

ICS 35.240
CCS L73

团 体 标 准

T/DGAG 035—2025

自然资源领域地理信息底座软件信创算力 资源需求评估规范

Specification for the assessment of information and innovation computing power
resource requirements for geographic information platform software in the natural
resources field

2025-11-28 发布

2025-12-01 实施

广东省数字政务协会 发布

目 次

前言	11
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 评估原则	2
6 评估指标	2
7 评估实施	2
7.1 地理信息数据库类底座软件	2
7.2 地理信息容器服务类底座软件	3
7.3 地理信息大模型类底座软件	4
8 评估结果	5
附录 A (资料性) 算力资源需求评估用表	6
A.1 需求评估数据采集表	6
A.2 压测结果样式表	6
A.3 评估结果样式表	8
附录 B (资料性) 算力资源实施评估示例	9
B.1 地理信息数据库评估示例	9
B.2 地理信息容器服务评估示例	10
B.3 地理信息大模型评估示例	10
参考文献	12

前　　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由广州市基础地理信息中心提出。

本文件由广东省数字政务协会归口。

本文件起草单位：广州市基础地理信息中心、北明软件有限公司、广州市城市规划勘测设计研究院有限公司、广州数据集团有限公司、广东省数字政务协会。

本文件主要起草人：水学民、蓝磊、林子熙、王习祥、周晓健、朱文胜、刘惠珍、付又和、简钰棠、陈展鹏、黄俊凯、林伟声、吴鸿飞、周丽娟、刘志超、谢明浩、刘学赟、葛海龙、董耀艺、罗畅、杨杰明、彭卓峰、彭晓平。

自然资源领域地理信息底座软件信创算力资源需求评估规范

1 范围

本文件规定了自然资源领域地理信息底座软件信创算力资源的评估要求，包括评估原则、评估指标、评估实施和评估结果等。

本文件适用于自然资源领域地理信息数据库、容器服务和大模型等底座软件部署过程中的信创算力资源估算工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 38634.1—2020 系统与软件工程 软件测试

GB/T 45288.1—2025 人工智能 大模型 第1部分：通用要求

GJB 1063A—2008 机载悬挂装置试验方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

算力资源 computing power resources

计算机执行运算的能力，主要取决于中央处理器（CPU）的时钟频率、核心架构及并行处理能力，图形处理器（GPU）的并行计算架构与浮点运算性能，内存模块的容量与数据传输速率等核心硬件参数的协同作用。

3.2

大模型 large-scale model

基于大量数据训练得到，具有复杂计算架构，能处理复杂任务，且具备一定泛化性的深度学习模型。

注1：大模型的参数量由其功能和模态决定，一般不低于1亿。大模型训练使用的数据总量受参数量的影响，达到收敛的大模型的参数量的对数与其训练数据总量的对数成正比。

注2：支持对地表结构、海洋状态、生物资源、地质特征等自然资源要素的智能识别、变化检测与趋势预测，支持资源管理、环境保护、灾害预警与可持续发展等领域的智能化分析与决策支持。

[来源：GB/T 45288.1—2025，3.1]

3.3

使用方 user

对地理信息底座软件有业务需要和使用场景的应用单位，负责发起地理信息底座软件算力资源评估任务。

3.4

评估方 evaluation party

受使用方委托，负责对项目所需的地理信息底座软件算力资源需求进行测算的组织或团队。

4 缩略语

下列缩略语适用于本文件。

CPU：中央处理器（Central Processing Unit）

GPU：图形处理器（Graphics Processing Unit）

5 评估原则

评估原则如下：

- a) 科学性：基于科学理论与方法，保障结果精准可靠；
- b) 客观性：避免主观干扰，真实反映算力需求实际；
- c) 全面性：覆盖功能、性能、安全等核心维度；
- d) 可操作性：适配自然资源领域业务场景，便于多方协作。

