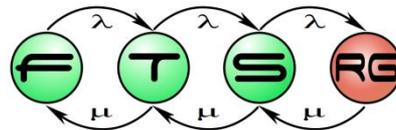


Benchmarking

Budapest University of Technology and Economics
Fault Tolerant Systems Research Group



KURZE WIEDERHOLUNG

aus aktuellem Anlass

Ariane 5 Trägerrakete

- Die leistungsfähigste europäische Trägerrakete



Ariane 5 Trägerrakete

- Am 4. Juni 1996 hat die Rakete sich 37 Sekunden nach dem Start zerstört
 - die gelieferten vier Cluster-Satelliten sind auch zerstört worden
 - \$370 Millionen Verlust



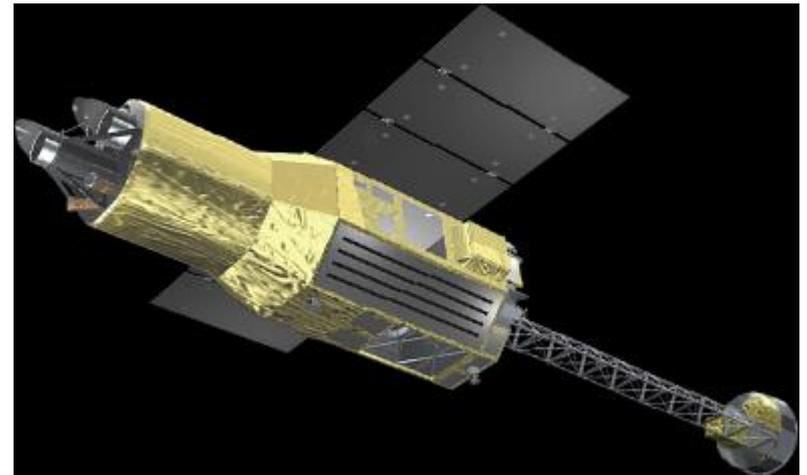
Ariane 5 Trägerrakete

- Am 4. Juni 1996 hat die Rakete sich 37 Sekunden nach dem Start zerstört
 - die gelieferten vier Cluster-Satelliten sind auch zerstört worden
 - \$370 Millionen Verlust
- Einer der teuersten Softwarefehler
 - primäre Ursache:
 - erfolglose Konversion einer Zahl
 - sekundäre Ursache:
 - Die Modulen wurden nie zusammen getestet**

„Die Menschen vergessen nicht. Aber sie lernen auch nicht.“

Hitomi Weltraumteleskop

- Auch als ASTRO-H Röntgenteleskop bekannt
 - Kosten: min. \$360 Millionen
 - 10 Jahren, 160 Wissenschaftler von 70 Firmen/Unis
- Gestartet am 17. Februar 2016 von JAXA
- Verbindung verloren am 26. März 2016



Hitomi Weltraumteleskop

- Auch als ASTRO-H Röntgenteleskop bekannt
 - Kosten: min. \$360 Millionen
10 Jahren, 160 Wissenschaftler von 70 Firmen/Unis
- Gestartet am 17. Februar 2016 von JAXA
- Verbindung verloren am 26. März 2016
- Primäre Ursache:
 - Falsche Detektierung der Rotation des Teleskopes
 - Richtige Gegensteuerung dieser „Rotation“
 - Notbremse: Raketenmanöver (in der falschen Richtung ☹)
 - Solarmodule und Hauptträger verloren/abgebrochen

Hitomi Weltraumteleskop

- Auch als ASTRO-H Röntgenteleskop bekannt
 - Kosten: min. \$360 Millionen
 - 10 Jahren, 160 Wissenschaftler von 70 Firmen/Unis
 - Gestartet am 17. Februar 2016 von JAXA
 - Verbindung verloren am 26. März 2016
 - Sekundäre Ursache:
 - Parameterdaten des Raketenbremsmanöver aktualisiert am 28. Februar 2016
 - (wegen „last minute“ Änderungen an der Struktur des Teleskopes)
- Wie richtig waren diese Parameterdaten getestet?

WARUM BENCHMARKING?

Warum benchmark?



Benchmarking

- Ziele: Vergleich der Leistung von Software- oder Hardwarekomponenten
 - Entscheidungsunterstützung
 - Welche Komponente soll gekauft/installiert werden?
 - Was kann das existierende System leisten? (Schwächen/Stärken)
 - Was können die Konkurrenten?
 - Leistungstesten
 - Soll die Leistung erhöht werden? Wo? (bei Entwicklung)
 - Ist eine gegebene Einstellung optimal?
 - Hat eine Einstellung tatsächliche Auswirkung auf die Gesamtleistung?

Erwartungen

- Wiederholbarkeit
 - Repeatability
- Reproduzierbarkeit
 - Reproducibility
- Relevanz
- Konformität mit Normen/Vereinbarungen
- Verallgemeinerter Anwendungsfall
 - Das Ergebnis sei für den Durchschnittsbenutzer interpretierbar

Benchmark-Belastungsmodelle

- **Wissenschaftliche/technische Systeme**
 - Verarbeitung von großen Datenmengen (number crunching)
 - Parallele Methoden
- **Transaktionsmanagement (OLTP)**
 - Klient-Server Umgebung
 - Viele schnelle, parallele Transaktionen
- **Batch-Datenverarbeitung**
 - Reporterstellung aus großen Datenmengen
- **Entscheidungsunterstützung**
 - Wenige, komplexe Abfragen
 - Ad-hoc Operationen
 - Viele Daten (z.B. OLAP)
- **Virtualisation**

Zu messende Parameter (Metriken)

- Laufzeit
 - Anfang, Ende?
 - Verteilung
 - CPU, I/O, Netzwerk,...
- Transaktionsgeschwindigkeit
 - Antwortzeit des Systems
 - Sogar verschachtelte Transaktionen
- (Grenz)Durchsatz
 - (*Maximale*) **Verarbeitete** Datenmenge / Laufzeit
 - nach Belastung

Zu messende Parameter(2)

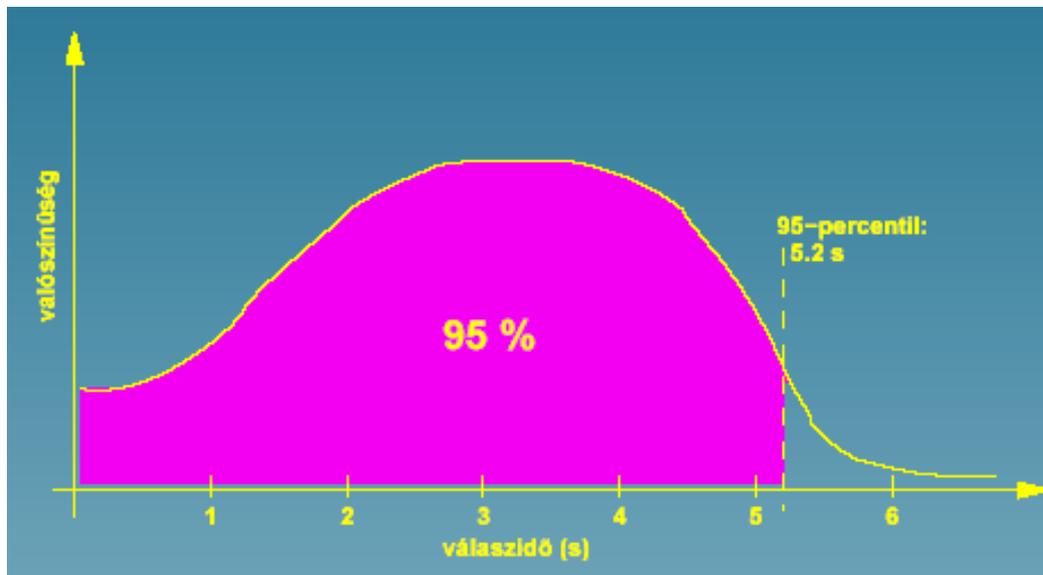
■ Antwortzeit

○ nach Belastung

- Anzahl der Benutzer
- Anzahl der Transaktionen, usw.

