

Optimize usage, and save on cloud infrastructure spending using ITOps analytics



Optimize usage, and save on cloud infrastructure spending using ITOps analytics

Introduction

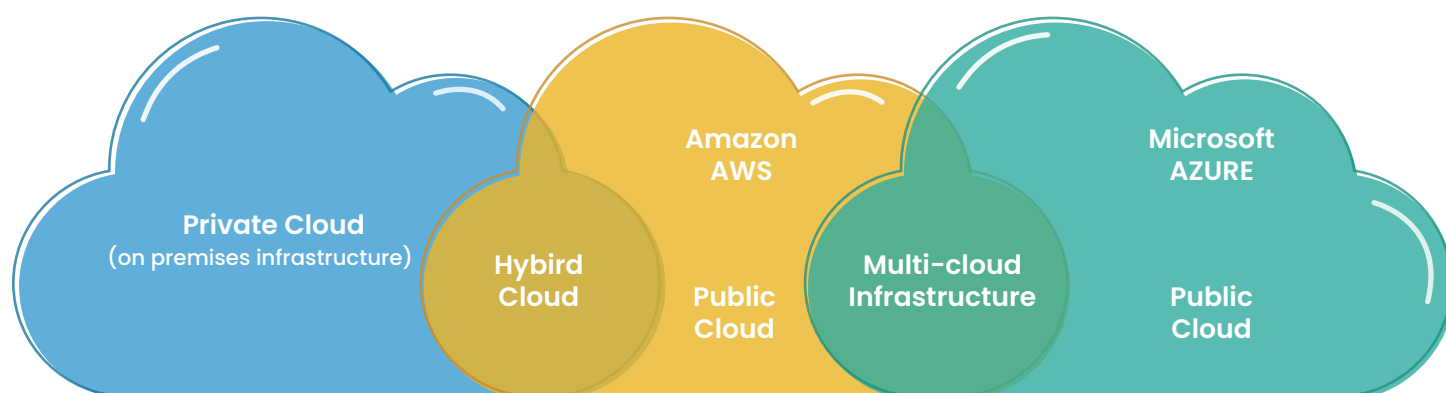
ITOps is in the midst of a rapid shift from on-premises infrastructures to cloud infrastructures. While the cloud infrastructure offers several benefits such as flexibility, reliability, cost savings, and security, it can quickly result in increased IT spending if not meticulously used, monitored, and managed.

Failure to monitor over provisioning or under utilization of cloud resources can quickly increase cloud costs. In this e-book, we'll discuss how organizations can optimize their cloud infrastructure usage, save costs, and avoid wasted spending on cloud infrastructure using analytics.

Understanding the cloud infrastructure

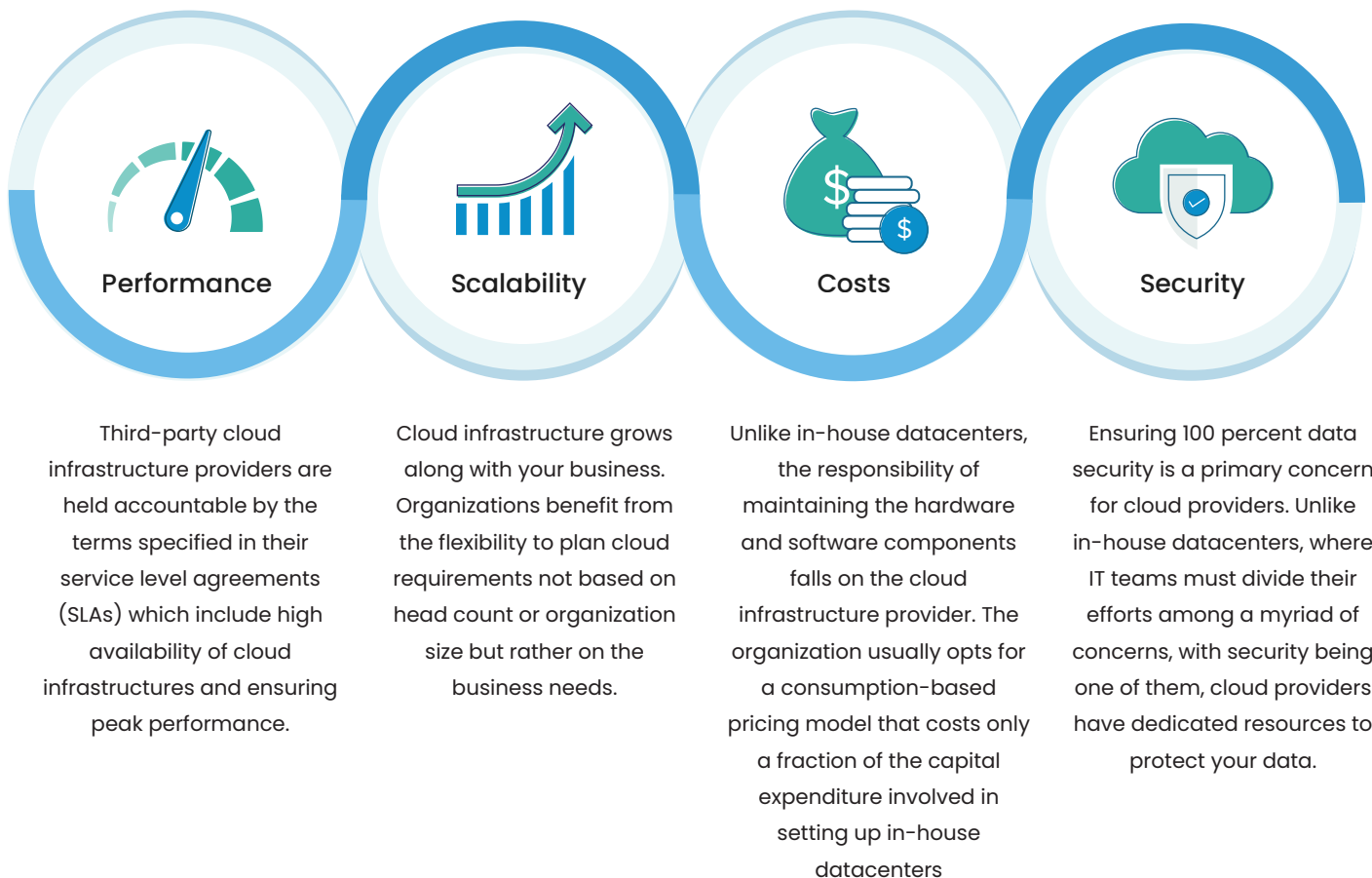
Migration to cloud is on the rise. **Gartner**^[1] predicts 80 percent of organizations will move entirely away from in-house datacenters to cloud infrastructures by 2025.

