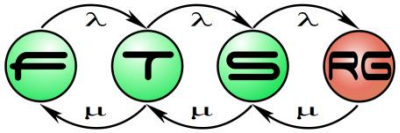
Benchmarking



Autóvásárlás

Technische Daten	Porsche Carrera GT	Porsche 911 GT2 RS
Motor	V10	B6, Biturbo
Einbaulage	Mitte längs	hinten längs
Hubraum	5733 cm3	3600 cm3
kW (PS) bei 1/min	450 (612)/8000	456 (620)/6500
Literleistung	107 PS/l	172 PS/l
Nm bei 1/min	590/5750	700/2250-5500
Antriebsart	Hinterrad	Hinterrad
Getriebe	6-Gang manuell	6-Gang manuell
Leistungsgewicht	2,3 kg/PS	2,2 kg/PS
Tankvolumen	92 l	90 l
Beschleunigung		
0-100 km/h	3,9 s	3,5 s
0-200 km/h	9,9 s	9,8 s
Testverbrauch		
Ø auf 100 km	17,9 l Super Plus	11,9 l Super Plus
Leergewicht	1380/255 kg	1370/310 kg
Höchstgeschwindigkeit	330 km/h	330 km/h
Preis Serienwagen	452.690 Euro (2006)	237.578 Euro

Benchmarking

- Célok: szoftver/hardver eszközök teljesítményének összehasonlítása
  - Döntéstámogatás
    - melyiket előnyösebb megvenni/telepíteni
    - mekkora terhelésre elég a meglévő rendszer
  - Teljesítménytesztelés
    - kell-e még teljesítményen javítani és hol (fejlesztésnél)
    - optimális-e egy konkrét beállítás
    - van-e egy beállításnak teljesítményre gyakorolt hatása

Bechmark környezet

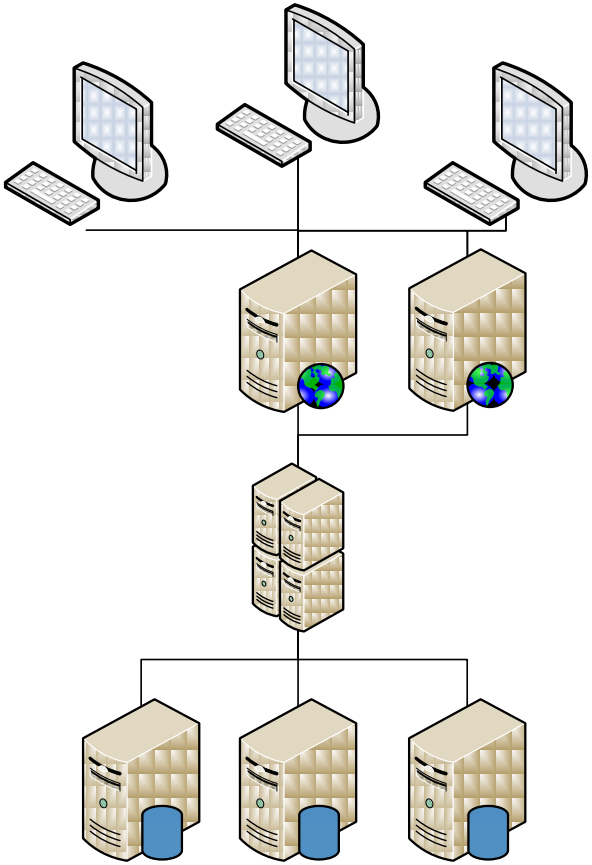
- Specifikáció
  - hardver
  - szoftver
  - üzemviszonyok
  - ütemezés
  - dokumentáció
- Alapelvek
  - kölcsönös zavarás minimalizálása (üzem közben)
  - Pareto elv (80/20) kihasználása
  - prekoncepció (mit mérünk)
  - terhelés közelítse a valós mintát (profilok)

Benchmark terhelési modellek

- Tudományos/műszaki rendszerek
  - nagy mennyiségű adat feldolgozása (number crunching)
  - párhuzamos módszerek
- Tranzakciókezelés (OLTP)
  - kliens-szerver környezet
  - sok gyors, párhuzamos tranzakció
- Batch jellegű adatfeldolgozás
  - riport készítés nagy mennyiségű adatból
- Döntéstámogatás
  - kevés, bonyolult lekérdezés
  - ad hoc műveletek
  - sok adat (pl. OLAP)
- Virtualizáció

Benchmark terhelési modellek példa

- OLTP architektúra példa (táblánál)