■ X-Perzentil

- x Prozent einer gegebenen Menge ist stärker als dieser Wert



Ausführung des Benchmark

- **Versicherung der Relevanz**
 - Wird die richtige Anwendung gemessen?
 - Art der Belastungsgenerierung soll die wirkliche Belastung annähern
 - Minimalisierung der störenden Faktoren

Benchmarking - Definition

■ Wikipedia

*„In **computing**, a benchmark is the **act of running** a computer program, a set of programs, or other operations, in order to **assess the relative performance** of an object, normally by running a number of **standard tests** and trials against it.“*

Benchmarking ist

- das **Ausführen** eines **Programmes** (von mehreren Programmen oder von anderen Operationen)
- **mit standardisierten Tests** oder Eingaben,
- **um die relative Performanz** eines Objektes zu **bestimmen**.

STANDARD BENCHMARKS

SPEC, TPC-C, ...

SPEC Benchmarks

- <http://www.spec.org/benchmarks.html>
 - Standard Performance Evaluation Corp.
- Ressource- und Anwendungsbenchmark
 - CPU
 - Anwendungen
 - Mailserver
 - Webserver usw.
- Benchmark: bestellbare Dienstleistung

SPEC CPU2006

- CPU-intensiv
- CINT2006
 - Rechenintensiv, mit Ganzzahlen
- CFP2006
 - Mit Gleitkommazahlen
- Ergebnisse: <http://spec.org/cpu2006/results/>
 - Test Sponsor (vendor), System Name (product)
 - Processor: enabled cores, enabled chips, cores/chip, threads/core
 - Results: base, peak

CINT2006 und CFP2006 Leistungsgeneratoren

■ CINT2006 :

400.perlbench	C	Programming Language
401.bzip2	C	Compression
403.gcc	C	C Compiler
429.mcf	C	Combinatorial Optimization
445.gobmk	C	Artificial Intelligence
456.hmmer	C	Search Gene Sequence
458.sjeng	C	Artificial Intelligence
462.libquantum	C	Physics / Quantum Computing
464.h264ref	C	Video Compression
471.omnetpp	C++	Discrete Event Simulation
473.astar	C++	Path-finding Algorithms
483.xalanbmk	C++	XML Processing

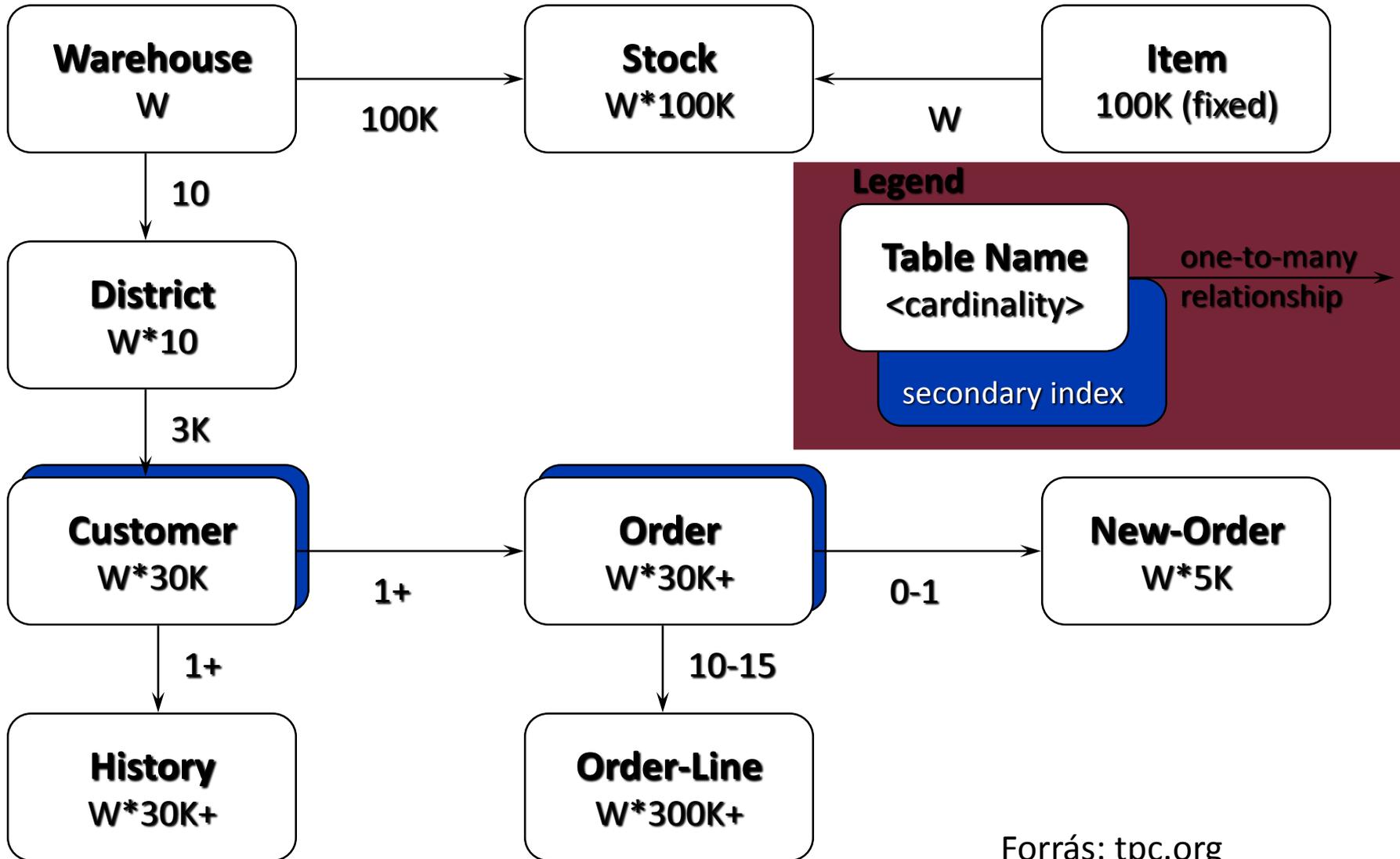
■ CFP2006:

410.bwaves	Fortran	Fluid Dynamics
416.gamess	Fortran	Quantum Chemistry
433.milc	C	Quantum Chromodynamics
434.zeusmp	Fortran	Fluid Dynamics
435.gromacs	C, Fortran	Molecular Dynamics
436.cactusADM	C, Fortran	General Relativity
437.leslie3d	Fortran	Fluid Dynamics
444.namd	C++	Molecular Dynamics
447.deall	C++	Finite Element Anal.
450.soplex	C++	Linear Programming
453.povray	C++	Image Ray-tracing
454.calculix	C, Fortran	Structural Mechanics
459.GemsFDTD	Fortran	Electromagnetics
465.tonto	Fortran	Quantum Chemistry
470.lbm	C	Fluid Dynamics
481.wrf	C, Fortran	Weather
482.sphinx3	C	Speech Recognition

Das TPC-Benchmark

- Beobachtung von Datenbanksystemen
 - RDBMS+OS+HW
- Messungsumgebung
 - Musterdatenbank: Kunden und Bestellungen
 - 5 Arten von Transakt. (Abfrage/Änderung) gemischt
 - Obergrenze für die Laufzeit
 - Realitätsnahe Bedingungen: ACID Transaktionen, Bedenkzeit der Benutzer
(Atomarität, Konsistenz, Isolation und Dauerhaftigkeit)
- Gemessene Daten
 - Durchsatz (tpmC) *(transaction per minute)*
 - „Effizienz“ (\$/tpmC)