There are different kinds of cloud infrastructures available to organizations: private cloud, public cloud, hybrid, and multi-cloud.



The benefits of cloud infrastructure

Third-party cloud infrastructure offers plenty of benefits over in-house datacenters or private cloud infrastructures.



Key challenges in cloud infrastructures

While cloud infrastructures offer several benefits to the organization, they do pose a few challenges organizations should consider.

- **Siloed systems and applications:** While cloud infrastructure offers greater flexibility and scalability, it's siloed. Organizations usually deploy their own infrastructure monitoring tools to track infrastructure performance, availability, and downtime. Additionally, public cloud providers offer their own monitoring services, such as CloudWatch from Amazon and Azure Monitor from Microsoft, that provide visibility into crucial metrics such as availability and downtime.
While each of these monitoring applications provides details on processor, memory, disk, network, and general health and availability parameters, it's important to have a comprehensive view of all cloud instances to truly understand and maximize the return on investment (ROI) on your cloud infrastructure, without which organizations will end up spending their IT budgets on underutilized, unused, or poorly performing cloud platforms.
- **Increased mean time to repair (MTTR):** Cloud service providers have complete control over their cloud stack, and take utmost care to prevent events from occurring. So, in the rare case that an event does occur, IT teams will not have visibility into the physical hardware upon which their workloads run. Reduced visibility means teams spend more time trying to figure out the root cause of issues rather than resolving them.

Gain visibility into cloud infrastructure using integrated analytics

Cloud infrastructures offer several benefits; visibility seems to be its only major downside. How do you manage your cloud infrastructure efficiently and tackle this challenge? The key is to use an integrated analytics solution that offers total visibility by consolidating all cloud infrastructure data into one console for analysis. This makes it possible for organizations to draw insights into cloud infrastructure usage, performance, and maintenance that can help optimize usage, improve performance, and reduce operational costs.

Reduce costs and optimize use of the cloud infrastructure

The top three cost drivers of cloud infrastructure management are procurement, maintenance, and downtime. Now let's see how you can leverage analytics to save costs on these three aspects of cloud infrastructure management:

- Cloud infrastructure procurement
- Cloud infrastructure maintenance
- Cloud infrastructure service disruption or downtime

1. How to save costs on cloud infrastructure procurement

Due to the diverse nature of cloud architecture and services that each organization subscribes to, there is no single way to approach cloud infrastructure procurement. However, effective cloud infrastructure management boils down to two key steps:

- a. Understanding cloud infrastructure usage
- b. Effective capacity planning

Consider a cloud platform that hosts several web-based applications. Effective capacity planning ensures that the platform offers sufficient application instances and databases that you need to maintain adequate performance of your applications.

