

# Industry Standard Benchmark Development

Meikel Poess

# Benchmark Categories

- Industry standard benchmarks
  - TPC: Transaction Processing Performance Council
  - SPEC: Standard Performance Evaluation Corporation
  - SPC: Storage Performance Council
- Application benchmarks
  - SAP's suite of standard application benchmarks
    - System and Server Power, Business Intelligence, ....
  - Similar Oracle Apps, JD Edwards, Exchange, Domino
- Special purpose benchmarks
  - Dhrystone, Whetstone, Linpak, Izone, Netperf, Stream

# TPC: Transaction Processing Performance Council

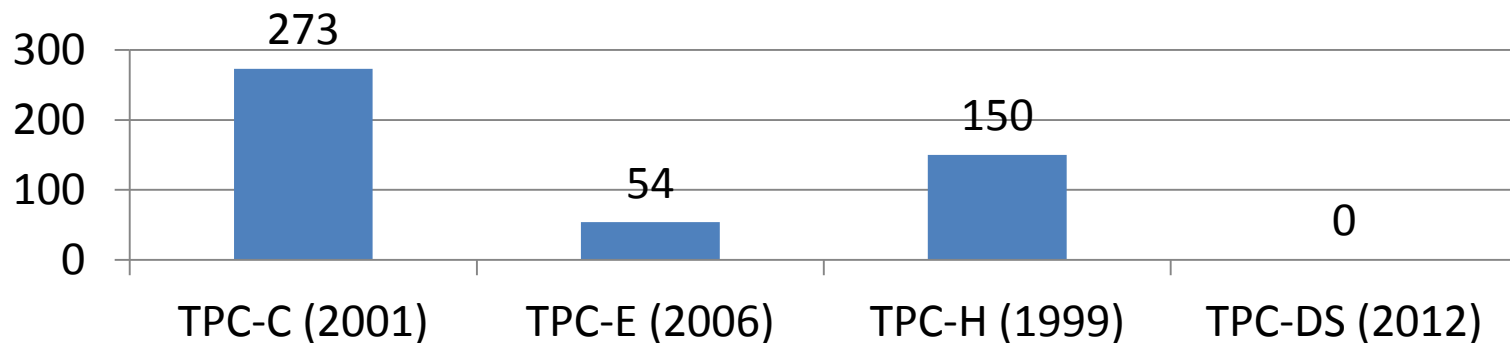
## Existing Benchmark Suite

- OLTP: TPC-C, TPC-E
- BI: TPC-H, TPC-DS
- Applicable to all above
  - Energy, Pricing

## Benchmarks under Development

- Virtualization
- Data Integration (ETL)

