Lessons from Automatic Incident Resolution for a Million Databases

SRECon EU, July 2016

Greg Burek
The Twelve-Factor App
~ Million Databases

Tens of thousands of AWS Instances
Some Databases

Hundreds of AWS Instances
“The goal is to build systems that can scale linearly with machines & sub-linearly with people” - Caitie McCaffrey
Monitor and alert on your business
Me: oh sorry, I got paged
Date: are you a doctor?
Me: uh
Nagios: holy SHIT this cert expires in SIXTY DAYS
Me: ...yes
Monitor and alert on your business

Usually, don’t alert on machine specific metrics
Write runbooks and playbooks
Turn playbooks into code
Stop writing playbooks and start writing scripts. If the steps to respond to an alert can be written down they can also be automated.

Jul 5, 2016, 1:58 PM

26 RETWEETS  36 LIKES
“The goal is not to never get paged, the goal is to never get paged for the same thing twice”
- Astrid Atkinson
Verify monitoring before restarting the world
Circuit breakers
Automation can’t handle the unknown

Wake someone up on exceptions and timeouts
Have a REPL/console
Aggregate and review trends
### Trusty Kernel Panics in Last Week

No Rows Returned

#### Stale Incidents (over 2 days old)

<table>
<thead>
<tr>
<th>type</th>
<th>state</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>old_wale_base_backup</td>
<td>await_resolution</td>
<td></td>
</tr>
<tr>
<td>resource_down</td>
<td>ignored</td>
<td></td>
</tr>
</tbody>
</table>

#### Total Incidents

<table>
<thead>
<tr>
<th>type</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>follower_behind</td>
<td></td>
</tr>
<tr>
<td>old_wale_base_backup</td>
<td></td>
</tr>
<tr>
<td>blocking_queries</td>
<td></td>
</tr>
<tr>
<td>resource_down</td>
<td></td>
</tr>
<tr>
<td>changeover</td>
<td></td>
</tr>
<tr>
<td>server_down</td>
<td></td>
</tr>
</tbody>
</table>
Humans can break

Automation can be simplistic

Humans + Automation for a resilient and operable system
1. Monitor and alert on your business
2. Write playbooks
3. Make playbooks into automation
4. Checks and balances of automation
5. Circuit breakers
6. Alert on exceptions and timeouts
7. Admin console
8. Aggregate and review trends
I'll do my best!

gregburek@heroku.com
@gregburek
State Machines