Amazon AWS dashboard

AWS Monitor

18

EC2 Instances

14

RDS Instances

18

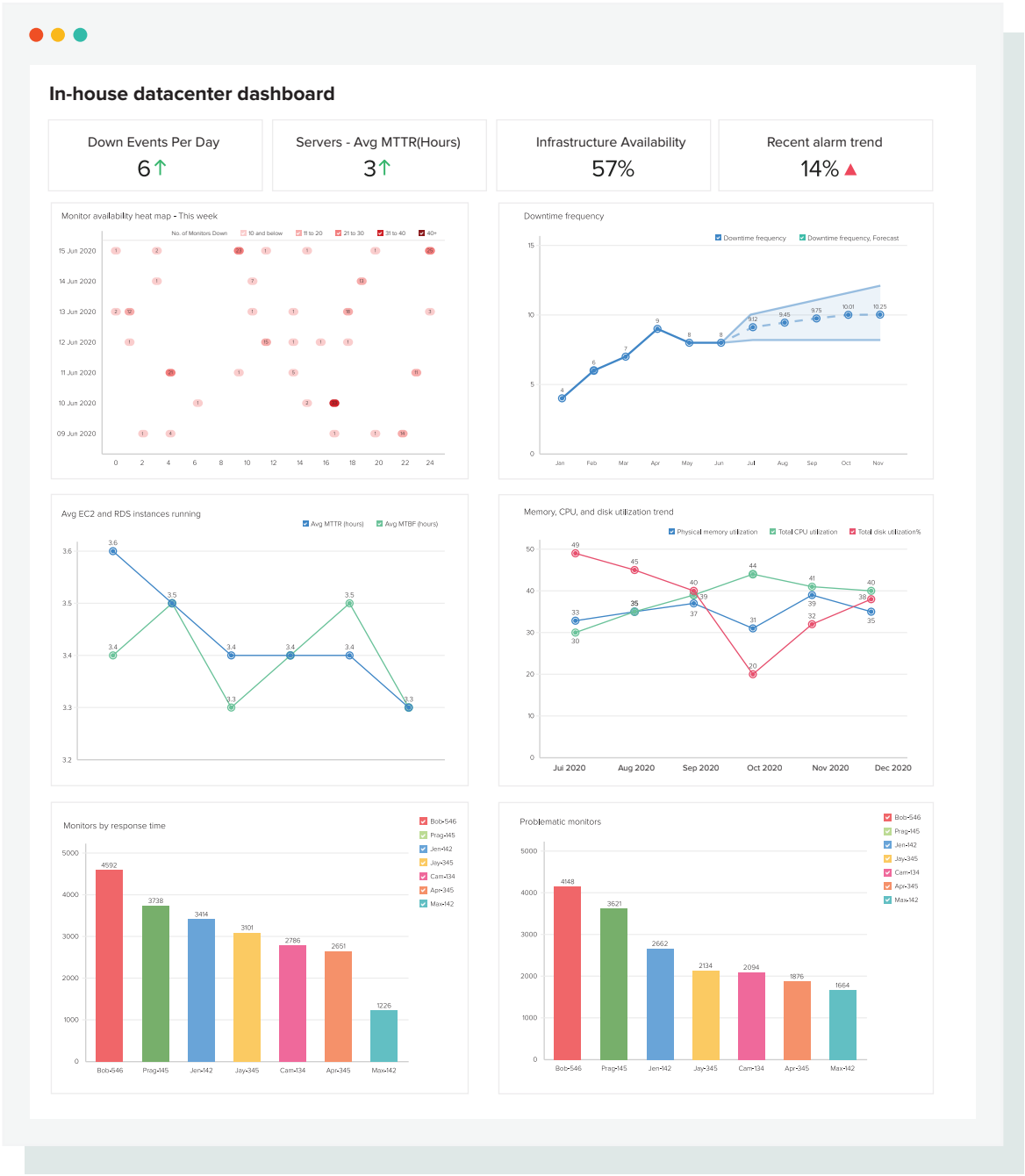


The above dashboard shows the critical usage metrics for Amazon AWS infrastructure for the last six months. Metrics for the number of AWS, EC2, and RDS instances in use, the trend of utilization, and the network traffic enable you to understand the current usage trend of the organization. This will also help you establish a baseline for capacity requirements that are essential to maintain adequate performance.

Next, you need to plan cloud capacity requirements after analyzing whether the CPU, memory, and disk utilization match the workload requirements.

A thorough comparison of usage and requirements helps you plan and invest intelligently in cloud infrastructure. In general, cloud infrastructure subscriptions follow a pay-per-use model, so subscriptions can be cancelled or upgraded at anytime. This means organizations don't have to wait for the beginning of a financial year to start saving on their cloud assets.

In case you're using in-house datacenters or private cloud to host your servers and applications, you need to look into reports that give you the availability and downtime of these platforms along with all the metrics shown in the earlier dashboard.



2. How to save costs on infrastructure maintenance

When you pay for a cloud service, you're investing to use their infrastructure, so it's normal to expect the service to be accessible at all times without disruptions. These ideal service levels are often difficult to attain and impossible to guarantee. This is why it's important for organizations to evaluate key cloud infrastructure maintenance metrics such as application availability, performance, and health, and understand the time, effort, and money involved in maintaining cloud infrastructure. This will enable you to make strategic infrastructure decisions to run business operations smoothly, minimize events or service disruptions, and manage maintenance costs.

Service availability is the percentage of time a service is available. It tells you whether the cloud service is available and performing adequately. The industry standard for cloud service availability is **99.9 percent**^[2]. This means that service providers must have less than 90 seconds of downtime in a day, 45 minutes in a month, and 9 hours in year.

