Leistungsverzeichnis “Kubernetes Cluster” für Geschäftskunden

teuto.net Netzdienste GmbH

Stand: **September 2024**

Leistungsverzeichnis Managed Kubernetes Cluster

Alle angegebenen Preise verstehen sich zzgl. der gesetzlichen Umsatzsteuer. Das Angebot richtet sich ausschließlich an gewerbliche Kunden.

Wir arbeiten ausschließlich auf der Grundlage unserer jeweils aktuell gültigen *Allgemeinen Geschäftsbedingungen für Geschäftskunden* sowie unserer *Besonderen Nutzungsbedingungen “teutoStack Cloud-Dienste” für Geschäftskunden*, die jederzeit von Ihnen unter <https://www.teuto.net> abgerufen werden können. Die Cluster werden in der teutoStack Public Cloud betrieben, daher gilt das *teutoStack Leistungsverzeichnis*. Bitte beachten Sie, dass die Auswahl und Preise der Instanzen sich unterscheiden.

Die Abrechnung wird Ihnen monatlich per E-Mail an die von Ihnen hinterlegte E-Mailadresse zugestellt.

Bei Fragen oder Anmerkungen zu Ihrer Rechnung zögern Sie bitte nicht direkt mit uns Kontakt aufzunehmen:

teuto.net Netzdienste GmbH
Niedernstraße 26, D-33602 Bielefeld
Tel: +49 521 96686 – 0
Fax: +49 521 96686 – 22
E-Mail: info@teuto.net

Allgemeines zum Kubernetes Cluster

teuto.net nutzt zur Bereitstellung von Managed Kubernetes die jeweils aktuelle Opensource Kubernetes Release Version. Zusätzlich installiert teuto.net die jeweils aktuelle Cilium Networking Erweiterung.

Auf Wunsch betreibt teuto.net weitere managed Software Komponenten, wie z. B. Datenbanken im Kundencluster oder den optimierten “Innenausbau” mit dem teuto.net Helm Chart base-cluster (siehe *Leistungsverzeichnis “teuto.net Helm Chart base-cluster” für Geschäftskunden*).

Cluster-Größe

Ein Kubernetes Cluster besteht stets aus Control-Plane und Compute-Plane Nodes. Dabei entsprechen einzelne Nodes immer einer Instanz (virtueller Host) in der teutoStack Public Cloud.

Wir nehmen keine Überbuchung der CPU (Cores) vor. Durch das eingeschaltete HyperThreading entsprechen 2 vCPUs in einer Instanz einem Core.

Der lokale Speicher (local Disc) ist flüchtig. Das heißt: bei dem Ausfall des jeweiligen Host-Systems (Compute-Knoten) ist der Verlust der hier gespeicherten Daten sehr wahrscheinlich.

Ein teuto.net Managed Kubernetes Cluster besteht immer aus 3 Control-Plane Nodes vom Flavor-Typ `standard.2.1905`.

Die Anzahl und Flavor der Compute-Plane Nodes sind frei konfigurierbar und kombinierbar. Es stehen die unten aufgeführten Flavor-Typen zur Auswahl.

Berechnungsgrundlage

	Beschreibung
Beginn	Create einer Instanz
Ende	Delete einer Instanz
Einheit	Instanz (Stück)
Granularität	1 Stück ¹
Preisangabe	€/Instanz/h
Berechnung	(angefangene Stunden im Abrechnungszeitraum) * Preis

Standard Instanzen Standard Instanzen sind für Aufgaben mit ausgewogenem CPU-/RAM-Verhältnis vorgesehen.

Flavor	vCPUs	RAM	Local Disc	Preis/Stunde	Preis/Monat ²
standard.2.1905	2	6 GiB	36 GiB	0,0746€	54,4215€
standard.4.1905	4	12 GiB	68 GiB	0,1491€	108,8430€
standard.8.1905	8	24 GiB	140 GiB	0,2982€	217,6860€
standard.16.1905	16	48 GiB	250 GiB	0,5963€	435,3209€
standard.20.1905	20	60 GiB	330 GiB	0,7454€	544,1639€

Zur Orientierung für die Performance der local Disc werden 48IOPS/GiB angegeben. Diese Werte verstehen sich als Richtwerte, nicht als garantierte Werte.

Der Preis pro Monat dient als Beispiel-Rechnung und legt einen Durchschnittsmonat mit 730h (365*24/12) als Abrechnungsperiode zugrunde. Berechnet wird der tatsächliche Verbrauch.