6 评估指标

由评估方根据评估对象的特点和需求，判断地理信息底座软件的类型，并根据对应的评估指标实施评估。评估指标如下：

- a) CPU 核数需求量；
- b) 内存容量需求；
- c) GPU 资源需求量。

7 评估实施

7.1 地理信息数据库类底座软件

7.1.1 评估数据采集

评估数据采集步骤如下：

- a) 使用方提供高峰时段并发用户数及估算高峰压力时长(需求评估数据采集表样式参见表 A.1)；
- b) 评估方根据并发用户数及高峰压力时长，在底座软件的信创部署环境按照 GB/T 38634.1—2020 开展软件压力测试（压测结果表样式参见附录 A.2.1）；
- c) 评估方采集 TPMC 值、单事务平均 CPU 消耗、热数据集、单连接平均内存消耗等数据。

7.1.2 CPU 核数需求评估

CPU核数需求测算见公式（1）。

$$C = \frac{T \times S}{P \times E} \times CPX \quad (1)$$

式中：

C——CPU核数需求量；

T——TPMC值（目标吞吐量）；

S——单事务平均CPU消耗；

P——单核处理能力；

E——预期CPU效率；

CPX——自然资源数据处理复杂度系数。

CPX值应结合具体应用场景进行合理选取。应用场景推荐CPX值表见表1。

表1 应用场景推荐 CPX 值表

应用场景类型	特点	推荐范围值
二维分析应用场景	数据维度低，计算简单	1.5
三维分析应用场景	数据维度增加，空间关系复杂	1.5≤CPX≤2.0
遥感影像处理应用场景	数据维度最高，文件体量大，数据要求时序性全面，高空间分辨率和高/多光谱特征	1.5≤CPX≤2.0

7.1.3 内存容量需求评估

内存容量需求测算见公式（2）。

$$M = (H \times 1.5 + N \times S) \times B \quad (2)$$

式中：

M——内存容量需求

H——热数据集；

N——并发连接数；

S——单连接平均内存消耗;

B——自然资源数据缓存冗余系数。

B值应结合具体应用场景进行合理选取。应用场景推荐B值表见表2。

表 2 应用场景推荐 B 值表

应用场景类型	特点	推荐范围值
二维分析应用场景	数据维度低, 计算简单	1.5
三维分析应用场景	数据维度增加, 空间关系复杂	$1.5 \leq B \leq 2.0$
遥感影像处理应用场景	数据维度最高, 文件体量大, 数据要求时序性全面, 高空间分辨率和高/多光谱特征	$1.5 \leq B \leq 2.0$

7.2 地理信息容器服务类底座软件

7.2.1 评估数据采集

评估数据采集步骤如下：

- a) 使用方提供高峰时段并发用户数及估算高峰压力时长(需求评估数据采集表样式参见表 A. 1);
 - b) 评估方根据并发用户数及高峰压力时长,在底座软件的原有部署环境(非信创)按照 GB/T 38634. 1—2020 标准要求开展软件压力测试(压测结果表样式参见附录 A. 2. 2);
 - c) 评估方采集时间序列数据集、P90_CPU、P90_内存等数据。

7.2.2 CPU 核数需求评估

CPU采集样本第90百分位位置取值数测算见公式（3）。

$$K = (Q - 1) \times 90\% \quad \dots \dots \dots \quad (3)$$

式中：

K——CPU在压力测试时，所采集的样本排序的第90百分位位置；

Q ——采集样本的总数量；

基于第90百分位位置，使用线性插值法估算百分位数值测算CPU核数需求，见公式（4）。

式中：

C——CPU核数需求；

P_K ——按升序排列后CPU采集样本，第K个采样值，第90百分位整数的数值；

R——冗余系数。

R值应结合具体服务类型及服务特点进行合理选取，冗余系数推荐取值参见表3。

表 3 服务类型及业务场景资源冗余系数推荐表

类型	服务特点	推荐范围值
动态地图服务	实时绘图, 动态变更	1.5
要素地图服务	属性数据交互	1.5
地理处理服务	空间分析计算	$1.5 \leq R \leq 2.0$
影像地图服务	大图加载	$1.5 \leq R \leq 2.0$
缓存地图服务	高并发, I/O集中	2.0
三维地图服务	倾斜摄影、BIM展示	2.0