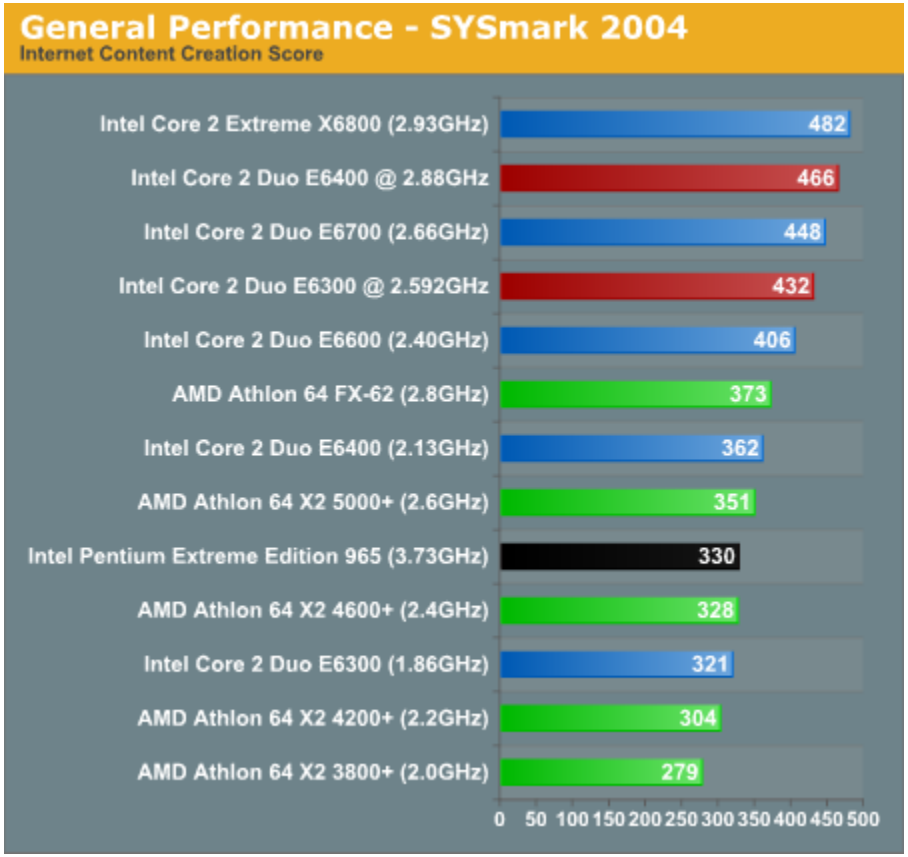


Kliens oldal - PC szintű benchmarkok

- BAPCo
- AMD, Intel, NVidia, MS, Dell, HP, Lenovo…
- SysMark2007
  - „E-Learning”, Video, Office, 3D…
- MobileMark2007
  - Notebook akku élettartam
  - DVD írás, dokumentum olvasás, …
- WebMark2004
  - Böngészés/tranzakciók

Példa

- SYSmark2004, tartalom létrehozása



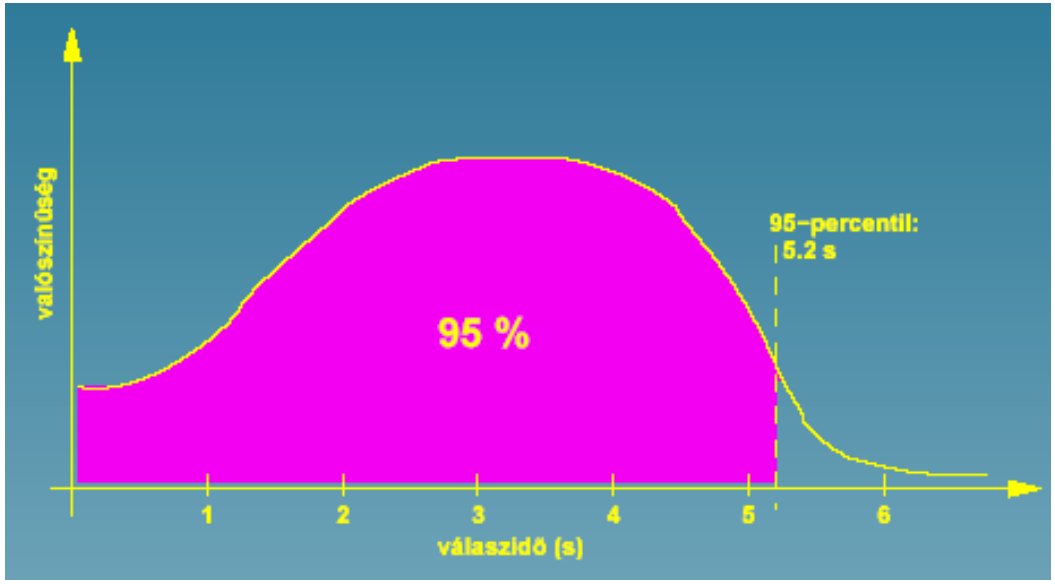
http://www.anandtech.com/cpuchipsets/showdoc.aspx?i=2802&p=4
Hasonló példa: http://www.tomshardware.com/reviews/wolfdale-steroids,1777-14.html

Mérendő paraméterek (Metrikák)

- Futási idő
  - kezdet, vég?
  - eloszlás
  - CPU, I/O, hálózat,…
- Tranzakiósebesség
  - rendszer reakcióideje
  - akár egymásba ágyazott tranzakciók
- Áteresztőképesség
  - ***feldolgozott*** adatmennyiség / futási idő
  - terhelés függvényében

Mérendő paraméterek (2)

- Válaszidő
  - terhelés függvényében
    - felhasználók
    - tranzakciók száma, stb.
- X-Percentil
  - Egy adott halmaz X százaléka ez alatt az érték alatt van



Eredmények összehasonlítása

- Pl. referenciarendszer alapján
- Több paraméter alapján ellentmondást kaphatunk
- Absztrakt referencia („standard benchmark”)
- Adatbányászati módszerek

- Nincs egy univerzális teljesítménymérő metrika még egy-egy szűkebb területen belül sem!

Tipikus problémák

- Túl kicsi problémaméret
- A felhasználási terület számára releváns eredmény?
- Elavult referenciák
- „Rejtett” paraméterek
  - konfigurációs beállítások
  - adott környezet specifikus tulajdonságai
- Elfogultság
- Hiányos specifikáció

Benchmark típusok

- Tudományos terminológia
  - Macrobenchmark – nagy összetett alkalmazás mérése
    - relevancia biztosításával közvetlenül használható eredményeket ad
  - Microbenchmark – alkalmazás kis részének kiemelt, analitikus mérése
    - analitikus felhasználás, profiling, teljesítmény előrejelzés
  - Nanobenchmark – egy-egy atomi művelet teljesítményének elkülönített mérése
    - főleg hibakeresés, profiling célra hasznos

