

Industry Standard benchmarks: Dos and Don'ts

*H. Reza Taheri, VMware, Inc.
rtaheri@vmware.com*

-
- **25 papers in VLDB 2010 with TPC results**
 - None audited, many were disputed
 - Much liberty taken in how benchmarks were run
 - **Industry standard benchmarks are useful only if everyone follows the rules**
 - **Let's learn from the experience of the past 30 years**

TPC

- + Longevity: TPC-C has carried the load for 20 years**
- + Comparability**
 - Audit requirements and strict detailed run rules mean one can compare results published by two different entities
- + Scaling**
 - Results just as meaningful at the high-end of the market as at the low-end; as relevant on clusters as on single servers
- Hard and expensive to run**
- No kit**
- DeWitt clauses**

SPEC

- + Kit provided
- + Benchmarks updated frequently
- + Many benchmarks
- + Base and Peak modes
- + Use of publically available and open source software
 - No DeWitt clause
- Not much in database benchmarks
- Strength in single system or even single component benchmark

Vendor-provided benchmarks

- **Examples: SAP SD, VMmark**
- + **Fast development process**
- + **Simple run rules**
- + **Kit**
- + **Built-in auditor**
- + **Does one thing, and does it well**

- **Does one thing**
- **Have to trust that one company**

Dos and Don'ts

- Provide a kit
- Base run against open source, prescribed software in *Base* mode; allow *Optimized* mode
- Don't try to develop a single benchmark to cover everything
 - Do one thing; do it well
 - Multiple benchmarks
- Scaling is paramount
- Ability to compare results from different entities also paramount
 - Motivate people to run the official version
- Limit academic use to new projects
- End of life