

SRE: The Cost of Failure



It's better to avoid a problem when given the chance, but when that is not possible try to minimize the impact.

The cost of failure can mean many different things, to some the cost is reflected by average cost per minute (CPM) multiplied by the average MTTR, loss of revenue or profit, regulatory risk/fines/scrutiny, SLA violation, or reputational risk or maybe just the overall embarrassment of being paid and trusted to provide a service and failing to do so. What is the failure in this context? The service is going down, a feature not working etc.

In every customer journey this question should be answered: What is the cost of failure? Literally what happens if this app is down, or the customer cannot do X? What does it cost us in dollars, in reputation? Will our user's comeback or switch services? What constitutes a failure? Does our availability make sense compared to the amount of financial risk we consider acceptable? What is our Error Budget Minutes x CPM (this is our projected failure cost for the year)? How much money are we willing to risk losing every year? Does our availability target reflect our risk appetite?

Investing in reliability has business benefits to the bottom-line. However, we need to ensure that engineers are working on the right thing. If we invest in Testing, Monitoring, Automation, Capacity, Engineering, RCA, Incident Management what do we get back? Compared to the projected profit gains of the new feature, is the gain of reliability higher? Is the ROI of doing it higher than the risk of not doing it? Which investments lower our CPM or increase our availability?

Example: We have a social media application that generates 400 million in profit. Each minute the app is down costs us \$761 per minute. We have set an SLO of 99% or an error budget of 5256 minutes (about 3 and a half days a year). We have projected yearly loss of \$3,999,816 from application downtimes. We have 1 feature release planned, feature A will increase profit by 100 million per year (\$951 cost per minute– increasing COF to 4.9M, 8-9M+ per month for every month after) and requires 9 months to develop. Feature B an investment in sharding to increase reliability would take 1 month to develop but will decrease the CPM by 50% (330K per month pre-A, post A 408K per month, post AB per 204K).

So below clearly the new feature should supersede the reliability feature. Now if we begin to consider the monetary investments in the feature you can see how the complexity grows or if we press the levers of profitability that consequently affect the CPM in this case.

A, B

P (M)	33	33	33	33	33	33	33	33	33	42	42	42	456 M
L (K)	330	330	330	330	330	330	330	330	330	408	204	204	3.7 M

B, A

P (M)	33	33	33	33	33	33	33	33	33	33	42	42	414M
L (K)	330	165	165	165	165	165	165	165	165	165	204	204	2.2M