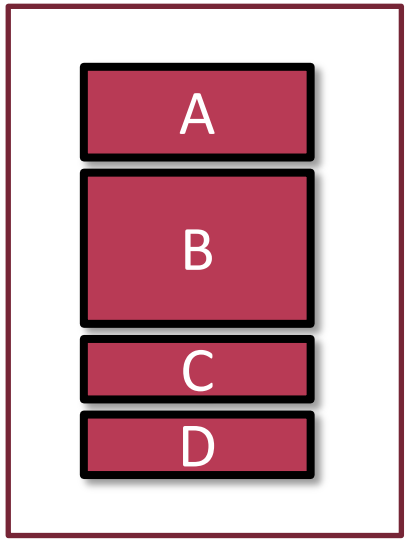
Benchmark elvégzése

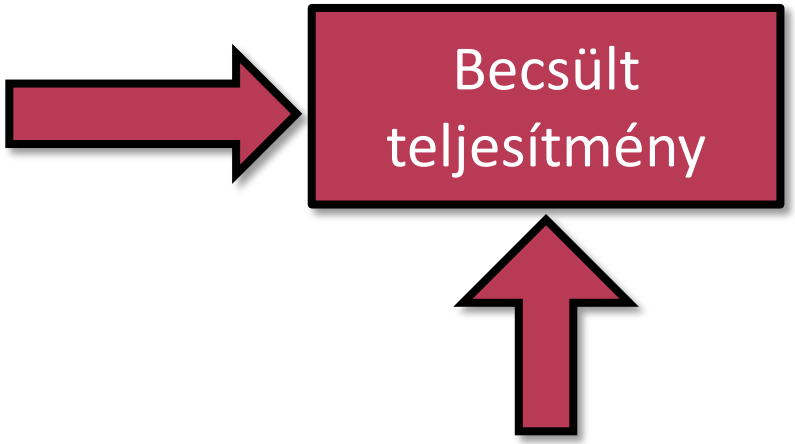
- Relevancia biztosítása
  - Tényleg azt az alkalmazást mérjük, amit kell
    - Különböző macrobenchmarkok eredményei többnyire nem vihetőek át egymás között.
  - Terhelésgenerálás jellege közelítse a valódi terhelést
    - Főleg macrobenchmarkoknál fontos, időben szétosztott valódi terhelést félrevezető lehet egy összefüggő batch terheléssel helyettesíteni.
    - Ügyeljünk a terhelésgenerátorra ható visszacsatolásokra
  - Minimalizáljuk a zavaró tényezőket
    - Főleg microbenchmarknál fontos, Pl.: ne mérjünk bele véletlenül diszk I/O-t a memória áteresztőképességébe, ne fusson más alkalmazás közben
    - Futási eredmények szórását kezelni kell

Eredmény analitikus felhasználása

- Cél: következtetünk jövőbeli (tehát nem mérhető) terhelésnél kapott eredményre

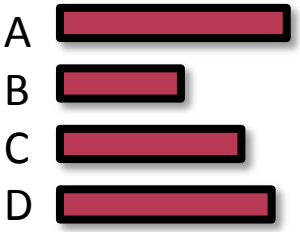
Összetett alkalmazás





Egyes részfeladatok aránya ismert (vagy legalább jól becsülhető) a terhelés során

Eredmények részfeladatonkénti microbenchmarkra:



SPEC benchmarkok

- http://www.spec.org/benchmarks.html
- Standard Performance Evaluation Corp.
- Erőforrás és alkalmazás szintű benchmarkok
  - CPU
  - Alkalmazások
  - Levelező szerverek
  - Web szerverek
  - Network File System, stb.
- Benchmark: megrendelhető szolgáltatás
  - Forráskódot adnak, nekünk kell fordítani
  - Licensz díjas!

SPEC CPU2006

- CPU intenzív
- CINT2006
  - Számításigényes egész számos
- CFP2006
  - Lebegőpontos

- Példa:

cpu2006.html

CINT2006 és CFP2006 terhelésgenerátorok

- 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.xalancbmk	C++	XML Processing

- CFP2006:

SPEC CFP2006 Benchmark Descriptions.htm

SPECweb2009

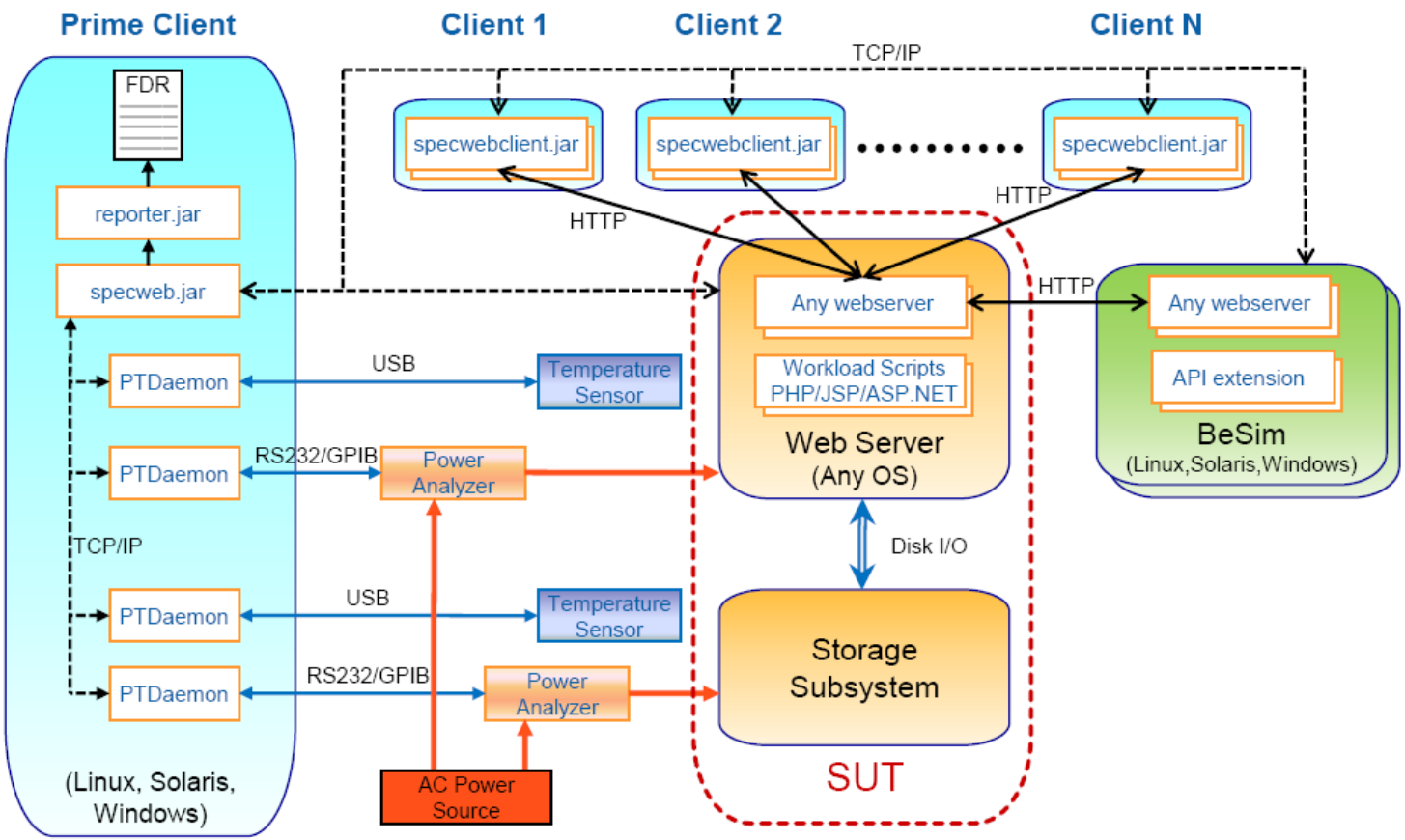
- Terhelés:
  - Biztonságos bankolás (SSL)
  - E-kereskedelem (SSL+ http letöltés pl)
  - Áteresztőképesség
  - Teljesítmény / energia metrikák