Top cloud service providers such as



Promise

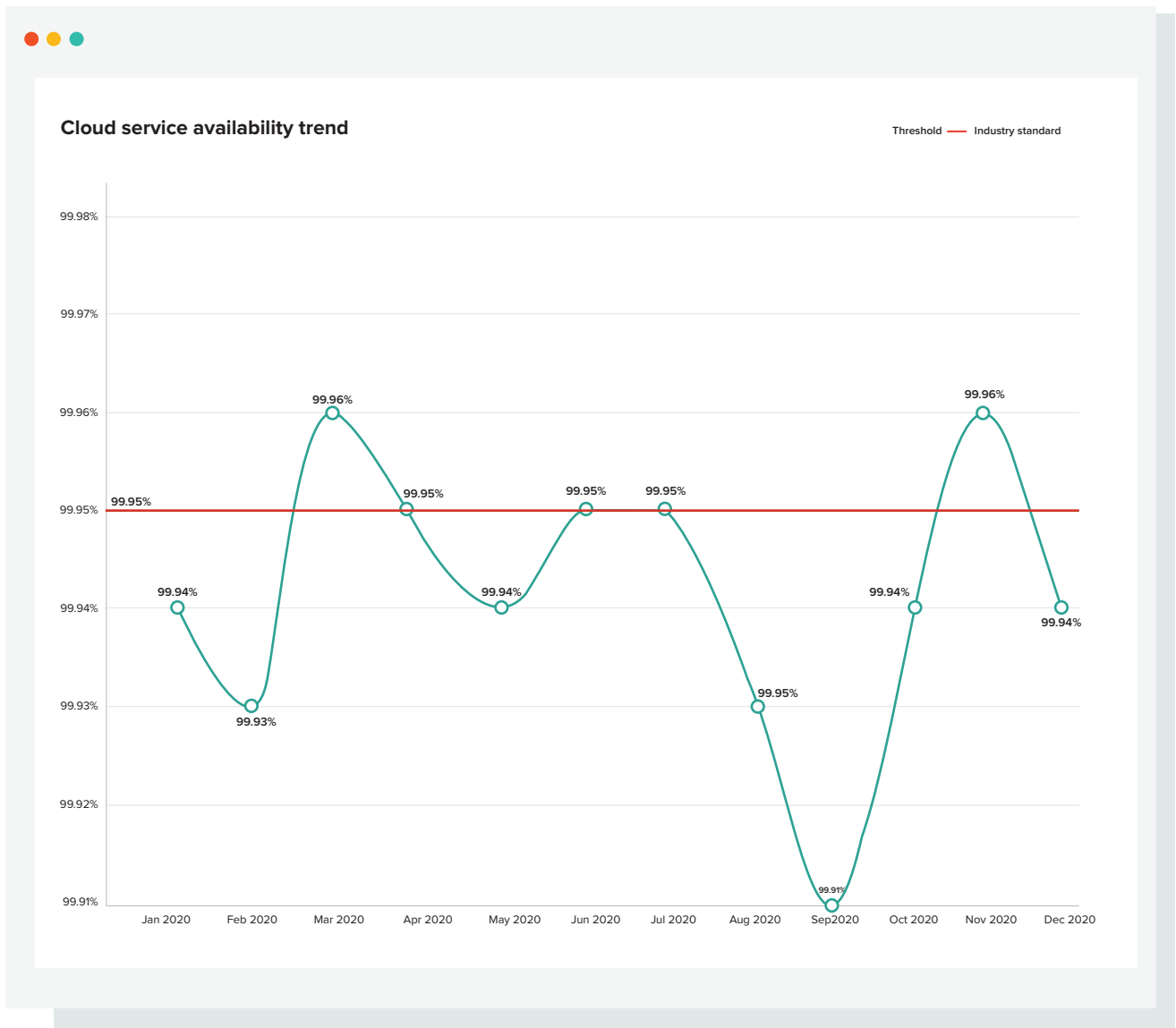
99.95%

service availability in their SLAs

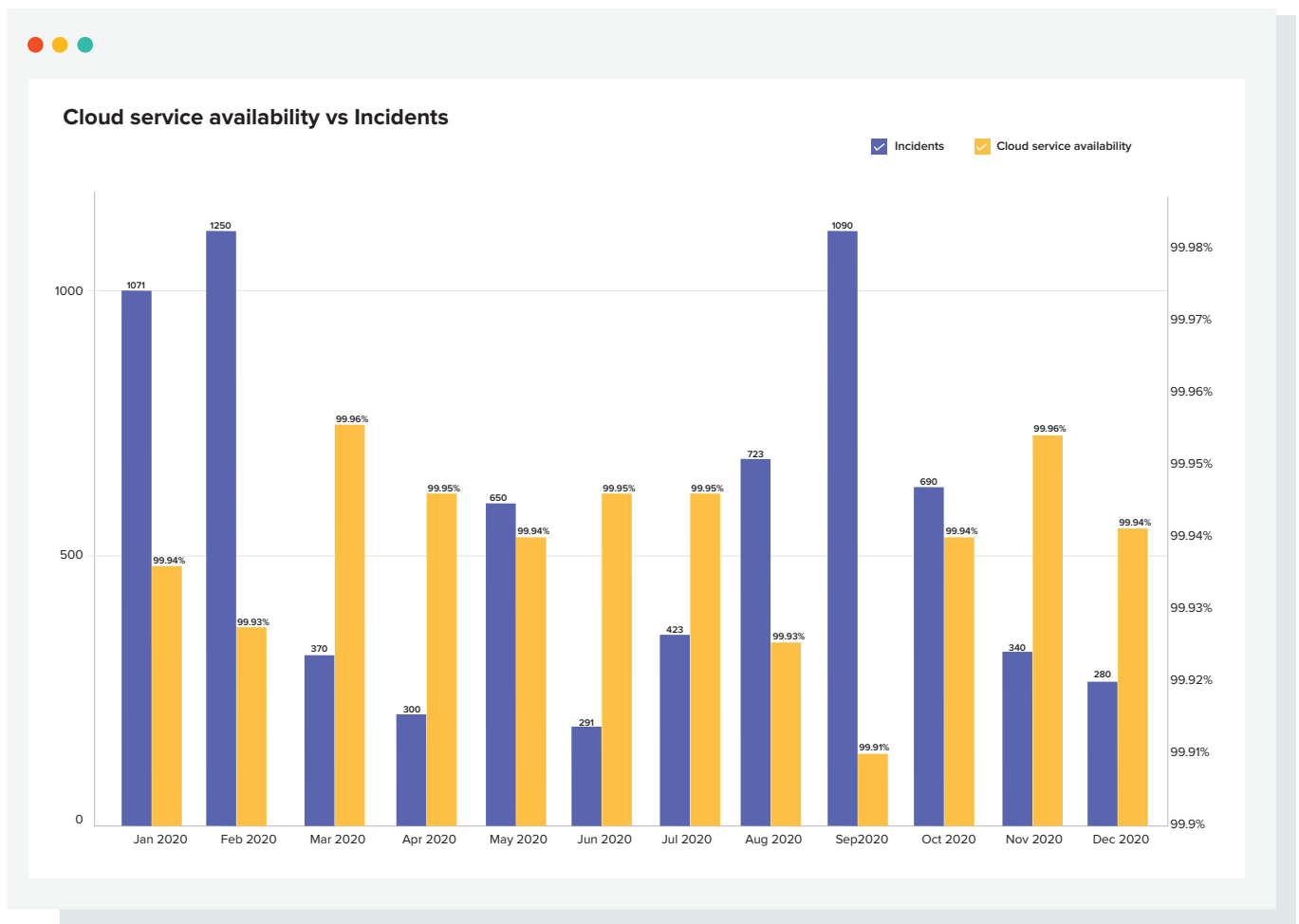
$$\text{Service availability} = \frac{\text{Uptime}}{(\text{Uptime} + \text{Downtime})} \times 100$$

$$\text{Service availability} = \frac{\text{MTBF}}{(\text{MTBF} + \text{MTTR})} \times 100$$

If you're evaluating ROI on your investment with your current cloud service provider, service availability is a good metric to assess the average time cloud services will be online, and what it would cost you to handle incidents during a downtime. The report below gives you the service availability of a sample cloud service provider for the past year.



Now, if you compare this against incidents logged for the same time frame, it's evident that as service availability dips, the number of incidents increases drastically.

