

## Challenges in capturing real World Workloads into Benchmarks

[www.ece.utexas.edu/~ljohn/wwc/wwc7](http://www.ece.utexas.edu/~ljohn/wwc/wwc7)

Renaissance Hotel

Austin, TX

October 25, 2004

**“Reputation of current bench marketing claims regarding system performance is on par with the promises made by politicians during elections”**



**Kaivalya M. Dixit**

**Chief Cook & Bottlewasher**  
***dixit@spec.org***

# Benchmarking Evils

Workloads Metrics Scalability Configuration Disclosure

R  
e  
v  
e  
n  
a  
n  
c  
e



© UFS, Inc.

F  
i  
r  
s  
t

Pace of Technology vs. Pace of Benchmark Development

Complexity



## Benchmarks & Workloads

Dhrystone

Whetstones

Linpack

STREAM

Webstones

SPECmark

MIPS

MOPS

FLOPS

BOPS

TOPS

JOPS

tpmC

FORTRAN

C/C++/C#

Java

Perl

PHP

TPC C

TPC D

TPC H

TPC H

SPEC {int,fp}

SPECweb

SPECsfs

SPECjava

SPECgpc

SPECchpg

Oracle Apps

SAP

## Towers of Babel

The Good, the Bad, the Ugly, and Popular

The Good, the Bad, the Ugly, and Popular

Zillions of Application Benchmarks

netbench

iobench

## Benchmarks & Workloads

- **Application Benchmarks**
  - ▶ **An EASY answer from a user perspective is "yes"**
  - ▶ **The CORRECT answer is "yes but inadequate"**
    - ◆ **Perormance spectrum is broad and deep**
    - ◆ **Many choices left on table**
    - ◆ **Too many applications**
    - ◆ **Industry and market dynamics**
  - ▶ **Challenges in application benchmarks**
    - ◆ **Workloads are treated like family jewels**
    - ◆ **Scalability across Configurations & Architectures**
    - ◆ **A'priori Knowledge Bias**
- **Ideal Benchmark in my opinion - an ILLUSION**
- **Good ones are hard to find and/or create**
  - ▶ **SPEC, TPC, SPC, EMBCC, Bapco, et al. working**

## Benchmark Development – standards organizations

- **The answer is an unqualified YES if you NEED:**
  - ▶ **to reduce the competitive benchmarking costs**
  - ▶ **to reproduce the competitor's results**
  - ▶ **to successfully challenge the competitor's results**
  - ▶ **to CREATE benchmarks scrutinized by all Performance Pundits (e.g., industry, academia, ...)**
  - ▶ **full disclosure, rigorous specifications, controlled source code , run and reporting rules and fair use guide lines to minimize BENCHMARKETING abuse**
  - ▶ **benchmarks CREATED by some of the BEST PERFORMANCE Gurus in the industry & academia**
  - ▶ **to IMPROVE your emerging hardware & software**
  - ▶ **benchmarks that keep up with fast changing technology**
  - ▶ **architects, researchers and performance pundits to complain about how inadequate they are**