7.2.3 内存容量需求评估

内存采集样本第90百分位位置取值数测算见公式（5）。

$$L = (Q - 1) \times 90\% \quad (5)$$

式中：

L——内存压力测试时，所采集的样本排序的第90百分位位置；

Q——采集样本的总数量；

基于第90百分位位置，使用线性插值法估算百分位数值测算内存容量需求，见公式（6）。

$$M = [P_L + (P_{L+1} - P_L) \times \{L\}] \times R \quad (6)$$

式中：

M——内存容量需求；

P_L——按升序排列后内存采集样本，第L个采样值，第90百分位整数的数值；

R——冗余系数（推荐取值参见表3）。

7.3 地理信息大模型类底座软件

7.3.1 评估数据采集

评估数据采集步骤如下：

- 使用方提供模型参数量、数据规模和预期时间（需求评估数据采集表样式参见表A.1）；
- 评估方根据模型参数量、数据规模和预期时间，在底座软件的原有部署环境（非信创）按照GB/T 38634.1—2020标准要求开展软件压力测试（压测结果表样式参见附录A.2.3）；
- 评估方采集每参数每样本浮点运算开销、Batch Size、GPU利用率、冗余系数（训练）、推理并发数、单次推理CPU周期、单核每秒处理能力、CPU利用率、冗余系数（推理）、模型加载大小、单次推理内存消耗、内存缓冲预留比例等数据。

7.3.2 GPU 算力需求评估

GPU需求测算见公式（7）。

$$G = \frac{FL \times PAR \times TDS \times BS}{TIME \times GU} \times R \quad (7)$$

式中：

G——GPU算力需求；

FL——每参数每样本浮点运算开销；

PAR——模型参数量；

TDS——训练数据样本数；

BS——每轮迭代一次处理的样本数量；

TIME——训练时间；

GU——GPU利用率；

R——冗余系数（推荐取值参见表3）。

7.3.3 CPU 核数需求评估

CPU核数需求测算见公式（8）。

$$C = \frac{I \times SC}{SP \times CU} \times R \quad (8)$$

式中：

C——CPU核数需求；

I——推理并发数；

SC——单次推理CPU周期；

SP——单核每秒处理能力；

CU——CPU利用率；

R——冗余系数（推荐取值参见表3）。

7.3.4 内存容量需求测算

内存容量需求测算见公式(9)。

式中：

M——内存容量；

MS——模型加载大小；

IC——推理并发数；

SI——单次推理内存消耗;

SR——缓冲预留。

8 评估结果

评估方将测算结果填写至资源推荐表并提交至使用方。各类地理信息底座软件资源推荐表模板参见表A.5。地理信息数据库评估示例参见附录B.1，地理信息容器服务评估示例参见附录B.2，地理信息大模型评估示例参见附录B.3。

附录 A
(资料性)
算力资源需求评估用表

A. 1 需求评估数据采集表

使用方填写的需求评估数据采集表样式见表A. 1。

表 A. 1 需求评估数据采集表

使用方单位名称			
评估对象类型		<input type="checkbox"/> 地理信息数据库 <input type="checkbox"/> 地理信息容器服务 <input type="checkbox"/> 地理信息大模型	
期望完成日期		_____年____月____日	
联系电话		邮箱	
类型	数据项	单位	使用方填写值
地理信息 数据库	预期用户数	个	
	自然资源数据处理复杂度系数	倍	
	预期CPU效率	百分比	
	单核处理能力	cpu-cycles/秒	
	自然资源数据缓存冗余系数	倍	
地理信息 容器服务	服务名称	-	
	并发用户数	人	
	冗余系数	倍	
地理信息大模型算力	模型名称	-	
	模型参数量	个	
	训练数据样本数	条	
	训练时间目标	小时	

