fasthosts ProActive

Auto-scaling - the downsides

Can you avoid spiralling costs?

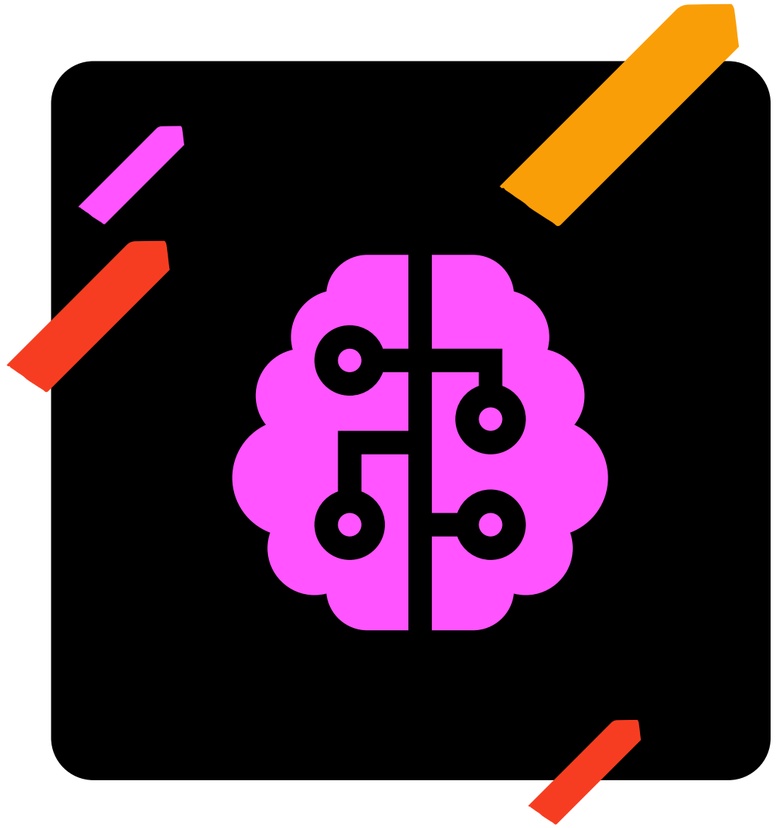
In a world where businesses are focused on efficiency and cost savings, auto-scaling of your SaaS applications sounds like a no-brainer. When you choose this option, it feels like you have total control of the resources you need for your operations to run smoothly. But is this strictly true?

Letting your software scale automatically—adding or removing resources as needed—may seem like a great way to ensure SaaS availability for your customers.

But there are some serious downsides to auto-scaling.

Because its strength—automation—is also a vulnerability. Either by accident or malicious intent, it can expose you to unexpected and uncontrolled costs.

Here, we explore some of the problems that could arise when using auto-scaling with your SaaS solution.



Contents

1 - What can go wrong with SaaS auto-scaling?

2 - Memory leaks and auto-scaling

3 - Logs and auto-scaling

4 - Malicious attacks and auto-scaling

Avoid unforeseen auto-scaling issues and spiralling costs by using an expert managed services provider (MSP) with the knowledge to make your infrastructure and solution as cost-efficient as possible.