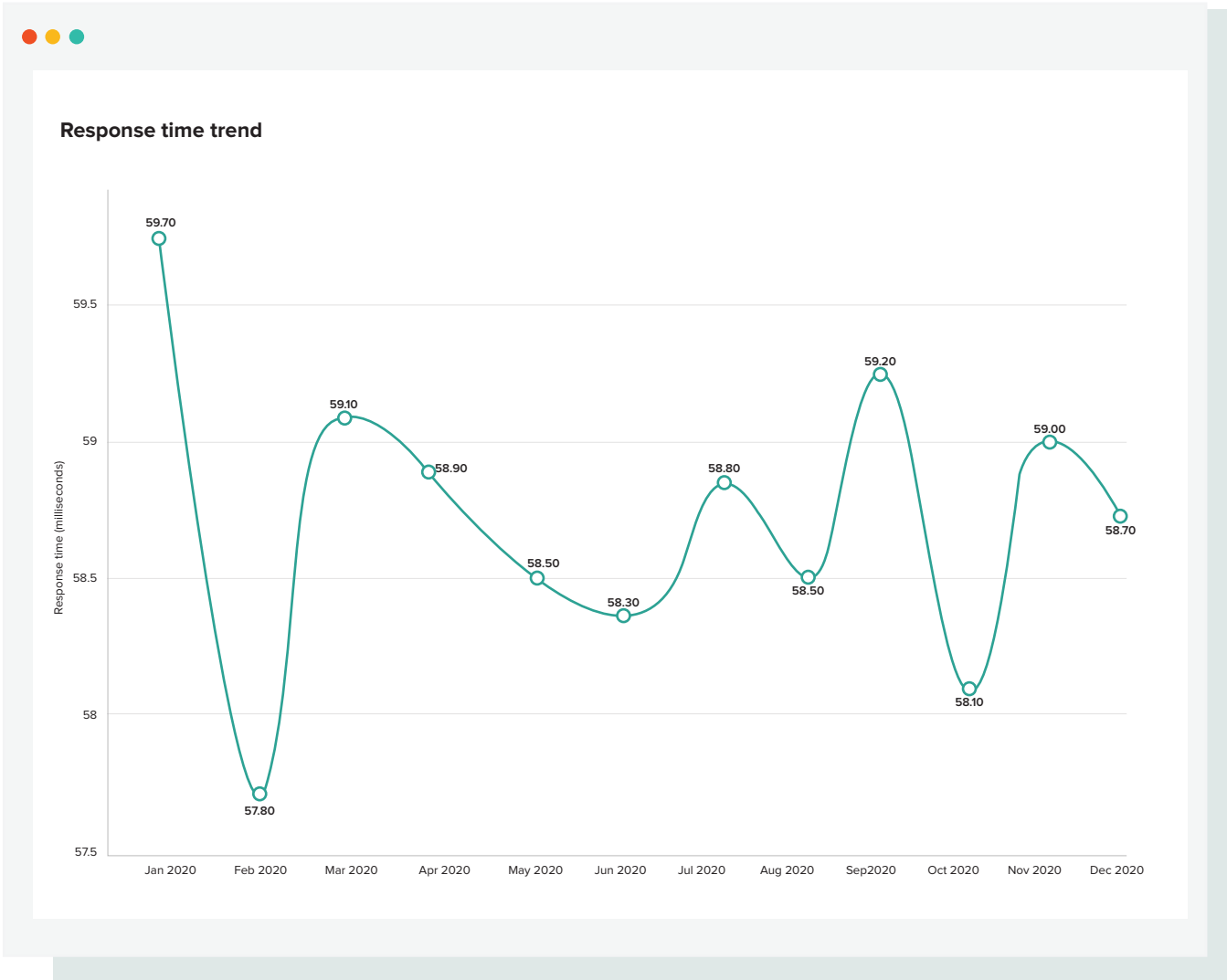


Organizations looking to save on cloud infrastructure spending should consider service availability along with the service cost of incident tickets while calculating the total cost of cloud infrastructure maintenance. Organizations can save costs in the maintenance phase by distributing workloads effectively between multiple cloud service providers to ensure uninterrupted services and minimize downtime, thereby reducing service costs.

Another approach to saving costs is to classify cloud maintenance needs based on core, sunset, and ancillary applications, then build relevant support models or roadmaps to resolve those tickets quickly. This allows organizations to streamline their service processes and trim service costs by eliminating redundant processes and protocols.

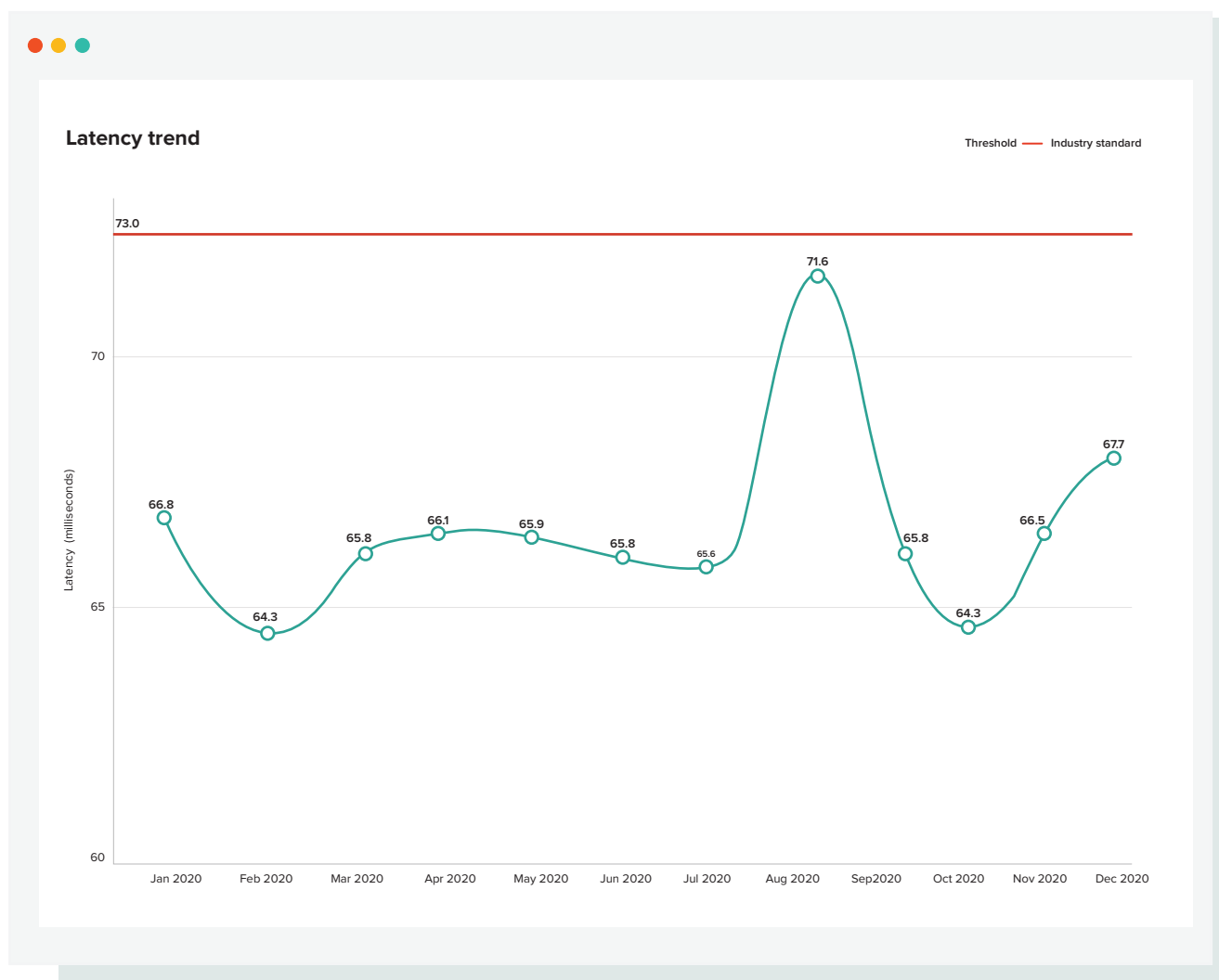
Responsiveness is a critical metric, as it impacts the performance and availability of applications hosted on a cloud server. Low response rates means applications and databases will be sluggish, creating several incidents that can add to your server maintenance costs. Setting up a shared services model with cloud service providers for infrastructure maintenance can help improve responsiveness and save on service costs.

