Anlage 3 – Preis- und Leistungsblatt

Verbindliche Bieterangaben zum Festpreis, wirtschaftlichen CU-Kontingent und technischen Benchmark

Vergabeart: Offenes Verfahren nach VgV | **Auftraggeber:** Universitätsklinikum Frankfurt (UKF)

Leistungszeitraum: 36 Monate ab vertragsgemäßer Bereitstellung gemäß verbindlicher Bereitstellungsbestätigung des Auftragnehmers; die Bereitstellung muss spätestens bis 15.12.2026 erfolgen.

Verbindliche Gesamtvergütung: 1.100.000,00 EUR brutto. Kein Preiswettbewerb; Wettbewerb über verbindlichen Leistungsumfang und objektiv messbare Qualitätsmerkmale.

1. Bieter und Angebot

Angabe	Eintrag Bieter
Unternehmen	___
Ansprechpartner	___
Angebotsdatum	___
Angebotskennung	___

2. Festpreisbestätigung

Position	Vorgabe / Bieterbestätigung
Verbindliche Gesamtvergütung brutto	1.100.000,00 EUR
Umsatzsteuerliche Behandlung	Umsatzsteuer zusätzlich zur Netto-Gesamtvergütung auszuweisen
Zusätzliche variable Entgelte für Mindestleistungen	0,00 EUR
Bestätigung	Sämtliche MUSS-Leistungen und angebotenen Wertungsmehrleistungen sind mit Gesamtvergütung abgegolten: JA/NEIN_

3. Verbindliches wirtschaftliches CU-Gesamtkontingent

Angabe	Bieterwert
Garantiertes CU-Gesamtkontingent für 36 Monate	___ CU
Feste CU/h-Faktoren	G1=1 / G2=3 / G3=4 / G4=12 / G5=32 bestätigt: JA/NEIN_
Einheitliche Nutzbarkeit	G1-G3 sowie angebotene G4/G5 ohne Teilkontingente: JA/NEIN_
Keine zeitlichen Teilkontingente/Verfallsregeln vor Laufzeitende	JA/NEIN_

Der angebotene CU-Wert ist verbindlich. Das Kontingent muss unter jeder technisch zulässigen tatsächlichen Nutzungsmischung der verpflichtenden G1-G3-Klassen sowie der angebotenen G4/G5-Capabilities einsetzbar sein. Die Anlage 4b-Benchmarkleistung verändert die CU/h-Verbrauchsfaktoren nicht.

3.1 Verbindlicher Accelerator-Ressourcenkatalog

Für jedes angebotene Profil sind Capability, technischer Typ, Mindestmerkmale, fester CU/h-Faktor und Anlage 4b-Rohwerte anzugeben. Für K1b werden je G1-G3 ausschließlich die verbindlichen Benchmarkwerte des in Anlage 3

benannten Referenzprofils gewertet. Für diese drei Referenzprofile ist der Nachweis nach der strengen Identitätsregel des Anlage 4b zu führen. Weitere Profile dürfen im Betrieb aufgenommen werden, ändern jedoch die Angebotswertung nicht.

Profil-ID	Capability	SKU/Instanztyp	VRAM GiB	CU/h	LLM Rohwert	IMG Rohwert	Anlage 4b-Nachweis-ID

3.2 Für die technische K1b-Wertung benannte Referenzprofile

Capability	Profil-ID	LLM Rohwert	IMG Rohwert	CU/h fix
G1				1
G2				3
G3				4

Je verpflichtender Capability ist genau ein Referenzprofil für K1b zu benennen. Dieses Profil muss die Capability vollständig erfüllen und während der gesamten Vertragslaufzeit grundsätzlich verfügbar sein. Ein späterer Profilwechsel ist zulässig, darf aber UKF technisch und wirtschaftlich nicht schlechter stellen und verändert die Zuschlagswertung nicht.