填写说明：

- a) 预期用户数：用于推导出并发连接数，并发连接数=预期用户数×并发比例/连接池复用率，如预期用户数为 10000 个，并发比例按 10% 计算，复用率 0.8 计算，可推导出并发连接数为 1250 连接；
- b) 自然资源数据处理复杂度系数：处理三维地质模型、高分辨率遥感影像等自然资源数据的复杂度，推荐值见表 1；
- c) 预期 CPU 效率：业务高峰期 CPU 使用率，建议 75%~85%，或根据属地信息化主管部门出台的政务云资源管理文件要求进行填写；
- d) 单核处理能力：单 CPU 核心每秒处理的周期数，即 CPU 主频，根据属地政务云 CPU 的型号决定；
- e) 自然资源数据缓存冗余系数：为自然资源特有的空间索引数据、元数据等预留的额外缓存空间，推荐值见表 2；
- f) 服务名称：具体需要开展的服务，如某自然资源图层查询服务；
- g) 并发用户数：通过分析历史监控数据（如访问日志、APM 工具等）获取业务高峰时段的并发用户数。如无历史数据，可根据业务总量和活跃用户比例进行估算；
- h) 冗余系数：GJB 1063A—2008 第 3.8 给出的冗余系数最小值取 1.5，推荐值见表 3；
- i) 模型名称：需要进行评估的地理信息大模型名称或版本标识；
- j) 模型参数量：模型总参数数量；
- k) 训练数据样本数：训练使用的数据样本总数量；
- l) 训练时间目标：希望完成训练的时间长度。

A. 2 压测结果样式表

A. 2. 1 地理信息数据库算力资源需求评估数据采集表

评估方通过监控/压测生成，压测结果表样式见表A. 2。

表 A. 2 压测结果表

数据项	单位	评估方填写值
TPMC值 (目标吞吐量)	事务数/每分钟	
单事务平均CPU消耗	cpu-cycles	
热数据集	GiB	
单连接平均内存消耗	MiB	

填写说明：

- a) TPMC 值（目标吞吐量）：每分钟事务数，需使用专门的测试工具，如 BenchmarkSQL 针对 TPC-C 执行事务， $TPMC = \text{New-Order 事务的总数} / \text{测试总时间 (分钟数)}$ ，如果在 60 分钟的测试中，系统处理了 120000 个 New-Order 事务，那么 TPMC 值就是 2000；
- b) 单事务平均 CPU 消耗：单个事务消耗的 CPU 周期数，利用 perf 工具记录事务执行时的 CPU 周期（perf stat -e cpu-cycles,cache-misses ./transaction_script），如一个 New-Order 事务消耗约 15000 CPU 周期；
- c) 热数据集：频繁访问的数据量，压测时通过监控工具获取，PG 类数据库可通过 pg_class 统计表大小，结合访问频率排序计算热数据大小；
- d) 单连接平均内存消耗：每个连接占用的内存，压测时通过监控工具获取，如 PG 类数据库约 200MB/连接。

A. 2. 2 地理信息容器服务算力资源需求评估数据采集表

评估方通过监控/压测生成，压测结果表样式见表A. 3。

表 A. 3 压测结果表

数据项	单位	评估方填写值
数据来源	-	
P90_CPU	核	
P90_内存	GiB	

填写说明：

- a) 数据来源：使用软件压测工具进行生产监控/压力测试的数据，如使用 Prometheus 测试 N 日后记录的时间序列数据集；
- b) P90_CPU：从 N 日监控 CPU 的 90% 分位数据；
- c) P90_内存：从 N 日监控内存的 90% 分位数据。

A. 2. 3 地理信息大模型算力资源需求评估数据采集表

评估方通过监控/压测生成，压测结果表样式见表A. 4。

表 A. 4 压测结果表

数据项	单位	评估方填写值
每参数每样本浮点运算开销	FLOPs/参数/样本	
BS (Batch Size)	条/批	
GPU利用率	比例	
冗余系数（训练）	倍	
推理并发数	个	
单次推理CPU周期	周期数	
单核每秒处理能力	周期数/秒	
CPU利用率	比例	
冗余系数（推理）	倍	
模型加载大小	GiB	