Highmem Instanzen Highmem Instanzen sind für Aufgaben mit besonders hohem Bedarf an Arbeitsspeicher und moderater CPU-Nutzung geeignet.

Flavor	vCPUs	RAM	Local Disc	Preis/Stunde	Preis/Monat ²
highmem.2.1905	2	12 GiB	50 GiB	0,0928€	67,7075€
highmem.4.1905	4	24 GiB	100 GiB	0,1855€	135,4150€
highmem.8.1905	8	48 GiB	200 GiB	0,3710€	270,8300€
highmem.16.1905	16	96 GiB	400 GiB	0,7420€	541,6600€
highmem.32.1905	32	192 GiB	650 GiB	1,2798€	934,2613€

Zur Orientierung für die Performance der local Disc werden 24IOPS/GiB angegeben. Diese Werte verstehen sich als Richtwerte, nicht als garantierte Werte.

Der Preis pro Monat dient als Beispiel-Rechnung und legt einen Durchschnittsmonat mit 730h (365*24/12) als Abrechnungsperiode zugrunde. Berechnet wird der tatsächliche Verbrauch.

Flüchtigkeit von Nodes Einzelne Nodes (Control- und Compute-Planes) können jederzeit durch teuto.net ausgetauscht werden. Dies passiert zum Beispiel um Updates des Betriebssystems oder der verwendeten Kubernetes Software einzuspielen, oder zur Durchführung von Wartungsarbeiten an der unterliegenden Cloud-Infrastruktur (Hypervisor).

Langzeitrabatt teuto.net bietet Kunden, die Ihren Cluster mindestens 12 Monate (8.760h am Stück) bei teuto.net hosten lassen möchten, einen Langzeitrabatt von 14% auf die Control- und Compute-Plane Nodes an.

Um diesen Rabatt in Anspruch nehmen zu können, muss der Kunde dies schriftlich über das Bestellformular oder per Mail an anfrage@teutostack.de mitteilen. Damit geht der Kunde zugleich eine Vereinbarung über eine Mindestlaufzeit des Clusters von mindestens 12 Monaten (8.760h) mit der teuto.net GmbH ein. Um diesen Rabatt in Anspruch nehmen zu können, muss der Kunde dies schriftlich über das Bestellformular oder per Mail an anfrage@teutostack.de mitteilen. Damit geht der Kunde zugleich eine Vereinbarung über eine Mindestvertragslaufzeit des Clusters über 12 Monate (8.760h) mit der teuto.net GmbH ein. Läuft der Cluster länger als 12 Monate (8.760h), gilt der Langzeitrabatt bis zur Abschaltung des Clusters.

Bei einer Beendigung des Geschäftsverhältnisses zwischen dem Kunden und teuto.net bzw. einer Abschaltung des Kubernetes Clusters vor Ablauf der 12 Monate (8.760h), hat der Kunde die Kosten für die verbliebene Restlaufzeit dennoch zu tragen. Der Kunde hat für die Restlaufzeit monatlich eine Pauschale in Höhe der bisherigen durchschnittlichen monatlichen Nutzung zu tragen. Diese wird dem Kunden von teuto.net nach Ablauf des letzten regulären Abrechnungszeitraumes in Rechnung gestellt.

SLA

teuto.net betreibt für die Managed Kubernetes Cluster eine 24/7 Aufzeichnung der betriebsrelevanten Parameter. Die Cluster haben eine Verfügbarkeit von min. 99,5 % im Monatsmittel. Ein Cluster gilt als verfügbar, solange jeweils n-1 der Control Plane- und Compute-Nodes verfügbar sind.

teuto.net stellt folgende Service Level Agreements zur Auswahl:

Typ	Verfügbarkeit	Wiederherstellung	inkl. Support-Zeit ¹⁰	Preis	Einheit
Basic	99,5%	Geschäftszeiten	0h	0,00	€/GB RAM/Monat
Advanced	99,5%	inh. eines Werktags	1h	7,50	€/GB RAM/Monat
Premium	99,5%	24x7, binnen 4h	2h	11,38	€/GB RAM/Monat

Die Verfügbarkeiten gelten im Monatsmittel.

Support-Zeiten per Ticket-System Werktags Mo.-Fr. 09:00 – 17:00 Uhr. Abrechnung der Arbeiten erfolgt nach tatsächlich entstandenem Aufwand mit 20€ je angefangene 10 Minuten. Nicht verwendete Stunden verfallen zum Monatsende.

Updates teuto.net installiert regelmäßig für den Kunden kostenlos die Updates der Kubernetes Cluster. Die aktuellste Minor-Version wird supported.

