



**«TASDIQLAYMAN»**  
**“Biznesni rivojlantirish banki” ATB**  
**Boshqaruv raisi o'rinbosari v.b.:**  
**O.Voxidov**

**«20» may 2026 y.**  
**№ 379**

**“Sun’iy intellekt vazifalari uchun GPUaaS hisoblash resurslarini  
ijaraga olish”**

**TEXNIK TOPSHIRIQ**

Toshkent 2026 y.

## TEXNIK TOPSHIRIQ

### 1. Buyurtmachi

Buyurtmachi – “Biznesni rivojlantirish banki” ATB.

Buyurtmachining manzili: O‘zbekiston Respublikasi, Toshkent shahri, Navoiy ko‘chasi, 18A uy. Tel: (+998 78) 150-00-55

### 2. Ijrochi

Ushbu tanlov bo‘yicha ijrochi tanlash natijalari asosida belgilanadi.

### 3. Texnik ta‘minotga qo‘yiladigan talablar

Sun‘iy intellekt modellarini ishga tushirish, o‘qitish, qayta o‘qitish (fine-tuning) va inference jarayonlari uchun GPUaaS formatidagi hisoblash GPU-resurslari 12 oy muddatga ijaraga taqdim etilishi kerak.

GPU-resurslar jismoniy server ko‘rinishida emas, balki NVIDIA DGX B200 GPU-klasteri asosidagi Run:AI Kubernetes platformasida ajratilgan loyiha (project/tenant) shaklida taqdim etilishi lozim.

DGX B200 tarkibiga 8 ta NVIDIA Blackwell GPU, jami 1 440 GB GPU xotirasi, umumiy 64 TB/s o‘tkazuvchanlikka ega HBM3e xotira, NVSwitch/NVLink texnologiyasi, 112 ta CPU yadrosi, 2 TB gacha tizim xotirasi hamda ichki NVMe saqlash qurilmalari kiradi. Platforma ma‘lumotlarni tayyorlashdan boshlab modelni o‘qitish va inference jarayonlarigacha bo‘lgan AI workload‘lari uchun mo‘ljallangan.

Texnik taklifda taklif etilayotgan GPUaaS platformasi, ajratiladigan GPU/CPU/RAM/storage resurslari hajmi hamda texnik xarakteristikalari ko‘rsatilishi kerak.

Taqdim etiladigan GPUaaS xizmatining texnik ko‘rsatkichlari texnik topshiriqda belgilangan minimal talablardan past bo‘lmasligi zarur.

Ijaraga taqdim etiladigan GPUaaS resurslari ro‘yxati va texnik talablari 1- va 2-jadvallarda keltirilgan. Texnik taklif barcha ko‘rsatilgan resurslar va xizmatlar bo‘yicha to‘liq taqdim etilishi lozim. Ayrim resurs yoki xizmat bo‘yicha texnik taklif taqdim etilmagan taqdirda, bu taklif rad etiladi va ishtirokchi tanlashdan chetlashtiriladi.

1-jadval

### GPUaaS resurslari ro‘yxati

№	Nomi	Soni	Muddati
1.	NVIDIA DGX B200 / Run:AI Kubernetes asosidagi AI workload‘lari uchun GPUaaS resurslar pool‘i.	1	12-oy

Shu bilan birga, tanlovda potensial ishtirokchilar doirasini kengaytirish maqsadida buyurtmachi tomonidan texnik topshiriqda ko'rsatilgan talablarga mos yoki undan yuqori texnik imkoniyatlarga ega GPUaaS yechimlari ham ko'rib chiqiladi. Shu munosabat bilan, shunga o'xshash yechim taqdim etilgan taqdirda, quyidagilarni taqdim etish shart:

- Unumdorlik va samaradorlik bo'yicha texnik-iqtisodiy ma'lumotlar;
- Mavjud infratuzilma va axborot tizimlari bilan integratsiya qilish bo'yicha texnik va moliyaviy hisob-kitoblar.

## 2-jadval

### GPUaaS resurslariga qo'yiladigan talablar

№	Parametrlar	Minimal majburiy talablar	Jami soni
1.	Taqdim etish turi	Run:AI Kubernetes platformasi asosidagi GPUaaS	1
2.	Protessor resurslari	Loyiha/workload kvotasi doirasida kamida 64 vCPU	1
3.	Operativ xotira hajmi	Loyiha/workload kvotasi doirasida kamida 512 GB RAM	1
4.	Storage tizimi	StorageClass / PVC / NFS / S3 asosidagi kelishilgan saqlash siyosati bilan	1
5.	Ma'lumotlar saqlash hajmi	Kamida 12,8 TB foydalanish mumkin bo'lgan PVC/NFS/S3 storage	1
6.	Tarmoq ulanishi	GPUaaS uchun fizik RJ-45/SFP/USB portlar talab etilmaydi; xizmatga umumiy tarmoq orqali kirish taqdim etiladi	1
7.	GPU VRAM sig'imi	Kamida 2 × NVIDIA B200, har bir GPU uchun kamida 180 GB VRAM	2
8.	NVIDIA vPC/vApps/RTX vWS	Bazaviy GPUaaS tarkibiga kirmaydi	-
9.	NVIDIA AI Enterprise	Tegishli litsenziya/obuna mavjud bo'lganda	-

	taqdim etiladi	
--	----------------	--

Tanlashda bo‘lg‘usi ishtirokchilar doirasi kengayganida Buyurtmachi tomonidan texnik xarakteristikalarini texnik topshiriqda keltirilgan texnik ko‘rsatkichlarga muvofiq keladigan yoki undan oshadigan o‘xshash yechimlar asosidagi takliflar ko‘rib chiqilishi mumkin.