The report below gives you the trend of response rate of various virtual monitors hosted on a cloud service provider.



Latency is a measure of the time interval between submitting a packet and it arriving at the destination. It tells you if the applications hosted on your cloud platform are available and usable. Let's say you have a homegrown web-based application hosted on your cloud server, and its average latency is around 50 milliseconds. If the latency exceeds 50 milliseconds, your applications will have a longer load time. This will result in users raising several tickets complaining about the application's load time, which will eventually increase your support and maintenance costs.

The below report gives you the trend of latency of an application hosted on a cloud server. Higher latency renders applications sluggish. So it's important to maintain lower latency.



Cloud platforms are prone to latency due to a host of factors such as delays in storage, propagation, and transmission. To avoid latency issues, organizations can adopt load balancing, that is, use multiple cloud platforms to host critical applications. This will cut down your service and support costs drastically. Additionally, organizations can compare cloud service providers on a variety of metrics and choose the one who provides the best services at optimal costs.

3. How to save costs by preventing downtime

Downtime is a major cost driver for organizations. It can result in lost productivity, missed opportunities, damaged brand value, and loss of data; depending on the type of business, it can also result in the organization shelling out massive payouts due to failure to adhere to SLA standards.

According to Gartner^[3]

The average cost of downtime is about

\$5,600

per minute

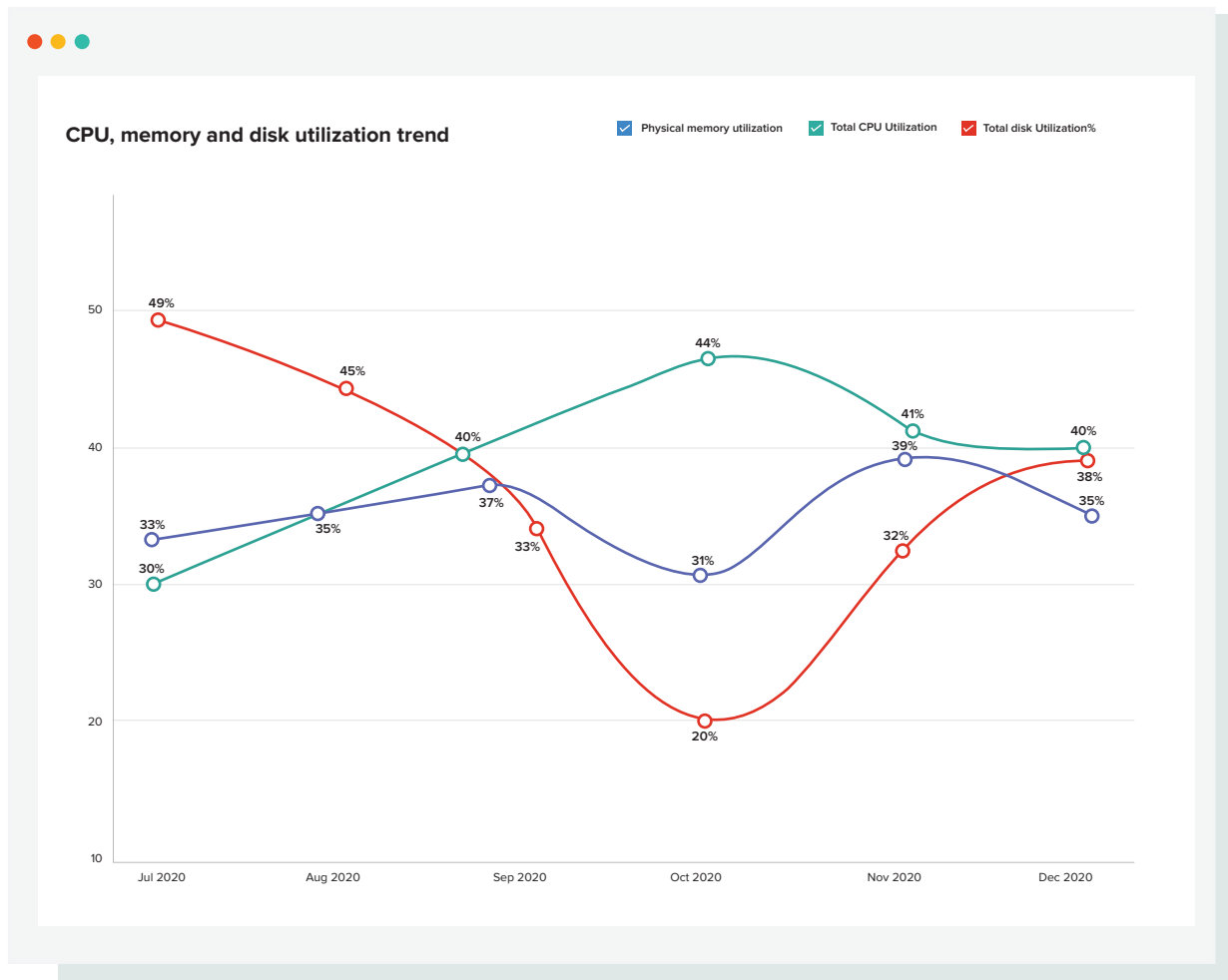
(or)

