

Softwarové architektury

Metriky pro Enterprise Aplikace

2008-06

David Toth

- **(Response time)** *Čas odpovědi (T_1)*
- **(Responsiveness)** *Doba reakce (T_2)*
- **(Latency)** *Zpoždění (L)* **(Efficiency)** *Efektivita*
- **(Throughput)** *Propust'* **(Performance)** *Výkon*
- **(Load)** *Zátěž* **(Degradation)** *Degradace*
- **(Scalability)** *Škálovatelnost* **(Capacity)** *Kapacita*
 - **(Vertical Scalability, scaling up)** *Vertikální škálovatelnost*
 - **(Horizontal Scalability, scaling out)** *Horizontální škálování*

Metriky, parametry softwarových systémů

- Uvedené metriky nejsou přímo ovlivněny návrhovými rozhodnutími
- Nýbrž jejich hodnoty jsou závislé na struktuře celého systému
 - Struktura jak statické, tak dynamické
- Závisí na architektonických rozhodnutích

Response Time

- Doba odpovědi (t_1)
- množství času, který uběhl od okamžiku vzniku žádosti po její (úplné) zpracování / vyřízení
- čas v řádech
 - mikro-sec, mili-sec, vteřiny, minuty, hodiny, dny

Responsivness

- Doba reakce / doba odezvy (t_2)
- čas, který uběhne, než systém projeví reakci na vzniklou událost
 - čím větší je t_1 , tím menší má být rozdíl $t_1 - t_2$
 - snažíme se, aby t_2 byla konstanta
 - => snižování frustrace uživatelů

Zpoždění

- Latence L
 - minimální doba potřebná pro odpověď aplikace
 - důležitá u vzdálených systémů
 - desktopová aplikace by měla mít $L \rightarrow 0$
 - Aplikační vývojář může udělat jen málo pro to, aby L změnil
 - Proto je tak důležité minimalizovat počet vzdálených volání.

Throughput

- Propust'
 - kolik práce lze dělat současně v jednom časovém intervalu
 - při kopírování souboru kB/s
 - Enterprise aplikace: metrika tps
 - (transaction per second)
 - Tento údaj je závislý na typu, rozsahu a komplexitě transakcí

Performance

- Výkonnost
 - Je dána propustí (KB/s)
 - Časem zpracování (s)

Load

- Zátěž
 - počet současně připojených uživatelů
 - kontext pro jiné parametry

Load sensitivity (Degradation)

- Citlivost na zátěž
 - jak se mění (derivace) doba reakce v pohledu na počet připojených uživatelů
- Degradace
 - systém A degraduje více než systém B, jestliže po zvýšení zátěže na oba systémy systém A reaguje citlivěji než B

Capacity

- Kapacita
 - maximální efektivní
 - propust' či
 - zátěž
 - bod
 - při kterém klesne výkonnost na 0,
 - či je menší než stanovený práh

Scalability

- Škálovatelnost
 - je měřítkem toho, jak se s přidávajícím (hardwarem) zdroji mění výkon
- Škálovatelný systéme
 - určitému zvýšení počtu (hardwarů) zdrojů, odpovídá lineární zvýšení výkonu
 - (2x více HW serverů
=> dvojnásobně lepší propustnost systému)

Vertical scalability, scaling up

- Vertikální škálovatelnost
 - Více síly na jeden stroj (počítač, typicky server)

Horizontal scalability, scaling out

- Horizontální škálovatelnost
 - Přidání více serverů

Konec

- Zdroje:
 - Fowler, M., Patterns of Enterprise Application Architecture, A-W, 2003