- Architektúra (ábra)
- Terhelésgenerálás (ábra)
- Példa:

web2009.html

SPECweb2009

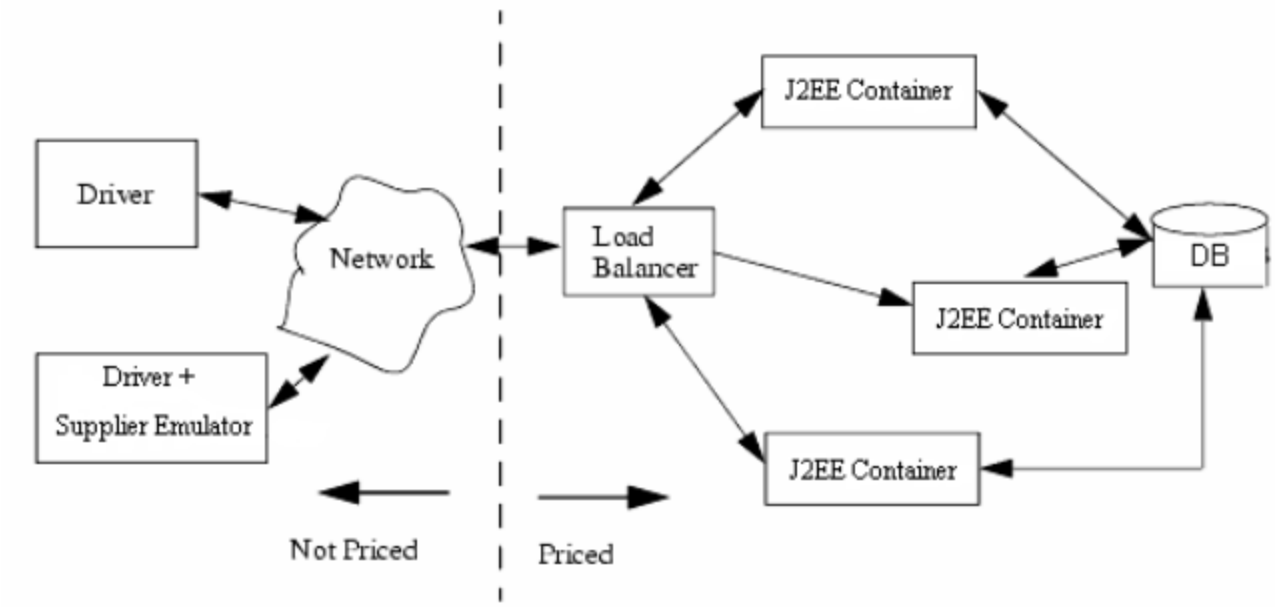
SPECweb2009 Setup (External Storage)



Copyrigth SPEC® 2009

SPECjAppServer2004

- Html oldal elemzése előadáson
- J2EE Benchmark
  - Alkalmazásszerver skálázhatósága és teljesítménye
- Architektúra (több szerveres, kliens - szerver arch.)
- Metrika:
  - Áteresztő-képesség
- Terhelés:
  - Autógyártás
  - Gyár – megrendelők
  - Autókatalógus



SPECjvm2008

- Ingyenes!
- JRE teljesytémény
  - CPU / memória
  - Kevés fájl I/O, nincs hálózati I/O

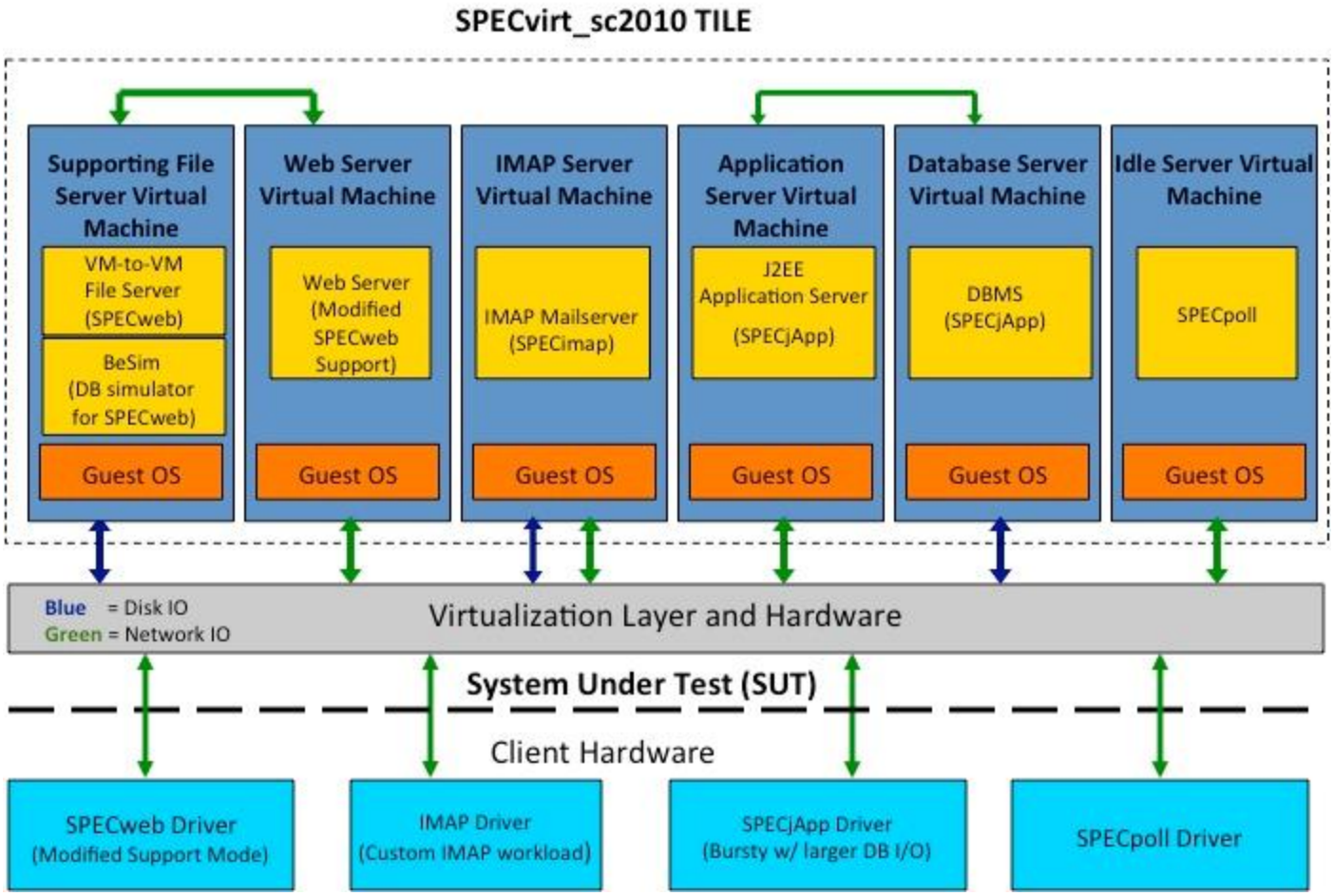
- Alkalmazások
  - Javac
  - LZW
  - Startup
  - MPEGaudio
  - Xml (transzformáció és validáció)
  - Crypto
    - Aes, rsa, signverify

