



中华人民共和国国家标准

GB/T 30975—2014/ISO/IEC 14756:1999

信息技术 基于计算机的 软件系统的性能测量与评级

Information technology—Measurement and rating of
performance of computer-based software systems

(ISO/IEC 14756:1999, IDT)

2014-09-03 发布

2015-02-01 实施

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

目 次

前言	III
引言	IV
1 范围	1
2 符合性	2
3 规范性引用文件	3
4 术语和定义	3
5 缩略语和符号	7
5.1 缩略语	7
5.2 符号	7
6 测量	8
6.1 配置需求	8
6.2 用户模拟	9
6.3 测量规程	10
6.4 测量有效性的证明	11
7 SUT 性能值的计算	12
7.1 平均执行时间	12
7.2 吞吐量	12
7.3 适时吞吐量	12
8 用于评级的基础数据	12
8.1 用户需求	12
8.2 用于软件效率评级的基准环境	12
9 性能值评级	13
9.1 计算性能参考值	13
9.2 计算性能评级值	13
9.3 SUT 总体性能评级	13
9.4 性能评估	13
10 输入需求	14
10.1 SUT 描述	14
10.2 工作负载参数集	15
10.3 用于测量确认的输入	17
11 测量	18
11.1 测量规程	18
11.2 单个评级区间	19
12 测量规程的输出	20
12.1 测量日志文件	20
12.2 计算结果文件	20

13	测量确认	20
13.1	SUT 计算正确性确认	21
13.2	RTE 准确度确认	21
13.3	测得的平均执行时间的统计显著性确认	21
14	SUT 性能值计算	22
14.1	平均执行时间	22
14.2	吞吐量	22
14.3	适时吞吐量	22
15	SUT 测得性能值的评级	23
15.1	评级级别规格说明	23
15.2	计算性能参考值	23
15.3	计算评级值	23
15.4	评级	24
附录 A (规范性附录)	RTE 基本功能规格说明	26
附录 B (规范性附录)	附加计算公式	27
附录 C (规范性附录)	工作负载描述格式	33
附录 D (规范性附录)	日志文件格式	36
附录 E (资料性附录)	实用程序	37
附录 F (资料性附录)	工作负载示例	39

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准使用翻译法等同采用 ISO/IEC 14756:1999《信息技术 基于计算机的软件系统的性能测量与评级》。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位：北京邮电大学、中国电子技术标准化研究院、山东省计算中心、山东省计算机网络重点实验室。

本标准主要起草人：史慧玲、杨美红、李刚、韩明军、邹丰义、周鸣乐、韩庆良、顾卫东、董火民、韩红强、刘建毅、胡宇、高翊、王枫。

引 言

在设计和使用数据处理系统时,执行速度是一个重要的性质。该性质主要受系统中所使用软件的效率影响。测量软件系统执行速度,以及软件效率影响,是最关心的问题。

为了测量软件对一个数据处理系统在时变行为方面的影响,有必要测量整个系统的时变行为。基于本标准建议的测量规程,可定义和计算软件时间效率的数值。

时变行为可重复估算是非常重要的。在实验中使用真实用户是不可能的,一个原因是真实用户不可能像计算机那样在没有偏差的前提下重复执行操作,另一个原因是当工作或任务流是来自很多用户时,若由真实用户来执行这样的实验,需要花费非常高的代价。因此,在测试中需要采用模拟器,该模拟器通过一个辅助数据处理系统来模拟多用户操作。

这意味着要根据本标准进行性能测量和评级,需要借助工具。该工具应是按照本标准规范要求来运行的模拟器。应证明实际使用的模拟器满足这些规范。

所有相关实验细节都通过用户模拟器记录在日志文件中。通过该日志文件,可计算出描述时变行为的值(如响应时间和吞吐量的值)。根据这些性能值,可计算出软件效率评级值。

并不是所有性能值都需要进行测量和评级。例如,对于一个具有简单负载的软件系统,只有很少的交互任务,或者只有很小批量的执行要求,则只需要定义部分性能指标及其值。该方法也适用于测量和评级大的、复杂的基于计算机的软件系统(CBSS),CBSS可处理由一系列不同用户产生的复杂 workflow 或任务流。定义中有必要包括一些数学术语,目的是为了计算性能值和评级值、校验测量运行和评级步骤的正确性,以及为了确定性能值和评级结果的(统计上的)意义获取精确的数学依据。

测量结果由计算的绩效值组成,包括吞吐量和执行时间的值。CBSS 性能评估的最终结果由评级值构成。评级值可通过比较绩效值与用户需求获得。另外,如果需要,通过与一个可参考的 CBSS(如,具有相同硬件配置,但却具有相同功能的应用程序的不同版本)绩效值相比较,可对待测 CBSS 绩效值进行评级。

评级结果是一组值,每个值或者大于 1,或者小于 1,或者等于 1,分别表示性能好于、坏于、等于定义的需求(或者是另一个作为参考的待测系统特性)。最终的一组评级值评估工作负载中分别定义的各种任务类型。

附录 E 和附录 F 包含软件以及特殊的数据,因为不能印刷,故本标准未提供。在这两个附录中都提供了简短的概述。

信息技术 基于计算机的 软件系统的性能测量与评级

1 范围

本标准界定了对于基于计算机的软件系统(CBSS)面向用户的性能如何测量和评级。从用户(例如各种不同终端的用户,或者数据处理中心的操作用户和商业用户)角度看,CBSS就是一种数据处理系统。

CBSS包括硬件及其所有软件(系统软件和应用软件),这些是实施用户所需的数据处理功能所要求的,或可影响到CBSS的时变行为。

本标准适用于测试所有受时间约束的系统或系统的一部分。同样,网络可以是某个系统的组成部分,或者可成为测试的主题。本标准所界定的方法不限于像经典的批处理或终端-宿主系统等特例,例如,也包括客户-服务器系统或按广义理解对“任务”界定的实时系统。但是就测试大环境所需花费来说,测试的实用性会受到限制。

本标准规定了面向用户的性能术语的关键指标,以及测量和评级这些性能值的方法。所规定的性能值描述了用户命令(任务)的执行速度的性能值,即如下“三位一体”的:

- 执行时间;
- 吞吐量;
- 适时性。

用户命令,以下称为任务,在内部结构上可以很简单,也可以很复杂。一项任务可能是一项作业、事务、过程,或更为复杂的结构,但都根据评价者的需要界定一个启动时间和结束时间。进行性能评价时,除单个响应时间外,还可参考商务事务处理完成时间,使用本标准来测量系统的时变行为。

评级可根据用户需求来完成,或者通过与两个或更多的已测量系统(类型或版本)相比较来完成。

对于测量内部值,例如:

- 利用值;
- 平均指令速率;
- 路径长度;
- 高速缓存命中率;
- 排队时间;
- 服务时间。

特意不给出建议,原因是内部值的界定取决于在测系统的硬软件体系结构。与此相反,本标准界定的面向用户性能值则独立于体系结构。对内部性能值的界定,可独立于对面向用户性能值的界定。除面向用户的性能值外,这些内部性能值可以使用,也能测量。同样,产生面向用户值的效率的术语也能自由定义。另外,本标准对如何在数据处理系统中建立稳定、可再现的操作状态给出了指导,这些可再现状态可用来测量其他性能值,例如上面提到的内部值。

本标准重点在于:

- 应用软件;
- 系统软件;
- 交钥匙系统(即由应用软件、系统软件以及为此设计的硬件所组成的系统);
- 通用数据处理系统。

本标准规定了有关模拟[通过技术系统,即所谓远程终端模拟器(RTE)]用户与数据处理系统交互