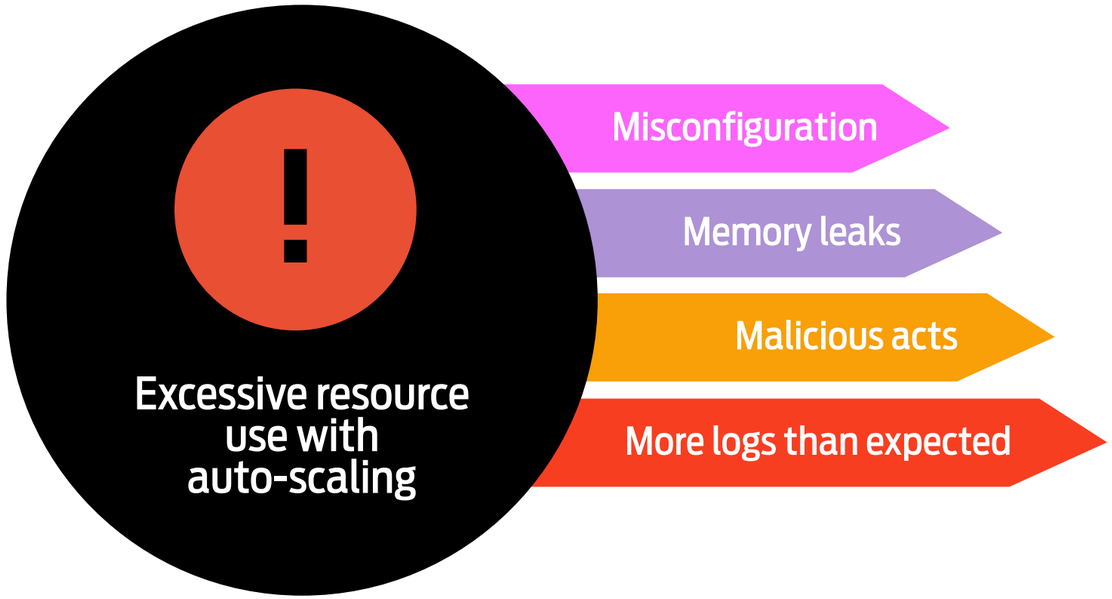
1 What can go wrong with SaaS auto-scaling?

With an underlying business model that's acutely sensitive to availability, overprovisioning can reduce your profitability, while underprovisioning can lose you customers—so for SaaS businesses, auto-scaling seems like an attractive option. But it's important to understand the downsides—because if your costs automatically increase dramatically without bringing in the corresponding ROI, the viability of your business could be at stake.

How can costs spiral?

Unexpected escalation of resource use can have various causes, from inadvertent misconfiguration to cyberattack. But while the causes may be varied, what they have in common is the ability to trigger the cloud provider's auto-scaling mechanism. So, if the issue is not identified in time, costs can quickly spiral out of control.

Causes of unplanned auto-scaling escalation



2 Memory leaks and auto-scaling

Memory leaks occur if memory is not released to the system after a program has finished using it. For a SaaS business, this can affect efficiency and prove costly.

If your application has a memory leak, it may consume extra:

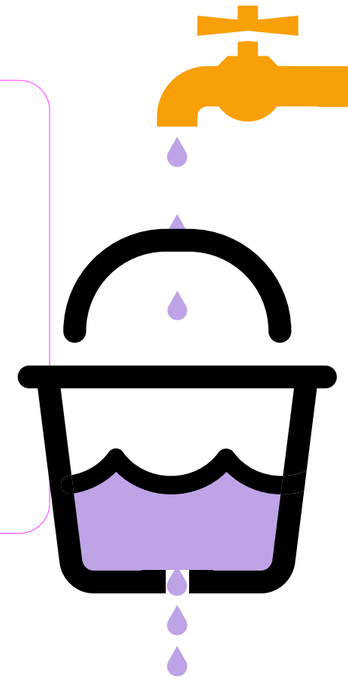
- Memory
- Cache
- Swap space

As this unreleased memory accumulates, it can be detrimental to your application's functionality, affecting performance and response times. It means you use more resources, but not through the additional user requests you want for business growth. Left unchecked, your solution's auto-scaling could kick in, and your costs creep up.

To prevent damaging memory leaks, use an expert MSP to monitor your infrastructure performance and flag up atypical memory usage patterns, even where it builds up gradually. This will enable you to fix leaky code before it causes excessive auto-scaling.

Use an expert MSP to monitor inftrastructure performance and alert you to abnormal usage trends.

Prevent unnecessary autoscaling due to overallocation of memory



3 Logs and auto-scaling

Any IT solution needs to create logs for efficient monitoring and operation. The problem comes if the number and size of logs becomes excessive. Efficient housekeeping of log files is essential to avoid log files clogging up your storage and causing inadvertent auto-scaling of your infrastructure. While event logs and audit logs can ramp up organically when your application is under heavy use, there are also other reasons for capacity needs to escalate.

