Whitepaper

IT-Verfügbarkeit: an sich ganz einfach und doch erstaunlich komplex

amasol

Dieser Artikel basiert auf einem Vortrag von Hans Maurer, Aufsichtsratsvorsitzender der amasol AG und Experte für Technology Business Management

Verfügbarkeit: an sich ganz einfach und doch erstaunlich komplex.

1. Kennzahlen und Berichtszeitraum

Warum ist „Verfügbarkeit" heute so wichtig?

Verfügbarkeit ist eine der zentralen Messgrößen für die Qualität eines Systems oder von IT-Services. Denn Geschäftsprozesse wie Produktion oder Logistik sind ohne IT nicht mehr möglich – und ein Systemausfall bringt oft den Stillstand des gesamten oder von Teilen eines Unternehmens mit sich, Produktionsausfälle und Umsatzverluste.

Verfügbarkeit spielt auch eine Rolle bei der Bewertung einer IT-Abteilung. Als „IT-Service-Provider" ist diese dafür verantwortlich, dass die Systeme laufen. Und sie muss – wie externe IT-Servicedienstleister auch – nachweisen, dass die in SLAs (Service Level Agreements) festgelegten Niveaus der Verfügbarkeit gehalten wurden, anderenfalls aber begründen, weshalb nicht. Verfügbarkeit ist schließlich die Grundlage für die Abrechnung der bereitgestellten IT-Systeme und Services.

Der Nachweis der Verfügbarkeit ist ein zentrales Thema bei der Zusammenarbeit von internen und externen Service-Providern sowie deren Kunden.

Verfügbarkeit – was ist das?

Der Begriff ist leicht definiert. Laut Wikipedia ist „die Verfügbarkeit eines technischen Systems […] die Wahrscheinlichkeit oder das Maß, dass das System bestimmte Anforderungen zu einem bestimmten Zeitpunkt bzw. innerhalb eines vereinbarten Zeitrahmens erfüllt. […] Sie ist ein Qualitätskriterium und eine Kennzahl eines Systems." Zu ihrer Berechnung ergibt sich daraus die Formel: Verfügbarkeit = (Gesamtzeit – Ausfallzeit)/Gesamtzeit. Das wird als Prozentsatz berechnet: Verfügbarkeit = (Gesamtzeit – Ausfallzeit)/Gesamtzeit × 100 %.

Diese Berechnung der Verfügbarkeit ist weit verbreitet. In der Regel ergeben sich daraus Werte von über 99 %. Systeme, zu denen eine Verfügbarkeit von 99,99 % und mehr vereinbart wird, werden als „hochverfügbar" bezeichnet. Die Bedeutung dieser Kennzahlen in der Praxis verdeutlichen folgende Beispiele: Für ein System, das 12 Stunden am Tag an 5 Wochentagen 52 Wochen verfügbar sein soll, bedeutet eine Verfügbarkeit von 99 % eine maximale Ausfallzeit von 31,2 Stunden. Bei einem System, das allerdings 365 Tage 24 Stunden verfügbar sein soll und eine Verfügbarkeit von 99,999 % („Five Nines") erzielen soll, kommt eine maximale Ausfallzeit von nur 5,26 Minuten in Frage – im gesamten Jahr!

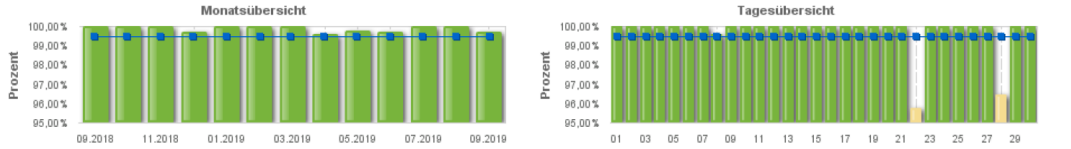
Außerdem gibt es eine Reihe alternativer Kennzahlen, die seltener vorkommen:

- **Anzahl einzelner Ausfälle**
- **Maximale Dauer einzelner Ausfälle**
- **Kumulierte Ausfallzeit**

Diese Kennzahlen kommen bei einem „Single Point of Failure" infrage, der Komponente eines übergeordneten Systems, deren Ausfall den des gesamten Systems verursacht. Dabei spielen Anzahl und Dauer von Ausfällen eine zentrale Rolle für die Bewertung der Verfügbarkeit. So kann ein einzelner Ausfall von einer Stunde andere Auswirkungen – beispielsweise auf die Produktivität des Gesamtsystems – haben als 60 Ausfälle von jeweils einer Minute.

