

Evaluarea performanțelor sistemelor informaticice

Curs 1

1

- florentina.suter@yahoo.com
- <http://groups.yahoo.com/group/epsi12/>
- <http://catless.ncl.ac.uk/Risks>
- <http://www5.in.tum.de/~huckle/bugse.html>

2



Bibliografie:

- Daniel A. Menascé, Virgilio A.F. Almeida, Lawrence W. Dowdy (2004) *Performance by Design: Computer Capacity Planning by Example*, Pearson Education, Inc.
- Jean-Yves Le Boudec (2009) *Performance Evaluation of Computer and Communication Systems*.
- Ion Văduva (2003) *Fiabilitatea programelor*, București, Editura Universității din București.

3



Definiții

- Sistem = un ansamblu format din una sau mai multe componente care interacționează pentru realizarea unui anumit scop
- Sistem informatic = procesor, memorie, rețea de calculatoare, server web, baze de date, alte softuri
- Evaluarea performanțelor = măsurarea calității serviciului furnizat de un sistem informatic (Quality of Service - QoS)

4



Definiții

- Elemente importante in EP:
 - **Încărcătura (load)** = cantitatea de cereri și natura acestora (web server – nr. de cereri pe secundă, cereri pentru obiecte folosite frecvent sau nu)
 - **Metrica** = este o cantitate măsurabilă care cuantifică ceea ce vrem să măsurăm (număr de task-uri finalizate în unitatea de timp, timp de răspuns, energia electrică consumată în unitatea de timp)
 - **Scopul** = obținerea unui echilibru cost-performanță, alegerea celei mai bune metode de proiectare, compararea a două sau mai multe sisteme.
 - **Factori** = elemente ale sistemului sau ale încărcării care pot afecta performanța

5



Metode de evaluare

- **Măsurarea sistemului real** = este dificil de făcut fără perturbarea sistemului. Dacă este făcută de sistemul însuși trebuie analizat impactul pe care îl are. Nu este întotdeauna posibilă
- **Simularea (Discrete Event Simulation)** = un model simplificat al sistemului și al încărcării pe care acesta o suportă, implementat într-un program. Performanța se măsoară ca în cazul sistemului real. Este metoda cea mai răspândită.
- **Metode analitice** = un model matematic al sistemului, model care este analizat numeric. Este mai rapidă și uneori mai precisă decât simularea, dar, în anumite cazuri sunt necesare ipoteze prea restrictive. Sunt folosite adesea în faza de proiectare.

6



Probleme ale sistemelor

- **Zone de blocaj (bottlenecks)** = performanța este puternic influențată de comportamentul celor mai slabe componente.
- **Căderi din cauza aglomerării** = aglomerarea apare atunci când capacitatea sistemului este depășită de încărcătură
- **Efectul secundar al competiției** = furnizarea mai multor resurse determină o performanță mai slabă serviciului acordat anumitor utilizatori
- **Căderi latente din cauza aglomerării** = înlăturarea unei zone de blocaj poate să ducă la o cădere

7



Necesitatea evaluării performanțelor

- Aplicații pentru urgențe
- Aplicații financiare
- Securitate națională, aplicații militare
- Sisteme de comunicații, servicii pentru clienți

8



Metriци QoS

- Timp de răspuns
- Ieşire
- Accesibilitate
- Fiabilitate
- Securitate
- Scalabilitate
- Extensibilitate

9



Analiza timpului de răspuns

Browser Time		Network Time			E-commerce Server Time		
Processing	I/O	Browser to ISP Time	Internet Time	ISP to Server Time	Processing	I/O	Networking

..... CONGESTION

Exista doua categorii de timpi, categorii regasite in toate cele trei componente:

- Timpul de servire (nu depinde de încărcare)
- Timp petrecut in sistem din cauza aglomerarii (depinde de încărcare) – reprezintă timpul de aşteptare la deferite resurse: procesoare, discuri, reţea.

