

# 团 体 标 准

T/CESA 1213—2022

## 通用计算 CPU 性能测试评价技术要求

Technical requirements for evaluation of general computing CPU performance test

2022-07-21 发布

2022-08-20 实施





版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

# 目 次

前言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	2
5 通用要求.....	2
5.1 可重复性.....	2
5.2 跨平台兼容性.....	2
5.3 结果合理性.....	2
5.4 扩展性.....	2
5.5 覆盖度.....	2
5.6 易用性.....	3
5.7 业务代表性.....	3
5.8 公平性.....	3
5.9 稳定性.....	3
6 运行规则.....	3
6.1 规则文档.....	3
6.2 连续运行.....	3
6.3 可获取性.....	3
6.4 结果输出.....	3
6.5 库函数替换.....	3
7 软件框架.....	4
7.1 运行.....	4
7.2 报告生成.....	4
8 基准负载.....	4
8.1 编译器支持.....	4
8.2 负载选择.....	4
9 运行环境.....	5
9.1 硬件.....	5
9.2 软件.....	5
10 工具应用.....	5

## 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。

本文件由中国电子技术标准化研究院提出。

本文件由中国电子技术标准化研究院、中国电子工业标准化技术协会归口。

本文件起草单位：中国电子技术标准化研究院、华为技术有限公司、飞腾信息技术有限公司、龙芯中科技术股份有限公司、上海兆芯集成电路有限公司、海光信息技术股份有限公司、无锡先进技术研究院、中国科学院计算技术研究所、中国人民解放军国防科技大学、中国移动通信有限公司研究院、北京计算机技术及应用研究所、浪潮电子信息产业股份有限公司、联想开天科技有限公司。

本文件主要起草人：钟伟军、逯海涛、任翔、尹航、李阳、葛小宇、包振忠、吴彤、王昭飞、靳国杰、李超、齐宗普、李红茜、从日本、王思博、刘杰、陈玉波、肖俊华、王鑫、王永文、邓林、陈佳媛、王怡惠、杨帆、孙大东、吕东波、牟茜、李伟招、王宁。



# 通用计算 CPU 性能测试评价技术要求

## 1 范围

本文件规定了通用计算CPU性能测试评价的通用要求，以及运行规则、软件框架、基准负载、运行环境及工具应用。

本文件适用于确定和评价通用计算CPU性能测试基准及CPU产品的设计、开发和使用。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### **基准测试 benchmark test**

基准测试是指通过设计科学的测试方法、测试工具和测试系统，实现对一类测试对象的某项性能指标进行定量的和可对比的测试。

### 3.2

#### **性能基准 performance benchmark**

基准测试时用于建立某个时刻的性能标准参考值，标定被测试硬件系统的性能指标。

### 3.3

#### **测试负载 test workload**

修改自实际应用软件，用于测试CPU性能的测试用例代码。

### 3.4

#### **负载套件 workload suites**

基准软件中用于特定测试模式的一组测试负载集合。

### 3.5

#### **单任务模式 single task model**

以运行时间作为评估指标，用来衡量单核运算性能的CPU执行模式。

### 3.6

#### **多任务模式 multiple task model**

以吞吐量作为评估指标，用来衡量多核运算性能的CPU执行模式。

### 3.7

#### **典型性能模式 typical performance model**

采用非定制优化措施的性能测试模式，同一种编程语言不同负载的性能相关的编译参数或优化措施必须一致。

### 3.8

#### 极限性能模式 extreme performance model

采用定制优化措施的性能测试模式,同一种编程语言不同负载的性能相关的编译参数或优化措施可以不同。

## 4 缩略语

下列缩略语适用于本文件。

AOCC: AMD C/C++优化编译器(AMD Optimized C/C++ Compiler)

CPU: 中央处理器(Central Processing Unit)

GCC: GNU编译器套装(GNU Compiler Collection)

ICC: Intel C++编译器(Intel C++ Compiler)

JDK: Java开发套件(Java Development Kit)

LLVM: 底层虚拟机(Low Level Virtual Machine)

## 5 通用要求

### 5.1 可重复性

在相同的测试条件下(测试环境信息、编译器工具链信息以及调优措施等),任一测试套件多次测试应有近似的测试结果,误差应不大于5%。

### 5.2 跨平台兼容性

跨平台兼容性要求为:

- a) 应支持不同的CPU指令集架构,如x86\_64、AArch64、LoongArch、SW\_64等;
- b) 应支持不同的操作系统,如Kylin、UOS、openEuler等;
- c) 应支持不同的编译器,如GCC、LLVM、ICC、AOCC等,C编译器要求支持C99及以上,C++编译器要求支持C++14及以上;
- d) 应兼容JDK1.8及以上版本和Python3.6及以上版本。

### 5.3 结果合理性

结果合理性要求为:

- a) 测试结果应与实际产品性能预期一致;
- b) 测试结果应能体现不同产品的性能情况;
- c) 测试结果应能匹配用户实际性能体验。

### 5.4 扩展性

扩展性要求为:

- a) 应支持负载增加、修改及删减;
- b) 应支持功能模块增加、修改及删减;
- c) 应支持数据集的增加、修改及删减。

### 5.5 覆盖度要求

覆盖度要求为：

- a) 应包含整型和浮点型（单精度、双精度）运算；
- b) 应支持 C/C++、Fortran、Java 编程语言；
- c) 测试内容应正交（体现指令解码、缓存访问、分支预测等不同方面的性能），避免过于强调某一方面的性能；
- d) 测试模式应支持单任务与多任务并发测试；
- e) 应能体现不同典型应用场景的特征；
- f) 应至少提供一种数据集，为满足不同测试目的，可提供不同规模数据集。

## 5.6 易用性

基准工具应支持单机运行，运行依赖少，运行时长适中，方便快速搭建测试环境和获取测试结果。

## 5.7 业务代表性

所采用的测试负载应来自于真实业务，能够代表通用服务器计算的 actual 业务场景。如编译、机器学习、数字媒体、数据库、大数据和高性能计算等典型的计算业务场景。

## 5.8 公平性

为了确保跨架构平台公平，应满足如下要求：

- a) 性能测试相关代码，禁止使用汇编代码实现；
- b) 消除锁机制对性能结果的影响。

## 5.9 稳定性

应运行稳定，无异常闪退、中断等不可用的情况。

## 6 运行规则

### 6.1 规则文档

应有明确的运行规则及配套支撑的中文文档。

### 6.2 连续运行

连续运行要求为：

- a) 同一个测试套件中的测试负载必须连续运行完成，不同的测试套件可以分开进行测试；
- b) 测试套件开始运行至结束运行期间，禁止修改系统参数。

### 6.3 可获取性

可获取性要求为：

优化措施应面向商用且通用，不应使用只对测试负载有效而对实际业务场景无效的优化措施；测试所使用的操作系统、编译器 etc 等硬件必须已正式发布，用户可通过公开渠道获取。

### 6.4 结果输出

运行输出结果应可视化、可验证且可比较。

### 6.5 库函数替换

库函数要求为：

- a) 不能通过特殊手段来替换性能基准中的实现函数，从而导致运行的代码非性能基准默认实现，获得不公平的测试分数。比如，使用自定义的 API 替换性能基准中实现的函数；
- b) 对于优化后的库函数，不能只在一个测试负载中使用，优化的库函数不能预置计算好的计算结果；
- c) 对于 ANSI/ISO 编程语言规范中提供的可选优化库或函数，允许通过编译选项启用。

## 7 软件框架要求

### 7.1 运行要求

CPU性能基准软件框架运行时应支持：

- a) 统一配置管理；
- b) 统一入口传参；
- c) 统一输出结果；
- d) 负载编译、构建和运行；
- e) 测试套件灵活组合；
- f) 离线和在线安装；
- g) 工具依赖的检测。

### 7.2 报告生成

在符合以下要求时，输出测试结果并生成报告：

- a) 对负载文件完整性进行校验；
- b) 对基准工具自身功能模块完整性进行校验；
- c) 对测试负载运行结果的正确性验证；
- d) 支持测试报告数字签名；
- e) 报告呈现的结果应明了简洁，便于识别；
- f) 支持 CSV、PDF 等不同格式的测试结果生成文件；
- g) 支持中文、英文测试报告生成。

## 8 基准负载

### 8.1 编译器支持

编译器支持度应符合以下要求：

- a) 支持业界通用的编译器，如 GCC、CLANG、AOCC 等；
- b) 支持业界通用编译器的通用版本，如 GCC7.3、GCC9.1、CLANG11.0.0、AOCC3.0.0 等。

### 8.2 负载选择

负载选择符合以下要求：

- a) 测试负载应包含整型计算负载和浮点运算负载；
- b) 测试负载应来源于实际的业务场景，应包括但不限于如下场景，整型应包括大数据、数据库、编译器、压缩、加解密、视频编解码等负载；浮点型应包括流体力学、分子动力学、基因工程、气动声学、电磁学、气象预测、天体物理等负载；

c) 负载设计应满足《通用计算 CPU 性能测试基准负载设计要求》要求。

## 9 运行环境

### 9.1 硬件

性能基准在单任务模式运行时至少需要2GB内存容量和20GB硬盘容量，在多任务模式运行时至少需要（任务数×2GB）内存容量和（任务数×20GB）硬盘容量。

### 9.2 软件

软件要求为：

- a) 基准工具应包含运行所需的基本依赖工具；
- b) 基准工具应以源代码方式向用户提供。

## 10 工具应用

应基于符合本文件要求的基准工具进行性能测试，如通用计算CPU性能测试基准工具CPUBench等，并提供工具生成的有效测试报告。

