5 Years of Metrics & Monitoring

Lindsay Holmwood
@auxesis
Cultural & Technical
Key retrospective questions

What did we do well?

What did we learn?

What should we do differently next time?

What still puzzles us?
What got us here won’t get us there
What did we do well?

(that if we don’t talk about, we might forget)
The Pipeline
collection → storage
collection → storage → checking → alerting
collection -> aggregation
storage -> checking -> alerting
graphing
collection & statsd
Graphite & OpenTSDB & InfluxDB
Riemann
Alert fatigue has become a recognised problem.
Cottage industry
PagerDuty & VictorOps & OpsGenie
#monitoringsucks
#monitoring
What would we do differently next time?
Graphs & Dashboards
Apparently the hardest problem in monitoring is graphing and dashboarding.
What we’re doing wrong
Strip charts
We have a problem
Strip charts: the PHP hammer of graphing
What can the data tell us?
What is the distribution?
It’s not a problem with the tools
Our approach is tainted
graphing problems we have

graphing problems serviced by strip charts
Basic graph layout
Black on white
bounding box with x + y axes labels
Colour
Differential colour engine
Maximum of 15 colours on-screen
Adjust saturation, not hue
Use minimal hue to call out data
Fucking Pie Charts
Choose appropriate scales when data on different panels are compared.

Figure 2.54 shows data from an experiment on graphical perception [34]. A group of 51 subjects judged 40 pairs of values on bar charts and the same 40 pairs on pie charts; each judgment consisted of studying the two values and visually judging what percent the smaller was of the larger. The top panel of Figure 2.54 shows the 40 average judgment errors (averaged across subjects) graphed against the true percents for the 40 pie chart judgments. The bottom panel shows the same variables for the bar chart judgments. To enhance the comparison of the bar chart and pie chart values, the scales on the two panels are the same; this allows us to see very clearly that the pie chart judgments are less accurate than the bar chart judgments. One result of the common scale is that the data do not fill either panel; we should always be prepared to forego the fill principle to achieve an effective comparison. But note that if all of the data were put on one of the panels, the data rectangle would nearly fill the scale-line rectangle.

Using the same scales in the two panels of Figure 2.54 allows a number of quantitative comparisons to be made of the two sets of data. We can compare the average level of the absolute error of each chart type for each true percent. For example, we can see that the average level is about the same for true percents less than about 35%, but generally for percents greater than 35%, the average level is greater for pie charts. Furthermore, we can compare the variation in the errors for the two chart types. There is greater change in the pie chart errors than in the bar chart errors as the true percent changes; the pie chart errors increase but the bar chart errors have a flat pattern. The same scales can be used in this example, in part, because the overall levels of the two sets of data are not radically different.
Experiment:

Compare segment sizes
This allows us to see very clearly that the pie chart judgements are less accurate than the bar chart judgements.

– William S. Cleveland, p.86 Principles of Graphing Data
Pie chart comparisons are more error prone
The only time you should use a pie chart
What did we learn?
Democratisation of graphing tool development
Scratch our itches
Same poor UX,
better paint job
We get the graphing tools we deserve
Nagios is here to stay (at least for ops)
Inertia
No

strong, compelling

alternative
Sensu
When I hear people say

“I'm not using Sensu because it's too complex”

I think

“and Nagios isn't hiding the same complexity from you?”
This is a problem
We don’t know stats
Numbers &
Strings &
Behaviour
Fault detection
(thresholding)
Anomaly detection
(trend analysis)
Monitoring is CI for Production
Continuous Integration
Continuous Integration

1. checkout
Continuous Integration

1. checkout

2. build
Continuous Integration

1. checkout
2. build
3. test
Continuous Integration

1. checkout
2. build
3. test
4. notify
Continuous Integration

Monitoring

1. checkout (crossed out)
2. build (crossed out)
3. test
4. notify
Monitoring

1. checkout
2. build
3. test
4. notify

can I see my app?
serverspec & sensu
What still puzzles us?
(or, what might the future look like?)
The future is analysing & acting on our alert data
Last 5 years

Building new tools

Formalising relationships

Search for parallels in other industries

Measuring the human impact
Next

Stabilisation of tools

Emerging standards

Exploiting parallels

Mitigating the human impact
Analysis:
Ops Weekly
## Update for week ending Sunday 15th June 2014

### On call report

#### Alerts received this week (Friday 6th June 2014 - Friday 13th June 2014)

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Host</th>
<th>Service</th>
<th>Output</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fri 13 Jun 14:04:29 EDT</td>
<td>logarchive02</td>
<td>Disk Space</td>
<td>DISK WARNING - free space: /logs 6616027 MB (10% inode=99%)</td>
<td>WARNING</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No action taken: Threshold adjuster</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Notes: Threshold is too low, increased</td>
<td></td>
</tr>
<tr>
<td>Fri 13 Jun 04:51:31 EDT</td>
<td>virt14</td>
<td>Disk Space</td>
<td>DISK CRITICAL - /var is not accessible: Input/output error</td>
<td>CRITICAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Action taken: Service Issue (View cli: )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Notes: RAID failure caused machine to die</td>
<td></td>
</tr>
<tr>
<td>Thu 12 Jun 20:36:19 EDT</td>
<td>localhost</td>
<td>Aggregate MySQL Slave</td>
<td>OK=46 WARNING=0 CRITICAL=1 UNKNOWN=0 services=/MySL Slave/ host=/~db(shard</td>
<td>CRITICAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Action taken: Service Issue (View cli: )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Notes: MySQL slave failure on host</td>
<td></td>
</tr>
<tr>
<td>Thu 12 Jun 19:12:43 EDT</td>
<td>database001b</td>
<td>Host Check</td>
<td>(Host Check Timed Out)</td>
<td>DOWN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Action taken: Service Issue (View cli: )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Notes: Machine died, hardware failure</td>
<td></td>
</tr>
<tr>
<td>Wed 11 Jun 12:57:21 EDT</td>
<td>api05</td>
<td>Memory</td>
<td>CHECK_NRPE: Socket timeout after 30 seconds.</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## 55 alerts handled on 2014-10-12