10



Ieșirea

- Măsurată în unități de lucru (operații) finalizate într-un anumit interval de timp. Este o rată. De exemplu, se poate măsura în:
 - Operații de I/O pe secundă
 - Descărcări de pagini pe secundă
 - Cereri HTTP pe secundă
 - Operații pe secundă
 - Tranzacții pe secundă (sisteme OLTP – Online Transaction Processing)
 - Pachete pe secundă, MB transferați pe secundă

11

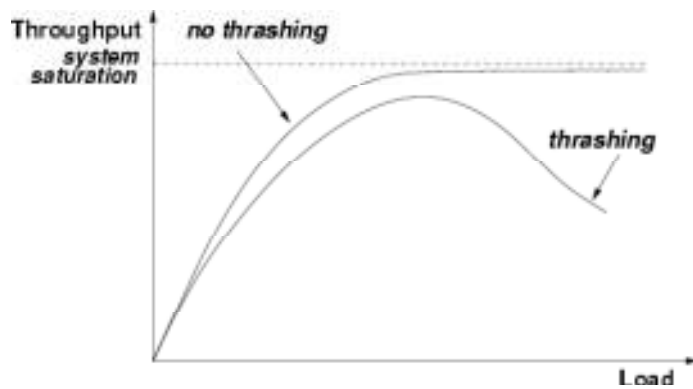


Ieșire - Exemplu

- O operație de I/O pentru un disc al unui sistem OLTP durează în medie 10 milisecunde
 - Care este ieșirea maximă de pe disc? – Dacă discul ar fi tot timpul ocupat, adică utilizarea sa este de 100% și execută operații de I/O în continuu → o rată de $1/0.01=100$ de tranzacții pe secundă
 - Care este ieșirea de pe disc, dacă acesta primește cereri de I/O la o rată de 80 de cereri pe secundă?
$$\text{ieșire} = \min\{\text{capacitate_serviciu}, \text{cereri}\}$$

12

Ieșire - Exemplu



Ieșirea în cazul în care în sistem există un fenomen de trashing, poate să scadă foarte mult chiar dacă încărcarea crește.

13

Ieșire - Exemplu

- Trashing (în general) = când sistemul petrece mai mult timp cu alocarea resurselor decât cu prelucrarea lor.
- Acest fenomen poate să apară în sistemele care folosesc tehnica memoriei virtuale, când sistemul petrece foarte mult timp cu transferarea blocurilor de memorie virtuală între memoria fizică și disc.

14

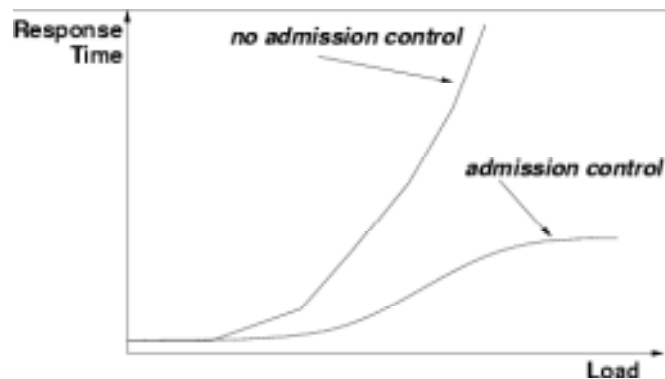
Accesibilitate (Availability)

- Frațiunea de timp în care sistemul funcționează și poate fi accesat
 - De exemplu un magazin online poate pierde clienți pentru că site-ul lui nu poate fi accesat
 - Un sistem cu 99.99% accesibilitate, în 30 de zile este inaccesibil $(1-0.9999) \times 30 \times 24 \times 60 = 4.32$ minute. Accesibilitate slabă pentru anumite sisteme (aparare, urgente).
 - Motivele lipsei de accesibilitate pot fi caderile sau încărcările prea mari.

15

Accesibilitate (Availability)

Controlul numărului de clienți (control de admitere) în sistem determină lipsa de accesibilitate pentru unii clienți, dar garantează calitatea serviciului pentru alții.