TPC-C Übersicht



Forrás: tpc.org

Vor der Analyse: Reinigung der Daten

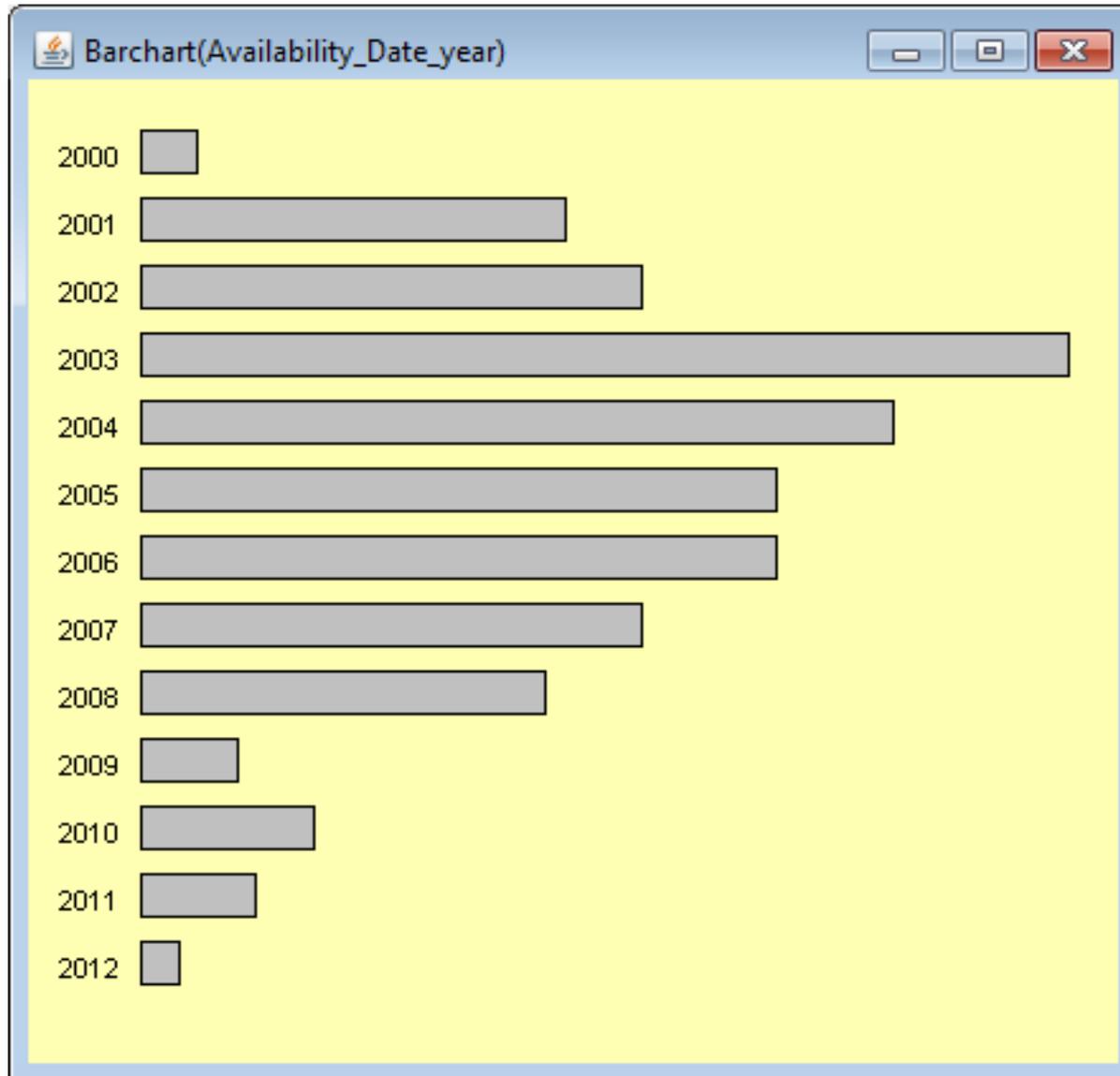
○ Ausgangsdatensatz:

	A	B	C	D	E	F	G	H	I	J	K
1	TPC-C BENCHMARK RESULTS										
2	These results are valid as of date 6/12/2012 10:04:24 PM										
3											
4	TPC-C Results - Revision 5.X										
5											
6	<u>Company</u>	<u>System</u>	<u>Spec. Revision</u>	<u>tpmC</u>	<u>Price/Perf</u>	<u>Total Sys. Cost</u>	<u>Currency</u>	<u>Database Software</u>	<u>Operating System</u>	<u>TP Monitor</u>	<u>Server CPU Type</u>
7	Acer	▶Altos R710	5.5	66543	12.42	826507.55	AUD	Microsoft SQL Server	▶Microsoft Windows Serv	▶Microsoft CO	▶Intel Xeon - 3.6 GHz
8	Bull	▶Bull Escal	5.9	6085166	2.81	17127928	USD	IBM DB2 9.5	▶IBM AIX 5L V5.3	▶Microsoft CO	▶IBM POWER6 - 5.0
9	Bull	▶Bull Escal	5.9	629159	2.49	1566664	USD	IBM DB2 9.5 Enterpri	▶IBM AIX 5L V5.3	▶Microsoft CO	▶IBM POWER6 - 4.2
10	Bull	▶Bull Escal	5.8	1616162	3.54	5716286	USD	IBM DB2 9.1	▶IBM AIX 5L V5.3	▶Microsoft CO	▶IBM POWER6 - 4.7
11	Bull	▶Bull Escal	5.8	404462	3.51	1417121	USD	Oracle Database 10g	▶IBM AIX 5L V5.3	▶Microsoft CO	▶IBM POWER6 - 4.7

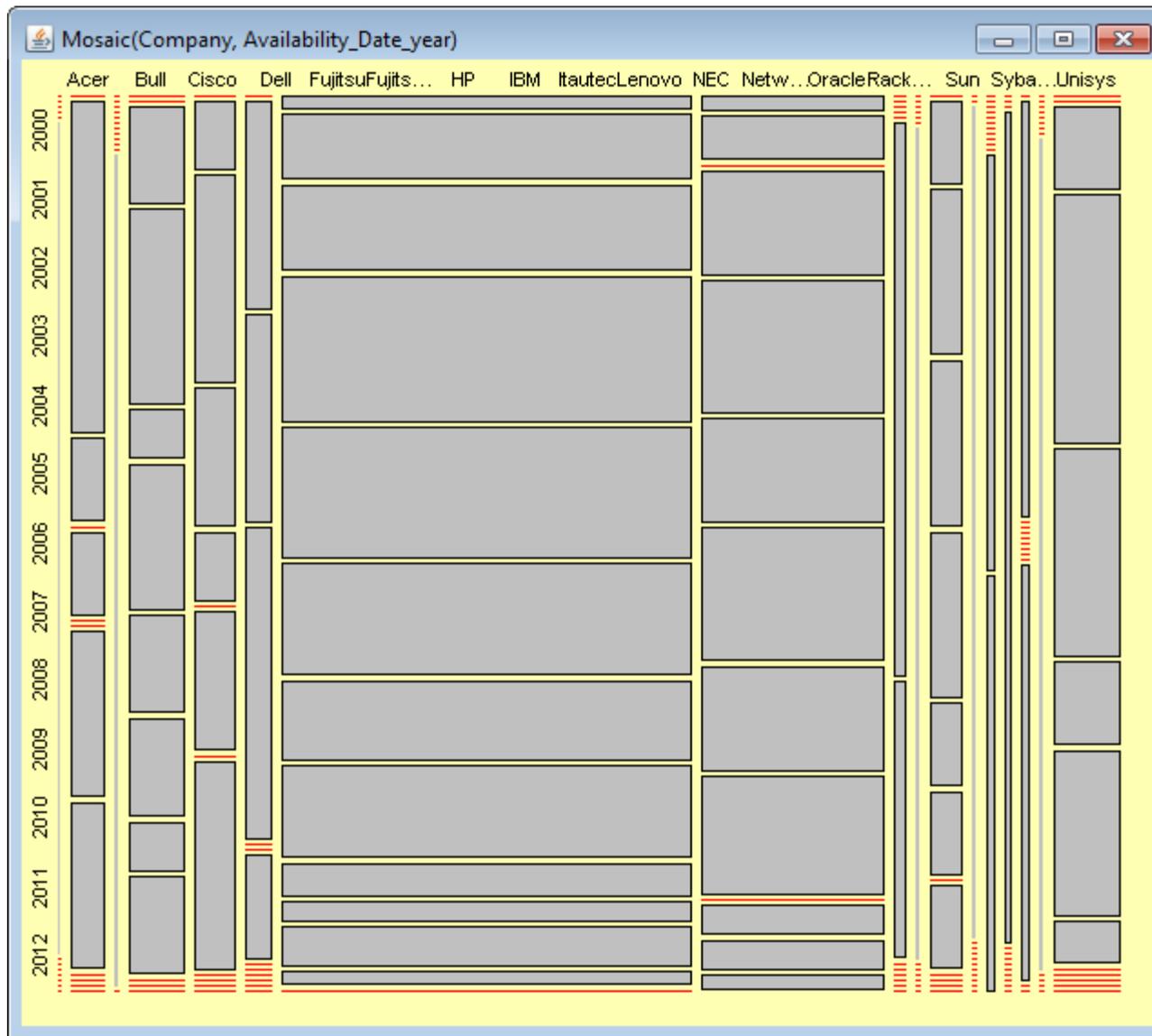
○ Überflüssige Daten:

- Zeilen (z.B. die ersten und letzten Zeilen, die nicht zum Ergebnis gehören)
- Spalten (z.B. „Server CPU Type“ interessant?)
- z.B. Kosten in verschiedenen Währungen
- Dezimalkomma vs. Dezimalpunkt
- *Fujitsu vs. Fujitsu-Siemens* (zu vereinigen?)

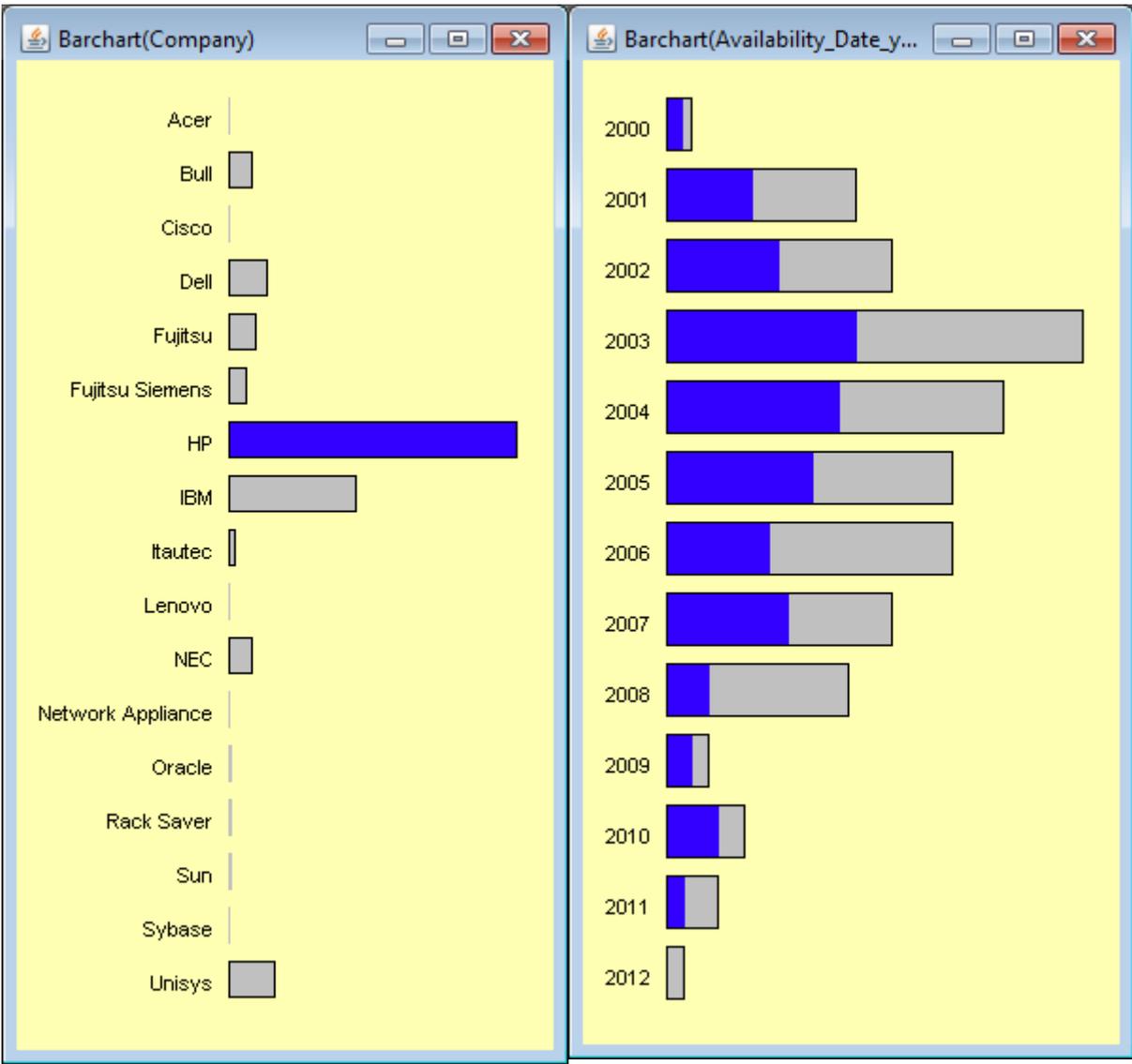
Welche Jahren werden im Benchmark betrachtet?



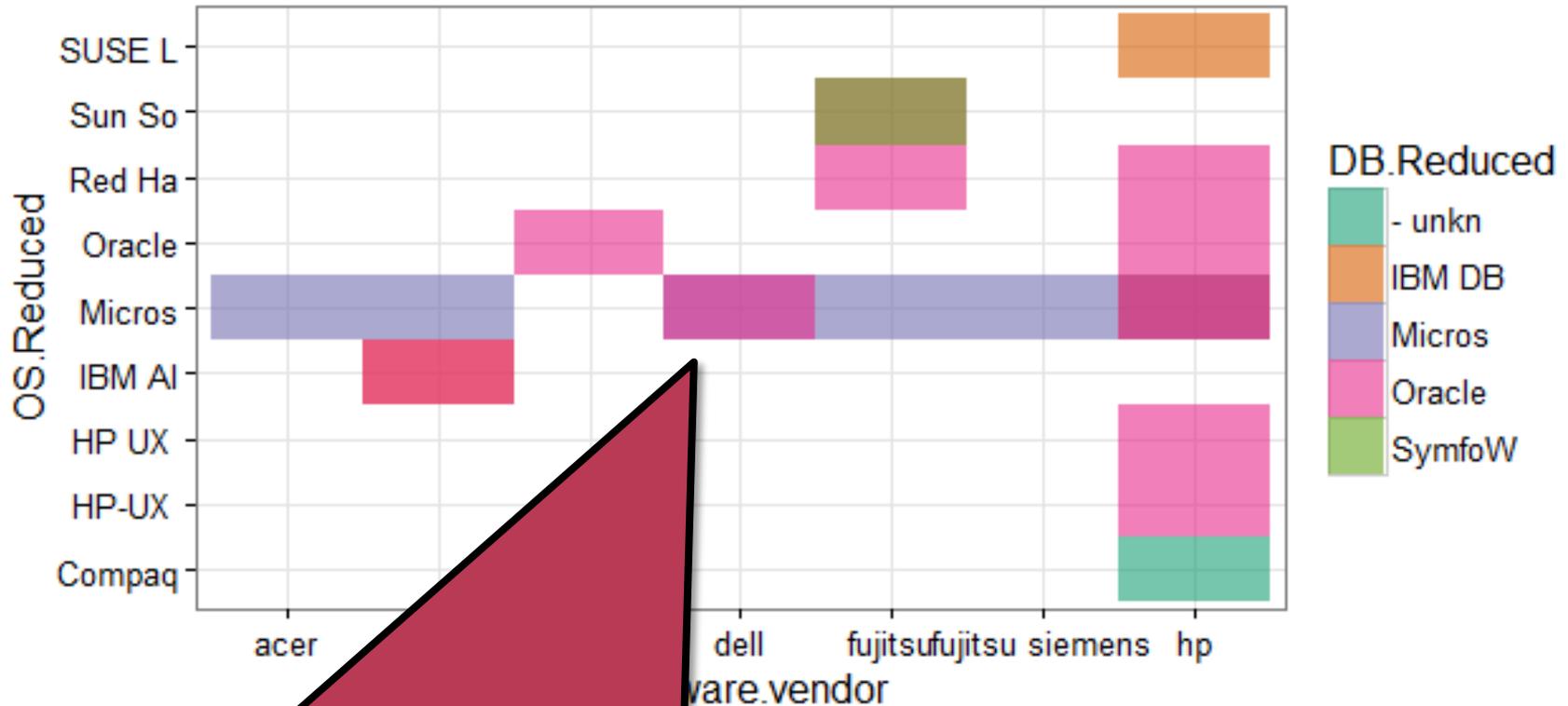
In welchen Jahren waren die Lieferanten aktiv?