Die Updates werden in der Regel monatlich, mindestes jedoch einmal pro Quartal, eingespielt. Die Updates finden jeweils am zweiten Dienstag des Monats, sofern dieser auf einen Werktag fällt, statt. Für das Update auf den neuesten Minor-Release wartet teuto.net mindestens den zweiten Patch-Release der neuesten Minor-Release ab.

Um die Updates reibungslos durchführen zu können, muss der Kunden die Best Practises, insbesondere die Konfiguration von Pod Disruption Budgets, einhalten. Eine genaue Beschreibung der Best Practises entnehmen Sie bitte unserer *Dokumentation*.

Bei nicht korrekt definierten Pod Disruption Budgets kann es während der Updates zu Ausfällen bzw. zu einem Non-Graceful Shutdown der betreffenden Anwendung kommen. Für Ausfälle und möglichen Folgen aufgrund von Nicht-Einhaltung der *Best Practises* übernimmt die teuto.net keine Haftung.

Durch den Austausch der Instanzen im Zuge der Updates können wir leider keinen Laufzeitrabatt gemäß *teutoStack Leistungsverzeichnis* gewähren.

Storage / Volumes

Volumes Um Daten zu persistieren stellt teuto.net StorageClasses zur Verfügung, für SSD-Speicher teutostack-ssd und für HDD-Speicher teutostack-hdd. Letztere ist nicht in jeder Cloud verfügbar.

Die Provisionierung eines PersistentVolumeClaims resultiert in einem entsprechend großen Volume im Openstack.

Berechnungsgrundlage

	Beschreibung
Beginn	Create eines Volume
Ende	Delete eines Volume ⁴
Veränderung	Resize eines Volume ⁵
Einheit	Gibibyte (GiB)
Granularität	1 angefangenes GiB ¹
Preisangabe	€/GiB/Tag
Berechnung	Es erfolgt eine tägliche Konsolidierung ⁶ : (angefangene Stunden an dem Tag) * (höchste Gesamtgröße eines Volumes an dem Tag) * Preis; zum Ende der Abrechnungsperiode werden alle angefallenen Tagessummen addiert
Rabatte	keine

Typ	Merkmale	Preis	Einheit	StorageClass
Ceph	HDD basierend	0,0020€	€/GiB/Tag	teutostack-hdd
Ceph-ssd	SSD basierend	0,0067€	€/GiB/Tag	teutostack-ssd

Load Balancer

Octavia Durch Anlegen eines Kubernetes Services vom Typ LoadBalancer wird ein Octavia Load Balancer erstellt. Dieser besteht aus zwei Amphora Instanzen und wird nach unten stehender Preistabelle abgerechnet.

Berechnungsgrundlage

	Beschreibung
Beginn	Create einer Instanz
Ende	Delete einer Instanz
Einheit	Octavia Instanz
Granularität	1 Stück ⁷
Preisangabe	€/Octavia Instanz(en)/h
Berechnung	(Anzahl der Load Balancer-Instanzen die an dem Tag existieren) * Preis
Rabatte	keine
Inklusive	1 Octavia-Instanz/Monat

Die Ausprägungen der verfügbaren Octavia Instanzen auf dem OpenStack Cluster sind im Folgenden aufgelistet.

Profil-Name	HA	Einzelpreis	Instanzen	Instanz-Name	Gesamtpreis	Gesamtpreis Einheit
ha-standard.2202	Ja	0,0251€	2	amphora	0,0502€	€/Octavia Instanzen/h

Das Merkmal HighAvailability HA gibt an, dass der Load Balancer 2-fach redundant im Cluster ausgelegt wird.

Wenn Sie einen Load Balancer starten, finden Sie diesen in der Detailabrechnung im Tab Load Balancer. Dort sind die Instanzen mit dem oben angegebenen Instanz-Name ausgewiesen. Bei Loadbalancern vom Typ HA werden zwei einzelne Amphora-Instanzen im Verbund gestartet und es wird die Menge von 2 ausgewiesen.

Zum Betrieb eines Kubernetes Clusters muss dessen API extern verfügbar sein. Zu diesem Zweck wird ein HA Load Balancer (zwei Instanzen) verwendet, welcher in der Abrechnung inklusive berechnet wird.

Floating IP

Floating IPs⁸ werden über einen angelegten Load Balancer automatisch angezogen und gebunden. Diese werden aus einem öffentlichen Pool der teutoStack public Cloud reserviert. Es handelt sich dabei um sogenannte "offizielle IP-Adressen".