4. Persistenter Storage

Angabe	Bieterwert
Mindestkapazität	100 TB nutzbar; davon >=10 TB mountbarer Block-/File-Storage
Garantierte Gesamtkapazität ohne Mehrpreis	___ TB
Davon über Mindestanforderung	___ TB
CU-Verbrauch Storage	0 CU
Davon mountbarer Block-/File-Storage	___ TB (MUSS >=10 TB)

5. Qualitäts-, Runtime- und Integrationswerte

Kriterium	Mindestwert / Status	Bieterwert
G1 Bereitstellungszeit	<=30 min	___ min
G2 Bereitstellungszeit	<=60 min	___ min
G3 Bereitstellungszeit	<=120 min	___ min
G4 Capability / Bereitstellungszeit	optional wertbar	JA/NEIN_; ___ h
G5 Capability / Bereitstellungszeit	optional wertbar	JA/NEIN_; ___ h
Platform-Runtime vCPU	>=164	___ vCPU
Platform-Runtime RAM	>=1.104 GiB	___ GiB
Runtime Bereitstellungszeit	<=4 h	___ h

Kriterium	Mindestwert / Status	Bieterwert
Plattformverfügbarkeit	>=99,7 %	___ %
Storage-Verfügbarkeit	>=99,7 %	___ %
P1 Reaktionszeit	<=120 min	___ min
P2 Reaktionszeit	<=240 min	___ min
Inkludierte Exit-Unterstützung	>=10 PT	___ PT
Zusätzlicher persistenter Storage	optional	___ TB gesamt
Projekt-/Teilbudgets	optional wertbar	JA/NEIN_
frei konfigurierbare Warnschwellen	optional wertbar	JA/NEIN_
erweiterte Audit-/Verbrauchs-API	optional wertbar	JA/NEIN_
automatisierter strukturierter Plattformexport	optional wertbar	JA/NEIN_
Sandbox-/Template-Automatisierung	optional wertbar	JA/NEIN_
Drittanbieter-Image-Lifecycle	optional wertbar	JA/NEIN_
mehrere private Partnerkopplungen	optional wertbar	JA/NEIN_
kontrollierte externe AI-/API-Anbindung	optional wertbar	JA/NEIN_
OpenAI-kompatibles/gleichwertiges Model Serving	optional wertbar	JA/NEIN_
Immutable Monatsnachweise	optional wertbar	JA/NEIN_
Plattformseitiger Netzübergabepunkt	>=10 Gbit/s	___ Gbit/s
Volumenabhängige UKF Ingress-/Egress-Entgelte	0,00 EUR	JA/NEIN_ (0,00 EUR bestätigt)
IaC-/Deployment-Portabilität	optional wertbar	JA/NEIN_
Exit-Test / technischer Probeexport	optional wertbar	JA/NEIN_

6. Kalkulations- und Leistungspausibilisierung

Der Bieter fügt eine nachvollziehbare, nicht margensensitive Plausibilisierung bei. Sie enthält mindestens die vorgesehenen Profile und Capability-Zuordnung, die festen CU/h-Faktoren, das angebotene CU-Gesamtkontingent, Anlage 4b-Rohwerte und Benchmarkprotokolle, Kapazitätsannahmen und Einlösbarkeit über 36 Monate, Runtime-Pool, Storage, Netzwerk, Ingress/Egress sowie die Bestätigung, dass alle Mindest- und angebotenen Mehrleistungen ohne nachträgliche variable Infrastrukturentgelte erbracht werden. Interne Kalkulationsmargen müssen nicht offengelegt werden.

7. Rechtsverbindliche Erklärung

Mit Angebotsabgabe werden sämtliche angegebenen Mindest- und Wertungswerte verbindlicher Angebotsbestandteil und im Zuschlagsfall Vertragsbestandteil. Abweichende AGB, Preislisten, Fair-Use-Regeln, Credits oder sonstige Vorbehalte gelten nur, soweit die Vergabeunterlagen dies ausdrücklich zulassen.