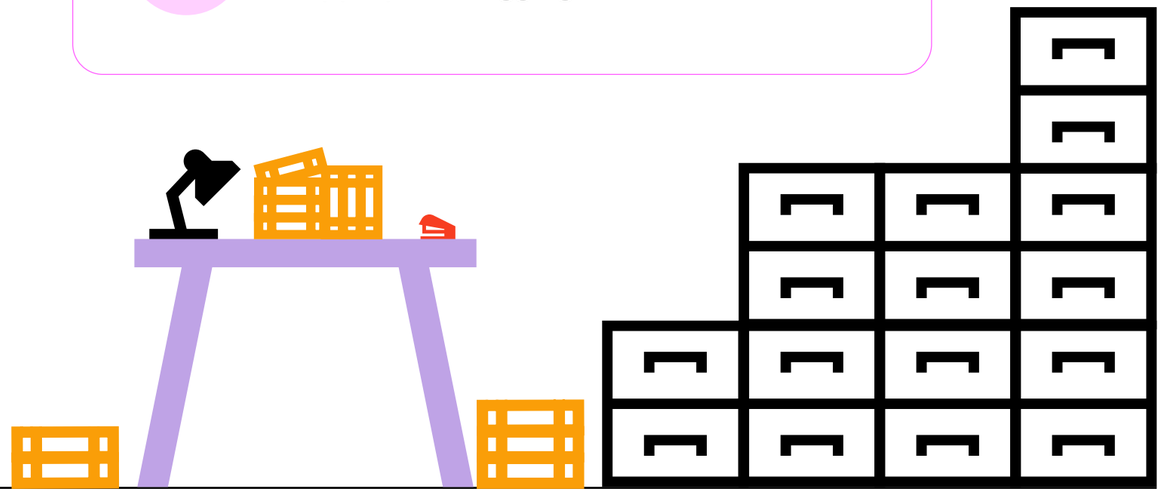
Logs can fill up excessively due to:

- Application misconfiguration
- Malicious exploitation of a vulnerability
- Denial of service (DoS) attack

This increase in log files can trigger undesired auto-scaling. Unless you're on top of your log files, this auto-scale trigger could go under the radar until you've already incurred major costs, particularly where misconfiguration has caused the increase to be gradual.

Use an expert MSP to monitor log file creation and advise you if unusual volumes of logs occur.

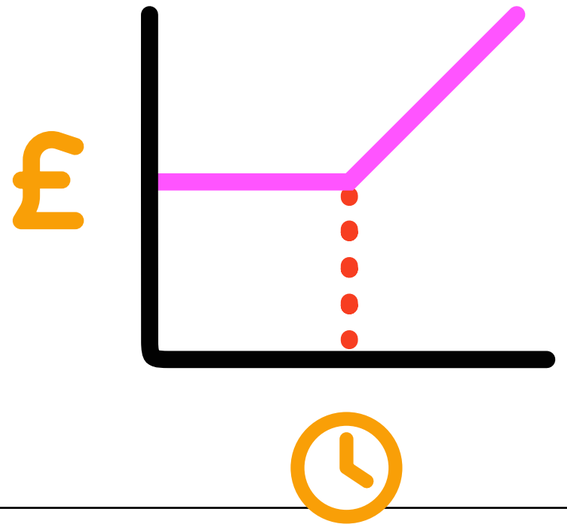
Prevent unnecessary autoscaling from inappropriate logging



4 Malicious attacks and auto-scaling

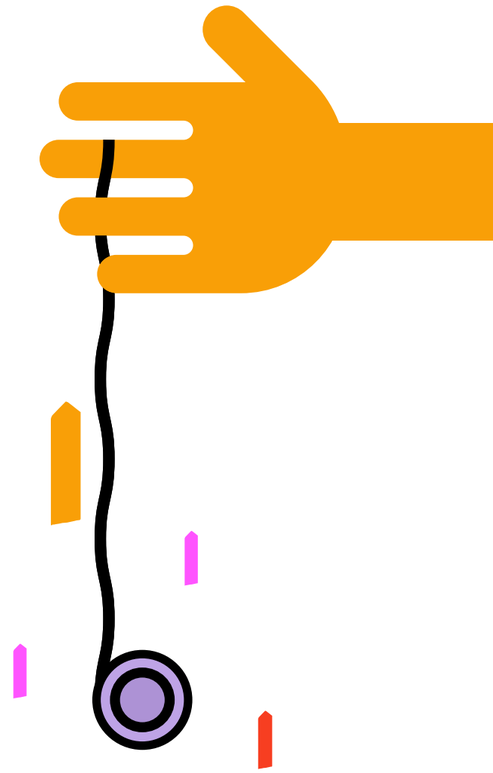
DDoS, EDoS, and FRC

Botnets have enabled a massive increase in DDoS attacks[1] over recent years[4], and this is now a cybersecurity risk for businesses of all sizes. If you have auto-scaling in place, when cybercriminals flood your infrastructure with DoS, it automatically increases capacity - and you'll end up footing the bill. The attack then becomes an Economic Denial of Sustainability (EDoS) or Fraudulent Resource Consumption (FRC) event, where you suffer financial pain from processing fake traffic[2].



