Streamlining Monitoring Infrastructure in IT-DB-IMS

Charles Newey
charlie@assemblyco.de
Part 1: Log Monitoring

Log Monitoring:

Getting, processing, analysing, visualising, and reacting to information in log files

...at scale?
Part 1: Log Monitoring

- Proprietary?
  - Splunk

- Open-source?
  - Elasticsearch
  - Logstash
  - Kibana

Also known as the "ELK stack"
Part 1: Log Monitoring

- Elasticsearch
  - High-performance scalable search engine
- Logstash
  - Log transport and processing daemon
- Logstash-Forwarder
  - Lightweight log shipper
- Kibana
  - Visualisation dashboard for Elasticsearch

Also known as the “ELK stack”
I’d never done anything with Puppet before
Puppet code is very easy to write
... badly

“The moment that Puppet goes wrong”
Part 2:

Metric Collection and Visualisation

› Need to store and visualise metrics for many machines

› What needs to be measured?
   › Load average (1, 5, 15 min)
   › CPU temperature
   › Network connectivity (latency between gateways, etc.)
   › ... and lots more

› OpenTSDB!
   › A time-series database running on top of HBase
   › Data replicated 3 times (inside HDFS)
Part 2: Metric Collection and Visualisation

- Metric collection agents?
  - tcollector (Python)
  - scollector (Go)
- Building and distributing RPMs with Koji
- Visualising with GNUPlot and Grafana

Charles Newey (charlie@assemblyco.de)
Part 3: WebLogic Log Analysis

› WebLogic’s logging architecture isn’t designed well...
› Parsing WebLogic log files is difficult

› What are the implications of this?
› Lots of regular expressions!

Source: XKCD
Part 3: WebLogic Log Analysis

It was a challenge getting Logstash to parse WebLogic logs...

› If you want logs to be readable and useful:
   › Don’t use several different formats for log messages
   › Don’t use several different (localised) formats for timestamps
   › Don’t make several different processes log to the same file
   › Don’t write multiline Java stack traces to a file... why?
     › Line-oriented tools won’t work (sed/awk/grep/etc)
     › Multiline logs *must* be parsed by a single Logstash thread
Part 3: WebLogic Log Analysis

› Most importantly...
› Don’t do all of these things at once
WebLogic Exceptions Dashboard

Charles Newey (charlie@assemblyco.de)
Impact

› Log files are searchable, easy to visualise and analyse

› Can analyse downtime, traffic spikes, load spikes, etc

› System metrics can be collected and stored indefinitely on HDFS for search, visualisation and diagnosis

› WebLogic exceptions can be pinpointed, searched and visualised in NRT, making it easy to report to developers

Charles Newey (charlie@assemblyco.de)