#### 4. Qo‘llab-quvvatlanadigan workload turlari

- Jupyter / interaktiv ish muhiti
- Paketli vazifalar (batch jobs)
- Modelni o‘qitish vazifalari
- Taqsimlangan o‘qitish (distributed training)
- Fine-tuning
- Inference / model servislarini ishga tushirish
- Maxsus Kubernetes workload’lari

#### 5. Model obrazlari bilan ishlash

Model obrazi Docker/OCI formatida taqdim etilishi hamda GPUaaS-klasterdan foydalanish mumkin bo‘lgan container registry’da joylashtirilishi lozim.

Workload’ni ishga tushirish uchun quyidagi parametrlarni ko‘rsatish imkoniyati mavjud bo‘lishi kerak:

- image URI;
- ishga tushirish komandasi (entrypoint);
- muhit o‘zgaruvchilari (environment variables);
- talab etiladigan GPU, CPU va RAM resurslari hajmi;
- ilovaning tarmoq portlari;
- ma’lumotlar uchun mount path;
- registry va tashqi tizimlarga ulanish uchun secrets.

#### 6. Tashqi ma’lumotlar bilan ishlash

Tashqi ma’lumotlar workload’ga quyidagi usullardan biri orqali ulanishi lozim:

**3-jadval**

Ma’lumot manbasi	Ulanish usuli
NFS / NAS	Data source yoki Kubernetes volume sifatida mount qilish
S3-compatible storage	Data source, credentials yoki sidecar/SDK orqali ulanish
PVC	Kubernetes PersistentVolumeClaim
Git	Kod va konfiguratsiyalarni Git orqali yuklash
Secret / ConfigMap	Konfiguratsiyalar, tokenlar va kalitlarni uzatish

Run:AI platformasi NFS, S3, Git, PVC, ConfigMap, Secret va HostPath kabi data source’larni qo‘llab-quvvatlashi hamda workload fayl tizimiga ma’lumotlarni mount qilish imkoniyatini ta’minlashi lozim.

#### 7. Kirish imkoniyatiga qo‘yiladigan talablar

Ijrochi tomonidan quyidagilar taqdim etilishi lozim:

- Run:AI UI platformasi URL manzili;

- foydalanuvchi hisoblari;
- loyiha (project/namespace);
- GPU/CPU/RAM/storage kvotalari;
- CLI/API orqali ulanish bo'yicha yo'riqnoma;
- container registry'dan foydalanish imkoniyati yoki tashqi registry'dan foydalanish bo'yicha yo'riqnoma;
- tashqi ma'lumotlar manbalariga ulanish parametrlari;
- workload'larni ishga tushirish uchun YAML manifest namunalari.

### 8. Qabul qilish talablari

Xizmat quyidagi tekshiruvlardan muvaffaqiyatli o'tganda taqdim etilgan deb hisoblanadi:

4-jadval

№	Tekshiruv	Qabul qilish mezonlari
1.	Run:AI UI platformasiga kirish	Foydalanuvchi avtorizatsiyadan o'tadi va o'z loyihasini ko'ra oladi
2.	GPU tekshiruvi	Test workload orqali kamida 2 × NVIDIA B200 GPU nvidia-smi yordamida aniqlanadi
3.	CPU/RAM tekshiruvi	Workload'ga belgilangan CPU/RAM resurslari ajratiladi
4.	Storage tekshiruvi	Ma'lumotlar container ichida kelishilgan mount path orqali mavjud bo'ladi
5.	Model obrazi tekshiruvi	Model joylashtirilgan container muvaffaqiyatli ishga tushadi
6.	Endpoint tekshiruvi	Inference endpoint kelishilgan manzil va port orqali foydalanish uchun mavjud bo'ladi
7.	Monitoring tekshiruvi	UI'da workload holati, loglar va metrikalar ko'rinadi

### 9. Xizmat ko'rsatish joyi

Xizmat ijrochining GPUaaS platformasiga himoyalangan tarmoq orqali masofaviy ulanish asosida taqdim etiladi.

Mazkur xizmat formati uchun server uskunalarini buyurtmachi manziliga jismoniy yetkazib berish talabi qo'llanilmaydi, chunki xizmat predmeti jismoniy uskuna yetkazib berish emas, balki hisoblash resurslarini ijaraga taqdim etish hisoblanadi.

kelishuvchilar: X.Abduraximov, S.Kayimov, Z.Orifkhojayev

<https://hujjat.brb.uz/?pin=nO94xF56&id=88d2c046-be8e-44f9-8cb1-c6d953bc145d>