In welchen Jahren waren die Lieferanten aktiv?

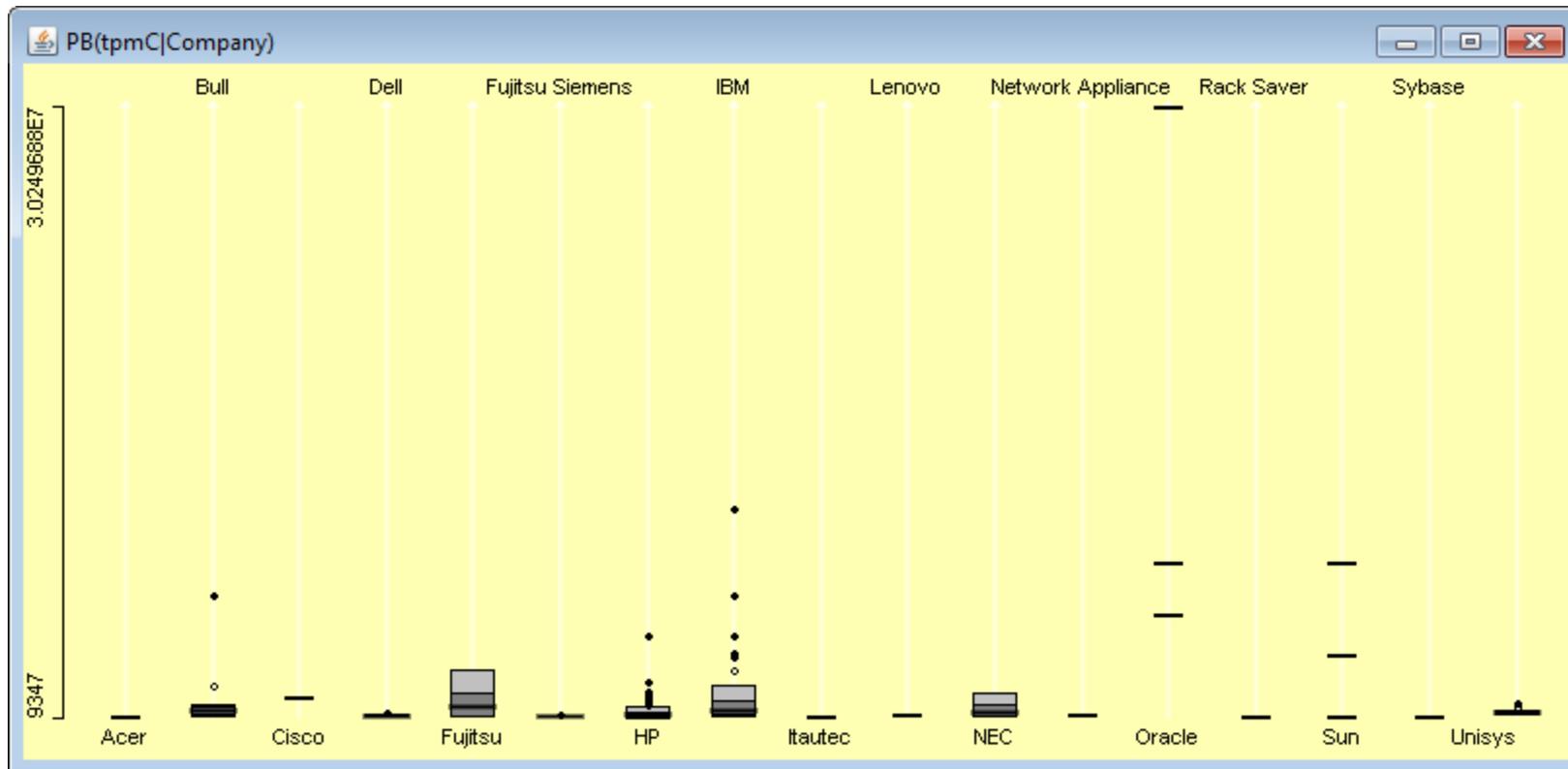


Gemessene Konfigurationsvariationen

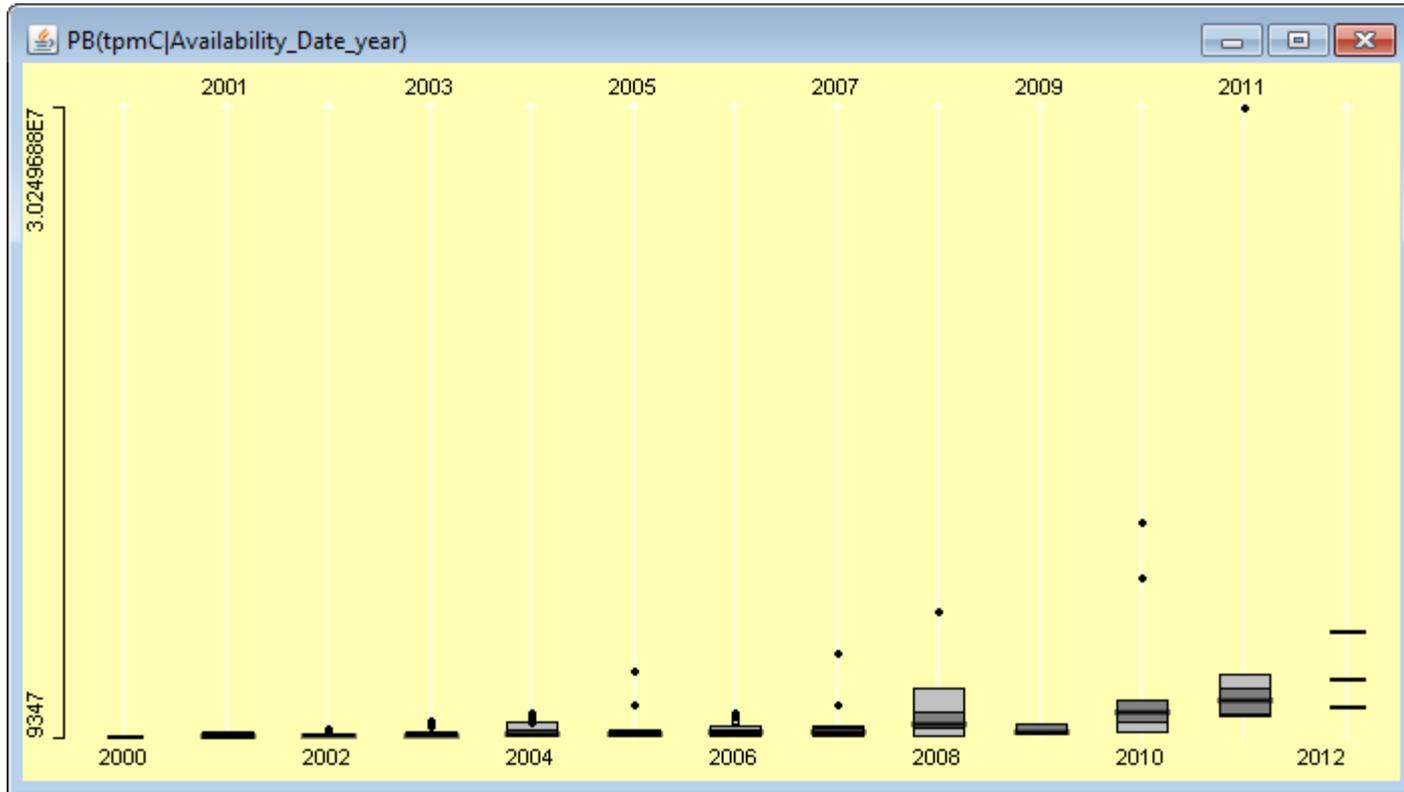


Datenbanksysteme können an mehreren Betriebssystemen laufen

