Availability Table

Availability is generally calculated based on how long a service was unavailable over some period. Assuming no planned downtime, Table 1-1 indicates how much downtime is permitted to reach a given availability level.

Table 1-1. Availability table

Availability level	Allowed unavailability window					
	per year	per quarter	per month	per week	per day	per hour
90%	36.5 days	9 days	3 days	16.8 hours	2.4 hours	6 minutes
95%	18.25 days	4.5 days	1.5 days	8.4 hours	1.2 hours	3 minutes
99%	3.65 days	21.6 hours	7.2 hours	1.68 hours	14.4 minutes	36 seconds

Availability level	Allowed unavailability window					
99.5%	1.83 days	10.8 hours	3.6 hours	50.4 minutes	7.20 minutes	18 seconds
99.9%	8.76 hours	2.16 hours	43.2 minutes	10.1 minutes	1.44 minutes	3.6 seconds
99.95%	4.38 hours	1.08 hours	21.6 minutes	5.04 minutes	43.2 seconds	1.8 seconds
99.99%	52.6 minutes	12.96 minutes	4.32 minutes	60.5 seconds	8.64 seconds	0.36 seconds
99.999%	5.26 minutes	1.30 minutes	25.9 seconds	6.05 seconds	0.87 seconds	0.04 seconds

Using an aggregate unavailability metric (i.e., "X% of all operations failed") is more useful than focusing on outage lengths for services that may be partially available—for instance, due to having multiple replicas, only some of which are unavailable—and for services whose load varies over the course of a day or week rather than remaining constant.

See Equations Time-based availability and Aggregate availability in Embracing Risk for calculations.

PREVIOUS
Chapter 34 - Conclusion

NEXT
Appendix B - A Collection of Best Practices for Production Services