Application AutoScaling

WATCH MY INSTANCES GROW AND SHRINK

Michael Fraenkel
Bo Yang
Dies Koper

Cloud Foundry Summit 2016
May 9, 2016
Current Status

• Incubating in cloudfoundry-incubator

• Distributed development team
  • IBM (China, US)
  • Fujitsu (Australia)
  • SAP (India)
Without AutoScaling

• cf scale MYAPP –i 3

• cf scale MYAPP –i 1
With AutoScaling

- Create a service instance
- Define your policy
- Bind the service

```bash
cf bind-service myapp myscaleinstance
   -c '{"type":"memory","upper-threshold":"80%","increment":"2","max":"12"}'
```
Memory based scaling

Max. instances

Min. instances

Memory

Instances
Scheduled scaling

- Time: 19:00 to 23:00
- Max. instances: 200
- Min. instances: 50

Diagram showing instances over time.
Scaling in/out

- Max. instances
- Scheduled Min. instances
- Default Min. instances
Dynamic Policy

Default Instance Limits

Minimum Instance Count: 1
Maximum Instance Count: 5

Scaling Rule(s)

Rule 1
Add 1 instance(s) if average Memory utilization exceeds 80% for 600 seconds.
Remove 1 instance(s) if average Memory utilization is below 30% for 600 seconds.

Metric Type: Memory

Scale Out: If average Memory utilization exceeds 80 %, then increase 1 instance(s).
Scale In: If average Memory utilization is below 30 %, then decrease 1 instance(s).
### Schedule Policy - Repeating

**Time Zone:** (GMT -07:00) US/Pacific

#### Recurring Schedules

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Repeat On</th>
<th>Minimum Instance Count</th>
<th>Maximum Instance Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>18:00</td>
<td>Mon Tue Wed Thr</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>
Schedule Policy - Fixed

Specific Date(s)

The specific date schedule takes precedence of the recurring schedule if there is overlap.

<table>
<thead>
<tr>
<th>Start Date &amp; Time</th>
<th>End Date &amp; Time</th>
<th>Minimum Instance Count</th>
<th>Maximum Instance Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/23/2016 07:00</td>
<td>5/26/2016 23:59</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
# AutoScaling: scaling history

<table>
<thead>
<tr>
<th>Configuration</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select Application:</td>
<td>sampleApp</td>
</tr>
</tbody>
</table>

## Past Week

<table>
<thead>
<tr>
<th>Status</th>
<th>Start Time</th>
<th>Duration</th>
<th>Event</th>
<th>Adjustment</th>
<th>Total instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed</td>
<td>Wednesday, April 09, 2014 2:59:04 PM</td>
<td>1 seconds 98 milliseconds</td>
<td>CPU Utilization lower than 30% for 60 seconds.</td>
<td>Remove 1 instance.</td>
<td>1</td>
</tr>
<tr>
<td>Completed</td>
<td>Tuesday, March 25, 2014 5:17:40 PM</td>
<td>5 seconds 147 milliseconds</td>
<td>CPU Utilization lower than 30% for 60 seconds.</td>
<td>Remove 1 instance.</td>
<td>1</td>
</tr>
<tr>
<td>Completed</td>
<td>Tuesday, March 25, 2014 5:11:40 PM</td>
<td>44 seconds 134 milliseconds</td>
<td>CPU Utilization lower than 30% for 60 seconds.</td>
<td>Remove 1 instance.</td>
<td>2</td>
</tr>
<tr>
<td>Completed</td>
<td>Monday, March 24, 2014 5:00:14 PM</td>
<td>1 seconds 508 milliseconds</td>
<td>CPU Utilization lower than 30% for 60 seconds.</td>
<td>Remove 1 instance.</td>
<td>1</td>
</tr>
</tbody>
</table>
Things to come

• Response time

• Custom metrics
  • Heap
  • Throughput

• CPU?
Thank you

• Questions?
  • @autoscaler on cloudfoundry.slack.com
    https://cloudfoundry.slack.com/messages/autoscaler/

  • Email: cf-dev@lists.cloudfoundry.org