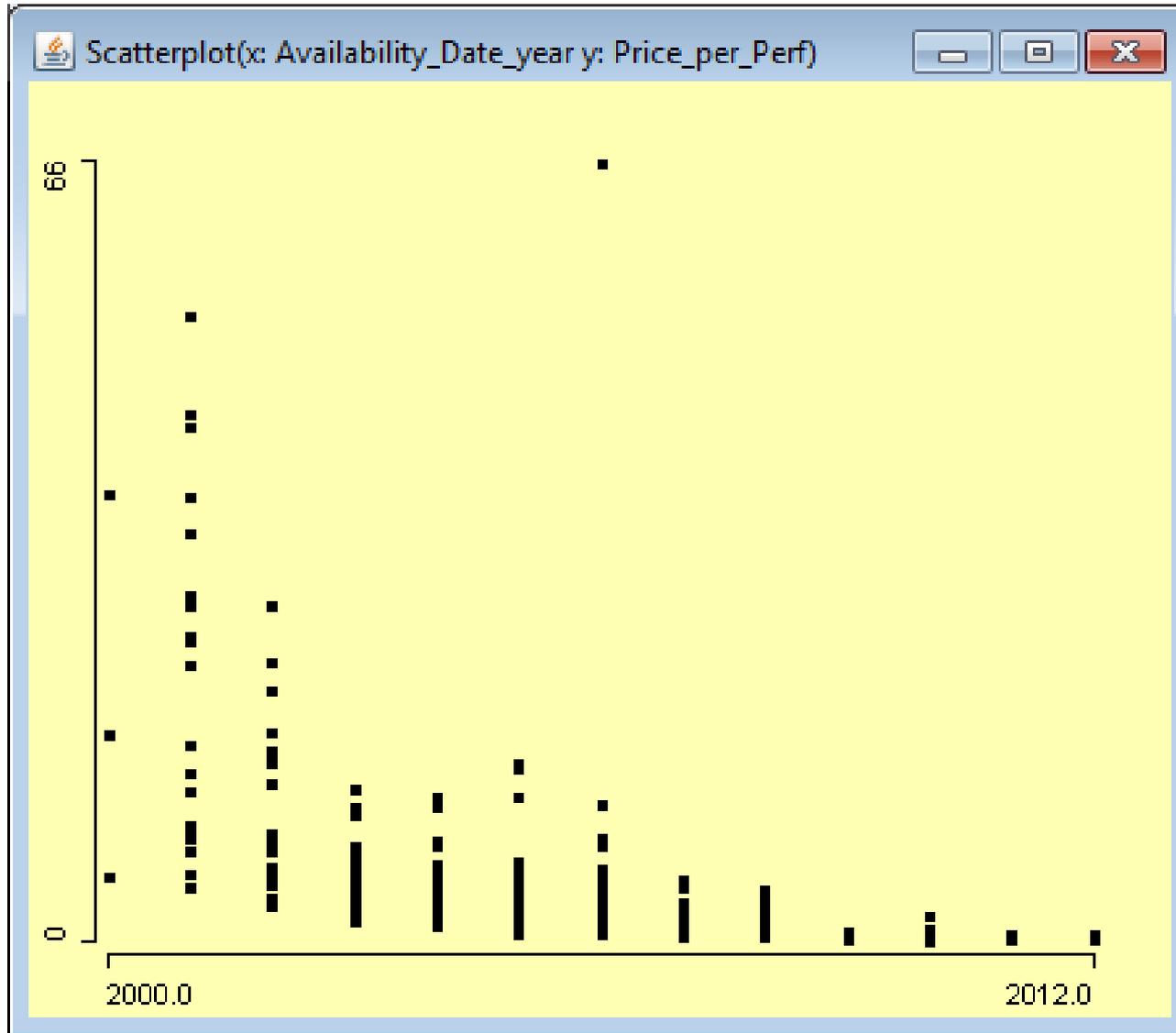
Gemessene Werte der Leistungsmetriken



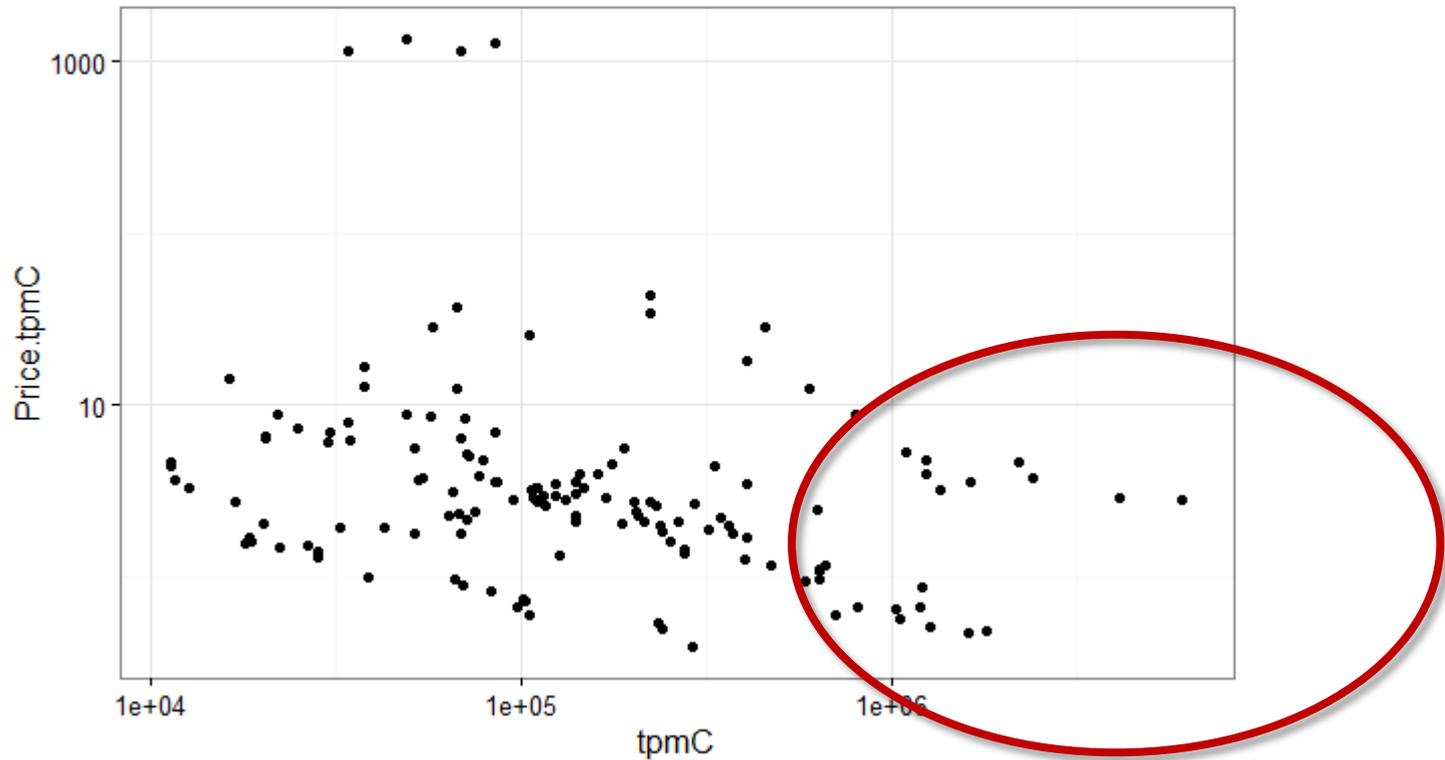
Wie hat sich Leistung mit der Zeit verändert?