表A.4 压测结果表（续）

数据项	单位	评估方填写值
单次推理内存消耗	GiB	
内存缓冲预留比例	比例	

填写说明：

- a) 每参数每样本浮点运算开销：每个参数处理一个样本所需的大约浮点运算次数；
- b) BS (Batch Size)：每轮迭代一次处理的样本数量；
- c) GPU 利用率：实际 GPU 使用效率，建议 70%~90%，或根据属地信息化主管部门出台的政务云资源管理文件要求进行填写；
- d) 冗余系数（训练）：用于容错及稳定性保障，GJB 1063A—2008 第 3.8 给出的冗余系数最小值取 1.5，推荐值见表 3；
- e) 推理并发数：部署时需同时处理的推理请求数量；
- f) 单次推理 CPU 周期：每次推理所需 CPU 时钟周期；
- g) 单核每秒处理能力：指单个 CPU 核心每秒可完成的时钟周期数，按实际主频 GHz×10⁹换算为周期/秒填写；
- h) CPU 利用率：业务高峰期 CPU 使用率，建议 75%~85%，或根据属地信息化主管部门出台的政务云资源管理文件要求进行填写；
- i) 冗余系数（推理）：推理容错与负载冗余推荐值，GJB 1063A—2008 第 3.8 给出的冗余系数最小值取 1.5，推荐值见表 3；
- j) 模型加载大小：FP32 精度时，每 1.2e8 (1.2×10⁸) 参数模型约 0.48GB；
- k) 单次推理内存消耗：推理中间张量、缓存等；
- l) 内存缓冲预留比例：总内存的预留比例，建议 20%~30%。

A.3 评估结果样式表

评估结果资源推荐表见表A.5。

表 A.5 评估结果资源推荐表

评估方名称			
评估对象类型		<input type="checkbox"/> 地理信息数据库 <input type="checkbox"/> 地理信息容器服务 <input type="checkbox"/> 地理信息大模型	
完成日期		____年____月____日	
联系电话		邮箱	
评估对象类型	数据项	单位	推荐值
地理信息	推荐CPU核数需求	核	
数据库	推荐内存容量需求	GiB	
地理信息	推荐CPU核数需求	核	
容器服务	推荐内存容量需求	GiB	
	推荐容器规格	-	
地理信息大模型算力	推荐GPU算力	Flops	
	推荐CPU核数	核	
	推荐内存总需求	GiB	

填写说明：

- a) 推荐 CPU 核数需求：根据本文件 7.1.2 或 7.2.2 计算公式所得出的 CPU 核数，应依据属地政务信息化主管部门制定的政务云资源目录进行匹配。如计算得出 CPU 核数需求为 1.455 核，则应向上取整为 2 核；
- b) 推荐内存容量需求：根据本文件 7.1.3 或 7.2.3 计算公式所得出的内存容量需求，应依据属地政务信息化主管部门制定的政务云资源目录进行匹配。如计算得出内存容量需求为 1.86GiB，则应向上匹配到 2GiB；
- c) 推荐容器规格：应依据属地政务信息化主管部门制定的政务云资源目录匹配可供参考的容器或实例配置；
- d) 推荐 GPU 算力需求：根据本文件 7.3.2 计算公式所得出的 GPU 算力。