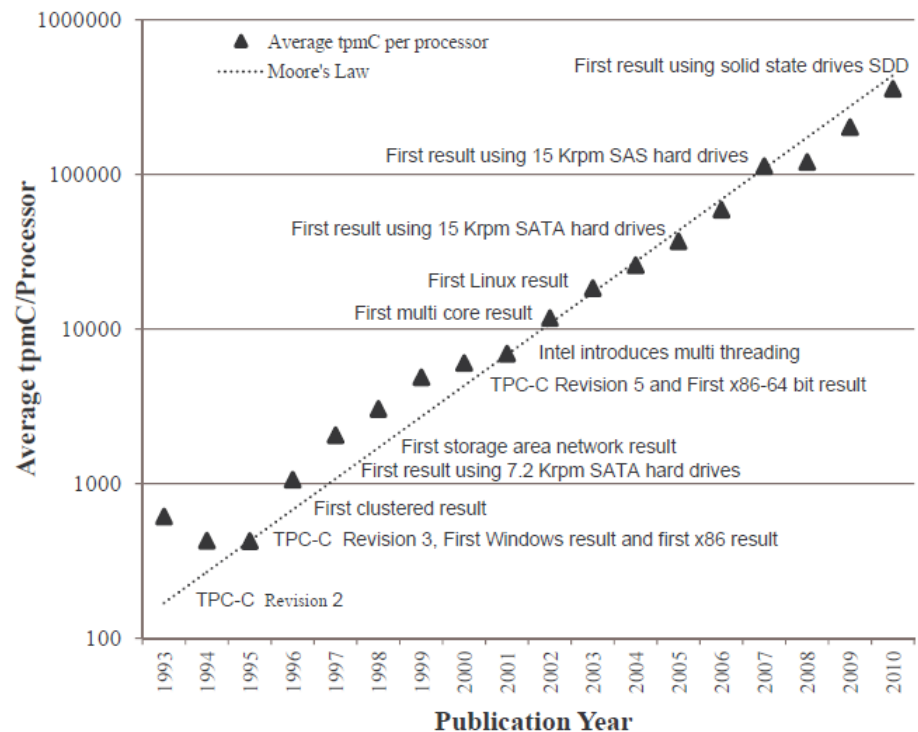
**Number of Benchmark Results**



# Aspects about TPC Benchmarks

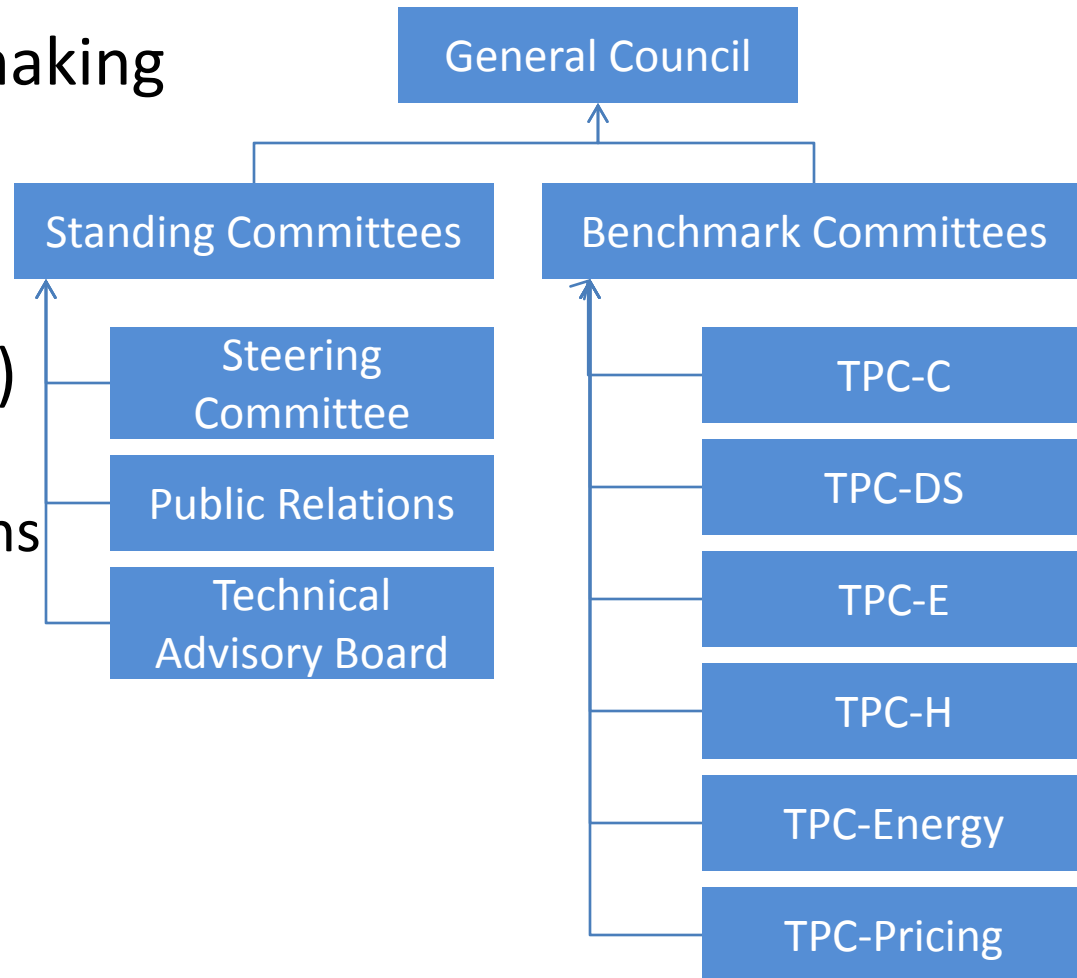
- Benchmarks
  - Are free to download
  - Have standardized metrics
    - Price performance metric
    - Energy metric (optional)
  - Test whole system performance in transactions/queries, A
  - Are Software and Hardware
  - Have a long shelf life
- Benchmark publications
  - Are subject to fee
  - Require full disclosure
  - Are Independently audited
  - Can be challenged by non-members available