Wie haben sich die Preise verändert?

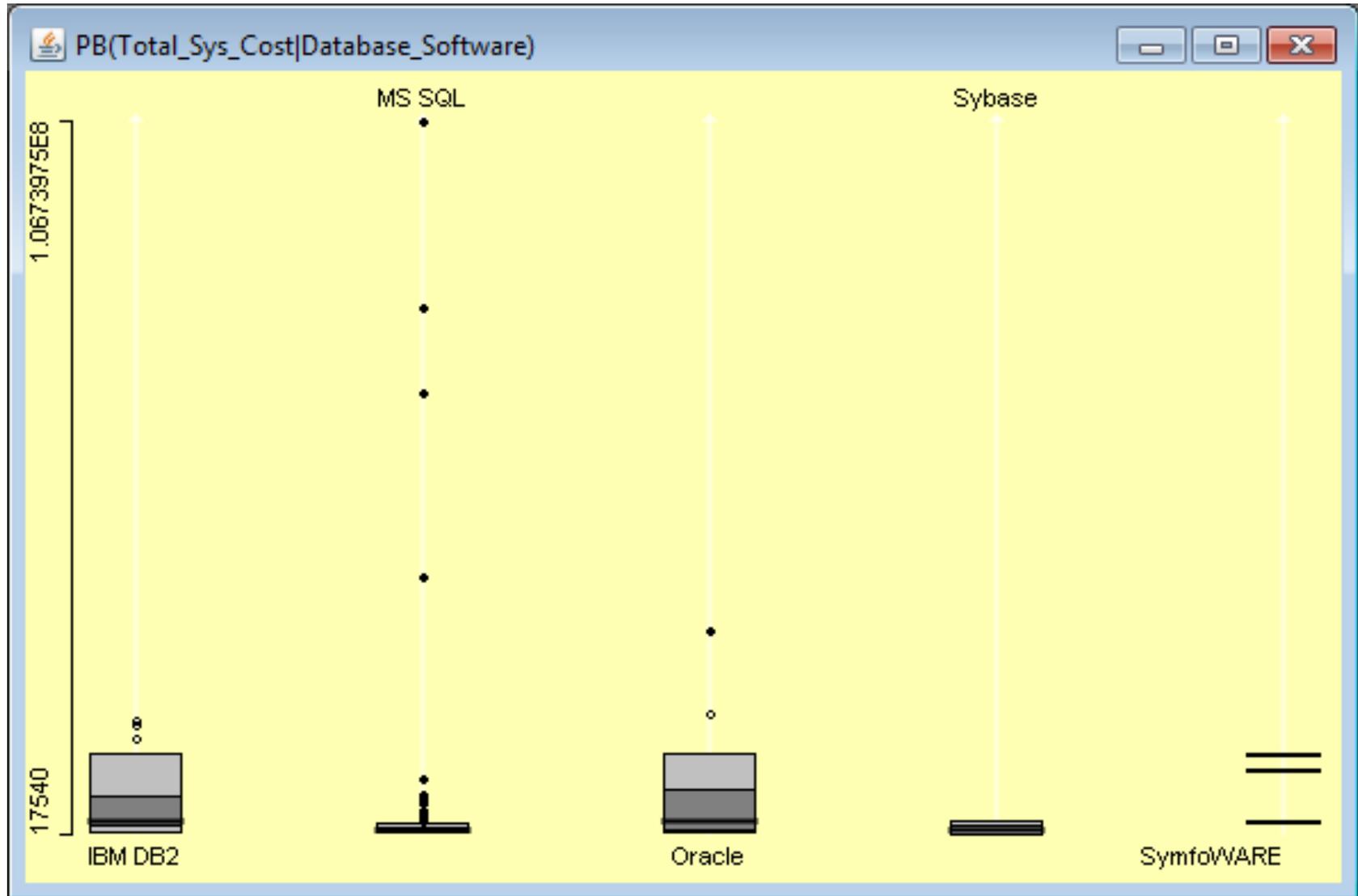


Benchmark Ergebnisse

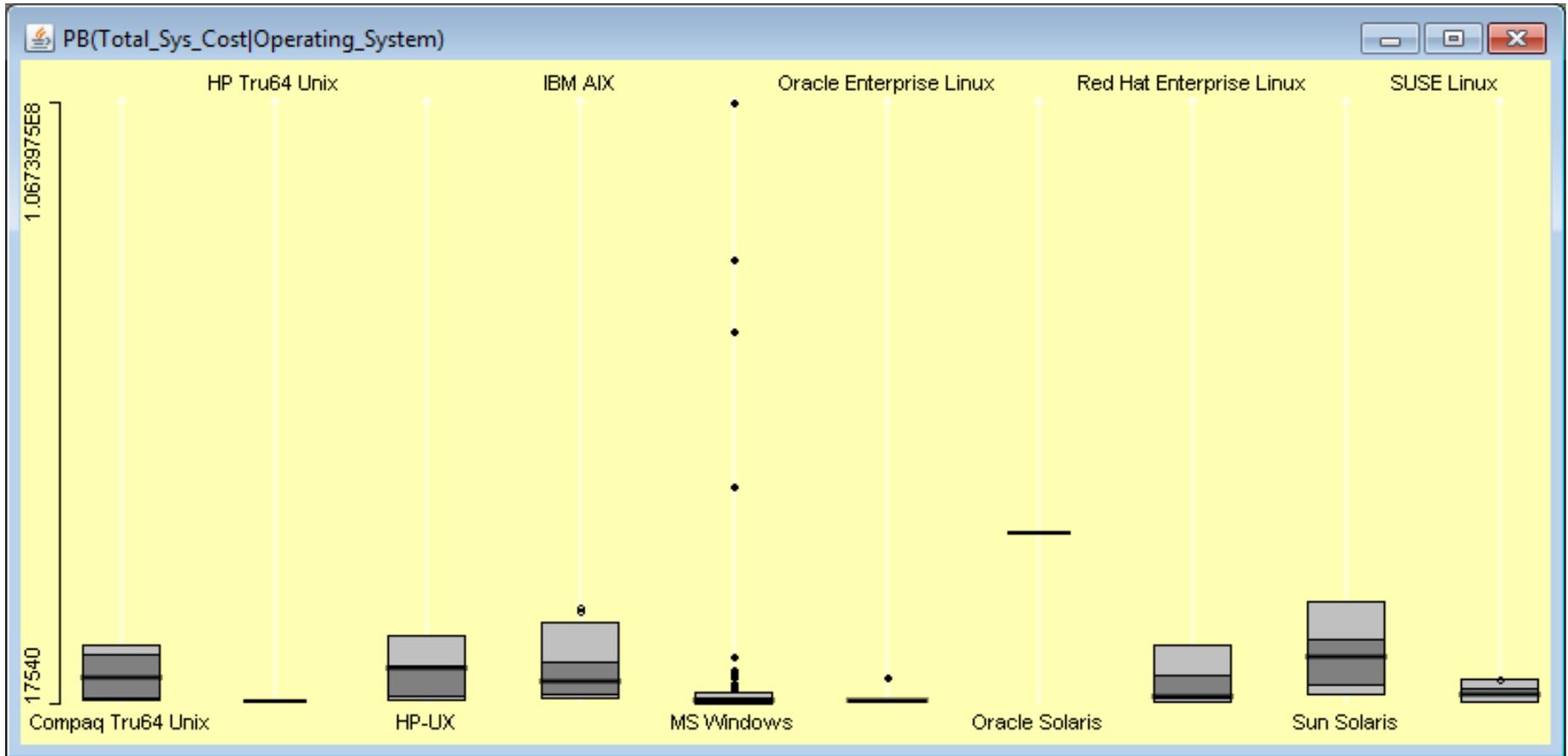


Logarithmische Skalierung?