about

\$336,000

per hour

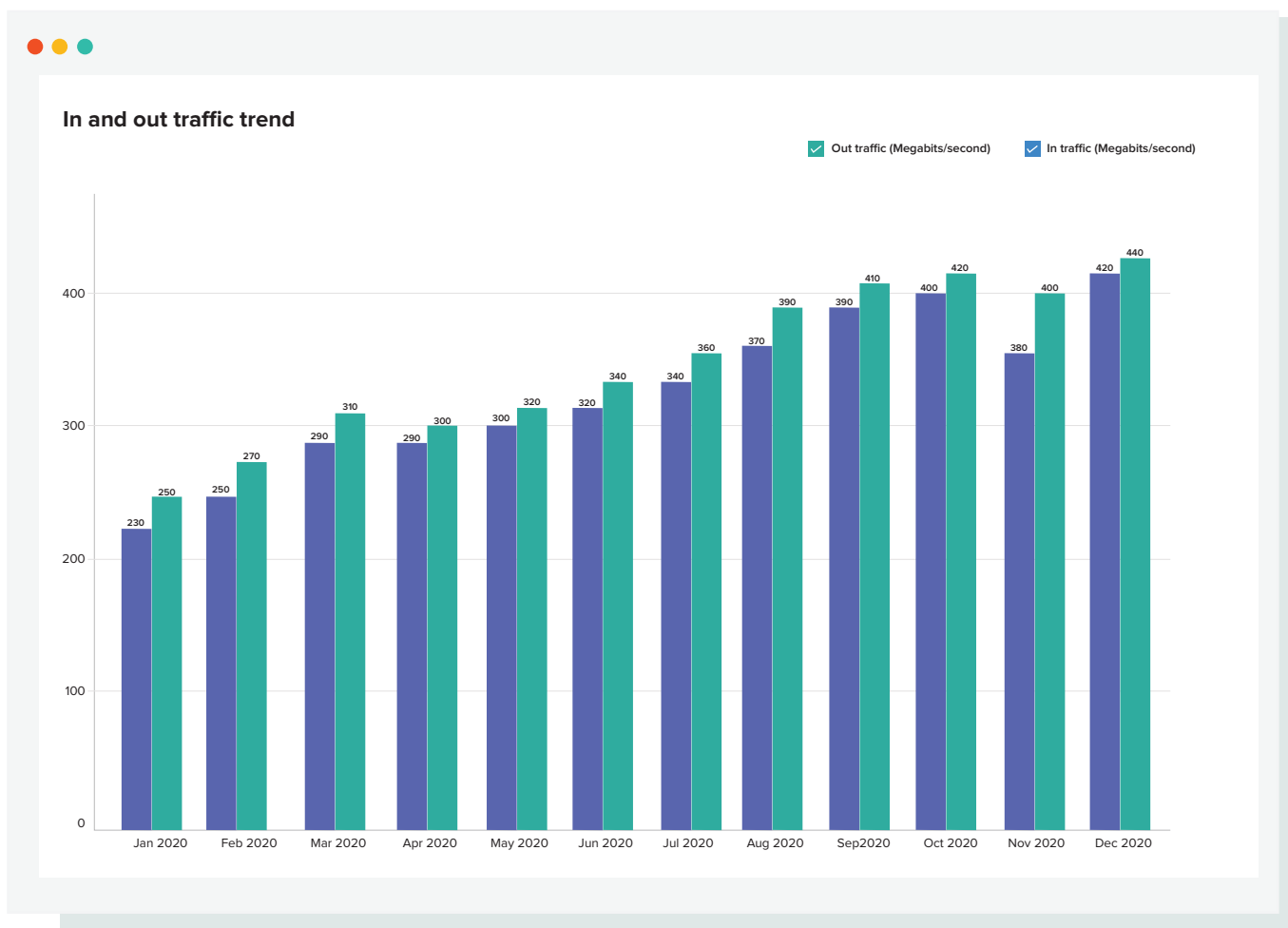
The good news is there are a few steps that you can take to avoid downtime. First, identify the major reasons for downtime. This can vary from one organization to another. However, the most common reasons for downtime are **non-availability of disk, memory, or CPU**. Thoughtful capacity planning should prevent this from happening in the first place. If your resource consumption varies a lot, and you're unable to determine the actual disk, memory, or CPU usage based on historical volumes, it'd be best to set up threshold-based automatic processes to purchase additional disk, memory, or CPU based on everyday usage.



The report above shows you the trend of disk, memory, and CPU utilization by the hour. The threshold lines (created using the average disk, memory, and CPU utilization values) for the last 24 hours can be used to set up alerts and get notified anytime disk, memory, or CPU space runs out.

Monitoring **in-and-out traffic** is important to detect early signs of latency, downtime, and security problems in your cloud platforms. This is significant for e-commerce businesses that rely heavily on the network. Tracking in-and-out traffic can clue you in on website slowness or application latency. This will help operations staff run diagnostics on their internal systems and networks, quickly determine the cause for high or low traffic, and fix problems before they snowball into downtime.

The report below gives you the trend of in and out traffic for the last past year.



The most important metric to consider for cloud infrastructure services is **reliability**. Often reliability is used interchangeably with availability, however, they are entirely different. As explained earlier, availability is the percentage of time a cloud service is available whereas reliability is the probability that a cloud service will meet SLA-specified performance standards.

