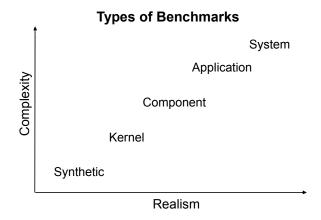# Building A Great Benchmark

## What is a benchmark?

- A program (or suite of programs) used to assess the performance characteristics of one or more computer systems

## Why do we need benchmarks?

- Measure and study the behavior of a class of applications (or systems) in a practical and repeatable way

## What are they used for?

- Marketing
- Product development
- Quality assurance
- Environmental verification
- Research

**Types of Benchmarks**



## What characteristics should they share?

- **Relevance**
  - They should mimic the behavior of some class of real applications, without being tied to or based on just one
- **Reproducibility**
  - They should produce results that can be reproduced by others deterministically given the same hardware
- **Fairness**
  - They enable competition between systems without imposing artificial constraints or bias
- **Verifiability**
  - All results produced can be verified to be accurate
- **Usability**
  - Ease of use encourages greater use, leading to more results and better accuracy

## What components should they include?

- Workload(s)
- Harness: hosts the workloads and handles environment-specific configuration and UI
- Reporting: human and / or machine-readable
- Documentation
- Run rules: how should the benchmark be use
- Peer review: any and all results should be available for review by competitors and customers
- Fair Use guidelines: under what circumstances and with what caveats may the results be used or shared