16



Fiabilitate

- Probabilitatea ca un sistem să funcționeze fără căderi într-un anumit mediu, pentru un anumit interval de timp dat.

17



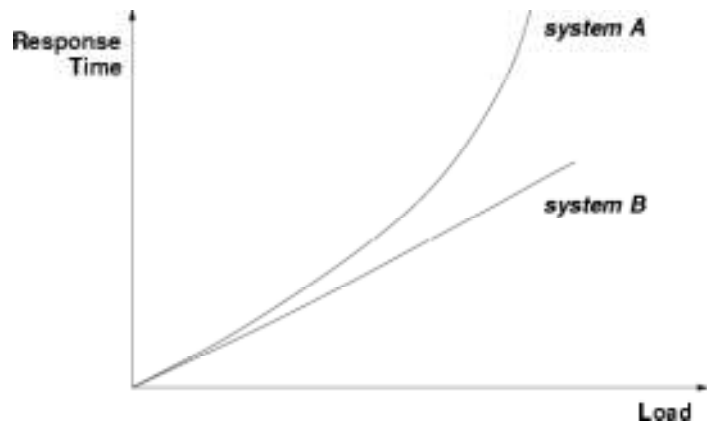
Securitatea

- O combinație de trei atribute de bază:
 - Confidențialitate: li se permite accesul la informații relevante doar persoanelor autorizate
 - Integritatea datelor: informațiile nu pot fi modificate de utilizatori neautorizați
 - Ne-repudiare: cel care trimite un mesaj nu poate nega acest lucru.
- Se implementează prin mecanisme de autentificare bazate pe una sau mai multe forme de criptare.

18

Scalabilitate

Scăderea performanțelor sistemului nu este abruptă atunci când încărcarea crește.



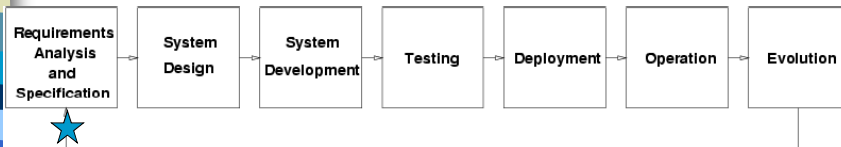
19

Extensibilitate

- Capacitatea unui sistem de a evolua și de a îndeplini specificațiile funcționale și de performanță si in cazul extinderii lui. Extinderea sistemului nu determină scăderea performanțelor.

20

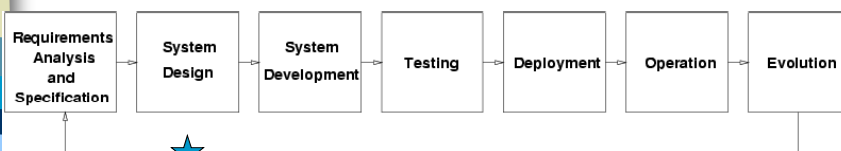
Ciclul de viață al unui sistem informatic



- Două tipuri de cerințe pentru sistem:
 - Cerințe funcționale: ce trebuie să facă sistemul și pe ce tipuri de platforme
 - Cerințe nefuncționale: cât de bine trebuie sistemul să-și îndeplinească funcțiile. În multe cazuri cerințele nefuncționale sunt neglijate sau luate în considerare doar la testarea sistemului.
- Sunt stabilite SLA (Service Level Agreements) – stabilirea lor este importantă pentru evaluarea performanțelor.
- O analiză cantitativă și calitativă a încărcării sistemului

21

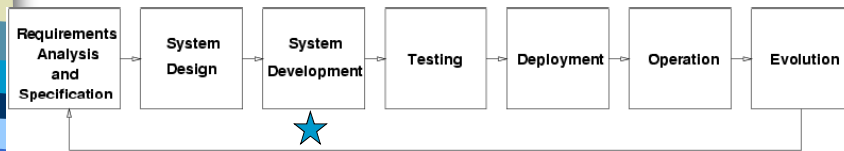
Ciclul de viață al unui sistem informatic



