>	<a href="#">System Health</a>	CRITICAL	2011-09-29 22:11:42	0d 0h 19m 39s	4/4	CRIT: TotalProcesses.message: 272 > 200
<a href="#">Widmore-2</a>	<a href="#">Status</a>	WARNING	2011-09-29 22:11:53	0d 0h 0m 25s	1/4	WARN: PercentMemoryUsed: 67.7933238% > 65%
<a href="#">[HYDRA2-STAGING]-Hydra-1</a>	<a href="#">Fetch System Status</a>	CRITICAL	2011-09-29 22:11:24	7d 11h 9m 25s	4/4	CRIT - no result after 20 seconds
<a href="#">[STAGING]-Hydra-2</a>	<a href="#">Fetch System Status</a>	CRITICAL	2011-09-13 22:53:33	15d 23h 18m 45s	4/4	CRIT - no result after 20 seconds
	<a href="#">Ping</a>	CRITICAL	2011-09-29 22:07:43	15d 23h 12m 54s	4/4	PING CRITICAL - Packet loss = 100%
<a href="#">backfill-9</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:35	0d 0h 18m 44s	4/4	WARN: BackfillErrors: 2 > 1
<a href="#">basecamp-1</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:14	0d 0h 49m 6s	4/4	WARN: FrontfillErrors: 4 > 1
<a href="#">campfire-1</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:35	0d 0h 18m 44s	4/4	WARN: FrontfillErrors: 11 > 1
<a href="#">dropbox-6</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:28	0d 0h 18m 51s	4/4	WARN: FrontfillErrors: 4 > 1
<a href="#">evernote-1</a>	<a href="#">Check Staff status</a>	CRITICAL	2011-09-29 22:10:28	0d 0h 18m 51s	4/4	CRIT: FrontfillErrors: 684 > 100
<a href="#">facebook-1</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:11:18	0d 23h 10m 21s	4/4	WARN: FrontfillErrors: 19 > 1
<a href="#">facebook-2</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:27	0d 0h 18m 52s	4/4	WARN: FrontfillErrors: 18 > 1
<a href="#">facebook-3</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:29	0d 0h 18m 50s	4/4	WARN: FrontfillErrors: 22 > 1
<a href="#">facebook-4</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:11:22	1d 16h 48m 39s	4/4	WARN: FrontfillErrors: 21 > 1
<a href="#">facebook-5</a>	<a href="#">Check Staff status</a>	WARNING	2011-09-29 22:10:28	0d 0h 18m 51s	4/4	WARN: FrontfillErrors: 21 > 1
	<a href="#">Fetch staff Status</a>	CRITICAL	2011-09-29 22:11:26	0d 0h 0m 52s	1/4	CRIT - no result after 30 seconds

# Step VI

**We made a list of all we had  
harmed, and made direct  
amends by fixing underlying  
problems.**

- **Fixit Friday**
- **3 Day Call**
- **Internal Libraries**
- **GEC monitors all user-facing 500s**

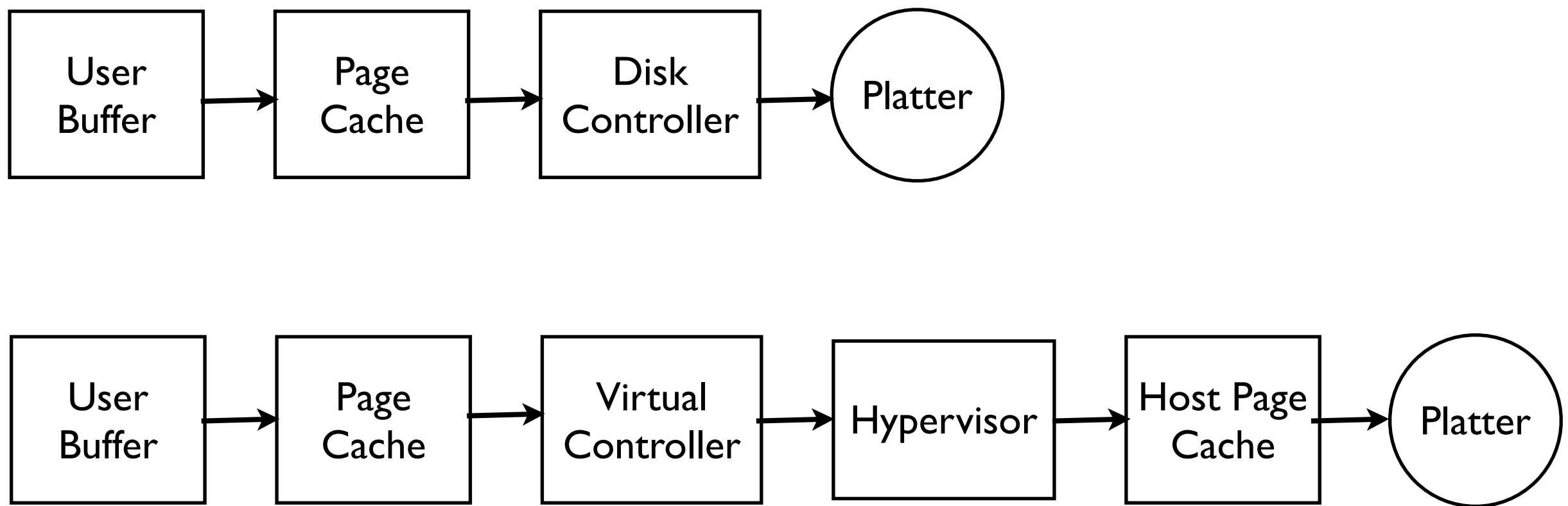
# Bonus!

## Avoid Introducing Accidental Complexity

**DNS == EVIL**



# EBS == EVIL



# Shaneal Manek

@smanek  
shaneal@greplin.com