Wichtig ist, dass die mögliche Abrechnung erfolgt, sobald eine Floating IP aus dem Pool für das Projekt reserviert wurde und nicht erst, wenn die IP gebunden wird.

IP-Adressen werden zur Zeit nicht berechnet, wir erwarten allerdings im Gegenzug von den Kunden einen verantwortungsvollen Umgang mit der Ressource.

Nicht verwendete IP-Adressen müssen daher vom Kunden zeitnah in den öffentlichen Pool zurück gegeben werden. Nur so können wir diesen Service weiterhin ohne gesonderte Berechnung erbringen.

Traffic

Traffic ist der ein- und ausgehende Datenverkehr von den, im Projekt genutzten, Ressourcen. Diese sind beispielsweise Instanzen und Load Balancer.

Berechnungsgrundlage

	Beschreibung
Beginn	Datennutzung 00:00 Uhr zum ersten eines jeden Monats
Ende	Datennutzung 24:00 Uhr zum letzten eines jeden Monats
Einheit	Gibibyte (GiB)
Granularität	1 angefangenes GiB
Preisangabe	€/GiB (Gilt bei einer Peakbandbreite von 1Gbit/s)
Berechnung	Auf die Gesamt-Gibibyte abzgl. etwaiger Inklusiv-Volumina wird eine additive Trafficstaffel angewendet
Inklusiv	1 GB Inklusivvolumen / Monat

Typ	Merkmale	Preis	Einheit
Traffic	Summierter Datenverkehr: 0-300 GiB/Monat (entspricht etwa 1 Mbit/s)	0,15€	€/GiB
Traffic	Summierter Datenverkehr: 300-3.000 GiB/Monat (entspricht etwa 10 Mbit/s)	0,12€	€/GiB
Traffic	Summierter Datenverkehr: >3.000 GiB/Monat (entspricht etwa 100 Mbit/s)	0,08€	€/GiB

Es wird keine Drosselung der Geschwindigkeit vorgenommen.

Für Datenvolumen größer 30.000 Gibibyte bzw. 100 Mbit/s und Peakbandbreiten größer als 1 Gbit/s sprechen Sie uns bitte an.

Datenverkehr innerhalb der Cloud-Infrastruktur des jeweiligen Standortes ist im Instanzpreis inbegriffen.

Die Abrechnung erfolgt immer anhand der tatsächlichen Nutzung im Monat. Sofern ein Inklusivvolumen für den Account existiert, wird dieses am Ende des Monats von der Gesamtsumme des gezählten Traffics abgezogen.

Auf die verbleibende Summe wird additiv die vereinbarte Trafficstaffel angewendet.

Beispiel:

	Beschreibung
Traffic Staffel	0-300; 301-3.000; 3.001-30.000
Inklusivvolumen	100 Gibibyte
Verbrauchtes Volumen	3.400 Gibibyte

Berechnung:

3.400 GiB - 100 GiB (Inklusivvolumen) = 3.300 GiB zu berechnen

- 300 GiB á 0,15€ = 45,00€
- 2700 GiB á 0,12€ = 324,00€
- 300 GiB á 0,08€ = 24,00€
- Gesamt 3.300 GiB = 393,00€

Abrufkontingente / Projekt Quotas

Zur Bereitstellung des Kubernetes Clusters wird ein Projekt in der teutoStack Public Cloud erstellt. Abrufkontingente regeln die maximal nutzbaren Ressourcen durch den Kunden innerhalb des teutoStack Public Cloud Projekts. Im Rahmen des Abrufkontingents kann der Kunde kostenpflichtige Cloud Ressourcen z.B. Instanzen nutzen.

Wenn nicht anders vereinbart, gelten die folgenden Standardkontingente:

Merkmal	Regulär
Instanzen	10
vCores	40 vCores
RAM	100 GiB RAM
Storage	1000 GiB Ceph
Floating IPs ⁸	10
Volumes	100
Security Groups	20

(Security Groups resultieren aus der Anzahl der Floating IPs + 10)

Abrufkontingente unterliegen häufig individuellen Anforderungen. Sollten Sie abweichende Anforderungen haben, zögern Sie nicht mit uns Kontakt aufzunehmen.

Berechnung

In der Preisliste werden Preise mit maximal 4 Nachkommastellen ausgewiesen. Im Berechnungsprozess für die endgültige Rechnungssumme wird mit doppelt genauer Floatingpointarithmetik nach IEEE 754 gearbeitet. Die Rechnungssumme wird kaufmännisch gerundet mit zwei Nachkommastellen ausgewiesen.