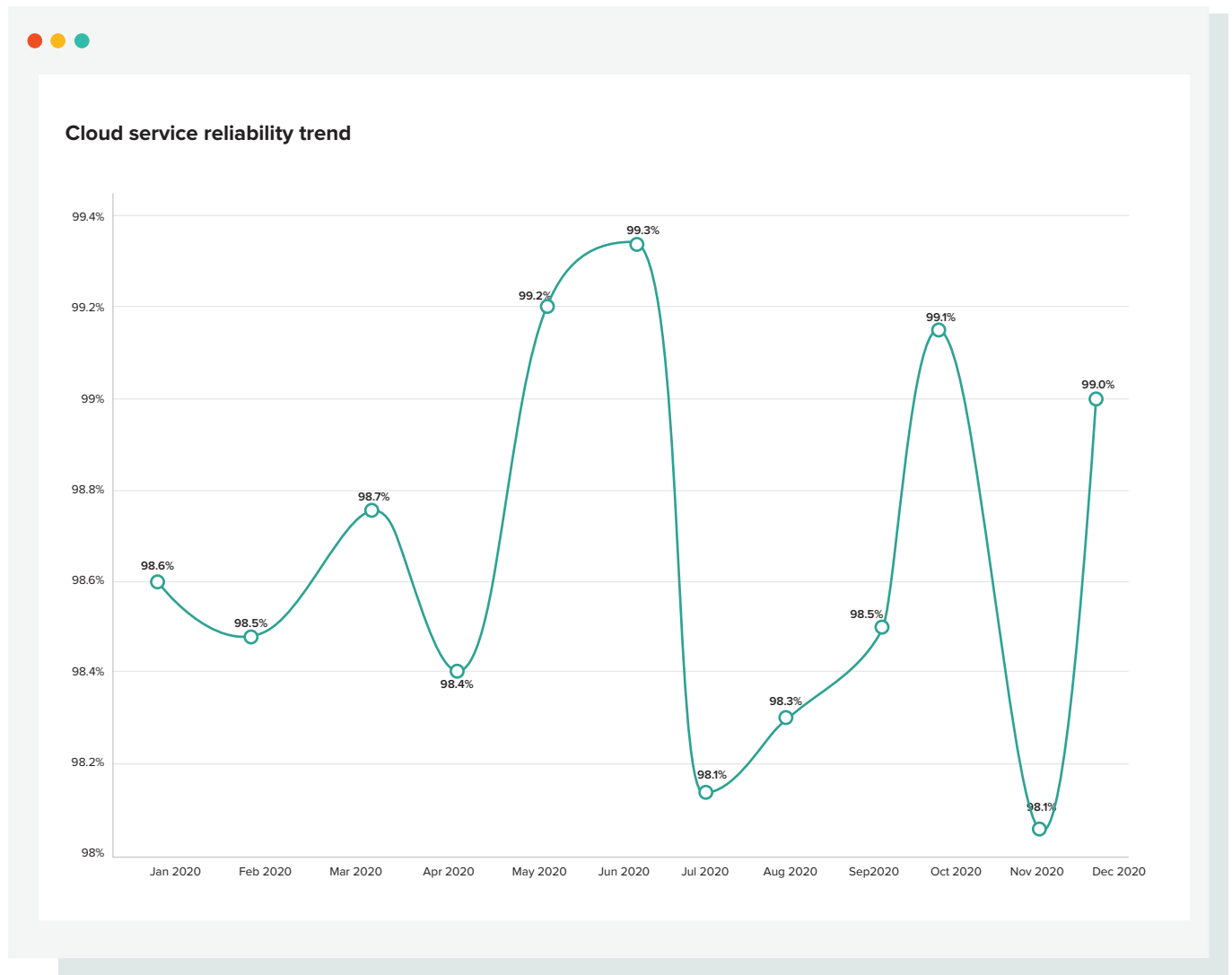
$$\text{Reliability} = 100 - \text{failure rate}$$

To calculate failure rate, use the below formula:

$$\text{Failure rate} = \frac{1}{\text{MTBF}}$$

Why is reliability important? While cloud service providers try their best to adhere to SLA availability times, they're not exempt from real-world conditions that impact availability. Back in 2017, Amazon AWS had a major outage that lasted **over four hours**^[4] and cost businesses millions of dollars in punitive damage and lost business hours. So it's important to understand how reliable your cloud service provide really is. This can help you plan and cushion the impact of downtime by investing in a multi-cloud environment with a distributed cloud framework or moving to alternate cloud platforms that suit your budgets and offer better reliability scores.

The report below gives the trend of reliability for a sample cloud service provider.



Conclusion

For most organizations, moving to the cloud is a drastic step to cut IT operational costs and still ensure peak infrastructure performance. Quite often, this move doesn't pan out as planned and instead, ends up draining up the organization's IT budget. However, with careful analysis of cloud monitoring data, organizations can get critical insights to optimize cloud usage and trim unwanted infrastructure costs.



About

ManageEngine Analytics Plus

Analytics Plus is an IT analytics and business intelligence application that offers organizations a consolidated view of their entire IT infrastructure. It enables them to build key metrics and monitor cloud usage and performance, and draw actionable insights to save costs. Analytics Plus comes with a built-in AI-assistant, Zia, that enables ITOps teams to gain instant insights just by asking voice or text-based questions.

Analytics Plus integrates out-of-the-box with several popular IT monitoring and management tools including ManageEngine's Applications Manager and OpManager. **Download a free**, 30-day trial to test out Analytics Plus' reporting capabilities. Or, **sign up** for a guided tour to see how quickly you can set up Analytics Plus and start analyzing your cloud infrastructure monitoring data.

180K
customers
across the world

18+
years of IT
management experience

90+
products
and free tools

190+
countries
served



Reference

1. <https://www.gartner.com/en/newsroom/press-releases/2018-12-04-gartner-identifies-the-top-10-trends-impacting-infras>
2. <https://www.intelliwavetechnologies.com/maintaining-99-9-uptime-in-the-cloud-reliability-security-monitoring-and-recovery/#:~:text=To%20rely%20on%20cloud%20services,of%20downtime%20in%20a%20year.>
3. <https://www.the20.com/blog/the-cost-of-it-downtime/#:~:text=According%20to%20Gartner%2C%20the%20average,hour%20at%20the%20higher%20end.>
4. <https://aws.amazon.com/message/41926/>

ManageEngine Analytics Plus

© ManageEngine, a division of Zoho Corporation