% Verfügbarkeit Webserverdienst CMS Cluster

Vertrag	CMS Hosting TYPO/3	Servicezeit	SZ3 7x24
Metrik	% Verfügbarkeit Webserverdienst CMS Cluster	Ausschluss	keine
Produkt	Managed_Service	Sollwert	99,50%
SLA-Relevanz	Ja		



Einzelverfügbarkeiten

Element	Istwert	Sollwert	Servicezeit	Istzeit	Sollzeit	Status
SLA_WEBCMS	99,73%	**99,50%**	**720:00:00**	**718:02:14**	**716:24:00**	erfüllt

SLA-relevante Ausfälle

Element	Start	Ende	Dauer	Kommentar	Referenz
SLA_WEBCMS	08.09.19 09:37:53	08.09.19 09:43:53	0:06:00		
SLA_WEBCMS	22.09.19 01:25:04	22.09.19 02:25:53	1:00:49		
SLA_WEBCMS	28.09.19 05:54:40	28.09.19 06:45:37	0:50:57	Fehlende Messdaten	

Der Berichtszeitraum – zweite wichtige Bemessungsgrundlage

Eine weitere Kennzahl ist der Berichtszeitraum, für den die Verfügbarkeit gemessen wird. Die Nutzung von Systemen oder Services wird in der Regel auf Monatsbasis abgerechnet. Kürzere (je Tag) oder längere (je Quartal, je Jahr) Zeiträume sind selten. Ein kurzer Berichtsraum kommt für einen Single Point of Failure infrage. Im Bericht werden die Monatswerte mit ihrer Zielvorgabe oft um eine grafische Darstellung der Tageswerte ergänzt. Damit sieht der Kunde auch die Verteilung von Nichtverfügbarkeiten. Ein längerer Zeitraum ist sinnvoll, wenn die Bedeutung des Systems geringer ist oder es selten genutzt wird. Ein gängiger Wert ist hier „Year to Date", die Berechnung der Verfügbarkeit im laufenden Berichtsjahr bis zum aktuellen Datum.

Gesamtzeit vs. Betriebs- oder Servicezeit – der „Nenner" in der Formel

Weitere Kriterien sind die Betriebs- und Servicezeit. In der Formel stellt die Gesamtzeit den Nenner dar, die Zeit, in der das System bereitgestellt werden muss. Bei der Kalkulation dieser Größe sind einige Bausteine zu berücksichtigen, etwa die Geschäftszeiten. Ein Webshop wird an 7 Tagen rund um die Uhr bereitgestellt, eine andere Anwendung wird werktags von 9.00 bis 17.00 Uhr benötigt. Ein weiterer Baustein sind die Wartungsintervalle. Es muss klar sein, wenn sie eingeplant sind, ob solche Phasen in die Betriebszeit einfließen und diese entsprechend reduziert wird. Das wirkt sich dann auf Ausfälle während der Wartungsintervalle aus.

Außerdem sind nationale wie internationale Feiertage zu berücksichtigen, an denen eine Verfügbarkeit regional nicht erforderlich ist. Dasselbe gilt für unterschiedliche Zeitzonen, denn „9.00 bis 17.00 Uhr in Deutschland" ist nicht identisch mit 9.00 bis 17.00 Uhr in den USA oder Japan. Sommer- und Winterzeit müssen ebenfalls bei den Definitionen der Betriebszeit und des Berichtszeitraums für SLA-Vereinbarungen überprüft werden. Bei global tätigen Unternehmen, die über die Datumsgrenzen hinaus arbeiten, ist auch der genaue Monatsanfang relevant.

Beim Festlegen der Größe Betriebszeit/Servicezeit spielt auch die Zuordnung zu bestimmten Services, Verträgen oder Mitarbeitern eine Rolle. Wechseln Mitarbeiter innerhalb eines Berichtszeitraums zwischen unterschiedlichen Niederlassungen und greifen dort auf unterschiedliche Systeme und Services zu, kann das Auswirkungen auf die zugesicherte Verfügbarkeit haben.