Erläuterungen zu technischen Größen

In der Preisliste finden sich technische Größenangaben. Die verwendeten Bezugsgrößen unterscheiden sich teilweise. Grund dafür sind die jeweiligen Quell- bzw. verarbeitenden Systeme.

Datenmengen Angaben mit **Dezimalpräfixen** (Megabyte, Gigabyte, Terabyte, etc.) erfolgen, wenn der Größenangabe die Basis 10 zugrunde liegt.

Beispiel: Gigabyte (GB) $10^9 = 1.000.000.000$ bytes

Angaben mit **Binärpräfixen** (Mebibyte, Gibibyte, Tebibyte, etc.) erfolgen, wenn der Größenangabe die Basis 2 und ein 10er Exponent zugrunde liegen.

Beispiel: Gibibyte (GiB) $2^{30} = 1.073.741.824$ bytes

Im Vergleich der Größenangabe GB zu GiB ergibt sich ein gerundeter Unterschied von 7,4%. Aus diesem Grund weisen wir die Werte in unserem Leistungsverzeichnis entsprechend differenziert aus.

Zusätzliche Features

teuto.net behält sich vor zusätzliche Features im Cluster **vorübergehend**, im Rahmen von Beta Tests, zur Verfügung zu stellen, ohne dabei die Stabilität des Clusters zu beeinflussen.

Alle im Leistungsverzeichnis nicht näher spezifizierten Dienstklassen sind **ausdrücklich nicht für den Produktionsbetrieb** ausgewiesen.

Fragen dazu beantworten wir gerne in einem persönlichen Gespräch.

Mängelansprüche

Um einen Mängelanspruch erheben zu können, muss der Kunde seinen Anspruch per E-Mail ggü. der teuto.net Netzdienste geltend machen. Diese können per E-Mail an regress@teuto.net eingereicht werden. Dies muss innerhalb von zwei Wochen nach dem Abrechnungsmonat erfolgen, in dem der Mangel entstanden ist. Dabei müssen der Unique Identifier (UID) der betroffenen Instanz, Volumes bzw. S3 Buckets sowie Zeitangaben zu jeder Nichtverfügbarkeit, die geltend gemacht werden soll, aufgeführt werden. Des Weiteren sind die Vorfälle durch mitgeführte Request Logs zu dokumentieren. Vertrauliche Informationen sind dabei unkenntlich zu machen.

Die teuto.net Netzdienste GmbH wird den Mängelanspruch innerhalb von sechs Wochen nach der Eingangsbestätigung des vollständigen Antrags gewähren oder zurückweisen. Sollte der Anspruch nicht die oben vorgegebenen Informationen erhalten, kann der Anspruch nicht geltend gemacht werden. Die ausgefallene Leistung wird nach Gewährung mit der nächsten Rechnung gut geschrieben.

Für Mängel, die aufgrund einer Nicht-Einhaltung unserer *Best Practises* entstehen, übernimmt die teuto.net Netzdienste GmbH keine Haftung.

Fußnoten

- ¹: Kleinste abrechenbare Zeiteinheit ist eine Stunde (h).
- ²: Der Preis pro Monat dient als Beispiel-Rechnung und legt einen Durchschnittsmonat mit 730h ($365 \cdot 24 / 12$) als Abrechnungsperiode zugrunde. Berechnet wird der tatsächliche Verbrauch.
- ³: Kosten nach Aufwand werden mit 20€ je angefangene 10 Minuten berechnet.
- ⁴: Volumes die nicht im Status `attached` sind, werden äquivalent zu `attached` Volumes berechnet.
- ⁵: Es wird jeweils die höchste Volumengröße im betrachteten Zeitraum für den ganzen Tag abgerechnet.
- ⁶: Aus Verwaltungsgründen werden täglich die gewonnenen Daten konsolidiert und für die weitere Berechnung vorbereitet. D.h., in der Regel wird bspw. für ein Volume wie folgt gerechnet: (höchste Gesamtgröße eines Volumes an dem Tag) * 24h * Preis/h. Abweichend davon beginnt oder endet der Tag jeweils beim Anlegen oder Löschen eines Volumes, es werden also die tatsächlich angefallenen Stunden für die Berechnung dieses Tages bzw. dieser beiden Tage heran gezogen.
- ⁷: Kleinste abrechenbare Zeiteinheit ist ein Tag (d).
- ⁸: Vergabe Gemäß der Richtlinien des RIPE.
- ⁹: Kleinste abrechenbare Zeiteinheit ist ein Monat (m)
- ¹⁰: Nicht verwendete Stunden verfallen zum Monatsende.