附录 B
(资料性)
算力资源实施评估示例

B. 1 地理信息数据库评估示例

B. 1. 1 示例说明

本评估示例均基于地理信息数据库在信创环境中的实际压力测试数据，并按照以下计算方法，结合服务性能和资源消耗情况进行推荐资源配置估算。

B. 1. 2 调研业务需求

用户根据业务实际需求填写使用方填写表，示例见表B. 1。

表 B. 1 地理信息数据库使用方填写示例

数据项	单位	填写值	说明
预期用户数	个	5000	最大并发用户数，计算公式：预期用户数×并发比例/连接池复用率。
并发比例	百分比	10	并发连接数为：5000用户×10%并发/0.8复用率≈625连接。
连接池复用率	百分比	80	
自然资源数据处理复杂度系数	倍	1. 5	处理三维地质模型、高分辨率遥感影像等自然资源数据的复杂度，推荐值见表1。
预期CPU效率	百分比	80	业务高峰期CPU使用率，建议75%~85%，或根据属地信息化主管部门出台的政务云资源管理文件要求进行填写。
单核处理能力	cpu-cycles/秒	25亿	单CPU核心每秒处理的周期数，即CPU主频：单位Hz（每秒周期数），根据属地政务云CPU的型号决定。 如CPU主频：2.5 GHz，主频转换：2.5 GHz=2.5×10 ⁹ Hz=25亿周期/秒。
自然资源数据缓存冗余系数	倍	1. 5	为自然资源特有的空间索引数据、元数据等预留的额外缓存空间，推荐值见表2。

B. 1. 3 评估数据采集

评估方技术团队通过监控/压测采集，示例见表B. 2。

表 B. 2 地理信息类数据库压测结果表

数据项	单位	采集值	说明
TPMC值 (目标吞吐量)	事务数/每分钟	1,800,000	每分钟事务数，需使用专门的测试工具，如BenchmarkSQL针对TPC-C执行事务，TPMC = New-Order事务的总数/测试总时间(分钟数)
单事务平均CPU消耗	cpu-cycles	15,000	单个事务消耗的CPU周期数，利用perf工具记录事务执行时的CPU周期：perf stat -e cpu-cycles, cache-misses ./transaction_script
热数据集	GiB	36	频繁访问的数据量，压测时通过监控工具获取
单连接平均内存消耗	MiB	200	每个连接占用的内存，压测时通过监控工具获取

B. 1. 4 实施评估

B. 1. 4. 1 CPU 核数需求

理论值：(TPMC值×单事务平均CPU消耗)/(单核处理能力×预期CPU效率) = (1,800,000×15,000)/(25 亿×0.8)=13.5 核

数据复杂度处理后（自然资源数据处理复杂度系数=1.5）：13.5×1.5≈20 核

B. 1.4.2 内存容量需求

并发连接数=预期用户数×并发比例/连接池复用率=5,000×0.1÷0.8=625

自然资源数据缓存冗余系数: 1.5

总内存=(热数据集×1.5+并发连接数×单连接平均内存消耗)×自然资源数据缓存冗余系数=(36 GB×1.5+625×0.2 GB)×1.5), 通过计算得: 总内存≈268 GB

B. 2 地理信息容器服务评估示例

本章节评估示例均基于地理信息类服务在信创容器环境中的实际压力测试数据, 并按照以下计算方法, 结合服务性能和资源消耗情况进行推荐资源配置评估。

资源推荐计算公式如下:

推荐CPU配置=P90_CPU×冗余系数

推荐内存配置=P90_内存×冗余系数

所有测试服务均部署为单容器服务, 通过容器方式运行, 运行环境为国产化架构。

注: 以下评估示例所用的数据均基于实际项目测试场景合理估算, 暂未结合连续7日以上的P90监控指标。真实项目中, 应结合 Prometheus 或容器平台中采集的7日以上指标数据进行填报。

示例一: 动态地图服务(轻负载)

动态地图服务是一种通过标准HTTP接口, 根据客户端请求实时生成并返回地理参考地图图像的网络服务。该服务通过定义地理图层与空间范围来动态处理数据, 其响应为可在浏览器中显示的栅格图像(如JPEG、PNG格式), 并支持透明背景等特性, 便于来自多个服务器的图层叠加。

资源配置	用户并发数	压测时长	平均响应时间	冗余系数	P90_CPU	推荐CPU	P90内存	推荐内存
4C8G	10	10 min	455 ms	1.5	0.92 核	1.4 核	2.14 G	3.2 G