Es ist auch möglich, dass Services innerhalb eines Berichtszeitraums in Betrieb genommen oder abgeschaltet werden – insbesondere Services aus der Cloud. Diese werden meist kurz oder zu Spitzenlastzeiten eingesetzt und müssen bei der Berechnung der Gesamtverfügbarkeit entsprechend berücksichtigt werden. Dabei stellt sich die Frage, wie „Gesamtzeit" definiert wird. Mögliche Optionen sind die tatsächliche Laufzeit, die reservierte Zeit, in der Ressourcen garantiert bereitstehen, oder der Monat. Abhängig von der gewählten Option kann ein Ausfall zu erheblich unterschiedlichen Verfügbarkeitswerten führen. Es wird damit unter Umständen schwierig, eine faire Regelung zu finden.

Vertragspartner sollten die wichtigsten Kennzahlen wie Gesamtzeit, Betriebs- und Servicezeit, Ausfallzeit, Verfügbarkeitszielwert und Berichtszeitraum gemeinsam vorab definieren. So entstehen im Rahmen des Reportings valide Daten, auf deren Basis das Erreichen der Zielwerte zuverlässig beurteilt werden kann.

2. Datenquellen und Ergebnisse

Endergebnisse oder Detaildaten

Daten und Messwerte für die Kalkulation der Verfügbarkeit eines IT-Systems oder -Services kommen in der Regel aus zwei Quellen: Es werden „fertige" Werte in Prozent geliefert oder es stehen Rohdaten zur Berechnung bereit. Fertige Ergebnisse können manuell oder automatisch anhand eines Monitoring-Systems erhoben werden. Die manuelle Methode ist sinnvoll bei isolierten Systemen. In der Regel werden Monitoring-Tools eingesetzt. Allerdings verfügen nicht alle dieser Tools über Datenschnittstellen für eine automatische Integration relevanter Daten. Dann ist auch ein manueller Aufwand nötig, um die aus verschiedenen Systemen erhobenen Messwerte zu ergänzen für ein plausibles Endergebnis.

Ein Sonderfall besteht darin, dass ein IT-Service-Provider nicht nur eigene Services bereitstellt, sondern als Vermittler von Services Dritter auftritt. Wenn er als Generalunternehmer SLAs (Service Level Agreements) vereinbart hat, ist er dafür verantwortlich, dass seine Subunternehmer die Leistungsdaten liefern, damit er als Vertragspartner die Verfügbarkeit des Systems nachweisen kann. Auch die Ermittlungsmethode wird vertraglich festgelegt. Das kann zu Komplexitäten führen. Denn es müssen die Methoden des „Vermittlers" und die seiner Lieferanten festgelegt werden. Ferner sollten alle Daten rechtzeitig kommen, damit die Berichte zur vereinbarten Zeit bereitstehen.

Komponenten, Services, Tickets – Quellen für Detaildaten

Basiert ein Verfügbarkeitsnachweis auf eigenen Berechnungen, dann stellt sich die Frage, auf welcher Grundlage die Daten erhoben werden. Die Status- oder Ausfalldaten für IT-Komponenten stammen häufig aus unterschiedlichsten Infrastructure-Management-Systemen, in seltenen Fällen werden sie auch manuell erfasst. So unterschiedlich wie die Monitoring-Systeme sind die Methoden der Erfassung und die Formate, in denen Daten aufbereitet werden. Gängige Methoden sind das regelmäßige Erfassen von Statuswerten wie das Prüfen der Erreichbarkeit im Minutenrhythmus oder das Auswerten von Downtime-Listen (Event Logs). Auch die Verfügbarkeit je Messintervall kann als Bewertungskriterium herangezogen werden. Dabei wird sichtbar, ob ein System oder eine Komponente während eines bestimmten Zeitraums kontinuierlich oder nur zeitweise (prozentual) verfügbar war.

Aufgrund der zahlreichen unterschiedlichen Systeme, Methoden und Formate, die generell für das Monitoring und Erfassen von Verfügbarkeitsdaten gängig sind, müssen die Ergebnisse in ein definiertes Standardformat überführt werden. Nur so kann eine valide und nachvollziehbare Aussage zur Gesamtverfügbarkeit entstehen.