- Se stabilește modul în care se vor realiza cerințele:
 - Arhitectura sistemului
 - Descompunerea sistemului în părți componente
 - Sunt proiectate structurile de date, fișierele și bazele de date importante
 - Sunt specificate interfețele între diferitele componente.
- Pentru evaluarea performanțelor: luarea în considerare a cerințelor nefuncționale pentru diverse componente, mai ales pentru cele preluate din alte sisteme (motoare de căutare, mesaje middleware). Aici pot să scadă performanțele sistemului

22

Ciclul de viață al unui sistem informatic



Componentele sunt implementate

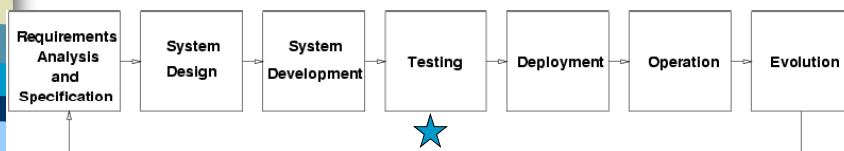
- Unele sunt noi
- Unele sunt refolosite
- Unele sunt adaptate

Componentele sunt interconectate pentru a forma un sistem

Pentru evaluarea performanțelor: decizii de implementare transmise din faza de proiectare, adunarea datelor necesare în faza de testare.

23

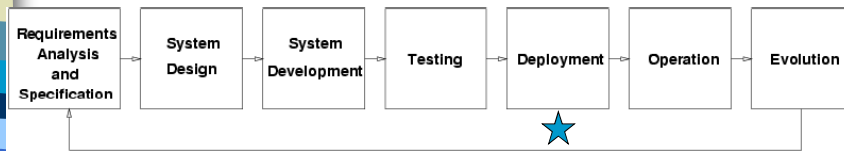
Ciclul de viață al unui sistem informatic



- În paralel cu dezvoltarea sistemului, pe măsură ce componentele devin disponibile, are loc testarea pe unități
- Sunt desfășurate teste pe subsisteme, apoi pe întreg sistemul atunci când acesta este finalizat
- Adesea se consumă mai mult timp cu testarea cerințelor funcționale, decât cu testarea celor nefuncționale
- Nu este posibilă testarea tuturor scenariilor din cauza constrângerilor de timp și buget.

24

Ciclul de viață al unui sistem informatic

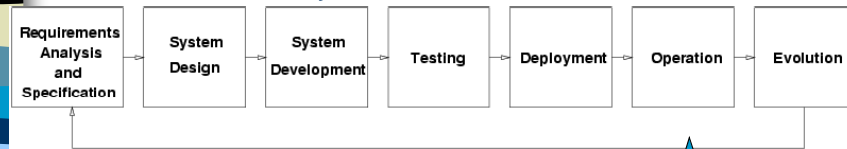


Parametrii de configurare trebuie să fie stabiliți pentru a îndeplini SLA-urile

- De exemplu parametrii pentru TCP, mărimea bazei de date, numărul maxim de fire de execuție (thread-uri) se pot seta pentru o cât mai buna performanță.
- Predictii de performanta pentru diferite scenarii de configurare

25

Ciclul de viață al unui sistem informatic

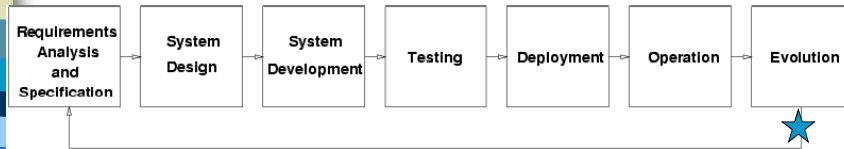


Monitorizarea constantă a anumitor metrici pentru a verifica dacă sistemul îndeplinește cerințele