SPECvirt_sc2010

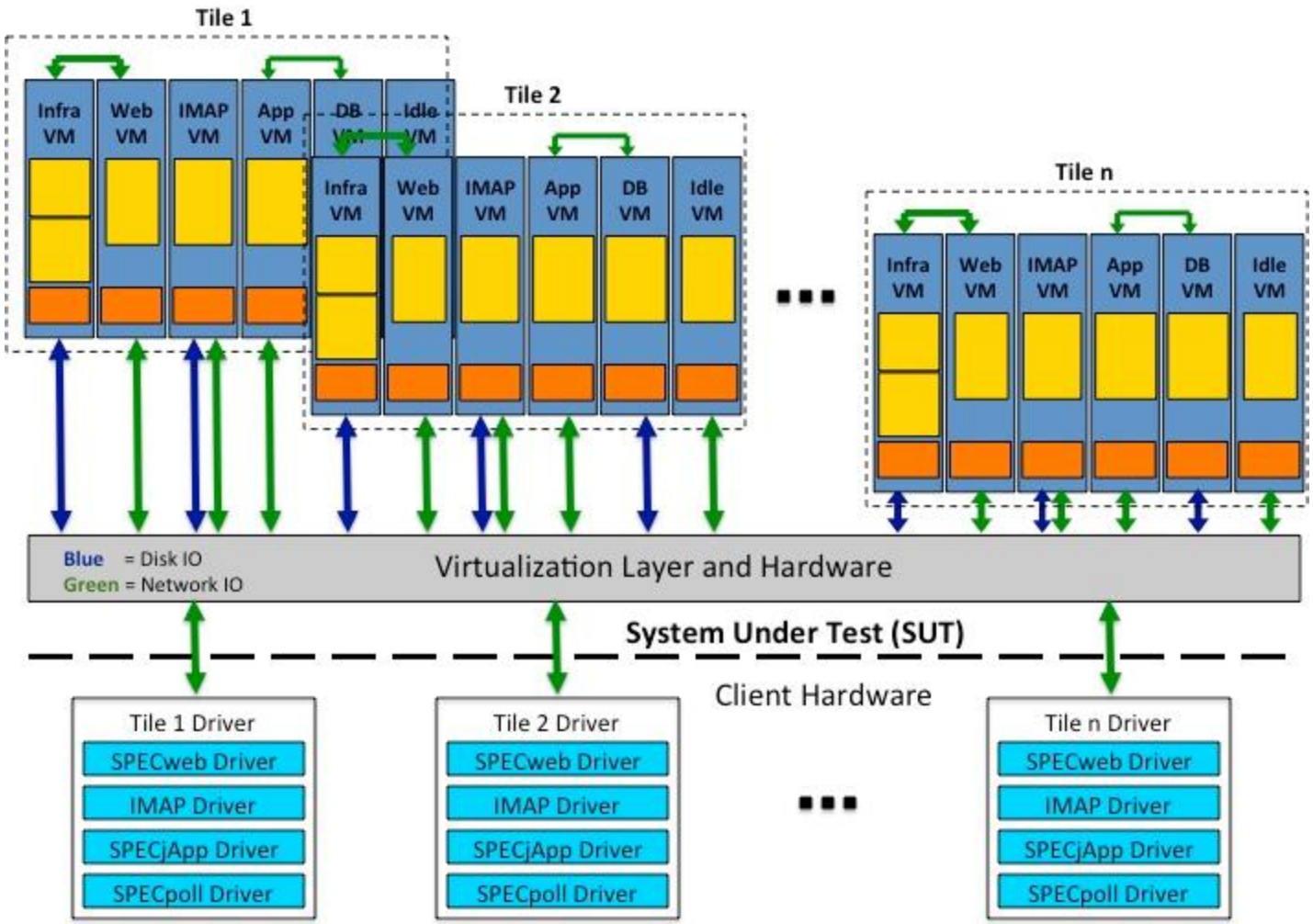
- Szerverek 18%-a volt virtualizált 2009q4-ben

- Cél: skálázhatóság / energiahatékonyság
  - Hány VM konszolidálható egy HW-re
- Architektúra (ld. ábra)
  - 1 „tile” 6 VM
  - Függőségek!
- Terhelés
  - Tile-ok számát növeli
  - Amíg QoS megvan, vagy amíg a metrikák már nem nőnek
- Metrika
  - Szubkomponensek normalizált metrikái
  - Pl kérés/s a webszerveren, alkalmazásszerver burstössége