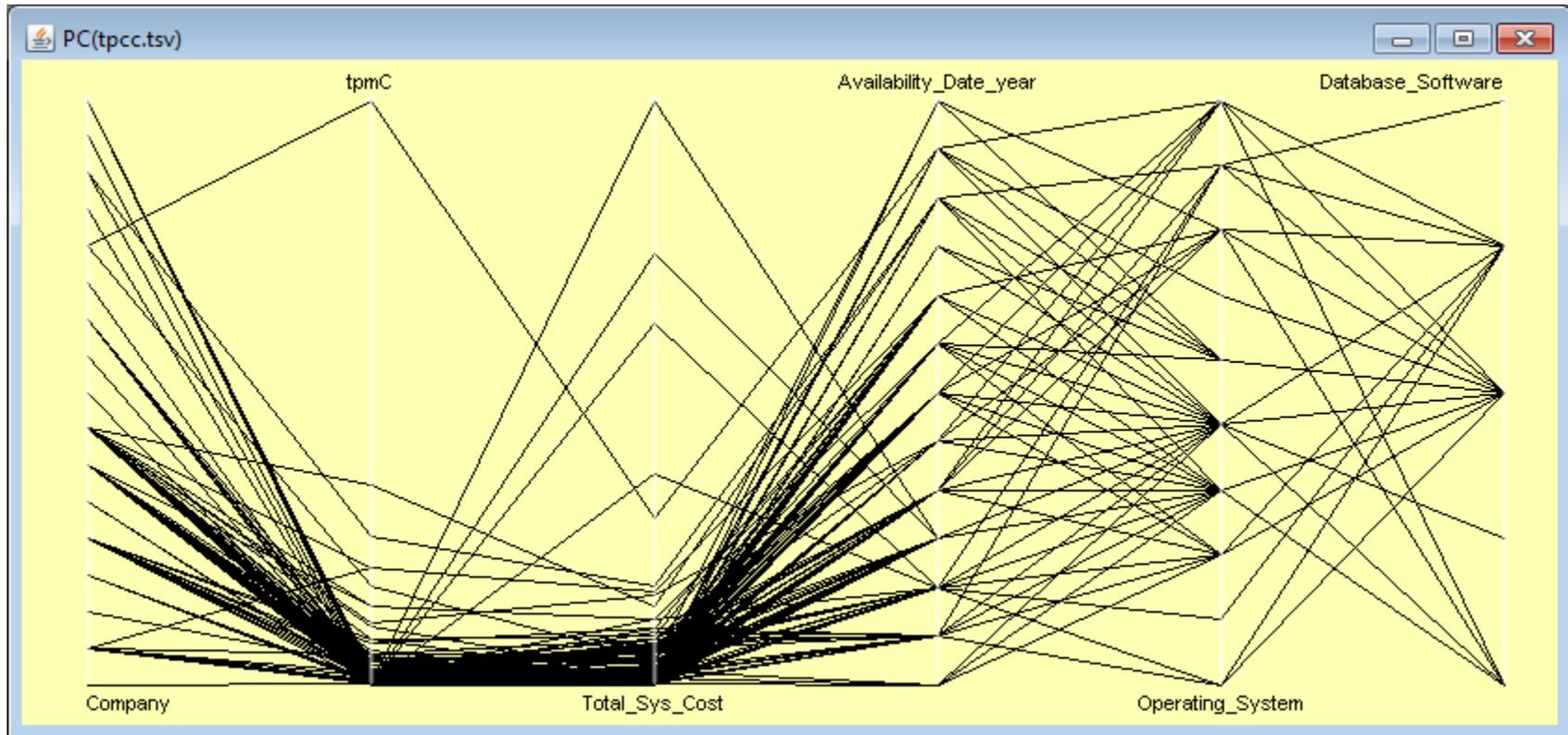
DB-Management SW wählen



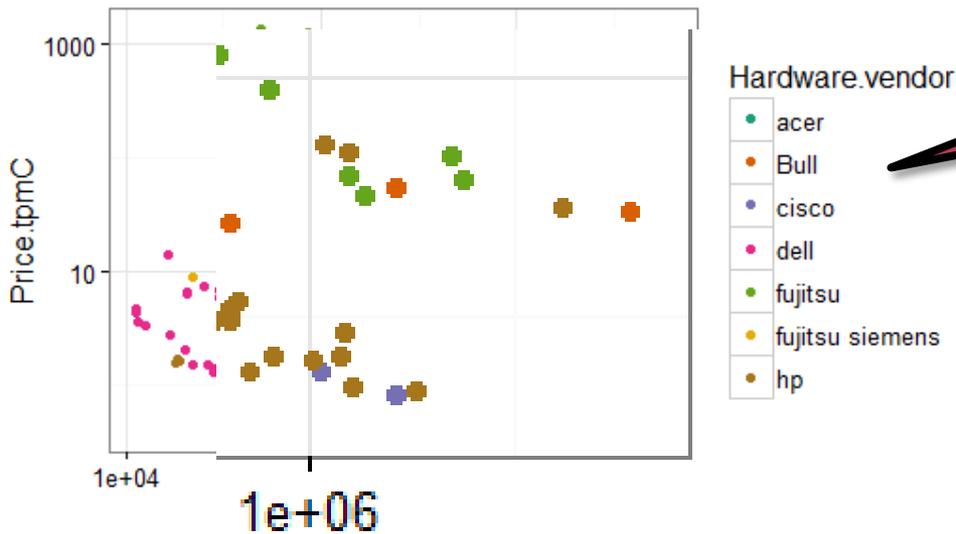
OS wählen



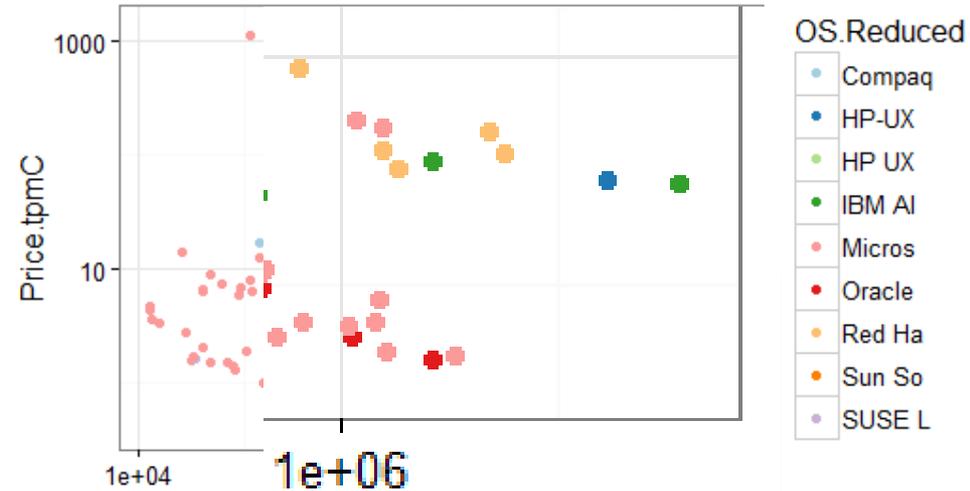
Das „big picture“



Benchmark Ergebnisse



Das Spitzelfeld ist eher vielfältig



Es gibt weder ein bestes OS, noch eine beste DB-Konfiguration