## KNOWN PROBLEMS

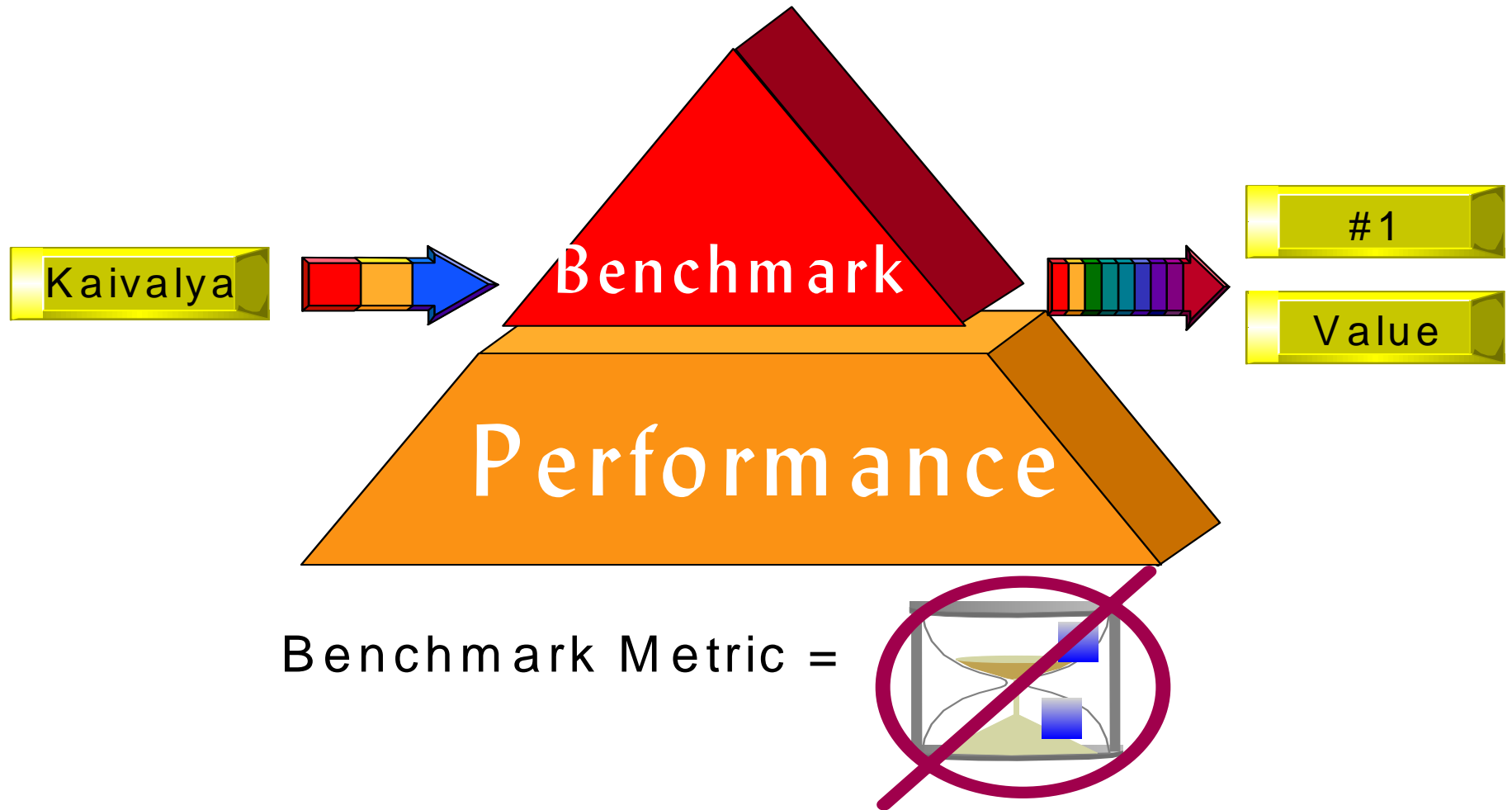
- ◆ Age of benchmarks & workloads
  - ◆ SPEC – 2 to 5 years
  - ◆ TPC – 2 to 15 years
- ◆ Confusion re: Speed vs. Throughput
- ◆ Orthogonal interests
  - ◆ Performance, Scalability, Open Source, Pricing
  - ◆ Diverse interests of Chip, H/W, S/W, System vendors
  - ◆ Cost of benchmarking
- ◆ Yesterday's workloads used in designing tomorrow's architectures (chip, h/w, s/w, compilers, ..)
- ◆ Technology is MOVING too FAST

## What is SPEC doing about this?

- ◆ SPECjAppServer2004
  - ◆ 34 digit decimals in database
- ◆ Next release of HPC suite
  - ◆ Games with relevant workloads (e.g., taxol, penicillin, ..)
- ◆ CPU2005 Grant Program
  - ◆ challenge to academia & industry for better benchmarks and workloads
  - ◆ Strong emphasis on analysis, portability, run rules.
- ◆ Next release of SPECweb
  - ◆ eCommerce, eBanking , eSupport workloads, QoS, ...
  - ◆ SPECcapc uses application workloads

Participation from members, academia & pundits

## Kaivalya & Other User's Performance Needs



## Credits

Thank You	Thank You
Dr. Men-chow	IBM Corp
John Stecher	IBM Corp
John Makis	IBM Corp
Dr. Erich Nahum	IBM Corp
SPEC	<a href="http://www.spec.org">http://www.spec.org</a>
Lizy Kurian John	University of Texas

## Real Life Events to compare SPECweb99

### Sample Logs for Illustration

Name:	Chess 1997	Olympics 1998	IBM 1998	World Cup 1998	Dept. Store 2000	IBM 2001
Description:	Kasparov- Deep Blue Event Site	Nagano 1998 Olympics Event Site	Corporate Presence	Sporting Event Site	Online Shopping	Corporate Presence
Period:	2 weeks in May 1997	2 days in Feb 1998	1 day in June 1998	31 days in Jun-Jul 1998	12 days in June 2000	1 day in Feb 2001
Hits:	1,586,667	5,800,000	11,485,600	1,111,970,278	13,169,361	12,445,739
Bytes:	14,171,711	10,515,507	54,697,108	54,697,108	54,697,108	28,804,852
Clients:	256,382	80,921	86,021	2,240,639	86,021	319,698
URLS:	2,293	30,465	15,788	89,997	15,788	42,874

We'll use statistics generated from these logs as examples.

## Author's view of Reality vs. SPECweb99

### Summary and Conclusions

- SPECweb99 has a mixed record depending on characteristic:
  - Methods: OK
  - Response codes: bad
  - Document popularity: good
  - File sizes: OK to bad
  - Transfer sizes: bad
  - Inter-arrival times: good
- Main problems are:
  - Needs to capture conditional GETs with IMS for 304's
  - Better file size distribution (smoother, larger)