<table>
<thead>
<tr>
<th>#</th>
<th>Incident</th>
<th>Description</th>
<th>Time</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>PJ9Y4LE</td>
<td>Problem: &quot;PING&quot; on guildlink-dvmh-web-02 is Warning, PING WARNING - Packet loss = 28%, RTA = 6.72 ms</td>
<td>2014-10-12 22:33:46 +1100</td>
<td>recovered without int</td>
</tr>
<tr>
<td>3</td>
<td>P3PK4UF</td>
<td>Problem: &quot;DISK F: Utilisation&quot; on bslhyp-sqitst01 is Warning, WARNING: 92.30% disk used (f: 16.9GB, u: 203.1GB) (should be &lt; 90.0%)</td>
<td>2014-10-12 22:06:26 +1100</td>
<td>real</td>
</tr>
<tr>
<td>4</td>
<td>P6BU3S1</td>
<td>Problem: &quot;HTTP Response Time&quot; on milan-avmh-proxy-01 is Critical, CRITICAL - Socket timeout after 16 seconds</td>
<td>2014-10-12 20:29:46 +1100</td>
<td>not set</td>
</tr>
<tr>
<td>5</td>
<td>PVLUIMC</td>
<td>Minimum FreeStorageSpace LessThanThreshold 1.073741824E9 for DBInstanceIdentifier masonlinestage</td>
<td>2014-10-12 19:59:47 +1100</td>
<td>not set</td>
</tr>
<tr>
<td>6</td>
<td>PGGR5L3F</td>
<td>Minimum FreeStorageSpace LessThanThreshold 2.147483648E9 for DBInstanceIdentifier masonlinestage</td>
<td>2014-10-12 19:59:46 +1100</td>
<td>not set</td>
</tr>
</tbody>
</table>
Context:
Nagios Herald
** PROBLEM Service ALERT: devking02.ny4dev/Disk Space is WARNING **

PROBLEM
Host: devking02.ny4dev
Service: Disk Space

State is now: WARNING for 0d 0h 2m 1s (was WARNING) after 3 / 3 checks

Threshold exceeded

Problematic volume

Ganglia graph

** THRESHOLDS - WARNING:10%;CRITICAL:5%; **

Filesystem Size Used Avail Use% Mounted on
/dev/sda2 29G 17G 11G 62% /
tmpfs 12G 0 12G 0% /dev/shm
/dev/sda1 194M 79M 106M 43% /boot
/dev/sda4 516G 462G 54G 90% /var  Free disk space (10%) is <= WARNING threshold (10%).
/dev/mapper/VolGrp_homes01-homes01
3.9T 1.5T 2.3T 39% /homes01
/dev/mapper/VolGrp_Deepersrl-Deepersrl
9.8T 4.4T 5.5T 45% /mnt/deepersrl
/dev/mapper/VolGrp_github-github
2.9T 569G 2.4T 20% /mnt/github
The future is richer metadata about our metrics
Metrics 2.0
{
    server: dfs1
    what: diskspace
    mountpoint: srv/node/dfs10
    unit: B
    type: used
    metric_type: gauge
}

meta: {
    agent: diamond,
    processed_by: statsd2
}
Self-describing
The future is richer metadata about our metrics
The future is richer metadata about our metrics to automatically build appropriate visualisations.
Aggregation &
Grouping &
Unit conversions &
Scaling &
Axes labelling &

...
Death to strip charts
The future is monitoring tools for devs
Ops must be enablers, not gatekeepers.
What has made sense about ops being gatekeepers?
Monitoring is treated as an operational responsibility
Ops team

own ops
We’ve won the battles
Ops team

own ops
This is no longer the world we live in
How do we become enablers?
Technical & Cultural
Technical
Technical

Ops provide the platform
Technical

Ops provide the platform

Maintain, monitor, and scale the platform
Rule #4: Monitoring systems need to be more available and scalable than the systems being monitored.

— Adrian Cockcroft
Cultural
Cultural

Coach on what makes a good check

Coach on what is good alert design

Listen to the needs of the end-user
Provide monitoring as a service
Monitoring is a core deliverable on every service
Ship checks & config with your applications
Example: Yelp
Situational awareness

Don't Panic! Ops already knows! Failed Sensu checks on this host:

Crit: crong_iproute2.git_commit_staleness: FILE_AGE CRITICAL: /nail/run/su...

Last login: Sat Sep 20 04:46:15 2014 from 10.250.104.14

relengsrv3-devc - 🌟 🌟 🌟
What’s the barrier to entry?
Does the idea just not have traction?
Are the tools not up to scratch?
Does monitoring need to be SaaS (or SaaS-like) to make this achievable at scale?
The future is here – it’s just not very evenly distributed

– William Gibson
Monitoring is still insular
We’re building tools for operations teams
Not the developers who need them most
Monitoring is like a joke.
Monitoring is like a joke.
If you have to explain it, it’s not that good.
aggregation

collection  storage  checking  alerting

graphing
What can we do better?
I’m Lindsay

@auxesis
Dank je wel!
Dank je wel!

Liked the talk? Let @auxesisis know.