Yo-yo attacks

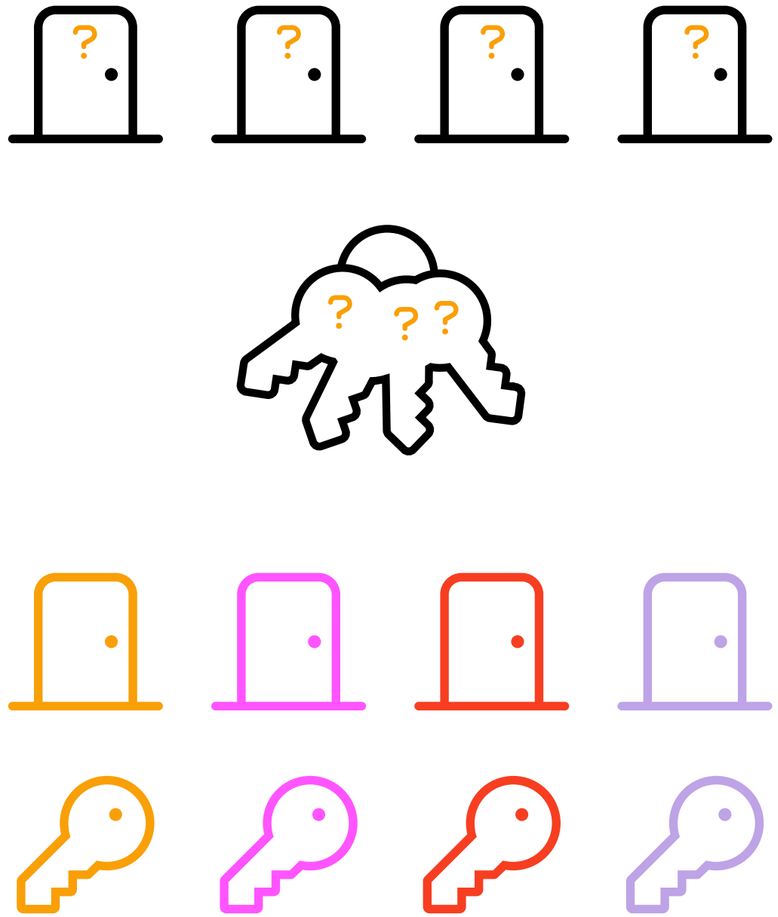
Another form of EDoS cyberattack specifically designed to exploit the auto-scaling facility is called a 'Yo-yo attack'. The attacker fluctuates between overloading your system with a burst of traffic and releasing the pressure - so your SaaS auto-scaler repeatedly scales your resource capacity up then down again. The timings of the traffic spikes are designed to maximise financial damage to its target—your business.



4 Malicious attacks and auto-scaling

Increased attack surface

Auto-scaling invites another cybersecurity risk too: if you are unfortunate enough to be the target of a brute force attack, your attacker can create a greater attack surface by getting your SaaS to auto-scale, and with a larger number of instances that need to be secured, this allows them to try passwords faster, increasing the likelihood of them breaching your system.[4]



- Use an expert MSP to keep you up to date with the latest security standards and avoid gaps in your security
- Protect your business from malicious runaway auto-scaling

Get in touch

Find out how your SaaS business can avoid the pitfalls of unrequired auto-scaling.

TALK TO US

We manage your cloud infrastructure so you don't have to, providing scalable services to increase availability, secure your solutions and improve cost efficiency, giving you more time to run your business.

ProActive – it's all in the name

References

1 https://www.sciencedirect.com/science/article/abs/pii/S1389128623003407.

2 https://research.redhat.com/blog/research_project/ddos-attacks-on-cloud-auto-scaling-mechanisms/