You get what you measure—why metrics are important

Kumar Srinivasamurthy
Bing & Cortana Engineering
@00kumars
Examples
Metric definition, education and reliability of monitors


Availability & Downtime

Simplified formula:

\[
\text{Availability} = \frac{(\text{Total Requests} - \text{Failed Requests})}{\text{Total Requests}} \times 100\% \\
A = \frac{T-F}{T} \times 100\%
\]

<table>
<thead>
<tr>
<th>Availability %</th>
<th>Downtime per year</th>
<th>Downtime per month</th>
<th>Downtime per week</th>
<th>Downtime per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% (&quot;one nine&quot;)</td>
<td>36.5 days</td>
<td>72 hours</td>
<td>16.8 hours</td>
<td>2.4 hours</td>
</tr>
<tr>
<td>95%</td>
<td>18.25 days</td>
<td>38 hours</td>
<td>8.4 hours</td>
<td>1.2 hours</td>
</tr>
<tr>
<td>97%</td>
<td>10.56 days</td>
<td>21.6 hours</td>
<td>5.04 hours</td>
<td>43.2 minutes</td>
</tr>
<tr>
<td>98%</td>
<td>7.30 days</td>
<td>14.4 hours</td>
<td>3.36 hours</td>
<td>28.8 minutes</td>
</tr>
<tr>
<td>99% (&quot;two nines&quot;)</td>
<td>3.65 days</td>
<td>7.20 hours</td>
<td>1.68 hours</td>
<td>14.4 minutes</td>
</tr>
<tr>
<td>99.5%</td>
<td>1.83 days</td>
<td>3.60 hours</td>
<td>50.4 minutes</td>
<td>7.2 minutes</td>
</tr>
<tr>
<td>99.8%</td>
<td>17.52 hours</td>
<td>86.23 minutes</td>
<td>20.16 minutes</td>
<td>2.88 minutes</td>
</tr>
<tr>
<td>99.9% (&quot;three nines&quot;)</td>
<td>8.76 hours</td>
<td>43.8 minutes</td>
<td>10.1 minutes</td>
<td>1.44 minutes</td>
</tr>
<tr>
<td>99.95%</td>
<td>4.38 hours</td>
<td>21.56 minutes</td>
<td>5.04 minutes</td>
<td>43.2 seconds</td>
</tr>
<tr>
<td>99.99% (&quot;four nines&quot;)</td>
<td>52.56 minutes</td>
<td>4.33 minutes</td>
<td>1.01 minutes</td>
<td>8.66 seconds</td>
</tr>
<tr>
<td>99.995%</td>
<td>26.28 minutes</td>
<td>2.16 minutes</td>
<td>30.24 seconds</td>
<td>4.32 seconds</td>
</tr>
<tr>
<td>99.999% (&quot;five nines&quot;)</td>
<td>5.26 minutes</td>
<td>25.9 seconds</td>
<td>6.05 seconds</td>
<td>864.3 milliseconds</td>
</tr>
<tr>
<td>99.9999% (&quot;six nines&quot;)</td>
<td>31.5 seconds</td>
<td>2.59 seconds</td>
<td>604.8 milliseconds</td>
<td>86.4 milliseconds</td>
</tr>
<tr>
<td>99.99999% (&quot;seven nines&quot;)</td>
<td>3.15 seconds</td>
<td>262.97 milliseconds</td>
<td>60.48 milliseconds</td>
<td>8.64 milliseconds</td>
</tr>
<tr>
<td>99.999999% (&quot;eight nines&quot;)</td>
<td>315.569 milliseconds</td>
<td>26.297 milliseconds</td>
<td>60.48 milliseconds</td>
<td>0.964 milliseconds</td>
</tr>
<tr>
<td>99.9999999% (&quot;nine nines&quot;)</td>
<td>31.5569 milliseconds</td>
<td>2.6297 milliseconds</td>
<td>60.48 milliseconds</td>
<td>0.0864 milliseconds</td>
</tr>
</tbody>
</table>

https://en.wikipedia.org/wiki/High_availability
Hitting the target and missing the point

\[ A = \frac{T-F}{T} \% \]

- Bad config led to no web results to the user, just ads
- Remember that one metric may not tell the whole story
- Opposing Metrics
How and where you measure matters

\[ A = \frac{T-F}{T} \% \]

- Traffic black-holing - Drop at CDN/Edge
- High availability at service level, but drop in incoming volume
- *Anomaly detection* can identify changes in volume
- Outside-in monitoring should supplement internal metrics
- Incorrect sample sizes can also affect your metric
Gaps in measurement

- Terrible slowness for a small segment of users
- PLT tracked at 75\textsuperscript{th} or 95\textsuperscript{th} didn’t show issues
- Investigation and findings led to additional dimensions
Performance measurement journey

- Periodic update of metrics
- Above fold vs below fold
The great Hanoi rat massacre

https://flic.kr/p/ovusTD

https://www.atlasobscura.com/articles/hanoi-rat-massacre-1902
Unintended consequences - Window tax
Closing thoughts

• High data quality
  • Correctness, completeness and cleanliness
• Periodic metrics update
  • Metrics do get “gamed”
• Wrong incentives – unintended consequences
• Metrics overload
• Percentiles can hide problems
• Some metrics do flat-line. Important to let them not regress
Questions?

• kumars@microsoft.com
• @00kumars