SPECvirt_sc2010



SPECvirt_sc2010



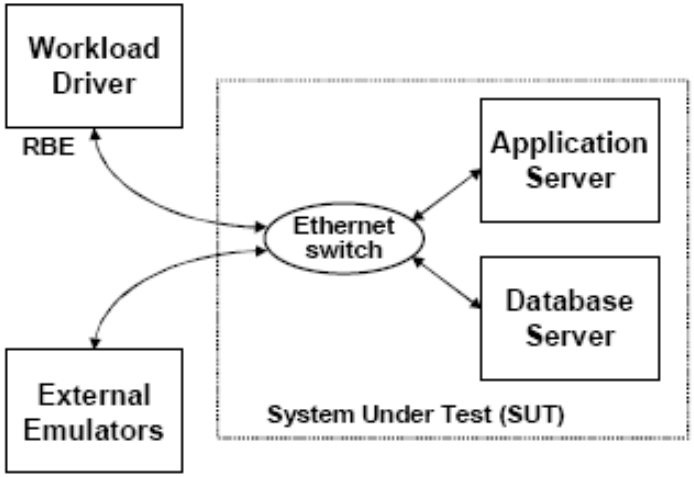
TPC benchmarkok

- http://www.tpc.org/information/benchmarks.asp
- Transaction Processing Council

- TPC-C (elektronikus kereskedelem, banki rendszer):
  - felhasználók tranzakciókat hajtanak végre
  - rendelés/lemondás, lekérdezés, stb.
  - szervereket hasonlít össze
    - HW
    - OS
    - DBMS
    - egyéb paraméterek:
      - tervezett rendelkezésre állás, elérhetőség (pl. 24/7 vs. 8/5)
  - OLTP rendszerek mérőszámai:
    - tranzakciós ráta (tpmC): 5 különböző fajta tranzakció alapján
    - ár / tranzakció ($/tpmC): fenntartási költségek / tranzakciók
  - Bővebben: info itt, pdf itt.

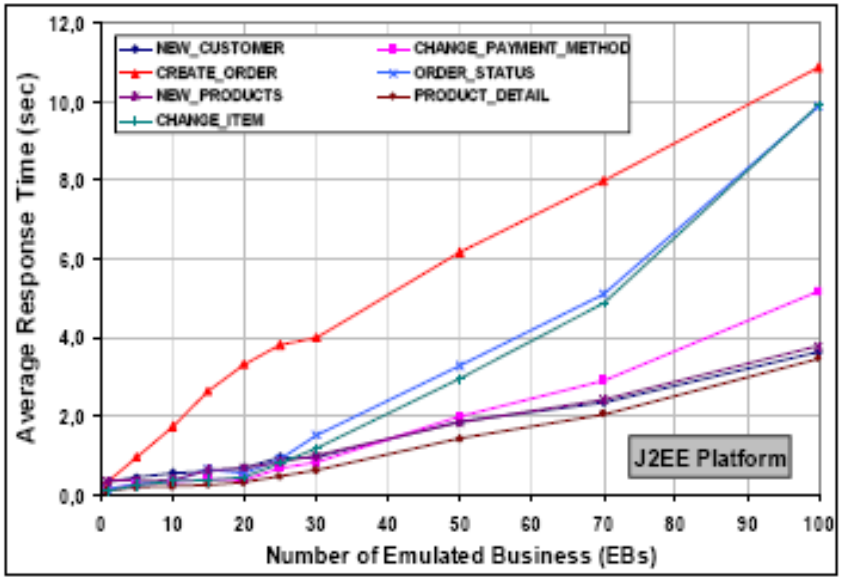
Minta rendszerek

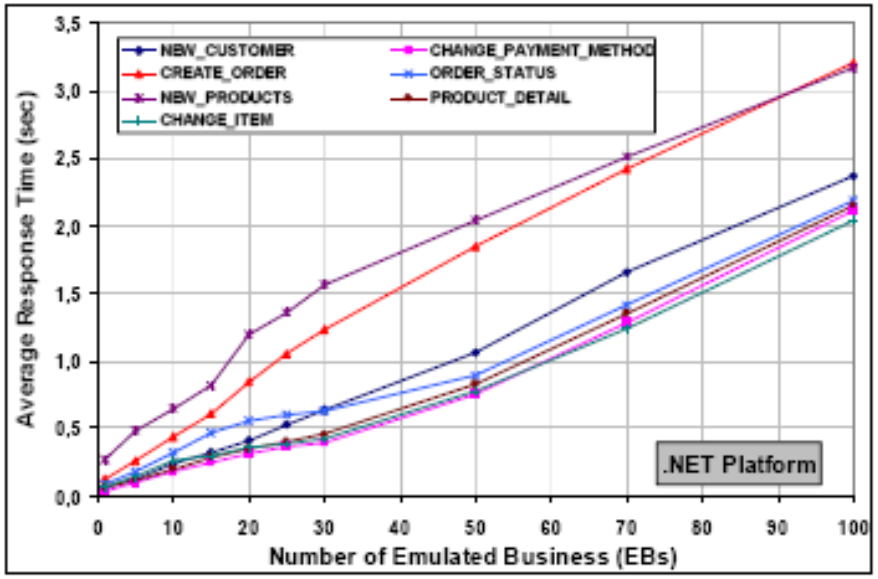
- J2EE megvalósítás
  - JBOSS + Message Queuing
- .NEt megvalósítás
  - COM+, ADO.NET
  - MSMQ
- Adatbázis közös (MSSQL)
- Skálázhatóság?
  - Válaszidő
  - Áteresztőképesség
  - Kihasználtság



Forrás: D.F. García, J. García, M. García, I. Peteira, R. García, P. Valledor: Benchmarking of Web Services platforms (An evaluation of the TCP-APP benchmark) (WEBIST 2006)