Detaildaten für IT-Services

Anders verhält es sich mit dem Erfassen von Detaildaten für IT-Services wie E-Mail, CRM oder SAP. Bei solchen komplexen Services werden häufig die Daten einzelner Teil-Services zu einem Gesamtergebnis korreliert. Dafür gibt es wiederum unterschiedliche Methoden. Eine Option ist der Single Point of Failure, wenn bei Ausfall einer Komponente der gesamte Service bedroht ist. In Systemen, in denen wesentliche Bausteine mehrfach vorhanden sind, bietet sich die Redundanzmethode an. Von Verfügbarkeit wird dann ausgegangen, wenn zumindest eine der Komponenten verfügbar ist. Eine weitere Methode zum Korrelieren der Komponenten-Detaildaten ist das Quorum. Dieses wird beispielsweise bei Terminal-Server-Farmen eingesetzt. Hier kann per Definition eine Verfügbarkeit bestehen, wenn alleine die Hälfte der Server bereitsteht.

				11/2018				
Path			Corr.	Availability	Planned	Unplanned	Total	Tickets
System			Quorum 75%	99,98%	0,00%	0,02%	720:00:00	INC#55172, INC#57121, INC#57125, INC#57242
	München		Redundant	100,00%	0,00%	0,00%	720:00:00	
		Terminal Services A		99,86%	0,14%	0,00%	720:00:00	
		Terminal Services B		99,86%	0,14%	0,00%	720:00:00	REQ#44548
	Berlin		Redundant	100,00%	0,00%	0,00%	720:00:00	
		Terminal Services A		99,86%	0,14%	0,00%	720:00:00	
		Terminal Services B		84,78%	0,36%	14,86%	720:00:00	
	Hamburg		Redundant	99,85%	0,00%	0,15%	720:00:00	INC#55172, INC#57125, REQ#44557
		Terminal Services A		99,85%	0,14%	0,01%	720:00:00	INC#57125, REQ#44557
		Terminal Services B		0,00%	0,00%	100,00%	720:00:00	INC#55172
	Stuttgart		Redundant	100,00%	0,00%	0,00%	720:00:00	
		Terminal Services A		99,79%	0,21%	0,00%	720:00:00	
		Terminal Services B		100,00%	0,00%	0,00%	720:00:00	
	Düsseldorf		Redundant	98,32%	0,00%	1,68%	720:00:00	INC#57121, INC#57242, REQ#44892
		Terminal Services A		78,51%	0,14%	21,35%	720:00:00	INC#57121, REQ#44554
		Terminal Services B		92,35%	0,15%	7,50%	720:00:00	INC#57242, REQ#44554, REQ#44892

Bei der Berechnung des Gesamtergebnisses sind auch Korrelationsstrukturen wichtig. Dabei kann es sich um einfache, feste Strukturen handeln, wie beispielsweise System – Cluster – Server – Prozess, die den

Gesamtservice abbilden. Bei komplexeren Services wie SAP kommt es oft zu vielfältigen Servicebäumen, die Sub-Services enthalten, die ihrerseits aus externen Datenquellen korrelierte Ergebnisse liefern.

Daten zur Verfügbarkeit werden auch aus Anwendersicht erfasst. Dazu dienen Messwerte aus End-User-Experience-Management- beziehungsweise Application-Performance-Management-Systemen. Diese führen über Monitoring Agents synthetische Messungen durch für das permanente Erfassen vergleichbarer Werte. Die Berechnungen erfolgen, indem parallele Messungen an mehreren Clients zu einem Gesamtergebnis aggregiert werden. Auch hier kommen die beschriebenen Korrelationsmethoden zur Bewertung der Verfügbarkeit in Frage.

Beim Real User Monitoring (RUM) wird das Verhalten aktiver Anwender erfasst. Dabei treten Lücken auf, wenn beispielsweise gerade niemand mit dem System arbeitet. Zudem hängen die Messwerte vom individuellen Anwenderverhalten ab. Im Vergleich zum synthetischen Messen eignet sich RUM daher sehr gut für andere Messungen wie von Abbruchraten oder des Anteils erfolgreicher Transaktionen.

Der Anwender hat das Wort: Detaildaten aus Support-Tickets

Eine weitere Möglichkeit zur Verfügbarkeitsberechnung ist das Auswerten der automatisch oder manuell erzeugten Incident-Tickets aus dem IT-Support oder User Helpdesk. Bei der manuellen Bearbeitung müssen einzelne Tickets jeweiligen Komponenten oder Services zugeordnet werden. Das ist nicht einfach, denn der Anwender meldet nur den Vorfall; der Service-Mitarbeiter muss die betroffenen Komponenten und Services finden. Eine weitere Herausforderung ist die Zuordnung der Tickets zu Berichtszeiträumen, denn der Lebenszyklus eines Tickets kann einen Zeitraum von mehreren Wochen oder Monaten umfassen. Es muss klar sein, ob und wie offene Tickets in die Verfügbarkeitsberechnung einfließen. Darüber hinaus sollte festgelegt sein, ob sich Änderungen an aktiven Tickets rückwirkend auf die Ergebnisse eines bereits abgeschlossenen Berichtszeitraums auswirken.