结论: 适用于并发不高、数据量适中服务场景, 可选用 2C4G 配置部署。

示例二: 缓存地图服务(中等负载)

缓存地图服务是一种使用预渲染、预切分的地图图片瓦片来提供地图可视化的网络服务。该服务遵循标准化的瓦片矩阵集, 客户端通过标准化接口请求特定级别、行列号的瓦片, 从而实现高效的地图浏览和高并发访问。

资源配置	用户并发数	压测时长	平均响应时间	冗余系数	P90_CPU	推荐CPU	P90内存	推荐内存
8C16G	100	10 min	386 ms	2.0	3.12 核	6.2 核	4.17 G	8.3 G

结论: 该配置下可支撑中小型并发的缓存地图查询服务, 建议使用 7C9G 规格部署, 满足响应性能与资源性价比平衡。

示例三: 要素地图服务(资源紧张型)

要素地图服务是一种支持对地理要素进行细粒度访问与编辑的网络服务。它允许客户端执行要素的查询、锁定及事务(创建、修改、删除)操作, 传输的是要素数据本身而非地图图像。

资源配置	用户并发数	压测时长	平均响应时间	冗余系数	P90_CPU	推荐CPU	P90内存	推荐内存
2C2G	10	10 min	553 ms	1.5	0.97 核	1.5 核	1.21 G	1.8 G

结论: 尽管2核能支撑10用户并发数, 但资源接近瓶颈, 建议升级至 2C4G 或以上配置以提升稳定性。

B. 3 地理信息大模型评估示例

本章节评估基于大模型在信创算力平台上的示例测试数据, 结合模型训练与推理全过程的资源使用情况, 采用统一指标体系进行推荐配置评估。

资源推荐计算公式如下:

推荐GPU卡数=(每参数每样本浮点运算开销×模型参数量×样本数×Batch Size)÷(训练时间×GPU利用率)×冗余系数

推荐CPU核数=(推理并发数×单次推理CPU周期)÷(单核每秒处理能力×CPU利用率)×冗余系数

推荐内存容量=模型加载大小+推理并发数×单次推理内存消耗+缓冲预留

测试环境采用国产化架构，模型训练与部署均在信创平台上完成，资源配置建议结合标准计算公式和业务场景压力测试结果进行核算。示例见表B.3。

表 B.3 遥感目标识别大模型表

参数项	示例值
模型名称	RS-TargetNet-L
每参数每样本浮点运算开销	3000 FLOPs/参数/样本
模型参数量	5×10^8 个
训练数据样本数	1×10^7 条
Batch Size	256 条/批
训练时间目标	72 小时
GPU利用率	0.8
冗余系数（训练）	1.5倍
推荐GPU算力	≈ 0.1 PFLOPs
推理并发数	10 个
单次推理CPU周期	2×10^7 周期
单核处理能力	3×10^9 周期/秒
CPU利用率	0.8
冗余系数（推理）	1.5倍
推荐CPU核数	≈ 1 核
模型加载大小	2.0 GiB
单次推理内存消耗	0.2 GiB
内存缓冲预留比例	0.25
推荐内存容量	≈ 5.0 GiB

结论：针对遥感目标识别大模型场景，根据计算结果，本模型在训练阶段单次训练至少需要约0.1PFLOPS GPU算力，推理阶段资源需求较轻，总体需配置约1核CPU和5GiB内存，可支持训练与小规模推理任务的并行执行。但在资源紧张或长时间高负载运行场景下，该配置已接近瓶颈，建议升级为：12GPU+2C8G，以提升系统稳定性、任务调度效率与整体服务保障能力。

参 考 文 献

- [1] GB/T 17694—2023 地理信息 术语
 - [2] GB/T 39623—2020 基础地理信息数据库系统质量测试与评价
 - [3] T/EGAG 019—2023 信息技术应用创新项目验收规范
-