Teszt eredmények



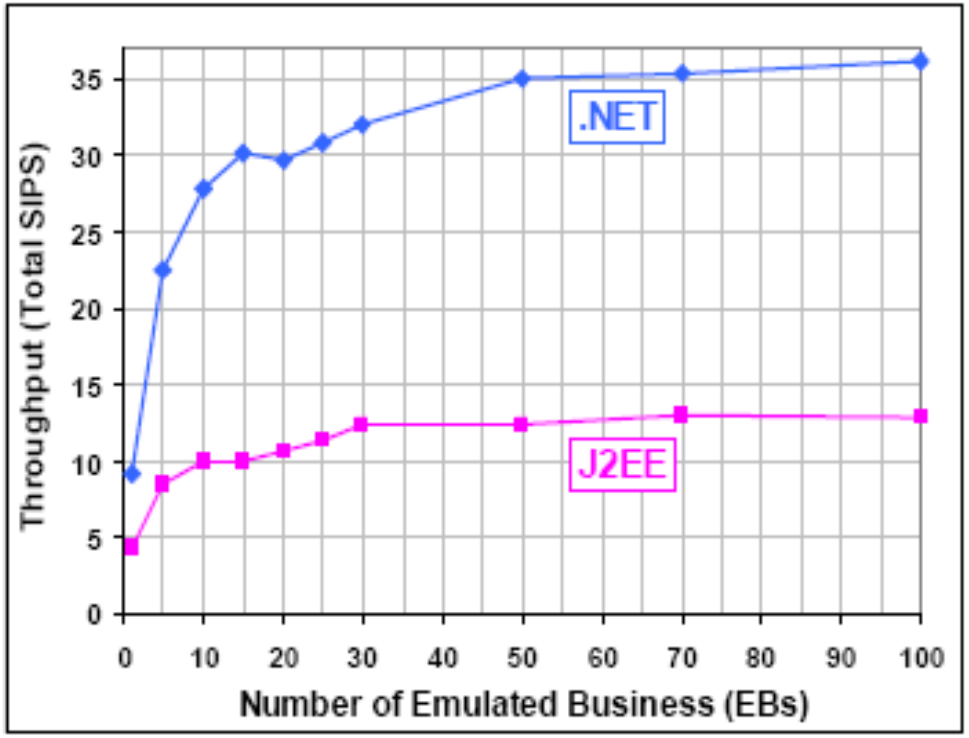


Válaszidő:

a .NET gyorsabb és jobban skálázódik

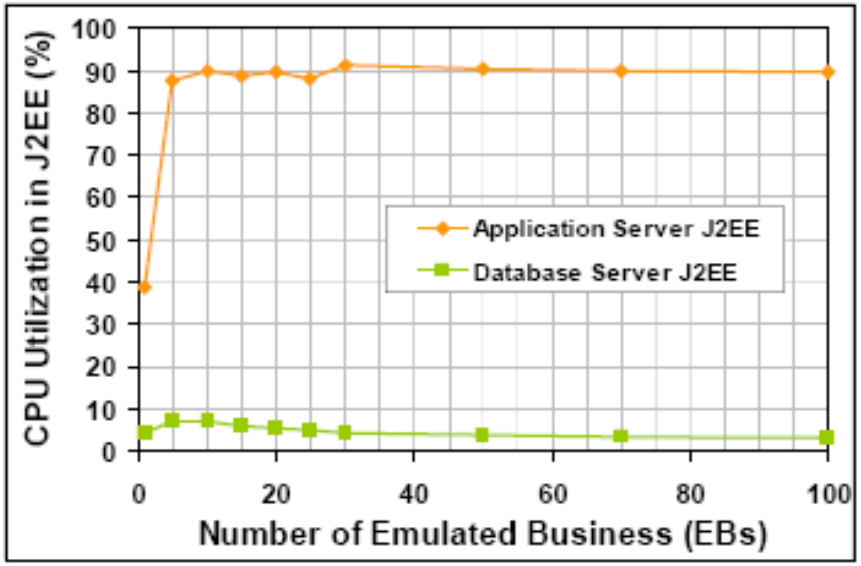
Hasonló funkciók igénylik a legtöbb időt (pl. Create order)

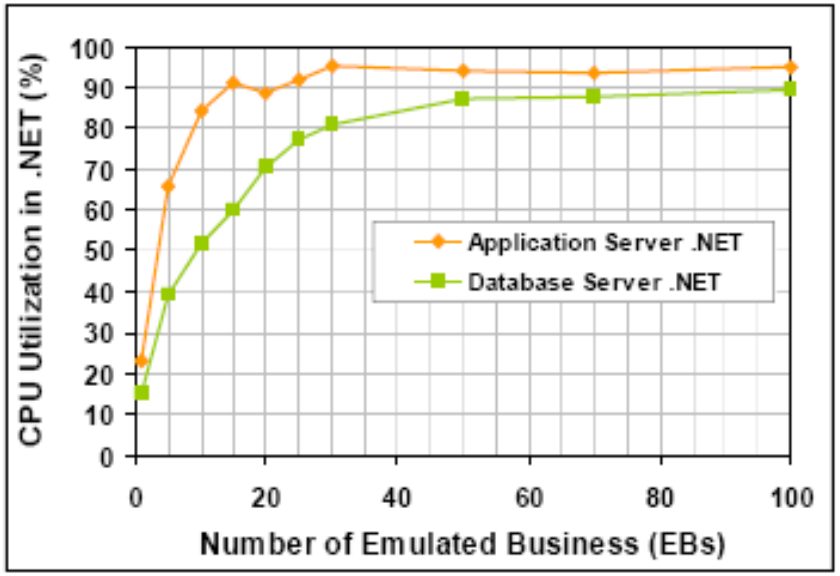
Teszt eredmények folyt.



Áteresztőképessége jobb a .NET alapú megoldásnak

Teszt eredmények folyt.





CPU kihasználtság:
a J2EE itt takarékosabb az erőforrásokkal (DB)

NEM következik belőle, hogy általában
„jobb” bármelyik megvalósítás

Pl. http://www.theserverside.com/news/thread.tss?thread_id=27341

The Autonomic Computing Benchmark

- Nagyvállalati környezet „rugalmasságának” vizsgálata (resilience)
  - Self * mérése (önmenedzselés)
  - Hibainjektálással
  - A rendszer robosztusságát és automatizáltságát vizsgálja

- Cél: kevés metrika
  - Rendszer evolúciójának vizsgálata
  - Különböző rsz-ek összehasonlítása

Architektúra

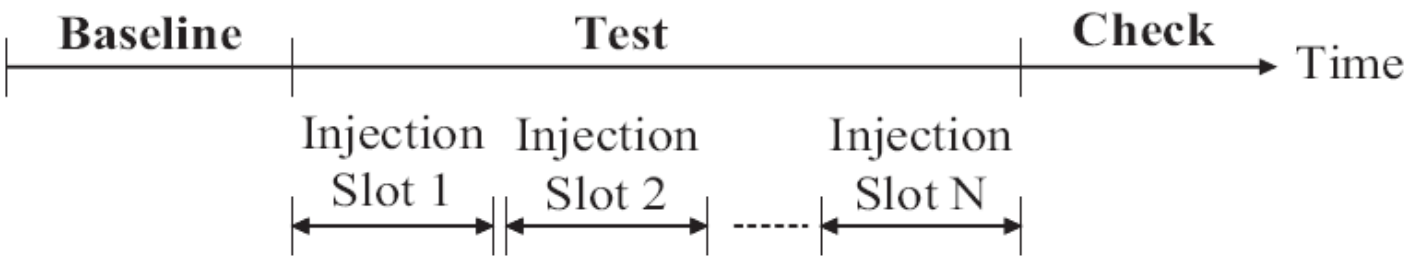
- Elvben bármi lehet
- Példa: SPECjAppServer2004 architektúra / benchmark
  - Webszerver
  - Alkalmazásszerver
  - Adatbázis
  - Üzenetküldő szerver