Ausfall von Messgeräten: Mut zur Lücke?

Es muss feststehen, wie mit Datenlücken umgegangen wird. Denn es kann jederzeit vorkommen, dass Messgeräte ausfallen oder Ergebnisse aus anderen technischen Gründen nicht vorhanden sind. Es muss entschieden werden, ob solche Lücken hinsichtlich der Performance als

„verfügbar", „nicht verfügbar" oder „nicht gemessen" bewertet werden und ob sie zu einer Reduktion der Servicezeit führen, was wiederum die Kalkulation der Verfügbarkeit beeinflusst. Außerdem sollte klar sein, ob eine Mindestlänge von Einzelausfällen eine Rolle bei der Berechnung spielt. So könnten Ausfälle unter einer bestimmten Länge nicht berücksichtigt werden.

Es gibt eine Vielzahl an Datenquellen sowie Methoden und Formaten, die zur Verfügbarkeitsberechnung ausgewählt werden können. Um allerdings eine valide Aussage zur Verfügbarkeit des IT-Systems oder IT-Services zu treffen, sind diese Daten einvernehmlich zu korrelieren und in ein Standardformat zu überführen.

3. Korrigieren und Kommentieren

Nachträgliches Korrigieren und Kommentieren der Ergebnisse

Liegen die konsolidierten Ergebnisse zur Verfügbarkeit einmal vor, kann es unter Umständen notwendig sein, Korrekturen durchzuführen. In der Regel werden die Werte nochmals überprüft, wenn ein (vermeintlich zu) schlechtes Ergebnis berechnet wurde – das unter dem im SLA vereinbarten Ziel liegt. Zudem ist es erforderlich, die Zahlen zu kommentieren, insbesondere dann, wenn auch Korrekturen nicht zu einem besseren Ergebnis führen.

Seriöse Korrekturen sind am Endergebnis und an den Detaildaten möglich – sowohl vor als auch nach deren Normalisierung. In welcher Phase des Normalisierungs- und Berechnungsprozesses Korrekturen erfolgen, unterscheidet sich von Projekt zu Projekt. Generell ist es allerdings sinnvoll, sich für eine Vorgehensweise zu entscheiden, da der Korrekturvorgang sonst unübersichtlich und intransparent wird.

Als Methode bietet sich die manuelle Korrektur an. Das übernimmt der Service-Lieferant oder die Fachabteilung, das Service-, das Availability- oder das Customer-Management. Auch die Bewertung der Ergebnisse sollte die für die Systeme verantwortliche Person oder Abteilung übernehmen. Häufig ist dies nicht möglich. Dann erfolgt der Prozess im abteilungs- oder unternehmensübergreifenden Team. Dabei ist es sehr wichtig, dass der Input gesammelt, gemeinsam diskutiert, bewertet und schließlich freigegeben wird. Das gilt für Korrekturen und Kommentare.

Eine automatisierte Korrekturmethode besteht in der Auswertung von Change- oder Incident-Tickets sowie der Listen von Wartungsintervallen.

Beispielsweise sind Systemausfälle während laufender Changes für die Berechnung der Verfügbarkeit nicht relevant und müssen herausgerechnet werden. Dabei sind die Vorfälle den entsprechenden Komponenten, Services und Verträgen korrekt zuzuordnen.

Die Auswirkungen der Korrekturen auf die Berechnung der Verfügbarkeit fallen unterschiedlich aus. So kann die Ausfallzeit verringert werden, was sich natürlich positiv auf das Ergebnis auswirkt. Manchmal ist es notwendig, Ausfälle in einzelne, für die Berechnung relevante und nicht relevante Phasen zu zerlegen, zum Beispiel wenn eine Störung über den vereinbarten Change-Zeitraum hinausgeht. Außerdem ist es möglich, dass Korrekturen zu einer Reduktion der Betriebszeit führen, was sich ebenfalls auf das Endergebnis auswirken kann.