## Transaction performance vs. Moore's Law, Milestones, 1993 to 2010



# TPC as an Organization

- Centralized decision making
- Full members (18)
  - One vote per company
- Associate members (5)
  - Universities
  - Research Organizations
- Affiliates
  - Consultants
  - Auditors



# Standards Not Just For Performance

- Step 0: Working group
  - Members meet informal to craft ideas for new benchmark
- Step 1: Draft Benchmark Submittal
  - Working group submits draft benchmark for consideration
- Step 2: Creation of a Subcommittee
- Step 3: Status and Direction
  - At every meeting the subcommittee reports to GC and gets direction
- Step 4: Authorizing Public Release of Draft Specification
- Step 5: Accepting a Standard for Review
  - Decision made in GC
- Step 6: Formal Review
  - 60-n days
- Step 7: Approval for Mail Ballot
- Step 8: Actual Mail Ballot
- Step 9: Creation of a Maintenance Subcommittee

# TPC Benchmarks

- They have a long shelf life, but they take a long time to develop
  - Centralized decision making
  - Realism/Complexity
  - Full system performance testing
  - Tie into TPC-Pricing and TPC-Energy
- They are realistic and holistic, but they take time to run
  - Engineering knowledge that is required
  - System complexity
- They are expensive
  - Audit
  - Engineering time

# TPC's Technology Conference on Performance Evaluation and Benchmarking

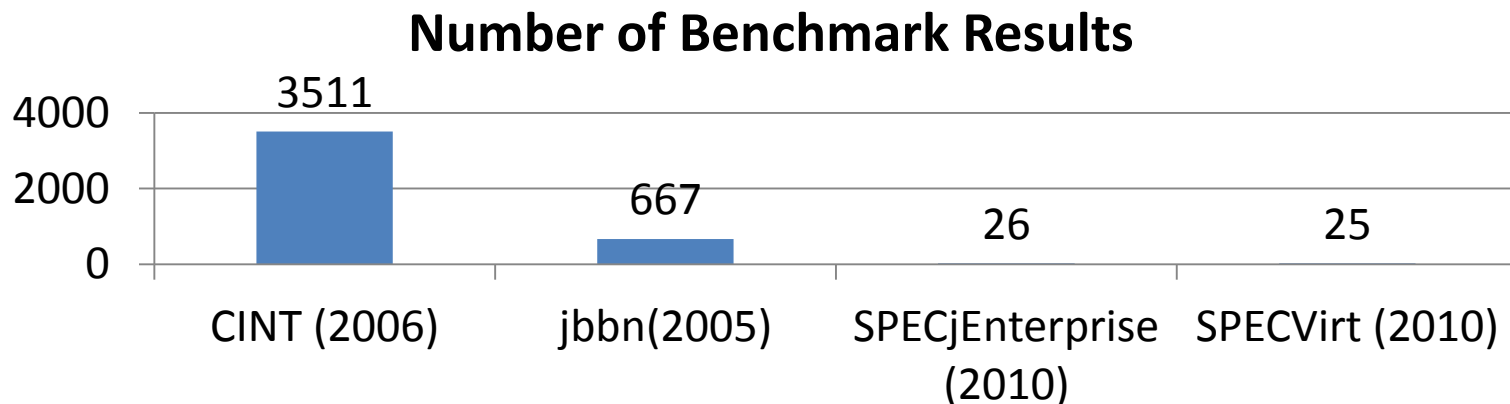
- International conference
- Started in 2009
- Co-located with International Conference on Very Large Data Bases (VLDB)
- Open to industry experts , researchers, vendors and customers
- Bridges the gap between industry and academia
- Is a forum do discuss novel ideas in benchmarking
- This years conference:
  - August 27, 2012 in Istanbul, Turkey
  - Keynote: Michael J. Carey, Professor of Computer and Information Sciences at the University of California, Irvine