Metrikák

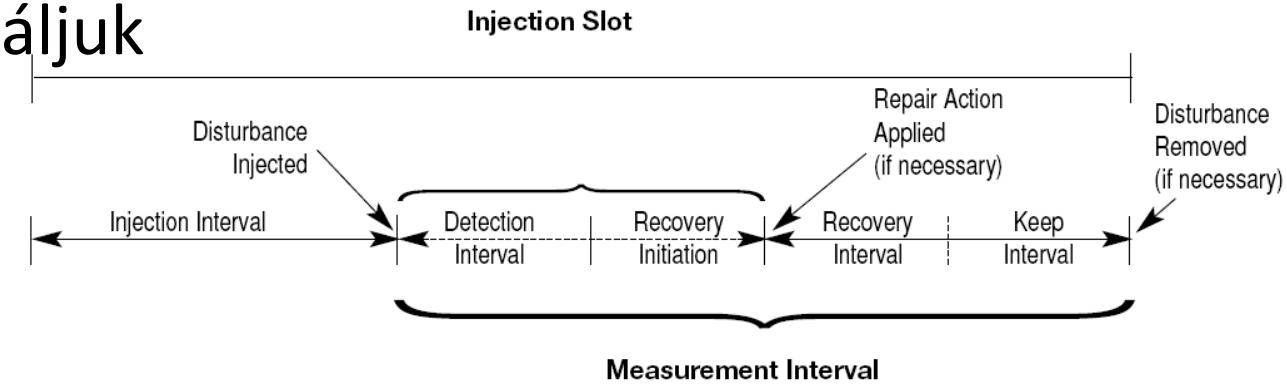
- Áteresztőképesség index
  - Zavar hatása az áteresztőképességre
- Fejlettség index (ld. később)
  - Emberi interakciók mennyiség
    - hibadetektálás,
    - hibaanalízis,
    - helyreállás során

Mechanizmus

- 3 fázis:
  - Alap fázis
  - Tesztfázis
  - Ellenőrzés
    - Konzisztencia ellenőrzése

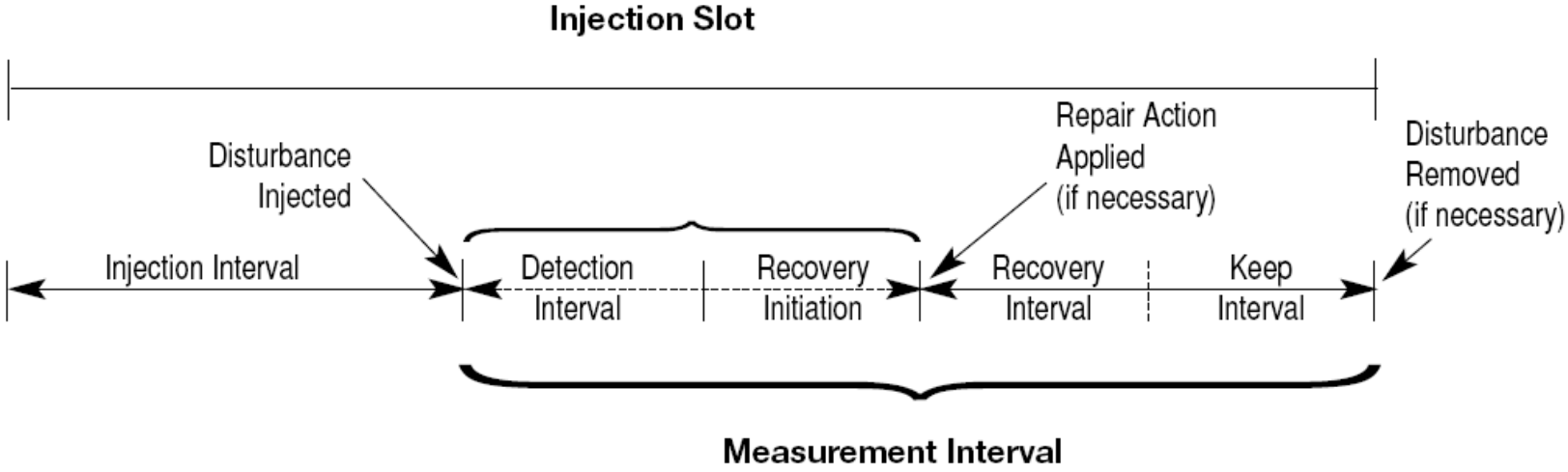


- Tesztfázist slotokra osztják
  - Egymás utáni slotokban hibákat injektálunk és azok hatását vizsgáljuk



Mechanizmus – 1 slot

- Állandósult állapotba kerül a rendszer
- Hibainjektálás
- Hibadetektálás
  - Automatikus
  - Bizonyos idő elteltével (emberi interakció szimulálása)
- Helyreállítás
- Újraindítás
- Rendszer futtatása
  - Degradálódott-e a szolgáltatás?



Hibafajták

- Hibák (példa):
  - Váratlan leállás (hw, os, sw)
  - Erőforrás versenyhelyzet (CPU, mem)
  - Adatvesztés (file, DBMS)
  - Terhelésváltozás („karácsonyi roham”)
  - Sikertelen újraindítás detektálása

Metrikák (folyt)

- Áteresztőképesség index: P_i / P_base
  - Figyelem! Nem rendelkezésre állás.
- Fejlettség index
  - Kifejezi mennyire automatizált a rendszer
  - Mennyi emberi beavatkozás kell
  - A SUT egy lista alapján pontokat kap
  - Nemlineáris skála
  - Átlagolunk, normalizáljuk
  - Index 0—1 között
    - 0 nincs automatizmus
    - 1 nem kell emberi közbeavatkozás

Példa

- Esettanulmány az órán
- Ábrák elemzése a cikkből

Nehézségek

- Zavar megtervezése
  - Zavarok katalógusának összegyűjtése
  - Szimuláció?
- Eredmények összehasonlítása
  - Terhelés
  - Skálázódás
    - a metrika nem skálázódik feltétlenül együtt a rendszerméretttel
  - Ár
    - Robosztus és automatizált rendszer nagyon költséges
- Benchmark során: hum. interakció becslés!