- Încărcarea (perioadele de vârf, tipare neobișnuite)
 - Metrici externe (care sunt percepute de utilizator): timp de raspuns, iesire
 - Metrici interne (care ajută în detectarea blocajelor și la reglarea sistemului): procentul de utilizare a resurselor
 - Accesibilitate (externă și internă)
- Poate fi necesară o ajustare a parametrilor de configurare

26

Ciclul de viață al unui sistem informatic



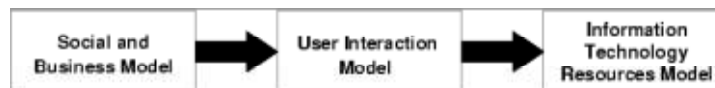
Uneori sistemele trebuie să evolueze din cauza multor tipuri de factori:

- pentru a respecta anumite legi noi
- pentru a realiza noi funcții

Ce impact are evoluția asupra QoS? Se fac modele predictive înainte de a realiza o modificare a sistemului

27

Exemplu de model referință (reference model) pentru un sistem informatic



Se dorește crearea unei aplicații bancare. Trebuie ținut cont de trei categorii de submodele care vor fi implicate:

- modelul business și social
- modelul utilizatorului
- modelul resurselor IT disponibile

28

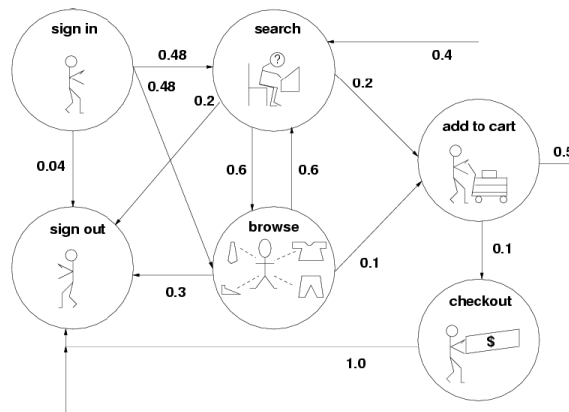
Exemplu de model referință (reference model) pentru un sistem informatic

- Modelul business: număr de filiale, numărul și locația ATM-urilor, numărul conturilor de orice tip, planurile de evoluție a business-ului (eventuale fuzionări)
- Modelul social: reglementări de confidențialitate, reglementări de accesibilitate

29

Exemplu de model referință (reference model) pentru un sistem informatic

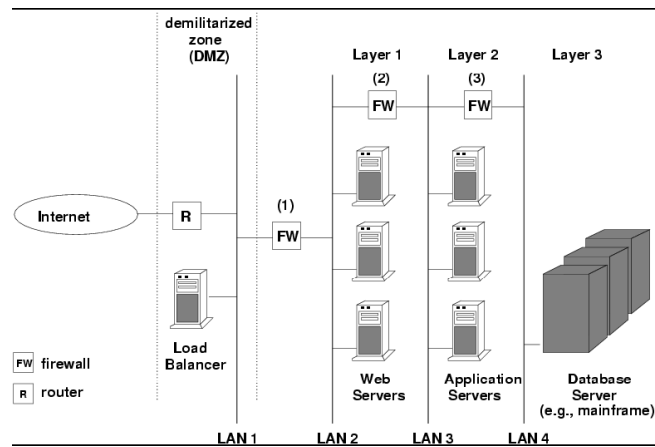
Modelul utilizatorului: graful model pentru comportamentul clientului



30

Exemplu de model referință (reference model) pentru un sistem informatic

Infrastructura IT: Exemplu



31

Exercitiu

Un sistem de vânzări online a înregistrat în ultimele două zile perioade de timp în care sistemul a fost căzut și s-a obținut următorul tabel:

Zi	Momentul căderii	Durata căderii (min)
1	1:25 AM	12
1	7:01 AM	1
1	8:31 PM	3
2	9:35 AM	10
2	9:12 PM	6

Care este accesibilitatea site-ului în cele două zile luate împreună? Dar separat? Cum diferă accesibilitatea în cele două zile? Când lucrează mai bine site-ul?

32