Multidimensionale Sichtweise durch Verdichtung der Einzelergebnisse

Mittlerweile hat das Volumen an monatlich erfassten Reporting-Daten ein Ausmaß angenommen, das es beinahe unmöglich macht, alle Daten auszuwerten. Vonseiten des Managements wird außerdem in der Regel ein einfaches Ampelsystem bevorzugt. Deshalb muss eine seriöse Verdichtung der Ergebnisse einen schnellen Überblick über die aktuelle Situation geben und die Möglichkeit bestehen, in kritischen Fällen einen Drilldown auf niedrigere Verdichtungsstufen durchzuführen.

Die Verdichtung von Einzelergebnissen erfolgt entlang unterschiedlicher Hierarchien:

- **Vom Service Level Objective (SLO) zum kompletten Vertrag**
- **Entlang technischer Strukturen (IT-Systeme, Services)**
- **Entlang geografischer oder organisatorischer Strukturen (Standorte, Abteilungen, Kunden etc.).**

Methodisch kann die Verdichtung durch die Berechnung gewichteter Mittelwerte erfolgen. Ein weiterer Weg ist das Berechnen der Anzahl verletzter oder erfüllter SLOs. Selbstverständlich ist auch eine Kombination mehrerer Methoden möglich. Bei Bedarf können außerdem noch fehlende Werte, zum Beispiel wenn Services in bestimmten Regionen nicht verfügbar sind, berücksichtigt und „aufgefüllt" werden.

Korrektur, Kommentierung und Verdichten der Einzelergebnisse in unterschiedlicher Art und Weise – diese Schritte erhöhen die Transparenz für Kunden und Service-Provider. Das Verdichten führt zu einer multidimensionalen Betrachtungsweise der Ergebnisse aus unterschiedlichen Sichtweisen und nach verschiedenen Bewertungskriterien. Werden die Verfügbarkeiten von Services je Kunde und Standort ermittelt, ist für die Service-Manager der standort- und kundenübergreifende Vergleich einzelner Services interessant, für den Standortverantwortlichen aber der service- und kundenübergreifenden Vergleich der Standorte.

United Kingdom \ Percentage fulfilled KPIs: 75,00% (6 of 8)

Häufig werden auf Basis der Ergebnisse der Verfügbarkeitsberechnung auch die im SLA festgelegten Bonus-Malus-Regelungen umgesetzt und die entsprechenden Pönalen berechnet. In einem laufenden Berichtszeitraum liefert die Berechnung darüber hinaus wichtigen Input für die proaktive Steuerung der Systeme und Services. So lassen sich auf der Grundlage verfügbarer Zwischenergebnisse Prognosen (Best Case, Worst Case) erstellen, auf deren Basis bei Bedarf noch rechtzeitig Maßnahmen ergriffen werden können.

Verfügbarkeit: an sich ganz einfach und doch erstaunlich komplex

Der Titel des Beitrags wurde mit Bedacht gewählt. Auf der einen Seite glaubt jeder zu wissen, was sich hinter dem Begriff „Verfügbarkeit" verbirgt. Auf der anderen Seite zeigt sich dann häufig erst in der Praxis, dass es mit der Formel „Verfügbarkeit = (Gesamtzeit – Ausfallzeit)/Gesamtzeit" allein längst nicht getan ist. Wie es auch die Praxis bei amasol zeigt, steckt der Teufel häufig im Detail. So ist es kein Wunder, dass es unterschiedlichste Regelungen, Datenquellen, Messmethoden, Darstellungsformate und Vorgehensweisen für das Erstellen von Reportings gibt. Vertragspartner müssen einvernehmlich klären, wer was wann wo wie misst und erfasst, denn nur so entstehen valide Daten, die es gewährleisten, gemeinsam Maßnahmen zu ergreifen für ein nachhaltiges Optimieren der Verfügbarkeit von IT-Systemen und -Services.

„IT-Verfügbarkeit ist erfolgskritisch: Ein System, das an 365 Tagen rund um die Uhr eine Verfügbarkeit von 99,999 % erzielen soll, darf maximal 5,26 Minuten ausfallen. Im gesamten Jahr!“

Hans Maurer, Aufsichtsratsvorsitzender der amasol AG und Experte für Technology Business Management

amasol AG | Campus Neue Balan | Claudius-Keller-Str. 3 B | 81669 München
Telefon: +49 (0)89 1894743-0 | info@amasol.de | www.amasol.de