# SPEC: Standard Performance Evaluation Corporation

## Existing Benchmark Suite

- CPU: new version every 3-5 years
- High Performance Computing: new version every 3-6 years
- Enterprise benchmark
- Power
- Virtualization
- Web



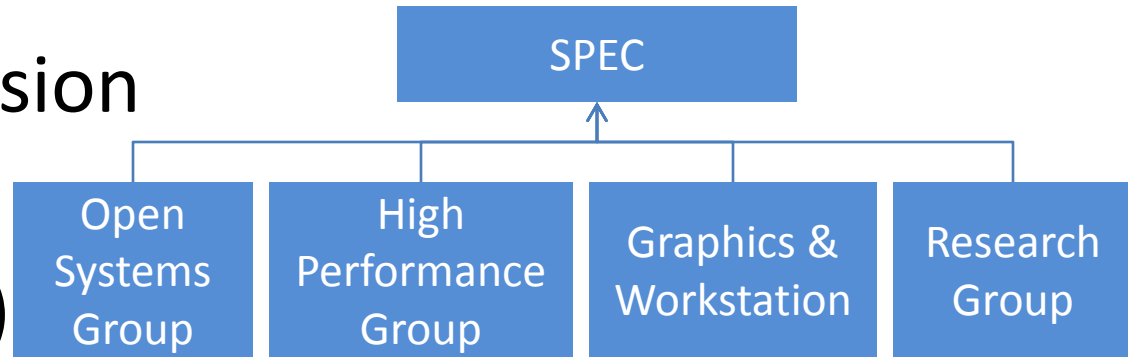
# Aspects about SPEC Benchmarks



- Benchmarks
  - Can be downloaded for a fee
  - Each defines its own metric
  - Test performance of small systems or components of systems
  - Server centric
  - Have a short shelf life
- Benchmark publications
  - Are free to members and subject to fee to non-members
  - Are peer reviewed
  - Require disclosure of summary

# SPEC as an Organization

- De-centralized decision making
- SPEC members (42)
- SPEC Associates (30)
  - Universities
  - Research Organizations
- SPEC Research Group (33)

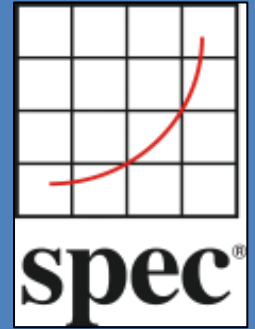


# SPEC Benchmarks

- Tend to be easy to run, but they do not measure the full system performance
- Tend to be quick to develop, but they cannot be used for long term capacity planning or performance projections

# SPEC Research Group (RG)

*The Research Group of the  
Standard Performance Evaluation Corporation*



## Mission Statement

- ▶ Provide a **platform for** collaborative research efforts in the areas of computer benchmarking and quantitative system analysis
- ▶ Provide metrics, tools and benchmarks for evaluating early prototypes and research results as well as full-blown implementations
- ▶ Foster interactions and collaborations btw. industry and academia





# Current Members (Mar 2012)



Find more information on: <http://research.spec.org>



# If you have an interest in new performance methods, you should join the SPEC RG

## ▶ Find a new venue to discuss your work

- ▶ Exchange with experts on how the performance of systems can be measured and engineered
- ▶ Find out about novel methods and current trends in performance engineering
- ▶ Get in contact with leading organizations in the field of performance evaluation

## ▶ Find a new group of potential employees

## ▶ Join a SPEC standardization process

## ▶ Performance in a broad sense:

- ▶ *Classical performance metrics:* Response time, throughput, scalability, resource/cost/energy, efficiency, elasticity
- ▶ *Plus dependability in general:* Availability, reliability, and security

# TPC vs. SPEC

| TPC Model                                   | SPEC Model   |
|---|--|
| Specification based                         | Kit based  |
| Performance, Price, Energy in one benchmark | Performance and Energy in separate benchmarks                                    |
| End-to-end                                  | Server-centric   |
| Multiple tests (ACID, Load, etc.)           | Single test  |
| Independent review                          | Peer review  |
| Full disclosure                             | Summary disclosure   |
| TPC Technology Conference                   | SPEC Research Group + ICPE (International Conference on Performance Engineering) |





# Standards not just for Benchmarks

